Stat Methods Appl (2015) 24:569–596
DOI 10.1007/s10260-015-0315-x

On the efficiency of Gini's mean difference

Carina Gerstenberger[1] · **Daniel Vogel**[2]

Accepted: 23 April 2015 / Published online: 7 May 2015

Abstract The asymptotic relative efficiency of the mean deviation with respect to the standard deviation is 88 % at the normal distribution. In his seminal 1960 paper *A survey of sampling from contaminated distributions*, J. W. Tukey points out that, if the normal distribution is contaminated by a small ϵ-fraction of a normal distribution with three times the standard deviation, the mean deviation is more efficient than the standard deviation—already for $\epsilon < 1\,\%$. In the present article, we examine the efficiency of Gini's mean difference (the mean of all pairwise distances). Our results may be summarized by saying Gini's mean difference combines the advantages of the mean deviation and the standard deviation. In particular, an analytic expression for the finite-sample variance of Gini's mean difference at the normal mixture model is derived by means of the residue theorem, which is then used to determine the contamination fraction in Tukey's 1:3 normal mixture distribution that renders Gini's mean difference and the standard deviation equally efficient. We further compute the influence function of Gini's mean difference, and carry out extensive finite-sample simulations.

Keywords Influence function · Mean deviation · Median absolute deviation · Normal mixture distribution · Residue theorem · Robustness · Q_n · Standard deviation

Mathematics Subject Classification 62G35 · 62G05 · 62G20

JEL Classification C13

✉ Carina Gerstenberger
carina.gerstenberger@rub.de

1 Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany

2 Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, UK

1 Introduction

Let X be a random variable with distribution F, and define $F^{\star}_{a,b}$ as the distribution of $aX + b$. We call any function s that assigns a non-negative number to any univariate distribution F (potentially restricted to a subset of distributions, e.g. with finite second moments) a *measure of variability*, (or a *measure of dispersion* or simply a *scale measure*) if it satisfies

$$s\left(F^{\star}_{a,b}\right) = |a|\, s(F) \quad \text{for all } a, b \in \mathbb{R}. \tag{1}$$

In this article, our focus is on three very common descriptive measures of variability,

(i) the standard deviation $\sigma(F) = \{E(X - EX)^2\}^{1/2}$,
(ii) the mean absolute deviation (or mean deviation for short) $d(F) = E|X - md(F)|$, where $md(F)$ denotes the median of F, and
(iii) Gini's mean difference $g(F) = E|X - Y|$.

Here, X and Y are independent and identically distributed random variables with distribution function F. Recall that the variance can also be written as $\sigma^2(F) = E(X - Y)^2/2$. We define the median $md(F)$ as the center point of the set $\{x \in \mathbb{R} \mid F(x-) \leq 1/2 \leq F(x)\}$, where $F(x-)$ denotes the left-hand side limit. Suppose now we observe data $\mathbb{X}_n = (X_1, \ldots, X_n)$, where the X_i, $i = 1, \ldots, n$, are independent and identically distributed with cdf F. Let $\hat{F}_n$ be the corresponding empirical distribution function. The natural estimates for the above scale measures are the functionals applied to $\hat{F}_n$. However, we define the sample versions of the standard deviation and the mean deviation slightly different. Let

(i) $\sigma_n = \sigma_n(\mathbb{X}_n) = \left\{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2\right\}^{1/2}$
denote the sample standard deviation,
(ii) $d_n = d_n(\mathbb{X}_n) = \frac{1}{n-1}\sum_{i=1}^{n}|X_i - md(\hat{F}_n)|$ the sample mean deviation and
(iii) $g_n = g_n(\mathbb{X}_n) = \frac{2}{n(n-1)}\sum_{1 \leq i < j \leq n}|X_i - X_j|$ the sample mean difference.

While it is common practice to use $1/(n-1)$ instead of $1/n$ in the definition of the sample variance, due to the thus obtained unbiasedness, it is not so clear which finite-sample version of the mean deviation to use. The factor $1/(n-1)$ does generally not yield unbiasedness, but it leads to a significantly smaller bias in all our finite-sample simulations, see Sect. 4. Little appears to be known for which distributions d_n as defined above is indeed unbiased. The computation of $E(d_n)$ requires the knowledge of the expectations of the order statistics, which are known in principle, but generally rather cumbersome to evaluate analytically. An exception is the uniform distribution, where the order statistics are known to follow a beta distribution, and it turns out that d_n is unbiased for odd n, but not for even n. For details, see Lemma 1 in "Appendix". This is also in line with the simulation results reported in Table 7.

Furthermore, there is the question of the location estimator, which applies, in principle, to the mean deviation as well as to the standard deviation, and also to their population versions. While it is again established to use the mean along with the standard deviation, the picture is less clear for the mean deviation. We propose to use the median, mainly due to conceptual reasons: the median minimizes the mean deviation as the mean minimizes the standard deviation. This also suggests to apply the $1/(n-1)$ bias correction in both cases. However, our main results concern asymptotic efficiencies at symmetric distributions, for which the choice of the location measure as well as n versus $n-1$ question is largely irrelevant.

The standard deviation is, with good cause, the by far most popular measure of variability. One main reason for considering alternatives is its lack of robustness, i.e. its susceptibility to outliers and its low efficiency at heavy-tailed distributions. The two alternatives considered here are—in the modern understanding of the term—not robust, but they are more robust than the standard deviation. The extreme non-robustness of the standard deviation, which also emerges when comparing it with the mean deviation, played a vital role in recognizing the need for robustness and thus helped to spark the development of robust statistics, cf. e.g. Tukey (1960). The purpose of this article is to introduce Gini's mean difference into the old debate of mean deviation versus standard deviation (e.g. Gorard 2005)—not as a compromise, but as a consensus. We will argue that Gini's mean difference combines the advantages of the standard deviation and the mean deviation.

When proposing robust alternatives to any normality-based standard estimator, the gain in robustness is usually paid by a loss in efficiency at the normal model. The two aspects, robustness and efficiency, have to be analyzed and be put into relation with each other.

As far as the robustness properties are concerned, it is fairly easy to see that all three estimators have an asymptotic breakdown point of zero and an unbounded influence function. There are some slight advantages for the mean deviation and Gini's mean difference: their influence functions increase linearly as compared to the quadratic increase for the standard deviation, and they require only second moments to be asymptotically normal, as compared to the 4th moments for the standard deviation. The influence functions of the three estimators are given explicitly in Sect. 3. For the standard normal distribution, they are plotted (Fig. 2) and compared to the respective empirical sensitivity curves (Fig. 3). The influence function of Gini's mean difference appears to not have been published elsewhere.

However, the main concern in this paper is the efficiency of the estimators. We compute and compare their asymptotic variances at several distributions. We restrict our attention to symmetric distributions, since we are interested primarily in the effect of the tails of the distribution, which arguably have the most decisive influence on the behavior of the estimators. We consider in particular the t_ν distribution and the normal mixture distribution, which are both prominent examples of heavy-tailed distributions, and are often employed in robust statistics to investigate the behavior of estimators in heavy-tailed scenarios. To summarize our findings, in all relevant situations where Gini's mean difference does not rank first among the three estimators in terms of efficiency, it does rank second with very little difference to the respective winner. A more detailed discussion is deferred to Sect. 5.

We complement our findings by also giving the respective values for the median absolute deviation[1] (MAD, Hampel 1974) and the Q_n by Rousseeuw and Croux (1993). The sample version of the median absolute deviation, which we denote by $m_n = m_n(\mathbb{X}_n)$ is the median of the values $|X_i - md(\hat{F}_n)|$, $1 \le i \le n$, and the corresponding population value $m(F)$ is the median of the distribution of $|X - md(F)|$, where $X \sim F$. The Q_n scale estimator is the kth order statistic of the $\binom{n}{2}$ values $|X_i - X_j|$, $1 \le i < j \le n$, with $k = \binom{\lfloor n/2 \rfloor + 1}{2}$ and will be denoted by $Q_n(\mathbb{X}_n)$. Its population version $Q(F)$ is the lower quartile of the distribution of $|X - Y|$, where X and Y are independent with distribution F.[2] So for the MAD as well as the Q_n, we omit any consistency factors, which are often included to render them consistent for σ at the normal distribution. These can be deduced from Table 4. However, these estimators are included in the comparison, but not studied here in detail. For the derivation of the respective results, we will refer to the literature. We neither attempt a complete review of scale measures. For background information on robust scale estimation see, e.g., Huber and Ronchetti (2009, Chapter 5). A numerical study comparing many robust scale estimators is given, e.g., by Lax (1985).

The paper is organized as follows: In Sect. 2, asymptotic efficiencies of the scale estimators are compared. We study in particular their asymptotic variances at the normal mixture model. In Sect. 3, the influence functions are computed, and finite-sample simulations results are reported in Sect. 4. Section 5 contains a summary. Proofs are given in “Appendix”.

2 Asymptotic efficiencies

We gather the general expressions for the population values and asymptotic variances of the three scale measures (Sect. 2.1) and then evaluate them at several distributions (Sect. 2.2). We study the two-parameter family of the normal mixture model in some detail in Sect. 2.3.

2.1 General results

If $EX^2 < \infty$, Gini's mean difference and the mean deviation are asymptotically normal. For the asymptotic normality of σ_n, fourth moments are required. Strong consistency and asymptotic normality of g_n and σ_n^2 follow from general U-statistic theory (Hoeffding 1948), and thus for σ_n by a subsequent application of the continuous mapping theorem and the delta method, respectively.
Letting

$$d_n(\mathbb{X}_n, t) = \frac{1}{n-1} \sum_{i=1}^{n} |X_i - t|,$$

[1] Here, the choice of the location estimator is unambiguous: high breakdown point robustness is the main selling feature of the MAD.

[2] For simplicity, we define the p-quantile of distribution F as the value of the quantile function $F^{-1}(p) = \inf\{x | F(x) \le p\}$. For all population distributions we consider, there is no ambiguity, but note that $\hat{F}_n^{-1}(1/2)$ and the sample median $md(\hat{F}_n)$ as defined above are generally different.

the asymptotic normality of $d_n(\mathbb{X}_n, t)$ for any fixed location t holds also under the existence of second moments and is a simple corollary of the central limit theorem. Consistency and asymptotic normality of $d_n(\mathbb{X}_n, t_n)$, where t_n is a location estimator, is not equally straightforward (cf. e.g. Bickel and Lehmann 1976, Theorem 5 and the examples below). A set of sufficient conditions is that $\sqrt{n}(t_n - t)$ is asymptotically normal and F is symmetric around t. See also Babu and Rao (1992, Theorem 2.5).

Letting s_n be any of the estimators above and s the corresponding population value, we define the asymptotic variance $ASV(s_n) = ASV(s_n; F)$ of s_n at the distribution F as the variance of the limiting normal distribution of $\sqrt{n}(s_n - s)$, when s_n is evaluated at an independent sample $X_1, \ldots, X_n$ drawn from F. We note that, in general, convergence in distribution does not imply convergence of the second moments without further assumptions (uniform integrability), but it is usually the case in situations encountered in statistical applications. Specifically it is true for the estimators considered here, and we may write

$$ASV(s_n) = \lim_{n\to\infty} n \operatorname{var}(s_n).$$

We are going to compute asymptotic relative efficiencies of g_n and d_n with respect to σ_n. Generally, for two estimators a_n and b_n with $a_n \xrightarrow{p} \theta$ and $b_n \xrightarrow{p} \theta$ for some $\theta \in \mathbb{R}$, the asymptotic relative efficiency of a_n with respect to b_n at distribution F is defined as

$$ARE(a_n, b_n; F) = ASV(b_n; F)/ASV(a_n; F).$$

In order to make two scale estimators $s_n^{(1)}$ and $s_n^{(2)}$ comparable efficiency-wise, we have to standardize them appropriately, and define their asymptotic relative efficiency at the population distribution F as

$$ARE(s_n^{(1)}, s_n^{(2)}; F) = \frac{ASV(s_n^{(2)}; F)}{ASV(s_n^{(1)}; F)} \left\{ \frac{s^{(1)}(F)}{s^{(2)}(F)} \right\}^2, \tag{2}$$

where $s^{(1)}(F)$ and $s^{(2)}(F)$ denote the corresponding population values of the scale estimators $s_n^{(1)}$ and $s_n^{(2)}$, respectively.

The exact finite-sample variance of the empirical variance σ_n^2 is

$$\operatorname{var}(\sigma_n^2) = \frac{1}{n} \left\{ \mu_4 - 4\mu_3\mu_1 + 3\mu_2^2 - 2\sigma^4 \frac{2n-3}{n-1} \right\}, \tag{3}$$

where $\mu_k = EX^k$, $k \in \mathbb{N}$, is the kth non-central moment of X, in particular $\sigma^2 = \sigma^2(F) = \mu_2 - \mu_1^2$. Thus $ASV(\sigma_n^2) = \mu_4 + 3\mu_2^2 - 4\left\{\mu_3\mu_1 + \sigma^4\right\}$, and hence we have by the delta method

$$ASV(\sigma_n) = \frac{\mu_4 - 4\mu_3\mu_1 + 3\mu_2^2}{4\sigma^2} - \sigma^2. \tag{4}$$

Formula (3) appears to be a classical textbook example, but we did not find a reference for this general form. The special case $\mu_1 = 0$ is stated, e.g., in Kenney and Keeping (1952, p. 164).

If the distribution F is symmetric around $E(X) = \mu_1$ and has a Lebesgue density f, the mean deviation $d = d(F)$ can be written as

$$d = \int_{-\infty}^{\infty} |x - \mu_1| f(x)\, dx = 2 \int_{\mu_1}^{\infty} (x - \mu_1) f(x)\, dx \tag{5}$$

The asymptotic variance of d_n is $ASV(d_n) = \sigma^2 - d^2$. See, e.g., Pham-Gia and Hung (2001) for a review on the properties of the mean deviation.

For any F possessing a Lebesgue density f, Gini's mean difference $g = g(F)$ is

$$g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y|\, f(x)\, f(y)\, dy\, dx = 2 \int_{-\infty}^{\infty} \int_{x}^{\infty} (y - x)\, f(x)\, f(y)\, dy\, dx, \tag{6}$$

which can be further reduced to

$$g = 4 \int_{-\infty}^{\infty} \int_{x}^{\infty} y\, f(y)\, dy\, f(x)\, dx = 8 \int_{0}^{\infty} \int_{x}^{\infty} y\, f(y)\, dy\, f(x)\, dx \tag{7}$$

if F is symmetric around 0. Lomnicki (1952) gives the variance of the sample mean difference g_n as

$$\operatorname{var}(g_n) = \frac{1}{n(n-1)} \left\{ 4(n-1)\sigma^2 + 16(n-2)J - 2(2n-3)g^2 \right\}, \tag{8}$$

where

$$J = J(F) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{x} \int_{z=x}^{\infty} (x - y)(z - x) f(z) f(y) f(x)\, dz\, dy\, dx. \tag{9}$$

Thus, the asymptotic variance of g_n is $ASV(g_n) = 4\{\sigma^2 + 4J - g^2\}$.

2.2 Specific distributions

Table 1 lists the densities and first four moments of the following distribution families: normal, Laplace, uniform, t_ν and normal mixture.

The normal mixture distribution $NM(\lambda, \epsilon)$, sometimes also referred to as *contaminated normal distribution*, is defined as

$$NM(\lambda, \epsilon) = (1 - \epsilon) N(0, 1) + \epsilon N(0, \lambda^2), \quad 0 \le \epsilon \le 1, \lambda \ge 1.$$

For these distribution families, expressions for $\sigma(F)$, $d(F)$ and the asymptotic variances of their sample versions are given in Table 2, and for Gini's mean difference, including the integral J, in Table 3. The contents of Table 2 are well known and

Table 1 Densities and non-central moments of several parametric families

Distribution	Density $f(x)$	Parameters	Moments
Normal	$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$	$\mu \in \mathbb{R}, \sigma^2 > 0$	$\mu_1 = \mu,\ \mu_2 = \sigma^2 + \mu^2,$ $\mu_3 = \mu^3 + 3\mu\sigma^2,$ $\mu_4 = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$
Laplace	$\frac{1}{2\alpha} \exp\left\{\frac{-\|x-\mu\|}{\alpha}\right\}$	$\mu \in \mathbb{R}, \alpha > 0$	$\mu_1 = \mu,\ \mu_2 = \mu^2 + 2\alpha^2,$ $\mu_3 = \mu^3 + 6\alpha^2\mu,$ $\mu_4 = \mu^4 + 12\alpha^2\mu^2 + 24\alpha^4$
Uniform	$\frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$	$-\infty < a < b < \infty$	$\mu_1 = \frac{1}{2}(a+b),$ $\mu_2 = \frac{1}{3}\left\{(a+b)^2 - ab\right\},$ $\mu_3 = \frac{1}{4}(a+b)(a^2+b^2),$ $\mu_4 = \frac{1}{5}\left\{(a+b)(a^3+ab^2)+b^4\right\}$
t_ν	$c_\nu \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\nu > 0$	$\mu_1 = \mu_3 = 0,$ $\mu_2 = \nu/(\nu-2),$ $\mu_4 = 3\nu^2/\{(\nu-2)(\nu-4)\}$
Normal mixture	$\epsilon \frac{1}{\sqrt{2\pi}\lambda} \exp\{-\frac{x^2}{2\lambda^2}\} +$ $(1-\epsilon) \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$	$0 \le \epsilon \le 1,$ $\lambda \ge 1$	$\mu_1 = \mu_3 = 0,$ $\mu_2 = \epsilon\lambda^2 + (1-\epsilon),$ $\mu_4 = 3\epsilon\lambda^4 + 3(1-\epsilon)$

The scaling factor for the t_ν distribution is $c_\nu = \Gamma\left(\frac{\nu+1}{2}\right) \Big/ (\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2}))$

Table 2 Specific values of σ, d and the respective asymptotic variances for the distribution families given in Table 1

Distribution	$\sigma(F)$	$ASV(\sigma_n)$	$d(F)$	$ASV(d_n)$
Normal	σ	$\frac{\sigma^2}{2}$	$\frac{2\sigma}{\sqrt{2\pi}}$	$\sigma^2\left\{1 - \frac{2}{\pi}\right\}$
Laplace	$\sqrt{2}\alpha$	$\frac{5}{2}\alpha^2$	α	α^2
Uniform	$\frac{b-a}{2\sqrt{3}}$	$\frac{1}{60}(b-a)^2$	$\frac{b-a}{4}$	$\frac{1}{48}(b-a)^2$
t_ν	$\sqrt{\frac{\nu}{\nu-2}}$	$\frac{\nu(\nu-1)}{2(\nu-2)(\nu-4)}$	$\frac{2\nu c_\nu}{\nu-1}$	$\frac{\nu}{\nu-2} - \left\{\frac{2\nu c_\nu}{\nu-1}\right\}^2$
Normal mixture	$\sqrt{\epsilon\lambda^2 + (1-\epsilon)}$	$\{4(\epsilon\lambda^2+1-\epsilon)\}^{-1}$ $\{3(\epsilon\lambda^4+1-\epsilon)$ $-(\epsilon\lambda^2+1-\epsilon)^2\}$	$\sqrt{\frac{2}{\pi}}\{\epsilon\lambda +$ $(1-\epsilon)\}$	$\epsilon\lambda^2 + 1 - \epsilon$ $-\frac{2}{\pi}\{\epsilon\lambda + (1-\epsilon)\}^2$

$c_\nu = \Gamma\left(\frac{\nu+1}{2}\right) \Big/ \left(\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)\right)$

straightforward to derive. The results for Gini's mean difference require the evaluation of the integrals (7) and (9), which is non-trivial for many distributions. The expressions for the normal case are due to Nair (1936). Results for the normal mixture distribution and the t_ν are subject of the following two theorems.

Table 3 Population values, cf. (6), expressions for J, cf. (9), and resulting asymptotic variances for Gini's mean difference at the parametric families of Table 1

Distribution	$g(F)$	J	$ASV(g_n)$
Normal	$\frac{2\sigma}{\sqrt{\pi}}$	$\left(\frac{\sqrt{3}}{2\pi} - \frac{1}{6}\right)\sigma^2$	$\{\frac{4}{3} + \frac{8}{\pi}(\sqrt{3}-2)\}\sigma^2$
Laplace	$\frac{3}{2}\alpha$	$\frac{5}{24}\alpha^2$	$\frac{7}{3}\alpha^2$
Uniform	$\frac{1}{3}(b-a)$	$\frac{1}{120}(b-a)^2$	$\frac{1}{45}(b-a)^2$
t_ν	$\frac{4\sqrt{\nu}}{\nu-1}\frac{B\left(\frac{\nu}{2}+\frac{1}{2},\nu-\frac{1}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)B\left(\frac{\nu}{2},\nu\right)}$	$\frac{2\nu}{(\nu-1)^2}\frac{B\left(\frac{3\nu}{2}-1,\frac{1}{2}\right)}{B\left(\frac{\nu}{2},\frac{1}{2}\right)^3} - \frac{\nu}{2(\nu-2)} + K_\nu$	$4\{\sigma^2 + 4J - g^2\}$
Normal mixture	$\frac{2}{\sqrt{\pi}}\Big\{\lambda\epsilon^2 + (1-\epsilon)^2 + \epsilon(1-\epsilon)\sqrt{2(1+\lambda^2)}\Big\}$	$\left(\frac{1}{3} + \frac{\sqrt{3}}{2\pi}\right)\{\epsilon^3\lambda^2 + (1-\epsilon)^3\} - \frac{\epsilon\lambda^2+1-\epsilon}{2} + \epsilon^2(1-\epsilon)\Big[\frac{\lambda^2}{2} + \frac{1}{4} + \frac{3\lambda\zeta(\lambda)}{2\pi} + \frac{\lambda^2}{\pi}\operatorname{atan}\left(\frac{\lambda}{\zeta(\lambda)}\right) + \frac{1}{2\pi}\operatorname{atan}\left(\frac{1}{\lambda\zeta(\lambda)}\right)\Big] + \epsilon(1-\epsilon)^2\Big[\frac{\lambda^2}{4} + \frac{1}{2} + \frac{3\sqrt{1+2\lambda^2}}{2\pi} + \frac{\lambda^2}{2\pi}\operatorname{atan}\left(\frac{\lambda}{\zeta(1/\lambda)}\right) + \frac{1}{\pi}\operatorname{atan}\left(\frac{1}{\lambda\zeta(1/\lambda)}\right)\Big]$	$4\{\sigma^2 + 4J - g^2\}$

$\zeta(\lambda) = \sqrt{2+\lambda^2}$, $K_\nu = \int_{-\infty}^{\infty} x^2 f_\nu(x) F_\nu^2(x)\,dx$ with f_ν, F_ν being density and cdf of the t_ν distribution. $B(\cdot,\cdot)$ denotes the beta function

Theorem 1 *At the normal mixture distribution* $NM(\lambda, \epsilon)$, $0 \le \epsilon \le 1$, $\lambda \ge 1$, *the value of Gini's mean difference is*

$$g(NM(\lambda, \epsilon)) = \frac{2}{\sqrt{\pi}} \left\{ \lambda \epsilon^2 + (1-\epsilon)^2 + \epsilon(1-\epsilon)\sqrt{2\left(1+\lambda^2\right)} \right\}$$

and the value of the integral J, *cf.* (9), *is*

$$\begin{aligned} J(NM(\lambda, \epsilon)) \;=\; & \left(\frac{1}{3} + \frac{\sqrt{3}}{2\pi}\right)\{\epsilon^3\lambda^2 + (1-\epsilon)^3\} \;-\; \frac{\epsilon\lambda^2 + 1 - \epsilon}{2} \\ & + \epsilon^2(1-\epsilon)\left[\frac{\lambda^2}{2} + \frac{1}{4} + \frac{3\lambda\zeta(\lambda)}{2\pi} + \frac{\lambda^2}{\pi}\,\mathrm{atan}\left\{\frac{\lambda}{\zeta(\lambda)}\right\} + \frac{1}{2\pi}\,\mathrm{atan}\left\{\frac{1}{\lambda\zeta(\lambda)}\right\}\right] \\ & + \epsilon(1-\epsilon)^2\left[\frac{\lambda^2}{4} + \frac{1}{2} + \frac{3\sqrt{1+2\lambda^2}}{2\pi} + \frac{\lambda^2}{2\pi}\,\mathrm{atan}\left\{\frac{\lambda}{\zeta(1/\lambda)}\right\} + \frac{1}{\pi}\,\mathrm{atan}\left\{\frac{1}{\lambda\zeta(1/\lambda)}\right\}\right], \end{aligned}$$

where $\zeta(\lambda) = \sqrt{2+\lambda^2}$.

Theorem 2 *The value of Gini's mean difference at the* t_ν *distribution,* $\nu > 1$, *is*

$$g(t_\nu) = \frac{4\sqrt{\nu}}{\nu-1} \frac{B\left(\frac{\nu}{2}+\frac{1}{2},\, \nu-\frac{1}{2}\right)}{B\left(\frac{\nu}{2},\, \frac{1}{2}\right) B\left(\frac{\nu}{2},\, \nu\right)},$$

where $B(\cdot\,,\,\cdot)$ *denotes the beta function. The value of the integral* J, *cf.* (9), *at the* t_ν *distribution,* $\nu > 2$, *is*

$$J(t_\nu) = \frac{2\,\nu}{(\nu-1)^2} \frac{B\left(\frac{3\nu}{2}-1,\, \frac{1}{2}\right)}{B\left(\frac{\nu}{2},\, \frac{1}{2}\right)^3} - \frac{\nu}{2(\nu-2)} + \int_{-\infty}^{\infty} x^2 f_\nu(x) F_\nu^2(x)\, dx.$$

where F_ν *and* f_ν *are the cdf and the density, respectively, of the* t_ν *distribution.*

Resulting numerical values of the three scale measures and their asymptotic variances are listed in Tables 4 and 5. Table 6 contains the corresponding asymptotic relative efficiencies, cf. (2), with respect to the standard deviation.
In particular, we have at the normal model

$$\begin{aligned} ARE(g_n, \sigma_n) &= \left\{\frac{2}{3}\pi + 4(\sqrt{3}-2)\right\}^{-1} = 0.9779, \\ ARE(d_n, \sigma_n) &= \frac{1}{\pi-2} = 0.876, \end{aligned}$$

and at the Laplace (or double exponential) model

$$ARE(g_n, \sigma_n) = 135/112 = 1.2054, \quad ARE(d_n, \sigma_n) = 5/4.$$

Table 4 Values of the standard deviation σ, Gini's mean difference g, the mean absolute deviation d, the median absolute deviation m and the Q_n scale measure at the standard normal distribution $N(0, 1)$, the standard Laplace distribution $L(0, 1)$, the uniform distribution $U(0, 1)$ and several members of the t_ν family and the normal mixture family $NM(\lambda, \epsilon)$

Distribution	σ	g	d	m	Q
$N(0, 1)$	1	1.128	0.798	0.675	0.451
$L(0, 1)$	1.414	1.5	1	0.693	0.518
$U(0, 1)$	0.289	0.333	0.25	0.25	0.134
t_5	1.291	1.384	0.949	0.727	0.512
t_6	1.225	1.332	0.919	0.718	0.501
t_7	1.183	1.297	0.898	0.711	0.494
t_{10}	1.118	1.240	0.865	0.700	0.480
t_{15}	1.074	1.200	0.841	0.691	0.470
t_{16}	1.069	1.195	0.838	0.690	0.469
t_{25}	1.043	1.170	0.823	0.684	0.462
t_{40}	1.026	1.154	0.813	0.681	0.458
t_{41}	1.025	1.153	0.813	0.681	0.458
t_{100}	1.010	1.138	0.804	0.677	0.454
$NM(3, 0.008)$	1.032	1.151	0.811	0.679	0.457
$NM(3, 0.00175)$	1.007	1.133	0.801	0.675	0.452
$NM(3, 0.000309)$	1.001	1.129	0.798	0.675	0.451

Table 5 Asymptotic variances of the standard deviation σ_n, Gini's mean difference g_n, the mean absolute deviation d_n, the median absolute deviation m_n and the Q_n scale estimator at the standard normal distribution $N(0, 1)$, the standard Laplace distribution $L(0, 1)$, the uniform distribution $U(0, 1)$ and several members of the t_ν family and the normal mixture family $NM(\lambda, \epsilon)$

Distribution	$ASV(\sigma_n)$	$ASV(g_n)$	$ASV(d_n)$	$ASV(m_n)$	$ASV(Q_n)$
$N(0, 1)$	0.5	0.651	0.36	0.619	0.124
$L(0, 1)$	2.5	2.333	1	1	0.332
$U(0, 1)$	0.017	0.022	0.021	0.063	0.002
t_5	3.333	1.784	0.766	0.792	0.224
t_6	1.875	1.453	0.656	0.759	0.204
t_7	1.4	1.269	0.593	0.737	0.188
t_{10}	0.938	1.014	0.502	0.698	0.168
t_{15}	0.734	0.865	0.447	0.670	0.152
t_{16}	0.714	0.848	0.441	0.667	0.148
t_{25}	0.621	0.768	0.410	0.649	0.140
t_{40}	0.570	0.721	0.391	0.638	0.132
t_{41}	0.568	0.719	0.391	0.637	0.132
t_{100}	0.526	0.678	0.374	0.626	0.128
$NM(3, 0.008)$	0.890	0.791	0.407	0.628	0.132
$NM(3, 0.00175)$	0.590	0.682	0.373	0.621	0.125
$NM(3, 0.000309)$	0.516	0.656	0.365	0.619	0.124

Table 6 Asymptotic relative efficiencies of Gini's mean difference g_n, the mean absolute deviation d_n, the median absolute deviation m_n and the Q_n scale estimator at the standard normal distribution $N(0, 1)$, the standard Laplace distribution $L(0, 1)$, the uniform distribution $U(0, 1)$ and several members of the t_ν family and the normal mixture family $NM(\lambda, \epsilon)$

Distribution	$ARE(g_n, \sigma_n)$	$ARE(d_n, \sigma_n)$	$ARE(m_n, \sigma_n)$	$ARE(Q_n, \sigma_n)$
$N(0, 1)$	0.9779	0.876	0.3675	0.8206
$L(0, 1)$	1.2054	1.25	0.6006	1.0103
$U(0, 1)$	1	0.6	0.2	1.7950
t_5	2.1468	2.3514	1.3332	2.3403
t_6	1.5250	1.6071	0.8477	1.5374
t_7	1.3256	1.3607	0.6864	1.2985
t_{10}	1.1373	1.1163	0.5259	1.0292
t_{15}	1.0591	1.0064	0.4534	0.9248
t_{16}	1.0517	0.9954	0.4462	0.9286
t_{25}	1.0181	0.9440	0.4123	0.8703
t_{40}	1.0006	0.9156	0.3936	0.8605
t_{41}	1	0.9145	0.3929	0.8591
t_{100}	0.9862	0.8908	0.3773	0.8303
$NM(3, 0.008)$	1.3995	1.3511	0.6130	1.3220
$NM(3, 0.00175)$	1.0953	0.9998	0.4272	0.9510
$NM(3, 0.000309)$	0.9999	0.8988	0.3783	0.8442

The mean deviation (with scaling $1/n$) is the maximum likelihood estimator of the scale parameter α of the Laplace distribution, cf. Table 1. Thus, at the normal as well as the Laplace distribution, Gini's mean difference has an efficiency of more than 96 % with respect to the respective maximum likelihood estimator.

Furthermore, we observe that Gini's mean difference g_n is asymptotically more efficient than the standard deviation σ_n at the t_ν distribution for $\nu \leq 40$. The mean deviation d_n is asymptotically more efficient than σ_n for $\nu \leq 15$ and more efficient than g_n for $\nu \leq 8$. Thus in the range $9 \leq \nu \leq 40$, Gini's mean difference is the most efficient of the three.

One can view the uniform distribution as a limiting case of very light tails. While our focus is on heavy-tailed scenarios, we include the uniform distribution in our study as a simple approach to compare the estimators under light tails. We find a similar picture as under normality: Gini's mean difference and the standard deviation perform equally well, while the mean deviation has a substantially lower efficiency. However, it must be noted that the uniform distribution itself is rarely encountered in practice. The limited range is a very strong information, which allows a super-efficient inference.

The numerical results of Tables 1, 2 and 3 are rounded off by the respective values for the MAD and Q_n. Analytical expressions are generally not available for these estimators, and their population values and asymptotic variances are obtained from the general expressions given in Hall and Welsh (1985) and Rousseeuw and Croux (1993), respectively.

Finally, we take a closer look at the normal mixture distribution and explain our choices for λ and ϵ in Table 6.

2.3 The normal mixture distribution

The normal mixture distribution captures the notion that the majority of the data stems from the normal distribution, except for some small fraction ϵ which stems from another, usually heavier-tailed, contamination distribution. In case of the normal mixture model, this contamination distribution is the Gaussian distribution with standard deviation λ. This type of contamination model has been popularized by Tukey (1960), who also argues that $\lambda = 3$ is a sensible choice in practice.

It is sufficient to consider the case $\lambda \geq 1$, since the parameter pair (λ, ϵ) yields (up to scale) the same distribution as $(1/\lambda, 1 - \epsilon)$. Now, letting $\lambda > 1$, the case where ϵ is small is the interesting one. In this case the mixture distribution is heavy-tailed (measured, say, by the kurtosis) which strongly affects the behavior of our scale measures. The case ϵ close to 1 is of lesser interest: it corresponds to a normal distribution with a contamination concentrated at the origin, which affects the scale measures to a much lesser extent.

From the expressions for σ, d and the corresponding asymptotic variances, as given in Table 2, we obtain the asymptotic relative efficiency $ARE(d_n, \sigma_n)$ as a function of λ and ϵ. This function is plotted in Fig. 1 (top left). The parameter ϵ is on a log-scale since we are primarily interested in small contamination fractions. Fixing $\lambda = 3$, we find that for $\epsilon = 0.00175$, the mean deviation is as efficient as the standard deviation. It is interesting to note that Tukey (1960) gives a value of $\epsilon = 0.008$, which is frequently reported. In Huber and Ronchetti (2009, p. 3), correct values are given. The more precise value of 0.00175 is also in line with the simulation results of Sect. 4, and it supports even more so Tukey's main message: the percentage of contamination in the 1:3 normal mixture model for which the mean deviation becomes more efficient than the standard deviation is surprisingly low.

The asymptotic relative efficiency $ARE(g_n, \sigma_n)$ of Gini's mean difference with respect to the standard deviation is depicted in the upper right plot of Fig. 1. For $\lambda = 3$, Gini's mean difference is as efficient as the standard deviation for ϵ as small as 0.000309. In the lower plot of Fig. 1, equal-efficiency curves are drawn. They represent those parameter values (λ, ϵ) for which each two of the scale measures have equal asymptotic efficiency. So for instance, the solid black line corresponds to the contour line at height 1 of the 3D surface depicted in the top right plot.

3 Influence functions

The influence function $IF(\cdot, s, F)$ of a statistical functional s at distribution F is defined as

$$IF(x, s, F) = \lim_{\epsilon \searrow 0} \frac{1}{\epsilon} \{s(F_{\epsilon,x}) - s(F)\},$$

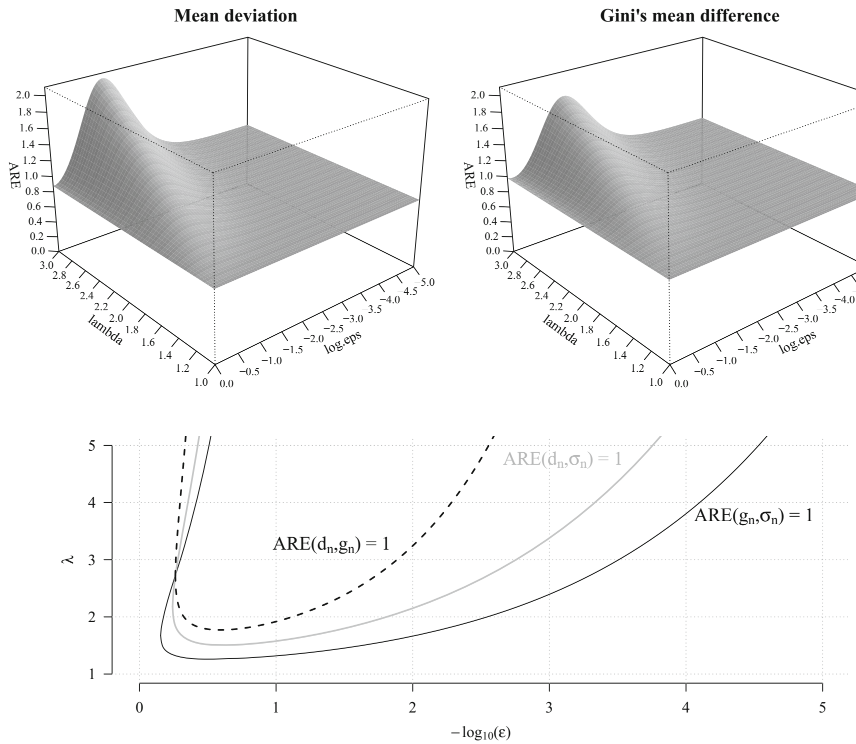

Fig. 1 *Top row* asymptotic relative efficiencies of the mean deviation (*left*) and Gini's mean difference (*right*) wrt the standard deviation in the normal mixture model as a function of λ and $\log(\epsilon)$. *Bottom* the curves for which values of λ and ϵ the scale measures have the same asymptotic efficiency

where $F_{\epsilon,x} = (1-\epsilon)F + \epsilon\Delta_x, 0 \le \epsilon \le 1, x \in \mathbb{R}$, and Δ_x denotes Dirac's delta, i.e., the probability measure that puts unit mass in x. The influence function describes the impact of an infinitesimal contamination at point x on the functional s if the latter is evaluated at distribution F. For further reading see, e.g., Huber and Ronchetti (2009) or Hampel et al. (1986). The influence functions of the standard deviation and the mean deviation are well known:

$$\begin{aligned} IF(x, \sigma(\cdot); F) &= (2\sigma(F))^{-1}\{(E(X)-x)^2 - \sigma^2(F)\}, \\ IF(x, d(\cdot); F) &= |x - md(F)| - d(F). \end{aligned}$$

For the formula for $d(\cdot)$ to hold in the last display, F has to fulfill certain regularity conditions in the vicinity of its median $md(F)$. Specifically, $(md(F_{\epsilon,x}) - md(F)) = O(\epsilon)$ as $\epsilon \to 0$ for all $x \in \mathbb{R}$ and $F(md(F_{\epsilon,x})) \to 1/2$ are a set of sufficient conditions. They are fulfilled, e.g., if F possesses a positive Lebesgue density in a neighborhood of $md(F)$. The influence function of Gini's mean difference appears to not have been published before.

Proposition 1 *The influence function of Gini's mean difference g at the distribution is*

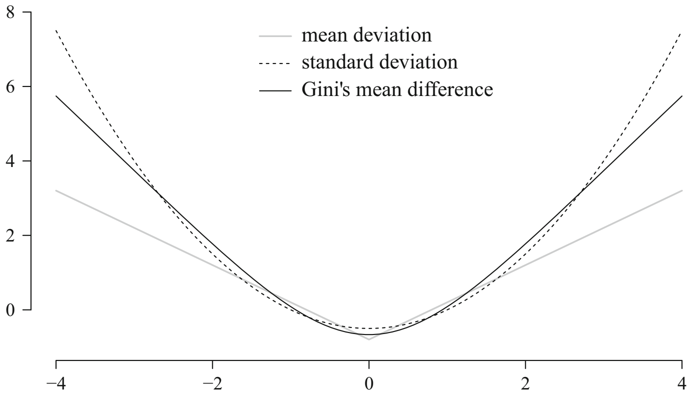

Fig. 2 Influence functions of the standard deviation, the mean deviation and Gini's mean difference at the standard normal distribution

$$IF(x, g(\cdot); F) = 2\left\{x[F(x) + F(x-) - 1] + E[X\mathbb{1}_{\{X\geq x\}}] - E[X\mathbb{1}_{\{X\leq x\}}] - g(F)\right\}.$$

For the standard normal distribution, these expressions for the influence functions of the three scale measures reduce to

$$\begin{aligned} IF(x, \sigma(\cdot); N(0,1)) &= (x^2 - 1)/2, \\ IF(x, d(\cdot); N(0,1)) &= |x| - \sqrt{2/\pi}, \\ IF(x, g(\cdot); N(0,1)) &= 4\phi(x) + 2x\{2\Phi(x) - 1\} - 4/\sqrt{\pi}, \end{aligned}$$

where ϕ and Φ denote the density and the cdf of the standard normal distribution, respectively. These curves are depicted in Fig. 2. Figure 3 shows empirical versions of the influence functions. Let $\mathbb{X}_n$ be a sample of size n drawn from $N(0, 1)$, and let $\mathbb{X}_n'(x)$ be the sample obtained from $\mathbb{X}_n$ by replacing the first observation by the value $x \in \mathbb{R}$. Then $n\{s_n(\mathbb{X}_n'(x)) - s_n(\mathbb{X}_n)\}$ is called a sensitivity curve for the estimator s_n (e.g. Huber and Ronchetti 2009, p. 15). Sensitivity curves usually strikingly resemble the corresponding influence function also for very moderate n. In Fig. 3, average sensitivity curves for σ_n, d_n and g_n are drawn (averaged over 10,000 samples of size $n = 100$). Figures 2 and 3 confirm the general impression mediated by Table 6 that Gini's mean difference is in-between the standard and the mean deviation, and support our claim that it combines the advantages of the other two: its influence function grows linearly for large $|x|$, but it is smooth at the origin.

The influence functions of the MAD and the Q_n can be found in Huber and Ronchetti (2009, p. 136) and Rousseeuw and Croux (1993), respectively.

4 Finite sample efficiencies

In a simulation study we want to check if the asymptotic efficiencies computed in Sect. 2 are useful approximations for the actual efficiencies in finite samples. For this purpose we consider the following nine distributions: the standard normal $N(0, 1)$, the

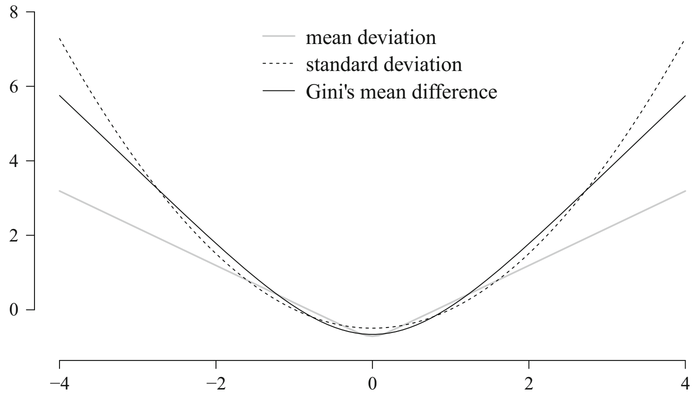

Fig. 3 Empirical influence functions (averaged sensitivity curves for $n = 100$, averaged over 10,000 samples) of the standard deviation, the mean deviation and Gini's mean difference at the standard normal distribution

standard Laplace $L(0, 1)$ (with parameters $\mu = 0$ and $\alpha = 1$, cf. Table 1), the uniform distribution $U(0, 1)$ on the unit interval, the t_ν distribution with $\nu = 5, 16, 41$ and the normal mixture with the parameter choices as in Tables 4, 5 and 6. The choice $\nu = 5$ serves as a heavy-tailed example, whereas for $\nu = 16$ and $\nu = 41$ we have witnessed at Table 6 that the mean deviation and the Gini mean difference, respectively, are asymptotically as efficient as the standard deviation.

For each distribution and each of the sample sizes $n = 5, 8, 10, 50, 500$, we generate 100,000 samples and compute from each sample five scale measures: the three moment-based estimators σ_n, d_n, g_n, and the two quantile-based estimators m_n and Q_n. The results for $N(0, 1)$, $L(0, 1)$ and $U(0, 1)$ are summarized in Table 7, for the t_ν distributions in Table 8, and for the normal mixture distributions in Table 9.

For each estimate, population distribution and sample size, the following numbers are reported: the sample variance of the 100,000 estimates multiplied by the respective value of n (the "n-standardized variance" which approaches the asymptotic variance given in Table 5 as n increases), the squared bias relative to the variance, and the relative efficiencies with respect to the standard deviation. With this information (variance and the squared-bias-to-variance ratio) the mean squared error is also implicitly given. For the relative efficiency computation, it is important to note that the standardization, cf. (2), is done not by the asymptotic value, but by the empirical finite-sample value, i.e. the sample mean of the 100,000 estimates. For Gini's mean difference, the simulated variances are also compared to the true finite-sample variances, cf. (8).

We observe the following: For large and moderate sample sizes ($n = 50, 500$), the simulated values are near the asymptotic ones from Tables 4, 5 and 6, and we may conclude that the asymptotic efficiency generally provides a useful indication for the actual efficiency in large samples, although to a much lesser extent for the quantile-based estimators.

In small samples, however, the simulated relative efficiencies may substantially differ from the asymptotic values. The ranking of the three moment-based estimators stays the same, but for the quantile-based estimators the picture is different: they exhibit

Table 7 Simulated variances, biases and relative efficiencies of σ_n, g_n, d_n, m_n and Q_n at $N(0,1)$, $L(0,1)$ and $U(0,1)$ for several sample sizes, d_n^*: mean deviation with $1/n$ scaling

Estimator	$n=5$	$n=8$	$n=10$	$n=50$	$n=500$
$N(0,1)$					
σ_n					
n·variance	0.577	0.548	0.541	0.507	0.505
bias^2/variance	0.031	0.019	0.014	0.003	1.0e−04
g_n					
n·variance (empirical)	0.850	0.767	0.743	0.666	0.655
n·variance (true)	0.852	0.766	0.740	0.667	0.653
bias^2/variance	3.4e−08	4.7e−07	7.8e−06	1.0e−05	4.7e−06
rel. efficiency wrt σ_n	0.986	0.982	0.980	0.979	0.978
d_n					
n·variance	0.482	0.454	0.427	0.374	0.365
bias^2/variance	0.009	0.020	0.012	0.001	1.2e−04
rel. efficiency wrt σ_n	0.938	0.902	0.894	0.880	0.876
d_n^*					
bias^2/variance	0.296	0.118	0.101	0.021	0.002
m_n					
n·variance	0.524	0.486	0.521	0.603	0.615
bias^2/variance	0.140	0.095	0.063	0.010	0.001
rel. efficiency wrt σ_n	0.385	0.431	0.415	0.375	0.370
Q_n					
n·variance	0.410	0.431	0.351	0.163	0.126
bias^2/variance	0.082	0.912	0.873	0.344	0.042
rel. efficiency wrt σ_n	0.453	0.619	0.634	0.746	0.810
$L(0,1)$					
σ_n					
n·variance	1.946	2.076	2.134	2.387	2.495
bias^2/variance	0.055	0.034	0.027	0.006	4.8e−04
g_n					
n·variance (empirical)	2.629	2.514	2.456	2.359	2.345
n·variance (true)	2.625	2.500	2.463	2.357	2.336
bias^2/variance	2.8e−06	8.4e−09	8.4e−08	1.3e−05	8.3e−10
rel. efficiency wrt σ_n	1.037	1.071	1.088	1.167	1.201
d_n					
n·variance	1.343	1.232	1.169	1.041	1.005
bias^2/variance	0.025	0.028	0.021	0.005	4.5e−04
rel. efficiency wrt σ_n	1.061	1.101	1.123	1.206	1.245
d_n^*					
bias^2/variance	0.106	0.040	0.031	0.006	0.001

Table 7 continued

Estimator	$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
m_n					
n·variance	0.979	0.881	0.897	0.975	0.994
bias^2/variance	5.9e−03	1.3e−03	7.6e−04	1.7e−05	2.4e−07
rel. efficiency wrt σ_n	0.537	0.627	0.622	0.599	0.610
Q_n					
n·variance	0.869	1.006	0.833	0.412	0.338
bias^2/variance	0.116	0.729	0.685	0.249	0.031
rel. efficiency wrt σ_n	0.607	0.798	0.827	0.944	1.004
$U(0, 1)$					
σ_n					
n·variance	0.031	0.025	0.022	0.018	0.017
bias^2/variance	0.021	0.010	0.007	0.001	2.6e−04
g_n					
n·variance (empirical)	0.045	0.035	0.032	0.024	0.023
n·variance (true)	0.044	0.035	0.032	0.024	0.022
bias^2/variance	1.9e−05	6.2e−07	9.4e−07	3.0e−05	5.1e−08
rel. efficiency wrt σ_n	0.985	0.967	0.962	0.985	0.998
d_n					
n·variance	0.030	0.028	0.026	0.022	0.021
bias^2/variance	6.1e−06	4.7e−03	2.3e−03	6.8e−05	1.7e−05
rel. efficiency wrt σ_n	0.829	0.694	0.672	0.614	0.603
d_n^*					
bias^2/variance	0.657	0.285	0.236	0.059	0.006
m_n					
n·variance	0.040	0.042	0.046	0.059	0.062
bias^2/variance	0.880	0.575	0.456	0.082	0.009
rel. efficiency wrt σ_n	0.275	0.272	0.255	0.210	0.200
Q_n					
n·variance	0.027	0.024	0.019	0.004	0.002
bias^2/variance	0.067	1.310	1.296	0.944	0.210
rel. efficiency wrt σ_n	0.340	0.483	0.500	0.963	1.875

quite a heavy bias for small samples, potentially of the same magnitude as the standard deviation of the estimator, complicating the comparison of the estimators. It is known that the finite-sample behavior, in terms of bias as well as variance, of robust quantile-based estimators in general may largely differ from the asymptotic approximation, particularly so in the case of the Q_n. Most striking certainly is the bias increase from $n = 5$ to $n = 8$ for the mean deviation d_n and, much more tremendously, for the Q_n. In case of the mean deviation, the reason lies in the different behavior of the sample median for odd and even numbers of observations, cf. also Lemma 1 in "Appendix".

Table 8 Simulated variances, biases and relative efficiencies of σ_n, g_n, d_n, m_n and Q_n at t_ν distributions for several sample sizes and values of ν; d_n^*: mean deviation with $1/n$ scaling

Estimator	$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
t_5					
σ_n					
n·variance	1.584	1.686	1.762	2.313	2.880
bias^2/variance	0.050	0.034	0.028	0.007	0.001
g_n					
n·variance (empirical)	2.050	1.942	1.890	1.805	1.790
n·variance (true)	2.047	1.935	1.901	1.806	1.787
bias^2/variance	4.0e−06	1.3e−05	2.5e−05	5.1e−06	1.4e−05
rel. efficiency wrt σ_n	1.073	1.150	1.185	1.499	1.811
d_n					
n·variance	1.036	0.949	0.901	0.791	0.760
bias^2/variance	0.014	0.018	0.014	0.003	2.3e−04
rel. efficiency wrt σ_n	1.105	1.208	1.282	1.673	1.977
d_n^*					
bias^2/variance	0.160	0.066	0.053	0.011	0.001
m_n					
n·variance	0.788	0.692	0.717	0.786	0.792
bias^2/variance	0.052	0.036	0.025	0.003	2.5e−04
rel. efficiency wrt σ_n	0.577	0.759	0.781	0.954	1.160
Q_n					
n·variance	0.678	0.741	0.611	0.282	0.228
bias^2/variance	0.101	0.808	0.750	0.294	0.036
rel. efficiency wrt σ_n	0.664	0.967	1.024	1.560	2.001
t_{16}					
σ_n					
n·variance	0.745	0.722	0.722	0.710	0.705
bias^2/variance	0.034	0.021	0.015	0.003	2.3e−04
g_n					
n·variance (empirical)	1.064	0.977	0.949	0.862	0.850
n·variance (true)	1.065	0.972	0.945	0.866	0.850
bias^2/variance	1.4e−07	5.0e−06	7.9e−07	7.6e−06	2.4e−05
rel. efficiency wrt σ_n	0.999	1.009	1.016	1.043	1.050
d_n					
n·variance	0.588	0.547	0.517	0.454	0.445
bias^2/variance	0.012	0.018	0.012	0.002	1.3e−04
rel. efficiency wrt σ_n	0.972	0.956	0.963	0.989	0.991
d_n^*					
bias^2/variance	0.259	0.106	0.085	0.017	0.002

Table 8 continued

Estimator	$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
m_n					
n·variance	0.596	0.546	0.574	0.649	0.667
bias^2/variance	0.105	0.069	0.051	0.007	0.001
rel. efficiency wrt σ_n	0.421	0.486	0.474	0.452	0.446
Q_n					
n·variance	0.480	0.513	0.421	0.198	0.154
bias^2/variance	0.089	0.877	0.822	0.329	0.039
rel. efficiency wrt σ_n	0.488	0.670	0.687	0.836	0.906
t_{41}					
σ_n					
n·variance	0.640	0.611	0.605	0.574	0.575
bias^2/variance	0.032	0.020	0.014	0.003	8.0e−05
g_n					
n·variance (empirical)	0.925	0.835	0.817	0.740	0.720
n·variance (true)	0.925	0.837	0.811	0.736	0.720
bias^2/variance	1.1e−05	3.6e−06	1.5e−06	9.5e−08	7.1e−07
rel. efficiency wrt σ_n	0.990	0.992	0.991	0.999	1.001
d_n					
n·variance	0.519	0.482	0.462	0.399	0.390
bias^2/variance	0.010	0.018	0.013	0.002	1.1e−04
rel. efficiency wrt σ_n	0.950	0.918	0.921	0.916	0.919
d_n^*					
bias^2/variance	0.276	0.113	0.094	0.019	0.002
m_n					
n·variance	0.555	0.509	0.538	0.617	0.638
bias^2/variance	0.120	0.086	0.059	0.008	0.001
rel. efficiency wrt σ_n	0.396	0.451	0.435	0.403	0.393
Q_n					
n·variance	0.439	0.457	0.377	0.172	0.137
bias^2/variance	0.087	0.910	0.867	0.338	0.041
rel. efficiency wrt σ_n	0.470	0.638	0.650	0.771	0.841

As for the Q_n, the definition of its sample version (see end of Sect. 1) also implies a qualitatively different behavior depending on whether n is odd or even. Specifically, for $n = 5$, the 3rd order statistic of 10 values is taken, whereas for $n = 8$, the 10th order statistic out of 28 observations is taken, both being compared to the 1/4 quantile of the respective population distribution. To reduce the bias as well as finite-sample variance, a smoothed version of the Q_n (i.e. a suitable linear combination of several order statistics) is certainly worth considering, for which the price to pay would be a small loss in the breakdown point.

Table 9 Simulated variances, biases and relative efficiencies of σ_n, g_n, d_n, m_n and Q_n at normal mixture distributions for $\lambda = 3$ and $\epsilon = 0.008, 0.00175, 0.000309$; d_n^*: mean deviation with $1/n$ scaling

Estimator	$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
$NM(3, 0.008)$					
σ_n					
n·variance	0.710	0.698	0.711	0.815	0.875
bias^2/variance	0.034	0.024	0.018	0.004	0.001
g_n					
n·variance (empirical)	0.997	0.910	0.876	0.804	0.790
n·variance (true)	0.996	0.908	0.882	0.808	0.793
bias^2/variance	4.6e−06	2.1e−10	1.6e−05	3.4e−06	8.4e−07
rel. efficiency wrt σ_n	1.023	1.060	1.083	1.257	1.385
d_n					
n·variance	0.540	0.507	0.480	0.423	0.405
bias^2/variance	0.010	0.016	0.013	0.002	1.6e−04
rel. efficiency wrt σ_n	1.000	1.016	1.039	1.204	1.332
d_n^*					
bias^2/variance	0.264	0.112	0.087	0.020	0.002
m_n					
n·variance	0.541	0.492	0.526	0.612	0.627
bias^2/variance	0.132	0.094	0.067	0.008	0.001
rel. efficiency wrt σ_n	0.442	0.538	0.527	0.562	0.601
Q_n					
n·variance	0.429	0.448	0.367	0.168	0.133
bias^2/variance	0.079	0.877	0.832	0.300	0.005
rel. efficiency wrt σ_n	0.523	0.760	0.779	1.092	1.312
$NM(3, 0.00175)$					
σ_n					
n·variance	0.617	0.587	0.576	0.573	0.590
bias^2/variance	0.032	0.019	0.017	0.003	3.0e−04
g_n					
n·variance (empirical)	0.889	0.791	0.764	0.704	0.675
n·variance (true)	0.883	0.797	0.771	0.698	0.684
bias^2/variance	1.6e−07	3.0e−07	4.8e−08	1.8e−05	1.0e−05
rel. efficiency wrt σ_n	0.995	1.002	1.009	1.056	1.092
d_n					
n·variance	0.500	0.462	0.441	0.385	0.370
bias^2/variance	0.011	0.017	0.013	0.002	3.9e−05
rel. efficiency wrt σ_n	0.951	0.931	0.931	0.971	0.992
d_n^*					
bias^2/variance	0.283	0.115	0.100	0.022	0.003

Table 9 continued

Estimator	$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
m_n					
n·variance	0.532	0.491	0.519	0.602	0.623
bias^2/variance	0.133	0.092	0.068	0.010	0.001
rel. efficiency wrt σ_n	0.397	0.457	0.441	0.418	0.424
Q_n					
n·variance	0.415	0.433	0.353	0.163	0.128
bias^2/variance	0.082	0.911	0.876	0.333	0.031
rel. efficiency wrt σ_n	0.470	0.648	0.670	0.831	0.935
$NM(3, 0.000309)$					
σ_n					
n·variance	0.584	0.558	0.543	0.517	0.515
bias^2/variance	0.031	0.017	0.014	0.003	3.2e−04
g_n					
n·variance (empirical)	0.853	0.775	0.744	0.667	0.655
n·variance (true)	0.857	0.771	0.746	0.673	0.658
bias^2/variance	1.3e−05	4.8e−06	5.1e−07	1.3e−06	8.3e−06
rel. efficiency wrt σ_n	0.986	0.986	0.985	0.993	0.999
d_n					
n·variance	0.484	0.452	0.434	0.375	0.365
bias^2/variance	0.009	0.018	0.012	0.002	1.8e−04
rel. efficiency wrt σ_n	0.941	0.900	0.903	0.899	0.903
d_n^*					
bias^2/variance	0.291	0.122	0.096	0.021	0.002
m_n					
n·variance	0.527	0.484	0.517	0.600	0.618
bias^2/variance	0.133	0.096	0.068	0.009	0.001
rel. efficiency wrt σ_n	0.388	0.439	0.421	0.384	0.379
Q_n					
n·variance	0.414	0.432	0.351	0.161	0.126
bias^2/variance	0.085	0.919	0.866	0.347	0.042
rel. efficiency wrt σ_n	0.459	0.626	0.643	0.764	0.835

We also include the mean deviation with factor $1/n$ instead of $1/(n-1)$ in the study, denoted by d_n^* in the tables. Since d_n and d_n^* differ only by multiplicative factor, the efficiencies are the same, and we only report the (squared) bias (relative to the variance). We find that d_n^* is quite heavily biased for small samples for all distributions considered, whereas d_n has in all situations a smaller bias than σ_n. Particularly, note that the bias of d_n at the uniform distribution is indeed zero for $n = 5$, but not for even n, cf. Lemma 1 in "Appendix".

Finally, the simulations confirm the unbiasedness of Gini's mean difference and the formula (8), due to Lomnicki (1952), for its finite-sample variance.

The simulations were done in R (R Development Core Team 2010), using the function Qn() from the package robustbase (Rousseeuw et al. 2014), the function mad() from the standard package stats, and an implementation for Gini's mean difference by A. Azzalini.[3] The default setting for both functions Qn() and mad() is to multiply the result by the asymptotic consistency factor for the standard deviation at normality, which is, for both functions, controlled by the parameter `constant`. This parameter was set to 1 in our simulations.

5 Summary and discussion

Several authors have argued that, when comparing the standard deviation with the mean deviation, the better robustness of the latter is a crucial advantage, which outweighs its disadvantages, and that the mean deviation is hence to be preferred out of the two. We share this view. However, we recommend to use Gini's mean difference instead of the mean deviation. While it has qualitatively the same robustness and the same efficiency under long-tailed distributions as the mean deviation, it lacks its main disadvantage as compared the standard deviation: the lower efficiency at strict normality. For near-normal distributions—and also for very light-tailed distribution, as the results for the uniform distribution suggest—Gini's mean difference and the standard deviation are for all practical purposes equally efficient. For instance, at the normal and all t_ν distributions with $\nu \geq 23$, the (properly standardized) asymptotic variances of g_n and σ_n are within a 3 % margin of each other. At heavy-tailed distributions, Gini's mean difference is, along with the mean deviation, substantially more efficient than the standard deviation.

To summarize our efficiency comparison, Gini's mean difference performs well over a wide range of distributions, including much heavier than normal tails. Here we basically consider the range up to the t_5 distribution, where no higher than fourth moments exist, and within this range, Gini's mean difference is clearly non-inferior to all competitors considered here.

However, the main advantage of Gini's mean difference is its finite-sample performance. First of all, being a U-statistic, it is unbiased—at all distributions with finite first moments. We do not know any other scale measure satisfying (1) of practical relevance for which this is true. Second, its finite-sample variance is known, which allows for instance better approximative confidence intervals. Neither of that is true for the standard deviation, and one can consequently argue that Gini's mean difference is a superior scale estimator even under strict normality. The latter statement is also a remark on the discussion by Yitzhaki (2003), who compares Gini's mean difference primarily to the variance.

When comparing Gini's mean difference to the mean deviation, both being similar L_1-type measures, the question arises, if an intuitive explanation can be given to why the former is more efficient at the normal distribution but less efficient at heavy tails.

[3] https://stat.ethz.ch/pipermail/r-help/2003-April/032820.html.

We leave this as an open question here. However, since Gini's mean difference can be viewed as a symmetrized version of the mean deviation, we remark that a similar effect can be observed in many instances of symmetrization. Other examples include the Hodges–Lehmann location estimator as a symmetrized version of the median, or Kendall's tau as a symmetrized verion of the quadrant correlation. In both cases, the original estimator has a rather low efficiency under normality, which is considerably increased by symmetrization, but the symmetrized estimator performs slightly worse at very heavy-tailed models. The median, for instance, is more efficient than the Hodges–Lehmann estimator at a t_3 distribution. But in general, symmetrization is a successful technique to increase the efficiency of highly robust estimators while retaining a large degree of robustness. The most prominent example may be the Q_n, the symmetrized version of the MAD.

Acknowledgments We are indebted to Herold Dehling for introducing us to the theory of U-statistics, to Roland Fried for introducing us to robust statistics, and to Alexander Dürre, who has demonstrated the benefit of complex analysis for solving statistical problems. Both authors were supported in part by the Collaborative Research Centre 823 *Statistical modelling of nonlinear dynamic processes*.

Appendix 1: Proofs

Towards the proof of Theorem 1, we spare a few words about the derivation of the corresponding result for the normal distribution. When evaluating the integral J, cf. (9), for the standard normal distribution, one encounters the integral

$$I_1 = \int_{-\infty}^{\infty} x^2 \phi(x) \Phi(x)^2 dx,$$

where ϕ and Φ denote the density and the cdf of the standard normal distribution, respectively. Nair (1936) gives the value $I_1 = 1/3 + 1/(2\pi\sqrt{3})$, resulting in $J = \sqrt{3}/(2\pi) - 1/6$, but does not provide a proof. The author refers to the derivation of a similar integral (integral 8 in Table I, Nair 1936, p. 433), where we find the result as well as the derivation doubtful, and to an article by Hojo (1931), which gives numerical values for several integrals, but does not contain an explanation for the value of I_1 either. We therefore include a proof here. Writing $\Phi(x)$ as the integral of its density and changing the order of the integrals in thus obtained three-dimensional integral yields

$$I_1 = (2\pi)^{-3/2} \int_{y=-\infty}^{0} \int_{z=\infty}^{0} \int_{x=-\infty}^{\infty} x^2 e^{x^2/2} e^{(y+x)^2/2} e^{(z+x)^2/2} \, dx \, dz \, dy.$$

Solving the inner integral, we obtain

$$I_1 = (18\pi\sqrt{3})^{-1} \int_{y=0}^{\infty} \int_{z=0}^{\infty} [(y+z)^2 + 3] \exp\left\{ -\frac{1}{3} \left[y^2 + z^2 - yz \right] \right\} \, dz \, dy.$$

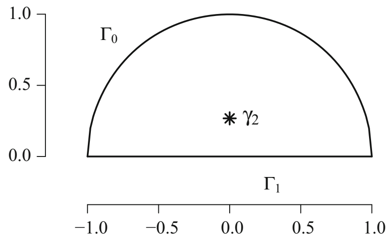

Fig. 4 Residue theorem: the line integral over h along the closed curve $\Gamma = \Gamma_0 \cup \Gamma_1$ is determined by the residue of h in γ_2

Introducing polar coordinates α, r such that $y = r\cos\alpha$, $z = r\sin\alpha$, and solving the integral with respect to r, we arrive at

$$I_1 = \frac{1}{4\pi\sqrt{3}} \int_{\alpha=0}^{\pi} \frac{4+\sin\alpha}{(2-\sin\alpha)^2}\, d\alpha.$$

This remaining integral may be solved by means of the residue theorem (e.g. Ahlfors 1966, p. 149). Substituting $\gamma = e^{i\alpha}$ and using $\sin\alpha = (e^{i\alpha} - e^{-i\alpha})/(2i)$, we transform I_1 into the following line integral in the complex plane,

$$I_1 = \frac{1}{4\pi\sqrt{3}} \int_{\Gamma_0} \frac{\gamma^2 + 8i\gamma - 1}{(\gamma^2 - 4i\gamma - 1)^2}\, d\gamma, \qquad (10)$$

where Γ_0 is the upper unit half circle in the complex plane, cp. Fig. 4. Let us call h the integrand in (10), its poles (both of order two) are $\gamma_{1/2} = (2 \pm \sqrt{3})i$, so that γ_2 lies within the closed upper half unit circle Γ. The residue of h in γ_2 is $-\sqrt{3}i/2$. Integrating h along Γ_1, i.e. the real line from -1 to 1, cf. Fig. 4, and applying the residue theorem to the closed line integral along Γ completes the derivation.

Proof (Proof of Theorem 1)

Evaluating the integral J for the normal mixture distribution, we arrive after lengthy calculations at

$$\begin{aligned} J = &\Big[\epsilon^3\lambda^2 + (1-\epsilon)^3\Big]\Big[2A(1) + C(1) + E(1)\Big] - (\epsilon\lambda^2 + 1 - \epsilon)B \\ &+ \epsilon^2(1-\epsilon)\Big[2(2+\lambda^2)A(1/\lambda) + C(\lambda) + 2\lambda^2 D(1/\lambda) + \lambda(2+\lambda^2)E(1/\lambda)\Big] \\ &+ \epsilon(1-\epsilon)^2\Big[2(2\lambda^2+1)A(\lambda) + \lambda^2 C(1/\lambda) + 2D(\lambda) + (\lambda^{-1} + 2\lambda)E(\lambda)\Big], \end{aligned}$$

where

$$\begin{aligned} A(\lambda) &= \int_{-\infty}^{\infty} x\phi^2(x)\Phi(x/\lambda)dx = \frac{1}{4\pi\sqrt{1+2\lambda^2}}, \\ B &= \int_{-\infty}^{\infty} x^2\phi(x)\Phi(x)dx = \frac{1}{2}, \\ C(\lambda) &= \int_{-\infty}^{\infty} x^2\phi(x)\Phi^2(x/\lambda)dx = \frac{1}{4} + \frac{\lambda}{\pi(1+\lambda^2)\sqrt{2+\lambda^2}} \end{aligned}$$

$$+\frac{1}{2\pi}\arctan\left(\frac{1}{\lambda\sqrt{2+\lambda^2}}\right),$$

$$D(\lambda) = \int_{-\infty}^{\infty} x^2\phi(x)\Phi(x)\Phi(x/\lambda)dx \;=\; \frac{1}{4} + \frac{3\lambda^2+1}{4\pi(1+\lambda^2)\sqrt{2\lambda^2+1}}$$
$$+\frac{1}{2\pi}\arctan\left(\frac{1}{\sqrt{2\lambda^2+1}}\right),$$

$$E(\lambda) = \int_{-\infty}^{\infty} \phi^2(x)\phi(x/\lambda)dx \;=\; \frac{1}{2\pi\sqrt{1+2\lambda^2}},$$

for all $\lambda > 0$. As before, ϕ and Φ denote the density and the cdf of standard normal distribution. The tricky integrals are $C(\lambda)$ and $D(\lambda)$, which, for $\lambda = 1$, both reduce to the integral I_1 above. Proceeding as before for the integral I_1, solving the respective two inner integrals yields

$$C(\lambda) = \frac{\lambda^3}{2\pi\sqrt{2+\lambda^2}}\int_0^{\pi/2}\frac{3+\lambda^2+\sin(2\alpha)}{\{1+\lambda^2-\sin(2\alpha)\}^2}d\alpha,$$

$$D(\lambda) = \frac{1}{2\pi\sqrt{1+2\lambda^2}}\int_0^{\pi/2}\frac{2+\lambda^2(2+\sin(2\alpha))+(3\lambda^4-\lambda^2-2)\sin^2(\alpha)}{\{2-\sin(2\alpha)+(\lambda^2-1)\sin^2(\alpha)\}^2}d\alpha.$$

These integrals are again solved by the residue theorem, which completes the proof. □

For the proof of Theorem 2, the following identities are helpful:

$$\int x\left(1+\tfrac{x^2}{\beta}\right)^{\alpha}dx \;=\; \tfrac{\beta}{2(\alpha+1)}\left(1+\tfrac{x^2}{\beta}\right)^{\alpha+1}, \quad \alpha\neq -1,\ \beta\neq 0. \tag{11}$$

$$\int_{-\infty}^{\infty}\left(1+\tfrac{x^2}{\nu}\right)^{-\nu}dx \;=\; \tfrac{1}{c_{2\nu-1}}\sqrt{\tfrac{\nu}{2\nu-1}}, \quad \nu > 0, \tag{12}$$

$$\int_{-\infty}^{\infty}\left(1+\tfrac{x^2}{\nu}\right)^{-\frac{3\nu-1}{2}}dx \;=\; \tfrac{1}{c_{3\nu-2}}\sqrt{\tfrac{\nu}{3\nu-2}}, \quad \nu > 0, \tag{13}$$

where c_ν is the scaling factor of the t_ν density, cf. Table 1. The identities (12) and (13) can be obtained by transforming the respective left-hand sides into a t_ν-densities by substituting $y = ((2\nu-1)/\nu)^{1/2}x$ and $y = ((3\nu-2)/\nu)^{1/2}x$, respectively.

Proof (Proof of Theorem 2) For computing g, we evaluate (7), successively making use of (11) and (12), and obtain

$$g \;=\; 4\frac{\nu c_\nu^2}{\nu-1}\int_{-\infty}^{\infty}\left(1+\frac{x^2}{\nu}\right)^{-\nu}dx \;=\; \frac{4\,\nu^{3/2}c_\nu^2}{(\nu-1)\sqrt{2\nu-1}\,c_{2\nu-1}},$$

which can be written as in Theorem 2 by using $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. For evaluating J, we write J as $J = \int_{\mathbb{R}} A(x) f_\nu(x)\, dx$ with f_ν being the t_ν density and

$$\begin{aligned} A(x) &= \int_{-\infty}^{x}\int_{x}^{\infty} xz f_\nu(z) f_\nu(y)\, dz\, dy \;-\; \int_{-\infty}^{x}\int_{x}^{\infty} yz f_\nu(z) f_\nu(y)\, dz\, dy \\ &\quad - \int_{-\infty}^{x}\int_{x}^{\infty} x^2 f_\nu(z) f_\nu(y)\, dz\, dy \;+\; \int_{-\infty}^{x}\int_{x}^{\infty} xy f_\nu(z) f_\nu(y)\, dz\, dy \\ &= A_1(x) - A_2(x) - A_3(x) + A_4(x). \end{aligned}$$

Using (11), we obtain

$$A_1(x) + A_4(x) \;=\; \frac{c_\nu\, \nu\, x}{\nu - 1}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu-1}{2}} \int_{-x}^{x} f_\nu(y)\, dy,$$

and

$$-A_2(x) \;=\; \left(\frac{c_\nu\, \nu}{\nu - 1}\right)^2 \left(1 + \frac{x^2}{\nu}\right)^{-\nu+1}.$$

Hence, $J = B_1 + B_2 - B_3$ with

$$\begin{aligned} B_1 &= \int_{-\infty}^{\infty} \frac{c_\nu\, \nu\, x}{\nu - 1}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu-1}{2}} f_\nu(x) \int_{-x}^{x} f_\nu(y)\, dy\, dx, \\ B_2 &= \int_{-\infty}^{\infty} \left(\frac{c_\nu\, \nu}{\nu-1}\right)^2 \left(1 + \frac{x^2}{\nu}\right)^{-\nu+1} f_\nu(x)\, dx, \\ B_3 &= \int_{-\infty}^{\infty} x^2 F_\nu(x)\,(1 - F_\nu(x))\, f_\nu(x)\, dx \;=\; \frac{\nu}{2(\nu-2)} - \int_{-\infty}^{\infty} x^2 f_\nu(x) F_\nu^2(x)\, dx, \end{aligned}$$

where F_ν is the cdf of the t_ν distribution. By employing (11) and (13), we find

$$B_1 \;=\; B_2 \;=\; \frac{2}{c_{3\nu-2}} \left(\frac{c_\nu\, \nu}{\nu - 1}\right)^2 \sqrt{\frac{\nu}{3\nu - 2}}$$

and arrive, again by employing $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$, at the expression for J given in Theorem 2. □

The remaining integral

$$K_\nu \;=\; \int_{-\infty}^{\infty} x^2 f_\nu(x) F_\nu^2(x)\, dx$$

cannot be solved by the same means as the analogous integral I_1 for the normal distribution, and we state this as an open problem. However, this one-dimensional integral can easily be approximated numerically, and the expression is quickly entered into a mathematical software like R (R Development Core Team 2010).

Proof (Proof of Proposition 1) We have

$$g(F_{\epsilon,x}) = 2\int_{-\infty}^{\infty}\int_{y}^{\infty}(z-y)\,dF_{\epsilon,x}(z)\,dF_{\epsilon,x}(y),$$
$$= (1-\epsilon)^2 g(F) + 2\epsilon(1-\epsilon)\int_{-\infty}^{\infty}(x-z)\left\{\mathbb{1}_{(-\infty,x]}(y) - \mathbb{1}_{[x,\infty)}(y)\right\}\,dF(y)$$

and hence

$$IF(x, g(\cdot); F) = \lim_{\epsilon\searrow 0}\frac{1}{\epsilon}\{g(F_{\epsilon,x}) - g(F)\}$$
$$= -2g(F) + 2\left\{x[F(x) + F(x-) - 1] + E[X\mathbb{1}_{\{X\geq x\}}] - E[X\mathbb{1}_{\{X\leq x\}}]\right\},$$

which completes the proof. □

With the influence function known, it is also possible use the relationship

$$ASV(s_n; F) = \int_{\mathbb{R}} IF(x, s, F)^2 F(dx)$$

instead of referring to the terms given in Sect. 2 to compute the asymptotic variance of the estimators. This leads to the same integrals.

Appendix 2: Miscellaneous

Lemma 1 *For $X_1, \ldots, X_n$ being independent and $U(a, b)$ distributed for $a, b \in \mathbb{R}$, $a < b$, we have for the sample mean deviation (about the median)*

$$E(d_n) = \begin{cases} (b-a)/4 & \textit{for odd } n\ (n \geq 3), \\ \dfrac{b-a}{4}\dfrac{n^2}{n^2-1} & \textit{for even } n. \end{cases}$$

Proof For notational convenience we restrict our attention to the case $a = 0, b = 1$. Let $X_{(i)}$ denote the ith order statistic, $1 \leq i \leq n$. The random variable $X_{(i)}$ has a Beta(α, β) distribution with parameters $\alpha = i$ and $\beta = n+1-i$, and hence $E(X_{(i)}) = i/(n+1)$. If n is odd, we write d_n as $d_n = (n-1)^{-1}\sum_{i=1}^{\lfloor n/2\rfloor}(X_{(n+1-i)} - X_{(i)})$ and obtain

$$E(d_n) = \frac{1}{n-1}\sum_{i=1}^{\lfloor n/2\rfloor}\left(\frac{n+1-i}{n+1} - \frac{i}{n+1}\right) = \frac{1}{4}.$$

If n is even, we have $d_n = (n-1)^{-1}\sum_{i=1}^{n/2}(X_{(n+1-i)} - X_{(i)})$, and hence

$$E(d_n) = \frac{1}{n-1}\sum_{i=1}^{n/2}\left(\frac{n+1-i}{n+1} - \frac{i}{n+1}\right) = \frac{n^2}{4(n^2-1)},$$

which completes the proof. □

References

Ahlfors LV (1966) Complex analysis, 2nd edn. McGraw-Hill, New York
Babu GJ, Rao CR (1992) Expansions for statistics involving the mean absolute deviations. Ann Inst Stat Math 2(44):387–403
Bickel PJ, Lehmann EL (1976) Descriptive statistics for nonparametric models, III. Dispersion. Ann Stat 6:1139–1148
Gorard S (2005) Revisiting a 90-year-old debate: the advantages of the mean deviation. Br J Educ Stud 4:417–430
Hall P, Welsh A (1985) Limit theorems for the median deviation. Ann Inst Stat Math 1(37):27–36
Hampel FR (1974) The influence curve and its role in robust estimation. J Am Stat Assoc 69:383–393
Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics. The approach based on influence functions. Wiley series in probability and mathematical statistics. Wiley, New York
Hoeffding W (1948) A class of statistics with asymptotically normal distribution. Ann Math Stat 19:293–325
Hojo T (1931) Distribution of the median, quartiles and interquartile distance in samples from a normal population. Biometrika 3–4(23):315–360
Huber PJ, Ronchetti EM (2009) Robust statistics. Wiley series in probability and statistics, 2nd edn. Wiley, Hoboken
Kenney F, Keeping E (1952) Mathematics of statistics. Part two. D. Van Nostrand Company, Inc., Princeton
Lax DA (1985) Robust estimators of scale: finite-sample performance in long-tailed symmetric distributions. J Am Stat Assoc 391(80):736–741
Lomnicki ZA (1952) The standard error of Gini's mean difference. Ann Math Stat 4(23):635–637
Nair US (1936) The standard error of Gini's mean difference. Biometrika 28:428–436
Pham-Gia T, Hung T (2001) The mean and median absolute deviations. Math Comput Model 7(34):921–936
R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http://www.R-project.org/. ISBN:3-900051-07-0
Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Manuel Koller MM (2014) robustbase: basic robust statistics. . R package version 0.91-1. http://CRAN.R-project.org/package=robustbase
Rousseeuw PJ, Croux C (1993) Alternatives to the median absolute deviation. J Am Stat Assoc 424(88):1273–1283
Tukey JW (1960) A survey of sampling from contaminated distributions. In: Olkin I et al (eds) Contributions to probability and statistics. Essays in honor of Harold Hotteling. Stanford University Press, Stanford, pp 448–485
Yitzhaki S (2003) Gini's mean difference: a superior measure of variability for non-normal distributions. Metron 2(61):285–316