Ann. Inst. Statist. Math.
Vol. 49, No. 4, 615–644 (1997)

ON MAD AND COMEDIANS

MICHAEL FALK

Mathematisch-Geographische Fakultät, Katholische Universität Eichstätt, Ostenstrasse 26–28, D-85071 Eichstätt, Germany

(Received March 8, 1996; revised October 15, 1996)

Abstract. A popular robust measure of dispersion of a random variable (rv) X is the *median absolute deviation from the median* $\text{med}(|X - \text{med}(X)|)$, MAD for short, which is based on the median $\text{med}(X)$ of X. By choosing $Y = X$, the MAD turns out to be a special case of the *comedian* $\text{med}((X - \text{med}(X))(Y - \text{med}(Y)))$, which is a robust measure of covariance between rvs X and Y. We investigate the comedian in detail, in particular in the normal case, and establish strong consistency and asymptotic normality of empirical counterparts. This leads to a robust competitor of the coefficient of correlation as an asymptotic level-α-statistic for testing independence of X and Y. An example shows the weird fact that knowledge of the population $\text{med}(X)$ does not necessarily pay (in the sense of asymptotic relative efficiency) when estimating the MAD.

Key words and phrases: Median absolute deviation from the median, robust measure of correlation, comedian, breakdown point, covariance, correlation coefficient, bivariate normal vectors, strong consistency, asymptotic normality.

1. Introduction

A widely accepted robust measure of location of a random variable (rv) X is the *median* $\text{med}(X)$, which satisfies the inequalities $P\{X \le \text{med}(X)\} \ge 1/2 \le P\{X \ge \text{med}(X)\}$. The median is in general not unique, but $\text{med}(X) = F^{-1}(1/2)$ is a possible choice, where

$$F^{-1}(q) := \inf\{t \in \mathbb{R} : F(t) \ge q\}, \qquad q \in (0,1),$$

denotes the generalized inverse of the distribution function (cdf) $F(t) := P\{X \le t\}$, $t \in \mathbb{R}$, of X. Consequently, we say $\text{med}(X) =$ (or $\le, \ge$) $b \in \mathbb{R}$, if there exists at least one median of X, which satisfies this relation.

In case of a symmetric X, the median of X is the center of symmetry and coincides in this case with the expectation $E(X)$ of X, if it exists; the normal case is a popular example.

A robust alternative to the standard deviation $\text{VAR}(X)^{1/2} := E((X - E(X))^2)^{1/2}$ as a measure of scale is the *median absolute deviation from the median*

(MAD), defined by

$$\text{MAD}(X) := \text{med}(|X - \text{med}(X)|). \tag{1.1}$$

As Huber ((1981), p. 107) points out, the "MAD has emerged as the single most useful ancillary estimate of scale". Hampel (1974), who called MAD the *median deviation*, showed it to be an M-estimate of scale, which facilitates calculation of the influence curve (see e.g. Huber ((1981), p. 137)). In particular for a normal rv X with mean μ and variance $\text{VAR}(X) = \sigma^2$, we obtain from Lemma 1.1 below the following linear correspondence between MAD(X) and the standard deviation of X:

$$\text{MAD}(X) = \sigma\,\text{med}(|(X - \mu)/\sigma|) = \sigma\Phi^{-1}(3/4),$$

where Φ denotes the standard normal cdf. Our approach towards robustness of statistical functionals relies on the data based concept of *breakdown points* (Rousseeuw and Leroy (1987)) i.e., the least proportion of observations in a data set that must be moved in order to let the value of the statistical functional under consideration break down i.e., tend to infinity.

The MAD has the highest possible breakdown point, since roughly one half of the data have to be pushed to infinity to let the median follow. While the MAD or the (in case of a symmetric rv X equivalent) *interquartile range* $IQR := F^{-1}(3/4) - F^{-1}(1/4)$ as a further robust measure of dispersion underlying a box-plot (Tukey (1977)) have become quite popular in recent years and are now built-in procedures in statistical software packages, a median based and therefore robust alternative to the covariance $\text{COV}(X, Y) = E((X - E(X))(Y - E(Y)))$ as a measure of covariance between rvs X, Y is not ready at hand, though it plays a crucial role in various topics of statistics.

The covariance is the center of interest in correlation analysis, it is essentially the (simple) regression coefficient in *regression analysis* and it is of cause the component of the *covariance matrix* $\boldsymbol{S}$ of a random vector $\boldsymbol{X}$. The covariance matrix is basic to different multidimensional techniques such as *principal component analysis*, where the eigenvectors of $\boldsymbol{S}$ are the (orthogonal) directions of largest variation or in *factor analysis* based on principal components. The generalized squared distance $(\boldsymbol{x} - \boldsymbol{y})\boldsymbol{S}^{-1}(\boldsymbol{x} - \boldsymbol{y})$ between vectors $\boldsymbol{x}$, $\boldsymbol{y}$ is known as their *Mahalanobis distance*, typically applied in *discriminant analysis* and *cluster analysis*. The linear transformation $\boldsymbol{Y} := \boldsymbol{S}^{-1/2}\boldsymbol{X}$ is known as *data sphering*, resulting in a random vector $\boldsymbol{Y}$ whose covariance matrix is the unity matrix. Here $\boldsymbol{S}^{-1/2}$ denotes the usual symmetric root of $\boldsymbol{S}^{-1/2}$.

The common empirical counterpart of $\text{COV}(X, Y)$, based on n independent copies $(X_1, Y_1), \ldots, (X_n, Y_n)$ of the bivariate random vector (X, Y), is the empirical covariance

$$\widehat{\text{COV}}_n(X, Y) = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)(Y_i - \bar{Y}_n),$$

where $\bar{X}_n := n^{-1} \sum_{i=1}^{n} X_i$ and $\bar{Y}_n := n^{-1} \sum_{i=1}^{n} Y_i$ are the sample means. The empirical covariance (matrix) is obviously highly sensitive to extreme observations as it can seriously be affected by only one extreme observation among

$(X_1, Y_1), \ldots, (X_n, Y_n)$ i.e., its breakdown point is $1/n$. Nevertheless it is typically in use as a black box in statistical software packages. An exposition of robust covariances is given by Huber ((1981), Chapter 8).

In this paper we propose an alternative measure of dependence between rvs X, Y, which we will call *comedian* of X and Y. It generalizes the MAD since it equals MAD^2 in case $X = Y$. In this sense, MAD and comedian parallel σ and covariance which satisfy $\sigma^2 = \text{COV}(X,Y)$ in case $X = Y$. As the comedian is actually a median, it also has the highest possible breakdown point.

In the following we compile several basic results for general rvs X, Y, thus evaluating parallel and different properties of $\text{COV}(X,Y)$ and the comedian. In the next section we investigate the particular case where (X,Y) is bivariate normal. We will see that there is a (continuously differentiable) one-to-one correspondence between the coefficient of correlation ϱ of normal vectors (X,Y) and the comedian based *correlation median*. This observation offers a way to define a robust estimator of ϱ. In Sections 3 and 4 we establish limit results for estimates of the comedian and show by an example the weird fact that knowledge of the population $\text{med}(X)$ does not necessarily pay (in the sense of asymptotic relative efficiency) when $\text{MAD}(X)$ has to be estimated. By $\mathcal{L}(X)$ we denote in the following the distribution of X. Our first lemma is obvious but nevertheless quite useful.

LEMMA 1.1. *For real numbers $a, b \in \mathbb{R}$*

$$\text{med}(aX + b) = a\,\text{med}(X) + b.$$

Note that the preceding equality includes a negative a. By

$$\text{COM}(X,Y) := \text{med}((X - \text{med}(X))(Y - \text{med}(Y))) \tag{1.2}$$

we denote the comedian of X and Y. The comedian parallels $\text{COV}(X,Y)$ and is therefore a measure of covariance. But $\text{COV}(X,Y)$ requires the existence of the first two moments of X and Y, whereas $\text{COM}(X,Y)$ always exists. Take for example X, Y independent, each following a standard Cauchy distribution. Then $\text{COV}(X,Y)$ does not exist but their comedian is zero. This is immediate from the following lemma.

LEMMA 1.2. *If X and Y are independent, then* $\text{COM}(X,Y) = 0$.

PROOF. The assertion follows from Fubini's theorem:

$$\begin{aligned}
&P\{(X - \text{med}(X))(Y - \text{med}(Y)) \underset{(\le)}{\ge} 0\} \\
&= \int_{(0,\infty)} P\{X - \text{med}(X) \underset{(\le)}{\ge} 0\}\mathcal{L}(Y - \text{med}(Y))(dy) \\
&\quad + \int_{(-\infty,0)} P\{X - \text{med}(X) \underset{(\le)}{\ge} 0\}\mathcal{L}(Y - \text{med}(Y))(dy) \\
&\quad + P\{Y = \text{med}(Y)\} \\
&\ge 1/2
\end{aligned}$$

i.e., $\text{COM}(X, Y) = 0$. □

The proof of the following lemma uses the fact that $\text{med}(X) \le \text{med}(Y)$ if $P\{Y \le t\} \le P\{X \le t\}$ for each $t \in \mathbb{R}$, which is in particular true if $P\{X \le Y\} = 1$.

LEMMA 1.3. *We have* $|\text{med}(X)| \le \text{med}(|X|) = (\text{med}(|X|^p))^{1/p}$ *for any* $p > 0$.

PROOF. The inequality follows from the preceding remark and Lemma 1.1:

$$0 = \text{med}(X - \text{med}(X)) \le \text{med}(|X| - \text{med}(X)) = \text{med}(|X|) - \text{med}(X),$$

and

$$0 = \text{med}(\text{med}(X) - X) \le \text{med}(\text{med}(X) + |X|) = \text{med}(X) + \text{med}\,|X|$$

which imply

$$-\text{med}(|X|) \le \text{med}(X) \le \text{med}(|X|).$$

The equality $\text{med}(|X|) = \text{med}(|X|^p)^{1/p}$ follows by utilizing the generalized inverse:

$$\begin{aligned}\text{med}(|X|) &= \inf\{t \ge 0 : P\{|X| \le t\} \ge 1/2\} = \inf\{t \ge 0 : P\{|X|^p \le t^p\} \ge 1/2\} \\ &= (\inf\{s \ge 0 : P\{|X|^p \le s\} \ge 1/2\})^{1/p} = \text{med}(|X|^p)^{1/p}.\end{aligned}$$ □

Lemmas 1.1 and 1.3 imply that $\text{MAD}(aX + b) = |a|\,\text{MAD}(X)$ and $\text{COM}(X, Y) = a\,\text{MAD}(X)^2$, if $Y = aX + b$ a.s. for some $a, b \in \mathbb{R}$. By choosing $X = Y$, the MAD therefore turns out to be a special case of the comedian. The comedian is moreover symmetric, location invariant and scale equivariant i.e., $\text{COM}(X, aY + b) = a\,\text{COM}(X, Y) = a\,\text{COM}(Y, X)$. A natural median based alternative to the *coefficient of correlation* $\varrho := \text{COV}(X, Y)/(\sigma_x\sigma_y)$ is therefore the *correlation median*

$$\delta(X, Y) := \delta := \frac{\text{COM}(X, Y)}{\text{MAD}(X)\,\text{MAD}(Y)}. \tag{1.3}$$

Since $\delta = 0$ if X and Y are independent by Lemma 1.2, and $\delta \in \{-1, 1\}$ in case of complete dependence $Y = aX + b$ a.s., the question naturally arises, whether $\delta \in [-1, 1]$ in general, just like ϱ. The answer is however "no" for general rvs X, Y, whereas in the bivariate normal case the answer is "yes" (see the discussion after Lemma 2.1 below). Even more, we will see in the next section that there is a smooth one-to-one correspondence between ϱ and δ in the normal case. Together with the limit results in Section 3 on empirical counterparts of MAD and COM, this observation offers therefore a way to estimate the coefficient of correlation of bivariate normal vectors by means of the comedian i.e., by an estimator with highest breakdown point.

2. The normal case

In this section we evaluate the comedian in the predominant case of bivariate normal vectors (X,Y), i.e. $\mathcal{L}((X,Y)) = \boldsymbol{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (\mu_1,\mu_2)^T \in \mathbb{R}^2$ is the vector of means and $\boldsymbol{\Sigma} = (\sigma_{ij})$ is the (2×2)-covariance matrix. We assume that $\sigma_{ii}^2 > 0$, $i = 1,2$. Since the distributions of X and Y are symmetric with respect to μ_1 and μ_2, we have $\text{med}(X) = \mu_1$, $\text{med}(X) = \mu_2$. Lemma 1.1 further entails

$$\begin{aligned}\text{COM}(X,Y) &= \text{med}((X-\mu_1)(Y-\mu_2))\\ &= \sigma_{11}\sigma_{22}\,\text{med}\left(\frac{X-\mu_1}{\sigma_{11}}\frac{Y-\mu_2}{\sigma_{22}}\right)\end{aligned}$$

and thus, we assume without loss of generality throughout this section that $\boldsymbol{\mu} = \mathbf{0}$ and that $\boldsymbol{\Sigma} = \begin{pmatrix}1 & \varrho\\ \varrho & 1\end{pmatrix}$ with $|\varrho| \le 1$. In the following we compile a list of auxiliary results on explicit representations of (X,Y) and XY via *independent* rvs.

By noting that $XY = ((X+Y)^2 - (X-Y)^2)/4$, the following auxiliary result is immediate from the well-known fact that if (X,Y) is bivariate normal $\boldsymbol{N}(\mathbf{0},\boldsymbol{\Sigma})$, then $X+Y$ and $X-Y$ are independent normal with means 0 and respective variances $2(1+\varrho)$ and $2(1-\varrho)$.

LEMMA 2.1. *Suppose (X,Y) is bivariate normal $\boldsymbol{N}(\mathbf{0},\boldsymbol{\Sigma})$. Then*

$$\mathcal{L}(XY) = \mathcal{L}\left(\frac{1}{2}((1+\varrho)Z_1^2 - (1-\varrho)Z_2^2)\right),$$

where Z_1, Z_2 are independent standard normal rvs.

Lemma 2.1 implies that the comedian of X and Y, whose joint distribution is $\boldsymbol{N}(\mathbf{0},\boldsymbol{\Sigma})$, is a strictly monotone increasing, symmetric and continuous function of $\varrho \in [-1,1]$. Define the comedian function by $g(\varrho) := \text{COM}(X,Y)$. Then we have

$$\begin{aligned}&g(\varrho_1) < g(\varrho_2) \quad \text{for} \quad -1 \le \varrho_1 < \varrho_2 \le 1, \qquad g(-\varrho) = -g(\varrho),\\ &g(1) = \text{med}(X^2) = \text{med}(|X|)^2 = (\Phi^{-1}(3/4))^2.\end{aligned} \tag{2.1}$$

(2.1) implies the inequality

$$\begin{aligned}|\text{COM}(X,Y)| &\le \text{med}((X-\text{med}(X))^2)^{1/2}\,\text{med}((Y-\text{med}(Y))^2)^{1/2}\\ &= \text{med}(|X-\text{med}(X)|)\,\text{med}(|Y-\text{med}(Y)|)\\ &= \text{MAD}(X)\,\text{MAD}(Y)\end{aligned} \tag{2.2}$$

for *general* bivariate normal vectors (X,Y). This inequality, which parallels the Cauchy-Schwarz inequality $|\text{COV}(X,Y)| \le E((X-E(X))^2)^{1/2}E((Y-E(Y))^2)^{1/2}$, is by Lemma 1.2 also true for general independent rvs X, Y and it is obviously true in the complete dependent case $Y = aX+b$ a.s. But one can also find examples of random vectors (X,Y) such that (2.2) does not hold. Fix for example $a > 1$ and let (X,Y) have the uniform distribution on the set $\{(x,y) : 1 \le x \le a, 1/a \le y \le 1$

and $xy > 1\} \cup \{(x,y) : -1 \le x \le -1/a, -a \le y \le -1$ and $xy > 1\}$. Then, by symmetry, $\mathrm{med}(X) = \mathrm{med}(Y) = 0$, $\mathrm{med}(X^2) = \mathrm{med}(Y^2) = 1$ which implies $\mathrm{MAD}(X) = \mathrm{MAD}(Y) = 1$, but $P\{XY \le 1\} = 0$ i.e., $\mathrm{COM}(X,Y) > 1$. The comedian has therefore the drawback, that the *comedian matrix* $(\mathrm{COM}(X_i, X_j))$ as a robust alternative to the covariance matrix of rvs $X_1, \dots, X_n$ is in general not positive (semi-)definite. The problem of non positive semidefiniteness of estimators frequently occurs in robust estimation of covariance matrices; see Rousseeuw and Molenberghs (1993), who extensively discuss and propose methods to transform such non positive semidefinite matrices to positive semidefinite ones. The investigation of the comedian matrix in bivariate models parametrized by some association parameters will be the content of future research work.

In case of arbitrary bivariate normal random vectors the preceding considerations imply however that the correlation median is always between -1 and 1, i.e.

$$\delta = \frac{\mathrm{COM}(X,Y)}{\mathrm{MAD}(X)\,\mathrm{MAD}(Y)} = g(\varrho)/g(1) \in [-1,1],$$

where ϱ is the usual correlation coefficient of X, Y. By (2.1), $\delta = \delta(\varrho)$ is a continuous and strictly increasing function from $[-1,1]$ onto $[1,1]$, satisfying $\delta(-\varrho) = -\delta(\varrho)$ and $\delta(-1) = -1$, $\delta(0) = 0$, $\delta(1) = 1$. We will see below that δ is continuously differentiable on $[-1,1]$ (Proposition 2.1).

We could therefore utilize the correlation median of normal vectors (X,Y) as a measure of dependence between (X,Y). In particular, the obvious empirical counterpart $\hat\delta_n$ can serve as a robust measure of the correlation coefficient by inversion:

$$\hat\varrho_n := g^{-1}(g(1)\hat\delta_n) \tag{2.3}$$

is a consistent estimate of ϱ, if $\hat\delta_n$ estimates δ consistently. This will be the content of Sections 3 and 4.

As a byproduct of Lemma 2.1 we obtain the characteristic function $\varphi_\varrho = E(\exp(itXY))$, $t \in \mathbb{R}$, of XY. Note that the characteristic function of the χ_1^2-distributed rv Z_1^2 is given by $E(\exp(itZ_1^2)) = (1-2it)^{-1/2}$, $t \in \mathbb{R}$. From Lemma 2.1 we obtain therefore

$$\begin{aligned}\varphi_\varrho(t) = E(\exp(itXY)) &= E\left(\exp\left(\frac{it}{2}((1+\varrho)Z_1^2 + (\varrho-1)Z_2^2)\right)\right) \\ &= E\left(\exp\left(\frac{it(1+\varrho)}{2}Z_1^2\right)\right) E\left(\exp\left(\frac{it(\varrho-1)}{2}Z_2^2\right)\right) \\ &= \frac{1}{(t^2 + (1-\varrho it)^2)^{1/2}}, \qquad t \in \mathbb{R}.\end{aligned} \tag{2.4}$$

The representation of a bivariate standard normal vector by polar coordinates, known as the *polar method* and going back to Box and Muller (1958), together with Lemma 2.1 yields the following representation of $\mathcal{L}(XY)$, which will be crucial for our further investigations.

COROLLARY 2.1. *Let again (X,Y) have distribution $\boldsymbol{N}(\mathbf{0},\boldsymbol{\Sigma})$. Then we have*

$$\mathcal{L}(XY)=\mathcal{L}(R(\cos(\pi U)+\varrho)),$$

where R and U are independent, U is uniformly on $(0,1)$ distributed and R follows the standard exponential distribution on $(0,\infty)$.

As an immediate consequence of the preceding representation of $\mathcal{L}(XY)$ we obtain the following well known fact (Feller (1971), p. 101, Huber (1981), p. 209):

$$P\{XY\geq 0\}=\frac{1}{\pi}\arccos(-\varrho),\qquad \varrho\in[-1,1].$$

Note that if we parametrize $\varrho=-\cos(\pi\alpha)$, $\alpha\in[0,1]$, then $P\{XY\geq 0\}=\alpha$ becomes the identity on $[0,1]$.

PROOF OF COROLLARY 2.1. Let U_1, U_2 be independent and uniformly on $(0,1)$ distributed rvs. The polar method implies that $Z_1:=\sqrt{-2\log U_1}\cos(2\pi U_2)$ and $Z_2:=\sqrt{-2\log U_1}\sin(2\pi U_2)$ are independent and standard normal rvs. As a consequence we obtain from Lemma 2.1 the representation

$$\begin{aligned}\mathcal{L}(XY)&=\mathcal{L}\left(\frac{1}{2}((1+\varrho)Z_1^2-(1-\varrho)Z_2^2)\right)\\&=\mathcal{L}(-\log U_1((1+\varrho)\cos^2(2\pi U_2)-(1-\varrho)\sin^2(2\pi U_2)))\\&=\mathcal{L}(-\log U_1(\cos^2(2\pi U_2)-\sin^2(2\pi U_2)+\varrho))\\&=\mathcal{L}(-\log U_1(\cos(4\pi U_2)+\varrho)).\end{aligned}$$

Finally, observe that $R:=-\log U_1$ has a standard exponential distribution and that $\mathcal{L}(\cos(4\pi U_2))=\mathcal{L}(\cos(\pi U_2))$. □

Corollary 2.1 enables us to write down the df of $\mathcal{L}(XY)$ in closed form. Its density can be derived from the following formula by interchanging differentiation and integration.

COROLLARY 2.2. *Under the conditions of Corollary 2.1 we have*

$$\begin{aligned}P\{XY\geq t\}&=\begin{cases}\displaystyle\int_0^{(1/\pi)\arccos(-\varrho)}\exp\left(-\frac{t}{\cos(\pi u)+\varrho}\right)du, & t\geq 0\\ \displaystyle 1-\int_{(1/\pi)\arccos(-\varrho)}^1\exp\left(-\frac{t}{\cos(\pi u)+\varrho}\right)du, & t\leq 0\end{cases}\\&=\begin{cases}\displaystyle\pi^{-1}\int_{-\varrho}^1\exp\left(-\frac{t}{u+\varrho}\right)\frac{1}{(1-u^2)^{1/2}}du, & t\geq 0\\ \displaystyle 1-\pi^{-1}\int_{-1}^{-\varrho}\exp\left(-\frac{t}{u+\varrho}\right)\frac{1}{(1-u^2)^{1/2}}du, & t\leq 0.\end{cases}\end{aligned}$$

PROOF. Corollary 2.1 together with Fubini's theorem yield

$$\begin{aligned} P\{XY \geq t\} &= P\{R(\cos(\pi U) + \varrho) \geq t\} \\ &= P\{R(\cos(\pi U) + \varrho) \geq t, \cos(\pi U) + \varrho \leq 0\} \\ &\quad + P\{R(\cos(\pi U) + \varrho) \geq t, \cos(\pi U) + \varrho \geq 0\} \\ &= \begin{cases} \int_0^{(1/\pi)\arccos(-\varrho)} P\left\{R \geq \frac{t}{\cos(\pi u) + \varrho}\right\} du, & t \geq 0 \\ \int_{(1/\pi)\arccos(-\varrho)}^1 P\left\{R \leq \frac{t}{\cos(\pi u) + \varrho}\right\} du & \\ \qquad + P\{\cos(\pi U) + \varrho \geq 0\}, & t \leq 0 \end{cases} \end{aligned}$$

from which the assertion follows. □

The preceding result together with some elementary analysis implies that the df $H_\varrho(t) = 1 \quad P\{XY \geq t\}$ of XY is continuously differentiable at $t \neq 0$ with derivative

$$\text{(2.5)} \quad h_\varrho(t) = \begin{cases} \int_0^{(1/\pi)\arccos(-\varrho)} \exp\left(-\frac{t}{\cos(\pi u) + \varrho}\right) \frac{1}{\cos(\pi u) + \varrho} du, & t > 0 \\ \int_{(1/\pi)\arccos(-\varrho)}^1 \exp\left(-\frac{t}{\cos(\pi u) + \varrho}\right) \frac{-1}{\cos(\pi u) + \varrho} du, & t < 0. \end{cases}$$

$$= \begin{cases} \pi^{-1} \int_{-\varrho}^1 \exp\left(-\frac{t}{u + \varrho}\right) \frac{1}{(u + \varrho)1 - u^2)^{1/2}} du, & t > 0 \\ \pi^{-1} \int_{\varrho}^1 \exp\left(-\frac{t}{\varrho - u}\right) \frac{1}{(u - \varrho)(1 - u^2)^{1/2}} du, & t < 0. \end{cases}$$

Observe that $h_\varrho(t) \to \infty$ as $t \to 0$, i.e. the density h_ϱ of XY has a pole at zero for any ϱ. This observation has the effect that asymptotic normality of an empirical counterpart of the comedian in case of *independent* normal rvs requires a different standardization and an extra proof (see Proposition 4.1 and Theorem 4.2). Next we will show that the function $g(\varrho) = \text{med}(XY)$ in (2.1) is continuously differentiable on $[-1, 1]$.

PROPOSITION 2.1. *The function* $g(\varrho) = \text{med}(XY)$, *where* (X, Y) *has distribution* $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$, *is continuously differentiable on* $[-1, 1]$ *with derivative satisfying*

$$g'(\varrho) = g(|\varrho|) \frac{\int_{-|\varrho|}^1 \exp\left(-\frac{g(|\varrho|)}{u + |\varrho|}\right) \frac{1}{(u + |\varrho|)^2(1 - u^2)^{1/2}} du}{\int_{-|\varrho|}^1 \exp\left(-\frac{g(|\varrho|)}{u + |\varrho|}\right) \frac{1}{(u + |\varrho|)(1 - u^2)^{1/2}} du}, \qquad \varrho \neq 0$$

and $g'(0) = 0$.

By noting that $g'(\varrho)/g(\varrho)$ is the derivative of $\log(g(\varrho))$, the preceding result implies that g satisfies for $\varrho \geq 0$ the integral equation

$$(2.6)\quad g(\varrho) = (\Phi^{-1}(3/4))^2 \cdot \exp\left(- \int_\varrho^1 \frac{\int_{-x}^1 \exp\left(-\frac{g(x)}{u+x}\right) \frac{1}{(u+x)^2(1-u^2)^{1/2}} du}{\int_{-x}^1 \exp\left(-\frac{g(x)}{u+x}\right) \frac{1}{(u+x)(1-u^2)^{1/2}} du} dx \right).$$

We are presently unable to solve this integral equation and to provide the function g in explicit form. A Monte Carlo approximation visualizing its shape is given at the end of this section.

Proof of Proposition 2.1. The following proof of the differentiability of $g(\varrho)$ for $\varrho \neq 0$ is based on implicit differentiation. Define the function H on $[-1,1] \times \mathbb{R}$ by

$$H(\varrho, t) := H_\varrho(t) - 1/2 = P\{XY > t\} - 1/2.$$

Then $H(\varrho, \cdot)$ is a continuous and strictly decreasing function with g being the unique solution of the equation $H(\varrho, g(\varrho)) = 0$, $\varrho \in [-1,1]$.

Fix now $\delta \in (0,1]$ and $t > 0$. Formula (2.5) implies

$$\frac{\partial}{\partial t} H(\varrho, t) = -\pi^{-1} \int_{-\varrho}^1 \exp\left(-\frac{t}{u+\varrho}\right) \frac{1}{(u+\varrho)(1-u^2)^{1/2}} du < 0.$$

Next we compute the partial derivative of H with respect to ϱ. By substituting $u \mapsto \pi^{-1}\arccos(u)$ we obtain from Corollary 2.2 the expansion

$$\begin{aligned}
&(\varrho + h, t) - H(\varrho, t) \\
&= \pi^{-1} \int_{-\varrho-h}^1 \exp\left(-\frac{t}{u+\varrho+h}\right) \frac{1}{(1-u^2)^{1/2}} du \\
&\quad - \pi^{-1} \int_{-\varrho}^1 \exp\left(-\frac{t}{u+\varrho}\right) \frac{1}{(1-u^2)^{1/2}} du \\
&= \pi^{-1} \int_{-\varrho}^1 \left(\exp\left(-\frac{t}{u+\varrho+h}\right) - \exp\left(-\frac{t}{u+\varrho}\right) \right) \frac{1}{(1-u^2)^{1/2}} du + o(h).
\end{aligned}$$

Fix now $a \in (-\varrho, 1)$. Splitting up the preceding integral into the sum $\int_{-\varrho}^a + \int_a^1$ of two integrals, substituting $u \mapsto u - h$ in the first integral and using Taylor's formula, we obtain that the above integral equals

$$\pi^{-1} \int_{-\varrho+h}^{a+h} \exp\left(-\frac{t}{u+\varrho}\right) \frac{1}{(1-(u-h)^2)^{1/2}} du - \pi^{-1} \int_{-\varrho}^a \exp\left(-\frac{t}{u+\varrho}\right) \frac{1}{(1-u^2)^{1/2}} du$$

$$
\begin{aligned}
&+\pi^{-1}\int_a^1 \exp\left(-\frac{1}{u+\varrho}\right)\left(\exp\left(\frac{t}{u+\varrho}-\frac{t}{u+\varrho+h}\right)-1\right)\\
&\cdot\frac{1}{(1-u^2)^{1/2}}du\\
=&\, h\pi^{-1}\exp\left(-\frac{t}{a+\varrho}\right)\frac{1}{(1-a^2)^{1/2}}+o(h)\\
&+\pi^{-1}\int_{-\varrho}^{a}\exp\left(-\frac{t}{u+\varrho}\right)\frac{(1-u^2)^{1/2}-(1-(u-h)^2)^{1/2}}{(1-(u-h)^2)^{1/2}(1-u^2)^{1/2}}du\\
&+\pi^{-1}\int_a^1\exp\left(-\frac{t}{u+\varrho}\right)\frac{th}{(u+\varrho)(u+\varrho+h)}\frac{1}{(1-u^2)^{1/2}}du\\
=&\, h\pi^{-1}\exp\left(-\frac{1}{a+\varrho}\right)\frac{1}{(1-a^2)^{1/2}}+o(h)\\
&-h\pi^{-1}\int_{-\varrho}^{a}\exp\left(-\frac{t}{u+\varrho}\right)\frac{u}{(1-u^2)^{3/2}}du\\
&+h\pi^{-1}\int_a^1\exp\left(-\frac{t}{u+\varrho}\right)\frac{t}{(u+\varrho)^2(1-u^2)^{1/2}}du.
\end{aligned}
$$

This representation implies

$$
\begin{aligned}
\frac{\partial}{\partial\varrho}H(\varrho,t)=&\frac{1}{\pi(1-a^2)^{1/2}}\exp\left(-\frac{t}{a+\varrho}\right)\\
&-\pi^{-1}\int_{-\varrho}^{a}\exp\left(-\frac{t}{u+\varrho}\right)\frac{u}{(1-u^2)^{3/2}}du\\
&+\pi^{-1}\int_a^1\exp\left(-\frac{t}{u+\varrho}\right)\frac{t}{(u+\varrho)^2(1-u^2)^{1/2}}du.
\end{aligned}
$$

By letting now a tend to $-\varrho$ we obtain

$$
\frac{\partial}{\partial\varrho}H(\varrho,t)=\pi^{-1}t\int_{-\varrho}^{1}\exp\left(-\frac{t}{u+\varrho}\right)\frac{1}{(u+\varrho)^2(1-u^2)^{1/2}}du>0.
$$

The implicit differentiability theorem implies therefore that the function g is differentiable on $(0,1]$ with derivative satisfying

$$
\frac{\partial}{\partial\varrho}H(\varrho,g(\varrho))+\frac{\partial}{\partial t}H(\varrho,g(\varrho))g'(\varrho)=0
$$

i.e.,

$$
g'(\varrho)=\frac{g(\varrho)\int_{-\varrho}^{1}\exp\left(-\frac{g(\varrho)}{u+\varrho}\right)\frac{1}{(u+\varrho)^2(1-u^2)^{1/2}}du}{\int_{\varrho}^{1}\exp\left(-\frac{g(\varrho)}{u+\varrho}\right)\frac{1}{(u+\varrho)(1-u^2)^{1/2}}du}.
$$

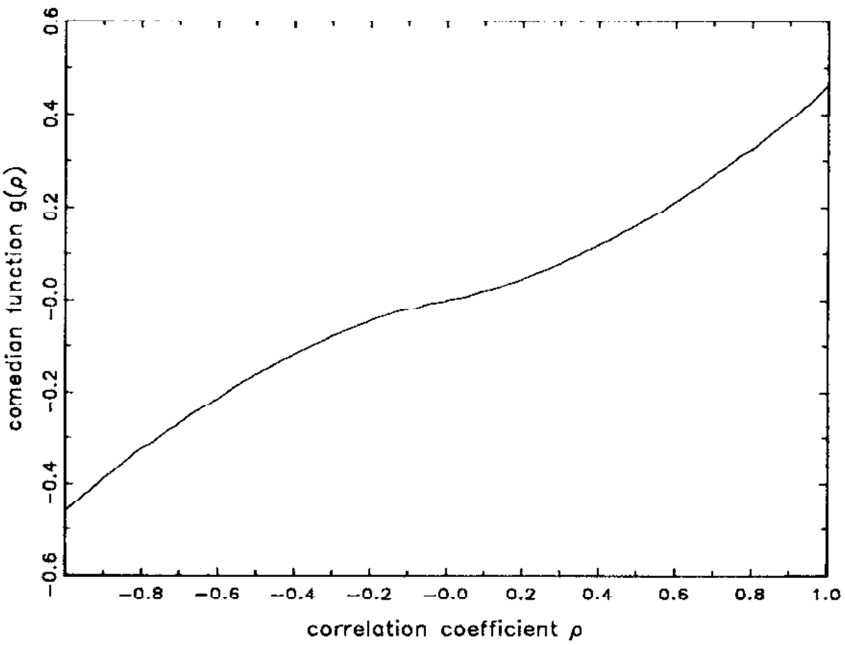

Fig. 1. Approximation (obtained by Monte Carlo) of the comedian function $g(\rho) = \mathrm{COM}(X,Y)$ as a function of the correlation coefficient ρ. Note that $g(\rho)$ ranges over $[-g(1), g(1)]$, where $g(1) = (\Phi^{-1}(0.75))^2 \approx 0.455$.

It remains to show that $g'(\varrho) \to 0$ as $\varrho \to 0$. Fix to this end $\varepsilon \in (0,1)$ and split up the integral in the numerator of $g'(\varrho)$ into the sum $\int_{-\varrho}^{\varepsilon} + \int_{\varepsilon}^{1}$. Then, with $C > 0$ denoting an appropriate constant we obtain for $\varrho \downarrow 0$ the bound

$$g'(\varrho) \leq \frac{C g(\varrho) \int_{-\varrho}^{\varepsilon} \exp\left(-\frac{g(\varrho)}{u+\varrho}\right) \frac{1}{(u+\varrho)^2} du + o(1)}{\int_{-\varrho}^{\varepsilon} \exp\left(-\frac{g(\varrho)}{u+\varrho}\right) \frac{1}{u+\varrho} du}.$$

The substitution $u \mapsto g(\varrho)u - \varrho$ implies that the right hand side equals

$$\frac{C \int_0^{(\varepsilon+\varrho)/g(\varrho)} \exp\left(-\frac{1}{u}\right) \frac{1}{u^2} du + o(1)}{\int_{-0}^{(\varepsilon+\varrho)/g(\varrho)} \exp\left(-\frac{1}{u}\right) \frac{1}{u} du} \to 0$$

if $\varrho \to 0$, since as $(\varepsilon+\varrho)/g(\varrho) \to \infty$, the numerator is bounded but the denominator converges to infinity. This proves $g'(\varrho) \to 0$ as $\varrho \to 0$.

Since g is continuous on $[0,1]$, the mean value theorem together with the preceding considerations implies with $\vartheta \in (0,\varrho)$

$$\frac{g(\varrho) - g(0)}{\varrho} = g'(\vartheta) \to 0 \quad \text{if} \quad \varrho \to 0$$

i.e., g is differentiable at $\varrho = 0$ with $g'(0) = 0$ and the proof is complete. □

Since we are presently unable to state $g(\varrho)$ explicitely as a function of ϱ, its evaluation as a solution t_ϱ of the equation $H_\varrho(t) = 0$ (see Corollary 2.2) requires approximation methods. Figure 1 is the result of a Monte Carlo simulation of g using GAUSS$_{\mathrm{TM}}$, version 2.1, based on the empirical medians of 8000 pseudo random numbers $-\log(u_i)(\cos(\pi w_i) + \varrho)$, with $\varrho = -1$ to 1 by 0.01. The function g is almost linear near 0.

3. Strong consistency of the comedian

In this section we establish strong consistency of the empirical comedian, based on independent copies $(X_1, Y_1), (X_2, Y_2), \ldots$ of (X, Y). Denote by $\widehat{\mathrm{med}}_n(X)$, $\widehat{\mathrm{med}}_n(Y)$ empirical medians of $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ and define the empirical MAD and comedian by

$$(3.1) \quad \begin{aligned} \widehat{\mathrm{MAD}}_n(X) &:= \mathrm{med}\{|X_i - \widehat{\mathrm{med}}_n(X)|, i = 1, \ldots, n\} \\ \widehat{\mathrm{COM}}_n(X, Y) &:= \mathrm{med}\{(X_i - \widehat{\mathrm{med}}_n(X))(Y_i - \widehat{\mathrm{med}}_n(Y)), i = 1, \ldots, n\}. \end{aligned}$$

In the following result we establish strong consistency of $\widehat{\mathrm{COM}}_n(X, Y)$ for general random vectors (X, Y) under suitable regularity conditions. By F, G and H we denote the cdfs of X, Y and $(X - \mathrm{med}(X))(Y - \mathrm{med}(Y))$, respectively.

THEOREM 3.1. *Suppose that F, G and H are continuous and strictly increasing at $F^{-1}(1/2)$, $G^{-1}(1/2)$ and $H^{-1}(1/2)$, respectively. Then,*

$$\widehat{\mathrm{COM}}_n(X, Y) \underset{n \to \infty}{\longrightarrow} \mathrm{COM}(X, Y) \qquad a.s.$$

From the fact that $\mathrm{COM}(X, X) = \mathrm{MAD}^2(X)$ for a general rv X (see the discussion after Lemma 1.3 (i)), strong consistency of $\widehat{\mathrm{MAD}}_n(X)$, which has been established by Hall and Welsh ((1985), Theorem 1), is a special case of the preceding result by choosing $Y_i = X_i$.

COROLLARY 3.1. *If F and H are continuous and strictly increasing at $F^{-1}(1/2)$ and $H^{-1}(1/2)$, respectively, we have*

$$\widehat{\mathrm{MAD}}_n(X) \underset{n \to \infty}{\longrightarrow} \mathrm{MAD}(X) \qquad a.s.$$

By the continuity and strong monotonicity of the comedian function g, defined in (2.1), the two preceding results imply strong consistency of the transformed empirical correlation median in the normal case.

COROLLARY 3.2. *Suppose that (X, Y) has an arbitrary nondegenerate bivariate normal distribution with correlation coefficient $\varrho \in [-1, 1]$. Then the transformed empirical correlation median*

$$\hat{\varrho}_n = g^{-1}\left(g(1) \frac{\widehat{\mathrm{COM}}_n(X,Y)}{\widehat{\mathrm{MAD}}_n(X)\widehat{\mathrm{MAD}}_n(Y)} \right),$$

with the convention $\hat{\varrho}_n = 1$ or -1 if the argument is above $g(1)$ or below $g(-1)$, is a strongly consistent estimate of ϱ i.e.,

$$\hat{\varrho}_n \underset{n\to\infty}{\longrightarrow} \varrho \qquad a.s.$$

Furthermore,

$$\frac{\widehat{\mathrm{MAD}}_n(X)\widehat{\mathrm{MAD}}_n(Y)}{g(1)} \hat{\varrho}_n \underset{n\to\infty}{\longrightarrow} \mathrm{COV}(X, Y) \qquad a.s.$$

Corollary 3.2 suggests as a highly robust and consistent alternative to the empirical correlation matrix of normal vectors $(X_1^{(1)}, \ldots, X_1^{(k)}), \ldots, (X_n^{(1)}, \ldots, X_n^{(k)})$ that matrix, whose entries are the pairwise transformed correlation medians $\hat{\varrho}_n = \hat{\varrho}_n(X^{(i)}, X^{(j)})$. This matrix, like the comedian matrix introduced in the preceding section, will in general not be positive semidefinite. To counteract this disadvantage it should therefore also be transformed as proposed in Rousseeuw and Molenberghs (1993).

For the proof of Theorem 3.1, which uses a truncation argument, we need the following lemma.

LEMMA 3.1. *Let $a_1 \le a_2 \le \cdots \le a_n$ and $b_1 \le b_2 \le \cdots \le b_n$ be two sets of real numbers such that at most $k \le (n-2)/4$ numbers in the two sets differ. Then we have*

$$|\,\mathrm{med}(a) - \mathrm{med}(b)| < b_{[n/2]+2k+1} - b_{[n/2]-2k},$$

where $[x] := \sup\{k \text{ integer} : k \le x\}$ denotes the integer part of $x \in \mathbb{R}$.

PROOF. Observe first that $\mathrm{med}(b) \in [b_{[n/2]}, b_{[n/2]+1}]$. Choose now the integer i minimal such that the interval $(b_{[n/2]+1}, b_{[n/2]+1+i}]$ contains at least k numbers b_j that belong also to the set $\{a_1, \ldots, a_n\}$. Note that $i \le 2k$. It follows that $\mathrm{med}(a) \le b_{[n/2]+1+i} \le b_{[n/2]+1+2k}$.

Choose on the other hand now the integer m minimal such that the interval $[b_{[n/2]-m}, b_{[n/2]})$ contains at least k numbers b_j that belong also to the set $\{a_1, \ldots, a_n\}$. Again $m \le 2k$ and $\mathrm{med}(a) \ge b_{[n/2]-m} \ge b_{[n/2]-2k}$ which implies the assertion. □

PROOF OF THEOREM 3.1. In the following we will make use of the quantile transform method i.e., a sequence of iid rvs with an arbitrary common cdf F can be assumed to be given in the form $F^{-1}(U_1), F^{-1}(U_2), \ldots,$ where $U_1, U_2, \ldots$ are independent and uniformly on $(0,1)$ distributed rvs (see e.g. Section 1.2 of Reiss (1989) for details). Further we will utilize the well-known fact that

$$\max_{1\le k\le n}\left|U_{k:n}-\frac{k}{n+1}\right|\underset{n\to\infty}{\longrightarrow}0 \quad \text{a.s.,} \tag{3.2}$$

where $Z_{1:n}\le Z_{2:n}\le\dots\le Z_{n:n}$ denote the ordered values pertaining to arbitrary rvs $Z_1,\dots,Z_n$.

Choose for a given small $\varepsilon>0$ a number $A=A(\varepsilon)>0$ large enough such that $P\{|X|<A\}>1-\varepsilon/2$ and $P\{|Y|<A\}>1-\varepsilon/2$ and define for $i=1,2,\dots$ the truncated rvs

$$\tilde X_i:=\begin{cases}A & \text{if } X_i>A\\ X_i & \text{if } |X_i|\le A\\ -A & \text{if } X_i<-A\end{cases} \quad\text{and}\quad \tilde Y_i:=\begin{cases}A & \text{if } Y_i>A\\ Y_i & \text{if } |Y_i|\le A\\ -A & \text{if } Y_i<-A\end{cases}.$$

Note that $\widehat{\text{med}}_n(X)\in[X_{[n/2]:n},X_{[n/2]+1:n}]=[\tilde X_{[n/2]:n},\tilde X_{[n/2]+1:n}]$ and $\widehat{\text{med}}_n(Y)\in[Y_{[n/2]:n},Y_{[n/2]+1:n}]=[\tilde Y_{[n/2]:n},\tilde Y_{[n/2]+1:n}]$ ultimately a.s., provided $F^{-1}(1/2),G^{-1}(1/2)\in(-A,A)$ which we assume (use the quantile transformation method, the convergence result (3.2) and the continuity of F^{-1},G^{-1} at $1/2$). Thus we obtain

$$\widehat{\text{med}}_n(X)=\widehat{\text{med}}_n(\tilde X) \quad \text{ultimately a.s.}$$

and

$$\widehat{\text{med}}_n(Y)=\widehat{\text{med}}_n(\tilde Y) \quad \text{ultimately a.s.}$$

The strong law of large numbers implies that there are a.s. ultimately more than $n(1-\varepsilon/2)$ observations among $X_1,\dots,X_n$ as well as among $Y_1,\dots,Y_n$ in the interval $(-A,A)$. This implies that in this case at least $n(1-\varepsilon)$ *pairs* (X_i,Y_i) are in $(-A,A)\times(-A,A)$. Thus we obtain that a.s. ultimately

$$(X_i-\widehat{\text{med}}_n(X))(Y_i-\widehat{\text{med}}_n(Y))=(\tilde X_i-\widehat{\text{med}}_n(\tilde X))(\tilde Y_i-\widehat{\text{med}}_n(\tilde Y))$$

for at least $n(1-\varepsilon)$ indices i among $1,\dots,n$. Put now for notational convenience $Z_i^{(n)}:=(\tilde X_i-\widehat{\text{med}}_n(\tilde X))(\tilde Y_i-\widehat{\text{med}}_n(\tilde Y))$, $i=1,\dots,n$. Lemma 3.1 then implies

$$|\widehat{\text{COM}}_n(X,Y)-\widehat{\text{med}}_n(Z^{(n)})|\le Z^{(n)}_{[n/2]+2[\varepsilon n]+1:n}-Z^{(n)}_{[n/2]-2[\varepsilon n]:n} \tag{3.3}$$

ultimately a.s.

Since $|\tilde X_i|$ and $|\tilde Y_i|$ are bounded by A and $\widehat{\text{med}}_n(\tilde X)\to_{n\to\infty}F^{-1}(1/2)$, $\widehat{\text{med}}_n(\tilde Y)\to_{n\to\infty}G^{-1}(1/2)$ a.s. (use the quantile transformation method), we obtain with $V_i:=(\tilde X_i-F^{-1}(1/2))(\tilde Y_i-G^{-1}(1/2))$, $i=1,2,\dots,$

$$\max_{1\le i\le n}|Z_i^{(n)}-V_i|\underset{n\to\infty}{\longrightarrow}0 \quad \text{a.s.}$$

which implies $\max_{1\le i\le n}|Z_{i:n}^{(n)} - V_{i:n}| \to_{n\to\infty} 0$ a.s. Thus, inequality (3.3) implies

$$|\widehat{\mathrm{COM}}_n(X,Y) - \widehat{\mathrm{med}}_n(V)| \le V_{[n/2]+2[\varepsilon n]+1:n} - V_{[n/2]+2[\varepsilon n]:n} + o(1) \tag{3.4}$$

ultimately a.s.

Observe now that $V_1, V_2, \dots$ are iid rvs whose cdf H_ε, say, converges to H as $\varepsilon \to 0$ and that a.s. $\widehat{\mathrm{med}}_n(V) \in [H_\varepsilon^{-1}(1/2-3\varepsilon), H_\varepsilon^{-1}(1/2+3\varepsilon)]$ ultimately and $\limsup_{n\to\infty} V_{[n/2]+2[\varepsilon n]+1} \le H_\varepsilon^{-1}(1/2+3\varepsilon)$, $\liminf_{n\to\infty} V_{[n/2]-2[\varepsilon n]} \ge H_\varepsilon^{-1}(1/2-3\varepsilon)$; (use the quantile transformation method, (3.2) and the monotonicity of H_ε^{-1} on $(0,1)$). Thus we obtain from (3.4) and the triangular inequality that a.s. (recall that $\mathrm{COM}(X,Y) = H^{-1}(1/2)$)

$$\begin{aligned}
&\limsup_{n\to\infty} |\widehat{\mathrm{COM}}_n(X,Y) - \mathrm{COM}(X,Y)| \\
&\quad\le \limsup_{n\to\infty} |\widehat{\mathrm{COM}}_n(X,Y) - \widehat{\mathrm{med}}_n(V)| + \limsup_{n\to\infty} |\widehat{\mathrm{med}}_n(V) - H^{-1}(1/2)| \\
&\quad\le H_\varepsilon^{-1}(1/2+3\varepsilon) - H_\varepsilon^{-1}(1/2-3\varepsilon) + |H_\varepsilon^{-1}(1/2+3\varepsilon) - H^{-1}(1/2)| \\
&\qquad + |H_\varepsilon^{-1}(1/2-3\varepsilon) - H^{-1}(1/2)| \\
&\quad\underset{\varepsilon\to 0}{\longrightarrow} 0
\end{aligned}$$

since H_ε converges to H pointwise, which implies the convergence $H_\varepsilon^{-1}(q) \to_{\varepsilon\to 0} H^{-1}(q)$ at every continuity point of H^{-1} (see e.g. Lemma 1.2.9 in Reiss (1989)), and since H^{-1} is continuous at $1/2$. This completes the proof of Theorem 3.1. □

4. Asymptotic normality of the comedian

In this section we will establish asymptotic normality of empirical counterparts of the comedian. It turns out that in the particular case, where X, Y are independent, the limiting distribution is not affected if the marginal medians $\mathrm{med}(X)$, $\mathrm{med}(Y)$ are known. This is different to the estimation of the MAD, where the phenomenon occurs that an *estimated* marginal median can even reduce the limiting variance. For reference we state the following theorem on asymptotic normality of the empirical MAD, proved by Hall and Welsh (1985). Further results for the MAD in a regression model were established by Welsh (1986).

THEOREM 4.1. *Suppose that the F is continuous near and differentiable at $F^{-1}(1/2)$, $F^{-1}(1/2)+\mathrm{MAD}(X)$ and $F^{-1}(1/2)-\mathrm{MAD}(X)$ with $f(F^{-1}(1/2)) > 0$ and $A := f(F^{-1}(1/2)+\mathrm{MAD}(X)) + f(F^{-1}(1/2)-\mathrm{MAD}(X)) > 0$, where $f = F'$. Then,*

$$\sqrt{n}(\widehat{\mathrm{MAD}}_n(X) - \mathrm{MAD}(X)) \underset{\mathcal{D}}{\to} N\left(0, \frac{1}{4A^2}\left(1 + \frac{B}{f(F^{-1}(1/2))^2}\right)\right),$$

where $C := f(F^{-1}(1/2)-\mathrm{MAD}(X)) - f(F^{-1}(1/2)+\mathrm{MAD}(X))$, and $B := C^2 + 4Cf(F^{-1}(1/2))(1 - F(F^{-1}(1/2)+\mathrm{MAD}(X)) - F(F^{-1}(1/2)-\mathrm{MAD}(X)))$.

If $f(F^{-1}(1/2)-\mathrm{MAD}(X))=f(F^{-1}(1/2)+\mathrm{MAD}(X))$, which is in particular true if the cdf F is symmetric with respect to $F^{-1}(1/2)$ i.e., $F(F^{-1}(1/2)-y)=1-F(F^{-1}(1/2)+y)$, $y\in\mathbb{R}$, then B equals zero and thus, the variance of the limiting normal distribution in the preceding result reduces to $1/(16f(F^{-1}(1/2)+\mathrm{MAD}(X))^2)$.

COROLLARY 4.1. *If in addition to the assumptions of Theorem* 4.1 *the cdf F is symmetric around $F^{-1}(1/2)$, then*

$$\sqrt{n}(\widehat{\mathrm{MAD}}_n(X)-\mathrm{MAD}(X))\underset{\mathcal{D}}{\to}N\left(0,\frac{1}{16f(F^{-1}(1/2)+\mathrm{MAD}(X))^2}\right).$$

If F is in particular the cdf of the normal distribution with arbitrary mean and variance $\sigma^2>0$, then

$$\sqrt{n}(\widehat{\mathrm{MAD}}_n(X)-\sigma\Phi^{-1}(3/4))\underset{\mathcal{D}}{\to}N\left(0,\frac{\sigma^2}{16\varphi(\Phi^{-1}(3/4))^2}\right),$$

where φ denotes the standard normal density.

Theorem 4.1 implies that under those assumptions on the cdf F of X

$$\sqrt{n}(\widehat{\mathrm{COM}}_n(X,Y)-\mathrm{COM}(X,Y))\underset{\mathcal{D}}{\to}N\left(0,\frac{a^2\,\mathrm{MAD}^2(X)}{A^2}\left(1+\frac{B}{f(F^{-1}(1/2))}\right)\right)$$

if $Y=aX+b$ a.s. for some $a,b\in\mathbb{R}$. The constants A, B are defined in Theorem 4.1.

If the median of F is *known* and unique, then we can replace $\widehat{\mathrm{MAD}}_n(X)$ by

$$\widetilde{\mathrm{MAD}}_n(X):=\mathrm{med}\{|X_i-\mathrm{med}(X)|,i=1,\dots,n\}.$$

If F is differentiable at $\mathrm{med}(X)+\mathrm{MAD}(X)$ and $\mathrm{med}(X)-\mathrm{MAD}(X)$ with $f(\mathrm{med}(X)+\mathrm{MAD}(X))+f(\mathrm{med}(X)-\mathrm{MAD}(X))>0$, then by Example 4.1.4 in Reiss (1989)

$$\sqrt{n}(\widetilde{\mathrm{MAD}}_n(X)-\mathrm{MAD}(X))\underset{\mathcal{D}}{\to}N\left(0,\frac{1}{4A^2}\right). \tag{4.1}$$

Since the term B in the variance of the limiting distribution of $\sqrt{n}(\widehat{\mathrm{MAD}}_n(X)-\mathrm{MAD}(X))$ in Theorem 4.1 does not vanish in general, the limiting distributions of $\widehat{\mathrm{MAD}}_n(X)$ and $\widetilde{\mathrm{MAD}}_n(X)$ differ in general. This is different to the limiting distribution of the sample variance, which is not affected by plugging in the sample mean.

It is easy to find examples of cdfs where the term B is positive. This is what we would expect as we replace the unknown median $\mathrm{med}(X)$ in $\widetilde{\mathrm{MAD}}_n(X)$ by the empirical one. But one can also find examples, where B is negative, which contradicts our intuition, since in this case $\widetilde{\mathrm{MAD}}_n(X)$ is for large n more concentrated around $\mathrm{MAD}(X)$ than $\widehat{\mathrm{MAD}}_n(X)$. The following example leads to a negative value of B. Put

$$f(x) := \begin{cases} 1/4, & -2 \le x \le 0 \\ 1/4 + 4/5x^3 - x^2, & 0 \le x \le x_0, \end{cases} \tag{4.2}$$

where x_0 satisfies $\int_0^{x_0} 1/4 + 4/5x^3 - x^2 dx = 1/2$. Then the median of the corresponding cdf F is 0 and $\mathrm{MAD} \in [1.3191, 1.3192]$. Consequently, $f(-\mathrm{MAD}) - f(\mathrm{MAD}) = \mathrm{MAD}^2 - 4/5\,\mathrm{MAD}^3 < 0$ and

$$\begin{aligned} B &= (f(-\mathrm{MAD}) - f(\mathrm{MAD}))^2 \\ &\quad + 4f(0)(f(-\mathrm{MAD}) - f(\mathrm{MAD}))(1 - F(\mathrm{MAD}) - F(-\mathrm{MAD})) \\ &= (\mathrm{MAD}^2 - 4/5\,\mathrm{MAD}^3)(\mathrm{MAD}^2 \quad 7/15\,\mathrm{MAD}^3 \quad \mathrm{MAD}^4/5) < 0. \end{aligned}$$

This example shows that knowledge of the population median does not necessarily pay (in the sense of asymptotic relative efficiency) when estimating the MAD.

It is interesting to compare the limiting normal distribution of $\widehat{\mathrm{MAD}}_n(X)/\Phi^{-1}(3/4)$ as an estimator of the standard deviation σ with that of $\widehat{\mathrm{VAR}}_n(X) = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$ in the particular case, where X has the normal distribution $N(\mu, \sigma^2)$. By Corollary 4.1 we have

$$\sqrt{n}\left(\frac{\widehat{\mathrm{MAD}}_n(X)}{\Phi^{-1}(3/4)} - \sigma\right) \underset{\mathcal{D}}{\to} N\left(0, \frac{\sigma^2}{16\varphi(\Phi^{-1}(3/4))^2\Phi^{-1}(3/4)^2}\right),$$

whereas

$$\sqrt{n}(\widehat{\mathrm{VAR}}_n^{1/2}(X) - \sigma) \underset{\mathcal{D}}{\to} N(0, \sigma^2/2).$$

The asymptotic relative efficiency (ARE) of MAD with respect to the empirical standard deviation, defined as the ratio of the variances of their limiting normal distributions equals therefore

$$\mathrm{ARE} = 8\varphi(\Phi^{-1}(3/4))^2\Phi^{-1}(3/4)^2 \sim 0.368. \tag{4.3}$$

As an estimator of σ, the MAD requires therefore about $3n$ observations to perform roughly as well as the empirical standard deviation based on n observations, if n is large (see, for example, Section 1.15.3 in Serfling (1980) for a discussion of the concept of ARE). This is the price one has to pay for the robustness of the estimator.

The price for robustness can actually be lowered by utilizing for example the estimator

$$S_n := c\,\mathrm{med}_{1\le i\le n}\{\mathrm{med}_{1\le j\le n} |x_i - x_j|\}, \tag{4.4}$$

introduced by Rousseeuw and Croux (1993) as an alternative to the empirical MAD. It was proved that S_n is consistent for the corresponding population functional

$$S(X) := c\,\mathrm{med}_{X_1}\{\mathrm{med}_{X_2}|X_1 - X_2|\}, \tag{4.5}$$

where X_1, X_2 are independent copies of X. Note that the constant c can be chosen so that $S(X)$ equals the standard deviation σ at normal distributions, just like the constant $1/\Phi^{-1}(3/4)$ for the MAD. The asymptotic efficiency of S_n is 58%, compared to 37%, roughly, for the empirical MAD. Moreover, S_n is *location-free* in the sense that it does not use any location estimator (like the sample median), so the effect that knowing med (X) in advance may be a disadvantage for the sample MAD does not occur for S_n.

In view of the general equality $\mathrm{med}(|X|) = \mathrm{med}(X^2)^{1/2}$ (see Lemma 1.3) one could also extend S_n to an estimator of covariance, given by

$$S_n(X,Y) := c^2\,\mathrm{med}_{1\le i\le n}\{\mathrm{med}_{1\le j\le n}(x_i - x_j)(y_i - y_j)\}. \tag{4.6}$$

The author is grateful to the referee for providing this idea, whose investigation has to be the content of future research work.

In the case, where $\mathrm{med}(X)$ and $\mathrm{med}(Y)$ are known, we can also modify the empirical comedian in an obvious way and obtain immediately from the asymptotic normality of sample quantiles (see e.g. Reiss (1989), Example 4.1.1) its limiting distribution, stated in the following result. In practice it is often assumed that X and Y are rvs which are symmetric with respect to some known real numbers. In this case, these centers of symmetry are $\mathrm{med}(X)$ and $\mathrm{med}(Y)$.

PROPOSITION 4.1. *Suppose that* $\mathrm{med}(X)$ *and* $\mathrm{med}(Y)$ *are unique and that the cdf* H *of* $(X-\mathrm{med}(X))(Y-\mathrm{med}(Y))$ *is differentiable at* $\mathrm{COM}(X,Y)$ *with positive derivative* $h(\mathrm{COM}(X,Y)) > 0$. *Put* $\widetilde{\mathrm{COM}}_n(X,Y) := \mathrm{med}\{(X_i-\mathrm{med}(X))(Y_i-\mathrm{med}(Y)), i = 1,\ldots,n\}$, *where* $(X_1,Y_1)(X_2,Y_2),\ldots$ *are independent copies of* (X,Y). *Then*

$$\sqrt{n}(\widetilde{\mathrm{COM}}_n(X,Y) - \mathrm{COM}(X,Y)) \underset{\mathcal{D}}{\to} N\left(0, \frac{1}{4h^2(\mathrm{COM}(X,Y))}\right).$$

Let now (X,Y) have an arbitrary bivariate normal distribution with positive variances σ_X^2, σ_Y^2 of X and Y and correlation coefficient $\varrho \neq 0$. In this case, the preceding result implies

$$\sqrt{n}(\widetilde{\mathrm{COM}}_n(X,Y) - \mathrm{COM}(X,Y)) \underset{\mathcal{D}}{\to} N\left(0, \frac{\sigma_X^2\sigma_Y^2}{4h_\varrho^2(g(\varrho))}\right), \tag{4.7}$$

where h_ϱ is given in (2.5). If X and Y are however independent, Proposition 4.1 does not entail nondegenerate asymptotic normality of $\widetilde{\mathrm{COM}}_n(X,Y)$, since in this case $\varrho = 0$, $g(\varrho) = 0$ and thus, $h_\varrho(g(\varrho)) = \infty$, see the remarks after (2.5).

The independence case requires therefore an extra investigation, provided in the following result.

THEOREM 4.2. *Suppose that the cdfs F, G of X and Y are differentiable near* $\mathrm{med}(X)$, $\mathrm{med}(Y)$ *with derivatives f, g that are Hölder-continuous at* $\mathrm{med}(X)$ *and* $\mathrm{med}(Y)$, *respectively, and satisfy* $f(\mathrm{med}(X)) > 0$, $g(\mathrm{med}(Y)) > 0$. *Then we have*

$$\sqrt{n}\log(n)\widetilde{\mathrm{COM}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0, 1/(4f(\mathrm{med}(X))^2 g(\mathrm{med}(Y))^2)).$$

The proof of Theorem 4.2 indicates that $2\sqrt{n}\log(n)f(\mathrm{med}(X))g(\mathrm{med}(Y))(1+2\log\log(n)/\log(n))\widetilde{\mathrm{COM}}_n(X,Y)$ approaches the standard normal distribution much faster than $2\sqrt{n}\log(n)f(\mathrm{med}(X))g(\mathrm{med}(Y))\widetilde{\mathrm{COM}}_n(X,Y)$. This is also verified by numerous Monte Carlo simulations.

COROLLARY 4.2. *Suppose that X, Y are independent normal rvs with positive variances σ_X^2, σ_Y^2. Then, based on n independent copies of (X,Y), we have*

$$\frac{\sqrt{n}}{\pi}\log(n)\widetilde{\mathrm{COM}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0, \sigma_X^2\sigma_Y^2).$$

This result entails that in the case $\varrho = 0$, the empirical comedian $\widetilde{\mathrm{COM}}_n(X,Y)$ as an estimator of $\mathrm{COM}(X,Y) = 0$ outperforms the sample covariance $\widetilde{\mathrm{COV}}_n(X,Y)$ of normal vectors with known means, since $\sqrt{n}\widetilde{\mathrm{COV}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0, \sigma_X^2\sigma_Y^2)$.

Proposition 4.1 and Corollary 4.2 indicate that the empirical comedian is "superefficient" at the parameter $\rho = 0$. Whether $\widetilde{\mathrm{COM}}_n(X,Y)$ is actually superefficient in the sense defined for example in Section 8.6 in Pfanzagl (1994), requires knowledge about its limiting distribution along a sequence of alternatives $\rho_n \to \rho \in (-1,1)$. But this is an open problem to the best of our knowledge. The same applies to the sample comedian $\widehat{\mathrm{COM}}_n(X,Y)$ with estimated marginal medians defined below and to the test for $\rho = 0$ defined in (4.12).

PROOF. We utilize again the quantile transformation technique and assume the representation $X_iY_i = H^{-1}(U_i)$, $i = 1, \dots, n$, where $U_1, \dots, U_n$ are independent and uniformly on $(0,1)$ distributed rvs and H denotes the cdf of XY. Without loss of generality we can suppose $\mathrm{med}(X) = \mathrm{med}(Y) = 0$.

Fix $t \in \mathbb{R}$. From Example 4.1.1 in Reiss (1989) we obtain

$$\begin{aligned} &P\{\sqrt{n}\log(n)\widetilde{\mathrm{COM}}_n(X,Y) \le t\} \\ &\quad = P\{2\sqrt{n}(\bar{F}_n^{-1}(1/2) - 1/2) \le 2\sqrt{n}(H(ta_n) - 1/2)\} + o(1) \\ &\quad = \Phi(2\sqrt{n}(H(ta_n) - 1/2)) + o(1), \end{aligned}$$

where $\bar{F}_n$ denotes the empirical cdf of $U_1, \dots, U_n$ and $a_n := 1/(\sqrt{n}\log(n))$. It remains to show that for $t \in \mathbb{R}$

$$\sqrt{n}(H(ta_n) - 1/2) \underset{n\to\infty}{\longrightarrow} f(0)g(0)t. \tag{4.8}$$

Fix therefore $t \in \mathbb{R}$. Fubini's theorem implies

$$\begin{aligned}(4.9)\qquad &\sqrt{n}(H(ta_n) - 1/2)\\ &= \sqrt{n}\int_0^\infty G(ta_n/x)\mathcal{L}(X)dx - \sqrt{n}\int_{-\infty}^0 G(ta_n/x)\mathcal{L}(X)dx\\ &= \sqrt{n}\int_0^\infty G(ta_n/x) - G(0)\mathcal{L}(X)(dx)\\ &\quad - \sqrt{n}\int_{-\infty}^0 G(ta_n/x) - G(0)\mathcal{L}(X)(dx).\end{aligned}$$

Choose a large constant $K > 0$ such that t/K is so close to zero that it is in the domain of Hölder-continuity of g. Then we have

$$\begin{aligned}(4.10)\qquad &\sqrt{n}\int_0^\infty G(ta_n/x) - G(0)\mathcal{L}(X)(dx)\\ &= \sqrt{n}\int_{Ka_n}^\infty G(ta_n/x) - G(0)\mathcal{L}(X)(dx) + o(1)\\ &= \sqrt{n}\int_{Ka_n}^\infty g(\vartheta ta_n/x)ta_n/x\mathcal{L}(X)(dx) + o(1)\\ &= \sqrt{n}\int_{Ka_n}^\infty g(0)ta_n/x\mathcal{L}(X)(dx)\\ &\quad + \sqrt{n}\int_{Ka_n}^\infty (g(\vartheta ta_n/x) - g(0))ta_n/x\mathcal{L}(X)(dx) + o(1),\end{aligned}$$

where $\vartheta \in (0,1)$. Fix $\varepsilon > 0$ small such that the interval $[0,\varepsilon]$ is in the domain of Hölder-continuity of f. The first integral then equals for large n

$$\begin{aligned}(4.11)\qquad &\frac{g(0)t}{\log(n)}\int_{Ka_n}^\varepsilon \frac{1}{x}f(x)dx + o(1)\\ &= \frac{g(0)t}{\log(n)}\int_{Ka_n}^\varepsilon \frac{1}{x}(f(0) + (f(x) - f(0)))dx + o(1)\\ &= \frac{g(0)f(0)t}{\log(n)}\int_{Ka_n}^\varepsilon \frac{1}{x}dx + o(1)\\ &= \frac{g(0)f(0)t}{\log(n)}(\log(\varepsilon) - \log(Ka_n)) + o(1)\\ &= \frac{1}{2}g(0)f(0)t + o(1).\end{aligned}$$

By the same arguments one shows that the second integral in (4.10) is of order $o(1)$. Together with (4.10) we have thus shown that the first integral in (4.9) equals $g(0)f(0)t/2$. In complete analogy one shows that the second integral equals $-g(0)f(0)t/2$, which yields (4.8) and thus the assertion. □

In the next result we establish asymptotic normality of the empirical comedian in the case, where $\operatorname{med}(X)$ and $\operatorname{med}(Y)$ are unknown.

THEOREM 4.3. *Suppose in addition to the assumptions in Theorem 4.2 that the derivatives f, g are Hölder-continuous in a neighborhood of* $\mathrm{med}(X)$ *and* $\mathrm{med}(Y)$, *respectively. Then we have*

$$\sqrt{n}\log(n)\widehat{\mathrm{COM}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0, 1/(4f(\mathrm{med}(X))^2 g(\mathrm{med}(Y))^2)).$$

The variance of the limiting distribution of the comedian with unknown marginal medians is therefore the same as in the case with known medians.

Formulas (4.16)–(4.19) in the proof of Theorem 4.3 indicate that with $b_n := 1 + 2\log\log(n)/\log(n)$, the distribution of $2b_n f(\mathrm{med}(X))g(\mathrm{med}(Y))\widehat{\mathrm{COM}}_n(X,Y)$ approaches the standard normal distribution much faster than without the variance correction term b_n.

Theorem 4.3 offers a way to test independence of arbitrary normal rvs X, Y by means of the comedian. Denote again by $\widehat{\mathrm{MAD}}_n(X) = \mathrm{med}\{|X_i - \widehat{\mathrm{med}}_n(X)| : i = 1, \dots, n\}$ the empirical MAD with *estimated* median of X, based on n independent copies of X. From Corollary 3.1 we obtain $\widehat{\mathrm{MAD}}_n(X) \to_{n\to\infty} \mathrm{MAD}(X) = \sigma_X \Phi^{-1}(3/4)$ almost surely and thus, the following result is an immediate consequence of Theorem 4.3.

COROLLARY 4.3. *For independent but arbitrary normal rvs X, Y we have*

$$\sqrt{n}\log(n)\widehat{\mathrm{COM}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0, \pi^2\sigma_X^2\sigma_Y^2)$$

and thus,

$$\frac{\Phi^{-1}(3/4)^2}{\pi}\sqrt{n}\log(n)\frac{\widehat{\mathrm{COM}}_n(X,Y)}{\widehat{\mathrm{MAD}}_n(X)\widehat{\mathrm{MAD}}_n(Y)} \underset{\mathcal{D}}{\to} N(0,1).$$

Again, it is preferable to multiply the left hand side above by $b_n = (1 + 2\log\log(n)/\log(n))$ in order to accelerate the rate of convergence to $N(0,1)$. If we denote again by $\hat{\delta}_n = \widehat{\mathrm{COM}}_n(X,Y)/(\widehat{\mathrm{MAD}}_n(X)\widehat{\mathrm{MAD}}_n(Y))$ the empirical correlation median, then the preceding result implies that

$$\varphi_n(\hat{\delta}_n) := 1_{(\Phi^{-1}(1-\alpha/2)\pi/(b_n\Phi^{-1}(3/4)^2\sqrt{n}\log(n)),\infty)}(|\hat{\delta}_n|) \tag{4.12}$$

is an asymptotic level α-test for testing independence of normal rvs X and Y.

From Theorem 3.1 we know that $\widehat{\mathrm{COM}}_n(X,Y)$ converges to $\mathrm{COM}(X,Y)$ a.s. for an arbitrary bivariate normal vector (X,Y) and from (2.1) that $\mathrm{COM}(X,Y) = 0$ iff X, Y are independent. The asymptotic level α-test $\varphi_n(\hat{\delta}_n)$ defined above is therefore also asymptotically consistent. These considerations are summarized in the next result.

COROLLARY 4.4. *If $\hat{\delta}_n$ is based on n independent copies of an arbitrary bivariate normal vector (X,Y) we have*

$$E(\varphi_n(\hat{\delta}_n)) \underset{n\to\infty}{\longrightarrow} \begin{cases} \alpha & \text{if } X, Y \text{ are independent} \\ 1 & \text{otherwise.} \end{cases}$$

It would clearly be interesting to evaluate the power function of $\varphi_n(\delta_n)$ along a sequence of alternatives $(X^{(n)}, Y^{(n)})$ converging to the independent case. This would enable us to compare $\varphi_n(\delta_n)$ with other tests for independence, in particular that one based on the empirical correlation coefficient. But to this end, we need the asymptotic distribution of $\hat{\delta}_n$ in the dependent case i.e., along a sequence of alternatives converging to zero. To the best of our knowledge, this is an open problem.

The standardization of the empirical comedian by $\widehat{\mathrm{MAD}}_n(X)\widehat{\mathrm{MAD}}_n(Y)$ is however taylor made for normal rvs. A nonparametric i.e., distributional robust standardization of $\widehat{\mathrm{COM}}_n(X,Y)$ can be based on estimators of the marginal densities at $\mathrm{med}(X)$, $\mathrm{med}(Y)$.

An obvious choice can be based on the kernel estimator of the quantile density function $(F^{-1})'(q) = 1/f(F^{-1}(q))$ at $q = 1/2$. Put

$$\hat{f}_n(q) := \left(\int_0^1 F_n^{-1}(x)\alpha_n^{-2} k((q-x)/\alpha_n)dx \right)^{-1}, \tag{4.13}$$

where the bandwidth $\alpha_n > 0$ converges to zero and the kernel function k has bounded support and satisfies $\int k(x)dx = 0$, $\int xk(x)dx = -1$. This kernel estimator has been extensively studied in the last decade, see e.g. Xiang (1994) and the literature cited therein. In particular Theorem 2 in Falk (1986) implies that if $f = F'$ exists and is positive near $\mathrm{med}(X) = F^{-1}(1/2)$, then

$$\sqrt{n\alpha_n}\left(\frac{\hat{f}_n(1/2)}{\mathrm{med}(X)} - 1 \right) \underset{\mathcal{D}}{\to} N(0, \int K^2(y)dy),$$

where $K(y) = \int_{-\infty}^y k(x)dx$, provided $n\alpha_n^2 \to \infty$ and $n\alpha_n^3 \to 0$.

Consequently, if $n\alpha_n^2 \to \infty$ and $n\alpha_n^3 \to 0$, we obtain under the conditions of Theorem 4.3

$$2\hat{f}_n(1/2)\hat{g}_n(1/2)\sqrt{n}\log(n)\widehat{\mathrm{COM}}_n(X,Y) \underset{\mathcal{D}}{\to} N(0,1),$$

where $\hat{g}_n(1/2)$ is defined the same way as $\hat{f}_n(1/2)$ but with $F_n^{-1}(x)$ replaced by the empirical quantile function of $Y_1, \ldots, Y_n$.

In particular for the choice $k_E(x) = -3x/2$, $x \in [-1,1]$, $k_E(x) = 0$ elsewhere, which is the derivative of the Epanechnikov (1969) kernel $K_E(x) = 3/4(1-x^2)$, $x \in [-1,1]$ we obtain $\int K_E^2(y)dy = 3/5$ and, moreover, with $\alpha_n \in (0, 1/2)$

$$\hat{f}_n(1/2) \sim \frac{2n\alpha_n^3}{3\sum_{i=[n(1/2-\alpha_n)]}^{[n(1/2+\alpha_n)]}(i/n - 1/2)X_{i:n}},$$

where $X_{1:n} \le \cdots \le X_{n:n}$ denote the ordered values of $X_1, \ldots, X_n$. Equally, one can approximate $\hat{g}_n(1/2)$ with $X_{i:n}$ replaced by $Y_{i:n}$. These estimators are obviously quite robust with asymptotic breakdown point 1/2.

An alternative standardization of the comedian can be based on the kernel density estimator directly. Put

$$\tilde{f}_n(\widehat{\text{med}}_n(X)) := \frac{1}{n\alpha_n}\sum_{i=1}^{n} K\left(\frac{\widehat{\text{med}}_n(x) - X_i}{\alpha_n}\right), \tag{4.14}$$

where K has bounded support and satisfies $\int K(x)dx = 1$, $\int xK(x)dx = 0$. Notice that $k(x) := K'(x)$, $x \in \mathbb{R}$, is a kernel as above, provided K is continuously differentiable on $\mathbb{R}$ and has bounded support. The following result can be proved by utilizing the conditioning technique of the proof of Theorem 4.3 below. Here we assume that for an even sample size $n = 2m$ the empirical median $\widehat{\text{med}}_n(X)$ is defined as a fixed convex combination $cX_{m:n} + (1-c)X_{m+1:n}$, $c \in [0,1]$, of the most central order statistics $X_{m:n}$, $X_{m+1:n}$. The common choice is $c = 1/2$.

LEMMA 4.1. *Suppose that the cdf F of X is differentiable near $F^{-1}(1/2)$ with positive and Lipschitz-continuous derivative f. If K has bounded support and satisfies $\int K(x)dx = 1$, $\int xK^3(x)dx = 0$, $\int K^2(x)dx < \infty$, then*

$$\sqrt{n\alpha_n}\left(\frac{\tilde{f}_n(\widehat{\text{med}}_n(X))}{f(\text{med}(X))} - 1\right) \underset{\mathcal{D}}{\rightarrow} N\left(0, \frac{\int K^2(x)dx}{f^2(\text{med}(x))}\right),$$

provided $n\alpha_n^3 \to 0$, $n\alpha_n \to \infty$ as $n \to \infty$.

Different to $\hat{f}_n(1/2)$, the limiting normal distribution of $\tilde{f}_n(\widehat{\text{med}}_n(X))$ depends on the value $f(\text{med}(X))$. In case of the Epanechnikov kernels k_E and K_E, $\hat{f}_n(1/2)$ outperforms asymptotically the standardization $\tilde{f}_n(\widehat{\text{med}}_n(X))$ iff $f(\text{med}(X)) \le 1$.

PROOF OF THEOREM 4.3. The proof is based on iterated conditioning and on the quantile transformation. Suppose in the following $X_i = F^{-1}(U_i)$, $Y_i = G^{-1}(W_i)$, $i = 1, 2, \dots$, where $U_1, U_2, \dots$ and $W_1, W_2, \dots$ are independent sequences of independent and uniformly on (0,1) distributed rvs.

By $\bar{F}_n$, $\bar{G}_n$ we denote the empirical cdfs of $U_1, \dots, U_n$ and $W_1, \dots, W_n$, respectively. For simplicity we assume that $n = 2m + 1$ is odd. Fix $t \ge 0$ and put $a_n := 1/(\sqrt{n}\log(n))$. Then we have

$$\begin{aligned}
&P\{\sqrt{n}\log(n)\widehat{\text{COM}}_n(X,Y) \le t\}\\
&\quad = P\left\{\sum_{i=1}^{n} 1_{(-\infty, ta_n]}((X_i - F_n^{-1}(1/2))(Y_i - G_n^{-1}(1/2))) > m\right\}\\
&\quad = \int P\Bigg(\sum_{i=1}^{n} 1_{(-\infty, ta_n]}((F^{-1}(U_i) - F^{-1}(\bar{F}_n^{-1}(1/2)))\\
&\qquad\qquad \times (G^{-1}(W_i) - G^{-1}(\bar{G}_n^{-1}(1/2))))\\
&\qquad\qquad > m \mid \sqrt{n}(\bar{F}_n^{-1}(1/2) - 1/2) = u\Bigg)\mathcal{L}(\sqrt{n}(\bar{F}_n^{-1}(1/2) - 1/2))(du)
\end{aligned}$$

$$= \int_{-\sqrt{n}/2}^{\sqrt{n}/2} P\Bigg\{\sum_{i=1}^{m} 1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)}) - F^{-1}(1/2+u/\sqrt{n})) \times (G^{-1}(W_i) - G^{-1}(\bar{G}_n^{-1}(1/2)))) + \sum_{i=1}^{m} 1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)}) - F_n^{-1}(1/2+u/\sqrt{n})) \times (G^{-1}(W_i) - G^{-1}(\bar{G}_n^{-1}(1/2)))) \ge m\Bigg\} \times \mathcal{L}(\sqrt{n}(\bar{F}_n^{-1}(1/2) - 1/2))(du),$$

since conditional on $\bar{F}_n^{-1}(1/2) = 1/2 + u/\sqrt{n}$, the sample $U_1, \ldots, U_n$ consists of two independent samples of independent rvs $U_1^{(1)}, \ldots, U_m^{(1)}$ and $U_1^{(2)}, \ldots, U_m^{(2)}$ and the value $1/2+u/\sqrt{n}$, where the $U_i^{(1)}$ are uniformly on $(0, 1/2+u/\sqrt{n})$ distributed and the $U_i^{(2)}$ are uniformly on $(1/2+u/\sqrt{n}, 1)$ distributed. Moreover, these rvs are independent of $W_1, W_2, \ldots$

By repeating this argument and conditioning on $\bar{G}_n^{-1}(1/2) = 1/2 + w/\sqrt{n}$, the preceding integral equals

$$= \int_{-\sqrt{n}/2}^{\sqrt{n}/2}\int_{-\sqrt{n}/2}^{\sqrt{n}/2} P\Bigg\{\sum_{i=1}^{m} 1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)}) - F^{-1}(1/2+u/\sqrt{n})) \times (G^{-1}(W_i^{(r(i))}) - G_n^{-1}(1/2+w/\sqrt{n}))) + \sum_{i=1}^{m} 1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)}) - F^{-1}(1/2+m/\sqrt{n})) \times (G^{-1}(W_i^{(r(i+m))}) \quad G^{-1}(1/2+w/\sqrt{n}))) \ge m\Bigg\} \times \mathcal{L}(\sqrt{n}(\bar{G}_n^{-1}(1/2) - 1/2))(dw)\mathcal{L}(\sqrt{n}(\bar{F}_n^{-1}(1/2) - 1/2))(du),$$

where $r(j) \in \{0,1,2\}$, $W_i^{(1)}$ is uniformly on $(0, 1/2+w/\sqrt{n})$ distributed, $W_i^{(2)}$ is uniformly on $(1/2+w/\sqrt{n}, 1)$ distributed and $W_1^{(1)}, \ldots, W_m^{(1)}$, $W_1^{(2)}, \ldots, W_m^{(2)}$ are independent rvs, $W_i^{(0)} \equiv 1/2 + w/\sqrt{n}$. The random vector $\boldsymbol{r} = (r(1), \ldots, r(n))$ is a random permutation of m ones, m twos and one zero, and independent of the U_i and W_i.

The number N of ones among the $r(i)$ in the first sum in the preceding integral has therefore a hypergeometric distribution with parameters $2m+1$, m, m i.e.

$$P\{N = k\} = \binom{m}{k}\binom{m+1}{m-k} \Big/ \binom{2m+1}{m}, \qquad k = 0, 1, \ldots, m.$$

By conditioning on $N = k$, this approximation together with the asymptotic normality of sample quantiles in variational distance (see e.g. Theorem 4.1.4 in

Reiss (1989)) implies that the above multiple integral equals

$$\int_{-\sqrt{n}/2}^{\sqrt{n}/2}\int_{-\sqrt{n}/2}^{\sqrt{n}/2}\int P\Bigg\{\sum_{i=1}^{k}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n})))$$
$$+\sum_{i=k+1}^{m}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))$$
$$+\sum_{i=1}^{m-k}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n})))$$
$$+\sum_{i=m-k+1}^{m}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))\geq m\Bigg\}$$
$$\times\mathcal{L}(N)(dk)\mathcal{L}(\sqrt{n}(\bar{G}_n^{-1}(1/2)-1/2))(dw)$$
$$\times\mathcal{L}(\sqrt{n}(\bar{F}_n^{-1}(1/2)-1/2))(du)+o(1)$$

$$=\int_{-K_m}^{K_m}\int_{-K_m}^{K_m}\int_{m/2-K_m\sqrt{m}}^{m/2+K_m\sqrt{m}}$$
$$\times P\Bigg\{\sum_{i=1}^{k}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+u/\sqrt{n})))$$
$$+\sum_{i=k+1}^{m}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))$$
$$+\sum_{i=1}^{m-k}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)})-F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n})))$$
$$+\sum_{i=m-k+1}^{m}1_{(-\infty,ta_n]}((F^{-1}(1/2+u/\sqrt{n}))$$
$$\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))\geq m\Bigg\}$$
$$\times\mathcal{L}(N)(dk)N(0,1/(4g^2(G^{-1}(1/2))))(dw)$$

$$\times N(0,1/(4f^2(F^{-1}(1/2))))(du)+o(1),$$

where K_m tends to infinity such that $K_m/\sqrt{m}\to_{n\to\infty}0$. The single 0 in the vector $\boldsymbol{r}$, i.e. $W_j^{(0)}$, can be neglected asymptotically as the subsequent arguments show.

Recall that $t\geq 0$ and observe that with probability one for n large $(F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))(G^{-1}(W_j^{(2)})-G^{-1}(1/2+w/\sqrt{n}))<0$ and $(F^{-1}(U_i^{(2)})-F^{-1}(1/2+w/\sqrt{n}))(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n}))<0$ and thus, the preceding multiple integral equals

$$\begin{aligned}
(4.15)\ &\int_{-K_m}^{K_m}\int_{-K_m}^{K_m}\int_{m/2-K_m\sqrt{m}}^{m/2+K_m\sqrt{m}}\\
&\quad\times P\Bigg\{\sum_{i=1}^{k}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))\\
&\qquad\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n})))\\
&\qquad+\sum_{i=m-k+1}^{m}1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)})-F^{-1}(1/2+u/\sqrt{n}))\\
&\qquad\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))\geq 2k-m\Bigg\}\\
&\quad\times\mathcal{L}(N)(dk)N(0,1/(4g^2(G^{-1}(1/2))))(dw)\\
&\quad N(0,1/(4f^2(F^{-1}(1/2))))(du)\\
&=\int_{-K_m}^{K_m}\int_{-K_m}^{K_m}\int_{m/2-K_m\sqrt{m}}^{m/2+K_m\sqrt{m}}\\
&\quad\times P\Bigg\{k^{-1/2}\sum_{i=1}^{k}(1_{(-\infty,ta_n]}((F^{-1}(U_i^{(1)})-F^{-1}(1/2+u/\sqrt{n}))\\
&\qquad\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n})))-p_n)\\
&\qquad+k^{-1/2}\sum_{i=m-k+1}^{m}(1_{(-\infty,ta_n]}((F^{-1}(U_i^{(2)})-F^{-1}(1/2+u/\sqrt{n}))\\
&\qquad\times(G^{-1}(W_i^{(2)})-G^{-1}(1/2+w/\sqrt{n})))-q_n)\\
&\qquad+k^{1/2}(p_n+q_n)\geq 2k^{-1/2}(k-m/2)\Bigg\}\mathcal{L}(N)(dk)\\
&\quad\times N(0,1/(4g^2(G^{-1}(1/2))))(dw)N(0,1/(4f^2(F^{-1}(1/2))))(du),
\end{aligned}$$

where

$$\begin{aligned}
p_n&:=p_n(u,w)\\
&:=P\{(F^{-1}(U_1^{(1)})-F^{-1}(1/2+u/\sqrt{n}))\\
&\qquad\times(G^{-1}(W_i^{(1)})-G^{-1}(1/2+w/\sqrt{n}))\leq ta_n\}
\end{aligned}$$

and

$$\begin{aligned} q_n &:= q_n(u,w) \\ &:= P\{(F^{-1}(U_1^{(2)}) - F^{-1}(1/2 + u/\sqrt{n})) \\ &\qquad \times (G^{-1}(W_1^{(2)}) - G^{-1}(1/2 + w/\sqrt{n})) \le ta_n\}. \end{aligned}$$

From the Hölder-continuity of f and g near $F^{-1}(1/2)$, $G^{-1}(1/2)$ we obtain by Fubini's theorem the expansion

$$\begin{aligned}
(4.16)\quad p_n &= \int_{-\infty}^{0} P\{F^{-1}(U_1^{(1)}) \ge F^{-1}(1/2 + u/\sqrt{n}) + ta_n/y\} \\
&\quad \cdot \mathcal{L}(G^{-1}(W_1^{(1)}) - G^{-1}(1/2 + w/\sqrt{n}))(dy) \\
&= \int_{-\infty}^{-La_n} P\{U_1^{(1)} \ge F(F^{-1}(1/2 + u/\sqrt{n}) + ta_n/y)\} \\
&\quad \cdot \mathcal{L}(G^{-1}(W_1^{(1)}) - G^{-1}(1/2 + w/\sqrt{n}))(dy) + O(a_n) \\
&= \int_{-\infty}^{-La_n} P\left\{U_1^{(1)} \ge \frac{1}{2} + \frac{u}{\sqrt{n}} + f\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right) + \frac{\vartheta a_n}{y}\right)\frac{ta_n}{y}\right\} \\
&\quad \cdot \mathcal{L}(G^{-1}(W_1^{(1)}) - G^{-1}(1/2 + w/\sqrt{n}))(dy) + O(a_n) \\
&= \int_{-\infty}^{-La_n} 1 - \frac{\frac{1}{2} + \frac{u}{\sqrt{n}} + f\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right) + \frac{\vartheta a_n}{y}\right)\frac{ta_n}{y}}{\frac{1}{2} + \frac{u}{\sqrt{n}}} \\
&\quad \cdot \mathcal{L}(G^{-1}(W_1^{(1)}) - G^{-1}(1/2 + w/\sqrt{n}))(dy) + O(a_n) \\
&= \frac{tf\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right)\right)}{\frac{1}{2} + \frac{u}{\sqrt{n}}} a_n \\
&\quad \cdot \int_{-\infty}^{-La_n} \frac{1}{-y}\mathcal{L}\left(G^{-1}(W_1^{(1)}) - G^{-1}\left(\frac{1}{2} + \frac{w}{\sqrt{n}}\right)\right)(dy) \\
&\quad + \frac{t}{\frac{1}{2} + \frac{u}{\sqrt{n}}} a_n \\
&\quad \cdot \int_{-\infty}^{-La_n} \frac{1}{-y}\left(f\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right) + \frac{\vartheta a_n}{y}\right) - f\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right)\right)\right) \\
&\quad \cdot \mathcal{L}\left(G^{-1}(W_1^{(1)}) - G^{-1}\left(\frac{1}{2} + \frac{w}{\sqrt{n}}\right)\right)(dy) + O(a_n) \\
&= \frac{tf\left(F^{-1}\left(\frac{1}{2} + \frac{u}{\sqrt{n}}\right)\right)}{\frac{1}{2} + \frac{u}{\sqrt{n}}} a_n
\end{aligned}$$

$$\cdot \int_{-\varepsilon}^{-La_n} \frac{1}{-y} \frac{g\left(y+G^{-1}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)\right)}{\frac{1}{2}+\frac{w}{\sqrt{n}}} dy + O(a_n)$$

$$= \frac{tf\left(F^{-1}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\right) g\left(G^{-1}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)\right)}{\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)} a_n$$

$$\cdot \int_{-\varepsilon}^{-La_n} \frac{1}{-y} dy + O(a_n)$$

$$= \frac{tf\left(F^{-1}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\right) g\left(G^{-1}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)\right)}{2\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)} \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right),$$

where $L > 0$ is chosen large enough such that $F^{-1}(1/2 + u/\sqrt{n}) + ta_n/y$ is in the domain of Hölder-continuity of f near $F^{-1}(1/2)$, $\vartheta \in (0,1)$ and $\varepsilon > 0$ is chosen such that $y + G^{-1}(1/2 + w/\sqrt{n})$ is in the domain of Hölder-continuity of g for $y \in [-\varepsilon, -La_n]$. Equally one shows

$$q_n = \frac{tf\left(F^{-1}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\right) g\left(G^{-1}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)\right)}{2\left(\frac{1}{2}-\frac{u}{\sqrt{n}}\right)\left(\frac{1}{2}-\frac{w}{\sqrt{n}}\right)} \frac{1}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right). \tag{4.17}$$

(4.16) and (4.17) imply that uniformly for $k \in [m/2 - K_m\sqrt{m}, m/2 + K_m\sqrt{m}]$ and $u, w \in [-K_m, K_m]$, where $K_m \to \infty$ and $K_m/\sqrt{m} \to 0$ as $m \to \infty$

$$\begin{aligned} (4.18)\quad & k^{1/2}(p_n + q_n) \\ &= \left(\frac{m}{2}\right)^{1/2} (1+o(1))(p_n+q_n) \\ &= \left(\frac{m}{2}\right)^{1/2} \frac{t}{2n^{1/2}} f\left(F^{-1}\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\right) g\left(G^{-1}\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)\right) \\ &\quad \times \left(\frac{1}{\left(\frac{1}{2}+\frac{u}{\sqrt{n}}\right)\left(\frac{1}{2}+\frac{w}{\sqrt{n}}\right)} + \frac{1}{\left(\frac{1}{2}-\frac{u}{\sqrt{n}}\right)\left(\frac{1}{2}-\frac{w}{\sqrt{n}}\right)} + o(1) \right) \\ &\quad \times (1+o(1)) \\ &= 2tf(F^{-1}(1/2))g(G^{-1}(1/2)) + o(1). \end{aligned}$$

By the Berry-Esseen theorem, the triple integral in (4.15) equals therefore up to $o(1)$

$$\int_{-K_m}^{K_m} \int_{-K_m}^{K_m} \int_{m/2-K_m\sqrt{m}}^{m/2+K_m\sqrt{m}} \tag{4.19}$$

$$
\begin{aligned}
&\cdot P\{\sqrt{p_n(1-p_n)}\xi + \sqrt{q_n(1-q_n)}\eta \\
&\qquad + 2tf(F^{-1}(1/2))g(G^{-1}(1/2)) + o(1) \\
&\qquad \geq 2(m/2)^{-1/2}(k - m/2)(1 + o(1))\} \\
&\cdot \mathcal{L}(N)(dk)N(0, 1/(4g^2(G^{-1}(1/2))))(dw) \\
&\cdot N(0, 1/(4f^2(F^{-1}(1/2))))(du) \\
= &\iint_{m/2-K_m\sqrt{m}}^{m/2+K_m\sqrt{m}} P\{(p_n(1-p_n) + q_n(1-q_n))^{1/2}x \\
&\qquad + 2tf(F^{-1}(1/2))g(G^{-1}(1/2)) + o(1) \\
&\qquad \geq 2(m/2)^{-1/2}(k - m/2)(1 + o(1))\} \\
&\cdot \mathcal{L}(N)(dk)N(0,1)(dx)(1 + o(1)) \\
&\underset{m\to\infty}{\longrightarrow} \Phi(2tf(F^{-1}(1/2))g(G^{-1}(1/2)))
\end{aligned}
$$

by the convolution theorem for the normal distribution, where ξ and η are independent standard normal rvs, and the fact that $2(m/2)^{-1/2}(N - m/2)$ is asymptotically standard normal. This follows from (7.6) in Section VII.7 in Feller (1970). A negative t can be dealt with in complete analogy. This completes the proof of Theorem 4.3. □

Acknowledgements

The author is indebted to the referee for his careful reading of the manuscript and his numerous constructive remarks.

References

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates, *Ann. Math. Statist.*, **29**, 610–611.

Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density, *Theory Probab. Appl.*, **14**, 153–158.

Falk, M. (1986). On the estimation of the quantile density function, *Statist. Probab. Lett.*, **4**, 69–73.

Feller, W. (1970). *An Introduction to Probability Theory and Its Applications*, Vol. I. 3rd ed. (revised printing), Wiley, New York.

Feller, W. (1971). *An Introduction to Probability Theory and Its Aplications*, Vol. II, 2nd ed., Wiley, New York.

Hall, P. and Welsh, A. H. (1985). Limit theorems for the median deviation, *Ann. Inst. Statist. Math.*, **37**, 27–36.

Hampel, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.*, **69**, 383–397.

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

Pfanzagl, J. (1994). *Parametric Statistical Theory*, De Gruyter, Berlin.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics (With Applications to Nonparametric Statistics)*, Springer Series in Statistics, Springer, New York.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation, *J. Amer. Statist. Assoc.*, **88**, 1273–1283.

Rousseeuw, P. J. and Leroy, A. (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Rousseeuw, P. J. and Molenberghs, G. (1993). Transformation of non positive semidefinite correlation matrices, *Comm. Statist. Theory Methods*, **22**, 965–984.

Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.

Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts.

Welsh, A. H. (1986). Bahadur representations for robust scale estimators based on regression residuals, *Ann. Statist.*, **14**, 1246–1251.

Xiang, X. (1994). A law of the logarithm for kernel quantile density estimators, *Ann. Probab.*, **22**, 1078–1091.