The Distance Standard Deviation

Edelmann, Dominic; Richards, Donald ; Vogel, Daniel

The Distance Standard Deviation

Dominic Edelmann,[*] Donald Richards,[†] and Daniel Vogel[‡]

May 17, 2017

Abstract

The distance standard deviation, which arises in distance correlation analysis of multivariate data, is studied as a measure of spread. New representations for the distance standard deviation are obtained in terms of Gini's mean difference and in terms of the moments of spacings of order statistics. Inequalities for the distance variance are derived, proving that the distance standard deviation is bounded above by the classical standard deviation and by Gini's mean difference. Further, it is shown that the distance standard deviation satisfies the axiomatic properties of a measure of spread. Explicit closed-form expressions for the distance variance are obtained for a broad class of parametric distributions. The asymptotic distribution of the sample distance variance is derived.

1   Introduction

In recent years, the topic of distance correlation has been prominent in statistical analyses of dependence between multivariate data sets. The concept of distance correlation was defined in the one-dimensional setting by Feuerverger [7] and subsequently in the multivariate case by Székely, et al. [25, 26], and those authors applied distance correlation methods to testing independence between random variables and vectors.

---

[*]German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany.

[†]Department of Statistics, Pennsylvania State University, University Park, PA 16802, U.S.A.

[‡]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB24 3UE, U.K.

[*]Corresponding author; e-mail address: dominic.edelmann@dkfz-heidelberg.de

1

Since the appearance of [25, 26], enormous interest in the theory and applications of distance correlation has arisen. We refer to the articles [22, 27, 28] on statistical inference; [8, 9, 14, 33] on time series; [4, 5, 6] on affinely invariant distance correlation and connections with singular integrals; [19] on metric spaces; and [23] on machine learning. Distance correlation methods have also been applied to assessing familial relationships [17], and to detecting associations in large astrophysical databases [20, 21].

For $z \in \mathbb{C}$, denote by $|z|$ the modulus of z. For any positive integer p and $s, x \in \mathbb{R}^p$, we denote by $\langle s, x \rangle$ the standard Euclidean inner product on $\mathbb{R}^p$ and by $\|s\| = \langle s, s \rangle^{1/2}$ the standard Euclidean norm. Further, we define the constant

$$c_p = \frac{\pi^{(p+1)/2}}{\Gamma\big((p+1)/2\big)}.$$

For jointly distributed random vectors $X \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$, let

$$f_{X,Y}(s,t) = \mathbb{E} \exp\big(\sqrt{-1}(\langle s, X \rangle + \langle t, Y \rangle)\big),$$

$s \in \mathbb{R}^p$, $t \in \mathbb{R}^q$, be the joint characteristic function of (X, Y) and let $f_X(s) = f_{X,Y}(s, 0)$ and $f_Y(t) = f_{X,Y}(0, t)$ be the corresponding marginal characteristic functions. The *distance covariance* between X and Y is defined as the nonnegative square root of

$$\mathcal{V}^2(X,Y) = \frac{1}{c_p c_q} \int_{\mathbb{R}^{p+q}} \big|f_{X,Y}(s,t) - f_X(s)f_Y(t)\big|^2 \frac{\mathrm{d}s\,\mathrm{d}t}{\|s\|^{p+1}\,\|t\|^{q+1}}; \qquad (1.1)$$

the *distance variance* is defined as

$$\mathcal{V}^2(X) := \mathcal{V}^2(X,X) = \frac{1}{c_p^2} \int_{\mathbb{R}^{2p}} \big|f_X(s+t) - f_X(s)f_X(t)\big|^2 \frac{\mathrm{d}s\,\mathrm{d}t}{\|s\|^{p+1}\,\|t\|^{p+1}}; \qquad (1.2)$$

and we define the *distance standard deviation* $\mathcal{V}(X)$ as the nonnegative square root of $\mathcal{V}^2(X)$. The *distance correlation coefficient* is defined as

$$\mathcal{R}(X,Y) = \frac{\mathcal{V}(X,Y)}{\sqrt{\mathcal{V}(X)\mathcal{V}(Y)}} \qquad (1.3)$$

as long as $\mathcal{V}(X), \mathcal{V}(Y) \neq 0$, and $\mathcal{R}(X,Y)$ is defined to be zero otherwise.

The distance correlation coefficient, unlike the Pearson correlation coefficient, characterizes independence: $\mathcal{R}(X,Y) = 0$ if and only if X and Y are mutually independent. Moreover, $0 \leq \mathcal{R}(X,Y) \leq 1$; and for one-dimensional random variables $X, Y \in \mathbb{R}$, $\mathcal{R}(X,Y) = 1$ if and only if Y is a linear function of X. The empirical distance correlation possesses a remarkably simple expression ([7], [25, Theorem 1]), and efficient algorithms for computing it are now available [13].

The objective of this paper is to study the distance standard deviation $\mathcal{V}(X)$. Since distance standard deviation terms appear in the denominator of the distance correlation coefficient (1.3) then properties of $\mathcal{V}(X)$ are crucial to understanding fully the nature

of $\mathcal{R}(X, Y)$. Now that $\mathcal{R}(X, Y)$ has been shown to be superior in some instances to classical measures of correlation or dependence, there arises the issue of whether $\mathcal{V}(X)$ constitutes a measure of spread suitable for situations in which the classical standard deviation cannot be applied.

As $\mathcal{V}(X)$ is possibly a measure of spread, we should compare it to other such measures. Indeed, suppose that $\mathbb{E}(\|X\|^2) < \infty$, and let X, X', and X'' be independent and identically distributed (i.i.d.); then, by [25, Remark 3],

$$\mathcal{V}^2(X) = \mathbb{E}(\|X - X'\|^2) + (\mathbb{E}\|X - X'\|)^2 - 2\mathbb{E}(\|X - X'\| \cdot \|X - X''\|), \qquad (1.4)$$

The second term on the right-hand side of (1.4) is reminiscent of the Gini mean difference [10, 31], which is defined for real-valued random variables Y as

$$\Delta(Y) := \mathbb{E}|Y - Y'|, \qquad (1.5)$$

where Y and Y' are i.i.d. Furthermore, if $X \in \mathbb{R}$ then one-half the first summand in (1.4) equals $\sigma^2(X)$, the variance of X:

$$\frac{1}{2}\mathbb{E}(|X - X'|^2) = \frac{1}{2}\mathbb{E}(X^2 - 2XX' + X'^2) = E(X^2) - E(X)E(X') \equiv \sigma^2(X).$$

Let X and Y be real-valued random variables with cumulative distribution functions F and G, respectively. Further, let F^{-1} and G^{-1} be the right-continuous inverses of F and G, respectively. Following [24, Definition 2.B.1], we say that X is *smaller than* Y *in the dispersive ordering*, denoted by $X \leq_{\text{disp}} Y$, if for all $0 < \alpha \leq \beta < 1$,

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha). \qquad (1.6)$$

According to [2], a *measure of spread* is a functional $\tau(X)$ satisfying the axioms:

(C1) $\tau(X) \geq 0$,

(C2) $\tau(a + bX) = |b|\,\tau(X)$ for all $a, b \in \mathbb{R}$, and

(C3) $\tau(X) \leq \tau(Y)$ if $X \leq_{\text{disp}} Y$.

The distance standard deviation $\mathcal{V}(X)$ obviously satisfies (C1). Moreover, Székely, et al. [25, Theorem 4] prove that:

1. If $\mathcal{V}(X) = 0$ then $X = \mathbb{E}[X]$, amost surely,

2. $\mathcal{V}(a + bX) = |b|\,\mathcal{V}(X)$ for all $a, b \in \mathbb{R}$, and

3. $\mathcal{V}(X + Y) \leq \mathcal{V}(X) + \mathcal{V}(Y)$ if X and Y are independent.

In particular, $\mathcal{V}(X)$ satisfies the dilation property (C2). In Section 5, we will show that $\mathcal{V}(X)$ satisfies condition (C3), proving that $\mathcal{V}(X)$ is a measure of spread in the sense

of [2]. However, we will also derive some stark differences between $\mathcal{V}(X)$, on the one hand, and the standard deviation and Gini's mean difference, on the other hand.

The paper is organized as follows. In Section 2, we derive inequalities between the summands in the distance variance representation (1.4). For real-valued random variables, we will prove that $\mathcal{V}(X)$ is bounded above by Gini's mean difference and by the classical standard deviation. In Section 3, we show that the representation (1.4) can be simplified further, revealing relationships between $\mathcal{V}(X)$ and the moments of spacings of order statistics. Section 4 provides closed-form expressions for the distance variance for numerous parametric distributions. In Section 5, we show that $\mathcal{V}(X)$ is a measure of spread in the sense of [2]; moreover, we point out some important differences between $\mathcal{V}(X)$, the standard deviation, and Gini's mean difference. Section 6 studies the properties of the sample distance variance.

2   Inequalities between the distance variance, the variance, and Gini's mean difference

The integral representation in equation (1.2) of the distance variance $\mathcal{V}^2(X)$ generally is not suitable for practical purposes. Székely, et al. [25, 26] derived an alternative representation; they show that if the random vector $X \in \mathbb{R}^p$ satisfies $\mathbb{E}\|X\|^2 < \infty$ and if X, X', and X'' are i.i.d. then

$$\mathcal{V}^2(X) = T_1(X) + T_2(X) - 2\,T_3(X), \tag{2.1}$$

where

$$\begin{aligned} T_1(X) &= \mathbb{E}(\|X - X'\|^2), \\ T_2(X) &= (\mathbb{E}\|X - X'\|)^2, \end{aligned} \tag{2.2}$$

and

$$T_3(X) = \mathbb{E}\big(\|X - X'\| \cdot \|X - X''\|\big), \tag{2.3}$$

Corresponding to the representation (2.1), a sample version of $\mathcal{V}^2(X)$ then is given by

$$\mathcal{V}_n^2(\boldsymbol{X}) = T_{1,n}(\boldsymbol{X}) + T_{2,n}(\boldsymbol{X}) - 2\,T_{3,n}(\boldsymbol{X}), \tag{2.4}$$

where

$$\begin{aligned} T_{1,n}(\boldsymbol{X}) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\|^2, \\ T_{2,n}(\boldsymbol{X}) &= \Big(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\|\Big)^2, \end{aligned} \tag{2.5}$$

and

$$T_{3,n}(\boldsymbol{X}) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \|X_i - X_j\| \cdot \|X_i - X_k\|. \qquad (2.6)$$

We remark that the version (2.4) is biased; indeed, throughout the paper, we work with biased sample versions to avoid dealing with numerous complicated, but unessential, constants in the ensuing results. In any case, an unbiased sample version can be defined in a similar fashion; see, e.g., [27]).

In the following we will study inequalities between the summands showing up in equations (2.1) and (2.4). In the one-dimensional case, these inequalities will lead to crucial results concerning the relationships between the distance standard deviation, Gini's mean difference and the standard deviation.

Lemma 2.1. *Let* $X = (X^{(1)}, \ldots, X^{(p)})^t \in \mathbb{R}^p$ *be a random vector. Moreover let* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *denote a random sample from* X *and let* $T_1(X)$, $T_2(X)$, $T_3(X)$, *and* $T_{1,n}(\boldsymbol{X})$, $T_{2,n}(\boldsymbol{X})$, $T_{3,n}(\boldsymbol{X})$ *be defined as in equations (2.1)-(2.6). Then*

$$T_{2,n}(\boldsymbol{X}) \le T_{3,n}(\boldsymbol{X}) \le T_{1,n}(\boldsymbol{X}), \qquad T_{1,n}(\boldsymbol{X}) \le 2T_{3,n}(\boldsymbol{X}). \qquad (2.7)$$

Further, if $\mathbb{E}\|X\|^2 < \infty$ *then*

$$T_2(X) \le T_3(X) \le T_1(X), \qquad T_1(X) \le 2T_3(X). \qquad (2.8)$$

PROOF. First note that

$$\begin{aligned}
T_{3,n}(\boldsymbol{X}) &= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \|X_i - X_j\| \cdot \|X_i - X_k\| \\
&= \frac{1}{n^3} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \|X_i - X_j\| \Big)^2.
\end{aligned}$$

By the Cauchy-Schwarz inequality, $(\sum_{i=1}^{n} a_i)^2 \le n \sum_{i=1}^{n} a_i^2$ for all $a_1, \ldots, a_n \in \mathbb{R}$; applying this inequality to the sums which define $T_{1,n}$, $T_{2,n}$ and $T_{3,n}$, we obtain

$$\begin{aligned}
T_{2,n}(\boldsymbol{X}) &= \frac{1}{n^4} \Big(\sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\| \Big)^2 \\
&\le \frac{n}{n^4} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \|X_i - X_j\| \Big)^2 = T_{3,n}(\boldsymbol{X})
\end{aligned}$$

and

$$\begin{aligned}
T_{3,n}(\boldsymbol{X}) &= \frac{1}{n^3} \sum_{i=1}^{n} \Big(\sum_{j=1}^{n} \|X_i - X_j\| \Big)^2 \\
&\le \frac{n}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\|^2 = T_{1,n}(\boldsymbol{X}).
\end{aligned}$$

The second assertion in (2.7) follows by the triangle inequality:

$$
\begin{aligned}
T_{1,n}(\boldsymbol{X}) &= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\|^2 \\
&= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \|X_i - X_j\| \cdot \|X_i - X_k + X_k - X_j\| \\
&\leq \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \|X_i - X_j\| \left(\|X_i - X_k\| + \|X_k - X_j\| \right) \\
&= 2\, T_{3,n}(\boldsymbol{X}).
\end{aligned}
$$

The corresponding inequalities (2.8) for the population measures follow from the strong consistency of the respective sample measures. Alternatively they can be derived by applying Jensen's inequality and the triangle inequality, respectively.   □

Using the inequalities in Lemma 2.1, we can derive upper bounds for the distance variance in terms of the variance of the components $X^{(1)}, \ldots, X^{(p)}$ and the Gini mean difference of the vector X.

Theorem 2.2. *Let* $X = (X^{(1)}, \ldots, X^{(p)})^t \in \mathbb{R}^p$ *be a random vector with* $\mathbb{E}\|X\| < \infty$, *and let* $X' = (X'^{(1)}, \ldots, X'^{(p)})^t$ *denote an i.i.d. copy of* X. *Then*

$$
\mathcal{V}^2(X) \leq \sum_{i=1}^{p} \sigma^2(X^{(i)}),
$$

and

$$
\mathcal{V}^2(X) \leq (\mathbb{E}\|X - X'\|)^2.
$$

PROOF. To prove the first assertion, we note that

$$
\begin{aligned}
\mathcal{V}^2(X) &= \lim_{n \to \infty} \left(T_{1,n}(\boldsymbol{X}) + T_{2,n}(\boldsymbol{X}) - 2T_{3,n}(\boldsymbol{X}) \right) \\
&\leq \lim_{n \to \infty} T_{2,n}(\boldsymbol{X}) \\
&= (\mathbb{E}\|X - X'\|)^2,
\end{aligned}
$$

where the inequality follows by Lemma 2.1.

To extablish the second inequality we can assume, without loss of generality, that

$\mathbb{E}\|X\|^2 < \infty$. Then

$$T_1(X) = \mathbb{E}\|X - X'\|^2$$

$$= \mathbb{E}\sum_{i=1}^{p}(X^{(i)} - X'^{(i)})^2$$

$$= \sum_{i=1}^{p}\mathbb{E}\Big[\big(X^{(i)} - \mathbb{E}X^{(i)}\big) + \big(\mathbb{E}X^{(i)} - X'^{(i)}\big)\Big]^2$$

$$= 2\sum_{i=1}^{p}\sigma^2(X^{(i)}).$$

Applying Lemma 2.1 yields

$$\mathcal{V}^2(X) = T_1(X) + T_2(X) - 2T_3(X)$$

$$\leq T_1(X) - T_3(X)$$

$$\leq \tfrac{1}{2}T_1(X)$$

$$= \sum_{i=1}^{p}\sigma^2(X^{(i)}).$$

The proof now is complete.   □

In the one-dimensional case, Theorem 2.2 implies that the distance variance is bounded above by the variance and the squared Gini mean difference.

Corollary 2.3. *Let X be a real-valued random variable with $\mathbb{E}\|X\| < \infty$. Then,*

$$\mathcal{V}^2(X) \leq \sigma^2(X), \qquad \mathcal{V}^2(X) \leq \Delta^2(X).$$

Let us note further that for $X \in \mathbb{R}$, the inequality $T_2(X) \leq T_1(X)$ can be sharpened.

Proposition 2.4. *Let X be a real-valued random variable with $\mathbb{E}(|X|^2) < \infty$. Then,*

$$T_2(X) \leq \tfrac{2}{3}T_1(X).$$

PROOF. By [31, p. 25],

$$1 \geq [\mathrm{Cor}(X, F(X))]^2 = \frac{\mathrm{Cov}^2(X, F(X))}{\sigma^2(X)\,\sigma^2(F(X))}. \tag{2.9}$$

By [30, equation (2.3)], $\mathrm{Cov}(X, F(X)) = \Delta(X)/4$; also, since $F(X)$ is uniformly distributed on the interval $[0, 1]$ then $\mathrm{Var}(F(X)) = 1/12$. By the definition of the Gini mean difference (1.5) and by (2.2), $\Delta^2(X) = T_2(X)$ and $\sigma^2(X) = T_1(X)/2$. Therefore, it follows from (2.9) that

$$1 \geq \frac{12}{16}\frac{\Delta^2(X)}{\sigma^2(X)} = \frac{3\,T_2(X)}{2\,T_1(X)},$$

and the proof now is complete. $\quad\square$

Interestingly, Gini's mean difference and the distance standard deviation coincide for distributions whose mass is concentrated on two points.

Theorem 2.5. *Let X be Bernoulli distributed with parameter p. Then*

$$\mathcal{V}^2(X) = \Delta^2(X) = 4p^2(1 - p)^2.$$

Conversely, if X is a non-trivial random variable for which $\mathcal{V}^2(X) = \Delta^2(X)$ then the distribution of X is concentrated on two points.

PROOF. It is straightforward from (2.1) to verify that, for a Bernoulli distributed random variable X, $\Delta(X) = 2\sigma^2(X) = 2T_3(X) = 2p(1 - p)$. Hence, by (2.1),

$$\mathcal{V}^2(X) = 2\sigma^2(X) + \Delta^2(X) - 2T_3(X) = 4p^2(1 - p)^2.$$

Conversely, if X is a non-trivial random variable for which $\mathcal{V}^2(X) = \Delta^2(X)$ then the conclusion that the distribution of X is concentrated on two points follows from Theorem 3.1. $\quad\square$

For the Bernoulli distribution with $p = \frac{1}{2}$, Theorem 2.5 implies immediately that $\mathcal{V}^2(X)$, $\sigma^2(X)$, and $\Delta^2(X)$ attain the same value, namely, 1/4. Hence, applying Corollary 2.3 and the dilation property $\mathcal{V}(aX) = |a|\mathcal{V}(X)$ in (C2), we obtain

Corollary 2.6. *Let $\mathcal{X}$ denote the set of all real-valued random variables and let $c > 0$. Then*

$$\max_{X \in \mathcal{X}}\{\mathcal{V}^2(X) : \sigma^2(X) = c\} = \max_{X \in \mathcal{X}}\{\mathcal{V}^2(X) : \Delta^2(X) = c\} = c,$$

and both maxima are attained by $Z = 2c^{1/2}Y$, where Y is Bernoulli distributed with parameter $p = \frac{1}{2}$.

This result answers a question raised by Gabor Székely (private communication, November 23, 2015).

We remark, that the second implication of Theorem 2.2 as well as Theorem 2.5 also follow directly from the result for the generalized distance variance in [19, Proposition 2.3]. However, the proof presented here provides a different and more elementary approach to these findings.

3 New representations for the distance variance

The representation of $\mathcal{V}$ given in (2.1), although more applicable than the expression given in equation (1.2), still has the drawback that it is undefined for random vectors

with infinite second moments. This problem can be circumvented by considering the representation

$$\mathcal{V}^2(X) = \Delta^2(X) + W(X), \tag{3.1}$$

where

$$W(X) = \mathbb{E}\Big[\|X - X'\| \cdot \big(\|X - X'\| - 2\,\|X - X''\|\big)\Big].$$

In the one-dimensional case, the representation (3.1) can be further simplified using the concept of order statistics.

Theorem 3.1. *Let X be a real-valued random variable with $\mathbb{E}|X| < \infty$, and let X, X', and X'' be i.i.d. copies of X. If $X_{1:3} \le X_{2:3} \le X_{3:3}$ are the order statistics of the triple (X, X', X'') then*

$$\mathcal{V}^2(X) = \Delta^2(X) - \tfrac{4}{3}\,\mathbb{E}[(X_{2:3} - X_{1:3})\,(X_{3:3} - X_{2:3})] \tag{3.2}$$
$$= \Delta^2(X) - 8\,\mathbb{E}[(X - X')_+\,(X'' - X)_+], \tag{3.3}$$

where $t_+ = \max(t, 0)$, $t \in \mathbb{R}$.

PROOF. We first prove the theorem for the case in which X is continuous. In this case, we apply the Law of Total Expectation and use the independence of the ranks and the order statistics [29, Lemma 13.1] to obtain

$$W(X)$$
$$= \mathbb{E}\Big[|X - X'|\,\big(|X - X'| - 2\,|X - X''|\big)\Big]$$
$$= \sum_{\substack{k,k',k''=1 \\ k,k',k''\,\text{are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\Big[|X - X'|\big(|X - X'| - 2|X - X''|\big)\Big|(r_X, r_{X'}, r_{X''}) = (k, k', k'')\Big]$$
$$\times \mathbb{P}\big((r_X, r_{X'}, r_{X''}) = (k, k', k'')\big).$$

Using the symmetry of X, X', and X'', it follows that

$$W(X) = \frac{1}{6} \sum_{\substack{k,k',k''=1 \\ k,k',k'' \text{ are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\Big[|X_{k:3} - X_{k':3}| \left(|X_{k:3} - X_{k':3}| - 2\,|X_{k:3} - X_{k'':3}|\right)\Big]$$

$$= \frac{1}{6} \sum_{\substack{k,k',k''=1 \\ k,k',k'' \text{ are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\Big[|X_{k:3} - X_{k':3}|^2\Big] - \mathbb{E}\Big[|X_{k:3} - X_{k':3}| \cdot |X_{k:3} - X_{k':3}|\Big].$$

Evaluating the first summand in the latter equation yields

$$\frac{1}{6} \sum_{\substack{k,k',k''=1 \\ k,k',k'' \text{ are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\big[|X_{k:3} - X_{k':3}|^2\big]$$

$$= \frac{1}{3}\left(\mathbb{E}\big[(X_{1:3} - X_{2:3})^2\big] + \mathbb{E}\big[(X_{1:3} - X_{3:3})^2\big] + \mathbb{E}\big[(X_{2:3} - X_{3:3})^2\big]\right).$$

Proceeding analogously with the second summand and simplifying the outcome, we obtain

$$W(X) = -\frac{4}{3}\,\mathbb{E}\big[(X_{2:3} - X_{1:3})\,(X_{3:3} - X_{2:3})\big].$$

This proves (3.2) in the continuous case.

For the case of general random variables, we now apply the method of quantile transformations. Let U be uniformly distributed on the interval $[0,1]$ and let U, U', and U'' be i.i.d.. Further, let F denote the cumulative distribution function of X. With $F^{-1}(p) = \inf\{x : F(x) \geq p\}$ denoting the right-continuous inverse of F, we define $\tilde{X} = F^{-1}(\tilde{U})$, $\tilde{X}' = F^{-1}(\tilde{U}')$, and $\tilde{X}'' = F^{-1}(\tilde{U}'')$. By [29, Theorem 21.1], the random variables $\tilde{X}$, $\tilde{X}'$, and $\tilde{X}''$ are i.i.d. copies of X and

$$W(X)$$
$$= \mathbb{E}\Big[|\tilde{X} - \tilde{X}'| \cdot \left(|\tilde{X} - \tilde{X}'| - 2\,|\tilde{X} - \tilde{X}''|\right)\Big]$$
$$= \sum_{\substack{k,k',k''=1 \\ k,k',k'' \text{ are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\Big[|\tilde{X} - \tilde{X}'| \cdot \left(|\tilde{X} - \tilde{X}'| - 2\,|\tilde{X} - \tilde{X}''|\right)\Big|(r_U, r_{U'}, r_{U''}) = (k, k', k'')\Big]$$
$$\times\; \mathbb{P}\big((r_U, r_{U'}, r_{U''}) = (k, k', k'')\big)$$

$$= \frac{1}{6} \sum_{\substack{k,k',k''=1 \\ k,k',k''\text{are pair-} \\ \text{wise distinct}}}^{3} \mathbb{E}\Big[|X_{k:3} - X_{k':3}| \cdot \big(|X_{k:3} - X_{k':3}| - 2\,|X_{k:3} - X_{k'':3}| \big) \Big]$$

$$= -\frac{4}{3}\,\mathbb{E}[(X_{2:3} - X_{1:3})\,(X_{3:3} - X_{2:3})].$$

The second representation for $W(X)$, (3.3), now follows by a combinatorial symmetry argument from the first representation. $\square$

In the continuous case with finite second moment, equation (3.3) is equivalent to

$$\mathbb{E}(|X - X'| \cdot |X'' - X'|) = \sigma^2(X) + 4J(X), \tag{3.4}$$

where

$$J(X) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{x} \int_{z=x}^{\infty} (x - y)\,(z - x)f(z)\,f(y)f(x)\,\mathrm{d}z\,\mathrm{d}y\,\mathrm{d}x.$$

Formula (3.4) is essentially the key result in the classical paper by Lomnicki [18], who also gave a simple expression for the variance of the empirical Gini mean difference,

$$\widehat{\Delta}_n(\boldsymbol{X}) = \frac{2}{n\,(n-1)} \sum_{1 \le i < j \le n} |X_i - X_j|. \tag{3.5}$$

Indeed, it is shown in [18] that

$$\mathrm{Var}\big(\widehat{\Delta}_n(\boldsymbol{X})\big) = \frac{1}{n\,(n-1)}\big(4\,(n-1)\,\sigma^2(X) + 16\,(n-2)J(X) - 2\,(2n-3)\Delta^2(X)\big). \tag{3.6}$$

We note two consequences of Theorem 3.1 and equation (3.6). First, Theorem 3.1 implies that the decomposition (3.6) holds in an analogous way for the non-continuous case. Second, for distributions with finite second moment, calculating the distance variance yields the variance of $\widehat{\Delta}_n$ and *vice versa*. These considerations imply that the asymptotic variance $ASV(\widehat{\Delta}(\boldsymbol{X})) = \lim_{n\to\infty} n\,\mathrm{Var}\big((\widehat{\Delta}(\boldsymbol{X}))\big)$ can be expressed alternatively as

$$ASV(\widehat{\Delta}(\boldsymbol{X})) = 4\,\sigma^2(X) - 2\,\mathcal{V}^2(X) - 2\,\Delta^2(X). \tag{3.7}$$

For a random sample $X_1, \ldots, X_n$ of real-valued random variables, the difference between successive order statistics, $D_{i:n} := X_{i+1:n} - X_{i:n}$, $i = 1, \ldots, n-1$, is called the ith *spacing* of $\boldsymbol{X} = (X_1, \ldots, X_n)$. Jones and Balakrishnan [15] (see also [30, 31]) studied closed-form expression for the moments of spacings and showed that

$$\sigma^2(X) = \mathbb{E}(D_{1:2}^2) = 2 \iint_{-\infty < x < y < \infty} F(x)(1 - F(y))\,\mathrm{d}x\,\mathrm{d}y \tag{3.8}$$

and

$$\Delta(X) = \mathbb{E}(D_{1:2}) = 2 \int_{-\infty}^{\infty} F(x)(1 - F(x))\,\mathrm{d}x. \tag{3.9}$$

By applying results in [15], we obtain an analogous representation for the distance variance.

Theorem 3.2. *Let X be a real-valued variable with $\mathbb{E}(|X|) < \infty$ and let X, X', X'', and X''' be i.i.d. Then,*

$$\mathcal{V}^2(X) = 8 \iint\limits_{-\infty < x < y < \infty} F^2(x)(1 - F(y))^2 \, \mathrm{d}x \, \mathrm{d}y \tag{3.10}$$

$$= \frac{2}{3} \mathbb{E}[(X_{3:4} - X_{2:4})^2], \tag{3.11}$$

where $X_{1:4} \leq X_{2:4} \leq X_{3:4} \leq X_{4:4}$ denote the order statistics of (X, X', X'', X''').

PROOF. By equation (3.9), we obtain

$$\Delta^2(X) = \left[2 \int_{-\infty}^{\infty} F(x) \left(1 - F(x)\right) \mathrm{d}x \right]^2$$

$$= 4 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(x) \left[1 - F(x)\right] F(y) \left[1 - F(y)\right] \mathrm{d}x \, \mathrm{d}y$$

$$= 8 \iint\limits_{-\infty < x < y < \infty} F(x) \left[1 - F(x)\right] F(y) \left[1 - F(y)\right] \mathrm{d}x \, \mathrm{d}y.$$

Moreover, by [15, equation (3.5)]

$$\mathbb{E}[(X_{2:3} - X_{1:3})(X_{3:3} - X_{2:3})]$$

$$= 8 \iint\limits_{-\infty < x < y < \infty} F(x) \left[F(y) - F(x)\right] \left[1 - F(y)\right] \mathrm{d}x \, \mathrm{d}y.$$

Hence,

$$\mathcal{V}^2(X) = \Delta^2(X) - \tfrac{4}{3} \mathbb{E}[(X_{(2)} - X_{(1)})(X_{(3)} - X_{(2)})]$$

$$= 8 \iint\limits_{-\infty < x < y < \infty} [F(x)]^2 \left[1 - F(y)\right]^2 \mathrm{d}x \, \mathrm{d}y,$$

which proves (3.10).

Finally, the formula (3.11) follows from (3.10) and from [15, equation (3.4)]. $\square$

Theorem 3.2 now yields for the distance variance a new sample version which is distinct from $\mathcal{V}_n^2(\boldsymbol{X})$, as follows.

Corollary 3.3. *Let X be a real-valued variable with $\mathbb{E}(|X|) < \infty$ and let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample from X. Then, a strongly consistent sample version for $\mathcal{V}^2(X)$ is*

$$\mathcal{U}_n^2(\boldsymbol{X}) = \binom{n}{2}^{-2} \sum_{i,j=1}^{n-1} \left(\min(i,j)\right)^2 \left(n - \max(i,j)\right)^2 D_{i:n} D_{j:n}, \tag{3.12}$$

where $D_{k:n} = X_{k+1:n} - X_{k:n}$ denotes the kth sample spacing of X, $1 \leq k \leq n-1$.

PROOF. Let $h : \mathbb{R}^4 \mapsto \mathbb{R}$ be the symmetric kernel defined by

$$h(X_1, \ldots, X_4) = \frac{2}{3}(X_{3:4} - X_{2:4})^2,$$

where $X_{1:4} \leq X_{2:4} \leq X_{3:4} \leq X_{4:4}$ are the order statistics of $X_1, \ldots, X_4$. By Theorem 3.2, we have $\mathbb{E}[h(X_1, \ldots, X_4)] < \infty$. Hence, by Hoeffding [12],

$$\widehat{\mathcal{U}}_n^2(\boldsymbol{X}) = \frac{2}{3}\binom{n}{4}^{-1} \sum_{1 \leq i_1 < i_2 < i_3 < i_4 \leq n} h(X_{i_1}, \ldots, X_{i_4})$$

is a strongly consistent estimator for $\mathcal{V}^2(X)$. Using a straightforward combinatorial calculation, we obtain

$$\widehat{\mathcal{U}}_n^2(\boldsymbol{X}) = \frac{2}{3}\binom{n}{4}^{-1} \sum_{1 \leq i < j \leq n} (i-1)(n-j)(X_{j:n} - X_{i:n})^2.$$

On inserting the definition of the spacings, the latter equation reduces to

$$\widehat{\mathcal{U}}_n^2(\boldsymbol{X}) = \frac{2}{3}\binom{n}{4}^{-1} \sum_{1 \leq i < j \leq n} (i-1)(n-j)(D_{i:n} + \cdots + D_{j-1:n})^2$$

$$\equiv \frac{2}{3}\binom{n}{4}^{-1} \sum_{1 \leq i < j \leq n} (i-1)(n-j) \sum_{k,l=i}^{j-1} D_{k:n}D_{l:n}.$$

Interchanging the above summations, we obtain

$$\widehat{\mathcal{U}}_n^2(\boldsymbol{X}) = \frac{2}{3}\binom{n}{4}^{-1} \sum_{k,l=1}^{n-1} D_{k:n}D_{l:n} \sum_{i=1}^{\min(k,l)} \sum_{j=\max(k,l)+1}^{n} (i-1)(n-j)$$

$$= \frac{1}{6}\binom{n}{4}^{-1} \sum_{k,l=1}^{n-1} D_{k:n}D_{l:n} \min(k,l)\big(\min(k,l) - 1\big)$$

$$\times \big(n - \max(k,l)\big)\big(n - \max(k,l) - 1\big),$$

where the latter equality follows from the fact that $\sum_{i=1}^{k} i = k(k-1)/2$. Since

$$\frac{1}{6}\binom{n}{4}^{-1} = \frac{4}{n(n-1)(n-2)(n-3)},$$

then we deduce that $\mathcal{U}_n^2(\boldsymbol{X}) = \widehat{\mathcal{U}}_n^2(\boldsymbol{X}) + o(1)$. This completes the proof. □

Denoting the vector of spacings by $D = (D_{1:n}, \ldots, D_{n-1:n})$, we can write the quadratic form in (3.12) as

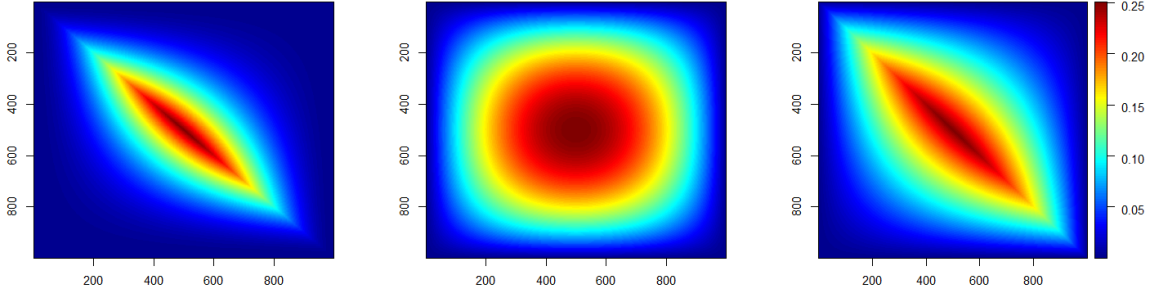$$\mathcal{U}_n^2(\boldsymbol{X}) = D^t V D,$$

Figure 3.1: Illustration of (from left to right) the sample distance variance $\mathcal{U}_n^2$, the squared sample Gini mean difference $\widehat{\Delta}^2$, and the sample variance $\widehat{\sigma}^2$ *via* their respective quadratic form matrices V, G, and S for sample size $n = 1,000$. The coordinate (i, j) corresponds to the (i, j)th entry of the corresponding matrix, and the size of the corresponding matrix element is specified *via* color code (see legend).

where the (i, j)th element of the matrix V is

$$V_{i,j} = \binom{n}{2}^{-2} \big(\min(i, j) \big)^2 \big(n - \max(i, j) \big)^2$$

Both the squared sample Gini mean difference and the sample variance

$$\widehat{\sigma}_n^2(\boldsymbol{X}) := \frac{1}{n(n-1)} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$$

can also be expressed as quadratic forms in the spacings vector D; specifically,

$$\widehat{\Delta}_n^2(\boldsymbol{X}) = D^t\, G\, D, \qquad \widehat{\sigma}_n^2(\boldsymbol{X}) = D^t\, S\, D,$$

where the elements of G and S are given by

$$G_{i,j} = \binom{n}{2}^{-2} i\, j\, (n - i)\, (n - j)$$

and

$$S_{i,j} = \frac{1}{2} \binom{n}{2}^{-1} \min(i, j)\, (n - \max(i, j)).$$

Hence, comparing $\mathcal{U}_n^2$, Δ_n^2, and σ_n^2 is equivalent to comparing the matrices V, G and S. We use this fact to graphically illustrate differing features of $\mathcal{V}$, Δ, and σ by plotting the values of the underlying matrices; see Figure 3.1.

Moreover, these quadratic form representations lead to the rediscovery of results from Section 2. For example, since V and G have the same diagonal entries then it

follows that $\mathcal{V}$ and Δ are equal for Bernoulli-distributed random variables. Also, if n is even then the elements $V_{n/2,n/2}$, $G_{n/2,n/2}$, and $S_{n/2,n/2}$ all coincide, representing the fact that the underlying measures coincide for the Bernoulli distribution with $p = \frac{1}{2}$. Finally, since $V_{ij} \leq G_{ij}$ and $V_{ij} \leq S_{ij}$ for all i, j then we obtain an alternative proof of Corollary 2.3.

It is also remarkable that V is twice the second Hadamard power of S and that V and S both are positive definite, while G is positive semidefinite with rank 1. Finally, we mention that there are numerous other statistics which can be written as quadratic forms or square-roots of quadratic forms in the spacings, e.g., the Greenwood statistic, the range, and the interquartile range.

4 Closed form expressions for the distance variance of some well-known distributions

Exploiting the different representations of the distance variance derived in the preceding sections, we can now state the distance variance of many well-known distributions. In the following result, we use the standard notation $_1F_1$ and $_2F_1$ for the classical confluent and Gaussian hypergeometric functions.

Theorem 4.1. *1. Let X be Bernoulli distributed with parameter p. Then $\mathcal{V}^2(X) = 4\,p^2\,(1-p)^2$.*

2. Let X be normally distributed with mean μ and variance σ^2. Then

$$\mathcal{V}^2(X) = 4\Big(\frac{1-\sqrt{3}}{\pi} + \frac{1}{3}\Big)\sigma^2.$$

3. Let X be uniformly distributed on the interval $[a,b]$. Then $\mathcal{V}^2(X) = 2(b-a)^2/45$.

4. Let X be Laplace-distributed with density function, $f_X(x) = (2\alpha)^{-1}$ $\exp(-|x-\mu|/\alpha)$, $x \in \mathbb{R}$, $\alpha > 0$, $\mu \in \mathbb{R}$. Then $\mathcal{V}^2(X) = 7\alpha^2/12$.

5. Let X be Pareto-distributed with parameters $\alpha > 1$ and $x_m > 0$, and density function $f_X(x) = \alpha x_m^\alpha x^{-(\alpha+1)}$, $x \geq x_m$. Then,

$$\mathcal{V}^2(X) = \frac{4\alpha^2 x_m^2}{(\alpha-1)\,(2\alpha-1)^2\,(3\alpha-2)}$$

6. Let X be exponentially distributed with parameter $\lambda > 0$ and density function $f_X(x) = \lambda\exp(-\lambda\,x)$, $x \geq 0$. Then, $\mathcal{V}^2(X) = (3\lambda^2)^{-1}$.

7. Let X be Gamma-distributed with shape parameter $\alpha > 0$ and scale parameter 1. Then
$$\mathcal{V}^2(X) = 2^{2(2-2\alpha)} \sum_{j,k=1}^{\infty} A_{j,k}(\alpha)^2,$$

where

$$A_{j,k}(\alpha) = 2^{-j-k} \left(\frac{(\alpha)_j \, (\alpha)_k}{j! \, k!} \right)^{1/2}$$
$$\times \frac{\Gamma(2\alpha + j + k - 1)}{\Gamma(\alpha + j) \, \Gamma(\alpha + k)} \, {}_2F_1\left(-j - k + 2, 1 - \alpha - j; 2 - 2\alpha - j - k; 2\right).$$

8. Let X be Poisson-distributed with parameter $\lambda > 0$. Then
$$\mathcal{V}^2(X) = \sum_{j,k=1}^{\infty} \frac{4^{j+k-1}}{j! \, k!} \lambda^{j+k} \, A_{jk}^2,$$

where

$$A_{jk} = \frac{1}{(j-1)!} \sum_{l=0}^{\lfloor (j-k)/2 \rfloor} \binom{j-k}{2l} (-1)^l (\tfrac{1}{2})_l \, (\tfrac{1}{2})_{j-l-1} \, {}_1F_1(j - l - \tfrac{1}{2}; j; -4\lambda).$$

9. Let X be negative binomially distributed with parameters c and β. Then
$$\mathcal{V}^2(X) = (1 - c)^{4\beta} \sum_{j,k=1}^{\infty} \frac{(\beta)_j \, (\beta)_k}{j! \, k!} (1 + c^2)^{-2\beta - 2j} 2^{2k} c^{j+k} A_{jk}^2,$$

where

$$A_{jk} = \sum_{l_1, l_2=0}^{j-k} \binom{j-k}{l_1}\binom{j-k}{l_2} (-c)^{l_1} (-1)^{l_2} (|l_1 - l_2|)! \sum_{l=0}^{\infty} \frac{(\beta + j) - l}{l!} \left(\frac{2c}{1 + c^2} \right)^l$$
$$\times \sum_{m=0}^{|l_1-l_2|} (-2)^m \frac{(m)_{|l_1-l_2|}}{(|l_1 - l_2| - m)! \, (2m)!} \frac{2^{k+m-1} \, (\tfrac{1}{2})_{k+m-1}}{(k + m - 1)!}$$
$$\times \, {}_2F_1(-l, k + m - \tfrac{1}{2}; k + m; 2).$$

10. Let $X = (X_1, \ldots, X_p)$ be a multivariate normally distributed random vector with mean $\mu = (\mu_1, \ldots, \mu_p)$ and identity covariance matrix $I_p := \mathrm{diag}(1, \ldots, 1)$. Then
$$\mathcal{V}^2(X) = 4\pi \frac{c_{p-1}^2}{c_p^2} \left[\frac{\Gamma(\tfrac{1}{2}p) \, \Gamma(\tfrac{1}{2}p + 1)}{\left[\Gamma\left(\tfrac{1}{2}(p+1)\right) \right]^2} - 2 \, {}_2F_1\left(-\tfrac{1}{2}, -\tfrac{1}{2}; \tfrac{1}{2}p; \tfrac{1}{4}\right) + 1 \right].$$

PROOF. 1. See Theorem 2.5.

2. See the proof of Theorem 7 in [25] or [4, p. 14].

3. and 4. These follow directly from Theorem 3.1 and the results in Table 3 in [10].

5. and 6. These results follow directly from the representation (2.1) and [32, equations (4.2) and (4.4)].

7., 8., and 9. See [6, Propositions 5.6, 5.7, and 5.8].

10. See [4, Corollary 3.3]. □

By equations (3.6) and (3.7), we can also derive expresssions for the variance and asymptotic variance of the sample Gini mean difference for the distributions 1.- 9. in Theorem 4.1. To the best of our knowledge, these expressions are novel for the Gamma, Poisson, and negative binomial distributions.

5 The distance standard deviation as a measure of spread

In this section, we show that the distance standard deviation $\mathcal{V}(X)$ satisfies the criteria (C1)-(C3) stated in Section 1 and therefore is an axiomatic measure of spread in the sense of [2]. Moreover, we point out some differences and commonalities between $\mathcal{V}$, Δ and σ. First, we state some additional preliminaries about stochastic orders.

Definition 5.1 ([24], Section 1.A.1)**.** A random variable X is said to be *stochastically smaller than* a random variable Y, or X is *smaller than Y in the stochastic ordering*, written $X \leq_{\mathrm{st}} Y$, if $\mathbb{P}(X > u) \leq \mathbb{P}(Y > u)$ for all $u \in \mathbb{R}$.

Proposition 5.2 ([24], Section 1.A.1)**.** *A necessary and sufficient condition that $X \leq_{st} Y$ is that*

$$\mathbb{E}[\phi(X)] \leq \mathbb{E}[\phi(Y)] \tag{5.1}$$

for all increasing functions ϕ for which these expectations exist.

Another important ordering of random variables is the *dispersive order*, $\leq_{\mathrm{disp}}$, which was stated earlier at (1.6) in the introduction. Bartoszewicz [1] proved the following result.

Proposition 5.3 ([1], Proposition 3)**.** *Let $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_n)$ be random samples from the random variables X and Y, respectively, and let $D_j = X_{j+1:n} - X_{j:n}$ and $E_j = Y_{j+1:n} - Y_{j:n}$, $j = 1, \ldots, n-1$ denote the corresponding sample spacings. If $X \leq_{\mathrm{disp}} Y$ then $D_{j:n} \leq_{\mathrm{st}} E_{j:n}$ for all $j = 1, \ldots, n-1$.*

Applying this result to the representation of the distance variance derived in Theorem 3.2, we conclude that the distance standard deviation $\mathcal{V}$ is indeed a measure of spread in the sense of [2].

Theorem 5.4. *If $X \leq_{\mathrm{disp}} Y$ then $\mathcal{V}(X) \leq \mathcal{V}(Y)$.*

PROOF. Let us consider i.i.d. replicates (X, Y), (X', Y'), (X'', Y''), and (X''', Y'''). Moreover, let $X_{1:4} \leq X_{2:4} \leq X_{3:4} \leq X_{4:4}$ and $Y_{1:4} \leq Y_{2:4} \leq Y_{3:4} \leq Y_{4:4}$ denote the respective order statistics. By Proposition 5.3,

$$(X_{3:4} - X_{2:4}) \leq_{\mathrm{st}} (Y_{3:4} - Y_{2:4}).$$

Applying equation (5.1) and Theorem 3.2 concludes the proof. $\quad\square$

Using similar arguments, we can show that the result of Theorem 5.4 holds analogously for the standard deviation and Gini's mean difference; see also [16].

Theorem 5.5 ([24], Theorem 3.B.7). *The random variable X satisfies the property*

$$X \leq_{\mathrm{disp}} X + Y \text{ for any random variable } Y \text{ which is independent of } X$$

if and only if X has a log-concave density.

Applying Theorem 5.5, we obtain the following corollary of Theorem 5.4.

Corollary 5.6. *Let X be a random variable with a log-concave density. Then*

$$\mathcal{V}(X + Y) \geq \mathcal{V}(X)$$

for any random variable Y independent of X.

In particular if X and Y are independently distributed, continuous, random variables with log-concave densities, then

$$\mathcal{V}(X + Y) \geq \max(\mathcal{V}(X), \mathcal{V}(Y)). \tag{5.2}$$

It is well known, both for the standard deviation and for Gini's mean difference, that analogous assertions hold without any restrictions on the distributions of X and Y. In particular, for any pair of independent random variables X and Y with existing first or second moments, respectively, there holds

$$\sigma^2(X + Y) = \sigma^2(X) + \sigma^2(Y) \geq \max(\sigma^2(X), \sigma^2(Y)). \tag{5.3}$$

Also, letting X' and Y' denote i.i.d. copies of X and Y, respectively, we have

$$\Delta(X + Y) = \mathbb{E}[\max(|X - X'|, |Y - Y'|)] \geq \max(\Delta(X), \Delta(Y)). \tag{5.4}$$

However we now show that this property does not hold generally for the distance standard deviation, $\mathcal{V}$, thereby answering a second question raised by Gabor Székely (private communication, November 23, 2015).

Example 5.7. *Let X be Bernoulli distributed with parameter $p = \frac{1}{2}$ and let Y be uniformly distributed on the interval $[0, 1]$ and independent of X. Then $\mathcal{V}(X) > \mathcal{V}(X + Y)$.*

PROOF. By a straightforward calculation using (2.1), we obtain

$$\mathcal{V}^2(X + Y) = T_1(X + Y) + T_2(X + Y) - 2\,T_3(X + Y)$$
$$= \frac{2}{3} + \frac{4}{9} - \frac{14}{15} = \frac{8}{45}.$$

However, by Theorem 2.5, $\mathcal{V}^2(X) = 1/4 > \mathcal{V}^2(X + Y)$.   □

Other common properties of the classical standard deviation and Gini's mean difference concerns differences and sums of independent random variables. From the representations of $\sigma^2(X + Y)$ and $\Delta(X + Y)$ given in (5.3) and (5.4), we see that

$$\Delta(X + Y) = \Delta(X - Y), \qquad \sigma(X + Y) = \sigma(X - Y)$$

for any independent random variables X and Y for which these expressions exist. On the other hand, these properties do not hold in general for the distance standard deviation.

Example 5.8. *Let X and Y be independently Bernoulli distributed with parameter $p \neq \frac{1}{2}$. Then $\mathcal{V}(X + Y) > \mathcal{V}(X - Y)$.*

PROOF. By a straightforward calculation using (2.1), we obtain

$$\mathcal{V}^2(X + Y) = 8\,(p - p^2)^2\,\left(2\,(p - p^2)^2 - 6\,(p - p^2) + 2\right)$$

and

$$\mathcal{V}^2(X - Y) = 8\,(p - p^2)^2\,\left(2\,(p - p^2)^2 - 2\,(p - p^2) + 1\right).$$

Hence,

$$\mathcal{V}^2(X + Y) - V^2(X - Y) = 8\,(p - p^2)^2\,(1 - 2p)^2,$$

and this difference obviously is positive for $p \neq \frac{1}{2}$.   □

However, an analogous property holds when either of the two variables has a symmetric distribution.

Theorem 5.9. *Let X and Y be independent real random variables with $\mathbb{E}|X + Y| < \infty$. Then $\mathcal{V}^2(X + Y) = \mathcal{V}^2(X - Y)$ if either X or Y is symmetric about μ.*

PROOF. Since $\mathcal{V}(X - Y) = \mathcal{V}(Y - X)$ then we can assume, without loss of generality, that Y is symmetric. Moreover since $\mathcal{V}^2(X + \mu) = \mathcal{V}^2(X)$ then we can assume that the point of symmetry is at 0.

By equation (1.2),

$$
\begin{aligned}
\mathcal{V}^2&(X-Y) \\
&= \frac{1}{\pi^2} \int_{\mathbb{R}^2} \left| f_{X-Y}(s+t) - f_X(s) f_{-Y}(t) \right|^2 \frac{\mathrm{d}s\,\mathrm{d}t}{|s|^2\,|t|^2} \\
&= \frac{1}{\pi^2} \int_{\mathbb{R}^2} \left| f_X(s+t)\,f_{-Y}(s+t) - f_X(s) f_{-Y}(s) f_X(t) f_{-Y}(t) \right|^2 \frac{\mathrm{d}s\,\mathrm{d}t}{|s|^2\,|t|^2} \\
&= \frac{1}{\pi^2} \int_{\mathbb{R}^2} \left| f_X(s+t)\,f_Y(s+t) - f_X(s) f_Y(s) f_X(t) f_Y(t) \right|^2 \frac{\mathrm{d}s\,\mathrm{d}t}{|s|^2\,|t|^2} \\
&= \mathcal{V}^2(X+Y),
\end{aligned}
$$

where the third equality follows from the fact that Y and $-Y$ have the same distribution. $\quad\square$

6   The distance standard deviation as an estimator

In this section, we investigate the properties of the sample distance variance $\mathcal{V}_n^2(\boldsymbol{X})$ and the sample distance standard deviation $\mathcal{V}_n(\boldsymbol{X})$ as estimators and derive their asymptotic distributions. For these purposes, we employ the representation (3.1), *viz.*,

$$
\mathcal{V}^2(X) = W(X) + \Delta^2(X)
$$

where

$$
W(X) = \mathbb{E}\left[\|X - Y\| \left(\|X - Y\| - 2\,\|X - Z\|\right)\right],
$$

and $\Delta(X)$ denotes the population value of Gini's mean difference. Throughout this section, X, Y, Z are i.i.d. p-variate random vectors with distribution F. For a sample of i.i.d. random vectors $\boldsymbol{X} = (X_1, \ldots, X_n)^t$, each with distribution F, we define the corresponding empirical quantities,

$$
\Delta_n(\boldsymbol{X}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \|X_i - X_j\|
$$

and

$$
W_n(\boldsymbol{X}) = \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} \|X_i - X_j\| \left(\|X_i - X_j\| - 2\|X_i - X_k\|\right).
$$

Note that

$$
W_n(\boldsymbol{X}) = T_{1,n}(\boldsymbol{X}) - 2\,T_{3,n}(\boldsymbol{X}),
$$

cf. (2.4), and

$$
\mathcal{V}_n^2(\boldsymbol{X}) = W_n(\boldsymbol{X}) + \Delta_n^2(\boldsymbol{X}).
$$

Further, it is straightforward to verify that

$$\mathbb{E}W_n(\boldsymbol{X}) = \frac{(n-1)(n-2)}{n^2}W(X). \tag{6.1}$$

The statistic $\Delta_n(\boldsymbol{X})$ does not wear a hat to distinguish it from $\widehat{\Delta}_n(\boldsymbol{X})$, the unbiased version of the sample Gini difference, defined for univariate observations in (3.5) and which we extend to multivariate observations by replacing the absolute value $|\cdot|$ by the Euclidean norm $\|\cdot\|$; thus,

$$\Delta_n^2(\boldsymbol{X}) = \frac{(n-1)^2}{n^2}\widehat{\Delta}_n^2(\boldsymbol{X}).$$

Similar to $\widehat{\Delta}_n(\boldsymbol{X})$, we define

$$\widehat{W}_n(\boldsymbol{X}) = \frac{1}{n(n-1)(n-2)} \sum_{\substack{1 \le i,j,k \le n \\ i \neq j, j \neq k, k \neq i}} \|X_i - X_j\|\,(\|X_i - X_j\| - 2\|X_i - X_k\|)$$

$$\equiv \frac{n^2}{(n-1)(n-2)}W_n(\boldsymbol{X}).$$

By (6.1), $\widehat{W}_n(\boldsymbol{X})$ is an unbiased sample version of $W(X)$.

Also, $\widehat{\mathcal{V}}_n(\boldsymbol{X}) = [\widehat{W}_n(\boldsymbol{X}) + \widehat{\Delta}_n^2(\boldsymbol{X})]^{1/2}$ is an alternative to $\mathcal{V}_n(\boldsymbol{X})$ as an empirical version of the distance standard deviation $\mathcal{V}(X)$. Although $\widehat{\mathcal{V}}_n(\boldsymbol{X})$ is based on the unbiased estimators $\widehat{W}_n(\boldsymbol{X})$ and $\widehat{\Delta}_n(\boldsymbol{X})$, the estimator $\widehat{\mathcal{V}}_n(\boldsymbol{X})$ itself has a larger bias than $\mathcal{V}_n(\boldsymbol{X})$. The results of Table 2 below indicate that $\mathcal{V}_n(\boldsymbol{X})$ is to be preferred over $\widehat{\mathcal{V}}_n(\boldsymbol{X})$ as an estimator of $\mathcal{V}(X)$ because it exhibits smaller finite-sample bias and smaller variance for scenarios considered in our simulations.

Nevertheless, $\mathcal{V}_n^2(\boldsymbol{X})$ and $\widehat{\mathcal{V}}_n^2(\boldsymbol{X})$ have the same asymptotic distribution. In order to establish that result, we define for $x \in \mathbb{R}^p$,

$$\begin{aligned} \psi_1(x) &= \mathbb{E}\|x - Y\|^2, & \psi_2(x) &= \mathbb{E}(\|x - Y\| \cdot \|x - Z\|), \\ \psi_3(x) &= \mathbb{E}(\|x - Y\| \cdot \|Y - Z\|), & \psi_4(x) &= \mathbb{E}\|x - Y\|, \end{aligned}$$

and, with $T_1(X)$, $T_2(X)$, and $T_3(X)$ as defined in (2.2)-(2.3), we also define

$$\begin{aligned} m_{11} &= 4\,\mathbb{E}\big[\psi_1(X) - \psi_2(X) - 2\psi_3(X)\big]^2 - 4\big(T_1(X) - 3T_3(X)\big)^2, \\ m_{12} &= 4\,\mathbb{E}\big[\psi_4(X)\big(\psi_1(X) - \psi_2(X) - 2\psi_3(X)\big)\big] - 4T_2(X)\big(T_1(X) - 3T_3(X)\big), \quad (6.2) \\ m_{22} &= 4\,\mathbb{E}\psi_4^2(X) - 4\big(T_2(X)\big)^2. \end{aligned}$$

and let

$$\gamma = m_{11} + 4m_{12}\Delta(X) + 4m_{22}\Delta^2(X).$$

We now provide in the following result the asymptotic distribution of $\mathcal{V}_n^2(\boldsymbol{X})$ and $\widehat{\mathcal{V}}_n^2(\boldsymbol{X})$.

Theorem 6.1. *Suppose that $\mathbb{E}(\|X\|^4) < \infty$. Then, as $n \to \infty$,*

$$\sqrt{n}\big(\mathcal{V}_n^2(\boldsymbol{X}) - \mathcal{V}^2(X)\big) \xrightarrow{d} N(0, \gamma) \tag{6.3}$$

and the same result holds for $\widehat{\mathcal{V}}_n^2(\boldsymbol{X})$.

PROOF. Consider the bivariate statistic $\widehat{B}_n(\boldsymbol{X}) = \big(\widehat{W}_n(\boldsymbol{X}), \widehat{\Delta}_n(\boldsymbol{X})\big)^t$, which has expected value $B(X) = (W(X), \Delta(X))^t$. Define the functions $K, L : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ such that

$$K(x, y, z) = \|x - y\|(\|x - y\| - 2\|x - z\|) + \|y - z\|(\|y - z\| - 2\|y - x\|)$$
$$+ \|z - x\|(\|z - x\| - 2\|z - y\|)$$

and

$$L(x, y, z) = (\|x - y\| + \|y - z\| + \|z - x\|),$$

$(x, y, z) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$. Then the statistic $\widehat{B}_n(\boldsymbol{X})$ can be written as a U-statistic with the bivariate, permutation-symmetric kernel of order three, $h : \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^2$, where

$$h(x, y, z) = \frac{1}{3}\begin{pmatrix} K(x, y, z) \\ L(x, y, z) \end{pmatrix},$$

$(x, y, z) \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^p$. Define the function $h_1 : \mathbb{R}^p \to \mathbb{R}^2$, where $h_1(x) = \mathbb{E}h(x, Y, Z) - B(X)$; then, h_1 is the linear part in the Hoeffding decomposition of the kernel h, and we calculate that

$$h_1(x) = \frac{2}{3}\begin{pmatrix} \psi_1(x) - \psi_2(x) - 2\psi_3(x) - T_1(X) + 3T_3(X) \\ \psi_4(x) - T_2(X) \end{pmatrix},$$

$x \in \mathbb{R}^p$. Since $\mathbb{E}(\|X\|)^4 < \infty$ then $\mathbb{E}[(h(X, Y, Z))^2] < \infty$; therefore, we deduce from a classical result of Hoeffding [11, Theorem 7.1] that

$$\sqrt{n}\big(\widehat{B}_n(\boldsymbol{X}) - B(X)\big) \xrightarrow{d} N_2\big(0, 9\,\mathbb{E}h_1(\boldsymbol{X})h_1(\boldsymbol{X})^t\big).$$

Denote the symmetric 2×2 matrix $9\,\mathbb{E}h_1(\boldsymbol{X})h_1(\boldsymbol{X})^t$ by $M = (m_{ij})_{i,j=1,2}$, where the elements m_{11}, m_{12}, and m_{22} are given in (6.2). Define $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(x, y) = x + y^2$; then $\widehat{\mathcal{V}}_n^2(\boldsymbol{X}) = g\big(\widehat{B}_n(\boldsymbol{X})\big)$. Since $\nabla h(x, y) = (1, 2y)^t$ then, by applying the Delta Method, we obtain $\sqrt{n}\big(\widehat{\mathcal{V}}_n^2(\boldsymbol{X}) - \mathcal{V}^2(X)\big) \xrightarrow{d} N(0, \gamma)$.

In the case of $\mathcal{V}_n^2(\boldsymbol{X})$, we need only to apply the formulas $W_n(\boldsymbol{X}) = (n-1)(n-2)\widehat{W}_n(\boldsymbol{X})/n^2$ and $\Delta_n^2(\boldsymbol{X}) = (n-1)^2\widehat{\Delta}_n^2(\boldsymbol{X})/n^2$ to deduce that $\mathcal{V}_n^2(\boldsymbol{X}) - \widehat{\mathcal{V}}_n^2(\boldsymbol{X}) = o(n^{-1})$. Then it follows by the Delta Method that $\mathcal{V}_n^2(\boldsymbol{X})$ has the same asymptotic distribution as $\mathcal{V}_n^2(\boldsymbol{X})$, as given in (6.3). $\square$

The asymptotic distribution of the sample distance standard deviation $\mathcal{V}_n(\boldsymbol{X})$ now follows from Theorem 6.1 by the Delta Method:

Distribution, F	$ARE(\mathcal{V}_n; F)$	$ARE(\widehat{\sigma}_n; F)$	$ARE(\widehat{d}_n; F)$	$ARE(\widehat{\Delta}_n; F)$
$N(0,1)$	0.784	1	0.876	0.978
$L(0,1)$	0.952	0.8	1	0.964
t_5	0.992	0.4	0.941	0.859
t_3	0.965	0	0.681	0.524

Table 1: Asymptotic relative efficiencies with respect to the respective maximum likelihood estimators of the distance standard deviation $\mathcal{V}_n$, the standard deviation $\widehat{\sigma}_n$, the mean devation $\widehat{d}_n$, and Gini's mean difference $\widehat{\Delta}_n$ at the normal distribution, the Laplace distribution, and the t_ν-distributions with $\nu = 5$ and $\nu = 3$.

Corollary 6.2. *Under the conditions of Theorem 6.1, we have*

$$\sqrt{n}(\mathcal{V}_n(\boldsymbol{X}) - \mathcal{V}(X)) \xrightarrow{d} N\big(0, \gamma/4\mathcal{V}^2(X)\big),$$

and the same result holds for $\widehat{\mathcal{V}}_n(\boldsymbol{X})$.

In the following, we study the empirical distance standard deviation $\mathcal{V}_n(\boldsymbol{X})$ as an estimator of spread in the univariate case. For any $\sqrt{n}$-consistent and asymptotically normal estimator $s_n(\boldsymbol{X})$, we define its asymptotic variance $ASV(s_n(\boldsymbol{X}); F)$ at the distribution F to be the variance of the limiting distribution of $\sqrt{n}(s_n(\boldsymbol{X}) - s)$, as $n \to \infty$, where $s_n(\boldsymbol{X})$ is evaluated at an i.i.d. sequence drawn from F and s denotes the corresponding population value of $s_n(\boldsymbol{X})$ at F. Any estimators $s_n^{(1)}(\boldsymbol{X})$ and $s_n^{(2)}(\boldsymbol{X})$ which estimate possibly different population values s_1 and s_2, respectively, at a given distribution F, and which obey the dilation property (C2) in Section 1, can be compared efficiency-wise by standardizing them through their respective population values. We define the *asymptotic relative efficiency of $s_n^{(1)}(\boldsymbol{X})$ with respect to $s_n^{(2)}(\boldsymbol{X})$* at the population distribution F as

$$ARE\big(s_n^{(1)}(\boldsymbol{X}), s_n^{(2)}(\boldsymbol{X}); F\big) = \frac{ASV(s_n^{(1)}(\boldsymbol{X}); F)/s_1^2}{ASV(s_n^{(2)}(\boldsymbol{X}); F)/s_2^2}.$$

Even in the univariate case with normally distributed data, the integrals underlying the parameters m_{11}, m_{12}, and m_{22} in (6.2) do not admit straightforward analytical expressions. Nevertheless, by means of numerical integration, we can obtain values for the asymptotic variance of $\mathcal{V}_n(\boldsymbol{X})$ for given population distributions and thus deduce properties of the efficiency of $\mathcal{V}_n(\boldsymbol{X})$ in relation to other widely-used estimators of scale.

In Table 1, we provide the asymptotic relative efficiency of the distance standard deviation with respect to the respective maximum likelihood estimator at the normal distribution, the Laplace distribution, and t_ν distributions with $\nu = 5$ and

Distribution, F		Sample size, n
		5	10	50	500	∞
$N(0,1)$	$\mathbb{E}(\mathcal{V}_n)$	0.663	0.658	0.640	0.634	0.633
	$\mathbb{E}(\widehat{\mathcal{V}}_n)$	0.701	0.665	0.639	0.634	0.633
	$n\mathrm{Var}(\mathcal{V}_n)$	0.297	0.276	0.255	0.255	0.256
	$n\mathrm{Var}(\widehat{\mathcal{V}}_n)$	0.359	0.298	0.259	0.255	0.256
$L(0,1)$	$\mathbb{E}(\mathcal{V}_n)$	0.888	0.861	0.790	0.767	0.764
	$\mathbb{E}(\widehat{\mathcal{V}}_n)$	0.942	0.864	0.785	0.766	0.764
	$n\mathrm{Var}(\mathcal{V}_n)$	0.955	0.836	0.668	0.605	0.613
	$n\mathrm{Var}(\widehat{\mathcal{V}}_n)$	1.136	0.858	0.663	0.604	0.613
t_5	$\mathbb{E}(\mathcal{V}_n)$	0.818	0.799	0.744	0.727	0.725
	$\mathbb{E}(\widehat{\mathcal{V}}_n)$	0.866	0.804	0.741	0.727	0.725
	$n\mathrm{Var}(\mathcal{V}_n)$	0.761	0.632	0.474	0.432	0.424
	$n\mathrm{Var}(\widehat{\mathcal{V}}_n)$	0.931	0.655	0.471	0.432	0.424
t_3	$\mathbb{E}(\mathcal{V}_n)$	1.003	0.960	0.861	0.817	0.810
	$\mathbb{E}(\widehat{\mathcal{V}}_n)$	1.074	0.967	0.855	0.816	0.810
	$n\mathrm{Var}(\mathcal{V}_n)$	5.762	2.001	1.089	0.777	0.680
	$n\mathrm{Var}(\widehat{\mathcal{V}}_n)$	8.420	2.177	1.067	0.774	0.680

Table 2: Simulated finite-sample values of the mean and the variance of the distance standard deviation $\mathcal{V}_n(\boldsymbol{X})$ for $n = 5, 10, 50, 500$ compared to asymptotic values (last column); $\widehat{\mathcal{V}}_n(\boldsymbol{X})$ refers to the version based on the unbiased estimates $\widehat{W}_n(\boldsymbol{X})$ and $\widehat{\Delta}_n(\boldsymbol{X})$; 10 000 repetitions.

$\nu = 3$. The maximum likelihood estimator of scale in the location-scale family generated by the $N(0,1)$, or standard normal, distribution is the standard deviation. In the Laplace model, the analogous estimator of scale is the mean deviation $\widehat{d}_n(\boldsymbol{X}) = n^{-1}\sum_{i=1}^n |X_i - m_n(\boldsymbol{X})|$, where $m_n(\boldsymbol{X})$ denotes the sample median of $\boldsymbol{X}$. In the case of the t_ν-distribution, the maximum likelihood estimator of the scale parameter does generally not admit an explicit representation.

In Table 1, the asymptotic efficiencies of the distance standard deviation are compared with those of the standard deviation $\widehat{\sigma}_n(\boldsymbol{X})$, the mean deviation $\widehat{d}_n(\boldsymbol{X})$, and Gini's mean difference $\widehat{\Delta}_n(\boldsymbol{X})$. The asymptotic variance of the maximum likelihood estimator of the scale parameter for the t_ν-distribution is $(\nu + 3)/2\nu$. The population values and asymptotic variances of the other estimators mentioned at the respective distributions are given by Gerstenberger and Vogel [10, Tables 2 and 3].

While the distance standard deviation has moderate efficiency at normality, it turns out to be asymptotically very efficient in the case of heavier-tailed populations. For the t_5- and t_3-distributions, the distance standard deviation outperforms its three competitors considered here and moreover is very close to the respective maximum likelihood estimator.

In Table 2, we complement our asymptotic analysis by finite-sample simulations. For sample sizes $n = 5, 10, 50$, and 500 and the same population distributions as above, the (simulated) expectations and variances (based on 10,000 observations) of the empirical distance standard deviation $\mathcal{V}_n(\boldsymbol{X}) = [W_n(\boldsymbol{X}) + \Delta_n^2(\boldsymbol{X})]^{1/2}$ and the alternative version $\widehat{\mathcal{V}}_n(\boldsymbol{X}) = [\widehat{W}_n(\boldsymbol{X}) + \widehat{\Delta}_n^2(\boldsymbol{X})]^{1/2}$ are given along with their respective asymptotic values. The corresponding values for the competing estimators $\widehat{\sigma}_n(\boldsymbol{X})$, $\widehat{d}_n(\boldsymbol{X})$, and $\widehat{\Delta}_n(\boldsymbol{X})$ are also provided by Gerstenberger and Vogel [10]. Table 2 indicates that $\mathcal{V}_n(\boldsymbol{X})$ indeed is to be preferred over $\widehat{\mathcal{V}}_n(\boldsymbol{X})$ in terms of bias as well as variance.

References

[1] Bartoszewicz, J. (1986). Dispersive ordering and the total time on test transformation. *Statist. Probab. Lett.*, **4**, 285–288.

[2] Bickel, P. J., and Lehmann, E. L. (2012). Descriptive statistics for nonparametric models, IV. Spread. In: *Selected Works of E. L. Lehmann* (J. Rojo, editor), 519–526. Springer, New York.

[3] David, H. A., and Nagaraja, H. N. (1981). *Order Statistics*. Wiley, New York.

[4] Dueck, J., Edelmann, D., Gneiting, T., and Richards, D. (2014). The affinely invariant distance correlation. *Bernoulli*, **20**, 2305–2330.

[5] Dueck, J., Edelmann, D., and Richards, D. (2015). A generalization of an integral arising in the theory of distance correlation. *Statist. Probab. Lett.*, **97**, 116–119.

[6] Dueck, J., Edelmann, D., and Richards, D. (2017). Distance correlation coefficients for Lancaster distributions, *J. Multivariate Anal.*, **154**, 19–39.

[7] Feuerverger, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.*, **61**, 419–433.

[8] Fiedler, J. (2016). *Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields*. Doctoral dissertation, University of Heidelberg.

[9] Fokianos, K., and Pitsillou, M. (2016). Consistent testing for pairwise dependence in time series. *Technometrics*, to appear.

[10] Gerstenberger, C., and Vogel, D. (2015). On the efficiency of Gini's mean difference. *Stat. Methods. Appt.*, **24**, 569–596.

[11] Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Stat.*, **19**, 293–325.

[12] Hoeffding, W. (1961). The strong law of large numbers for *U*-statistics. *Institute of Statistics, Mimeograph Series No. 302*. University of North Carolina, Chapel Hill, NC.

[13] Huo, X., and Székely, G. J. (2016). Fast computing for distance covariance. *Technometrics*, **58**, 435–447.

[14] Jentsch, C., Leucht, A., Meyer, M., and Beering, C. (2016). Empirical characteristic functions-based estimation and distance correlation for locally stationary processes. Preprint, Mannheim University.

[15] Jones, M. C., and Balakrishnan, N. (2002). How are moments and moments of spacings related to distribution functions? *J. Stat. Plan. Inference*, **103**, 377–390.

[16] Kochar, S. C. (1999) On stochastic orderings between distributions and their sample spacings. *Statist. Probab. Lett.*, **42**, 345–352.

[17] Kong, J., Klein, B. E. K., Klein, R., and Wahba, G. (2012). Using distance correlation and SS-ANOVA to access associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci.*, **109**, 20352–20357.

[18] Lomnicki, Z. A. (1952). The standard error of Gini's mean difference. *Ann. Math. Stat.*, **23**, 635-637.

[19] Lyons, R. (2013). Distance covariance in metric spaces. *Ann. Probobab*, **41**, 3284–3305.

[20] Martínez-Gómez, E., Richards, M. T., and Richards, D. St. P. (2014). Distance correlation methods for discovering associations in large astrophysical databases. *Astrophys. J.*, **781**, 39 (11 pp.).

[21] Richards, M. T., Richards, D. St. P., and Martínez-Gómez, E. (2014). Interpreting the distance correlation results for the COMBO-17 survey. *Ann. Appl. Stat.*, **784**, L34 (5 pp.).

[22] Rizzo, M. L., and Székely, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Stat.*, **4**, 1034–1055.

[23] Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Stat.*, **41**, 2263–2291.

[24] Shaked, M., and Shanthikumar, G. (2007) *Stochastic Orders.* Springer, New York.

[25] Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007). Measuring and testing independence by correlation of distances. *Ann. Stat.*, **35**, 2769–2794.

[26] Székely, G. J., and Rizzo, M. (2009). Brownian distance covariance. *Ann. Appl. Stat.*, **3**, 1236–1265.

[27] Székely, G. J., and Rizzo, M. (2013). The distance correlation *t*-test of independence in high dimension. *J. Multivariate Anal.*, **117**, 193–213.

[28] Székely, G. J., and Rizzo, M. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Stat.*, **42**, 2382–2412.

[29] Van der Vaart, A. W. (2000) *Asymptotic Statistics*, Volume 3. Cambridge University Press, New York.

[30] Yitzhaki, S. (2003). Gini's mean difference: A superior measure of variability for non-normal distributions. *Metron*, **61**, 285-316.

[31] Yitzhaki, S., and Schechtman, E. (2012). *The Gini Methodology: A Primer on a Statistical Methodology.* Springer, New York.

[32] Zenga, M., Polisicchio, M., and Greselin, F. (2004). The variance of Gini's mean difference and its estimators. *Statistica*, **64**, 455-475.

[33] Zhou, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.*, **33**, 438–457.