

THE DISTANCE STANDARD DEVIATION

BY DOMINIC EDELMANN,¹ DONALD RICHARDS² AND DANIEL VOGEL³

¹German Cancer Research Center, dominic.edelmann@dkfz-heidelberg.de

²Department of Statistics, Pennsylvania State University, richards@stat.psu.edu

³Institute for Complex Systems and Mathematical Biology, University of Aberdeen, daniel.vogel@tu-dortmund.de

The distance standard deviation, which arises in distance correlation analysis of multivariate data, is studied as a measure of spread. The asymptotic distribution of the empirical distance standard deviation is derived under the assumption of finite second moments. Applications are provided to hypothesis testing on a data set from materials science and to multivariate statistical quality control. The distance standard deviation is compared to classical scale measures for inference on the spread of heavy-tailed distributions. 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. New expressions for the distance standard deviation are obtained in terms of Gini's mean difference and the moments of spacings of order statistics. It is also shown that the distance standard deviation satisfies the axiomatic properties of a measure of spread.

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 by Székely, Rizzo and Bakirov [39] and Székely and Rizzo [36], and they applied distance correlation methods to testing independence and measuring association between collections of random vectors.

Since the appearance of the papers [36, 39], enormous interest in the theory and applications of distance correlation has arisen. We refer to the articles [31, 37, 38] on statistical inference, [14, 15, 22, 44] on time series, [8–10] on affinely invariant distance correlation and connections with singular integrals, [24] on metric spaces and [32] on machine learning. Distance correlation methods have also been applied to assessing associations between familial relationships, lifestyle factors, diseases and mortality [23], and to detecting associations in large astrophysical databases [25, 30].

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

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

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

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

where $s \in \mathbb{R}^p$, $t \in \mathbb{R}^q$, 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*

Received November 2018; revised November 2019.

MSC2020 subject classifications. Primary 60E15, 62H20; secondary 60E05, 60E10.

Key words and phrases. Asymptotic efficiency, characteristic function, dispersive ordering, distance correlation coefficient, distance variance, Gini's mean difference, measure of spread, order statistic, sample spacing, stochastic ordering, U-statistic.

covariance between X and Y is defined as the nonnegative square root of

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

the *distance variance* is defined as

$$(1.2) \quad \mathcal{V}^2(X) = \mathcal{V}^2(X, X) = \frac{1}{c_p^2} \int_{\mathbb{R}^{2p}} |f_X(s+t) - f_X(s) f_X(t)|^2 \frac{ds dt}{\|s\|^{p+1} \|t\|^{p+1}},$$

and the *distance standard deviation*, $\mathcal{V}(X)$, is defined as the nonnegative square root of $\mathcal{V}^2(X)$. (We note that this terminology differs from that of Székely et al. [36, 39], who refer to $\mathcal{V}(X)$ as the *distance variance*; we will refer to $\mathcal{V}(X)$ instead as the *distance standard deviation*, which is justified by the fact that $\mathcal{V}(X)$ satisfies an equivariance property that is given below in (1.4).) Also, the *distance correlation coefficient* is defined as

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

as long as $\mathcal{V}(X), \mathcal{V}(Y) \neq 0$, and zero otherwise. We remark that the weighted L^2 -norm in (1.1) was studied in the univariate setting by Feuerverger [13].

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 $X, Y \in \mathbb{R}$, we have $\mathcal{R}(X, Y) = 1$ if and only if Y is a linear function of X , almost surely. The empirical distance correlation possesses a remarkably simple expression ([39], Theorem 1), and efficient algorithms for computing it are now available [21].

We note that $\mathcal{R}(X, Y)$ is one of several coefficients characterizing independence that are applicable to hypothesis testing. Other concepts of dependence are, for example, the Hilbert–Schmidt Independence Criterion (HSIC) [19], ball covariance [27] and mutual information [3]. Each of these concepts satisfy numerous desirable properties, and the comparison of their properties and finite-sample performance is an active area of research [29, 32, 34].

An interesting property of the distance covariance is that its square is translation-invariant and scale-equivariant, which implies that the distance standard deviation satisfies

$$(1.4) \quad \mathcal{V}(a + bX) = |b|\mathcal{V}(X)$$

for all $a, b \in \mathbb{R}$ ([39], Theorem 4). Moreover, $\mathcal{V}(X)$ is defined for all random variables X with finite first moments, whereas the classical standard deviation requires the existence of finite second moments. These properties suggest that the distance standard deviation is a potentially interesting measure of scale for heavy-tailed distributions. As the term $\mathcal{V}(X)\mathcal{V}(Y)$ appears in the denominator of $\mathcal{R}(X, Y)$, a study of properties of the distance standard deviation may lead to a better understanding of the distance correlation.

In this paper, we study the distance standard deviation $\mathcal{V}(X)$ and provide applications to hypothesis testing and multivariate statistical quality control. We apply the distance standard deviation to a data set, originating from materials science, on a physical model for describing a crystal undergoing a structural phase transition when subjected to several cooling–heating cycles. In a different direction, we further show how $\mathcal{V}(X)$ can be applied in the statistical quality control of multivariate production processes.

We will also compare $\mathcal{V}(X)$ to other measures of spread. Indeed, suppose that $\mathbb{E}(\|X\|^2) < \infty$, and let X, X' , and X'' be independent and identically distributed (i.i.d.); then, by Székely et al. [39], Remark 3,

$$(1.5) \quad \mathcal{V}^2(X) = \mathbb{E}(\|X - X''\|^2) + (\mathbb{E}\|X - X'\|)^2 - 2\mathbb{E}(\|X - X'\| \cdot \|X - X''\|).$$

The second term on the right-hand side of (1.5) is reminiscent of Gini's mean difference [17, 43], which is defined for real-valued random variables Y as

$$(1.6) \quad \Delta(Y) = \mathbb{E}|Y - Y'|,$$

where Y and Y' are i.i.d. Furthermore, if $X \in \mathbb{R}$ then one-half the first summand in (1.5) equals $\sigma^2(X) = E(X^2) - E(X)^2$, the variance of X .

We provide a detailed comparison of $\mathcal{V}(X)$, $\Delta(X)$, and $\sigma(X)$. We demonstrate that when the distributions of interest are heavy-tailed, $\mathcal{V}(X)$ provides estimators of scale that are asymptotically more efficient than estimators based on $\Delta(X)$ or $\sigma(X)$. Moreover, several inequalities between $\mathcal{V}(X)$, $\Delta(X)$, and $\sigma(X)$ are derived.

We further show that the distance standard deviation is an axiomatic measure of spread in the sense of Bickel and Lehmann [5]. According to [5], 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 for all $0 < \alpha \leq \beta < 1$,

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha),$$

where F and G are the cumulative distribution functions of X and Y , respectively, and F^{-1} and G^{-1} are the corresponding right-continuous inverses.

The distance covariance obviously satisfies (C1) and (C2). We will show that $\mathcal{V}(X)$ also satisfies (C3), hence proving that $\mathcal{V}(X)$ is a measure of spread in the above sense. However, we will also establish some clear differences between $\mathcal{V}(X)$, on the one hand, and $\Delta(X)$ and $\sigma(X)$, on the other hand.

The paper is organized as follows. In Section 2, the asymptotic distribution of the empirical distance standard deviation under the existence of the second moment of X is derived. The asymptotic relative efficiency (ARE) of the empirical distance standard deviation with respect to competing estimators of spread is evaluated for various distributions. In Section 3, we apply the empirical distance standard deviation to perform two-sample hypothesis testing for a data set from materials science and we also show the applicability of the empirical distance standard deviation in multivariate statistical quality control. Further, we demonstrate the superior performance of tests based on the distance standard deviation when the underlying distributions are heavy-tailed. In Section 4, we derive inequalities between the summands in the distance variance representation (1.5). We will prove in the case of scalar random variables that $\mathcal{V}(X)$ is bounded above by $\Delta(X)$ and by $\sigma(X)$. In Section 5, we show that the representation (1.5) can be simplified further, revealing relationships between $\mathcal{V}(X)$ and the moments of spacings of order statistics. Using novel representations, we show that $\mathcal{V}(X)$ is a measure of spread in the sense of [5]; moreover, we identify crucial differences between $\mathcal{V}(X)$, $\Delta(X)$ and $\sigma(X)$. We conclude the paper in Section 6 with a discussion of the given results. All proofs are provided in the Supplementary Material [12].

2. The empirical distance standard deviation. In order to develop an empirical version of $\mathcal{V}^2(X)$, Székely et al. [36, 39] derived an alternative representation of $\mathcal{V}^2(X)$; they showed 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

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

where

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

and

$$(2.3) \quad T_3(X) = \mathbb{E}(\|X - X'\| \cdot \|X - X''\|).$$

For an i.i.d. sample $\mathbf{X} = (X_1, \dots, X_n)$ drawn from X , the empirical version of $\mathcal{V}^2(X)$ was given in [39] as

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

where

$$(2.5) \quad T_{1,n}(\mathbf{X}) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|X_i - X_j\|^2,$$

$$T_{2,n}(\mathbf{X}) = \left(\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \|X_i - X_j\| \right)^2$$

and

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

The version (2.4) is not unbiased; an unbiased estimator for $\mathcal{V}^2(X)$ was derived in [21], namely,

$$(2.7) \quad \widehat{\mathcal{V}}_n^2(\mathbf{X}) = \frac{n}{n-3} T_{1,n}(\mathbf{X}) + \frac{n^3}{(n-1)(n-2)(n-3)} T_{2,n}(\mathbf{X}) - \frac{2n^2}{(n-2)(n-3)} T_{3,n}(\mathbf{X}).$$

By [21], $\widehat{\mathcal{V}}_n^2(\mathbf{X})$ is a U-statistic of order four with kernel function

$$(2.8) \quad h(X_1, X_2, X_3, X_4) = \frac{1}{4} \sum_{\substack{1 \leq i, j \leq 4 \\ i \neq j}} \|X_i - X_j\|^2 - \frac{1}{4} \sum_{i=1}^4 \left(\sum_{\substack{j=1 \\ j \neq i}}^4 \|X_i - X_j\| \right)^2 + \frac{1}{24} \left(\sum_{\substack{1 \leq i, j \leq 4 \\ i \neq j}} \|X_i - X_j\| \right)^2.$$

In the sequel, we derive the asymptotic distribution of $\widehat{\mathcal{V}}_n^2(\mathbf{X})$ and $\mathcal{V}_n^2(\mathbf{X})$; further, we do so under conditions weaker than known previously. Hitherto, the asymptotic normality of $\widehat{\mathcal{V}}_n^2(\mathbf{X})$ was proved only under the assumption that the *fourth* moment of X is finite; see [20], Lemma 4.8 and Theorem 4.11. Here, we derive the asymptotic normality under the broader assumption that the second moment of X is finite. The following lemma provides an alternative representation for the kernel function $h(\cdot)$ in (2.8).

LEMMA 2.1. *The kernel function h in (2.8) can be written as*

$$h(X_1, X_2, X_3, X_4) = \frac{1}{12} \sum_{\substack{1 \leq i, j \leq 4 \\ i \neq j}} \|X_i - X_j\|^2 - \frac{1}{12} \sum_{\substack{1 \leq i, j, k \leq 4 \\ i, j, k \text{ all different}}} \|X_i - X_j\| \|X_i - X_k\| + \frac{1}{24} \sum_{\substack{1 \leq i, j, k, l \leq 4 \\ i, j, k, l \text{ all different}}} \|X_i - X_j\| \|X_k - X_l\|.$$

In applying Lemma 2.1 to establish the asymptotic normality of $\mathcal{V}_n^2(X)$ under the assumption of finite second moments of X , let

$$h_1(x) = \mathbb{E}[h(x, X_2, X_3, X_4)] - \mathcal{V}^2(X)$$

be the linear part in the Hoeffding decomposition ([41], Section 11.4) of the kernel h and let

$$(2.9) \quad \gamma = 16\mathbb{E}[h_1^2(X)].$$

We remark that an expansion of $h_1(X)$ in our setting is given in [20], equation (B.6).

Denote by X_{-k} the sample X with the k th observation deleted. Assuming that $\mathbb{E}[h^2(X_1, X_2, X_3, X_4)] < \infty$ (which is a consequence of $\mathbb{E}(\|X\|^2) < \infty$; cf. the proof of Theorem 2.2), it follows from Arvesen [2], Theorem 9, that the jackknife estimator

$$(2.10) \quad \hat{\gamma}(X) = (n - 1) \sum_{i=1}^n \left(\hat{\mathcal{V}}_{n-1}^2(X_{-i}) - \frac{1}{n} \sum_{j=1}^n \hat{\mathcal{V}}_{n-1}^2(X_{-j}) \right)^2$$

is a weakly consistent estimator of γ .

THEOREM 2.2. *Suppose that $\mathbb{E}(\|X\|^2) < \infty$. As $n \rightarrow \infty$,*

$$(2.11) \quad \sqrt{n}(\hat{\mathcal{V}}_n^2(X) - \mathcal{V}^2(X)) \xrightarrow{d} N(0, \gamma)$$

and

$$(2.12) \quad \frac{\sqrt{n}(\hat{\mathcal{V}}_n^2(X) - \mathcal{V}^2(X))}{\sqrt{\hat{\gamma}(X)}} \xrightarrow{d} N(0, 1),$$

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

The asymptotic distribution of $\mathcal{V}_n(X)$, the empirical distance standard deviation, now follows from Theorem 2.2 by the delta method. A weakly consistent estimator for the variance of the asymptotic distribution of $\sqrt{n}(\hat{\mathcal{V}}_n(X) - \mathcal{V}(X))$ is obtained analogously from Arvesen [2], Theorem 9, and is given by

$$(2.13) \quad \hat{\xi}(X) = (n - 1) \sum_{i=1}^n \left(\hat{\mathcal{V}}_{n-1}(X_{-i}) - \frac{1}{n} \sum_{j=1}^n \hat{\mathcal{V}}_{n-1}(X_{-j}) \right)^2.$$

COROLLARY 2.3. *Suppose that $\mathbb{E}(\|X\|^2) < \infty$ and $\mathcal{V}(X) > 0$. Then*

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

and

$$\frac{\sqrt{n}(\hat{\mathcal{V}}_n(X) - \mathcal{V}(X))}{\sqrt{\hat{\xi}(X)}} \xrightarrow{d} N(0, 1),$$

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

We now consider the problem of estimating scale in a location-scale family of the form $X \stackrel{d}{=} \mu + \lambda Z$, with $\mu \in \mathbb{R}$ and $\mathbb{E}|Z|^2 < \infty$, where $\stackrel{d}{=}$ denotes equality in distribution. In this location-scale setting, Corollary 2.3 enables the comparison of the efficiency of the distance standard deviation to other estimators of spread. For any \sqrt{n} -consistent and asymptotically normal estimator $s_n(X)$, we define the asymptotic variance $ASV(s_n(X); F)$ at the distribution

F to be the variance of the limiting distribution of $\sqrt{n}(s_n(\mathbf{X}) - s(X))$ as $n \rightarrow \infty$, where $s_n(\mathbf{X})$ is evaluated at an i.i.d. sequence drawn from $X \sim F$ and $s(X)$ denotes the corresponding population value of $s_n(\mathbf{X})$. While two scale estimators (i.e., estimators satisfying property (C2) in Section 1) $s_n^{(1)}(\mathbf{X})$ and $s_n^{(2)}(\mathbf{X})$ may converge to different population values $s_1(X)$ and $s_2(X)$, respectively, $s_n^{(2)}(\mathbf{X})$ can be made consistent for $s_1(X)$ within the considered location-scale family by multiplying it with the factor $s_1(Z)/s_2(Z) = s_1(X)/s_2(X)$. Thus we define the *asymptotic relative efficiency* for scale estimators ([4], equation (2.1)) $s_n^{(1)}(\mathbf{X})$ with respect to $s_n^{(2)}(\mathbf{X})$ at the population distribution F as

$$(2.14) \quad \text{ARE}(s_n^{(1)}(\mathbf{X}), s_n^{(2)}(\mathbf{X}); F) = \frac{\text{ASV}(s_n^{(2)}(\mathbf{X}); F)/(s_2(X))^2}{\text{ASV}(s_n^{(1)}(\mathbf{X}); F)/(s_1(X))^2}.$$

We consider as alternatives to $\mathcal{V}_n(\mathbf{X})$ the empirical standard deviation,

$$(2.15) \quad \hat{\sigma}_n(\mathbf{X}) = \left[\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \right]^{1/2},$$

where \bar{X}_n denotes the sample mean of \mathbf{X} , the empirical mean deviation

$$(2.16) \quad \hat{d}_n(\mathbf{X}) = \frac{1}{n} \sum_{i=1}^n |X_i - m_n(\mathbf{X})|,$$

where $m_n(\mathbf{X})$ denotes the sample median of \mathbf{X} , and Gini’s mean difference,

$$(2.17) \quad \hat{\Delta}_n(\mathbf{X}) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} |X_i - X_j|.$$

We remark that $(\frac{n-1}{n})^{1/2} \hat{\sigma}_n(\mathbf{X})$ is the maximum likelihood estimator of scale in the location-scale family generated by the normal distribution $N(0, 1)$. Also, $\hat{d}_n(\mathbf{X})$ is the analogous estimator of scale for the Laplace distribution $L(0, 1)$.

Let $\text{NM}(\lambda, \epsilon)$ denote the normal scale mixture distribution that is defined as

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

and is also known as the *contaminated normal distribution* [40].

In Table 1, we compare the asymptotic efficiencies of the distance standard deviation with the three alternative measures of spread at the Laplace distribution, normal distribution, the normal scale mixture distribution $\text{NM}(3, 0.1)$ and the t_ν -distributions with $\nu = 3$ and $\nu = 5$.

TABLE 1

The asymptotic relative efficiencies (2.14) with respect to the respective maximum likelihood estimators of the distance standard deviation \mathcal{V}_n , the standard deviation $\hat{\sigma}_n$, the mean deviation \hat{d}_n and Gini’s mean difference $\hat{\Delta}_n$ at the Laplace distribution, the normal distribution, the normal scale mixture distribution $\text{NM}(3, 0.1)$ and the t_ν -distributions with $\nu = 3$ and $\nu = 5$

Distribution, F	$\text{ARE}(\mathcal{V}_n; F)$	$\text{ARE}(\hat{\sigma}_n; F)$	$\text{ARE}(\hat{d}_n; F)$	$\text{ARE}(\hat{\Delta}_n; F)$
$L(0, 1)$	0.952	0.8	1	0.964
$N(0, 1)$	0.784	1	0.876	0.978
$\text{NM}(3, 0.1)$	0.887	0.398	0.757	0.641
t_3	0.965	0	0.681	0.524
t_5	0.992	0.4	0.941	0.859

TABLE 2
Simulated finite-sample values of the bias and the variance of the estimators $\mathcal{V}_n(\mathbf{X})$ and $\widehat{\mathcal{V}}_n(\mathbf{X})$ for $n = 5, 10, 50, 500$ compared to asymptotic values (last column); 10,000 replications

Distribution		Sample size				
		5	10	50	500	∞
$L(0, 1)$	$\sqrt{n}(\mathbb{E}(\mathcal{V}_n) - \mathcal{V})$	0.282	0.313	0.191	0.067	0
	$\sqrt{n}(\mathbb{E}(\widehat{\mathcal{V}}_n) - \mathcal{V})$	-0.270	-0.136	-0.057	-0.022	0
	$n \text{Var}(\mathcal{V}_n)$	0.953	0.834	0.668	0.605	0.613
	$n \text{Var}(\widehat{\mathcal{V}}_n)$	0.899	0.723	0.642	0.604	0.613
$N(0, 1)$	$\sqrt{n}(\mathbb{E}(\mathcal{V}_n) - \mathcal{V})$	0.067	0.085	0.049	0.022	0
	$\sqrt{n}(\mathbb{E}(\widehat{\mathcal{V}}_n) - \mathcal{V})$	-0.197	-0.082	-0.022	0.022	0
	$n \text{Var}(\mathcal{V}_n)$	0.299	0.280	0.256	0.253	0.256
	$n \text{Var}(\widehat{\mathcal{V}}_n)$	0.488	0.336	0.265	0.253	0.256
$NM(3, 0.1)$	$\sqrt{n}(\mathbb{E}(\mathcal{V}_n) - \mathcal{V})$	0.244	0.307	0.198	0.067	0
	$\sqrt{n}(\mathbb{E}(\widehat{\mathcal{V}}_n) - \mathcal{V})$	-0.246	-0.114	-0.049	-0.022	0
	$n \text{Var}(\mathcal{V}_n)$	0.905	0.807	0.514	0.440	0.426
	$n \text{Var}(\widehat{\mathcal{V}}_n)$	0.673	0.523	0.455	0.434	0.426
t_3	$\sqrt{n}(\mathbb{E}(\mathcal{V}_n) - \mathcal{V})$	0.409	0.471	0.368	0.157	0
	$\sqrt{n}(\mathbb{E}(\widehat{\mathcal{V}}_n) - \mathcal{V})$	-0.304	-0.158	-0.049	0.000	0
	$n \text{Var}(\mathcal{V}_n)$	4.231	2.027	1.123	0.746	0.680
	$n \text{Var}(\widehat{\mathcal{V}}_n)$	0.991	0.800	0.711	0.676	0.680
t_5	$\sqrt{n}(\mathbb{E}(\mathcal{V}_n) - \mathcal{V})$	0.212	0.234	0.148	0.045	0
	$\sqrt{n}(\mathbb{E}(\widehat{\mathcal{V}}_n) - \mathcal{V})$	-0.235	-0.114	-0.042	-0.022	0
	$n \text{Var}(\mathcal{V}_n)$	0.772	0.638	0.472	0.427	0.424
	$n \text{Var}(\widehat{\mathcal{V}}_n)$	0.708	0.517	0.435	0.418	0.424

The asymptotic relative efficiencies of these estimators with respect to the respective maximum likelihood estimator at each of the distributions are given in Table 1. Details on the calculations of the values in Table 1 are given in Appendix B in the Supplementary Material [12].

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 normal scale mixture, the t_3 - and the t_5 -distributions, the distance standard deviation outperforms its three competitors.

In Table 2, we complement our asymptotic analysis with finite-sample simulations. For sample sizes $n = 5, 10, 50, 500$ and the same population distributions as above, the simulated biases and variances (based on 10,000 replications) of the empirical versions of distance standard deviation $\mathcal{V}_n(\mathbf{X})$ and $\widehat{\mathcal{V}}_n(\mathbf{X})$ are given along with their respective asymptotic values. The corresponding values for the competing estimators $\widehat{\sigma}_n(\mathbf{X})$, $\widehat{d}_n(\mathbf{X})$ and $\widehat{\Delta}_n(\mathbf{X})$ are provided by Gerstenberger and Vogel [17], Tables 7, 8.

The values presented in Table 2 indicate that $\widehat{\mathcal{V}}_n(\mathbf{X})$ is preferable to $\mathcal{V}_n(\mathbf{X})$ as an estimator of $\mathcal{V}(X)$. We note that both estimators are biased and that $\widehat{\mathcal{V}}_n^2$ is a U-statistic whereas $\widehat{\mathcal{V}}_n$ is not; however, $\widehat{\mathcal{V}}_n$ shows considerably smaller bias than \mathcal{V}_n at the heavier-tailed distributions. In light of the efficiency comparison with the other standard scale estimators, it emerges that heavy-tailed distributions represent the most promising area for applications of the distance standard deviation.

Although the definition of the empirical distance standard deviation $\widehat{\mathcal{V}}_n$ does not make apparent its superior performance under heavy tails, an intuitive explanation for its superiority in that context is obtained in Section 5, where the scale estimators are expressed in terms of the spacings between data points; it is seen there that the distance standard deviation provides

comparably little weight to the extreme spacings at either end of the data range, and it is this property that leads to the superior performance of $\widehat{\mathcal{V}}_n$ with heavy-tailed data.

In concluding this section, we note that the main argument in the proof of Theorem 2.2, namely that $\mathbb{E}(\|X\|^2) < \infty$ implies $\mathbb{E}[h^2(X_1, X_2, X_3, X_4)] < \infty$, leads to a proof of the central limit theorem (CLT) for the *squared distance covariance* given in [20], Theorem 4.11, under weaker conditions than known previously. The intrinsic idea in our proof of this CLT is that the square of the corresponding U-statistic,

$$\begin{aligned}
 \widehat{\Omega}_n(\mathbf{X}, \mathbf{Y}) = & \frac{1}{n(n-3)} \left[\sum_{i,j=1}^n \|X_i - X_j\| \|Y_i - Y_j\| \right. \\
 (2.18) \quad & + \frac{1}{(n-1)(n-2)} \sum_{i,j=1}^n \|X_i - X_j\| \cdot \sum_{i,j=1}^n \|Y_i - Y_j\| \\
 & \left. - \frac{2}{(n-2)} \sum_{i,j,k=1}^n \|X_i - X_j\| \|Y_i - Y_k\| \right],
 \end{aligned}$$

is bounded above by $\widehat{\mathcal{V}}_n^2(\mathbf{X})\widehat{\mathcal{V}}_n^2(\mathbf{Y})$, where $\mathbf{Y} = (Y_1, \dots, Y_n)$ denotes an i.i.d. sample drawn from some random variable $Y \in \mathbb{R}^q$. The complete statement of this limit theorem is given in Section C of the Supplementary Material [12].

3. Applications of the distance variance. We consider two applications in detail: hypothesis testing and quality control charts. For other interesting applications of the distance standard deviation, we refer to Fiedler [14] who defined the distance variogram and gave a natural generalization of the usual variogram for α -stable distributions.

3.1. *Hypothesis testing.* For ease of exposition, we focus on two-sample hypothesis testing in the univariate case. One-sample tests and results for the multivariate setting can be derived analogously.

Let $\mathbf{X}_n = (X_1, \dots, X_n)$ and $\mathbf{Y}_m = (Y_1, \dots, Y_m)$ be two i.i.d., mutually independent random samples drawn from random variables $X, Y \in \mathbb{R}$ with finite second moments. We wish to test the null hypothesis $H_0 : \mathcal{V}(X) = \mathcal{V}(Y)$. For this purpose, we propose the test statistic

$$(3.1) \quad \widehat{T}_{\mathcal{V}} = \sqrt{\frac{nm}{n+m}} \frac{\widehat{\mathcal{V}}_n(\mathbf{X}_n) - \widehat{\mathcal{V}}_m(\mathbf{Y}_m)}{\sqrt{\widehat{\xi}_p(\mathbf{X}_n, \mathbf{Y}_m)}},$$

where $\widehat{\xi}_p(\mathbf{X}_n, \mathbf{Y}_m)$ is a pooled estimator of the form

$$\widehat{\xi}_p(\mathbf{X}_n, \mathbf{Y}_m) = \frac{n\widehat{\xi}(\mathbf{X}_n) + m\widehat{\xi}(\mathbf{Y}_m)}{n+m}$$

and $\widehat{\xi}(\cdot)$ is defined in equation (2.13). By Theorem 3.1, $\widehat{T}_{\mathcal{V}}$ and $|\widehat{T}_{\mathcal{V}}|$ can be directly applied to test H_0 against one-sided and two-sided alternatives, respectively. In stating this theorem, we denote by ξ_X and ξ_Y the asymptotic variances of the distributions of $\sqrt{n}(\widehat{\mathcal{V}}_n(\mathbf{X}_n) - \mathcal{V}(X))$ and $\sqrt{m}(\widehat{\mathcal{V}}_m(\mathbf{Y}_m) - \mathcal{V}(Y))$, respectively (see Corollary 2.3 for details).

THEOREM 3.1. *Let $\mathbb{E}|X|^2 < \infty$ and $\mathbb{E}|Y|^2 < \infty$. Then, for $n, m \rightarrow \infty$, such that $n/m \rightarrow r > 0$, it holds:*

- (i) *If $\mathcal{V}(X) = \mathcal{V}(Y)$, then $\widehat{T}_{\mathcal{V}} \xrightarrow{d} N(0, (\xi_X + r\xi_Y)/(r\xi_X + \xi_Y))$. In particular, if additionally $X + \mu \stackrel{d}{=} Y$ where $\mu \in \mathbb{R}$, or $n/m \rightarrow 1$, then $\widehat{T}_{\mathcal{V}} \xrightarrow{d} N(0, 1)$.*

- (ii) If $\mathcal{V}(X) < \mathcal{V}(Y)$, then $\widehat{T}_{\mathcal{V}} \xrightarrow{P} -\infty$.
- (iii) If $\mathcal{V}(X) > \mathcal{V}(Y)$, then $\widehat{T}_{\mathcal{V}} \xrightarrow{P} \infty$.

A weakly consistent estimator of the asymptotic variance in Theorem 3.1(i) is

$$[\widehat{\xi}(X_n) + (n/m)\widehat{\xi}(Y_m)] / [(n/m)\widehat{\xi}(X_n) + \widehat{\xi}(Y_m)],$$

and this estimator can be used to construct a Studentized statistic for testing $H_0 : \mathcal{V}(X) = \mathcal{V}(Y)$. For cases in which $n/m \rightarrow 1$ or if X and Y belong to a common location-scale family, the resulting asymptotic variance equals 1, identically, and hence no estimation is needed.

When the distributions of X and Y belong to the same location-scale family, that is, $\lambda X + \mu \stackrel{d}{=} Y$ with $\mu \in \mathbb{R}$, the null hypothesis can be expressed as $H_0 : \lambda = 1$. Within this location-scale setting, we can compare the distance standard deviation based two-sample test to analogously constructed tests based on the standard deviation and Gini’s mean difference. For scale measures s_1 and s_2 , we denote by $s_n^{(1)}(\cdot)$ and $s_n^{(2)}(\cdot)$ the respective empirical versions of these measures. Moreover, we assume for i.i.d. samples X_n drawn from random variables $X \sim F$ and for $j = 1, 2$ that:

(A1) If $y_i = bx_i + a$ for $a, b \in \mathbb{R}, i \in \{1, \dots, n\}, n \in \mathbb{N}$, then $s_n^{(j)}(y_n) = |b|s_n^{(j)}(x_n)$, where $x_n = (x_1, \dots, x_n)$ and $y_n = (y_1, \dots, y_n)$.

(A2) For $n \rightarrow \infty, \sqrt{n}(s_n^{(j)}(X_n) - s_j(X)) \xrightarrow{d} \mathcal{N}(0, \xi_j)$ with $\xi_j > 0$ and $\widehat{\xi}_j(X_n)$ is a consistent estimator for ξ_j .

(A3) The estimator of the asymptotic standard deviation $\sqrt{\xi_j}$ satisfies (A1) (with $s_n^{(j)}$ replaced by $\sqrt{\widehat{\xi}_j}$).

Test statistics analogous to (3.1) can then be constructed as

$$(3.2) \quad \widehat{T}_j(X_n, Y_m) = \sqrt{\frac{nm}{n+m}} \frac{s_n^{(j)}(X_n) - s_n^{(j)}(Y_m)}{\sqrt{\widehat{\xi}_{p,j}(X_n, Y_m)}}$$

where

$$\widehat{\xi}_{p,j}(X_n, Y_m) = \frac{n}{n+m} \widehat{\xi}_j(X_n) + \frac{m}{n+m} \widehat{\xi}_j(Y_m).$$

Theorem 3.2 provides a comparison of the efficiency of two scale tests of the form (3.2) under local alternatives. Let $\lambda_{n,m}$ be an array of real numbers satisfying

$$\sqrt{\frac{nm}{n+m}}(\lambda_{n,m} - 1) \rightarrow \Lambda$$

for some $\Lambda \in \mathbb{R}$ as $n, m \rightarrow \infty$ such that $n/m \rightarrow r > 0$. For $m, n \in \mathbb{N}$, denote by $X_n = (X_1, \dots, X_n), Z_m = (Z_1, \dots, Z_m)$ two mutually independent, i.i.d. samples drawn from random variables X and Z following the same distribution F . Moreover, for $k \leq m$, define

$$(3.3) \quad \mathbf{Y}_k^{(n,m)} = (\lambda_{n,m}Z_1 + \mu, \dots, \lambda_{n,m}Z_k + \mu),$$

with $\mu \in \mathbb{R}$. In the following theorem, $[t]$ will denote the integer part of $t \in \mathbb{R}$ and $\Phi^{-1}(\cdot)$ is the inverse of Φ , the standard normal distribution function.

THEOREM 3.2. *Let $\rho = \xi_2 s_1^2(X) / (\xi_1 s_2^2(X))$ denote the asymptotic relative efficiency of $s_n^{(1)}$ with respect to $s_n^{(2)}$ at F (cf. (2.14)), where we assume without loss of generality that $\rho \leq 1$. Then, under Assumptions (A1), (A2) and (A3), $\widehat{T}_1(X_n, \mathbf{Y}_m^{(n,m)})$ and $\widehat{T}_2(X_{[\rho n]}, \mathbf{Y}_{[\rho m]}^{(n,m)})$ both converge in distribution to $N(-s_1(X)\Lambda / \sqrt{\xi_1}, 1)$ as $n, m \rightarrow \infty$, such that $n/m \rightarrow r > 0$.*

Consequently, for the ratio of the power of two-sided tests with asymptotic size α ,

$$\lim_{\substack{n,m \rightarrow \infty \\ n/m \rightarrow r}} \frac{P(|\widehat{T}_1(\mathbf{X}_n, \mathbf{Y}_m^{(n,m)})| > \Phi^{-1}(1 - \alpha/2))}{P(|\widehat{T}_2(\mathbf{X}_{\lfloor \rho n \rfloor}, \mathbf{Y}_{\lfloor \rho m \rfloor}^{(n,m)})| > \Phi^{-1}(1 - \alpha/2))} = 1.$$

Two benefits of Theorem 3.2 are that it enables explicit calculation of the asymptotic power for alternatives of the form (3.3), and it establishes a direct link between the asymptotic relative efficiencies of the scale estimators studied in Section 2 and the efficiencies of corresponding two-sample tests.

Let \widehat{T}_σ and \widehat{T}_Δ denote test statistics according to (3.2) based on $\widehat{\sigma}_n$ and $\widehat{\Delta}_n$, respectively, where the asymptotic variance of each estimator is estimated by the jackknife method; cf. (2.13).

Tables 3 and 4 contain rejection frequencies (based on 10,000 replications) at the 5% level for two-sided asymptotic tests based on $|\widehat{T}_\nu|$, $|\widehat{T}_\sigma|$ and $|\widehat{T}_\Delta|$. The F -test is also included for the sake of completeness and to serve as a benchmark in the normal case. The sensitivity of the F -test with respect to the assumption of normality is well known and is confirmed by the tables. We consider the Laplace distribution, normal distribution, normal scale mixture distribution $NM(3, 0.1)$, and the t_ν -distributions with $\nu = 3$ and $\nu = 5$. The sample sizes n, m range from $n + m = 30$ to $n + m = 2000$. Table 3 (test size) contains results for the null hypothesis $\lambda = 1$ and Table 4 (test power) gives results for the sample-size-dependent

TABLE 3

Test size. Empirical rejection frequencies (%) under the null hypothesis $\lambda = 1$ of asymptotic two-sample scale tests (based on the distance standard deviation \widehat{V}_n , the standard deviation $\widehat{\sigma}_n$, Gini's mean difference $\widehat{\Delta}_n$, and the F -test) at the 5% significance level. Results are based on 10,000 replications

Distribution	Test	$n:$							
		$m:$	15	50	120	250	600	1000	∞
			15	50	120	250	600	1000	∞
			15	50	40	250	200	1000	∞
			Rejection frequencies (%)						
$L(0, 1)$	\widehat{V}_n	4.4	4.6	5.0	4.8	5.2	4.7	5.0	
	$\widehat{\sigma}_n$	3.3	4.1	4.4	4.6	4.9	5.0	5.0	
	$\widehat{\Delta}_n$	5.9	5.1	4.9	4.8	5.3	4.9	5.0	
	F -test	17.5	20.0	19.5	21.6	20.8	20.6		
$N(0, 1)$	\widehat{V}_n	4.5	5.1	5.1	5.3	5.1	5.2	5.0	
	$\widehat{\sigma}_n$	4.4	5.0	5.0	5.4	5.1	5.1	5.0	
	$\widehat{\Delta}_n$	5.8	5.3	5.3	5.4	5.3	5.0	5.0	
	F -test	5.4	5.3	4.9	5.3	5.2	5.1	5.0	
$NM(3, 0.1)$	\widehat{V}_n	4.0	4.2	4.3	4.9	4.8	5.0	5.0	
	$\widehat{\sigma}_n$	2.6	3.3	4.3	4.4	4.8	4.9	5.0	
	$\widehat{\Delta}_n$	4.7	5.2	4.8	4.8	5.0	5.0	5.0	
	F -test	21.0	27.3	27.9	30.6	31.3	31.1		
t_3	\widehat{V}_n	4.3	4.5	4.9	4.6	4.8	5.2	5.0	
	$\widehat{\sigma}_n$	3.0	2.7	4.1	3.3	4.3	3.5		
	$\widehat{\Delta}_n$	5.4	4.6	5.0	4.7	4.7	4.7	5.0	
	F -test	25.8	35.8	38.0	49.9	51.2	59.3		
t_5	\widehat{V}_n	4.4	4.8	4.7	4.7	4.8	5.2	5.0	
	$\widehat{\sigma}_n$	3.6	4.0	4.4	4.4	4.9	4.4	5.0	
	$\widehat{\Delta}_n$	5.6	5.2	5.0	5.0	5.0	5.1	5.0	
	F -test	14.5	19.4	18.2	24.2	24.5	26.7		

TABLE 4

Test power. Empirical rejection frequencies (%) under the alternative $\lambda_{n,m} = 1 + 3\sqrt{(n+m)/n/m}$ of asymptotic two-sample scale tests (tests based on the distance standard deviation \widehat{V}_n , the standard deviation $\widehat{\sigma}_n$, Gini's mean difference $\widehat{\Delta}_n$ and the F -test) at the 5% significance level. Results are based on 10,000 replications

Distribution	Test	$n:$							
		$m:$	15	50	120	250	600	1000	∞
			15	50	40	250	200	1000	∞
		Rejection frequencies (%)							
$L(0, 1)$	\widehat{V}_n		33.0	57.1	63.8	72.6	75.0	78.8	83.3
	$\widehat{\sigma}_n$		27.1	51.5	66.1	65.8	71.5	71.7	76.5
	$\widehat{\Delta}_n$		44.6	62.2	68.7	73.6	76.6	79.3	83.8
$N(0, 1)$	\widehat{V}_n		48.0	78.0	83.2	90.9	91.8	94.0	96.4
	$\widehat{\sigma}_n$		56.9	87.3	92.5	96.0	97.0	97.9	98.9
	$\widehat{\Delta}_n$		68.7	88.1	91.4	95.8	96.3	97.6	98.7
	F -test		76.1	90.3	92.0	96.2	96.6	97.9	
$NM(3, 0.1)$	\widehat{V}_n		41.5	68.2	73.1	82.7	83.9	87.2	91.1
	$\widehat{\sigma}_n$		28.2	40.2	55.0	51.0	56.2	54.7	60.0
	$\widehat{\Delta}_n$		46.3	60.2	66.0	70.8	73.0	74.8	80.3
t_3	\widehat{V}_n		36.3	58.3	64.4	74.2	75.2	79.0	83.8
	$\widehat{\sigma}_n$		23.1	31.9	45.1	31.3	35.8	24.8	
	$\widehat{\Delta}_n$		39.9	49.9	55.6	56.4	58.1	58.1	58.4
t_5	\widehat{V}_n		41.6	69.0	73.9	82.8	85.2	88.4	91.6
	$\widehat{\sigma}_n$		36.4	56.5	68.7	64.1	68.2	64.7	56.4
	$\widehat{\Delta}_n$		53.6	69.2	74.1	78.6	81.4	83.2	87.5

alternative with

$$\lambda_{n,m} = 1 + 3\sqrt{\frac{n+m}{nm}}.$$

Theorem 3.2 yields large-sample approximations for the power of the tests, which are provided in the last column of Table 3 and Table 4. The asymptotic power for the distance standard deviation test is $P(|N(3V/\sqrt{\xi}, 1)| > \Phi^{-1}(0.975))$, and similar expressions hold for $\widehat{\sigma}_n$ and $\widehat{\Delta}_n$. Note that $\widehat{\sigma}_n$ does not satisfy the conditions of Theorem 3.2 at the t_3 -distribution.

In Table 3, we observe that the tests $|\widehat{T}_V|$, $|\widehat{T}_\sigma|$, and $|\widehat{T}_\Delta|$ control the nominal level of 5% well for all distributions under consideration. The actual rejection frequencies for the distance standard deviation test $|\widehat{T}_V|$ range between 4.0 and 5.3. The F -test grossly exceeds the nominal level for nonnormal distributions and is therefore omitted from the power considerations in Table 4 except for the normal case.

In Table 4, we see that except for the small-sample case $(n, m) = (15, 15)$, the distance standard deviation test $|\widehat{T}_V|$ performs best at the heavier-tailed distributions $NM(3, 0.1)$, t_3 and t_5 . At the Laplace $L(0, 1)$, $|\widehat{T}_\Delta|$ performs best and outperforms $|\widehat{T}_V|$ and $|\widehat{T}_\sigma|$ for small sizes; for large sample sizes $|\widehat{T}_V|$ and $|\widehat{T}_\Delta|$ perform almost equally. At the normal distribution $N(0, 1)$, $|\widehat{T}_\sigma|$ and $|\widehat{T}_\Delta|$ dominate $|\widehat{T}_V|$.

For small sample sizes, a better performance of the considered two-sample scale tests may be achieved by using a permutation-based approach for obtaining critical values, which we investigate in Section D in the Supplementary Material [12]. We note that the permutation test requires both distributions to share a common location, which is a more restrictive assumption than is needed for the asymptotic test.

Finally, we remark that, in the univariate case, the distance variance and hence the distance standard deviation can be computed rapidly. For the asymptotic derivations in Section 2, we

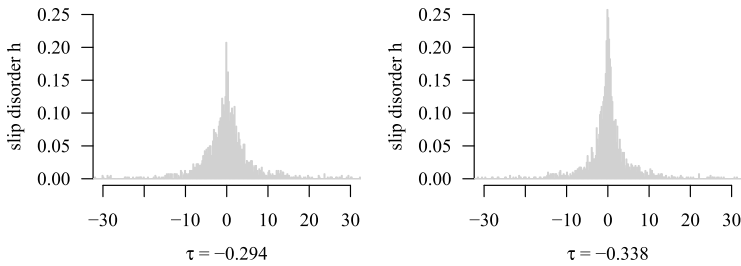


FIG. 1. Histograms of slip disorder for two values of the thermal driving parameter τ ; $n = 2000$ observations each.

used a fourth-order U-statistic representation of \widehat{V}_n^2 , which may suggest the opposite; however, Huo and Székely [21] devised an $O(n \log n)$ algorithm, which shows that the distance standard deviation has the same computational complexity as Gini's mean difference. All calculations for this article have been carried out using computationally efficient implementations of the distance standard deviation from the R package `dcortools` [11], available on <https://github.com/edelman21/dcortools>. An alternative $O(n \log n)$ implementation for the distance standard deviation is provided in the R package `energy` [35], available on the Comprehensive R Archive Network (CRAN).

Data example. We demonstrate the use of the two-sample distance variance test with an application to a data example. The data set stems from a physical model, studied by Perez-Reche et al. [28], for describing a crystal undergoing a structural phase transition between austenite and martensite phases when subject to several cooling-heating cycles. In this model, a quantity called the *slip disorder*, and denoted by h , is of particular interest. The slip disorder depends on a parameter τ that represents thermal driving and is referred to as the *temperature* within the model. The distribution of h for two values of τ is depicted in Figure 1.

An important issue is whether, and how, the distribution of the slip disorder h is affected by τ . As a consequence of the theoretical properties of the model, the distribution of h is symmetric around zero; this symmetry is also suggested by Figure 1. Hence, the main feature by which h may differ with respect to τ is in its scale. The distribution of observed values of h is very heavy-tailed and clearly nonnormal, with excess kurtoses exceeding 10 (cf. [28], Figures 10 and 11). Consequently, a F -test is inappropriate here.

Although arbitrarily large data sets can be obtained by letting the model run sufficiently long, the simulations are computationally costly. Hence, fast detection of a statistically significant difference is desirable. For the data depicted in Figure 1 (with parameter values $\tau = -0.294$ and $\tau = -0.338$, and sample sizes 2000 each), the two-sided asymptotic test based on the statistic $|\widehat{T}_V|$ yields a p -value of 0.0013. Corresponding tests based on the standard deviation and Gini's mean difference give p -values of 0.6940 and 0.0327, respectively. The large p -value of the test based on the classical standard deviation is consistent with the tendency of the test to underreject the null hypothesis in the presence of heavy tails; see Table 4. Moreover, as the standard deviation test remains persistently nonsignificant even for samples of size 10,000, we find again that the classical standard deviation is an inappropriate measure of spread for heavy-tailed distributions.

3.2. Multivariate statistical quality control. In statistical quality control [7, 26], the objective is to monitor quality characteristics in production processes using statistical methods. Important tools for process monitoring are the Shewhart control charts that are used to survey whether quality characteristics of the production process are under control.

A common tool used to monitor the dispersion of multivariate processes is the $|S|$ -chart [1], which applies the *generalized variance*, that is, the determinant of the covariance matrix of the process. In the following, we investigate the potential of a control chart based on \widehat{V}_n as an alternative tool for monitoring the dispersion of multivariate processes. The \mathcal{V} -chart, as we will call the corresponding chart, can also be applied in high-dimensional settings, where the dimension of the production process exceeds the number of samples per subgroup (as for example in molecular data).

To compare the performance of the $|S|$ -chart with the \mathcal{V} -chart, we conduct a simulation study. For simplicity, we assume throughout the simulation study that the process under consideration is bivariate and that the two components of the production process are independent. The covariance matrix of the process in control will always be given by

$$\Sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

For the covariance matrix of the process out of control, we will consider the matrices

$$\Sigma_1 = \begin{pmatrix} \delta^2 & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma_2 = \begin{pmatrix} \delta & 0 \\ 0 & \delta \end{pmatrix},$$

where $\delta = 1, 1.5, 2, 2.5, 3, 3.5, 4$ and the distribution of the components will follow either a normal, Laplace, t_3 - or t_5 -distribution. A Shewhart control chart consists of an upper control limit (UCL) and a lower control limit (LCL) for the characteristic under consideration. This characteristic (e.g., the generalized variance) is then computed for consecutive subgroups of a fixed size k . When the characteristic lies below the LCL or exceeds the UCL, this represents an out-of-control signal of the process; the corresponding subgroup can then be investigated and, if necessary, further action can be taken. Formally, this corresponds to consecutive testing of the null hypothesis that the characteristic in a subgroup under consideration equals the characteristic in-control.

To compare the $|S|$ - and \mathcal{V} -charts, we pursue a bootstrap approach. First, we simulate 10,000 i.i.d. samples of the process in control, that is, using the covariance matrix Σ_0 . In application, this is typically given by a phase-I sample of the process, for which it is known that the production process was in control. From these 10,000 samples, we now take $B = 100,000$ bootstrap samples of size $k = 25$, where k coincides with the fixed subgroup size. For each of the bootstrap samples, the generalized variance and the distance standard deviation are evaluated. The UCL for the $|S|$ -chart is then given by the 99.75%-quantile of the generalized variances of the B bootstrap samples, the corresponding LCL is given by the 0.25%-quantile. The respective UCL and LCL for the \mathcal{V} -chart are calculated analogously. Using the respective alternative distribution (i.e., using Σ_1 or Σ_2), we now generate 250,000 i.i.d. samples which are partitioned into 10,000 subgroups of size 25. For each subgroup, we evaluate if the generalized variance (or distance standard deviation respectively) exceeds the bounds given by the UCL or LCL of the respective chart. The empirical power for the $|S|$ -chart and the \mathcal{V} -chart are then calculated by the fraction of subgroups for which these bounds were exceeded. The procedure is replicated $N = 100$ times and the empirical power is averaged over these runs. The two methods are then compared using the average run length (ARL), which is the average number of subgroups one needs to test until an out-of-control signal is obtained. In the case of i.i.d. samples, the ARL is given by the reciprocal value of the empirical power.

Table 5 lists the ARLs of the $|S|$ -chart and the \mathcal{V} -chart for each of the covariance matrices Σ_1 and Σ_2 , where $\delta = 1, 1.5, 2, 2.5, 3, 3.5, 4$. The \mathcal{V} -chart shows substantial advantages compared to the $|S|$ -chart for heavy-tailed distributions, such as the t_3 - and t_5 -distributions. Moreover, while the $|S|$ -chart shows comparable performances for the different dispersion settings given by Σ_1 and Σ_2 (which is not surprising since they feature the same generalized

TABLE 5
Average run lengths for the \mathcal{V} -chart and $|S|$ -chart, respectively, for several alternative distributions representing the out-of-control state

Distribution	Chart	Effect size, δ						
		1	1.5	2	2.5	3	3.5	4
normal (Σ_1)	\mathcal{V}	198.77	2.72	1.19	1.01	1.00	1.00	1.00
	$ S $	198.77	3.77	1.30	1.05	1.01	1.00	1.00
normal (Σ_2)	\mathcal{V}	202.92	6.11	1.70	1.15	1.03	1.01	1.00
	$ S $	200.48	3.79	1.31	1.05	1.01	1.00	1.00
Laplace (Σ_1)	\mathcal{V}	196.23	5.85	1.53	1.10	1.02	1.00	1.00
	$ S $	190.30	13.16	3.18	1.69	1.27	1.12	1.05
Laplace (Σ_2)	\mathcal{V}	199.80	14.07	3.38	1.75	1.30	1.13	1.06
	$ S $	193.27	13.21	3.24	1.69	1.27	1.11	1.05
t_5 (Σ_1)	\mathcal{V}	205.38	4.70	1.32	1.03	1.00	1.00	1.00
	$ S $	184.20	26.99	5.99	2.55	1.59	1.27	1.15
t_5 (Σ_2)	\mathcal{V}	199.23	11.68	2.69	1.47	1.16	1.05	1.02
	$ S $	185.77	29.63	6.19	2.62	1.65	1.28	1.14
t_3 (Σ_1)	\mathcal{V}	199.80	12.78	2.20	1.22	1.04	1.01	1.00
	$ S $	169.00	94.04	40.40	18.41	9.17	5.87	4.51
t_3 (Σ_2)	\mathcal{V}	202.92	34.12	7.01	2.99	1.80	1.37	1.17
	$ S $	176.37	84.69	33.07	18.00	10.45	6.14	4.41

variance), the distance standard deviation seems to be more powerful against large changes in one component compared to moderate changes in both components. For a definitive statement about the potential of the distance standard deviation for multivariate statistical quality control, more detailed comparisons with the generalized variance are required that go beyond the scope of this paper. Yet, our results indicate that the \mathcal{V} -chart is a promising alternative to the generalized variance for multivariate statistical quality control in the presence of heavy-tailed distributions.

4. Inequalities between the distance variance, the variance and Gini’s mean difference. In the following, we will study inequalities between the summands appearing in (2.1) and (2.4). In the one-dimensional case, these inequalities will lead to crucial results concerning the relationships between $\mathcal{V}(X)$, $\Delta(X)$ and $\sigma(X)$.

LEMMA 4.1. *Let $T_{1,n}(X)$, $T_{2,n}(X)$, $T_{3,n}(X)$ be defined as in (2.5) and (2.6). Then there hold the algebraic inequalities*

$$(4.1) \quad T_{2,n}(X) \leq T_{3,n}(X) \leq T_{1,n}(X), \quad T_{1,n}(X) \leq 2T_{3,n}(X).$$

Further, if $X \in \mathbb{R}^p$ is a random vector such that $\mathbb{E}\|X\|^2 < \infty$, and if $T_1(X)$, $T_2(X)$, $T_3(X)$ are defined as in (2.2) and (2.3), then

$$(4.2) \quad T_2(X) \leq T_3(X) \leq T_1(X), \quad T_1(X) \leq 2T_3(X).$$

Using the inequalities in Lemma 4.1, we can derive upper bounds for the distance variance.

THEOREM 4.2. *Let $X \in \mathbb{R}^p$ be a random vector with $\mathbb{E}\|X\| < \infty$ and let $\mathbf{X} = (X_1, \dots, X_n)$ denote an i.i.d. sample drawn from X . Then*

$$\mathcal{V}_n^2(\mathbf{X}) \leq \frac{1}{n^4} \left(\sum_{i=1}^n \sum_{j=1}^n \|X_i - X_j\| \right)^2.$$

Moreover, denoting by X' an independent copy of X , we obtain $\mathcal{V}^2(X) \leq (\mathbb{E}\|X - X'\|)^2$. Further, if $\mathbb{E}\|X\|^2 < \infty$ then $\mathcal{V}^2(X) \leq \text{trace}(\Sigma_X)$, where Σ_X is the covariance matrix of X .

In the one-dimensional case, Theorem 4.2 implies that the distance variance is bounded above by the variance and the squared Gini mean difference.

COROLLARY 4.3. *Let X be a scalar random variable with $\mathbb{E}(|X|) < \infty$. Then, $\mathcal{V}^2(X) \leq \Delta^2(X)$. Moreover, if $\mathbb{E}(|X|^2) < \infty$ then $\mathcal{V}^2(X) \leq \sigma^2(X)$.*

We also note that for continuous variables $X \in \mathbb{R}$, the inequality $T_2(X) \leq T_1(X)$ can be sharpened.

PROPOSITION 4.4. *Let X be a real-valued continuous random variable with $\mathbb{E}(|X|^2) < \infty$. Then $T_2(X) \leq \frac{2}{3}T_1(X)$.*

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

THEOREM 4.5. *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 nontrivial random variable for which $\mathcal{V}^2(X) = \Delta^2(X)$ then the distribution of X is concentrated on two points.

For the Bernoulli distribution with $p = \frac{1}{2}$, Theorem 4.5 implies immediately that $\mathcal{V}^2(X)$, $\sigma^2(X)$ and $\Delta^2(X)$ attain the same value, namely, $\frac{1}{4}$. Hence, applying Corollary 4.3 and the dilation property $\mathcal{V}(bX) = |b|\mathcal{V}(X)$ in (C2), we obtain

COROLLARY 4.6. *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 Gábor Székely (private communication, November 23, 2015).

We remark that the second implication of Theorem 4.2 and Theorem 4.5 also follow from a result for the generalized distance variance in [24], Proposition 2.3. However, our presentation provides a more direct approach to these findings.

Since distance standard deviation terms appear in the denominator of the distance correlation coefficient, the inequalities derived in this section lead to new properties for the distance correlation. As an example, we now state a result, on the behavior of the empirical distance correlation in high dimensions that can be derived using Theorem 4.2 (see Appendix A in

the Supplementary Material [12] for full details). In [37], Appendix A.1, it is shown under certain assumptions that $\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})$ converges to 1 almost surely when the dimensions of \mathbf{X} and \mathbf{Y} tend to infinity, while the sample size n is fixed. We now show that a similar property can be derived when only the dimension of \mathbf{X} tends to infinity.

Before stating the result, we note (see [39], p. 2776, equation (2.18)) that the squared standard empirical distance covariance $\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})$ is expressible as

$$\begin{aligned}
 \mathcal{V}_n^2(\mathbf{X}, \mathbf{Y}) &= \frac{1}{n^2} \sum_{i,j=1}^n \|X_i - X_j\| \|Y_i - Y_j\| \\
 (4.3) \quad &+ \frac{1}{n^4} \sum_{i,j=1}^n \|X_i - X_j\| \sum_{i,j=1}^n \|Y_i - Y_j\| \\
 &- \frac{2}{n^3} \sum_{i,j,k=1}^n \|X_i - X_j\| \|Y_i - Y_k\|.
 \end{aligned}$$

The squared standard empirical distance correlation (see [39], p. 2774, Definition 5) is defined as

$$(4.4) \quad \mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = \frac{\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})}{\mathcal{V}_n(\mathbf{X})\mathcal{V}_n(\mathbf{Y})},$$

if both $\mathcal{V}_n(\mathbf{X})$ and $\mathcal{V}_n(\mathbf{Y})$ are different from 0, and $\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = 0$ otherwise.

PROPOSITION 4.7. *For fixed q , let Y be a q -dimensional random vector. For each $p \in \mathbb{N}$, let $X = (X^{(1)}, \dots, X^{(p)})^t$ be a p -dimensional random vector with $\mathbb{E}\|X\|^2 < \infty$ and i.i.d. coordinates $X^{(1)}, \dots, X^{(p)}$. For fixed $n \in \mathbb{N}$, let $(\mathbf{X}, \mathbf{Y}) = ((X_1, Y_1), \dots, (X_n, Y_n))$ denote a sample of size n drawn from (X, Y) . Then, almost surely,*

$$(4.5) \quad \lim_{p \rightarrow \infty} \frac{\mathcal{V}_n^2(\mathbf{X}, \mathbf{Y})}{\mathbb{E}\|X - X'\|^2} = n^{-3} \sum_{i,j=1}^n \|Y_i - Y_j\|^2$$

and

$$(4.6) \quad \lim_{p \rightarrow \infty} \mathcal{R}_n^2(\mathbf{X}, \mathbf{Y}) = (n-1)^{-1/2} \frac{n^{-2} \sum_{i,j=1}^n \|Y_i - Y_j\|^2}{\mathcal{V}_n(\mathbf{Y}, \mathbf{Y})} \geq (n-1)^{-1/2}.$$

To demonstrate the relevance of Proposition 4.7, we generate i.i.d. samples $(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) = ((X_1^{(k)}, Y_1^{(k)}), \dots, (X_n^{(k)}, Y_n^{(k)}))$ of size $n = 50$ drawn from (X, Y) , where $X \in \mathbb{R}^p$ with $p = 100$, $Y \in \mathbb{R}$ and (X, Y) follows a $p + 1$ -dimensional standard normal distribution with identity covariance matrix. For the average standard distance correlation over $K = 10,000$ simulation runs, we then obtain $K^{-1} \sum_{k=1}^K \mathcal{R}_n(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) = 0.4823$. Considering that we simulated \mathbf{X} and \mathbf{Y} to be independent, this reveals a heavy bias of the standard distance correlation in this setting, showing that this coefficient is hard to interpret when p is high. Even more, we note that the limiting value of $\mathcal{R}_n^2(\mathbf{X}, \mathbf{Y})$ depends only on the distribution of Y and not on the dependence between X and Y . Hence, we can expect similar results for random variables following the same distribution as Y even when they are strongly associated with X . Indeed, let $\mathbf{Z}^{(k)} = (Z_1^{(k)}, \dots, Z_n^{(k)})$, where $Z_i^{(k)} = p^{-1/2} \mathbf{1}'_p X_i^{(k)}$ with $\mathbf{1}_p = (1, \dots, 1)' \in \mathbb{R}^p$. Obviously, $Z_i^{(k)}$ shows the same variance as $Y_i^{(k)}$, but now $Z_i^{(k)}$ and $X_i^{(k)}$ are collinear. Yet, we obtain $K^{-1} \sum_{k=1}^K \mathcal{R}_n(\mathbf{X}^{(k)}, \mathbf{Z}^{(k)}) = 0.5112$, showing only a slight difference to the result

in the independent case. For an interpretable version of distance correlation when p is high, we propose using

$$(4.7) \quad \widehat{\mathcal{R}}_n(\mathbf{X}, \mathbf{Y}) = \text{sign}\left(\frac{\widehat{\Omega}_n(\mathbf{X}, \mathbf{Y})}{\widehat{\mathcal{V}}_n(\mathbf{X})\widehat{\mathcal{V}}_n(\mathbf{Y})}\right) \sqrt{\left|\frac{\widehat{\Omega}_n(\mathbf{X}, \mathbf{Y})}{\widehat{\mathcal{V}}_n(\mathbf{X})\widehat{\mathcal{V}}_n(\mathbf{Y})}\right|},$$

where $\text{sign}(t)$ denotes the sign of $t \in \mathbb{R}$ and $\widehat{\Omega}_n(\mathbf{X}, \mathbf{Y})$ is defined in (2.18).

Since this version is based on the U-statistic estimates of the squared distance covariance and the distance variance, it may be conjectured that it will generally not show a strong bias. Notably, $K^{-1} \sum_{k=1}^K \widehat{\mathcal{R}}_n(\mathbf{X}^{(k)}, \mathbf{Y}^{(k)}) = -0.00511$ and $K^{-1} \sum_{k=1}^K \widehat{\mathcal{R}}_n(\mathbf{X}^{(k)}, \mathbf{Z}^{(k)}) = 0.2897$; the population versions can be explicitly calculated using Corollary 3.2 and Corollary 3.3 in [8] and are given by $\mathcal{R}(X, Y) = 0$ and $\mathcal{R}(X, Z) \approx 0.2987$.

Examples in which the dimension of X is 100 or larger and Y is univariate occur in the analysis of genetic data, where it is often the goal to assess the association of a large number of molecular markers with some univariate clinical response, such as the development of a certain disease or response to treatment. One common approach [18] for this kind of data is to test for the association of the response with interesting *sets of markers* which may, for example, be defined *via* gene pathways or gene ontology (GO) [16] terms. While hypothesis testing itself gives little information about the effect size, distance correlation offers a way to quantify the strength of association between sets and univariate responses. Proposition 4.7 and the above considerations yield that the bias-corrected estimate (4.7) is to be preferred over the standard estimator (4.4) in these situations.

5. Properties of the distance standard deviation in one dimension. The representation of \mathcal{V}^2 given in (2.1), although more applicable than the expression given in equation (1.2), is undefined for random vectors with infinite second moments. This problem can be circumvented by considering the representation

$$(5.1) \quad \mathcal{V}^2(X) = \Delta^2(X) + W(X),$$

where

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

Note that since $0 \leq \mathcal{V}^2(X) \leq \Delta^2(X)$ then $W(X) \leq 0$ and $|W(X)| \leq \Delta^2(X)$; since $\Delta(X)$ exists under the assumption of finite first-order moments, then so does $W(X)$.

In the one-dimensional case, (5.1) gives rise to other representations that lead to crucial results about the distance standard deviation.

THEOREM 5.1. *Let X be a real-valued random variable with $\mathbb{E}|X| < \infty$, and let X, X', X'' and X''' be i.i.d.*

(i) *Let $X_{1:4} \leq X_{2:4} \leq X_{3:4} \leq X_{4:4}$ be the order statistics of the quadruple (X, X', X'', X''') . Then*

$$(5.2) \quad \mathcal{V}^2(X) = \frac{2}{3} \mathbb{E}[(X_{3:4} - X_{2:4})^2].$$

(ii) *Let F be the cumulative distribution function of X . Then*

$$(5.3) \quad \mathcal{V}^2(X) = 8 \iint_{-\infty < x < y < \infty} F^2(x)(1 - F(y))^2 dx dy.$$

(iii) *Let $t_+ = \max(t, 0)$, $t \in \mathbb{R}$. Then*

$$(5.4) \quad \mathcal{V}^2(X) = \Delta^2(X) - 8\mathbb{E}[(X - X')_+(X'' - X)_+].$$

(iv) Let $X_{1:3} \leq X_{2:3} \leq X_{3:3}$ be the order statistics of the triple (X, X', X'') . Then

$$(5.5) \quad \mathcal{V}^2(X) = \Delta^2(X) - \frac{4}{3}\mathbb{E}[(X_{2:3} - X_{1:3})(X_{3:3} - X_{2:3})].$$

Important properties of \mathcal{V} following from equation (5.2) are discussed in Theorem 5.2, and motivation for the representations provided in (5.3)–(5.5) are given in the Supplementary Material [12].

THEOREM 5.2. *The functional \mathcal{V} is an axiomatic measure of spread, that is,*

- (C1) $\mathcal{V}(X) \geq 0$,
- (C2) $\mathcal{V}(a + bX) = |b|\mathcal{V}(X)$ for all $a, b \in \mathbb{R}$, and
- (C3) $\mathcal{V}(X) \leq \mathcal{V}(Y)$ if for all $0 < \alpha \leq \beta < 1$,

$$F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha),$$

where F and G are the cumulative distribution functions of X and Y , respectively, and F^{-1} and G^{-1} are the corresponding right-continuous inverses.

Applying [33], Theorem 3.B.7, we obtain the following corollary of Theorem 5.2.

COROLLARY 5.3. *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

$$(5.6) \quad \mathcal{V}(X + Y) \geq \max(\mathcal{V}(X), \mathcal{V}(Y)).$$

It is well known, both for the standard deviation and for Gini's mean difference, that assertions analogous to (5.6) hold without restrictions on the distributions of X and Y .

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

EXAMPLE 5.4. 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 $\mathcal{V}(X) > \mathcal{V}(X + Y)$.

Other common properties of the classical standard deviation and Gini's mean difference concern differences and sums of independent random variables. Notably, it is well known that $\Delta(X + Y) = \Delta(X - Y)$ and $\sigma(X + Y) = \sigma(X - Y)$ 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 5.5. Let X and Y be independently Bernoulli distributed with parameter $p \neq \frac{1}{2}$. Then $\mathcal{V}(X + Y) > \mathcal{V}(X - Y)$.

While $\Delta(X)$, $\sigma(X)$ and $\mathcal{V}(X)$ are all measures of spread in the sense of [5], Examples 5.4 and 5.5 and the comparison of the asymptotic relative efficiencies in Section 2 suggest that there are substantial differences between these coefficients as measures of spread. To provide further understanding of these differences, we now derive representations that enable graphical comparisons of these three measures.

For this purpose, we apply equation (5.2) to derive a new empirical version for distance variance which is distinct from $\mathcal{V}_n^2(X)$ and $\widehat{\mathcal{V}}_n^2(X)$, as follows. For an i.i.d. sample X_1, \dots, X_n of real-valued random variables, denote by $D_{i:n} = X_{(i+1):n} - X_{i:n}$, $i = 1, \dots, n - 1$ the i th spacing of $\mathbf{X} = (X_1, \dots, X_n)$.

PROPOSITION 5.6. *Let X be a real-valued random variable with $\mathbb{E}(|X|) < \infty$ and let $\mathbf{X} = (X_1, \dots, X_n)$ be an i.i.d. sample from X . Then a strongly consistent empirical version for $\mathcal{V}^2(X)$ is*

$$(5.7) \quad \mathcal{U}_n^2(\mathbf{X}) = \binom{n}{2}^{-2} \sum_{i,j=1}^{n-1} (\min(i, j))^2 (n - \max(i, j))^2 D_{i:n} D_{j:n}.$$

Let $D = (D_{1:n}, \dots, D_{(n-1):n})$ denote the vector of spacings, then we can write the quadratic form in (5.7) as $\mathcal{U}_n^2(\mathbf{X}) = D^t V D$, where the (i, j) th element of the matrix V is

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

Both the squared empirical Gini mean difference and the empirical variance (see equations (2.15) and (2.17)) can also be expressed as quadratic forms in the spacings vector D ; specifically, $\widehat{\Delta}_n^2(\mathbf{X}) = D^t G D$ and $\widehat{\sigma}_n^2(\mathbf{X}) = D^t S D$, where the elements of G and S are given by

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

and

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

Comparing \mathcal{U}_n^2 , $\widehat{\Delta}_n^2$, and $\widehat{\sigma}_n^2$ clearly 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 2. These plots provide a descriptive explanation as to why \mathcal{V} and Δ are more suitable for heavy-tailed distribution since they place smaller weight than σ on extreme spacings; in particular, $V_{n-1,n-1} = G_{n-1,n-1} = 4n^{-2}$ while $S_{n-1,n-1} = n^{-1}$.

On the other hand, the *shape* of the plot for \mathcal{V} resembles the plot for σ more than the plot for Δ . Both \mathcal{V} and σ place comparably high weights on the diagonals; specifically, the highest entries in row i are the diagonal entries $V_{i,i}$ and $S_{i,i}$, respectively, and the highest entry in row i for Gini’s mean difference is $G_{i, \lfloor n/2 \rfloor}$. The intrinsic reason for this finding is that \mathcal{V}^2 and σ^2 are sums of *quadratic* differences of the observations, while Δ is a sum of the untransformed differences.

Since \mathcal{V} and σ place comparably high weights on the diagonals, one may conjecture that these measures are sensitive to variability in the spacings. Indeed, Yitzthaki [42], p. 291, points out that “the more equal are the distances between adjacent observations, the lower

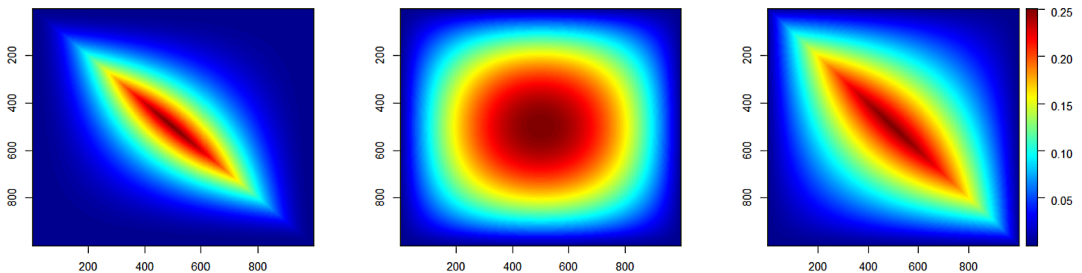


FIG. 2. *Illustration of, from left to right, the empirical distance variance \mathcal{U}_n^2 , the squared empirical Gini mean difference $\widehat{\Delta}_n^2$, and the empirical variance $\widehat{\sigma}_n^2$ via their respective quadratic form matrices V , G and S for sample size $n = 1000$. The 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.*

the variance.” This statement also holds for the distance standard deviations of light-tailed distributions.

For example, consider a distribution which is concentrated on three points, $\{0, \delta, 1\}$ and attains each value with probability $\frac{1}{3}$, similar to [42], p. 291. For this distribution, $\Delta = \frac{4}{9}$, irrespective of the value of δ . On the other hand, $\mathcal{V}^2 = 4(\frac{1}{3} + \delta^2 + (1 - \delta)^2)/27$ and $\sigma^2 = 2(1 + \delta^2 + (1 - \delta)^2)/9$. Hence, both measures are maximal for $\delta \in \{0, 1\}$ and minimal for $\delta = \frac{1}{2}$, that is, when the difference between the spacings is maximal and minimal, respectively.

The sensitivity of \mathcal{V} to differences between the spacings extends to other light-tailed distributions. For example, the most broadly spread distribution on $[0, 1]$ arguably is $B(1, \frac{1}{2})$, the Bernoulli distribution with $p = \frac{1}{2}$. The ratio of the population values between $B(1, \frac{1}{2})$ and the uniform distribution on $[0, 1]$ are $3/2 = 1.5$ (Δ), $\sqrt{3} \approx 1.73$ (σ) and $\sqrt{45/8} \approx 2.37$ (\mathcal{V}); see Appendix F for the value of \mathcal{V} for the uniform distribution.

In the case of σ , Yitzthaki [42], p. 291, states that the sensitivity to differences between the spacings “is translated to sensitivity to extreme observations.” However, this is not the case for \mathcal{V} as it places lower weights than Δ on the outer spacings, resulting in less sensitive behavior to extreme observations. For a demonstrative example, consider the t_ν -distributions. Specifically, the ratio of the respective population values between the t_3 - and t_5 -distribution is $(54\sqrt{3})/(35\sqrt{5}) \approx 1.20$ for Δ and $\sqrt{9/5} \approx 1.34$ for σ ; for \mathcal{V} , numerical evaluation yields that the ratio is approximately 1.12.

To summarize, the distance standard deviation is very sensitive to variability in the central spacings of an observation but relatively insensitive to changes in the extreme spacings. Consequently, the behavior of \mathcal{V} resembles the behavior of σ for light-tailed distributions where the central spacings are relatively large compared to the extreme spacings. On the other hand, the low sensitivity of \mathcal{V} to extreme observations makes it a very good measure of spread for heavy-tailed distributions.

6. Discussion. In this work, we have studied the statistical properties of the distance standard deviation, which arises as a special case of the distance covariance introduced by Székely et al. [39]. Notably, we have demonstrated that the empirical distance standard deviation has appealing statistical properties: it is less vulnerable to outliers and generally more appropriate for heavy-tailed distributions, more so than classical alternatives such as the mean deviation and Gini’s mean difference.

The distance standard deviation, and the distance correlation coefficient, may be seen as natural statistics for analyzing multivariate distributions, providing an alternative to the classical second-moment statistics and also being potentially more appropriate in light of their statistical properties.

For multivariate random variables, the distance standard deviation summarizes the spread as a single value. While this can be useful for many applications (see, e.g., the example on multivariate statistical quality control in Section 3), a referee has noted that the covariance matrix provides richer information, such as the spread of single components and the association between different components. To obtain an analogue of the covariance matrix based on the concept of distance covariance, one can define the *distance covariance matrix*,

$$\Gamma_X = (\mathcal{V}^2(X_i, X_j))_{i,j=1,\dots,p},$$

where X_1, \dots, X_p are the components of X ; a related concept is that of distance multivariate [6], which allows for testing the mutual independence of more than two sets of random vectors. It can be shown that Γ_X is positive semidefinite. Investigating properties of the distance covariance matrix is a promising direction for further research.

Acknowledgments. We are grateful to the reviewers, the associate editor and the editor for constructive and insightful comments that led to numerous revisions in the manuscript. Dominic Edelmann is funded by the German Research Foundation (Project No. 417754611).

SUPPLEMENTARY MATERIAL

Supplement to “The distance standard deviation” (DOI: [10.1214/19-AOS1935SUPP](https://doi.org/10.1214/19-AOS1935SUPP); .pdf). The supplementary material consists of six appendices. Appendix A contains various proofs to results in the main paper. Appendix B contains details on the derivation of Table 1 in the main document. In Appendix C, a limit theorem for the squared distance covariance is stated, under weaker assumptions than known previously. Simulation results for permutation-based two-sample scale tests are provided in Appendix D. Appendix E gives additional theoretical results for the distance standard deviation in one dimension. Appendix F tabulates the distance variances for a collection of distributions.

REFERENCES

- [1] APARISI, F., JABALOYES, J. and CARRIÓN, A. (1999). Statistical properties of the $|S|$ multivariate control chart. *Comm. Statist. Theory Methods* **28** 2671–2686. MR1729896 <https://doi.org/10.1080/03610929908832445>
- [2] ARVESEN, J. N. (1969). Jackknifing U -statistics. *Ann. Math. Stat.* **40** 2076–2100. MR0264805 <https://doi.org/10.1214/aoms/1177697287>
- [3] BERRETT, T. B. and SAMWORTH, R. J. (2019). Nonparametric independence testing via mutual information. *Biometrika* **106** 547–566. MR3992389 <https://doi.org/10.1093/biomet/asz024>
- [4] BICKEL, P. J. and LEHMANN, E. L. (2012). Descriptive statistics for nonparametric models, III. Dispersion. In *Selected Works of E. L. Lehmann* (J. Rojo, ed.) 499–518. *Selected Works in Probability and Statistics*. Springer, New York. MR2920510
- [5] BICKEL, P. J. and LEHMANN, E. L. (2012). Descriptive statistics for nonparametric models, IV. Spread. In *Selected Works of E. L. Lehmann*. *Selected Works in Probability and Statistics* (J. Rojo, ed.) 519–526. Springer, New York. MR2920510
- [6] BÖTTCHER, B., KELLER-RESSEL, M. and SCHILLING, R. L. (2019). Distance multivariate: New dependence measures for random vectors. *Ann. Statist.* **47** 2757–2789. MR3988772 <https://doi.org/10.1214/18-AOS1764>
- [7] CHANDRA, M. J. (2001). *Statistical Quality Control*. CRC Press, Melbourne, FL.
- [8] DUECK, J., EDELMANN, D., GNETING, T. and RICHARDS, D. (2014). The affinely invariant distance correlation. *Bernoulli* **20** 2305–2330. MR3263106 <https://doi.org/10.3150/13-BEJ558>
- [9] 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. MR3299759 <https://doi.org/10.1016/j.spl.2014.11.012>
- [10] DUECK, J., EDELMANN, D. and RICHARDS, D. (2017). Distance correlation coefficients for Lancaster distributions. *J. Multivariate Anal.* **154** 19–39. MR3588555 <https://doi.org/10.1016/j.jmva.2016.10.012>
- [11] EDELMANN, D. and FIEDLER, J. (2019). dcorntools: Providing fast and flexible functions for distance correlation analysis. R package version 0.0.0.9000.
- [12] EDELMANN, D., RICHARDS, D. and VOGEL, D. (2020). Supplement to “The distance standard deviation.” <https://doi.org/10.1214/19-AOS1935SUPP>.
- [13] FEUERVERGER, A. (1993). A consistent test for bivariate dependence. *Int. Stat. Rev.* **61** 419–433.
- [14] FIEDLER, J. (2016). Distances, Gegenbauer expansions, curls, and dimples: On dependence measures for random fields. Doctoral dissertation, Unive. Heidelberg.
- [15] FOKIANOS, K. and PITSILLOU, M. (2017). Consistent testing for pairwise dependence in time series. *Technometrics* **59** 262–270. MR3635048 <https://doi.org/10.1080/00401706.2016.1156024>
- [16] GENE ONTOLOGY CONSORTIUM (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* **32** D258–D261. <https://doi.org/10.1093/nar/gkh036>
- [17] GERSTENBERGER, C. and VOGEL, D. (2015). On the efficiency of Gini’s mean difference. *Stat. Methods Appl.* **24** 569–596. MR3421674 <https://doi.org/10.1007/s10260-015-0315-x>
- [18] GOEMAN, J. J. and BÜHLMANN, P. (2007). Analyzing gene expression data in terms of gene sets: Methodological issues. *Bioinformatics* **23** 980–987.

- [19] GRETTON, A., FUKUMIZU, K., TEO, C. H., SONG, L., SCHÖLKOPF, B. and SMOLA, A. J. (2008). Measuring statistical dependence with Hilbert–Schmidt norms. *Adv. Neural Inf. Process. Syst.* 585–592.
- [20] HUANG, C. and HUO, X. (2017). A statistically and numerically efficient independence test based on random projections and distance covariance. Preprint. Available at [arXiv:1701.06054](https://arxiv.org/abs/1701.06054).
- [21] HUO, X. and SZÉKELY, G. J. (2016). Fast computing for distance covariance. *Technometrics* **58** 435–447. [MR3556612 https://doi.org/10.1080/00401706.2015.1054435](https://doi.org/10.1080/00401706.2015.1054435)
- [22] JENTSCH, C., LEUCHT, A., MEYER, M. and BEERING, C. (2020). Empirical characteristic functions-based estimation and distance correlation for locally stationary processes. *J. Time Series Anal.* **41** 110–133. [MR4048684](https://doi.org/10.1080/00401706.2015.1054435)
- [23] KONG, J., KLEIN, B. E. K., KLEIN, R. and WAHBA, G. (2012). Using distance correlation and SS-ANOVA to assess associations of familial relationships, lifestyle factors, diseases, and mortality. *Proc. Natl. Acad. Sci. USA* **109** 20352–20357.
- [24] LYONS, R. (2013). Distance covariance in metric spaces. *Ann. Probab.* **41** 3284–3305. [MR3127883 https://doi.org/10.1214/12-AOP803](https://doi.org/10.1214/12-AOP803)
- [25] 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** Art. ID 39.
- [26] MONTGOMERY, D. C. (2009). *Introduction to Statistical Quality Control*. Wiley, New York.
- [27] PAN, W., WANG, X., ZHANG, H., ZHU, H. and ZHU, J. (2020). Ball covariance: A generic measure of dependence in Banach space. *J. Amer. Statist. Assoc.* **115** 307–317. [MR4078465 https://doi.org/10.1080/00401706.2015.1054435](https://doi.org/10.1080/00401706.2015.1054435)
- [28] PEREZ-RECHE, F. J., TRIGUERO, C., ZANZOTTO, G. and TRUSKINOVSKY, L. (2016). Origin of scale-free intermittency in structural first-order phase transitions. *Phys. Rev. B* **94** 144102.
- [29] RAMDAS, A., REDDI, S. J., PÓCZOS, B., SINGH, A. and WASSERMAN, L. (2015). On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [30] RICHARDS, M. T., RICHARDS, D. ST. P. and MARTÍNEZ-GÓMEZ, E. (2014). Interpreting the distance correlation results for the COMBO-17 survey. *Astrophys. J. Lett.* **784** Art. ID L34.
- [31] RIZZO, M. L. and SZÉKELY, G. J. (2010). DISCO analysis: A nonparametric extension of analysis of variance. *Ann. Appl. Stat.* **4** 1034–1055. [MR2758432 https://doi.org/10.1214/09-AOAS245](https://doi.org/10.1214/09-AOAS245)
- [32] SEJDINOVIC, D., SRIPERUMBUDUR, B., GRETTON, A. and FUKUMIZU, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Ann. Statist.* **41** 2263–2291. [MR3127866 https://doi.org/10.1214/13-AOS1140](https://doi.org/10.1214/13-AOS1140)
- [33] SHAKED, M. and SHANTHIKUMAR, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics. Springer, New York. [MR2265633 https://doi.org/10.1007/978-0-387-34675-5](https://doi.org/10.1007/978-0-387-34675-5)
- [34] SIMON, N. and TIBSHIRANI, R. (2014). Comment on “Detecting novel associations in large data sets” by Reshef et al Science Dec 16, 2011. Preprint. Available at [arXiv:1401.7645](https://arxiv.org/abs/1401.7645).
- [35] SZÉKELY, G. J. and RIZZO, M. (2019). energy: E-Statistics: Multivariate inference via the energy of data. R package version 1.7-6.
- [36] SZÉKELY, G. J. and RIZZO, M. L. (2009). Brownian distance covariance. *Ann. Appl. Stat.* **3** 1236–1265. [MR2752127 https://doi.org/10.1214/09-AOAS312](https://doi.org/10.1214/09-AOAS312)
- [37] SZÉKELY, G. J. and RIZZO, M. L. (2013). The distance correlation t -test of independence in high dimension. *J. Multivariate Anal.* **117** 193–213. [MR3053543 https://doi.org/10.1016/j.jmva.2013.02.012](https://doi.org/10.1016/j.jmva.2013.02.012)
- [38] SZÉKELY, G. J. and RIZZO, M. L. (2014). Partial distance correlation with methods for dissimilarities. *Ann. Statist.* **42** 2382–2412. [MR3269983 https://doi.org/10.1214/14-AOS1255](https://doi.org/10.1214/14-AOS1255)
- [39] SZÉKELY, G. J., RIZZO, M. L. and BAKIROV, N. K. (2007). Measuring and testing dependence by correlation of distances. *Ann. Statist.* **35** 2769–2794. [MR2382665 https://doi.org/10.1214/009053607000000505](https://doi.org/10.1214/009053607000000505)
- [40] TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics* 448–485. Stanford Univ. Press, Stanford, CA. [MR0120720](https://doi.org/10.1214/09-AOAS312)
- [41] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics **3**. Cambridge Univ. Press, Cambridge. [MR1652247 https://doi.org/10.1017/CBO9780511802256](https://doi.org/10.1017/CBO9780511802256)
- [42] YITZHAKI, S. (2003). Gini’s mean difference: A superior measure of variability for non-normal distributions. *Metron* **61** 285–316. [MR2025523](https://doi.org/10.1017/CBO9780511802256)
- [43] YITZHAKI, S. and SCHECHTMAN, E. (2013). *The Gini Methodology: A Primer on a Statistical Methodology*. Springer Series in Statistics. Springer, New York. [MR3012052 https://doi.org/10.1007/978-1-4614-4720-7](https://doi.org/10.1007/978-1-4614-4720-7)
- [44] ZHOU, Z. (2012). Measuring nonlinear dependence in time-series, a distance correlation approach. *J. Time Series Anal.* **33** 438–457. [MR2915095 https://doi.org/10.1111/j.1467-9892.2011.00780.x](https://doi.org/10.1111/j.1467-9892.2011.00780.x)