

## ON THE EFFICIENCY OF GINI'S MEAN DIFFERENCE

CARINA GERSTENBERGER AND DANIEL VOGEL

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  $\epsilon$ -fraction of a normal distribution with three times the standard deviation, the mean deviation is more efficient than the standard deviation—already for  $\epsilon < 1\%$ . This came as a surprise to most statisticians at the time, and the publication is today considered as one of the main pioneering works in the development of robust statistics. In the present article, we examine the efficiency of the mean deviation and Gini's mean difference (the mean of all pairwise distances). The latter is known to have an asymptotic relative efficiency of 98% at the normal distribution. Our findings support the viewpoint that Gini's mean difference combines the advantages of the mean deviation and the standard deviation. We also answer the question, what percentage of contamination in Tukey's 1:3 normal mixture model renders Gini's mean difference more efficient than the standard deviation.

2010 MSC: 62G35, 62G05, 62G20

## 1. INTRODUCTION

Let  $X$  be a random variable with distribution  $F$ , and define  $F_{a,b}^*$  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

$$s(F_{a,b}^*) = |a| s(F) \quad \text{for all } a, b \in \mathbb{R}.$$

In this article, we compare 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)  $d(F) = E|X - m(F)|$ , where  $m(F)$  denotes the median of  $F$ , and
- (iii) Gini's mean difference  $g = E|X - Y|$ .

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  $m(F)$  as the center point of the set  $\{x \in \mathbb{R} | F(x-) \leq 1/2 \leq F(x)\}$ , where  $F(x-)$  denotes the left-hand side limit. Suppose now we observe data  $\mathbb{X}_n = (X_1, \dots, X_n)$ , where the  $X_i$ ,  $i = 1, \dots, n$ , 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 differently. Let

---

*Date:* May 21, 2014.

*Key words and phrases.* influence function, mean deviation, normal mixture distribution, robustness, residue theorem, standard deviation.

$$(i) \sigma_n = \sigma_n(\mathbb{X}_n) = \left\{ \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right\}^{1/2} = \left\{ \frac{1}{n(n-1)} \sum_{1 \leq i < j \leq n} (X_i - X_j)^2 \right\}^{1/2}$$

denote the sample standard deviation,

$$(ii) d_n = d_n(\mathbb{X}_n) = \frac{1}{n-1} \sum_{i=1}^n |X_i - m(\hat{F}_n)| \quad \text{the sample mean deviation and}$$

$$(iii) g_n = g_n(\mathbb{X}_n) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |X_i - X_j| \quad \text{the sample mean difference.}$$

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 what finite-sample version of the mean deviation to use. Unfortunately, the factor  $1/(n-1)$  does not yield unbiasedness for any distribution, as it is the case for the variance, but it leads to a significantly smaller bias in all our finite-sample simulations, see Section 4.

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 simple  $1/(n-1)$  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  $n$  vs.  $n-1$  question is irrelevant.

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

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

the asymptotic normality of  $d_n(\mathbb{X}, t)$  for any fixed location  $t$  holds also under the existence of second moments and is a simple corollary of the central limit theorem. The 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  $t$ .

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 vs. 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. The theoretical robustness properties of the three estimators are quickly summarized: they all 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 all three estimators at the standard normal distribution are plotted in Figure 2.

We are thus left to study their efficiencies. This is the main concern in this paper. We compute and compare the asymptotic variances of the estimates 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  $t_\nu$  distribution and the normal mixture distribution, which are both popular outlier models in robust statistics. 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 Section 5.

The rest of the paper is organized as follows: In Section 2, asymptotic efficiencies of the scale estimators are compared. We study in particular their asymptotic variances at the normal mixture model. In Section 3, the influence functions are computed. We complement our findings with finite-sample simulations in Section 4. Section 5 contains a summary. Remarks on the computation of the asymptotic variances are given in the Appendix.

We close this section by introducing some further terms and notation. 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  $X_1, \dots, X_n$  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

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

We are going to compute asymptotic relative efficiencies of  $g_n$  and  $d_n$  with respect to  $\sigma_n$ . Generally, for two estimators  $a_n$  and  $b_n$  with  $a_n \xrightarrow{p} \mu$  and  $b_n \xrightarrow{p} \mu$  for some  $\mu \in \mathbb{R}$ , 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 the scale estimators comparable efficiency-wise, we introduce the  $\sigma$ -standardized versions of the estimators,

$$\tilde{g}_n = \frac{\sigma}{g} g_n, \quad \tilde{d}_n = \frac{\sigma}{d} d_n.$$

These estimators are of no practical use, since they require the knowledge of the parameter  $\sigma$ , which they aim to estimate, but since they estimate the same quantity as  $\sigma_n$ , we may compare their asymptotic variances. We then define the asymptotic relative efficiency of  $s_n$  (where  $s_n$  may be any scale estimator) with respect to the standard deviation at the population distribution  $F$  as

$$(1) \quad ARE(s_n, \sigma_n; F) = ARE(\tilde{s}_n, \sigma_n; F) = \frac{ASV(\sigma_n; F) s^2(F)}{ASV(s_n; F) \sigma^2(F)}.$$

Also, if we compare the efficiencies of two scale estimators  $s_n^{(1)}$  and  $s_n^{(2)}$ , the comparison shall refer to their  $\sigma$ -standardized versions.

## 2. ASYMPTOTIC EFFICIENCIES

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

**2.1. General expressions.** The exact finite-sample variance of the empirical variance  $\sigma_n^2$  is

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

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

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

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

$$(3) \quad d = \int_{-\infty}^{\infty} |x - \mu_1| f(x) dx = 2 \int_{\mu_1}^{\infty} (x - \mu_1) f(x) dx$$

The asymptotic variances of  $d_n$  and its  $\sigma$ -standardized version  $\tilde{d}_n$  are  $ASV(d_n) = \sigma^2 - d^2$  and  $ASV(\tilde{d}_n) = \sigma^2 \{\sigma^2/d^2 - 1\}$ , respectively. For any  $F$  possessing a Lebesgue density  $f$ , Gini's mean difference  $g = g(F)$  is

$$(4) \quad g = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x - y| f(x) f(y) dy dx = 2 \int_{-\infty}^{\infty} \int_x^{\infty} (y - x) f(x) f(y) dy dx,$$

which can be further reduced to

$$(5) \quad g = 4 \int_{-\infty}^{\infty} \int_x^{\infty} y f(y) dy f(x) dx = 8 \int_0^{\infty} \int_x^{\infty} y f(y) dy f(x) dx$$

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

$$(6) \quad \text{var}(g_n) = \frac{1}{n(n-1)} \{4(n-1)\sigma^2 + 16(n-2)J - 2(2n-3)g^2\},$$

where

$$(7) \quad J = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^x \int_{z=x}^{\infty} (x-y)(z-x)f(z)f(y)f(x) dz dy dx.$$

Thus, the asymptotic variances of  $g_n$  and its  $\sigma$ -standardized version  $\tilde{g}_n$  are  $ASV(g_n) = 4\{\sigma^2 + 4J - g^2\}$  and  $ASV(\tilde{g}_n) = 4\sigma^2\{(\sigma^2 + 4J)/g^2 - 1\}$ , respectively.

**2.2. Specific distributions.** Table 1 lists the densities and first four moments of the following distribution families: normal, Laplace, uniform,  $t_\nu$  and normal mixture. The resulting 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. While the contents of Table 2 are straightforward and stated here without proof, the results for Gini's mean difference require the evaluation of the integrals (5) and (7), which is non-trivial for many distributions. Details for the  $t_\nu$  and the normal mixture distribution are given in the Appendix. The expressions for the normal case are due to Nair (1936). For convenience, resulting numerical values of the three scale

measures and their asymptotic variances are listed in Table 4. Table 5 contains the asymptotic relative efficiencies, cf. (1). In particular, we have at the normal model

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

and at the Laplace (or double exponential) distribution

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

Thus, in both situations, 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  $\sigma_n$  at  $t_\nu$  distribution for  $\nu \leq 40$ . The mean deviation  $d_n$  is asymptotically more efficient than  $\sigma_n$  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.

We also include the interquartile range (the distance between the upper and the lower quartile) in the efficiency comparison of Table 5 — without examining this estimator in detail. The purpose is to give a rough impression of how the numbers given compare to another well-known scale measure. This comparison, though, must be into perspective with two aspects: Firstly, the interquartile range is primarily used as a descriptive statistic for data sets rather than an estimator for a true population value. Secondly, the interquartile range is much more robust, it has a bounded influence function and a breakdown point of about 0.25. Also, there are other highly robust scale measures which are more efficient than the interquartile range, for instance the median absolute deviation (MAD, Hampel, 1974) or the  $Q_n$  by Rousseeuw and Croux (1993). We do not attempt to give a complete review, which is clearly beyond the scope of this paper.

Finally, we take a closer look at the normal mixture distribution and explain our choices for  $\lambda$  and  $\epsilon$  in Table 5.

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

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

The resulting density is given in Table 1. The normal mixture distribution is a popular model in robust statistics. It captures the notion that the majority of the data stems from the normal distribution, except for some small fraction  $\epsilon$  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  $\lambda$ . 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  $(\lambda, \epsilon)$  yields (up to scale) the same distribution as  $(1/\lambda, 1 - \epsilon)$ . Now, letting  $\lambda > 1$ , the case where  $\epsilon$  is small is the interesting one. In this case the contamination is heavy tailed, which strongly affects the behavior of our scale measures. The case  $\epsilon$  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  $\sigma$ ,  $d$  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  $\lambda$  and  $\epsilon$ . This function is plotted in Figure 1 (top left). The parameter  $\epsilon$  is on a log-scale since we are primarily interested

TABLE 1. Densities and non-central moments of several parametric families. The scaling factor for the  $t_\nu$  distribution is  $c_\nu = \Gamma(\frac{\nu+1}{2})/(\sqrt{\nu\pi}\Gamma(\frac{\nu}{2}))$ .

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	$\frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$	$-\infty < a < b < \infty$	$\mu_1 = \frac{1}{2}(a+b),$ $\mu_2 = \frac{1}{3}\{(a+b)^2 - ab\},$ $\mu_3 = \frac{1}{4}(a+b)(a^2 + b^2),$ $\mu_4 = \frac{1}{5}\{(a+b)(a^3 + ab^2) + b^4\}$
$t_\nu$	$c_\nu \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\nu \in \mathbb{N}$	$\mu_1 = \mu_3 = 0,$ $\mu_2 = \nu/(\nu-2),$ $\mu_4 = 3\nu^2/\{(\nu-2)(\nu-4)\}$
normal mixture	$\epsilon \frac{1}{\sqrt{2\pi\lambda}} \exp\left\{-\frac{x^2}{2\lambda^2}\right\} + (1-\epsilon) \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}$	$0 \leq \epsilon \leq 1, \lambda \geq 1$	$\mu_1 = \mu_3 = 0,$ $\mu_2 = \epsilon\lambda^2 + (1-\epsilon),$ $\mu_4 = 3\epsilon\lambda^4 + 3(1-\epsilon)$

TABLE 2. Specific values of  $\sigma$ ,  $d$  and the respective asymptotic variances for the distribution families given in Table 1.  $c_\nu = \Gamma(\frac{\nu+1}{2})/(\sqrt{\nu\pi}\Gamma(\frac{\nu}{2}))$

distribution	$\sigma(F)$	$ASV(\sigma_n)$	$d(F)$	$ASV(d_n)$
normal	$\sigma$	$\frac{\sigma^2}{2}$	$\frac{2\sigma}{\sqrt{2\pi}}$	$\sigma^2 \left\{1 - \frac{2}{\pi}\right\}$
Laplace	$\sqrt{2}\alpha$	$\frac{5}{2}\alpha^2$	$\alpha$	$\alpha^2$
uniform	$\frac{b-a}{2\sqrt{3}}$	$\frac{1}{60}(b-a)^2$	$\frac{b-a}{4}$	$\frac{1}{48}(b-a)^2$
$t_\nu$	$\sqrt{\frac{\nu}{\nu-2}}$	$\frac{\nu(\nu-1)}{2(\nu-2)(\nu-4)}$	$\frac{2\nu c_\nu}{\nu-1}$	$\frac{\nu}{\nu-2} - \left\{\frac{2\nu c_\nu}{\nu-1}\right\}^2$
normal mixture	$\sqrt{\epsilon\lambda^2 + (1-\epsilon)}$	$\frac{3(\epsilon\lambda^4 + 1 - \epsilon) - (\epsilon\lambda^2 + 1 - \epsilon)^2}{4(\epsilon\lambda^2 + 1 - \epsilon)}$	$\sqrt{\frac{2}{\pi}}\{\epsilon\lambda + (1-\epsilon)\}$	$\epsilon\lambda^2 + 1 - \epsilon - \frac{2}{\pi}\{\epsilon\lambda + (1-\epsilon)\}^2$

TABLE 3. Population values, cf. (4), expressions for  $J$ , cf. (7), and resulting asymptotic variances for Gini's mean difference at the parametric families of Table 1. Abbreviations used:  $\zeta(\lambda) = \sqrt{2 + \lambda^2}$ ,  $K_\nu = \int_{-\infty}^{\infty} x^2 f_\nu(x) F_\nu^2(x) dx$  with  $f_\nu, F_\nu$  being density and cdf of the  $t_\nu$  distribution.  $B(\cdot, \cdot)$  denotes the beta function.

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

in small contamination fractions. Fixing  $\lambda = 3$ , we find that for  $\epsilon = 0.00175$ , the mean deviation is as efficient as the standard deviation. Tukey (1960) gives a value of  $\epsilon = 0.008$ . The more precise value of 0.00175 is also in line with the simulation results of Section 4, and it supports even more so Tukey's main message: that this values is surprisingly low. Tukey (1960) also points out that it is virtually impossible to distinguish a normal sample from a sample generated by a normal mixture distribution with such a low contamination fraction.

As for Gini's mean difference, the asymptotic relative efficiency  $ARE(g_n, \sigma_n)$  is depicted in the upper right plot of Figure 1. For  $\lambda = 3$ , Gini's mean difference is as efficient as the standard deviation for  $\epsilon$  as small as 0.000309. In the lower plot of Figure 1, equal-efficiency curves are drawn. They represent those parameter values  $(\lambda, \epsilon)$ , 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.

TABLE 4. Values and asymptotic variances of the standard deviation  $\sigma$ , Gini's mean difference  $g$  and the mean absolute deviation  $d$  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  $t_\nu$  family and the normal mixture family  $NM(\lambda, \epsilon)$ .

distribution	$\sigma$	$g$	$d$	$ASV(\sigma_n)$	$ASV(g_n)$	$ASV(d_n)$
$N(0, 1)$	1	1.128379	0.797884	0.5	0.651006	0.36338
$L(0, 1)$	1.414214	1.5	1	2.5	2.333333	1
$U(0, 1)$	0.288675	0.333333	0.25	0.016667	0.022222	0.020833
$t_5$	1.290994	1.383983	0.949017	3.333333	1.784415	0.766034
$t_6$	1.224745	1.331554	0.918559	1.875	1.453316	0.656250
$t_7$	1.183216	1.29694	0.898313	1.4	1.268881	0.593033
$t_{10}$	1.118034	1.239891	0.864685	0.9375	1.013824	0.502319
$t_{15}$	1.074172	1.199657	0.840757	0.734266	0.864755	0.446974
$t_{16}$	1.069045	1.194859	0.837891	0.714286	0.848464	0.440796
$t_{25}$	1.042572	1.169776	0.822862	0.621118	0.768006	0.409855
$t_{40}$	1.025978	1.153794	0.813245	0.570175	0.720625	0.391264
$t_{41}$	1.025320	1.153156	0.812861	0.568261	0.718794	0.390540
$t_{100}$	1.010153	1.138367	0.803932	0.526148	0.677563	0.374102
$NM(3, 0.008)$	1.031504	1.150661	0.810651	0.890015	0.791360	0.406845
$NM(3, 0.00175)$	1.006976	1.133259	0.800677	0.589695	0.681919	0.372916
$NM(3, 0.000309)$	1.001235	1.129241	0.798378	0.516028	0.656474	0.365065

### 3. INFLUENCE FUNCTIONS

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

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

where  $F_{\epsilon, x} = (1 - \epsilon)F + \epsilon\Delta_x$ ,  $0 \leq \epsilon \leq 1$ ,  $x \in \mathbb{R}$ , and  $\Delta_x$  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 three scale measures are

$$IF(x, \sigma(\cdot); F) = (2\sigma(F))^{-1} \{(E(X) - x)^2 - \sigma^2(F)\},$$

$$IF(x, d(\cdot); F) = |x - m(F)| - d(F),$$

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

The derivations are straightforward, and the results are stated without proof. For the formula of  $d(\cdot)$  to hold,  $F$  has to fulfill certain regularity conditions in the vicinity of its median  $m(F)$ . Specifically,  $(m(F_{\epsilon, x}) - m(F)) = O(\epsilon)$  as  $\epsilon \rightarrow 0$  for all  $x \in \mathbb{R}$  and  $F(m(F_{\epsilon, x})) \rightarrow 1/2$  are a set of sufficient conditions. They are fulfilled, e.g., if  $F$  possesses a positive Lebesgue density in a neighborhood of  $m(F)$ . For the standard normal distribution, above expressions reduce to

$$IF(x, \sigma(\cdot); N(0, 1)) = (x^2 - 1)/2,$$

$$IF(x, d(\cdot); N(0, 1)) = |x| - \sqrt{2/\pi},$$

$$IF(x, g(\cdot); N(0, 1)) = 4\phi(x) + 2x\{2\Phi(x) - 1\} - 4/\sqrt{\pi},$$



TABLE 5. Asymptotic relative efficiencies of Gini's mean difference  $g_n$ , the mean deviation and the interquartile range ( $IQR$ ) with respect to the standard deviation 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  $t_\nu$  family and the normal mixture family  $NM(\lambda, \epsilon)$ .

distribution	$ARE(g_n, \sigma_n)$	$ARE(d_n, \sigma_n)$	$ARE(IQR, \sigma_n)$
$N(0, 1)$	0.977901	0.875969	0.367529
$L(0, 1)$	1.205357	1.25	0.600566
$U(0, 1)$	1	0.6	0.2
$t_5$	2.146820	2.351417	1.333182
$t_6$	1.524991	1.607143	0.847705
$t_7$	1.325620	1.360745	0.686402
$t_{10}$	1.137276	1.116343	0.525866
$t_{15}$	1.059075	1.006384	0.453437
$t_{16}$	1.051672	0.995444	0.446226
$t_{25}$	1.018130	0.944031	0.412339
$t_{40}$	1.000643	0.915598	0.393605
$t_{41}$	0.9999998	0.914524	0.392898
$t_{100}$	0.986164	0.890804	0.377281
$NM(3, 0.008)$	1.399511	1.351120	0.627943
$NM(3, 0.00175)$	1.095255	0.999755	0.429469
$NM(3, 0.000309)$	0.999901	0.898767	0.378687

where  $\phi$  and  $\Phi$  denote the density and the cdf of the standard normal distribution, respectively. These curves are depicted in Figure 2. They confirm the general impression mediated by Table 5, 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.

#### 4. FINITE SAMPLE EFFICIENCIES

In a small simulation study we want to check if the asymptotic efficiencies computed in Section 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 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  $t_\nu$  distribution with  $\nu = 5, 16, 41$  and the normal mixture with the parameter choices as in Tables 4 and 5. The choice  $\nu = 5$  serves as a heavy-tailed example, whereas for  $\nu = 16$  and  $\nu = 41$  we have witnessed at Table 5 that the mean deviation and the Gini mean difference, respectively, are asymptotically equally 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 the three scale measures  $\sigma_n$ ,  $d_n$  and  $g_n$ . The results for  $N(0, 1)$ ,  $L(0, 1)$  and  $U(0, 1)$  are summarized in Table 6, for the  $t_\nu$  distributions in Table 7 and for the normal mixture distributions in Table 8. 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 4 as  $n$  increases), the squared bias relative to the variance, and, for  $g_n$  and  $d_n$ , the

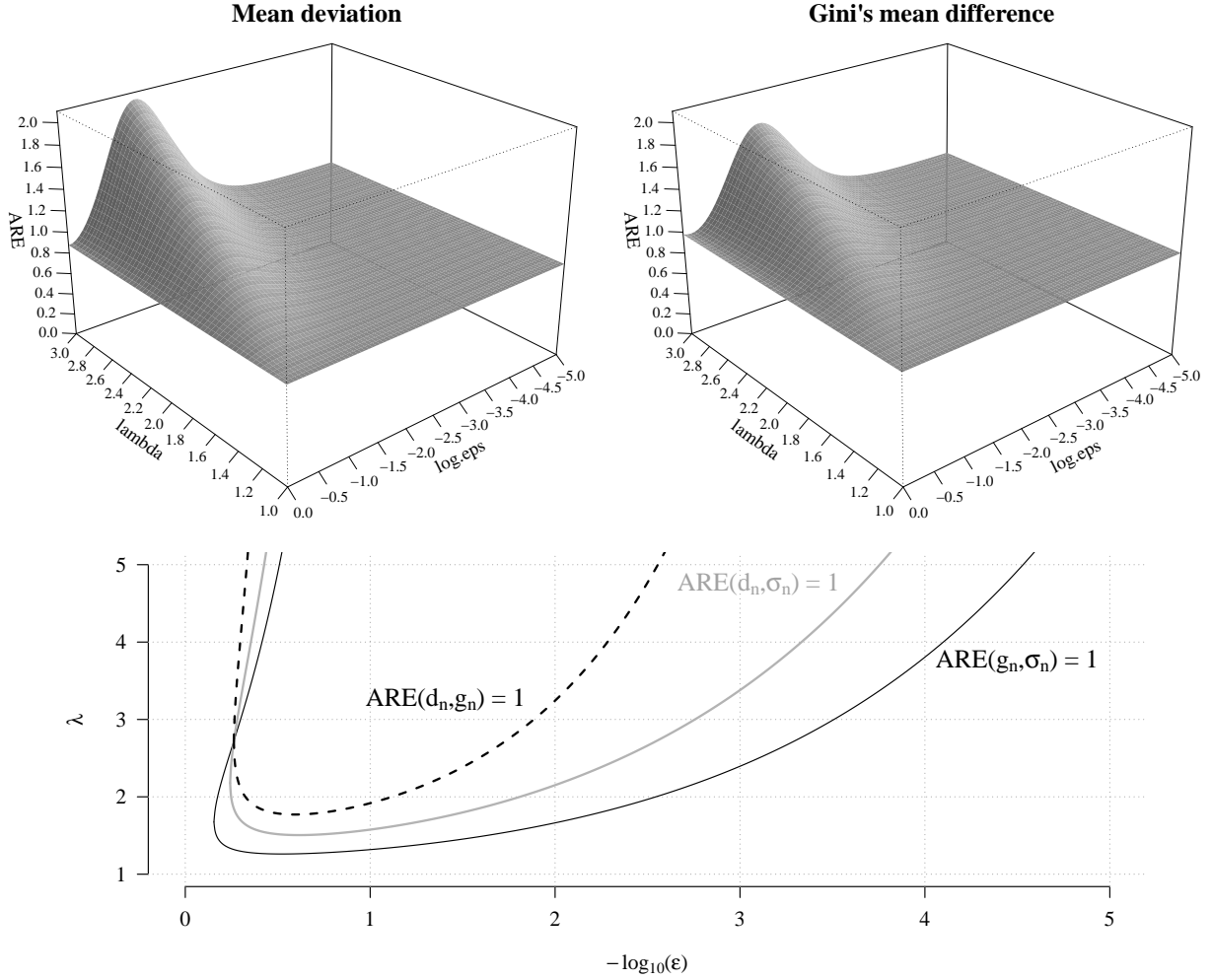


FIGURE 1. Top row: Asymptotic relative efficiencies of the mean deviation (left) and Gini's mean difference (right) in the normal mixture model as a function of  $\lambda$  and  $\log(\epsilon)$ . Bottom: The curves for which values of  $\lambda$  and  $\epsilon$  the scale measures have the same asymptotic efficiency.

relative efficiencies with respect to the standard deviation. For the relative efficiency computation, it is important to note that the standardizing, cf. (1), is done not by the true asymptotic values, but by the empirical finite-sample value, i.e. the sample mean of the 100,000 estimates. Also note that variances, not mean squared errors, are reported. For Gini's mean difference, the simulated variances are also compared to the true finite-sample variances, cf. (6).

We observe the following: For large and moderate sample sizes ( $n = 50, 500$ ), the simulated values are close to the asymptotic ones from Tables 4 and 5, and we conclude that the asymptotic efficiency generally provides a useful indication for the actual efficiency. In small samples, the simulated relative efficiencies may substantially differ from the asymptotic values, but the ranking of the three estimators stays the same. Furthermore, the simulations confirm the unbiasedness of Gini's mean difference and the formula (6), due to Lomnicki (1952), for its finite-sample variance.

Finally, 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 heavily biased for small samples for all distributions considered, whereas  $d_n$  has in all situations

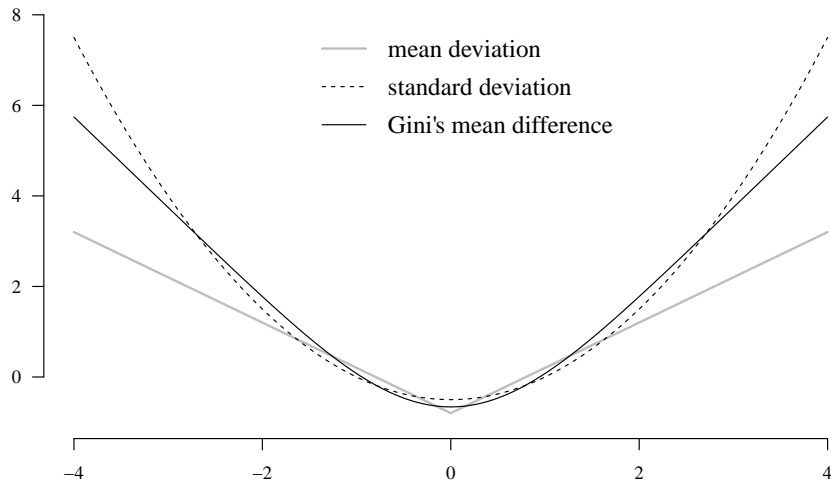


FIGURE 2. Influence functions of the standard deviation, the mean deviation and the Gini mean difference at the standard normal distribution.

a smaller bias than  $\sigma_n$ . Somewhat unexpected is the increase of the squared bias relative to the variance of  $d_n$  from  $n = 5$  to  $n = 8$ . The reason may lie in the different behavior of the sample median for odd and even numbers of observations.

The simulations were done in R (R Development Core Team, 2010), using an implementation for Gini's mean difference by A. Azzalini.<sup>1</sup>

## 5. SUMMARY AND CONCLUSION

Neither the standard deviation nor the mean deviation is a robust estimator. However, several authors have argued that, when comparing the standard deviation with the mean deviation, the (relatively) 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  $t_\nu$  distributions with  $\nu \geq 23$ , the (properly standardized) asymptotic variances of  $g_n$  and  $\sigma_n$  are within a three percent 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. However, it must also be noted that heavy tails are a bad case scenario for all three scale measure, and that, for heavy-tailed distributions, much more efficient estimators available, e.g.,  $M$ -estimators or trimmed or quantile-based scale estimators.

Gini's mean difference has further advantages, which particularly concern its finite-sample performance: it is unbiased, and the finite-sample variance is known. It either can be computed exactly or, if no specific model is assumed, estimated, which allows for instance better approximative confidence intervals. Neither of that is true for the standard deviation or the mean deviation, and one can consequently argue that Gini's mean difference is a superior scale estimator even under normality. Scale measures may serve different purposes: some, such as the interquartile range, are

<sup>1</sup><https://stat.ethz.ch/pipermail/r-help/2003-April/032820.html>

TABLE 6. Simulated variances, biases and relative efficiencies of  $\sigma_n$ ,  $g_n$ ,  $d_n$  at  $N(0, 1)$ ,  $L(0, 1)$  and  $U(0, 1)$  for several sample sizes,  $d_n^*$ : mean deviation with  $1/n$  scaling.

estimator		$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
$N(0, 1)$						
$\sigma_n$	$n$ -variance	0.577	0.548	0.541	0.507	0.505
	bias <sup>2</sup> /variance	0.031	0.019	0.014	0.003	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	3.4e-08	4.7e-07	7.8e-06	1.0e-05	4.7e-06
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.009	0.020	0.012	0.001	< 0.001
	rel. efficiency wrt $\sigma_n$	0.938	0.902	0.894	0.880	0.876
$d_n^*$	bias <sup>2</sup> /variance	0.296	0.118	0.101	0.021	0.002
$L(0, 1)$						
$\sigma_n$	$n$ -variance	1.946	2.076	2.134	2.387	2.495
	bias <sup>2</sup> /variance	0.055	0.034	0.027	0.006	<0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	2.8e-06	8.4e-09	8.4e-08	1.3e-05	8.3e-10
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.025	0.028	0.021	0.005	<0.001
	rel. efficiency wrt $\sigma_n$	1.061	1.101	1.123	1.206	1.245
$d_n^*$	bias <sup>2</sup> /variance	0.106	0.040	0.031	0.006	0.001
$U(0, 1)$						
$\sigma_n$	$n$ -variance	0.031	0.025	0.022	0.018	0.017
	bias <sup>2</sup> /variance	0.021	0.010	0.007	0.001	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	1.9e-05	6.2e-07	9.4e-07	3.0e-05	5.1e-08
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	6.1e-06	4.7e-03	2.3e-03	6.8e-05	1.7e-05
	rel. efficiency wrt $\sigma_n$	0.829	0.694	0.672	0.614	0.603
$d_n^*$	bias <sup>2</sup> /variance	0.657	0.285	0.236	0.059	0.006

primarily used for descriptive reasons. In this respect, the mean deviation and the standard deviation may be preferred (the latter due to its widespread use), but as an estimator, i.e. for inferring about an unknown population scale, Gini's mean difference has, in our opinion, clear advantages. Although there are quite a few articles that advocate the use of alternative scale estimators (specifically for Gini's mean difference, cf. e.g. Yitzhaki, 2003), we are aware that the standard deviation is so widespread and common that any other scale estimator taking its role seems as unlikely as a change from the decimal to another numeral system.

TABLE 7. Simulated variances, biases and relative efficiencies of  $\sigma_n$ ,  $g_n$ ,  $d_n$  at  $t_\nu$  distributions for several sample sizes and values of  $\nu$ ;  $d_n^*$ : mean deviation with  $1/n$  scaling.

estimator		$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
$t_5$						
$\sigma_n$	$n$ -variance	1.584	1.686	1.762	2.313	2.880
	bias <sup>2</sup> /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 <sup>2</sup> /variance	4.0e-06	1.3e-05	2.5e-05	5.1e-06	1.4e-05
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.014	0.018	0.014	0.003	< 0.001
	rel. efficiency wrt $\sigma_n$	1.105	1.208	1.282	1.673	1.977
$d_n^*$	bias <sup>2</sup> /variance	0.160	0.066	0.053	0.011	0.001
$t_{16}$						
$\sigma_n$	$n$ -variance	0.745	0.722	0.722	0.710	0.705
	bias <sup>2</sup> /variance	0.034	0.021	0.015	0.003	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	1.4e-07	5.0e-06	7.9e-07	7.6e-06	2.4e-05
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.012	0.018	0.012	0.002	< 0.001
	rel. efficiency wrt $\sigma_n$	0.972	0.956	0.963	0.989	0.991
$d_n^*$	bias <sup>2</sup> /variance	0.259	0.106	0.085	0.017	0.002
$t_{41}$						
$\sigma_n$	$n$ -variance	0.640	0.611	0.605	0.574	0.575
	bias <sup>2</sup> /variance	0.032	0.020	0.014	0.003	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	1.1e-05	3.6e-06	1.5e-06	9.5e-08	7.1e-07
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.010	0.018	0.013	0.002	< 0.001
	rel. efficiency wrt $\sigma_n$	0.950	0.918	0.921	0.916	0.919
$d_n^*$	bias <sup>2</sup> /variance	0.276	0.113	0.094	0.019	0.002

## ACKNOWLEDGMENT

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*.

TABLE 8. Simulated variances, biases and relative efficiencies of  $\sigma_n$ ,  $g_n$ ,  $d_n$  at normal mixture distributions for  $\lambda = 3$  and  $\epsilon = 0.008, 0.00175, 0.000309$ ;  $d_n^*$ : mean deviation with  $1/n$  scaling.

estimator		$n = 5$	$n = 8$	$n = 10$	$n = 50$	$n = 500$
<i>NM(3, 0.008)</i>						
$\sigma_n$	$n$ -variance	0.710	0.698	0.711	0.815	0.875
	bias <sup>2</sup> /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 <sup>2</sup> /variance	4.6e-06	2.1e-10	1.6e-05	3.4e-06	8.4e-07
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.010	0.016	0.013	0.002	< 0.001
	rel. efficiency wrt $\sigma_n$	1.000	1.016	1.039	1.204	1.332
$d_n^*$	bias <sup>2</sup> /variance	0.264	0.112	0.087	0.020	0.002
<i>NM(3, 0.00175)</i>						
$\sigma_n$	$n$ -variance	0.617	0.587	0.576	0.573	0.590
	bias <sup>2</sup> /variance	0.032	0.019	0.017	0.003	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	1.6e-07	3.0e-07	4.8e-08	1.8e-05	1.0e-05
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.011	0.017	0.013	0.002	3.9e-05
	rel. efficiency wrt $\sigma_n$	0.951	0.931	0.931	0.971	0.992
$d_n^*$	bias <sup>2</sup> /variance	0.283	0.115	0.100	0.022	0.003
<i>NM(3, 0.000309)</i>						
$\sigma_n$	$n$ -variance	0.584	0.558	0.543	0.517	0.515
	bias <sup>2</sup> /variance	0.031	0.017	0.014	0.003	< 0.001
$g_n$	$n$ -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 <sup>2</sup> /variance	1.3e-05	4.8e-06	5.1e-07	1.3e-06	8.3e-06
	rel. efficiency wrt $\sigma_n$	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 <sup>2</sup> /variance	0.009	0.018	0.012	0.002	< 0.001
	rel. efficiency wrt $\sigma_n$	0.941	0.900	0.903	0.899	0.903
$d_n^*$	bias <sup>2</sup> /variance	0.291	0.122	0.096	0.021	0.002

#### APPENDIX A. INTEGRALS FOR THE NORMAL DISTRIBUTION

When evaluating the integral  $J$ , cf. (7), for the standard normal distribution, one encounters the integral

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

where  $\phi$  and  $\Phi$  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 therefor 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

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

Introducing polar coordinates  $\alpha, r$  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,

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

where  $\Gamma_0$  is the upper unit half circle in the complex plane, cp. Figure 3. Let us call  $h$  the integrand in (8), its poles (both of order two) are  $\gamma_{1/2} = (2 \pm \sqrt{3})i$ , so that  $\gamma_2$  lies within the closed upper half unit circle  $\Gamma$ . The residue of  $h$  in  $\gamma_2$  is  $-\sqrt{3}i/2$ . Integrating  $h$  along  $\Gamma_1$ , i.e. the real line from -1 to 1, cf. Figure 3, and applying the residue theorem to the closed line integral along  $\Gamma$  completes the derivation.

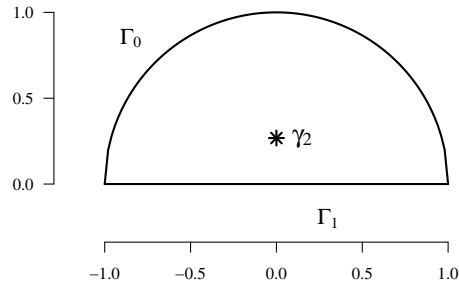


FIGURE 3. Residue theorem: the line integral over  $h$  along the closed curve  $\Gamma = \Gamma_0 \cup \Gamma_1$  is determined by the residue of  $h$  in  $\gamma_2$ .

## APPENDIX B. INTEGRALS FOR THE NORMAL MIXTURE DISTRIBUTION

Evaluating the integral  $J$  for the normal mixture distribution, we arrive after some lengthy but straightforward calculations at

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

where

$$\begin{aligned} A(\lambda) &= \int_{-\infty}^{\infty} x\phi^2(x)\Phi(x/\lambda)dx = \frac{1}{4\pi\sqrt{1+2\lambda^2}}, & B &= \int_{-\infty}^{\infty} x^2\phi(x)\Phi(x)dx = \frac{1}{2}, \\ C(\lambda) &= \int_{-\infty}^{\infty} x^2\phi(x)\Phi^2(x/\lambda)dx = \frac{1}{4} + \frac{\lambda}{\pi(1+\lambda^2)\sqrt{2+\lambda^2}} + \frac{1}{2\pi} \arctan\left(\frac{1}{\lambda\sqrt{2+\lambda^2}}\right), \\ D(\lambda) &= \int_{-\infty}^{\infty} x^2\phi(x)\Phi(x)\Phi(x/\lambda)dx = \frac{1}{4} + \frac{3\lambda^2+1}{4\pi(1+\lambda^2)\sqrt{2\lambda^2+1}} + \frac{1}{2\pi} \arctan\left(\frac{1}{\sqrt{2\lambda^2+1}}\right), \\ E(\lambda) &= \int_{-\infty}^{\infty} \phi^2(x)\phi(x/\lambda)dx = \frac{1}{2\pi\sqrt{1+2\lambda^2}}, \end{aligned}$$

for all  $\lambda > 0$ . As before,  $\phi$  and  $\Phi$  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. They can be solved by similar means as  $I_1$ . Proceeding as in Appendix A, solving the respective two inner integrals yields

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

These integrals are again solved by the residue theorem, for which we used the software Mathematica (Wolfram Research, Inc., 2012).

### APPENDIX C. INTEGRALS FOR THE $t_\nu$ DISTRIBUTION

In order to compute analytical expressions for  $g$  and  $J$  in case of the  $t_\nu$  distribution, the following identities are helpful:

$$(9) \quad \int x \left(1 + \frac{x^2}{\beta}\right)^\alpha dx = \frac{\beta}{2(\alpha+1)} \left(1 + \frac{x^2}{\beta}\right)^{\alpha+1}, \quad \alpha \neq -1, \beta \neq 0.$$

$$(10) \quad \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{m}\right)^{-m} dx = \frac{1}{c_{2m-1}} \sqrt{\frac{m}{2m-1}}, \quad m \in \mathbb{N},$$

$$(11) \quad \int_{-\infty}^{\infty} \left(1 + \frac{x^2}{m}\right)^{-\frac{3m-1}{2}} dx = \frac{1}{c_{3m-2}} \sqrt{\frac{m}{3m-2}}, \quad m \in \mathbb{N},$$

where  $c_\nu$  is the scaling factor of the  $t_\nu$  density, cf. Table 1. The identities (10) and (11) can be obtained by transforming the respective left-hand sides into a  $t_\nu$ -densities by substituting  $y = ((2m-1)/m)^{1/2}x$  and  $y = ((3m-2)/m)^{1/2}x$ , respectively.

For computing  $g$ , we evaluate (5), successively making use of (9) and (10), 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 Table 3 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  $f_\nu$  being the  $t_\nu$  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 (9), 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  $F_\nu$  is the cdf of the  $t_\nu$  distribution. By employing (9) and (11), 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 Table 3. 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.

#### REFERENCES

- L. V. Ahlfors. *Complex analysis*. New York: McGraw-Hill, 2nd edition, 1966.
- P. J. Bickel and E. L. Lehmann. Descriptive statistics for nonparametric models. III. Dispersion. *Annals of Statistics*, (6):1139–1148, 1976.
- S. Gorard. Revisiting a 90-year-old debate: the advantages of the mean deviation. *British Journal of Educational Studies*, 53(4):417–430, 2005.
- F. R. Hampel. The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, 69:383–393, 1974.
- F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust statistics. The approach based on influence functions*. Wiley Series in Probability and Mathematical Statistics. New York etc.: Wiley, 1986.
- W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.
- T. Hojo. Distribution of the median, quartiles and interquartile distance in samples from a normal population. *Biometrika*, 23(3-4):315–360, 1931.
- P. J. Huber and E. M. Ronchetti. *Robust statistics*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, 2nd edition, 2009.
- Z. A. Lomnicki. The standard error of Gini's mean difference. *Ann. Math. Statist.*, 23(4):635–637, 1952.
- U. S. Nair. The standard error of Gini's mean difference. *Biometrika*, 28:428–436, 1936.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL <http://www.R-project.org/>. ISBN 3-900051-07-0.
- P. J. Rousseeuw and C. Croux. Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283, 1993.
- J. W. Tukey. A survey of sampling from contaminated distributions. In Olkin, I. et al., editor, *Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling*, pages 448–485. Stanford, CA: Stanford University Press, 1960.
- Wolfram Research, Inc. *Mathematica*. Champaign, Illinois, version 9.0 edition, 2012.
- S. Yitzhaki. Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron*, 61(2): 285–316, 2003.

FAKULTÄT FÜR MATHEMATIK, RUHR-UNIVERSITÄT BOCHUM, 44780 BOCHUM, GERMANY  
*E-mail address:* `carina.gerstenberger@rub.de`

FAKULTÄT STATISTIK, TECHNISCHE UNIVERSITÄT DORTMUND, 44221 DORTMUND, GERMANY  
*E-mail address:* `daniel.vogel@tu-dortmund.de`