DIMACS Series in Discrete Mathematics
and Theoretical Computer Science

On Scale Curves for Nonparametric Description of Dispersion

Jin Wang and Robert Serfling

ABSTRACT. A two-dimensional visual device for description of the dispersion of a multivariate data set in any dimension is the scale curve, a depth-based method introduced by Liu (1990) and treated in detail by Liu, Parelius and Singh (1999). It offers an appealing alternative to covariance matrix methods and even provides an attractive new tool in the univariate case. To support applications such as uniform confidence bounds for the population scale curve, we develop some asymptotics. Uniform strong convergence of the sample scale curve is established under broad conditions and uniform weak convergence under somewhat restricted conditions. Open issues for further investigation are discussed.

1. Introduction and Preliminaries

The geometric structure of a multivariate probabililty distribution F may be described quite naturally in terms of its *contours*. To describe *local* structure, contours are defined by equal levels of *probability density*. On the other hand, for description in terms of *outlyingness* — a *globally* oriented feature — the contours should be defined by a function whose value at any point measures in some sense the outlyingness of that point. Equivalently, typical *depth functions* provide this kind of contour. (For detailed general background on depth functions, see Liu, Parelius and Singh, 1999, Zuo and Serfling [**19**], and Mosler [**9**], and see Serfling [**15**] for a recent overview.) In some cases, these two approaches yield the same family of contours, just indexed differently.

One useful application of the contours is to provide a nonparametric description of the *dispersion* of F, using the *volumes* of the enclosed regions. In particular, for *depth-based* contours based on a sample from F, Liu (1990) and Liu, Parelius and Singh (1999) introduce the "*scale curve*", which plots the enclosed volume as a function of the corresponding probability weight p, for $0 < p < 1$. This provides a two-dimensional visual device for viewing or comparing multivariate datasets of any dimension with respect to their dispersion. It offers an appealing alternative

2000 *Mathematics Subject Classification.* Primary 62G05; Secondary 62G20, 62H05.
Key words and phrases. Multivariate analysis, Nonparametric methods, Scale curve, Weak convergence.
The authors gratefully acknowledge support by NSF Grant DMS-0103698.

to covariance matrix methods and even provides a useful new tool in the univariate case.

Of particular interest is the asymptotic behavior of the sample scale curve. Does it consistently estimate the population scale curve? Suitably normed, does it converge in distribution, so that, for example, uniform confidence bands may be placed on the population scale curve? We establish uniform *strong* convergence of the sample scale curve under quite broad conditions, and uniform *weak* convergence, which poses some difficulties, under somewhat restricted conditions (extending a limited previous investigation of Serfling [**12**]). We also treat certain variants of the scale curve which can be handled in quite straightforward fashion, but which sacrifice some intuitive appeal and are not preferred.

The technical difficulties apply only in the higher dimensional case, and for the univariate case we obtain a complete treatment of asymptotics, both for the sample scale curve and for a modified version related to the familiar "shorth" statistic. This development is provided in Section 2. For the higher dimensional case, we obtain in Section 3 uniform strong convergence for both the unmodified scale curve and a modified version related to the familiar MVE estimator, but uniform weak convergence only for the modified version. Complementary topics are discussed in Section 4, including an approach toward weak convergence of the unmodified sample scale curve. We complete the present section by formulating sample and population scale curves, including modified versions, and discussing basic issues concerning asymptotics.

Formulation of depth-based scale curves. Given F on $\mathbb{R}^d$ and a depth function $D(\boldsymbol{x}, F)$ which provides an F-based center-outward ordering of points $\boldsymbol{x}$ in $\mathbb{R}^d$, a corresponding family of contours is given by the boundaries of "central regions" of form $\{\boldsymbol{x} : D(\boldsymbol{x}, F) \geq \alpha\}$, $\alpha > 0$. Letting $C_{F,D}(p)$ denote the central region having probability weight p, the associated "scale curve" is then a plot of

$$V_{F,D}(p) = \text{volume}(C_{F,D}(p)),\ 0 < p < 1.$$

Since the depth-based central regions are *nested*, the scale curve has an appealing interpretation as quantifying the expansion of the central regions with increasing probability weight p.

As pointed out in Serfling [**12**], a scale curve $V_{F,D}(\cdot)$ has the structure of a *generalized quantile function* in the sense of Einmahl and Mason (1992). Given a probability distribution F on $\mathbb{R}^d$, a class $\boldsymbol{\mathcal{A}}$ of Borel sets in $\mathbb{R}^d$, and a real-valued set function $\lambda(A)$ defined over $A \in \boldsymbol{\mathcal{A}}$, they define the corresponding "generalized quantile function" as

$$U_F(p) = \inf\{\lambda(A) : F(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}\},\ 0 < p < 1, \tag{1.1}$$

and establish weak convergence of an associated empirical process. With $\lambda(A)$ given by the *volume* of A, and $\boldsymbol{\mathcal{A}}$ given by $\boldsymbol{\mathcal{A}}_{F,D} = \{C_{F,D}(p),\ 0 < p < 1\}$, we readily obtain $U_F(p) = V_{F,D}(p)$, the scale curve. That is, $V_{F,D}(p)$ has the representation

$$V_{F,D}(p) = \inf\{\text{volume}(A) : F(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}_{F,D}\},\ 0 < p < 1. \tag{1.2}$$

Sample versions, and issues. A *sample scale curve* is defined by $V_{n,D}(\cdot)$, using the volumes of the sample central regions $C_{F_n,D}(p)$, with F_n the usual empirical distribution based on a sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from F. Thus $C_{F_n,D}(p)$ denotes the smallest region of form $\{\boldsymbol{x} : D(\boldsymbol{x}, F_n) \geq \alpha\}$ having F_n-probability $\geq p$. Via (1.2) we may write

$$V_{n,D}(p) = \inf\{\text{volume}(A) : F_n(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}_{F_n,D}\},\ 0 < p < 1. \quad (1.3)$$

On the other hand, the sample version of (1.1) considered by Einmahl and Mason (1992) is given by

$$U_n(p) = \inf\{\lambda(A) : F_n(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}\},\ 0 < p < 1, \quad (1.4)$$

with $\boldsymbol{\mathcal{A}}$ precisely as in (1.1), i.e., *fixed* and not random. Under regularity conditions they establish weak convergence of the corresponding process

$$n^{1/2}\, u_F(p)^{-1}\,(U_n(p) - U_F(p)),\ 0 < p < 1, \quad (1.5)$$

with $u_F(p)$ a suitable normalizing function (typically the derivative of $U_F(p)$). This does not, however, cover the empirical process based on (1.2) and (1.3), which involves an *empirical* choice of class $\boldsymbol{\mathcal{A}}$.

Of course, as seen in Serfling [**12**], for

$$\widetilde{V}_{n,D}(p) = \inf\{\text{volume}(A) : F_n(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}_{F,D}\},\ 0 < p < 1, \quad (1.6)$$

and $v_{F,D}(p)$ the derivative of $V_{F,D}(p)$, the process

$$n^{1/2}\, v_{F,D}(p)^{-1}\,(\widetilde{V}_{n,D}(p) - V_{F,D}(p)),\ a \leq p \leq b, \quad (1.7)$$

defined over any closed interval $[a, b]$ in $(0, 1)$ satisfies the regularity conditions of Einmahl and Mason (1992) and converges weakly to simply the Brownian bridge over $[a, b]$. Practical implementation of this result, is thwarted, however, by the unknown F being involved both in the class $\boldsymbol{\mathcal{A}}_{F,D}$ in (1.6) and in the normalization $v_{F,D}(p)$ in (1.7).

Thus, in order to be able to exploit the weak convergence result for (1.5), we consider *modified scale curves* defined by (1.2) using classes $\boldsymbol{\mathcal{A}}$ not depending on F (but possibly on $D(\cdot,\cdot)$), i.e.,

$$V^*_{F,D}(p) = \inf\{\text{volume}(A) : F(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}_D\},\ 0 < p < 1, \quad (1.8)$$

with sample version

$$V^*_{n,D}(p) = \inf\{\text{volume}(A) : F_n(A) \geq p,\ A \in \boldsymbol{\mathcal{A}}_D\},\ 0 < p < 1, \quad (1.9)$$

and we *studentize*, introducing uniformly consistent estimators for $v_{F,D}(p)$ or $v^*_{F,D}(p)$. Here we carry out this approach with F restricted to be unimodal and elliptically symmetric, in which case the density-based contours are ellipsoidal and so are the depth-based contours for typical choices of $D(\cdot,\cdot)$. The corresponding modification of (1.7) is seen to converge weakly to the Brownian bridge over $[a, b]$. This yields asymptotic normality of the modified sample scale curve for any fixed p as well as uniform confidence bands for the modified population scale curve over $[a, b]$. Such modified scale curves are not based on nested regions, however, so that lack an interpretation in terms of expanding central regions.

2. Scale curves in the univariate case

In Section 2.1 we formulate a very natural univariate scale curve having strong intuitive appeal and establish its convergence properties. An alternative version is treated in Section 2.2. Studentized versions are developed in Section 2.3, and uniform confidence bands are discussed.

2.1. A natural scale curve and its convergence properties. In the case of univariate continuous F, several typical depth functions — the *halfspace* (Tukey), *simplicial*, and *spatial* depths, for example — are equivalent and therefore generate the same central regions, which in particular are the *nested* intervals

$$C_F(p) = \left[F^{-1}\left(\frac{1}{2}-\frac{p}{2}\right),\ F^{-1}\left(\frac{1}{2}+\frac{p}{2}\right)\right],\ 0 \le p < 1, \tag{2.1}$$

which we note have *equiprobable tails* of weight $p/2$. For comparison, the *projection* depth generates the family of intervals

$$\left[F^{-1}\left(\frac{1}{2}\right)-\delta(p),\ F^{-1}\left(\frac{1}{2}\right)+\delta(p)\right],\ 0 \le p < 1,$$

where $\delta(p)$ is chosen for the given interval to have probability weight p. These intervals also are nested.

Here we adopt the family (2.1) and, accordingly, take as scale curve

$$V_F(p) = F^{-1}\left(\frac{1}{2}+\frac{p}{2}\right) - F^{-1}\left(\frac{1}{2}-\frac{p}{2}\right),\ 0 \le p < 1.$$

Both $V_F(\cdot)$ and its sample version,

$$V_n(p) = F_n^{-1}\left(\frac{1}{2}+\frac{p}{2}\right) - F_n^{-1}\left(\frac{1}{2}-\frac{p}{2}\right),\ 0 \le p < 1,$$

have natural interpretations as quantifying the expansion of nested central regions with increasing probability weight.

Convergence properties for the sample scale curve may be derived from results for the classical sample quantile function $F_n^{-1}(\cdot)$. We have

THEOREM 2.1 (Uniform strong convergence). *Let F^{-1} be continuous with finite $E(X^-)^{1/r}$ and $E(X^+)^{1/s}$ for some $r, s > 0$. Then, for any $[a,b] \subset (0,1)$,*

$$\sup_{p\in[a,b]} |V_n(p) - V_F(p)| \xrightarrow{a.s.} 0.$$

PROOF. The result follows immediately from

$$\sup_{p\in[a',b']} |F_n^{-1}(p) - F^{-1}(p)| \xrightarrow{\text{a.s.}} 0 \tag{2.2}$$

for any $[a',b'] \subset (0,1)$, which itself follows from a strong convergence result of Mason (1982) for F_n^{-1} via

$$\begin{aligned}\inf_{p\in[a',b']} p^r(1-p)^s & \sup_{p\in[a',b']} |F_n^{-1}(p) - F^{-1}(p)| \\ & \le \sup_{p\in[0,1]} |p^r(1-p)^s\,[F_n^{-1}(p) - F^{-1}(p)]| \xrightarrow{\text{a.s.}} 0.\end{aligned}$$

□

Assuming that F has density f and defining the Gaussian process

$$G_F(p) = \frac{1}{f(F^{-1}(\frac{1}{2}+\frac{p}{2}))}B\left(\frac{1}{2}+\frac{p}{2}\right) - \frac{1}{f(F^{-1}(\frac{1}{2}-\frac{p}{2}))}B\left(\frac{1}{2}-\frac{p}{2}\right),\ 0 \le p < 1,$$

with $B(\cdot)$ denoting the standard Brownian bridge, we have

THEOREM 2.2 (Uniform weak convergence). *Let $f(F^{-1}(p))$ be positive and continuous on an open subinterval of* $[0,1]$ *containing* $[a,b]$, *with* $a < 1/2 < b$. *Put* $p' = \min\{1-2a, 2b-1\}$. *Then*

$$\{n^{1/2}(V_n(p) - V_F(p)),\ 0 \le p \le p'\} \xrightarrow{d} \{G_F(p)),\ 0 \le p \le p'\}.$$

PROOF. The result follows by application of the functional delta method (e.g., van der Vaart, 1998, Thm. 20.8) in connection with weak convergence of the classical quantile process $n^{1/2}(F_n(p) - F(p))$ (e.g., Shorack and Wellner, 1986). □

For F *symmetric*, the scale curve becomes simply

$$V_F(p) = 2\left[F^{-1}\left(\frac{1}{2}+\frac{p}{2}\right) - F^{-1}\left(\frac{1}{2}\right)\right],$$

with derivative

$$v_F(p) = \frac{d}{dp}V_F(p) = \frac{1}{f(F^{-1}(\frac{1}{2}+\frac{p}{2}))}.$$

Then the covariance function of $G_F(\cdot)$ becomes

$$\mathrm{Cov}(G_F(p_1), G_F(p_2)) = v_F(p_1)v_F(p_2)(\min\{p_1,p_2\} - p_1p_2)$$

and it follows that the normalized scale curve process converges weakly to simply a Brownian bridge:

COROLLARY 2.1 (The symmetric case). *Let F be symmetric and $f(F^{-1}(p))$ positive and continuous on an open subinterval of* $[0,1]$ *containing* $[\frac{1}{2}-a, \frac{1}{2}+a]$. *Then*

$$\{n^{1/2}v_F(p)^{-1}(V_n(p) - V_F(p)),\ 0 \le p \le 2a\} \xrightarrow{d} \{B(p)),\ 0 \le p \le 2a\}.$$

2.2. An alternative scale functional. A related scale functional is given by the length of the *shortest interval* having F-probability at least p,

$$V_F^*(p) = \inf\{\ell > 0 : F(x+\ell) - F(x-) \ge p\},\ 0 < p < 1.$$

In this case, however, the relevant intervals are not nested, so this curve lacks the intuitive appeal of $V_F(p)$ as quantifying the expansion of central regions. The sample version is

$$V_n^*(p) = \inf\{\ell > 0 : F_n(x+\ell) - F_n(x-) \ge p\},\ 0 < p < 1,$$

i.e., the length of the shortest interval containing at least a fraction p of the data.

In general we have

$$V_F^*(p) \le V_F(p),\ 0 < p < 1,$$

with equality in the case of F *symmetric* with density f positive and continuous on $(\alpha,\beta) \subseteq (-\infty,\infty)$ and *strictly increasing* on $(\alpha, F^{-1}(\frac{1}{2}))$. (In this case equality of the *sample* versions does not hold, however.) Under these conditions on F, weak convergence of the process based on $V_n^*(\cdot)$ has been investigated by Grübel (1988) and Einmahl and Mason (1992), and the result of Corollary 2.1 applies with $V_n(\cdot)$ replaced by $V_n^*(\cdot)$, that is, these two scale curves are asymptotically equivalent in distribution.

In particular, for $p = 1/2$, these weak convergence results cover two well-known robust scale estimators, $V_n(\frac{1}{2})$ = the *interquartile range* and $V_n^*(\frac{1}{2})$ = the *shorth*, respectively. The corresponding *location* estimators given by the midpoints of these intervals behave quite differently with respect to asymptotic distribution, however, the former satisfying $n^{1/2}$-convergence but the latter only $n^{1/3}$-convergence. See Grübel (1988) and Rousseeuw and Leroy (1988) for detailed discussion. Thus, in comparison, the sample scale curve $V_n(\cdot)$ is somewhat more appealing than $V_n^*(\cdot)$, by virtue of its greater interpretability and its (infinitely) more efficient associated location estimator.

2.3. Studentized versions and uniform confidence bands. Practical use of these weak convergence results requires $v_F(p)$ to be replaced by a uniformly consistent estimator. For convenience now confining to symmetric F, we use

$$v_n(p) = \frac{1}{f_n(F_n^{-1}(\frac{1}{2}+\frac{p}{2}))},$$

where f_n is one of many available nonparametric density estimators (e.g., kernel, or nearest neighbor, or orthogonal series type) which satisfy

$$\sup_x |f_n(x) - f(x)| \xrightarrow{p} 0, \ n \to \infty. \tag{2.3}$$

Then, for $f(F^{-1}(\frac{1}{2}+\frac{p}{2}))$ positive and continuous on an open subinterval of $[0,1]$ containing $[a,b]$, straightforward manipulations again utilizing (2.2) yield

$$\sup_{p\in[a,b]} |v_n(p) - v_F(p)| \xrightarrow{p} 0, \tag{2.4}$$

which with standard techniques yields Corollary 2.1, for either $V_n(\cdot)$ or $V_n^*(\cdot)$, with $v_F(p)$ replaced by $v_n(p)$.

As an application, we obtain large sample *uniform* $1-\alpha$ *confidence bands* for the scale curve $V_F(p)$ over any interval $[a,b]$. Defining k_α by $P(\|B(\cdot)\|_a^b \le k_\alpha) = 1-\alpha$, where $\|g(\cdot)\|_a^b = \sup_{a\le p\le b} |g(p)|$, these are given by

$$V_n(p)(\text{or } V_n^*(p)) \pm k_\alpha \frac{v_n(p)}{\sqrt{n}}, \ a \le p \le b.$$

3. The multivariate case

In Section 3.1 uniform strong convergence is obtained both for the sample scale curve $V_{n,D}(p)$ under broad conditions on F and $D(\cdot,F)$ and for a special version $V_F^*(p)$ defined for the case that F is elliptically symmetric and $D(\cdot,F)$ generates ellipsoidal contours. For $V_F^*(p)$ and its studentized version, weak convergence is obtained in Section 3.2, yielding uniform confidence bands.

3.1. Uniform strong convergence of sample scale curves. First let us consider $V_{F,D}(p)$ with derivative denoted by $v_{F,D}(p)$ and with sample version $V_{n,D}(p)$ given by (1.3).

THEOREM 3.1. *Let F and $D(\cdot,F)$ satisfy*

(i) *For any $[a,b] \subset (0,1)$, $\sup_{p\in[a,b]} |v_{F,D}(p)| \le K_{a,b} < \infty$,*

(ii) *$D(x,F) \to 0$ as $\|x\| \to \infty$,*

(iii) *$\sup_S |D(x,F_n) - D(x,F)| \to 0$ a.s. for any bounded set $S \in \mathbb{R}^d$, and*

(iv) *$F^{-1}_{D(\cdot,F_n)}(1-p) \xrightarrow{a.s.} F^{-1}_{D(\cdot,F)}(1-p)$.*

Then, for any $[a,b] \subset (0,1)$,

$$\sup_{p\in[a,b]} |V_{n,D}(p) - V_{F,D}(p)| \xrightarrow{a.s.} 0, \; n \to \infty. \tag{3.1}$$

Proof. Under assumptions (ii), (iii) and (iv), we have by Theorem 4.1 of Zuo and Serfling [**20**] that, for any $\varepsilon > 0$ and all sufficiently large n,

$$C_{F,D}(p-\varepsilon) \subset C_{F_n,D}(p) \subset C_{F,D}(p+\varepsilon) \text{ a.s.}$$

uniformly in $p \in [a,b]$, and hence, using (i), we have

$$\begin{aligned}\sup_{p\in[a,b]} |V_{n,D}(p) - V_{F,D}(p)| &\le \sup_{p\in[a,b]} |V_{F,D}(p+\varepsilon) - V_{F,D}(p-\varepsilon)| \\ &\le 2K_{a,b}\,\varepsilon \text{ a.s.}\end{aligned}$$

This readily yields (3.1). □

Conditions (i) and (ii) are straightforward for typical (continuous) depth functions, and condition (iii) has been established for several popular cases (see Appendix B of Zuo and Serfling [**20**]). We note from the above proof that (i) may be replaced, alternatively, by continuity of $V_{F,D}(p)$ on $(0,1)$. Condition (iv) requires checking case by case; it holds, for example, under Conditions A below (see Lemma 3 and Theorem 1 of He and Wang [**4**] and Corollary 4.1 of Zuo and Serfling [**20**]).

Next we consider a modified scale curve and sample version, defined under

Conditions A.

(i) F is *elliptically symmetric*, i.e., has density of form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2} h((\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})), \; \boldsymbol{x} \in \mathbb{R}^d,$$

for some nonnegative scalar function $h(\cdot)$ and positive definite matrix $\boldsymbol{\Sigma}$,

(ii) $h(\cdot)$ in (i) is *continuous and strictly increasing*, and

(iii) $D(\cdot, F)$ is *affine invariant with maximum at* $\boldsymbol{\mu}$.

For F and $D(\cdot, F)$ satisfying Conditions A, the central regions are nested ellipsoids of form

$$\{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \le c\}$$

and we may define the scale curve $V_{F,D}(\cdot)$ equivalently via (1.1) or (1.8) with $\boldsymbol{\mathcal{A}}$ given by $\boldsymbol{\mathcal{A}}^* = \{\text{all ellipsoids in } \mathbb{R}^d\}$ instead of by $\boldsymbol{\mathcal{A}}_F$ or $\boldsymbol{\mathcal{A}}_D$, i.e., by

$$V_F^*(p) = \inf\{\text{volume}(A) : F(A) \ge p, \; A \in \boldsymbol{\mathcal{A}}^*\},$$

the volume of the smallest ellipsoid having probability weight p under F, which we note does not depend upon the particular choice of depth function $D(\cdot, F)$ satisfying Conditions A. We note that Condition A(iii) is satisfied by many typical depth functions. For detailed discussion, see Liu and Singh (1993, Lemma 3.1) and Zuo and Serfling [**20**, Theorems 3.3 and 3.4]. Consequently, we have

Lemma 3.1. *Under Conditions A,*

$$V_F^*(p) = V_{F,D}(p), \; 0 < p < 1.$$

The sample version

$$V_n^*(p) = \inf\{\text{volume}(A) : F_n(A) \ge p, \; A \in \boldsymbol{\mathcal{A}}^*\},$$

the volume of the smallest ellipsoid containing at least a fraction p of the data, provides a suitable alternative notion of scale curve for the case that F satisfies

Conditions A, but differs from the depth-based version $V_{n,D}(p)$ given by (1.9). It is not difficult to extend Theorem 3.1 to

THEOREM 3.2. *Let F and $D(\cdot, F)$ satisfy Conditions A. Suppose that, for any $[a, b] \subset (0, 1)$, the derivative $v_F^*(p)$ of $V_F^*(p)$ satisfies* $\sup_{p\in[a,b]} |v_F(p)| \leq K_{a,b} < \infty$. *Then, for any $[a, b] \subset (0, 1)$,*

$$\sup_{p\in[a,b]} |V_n^*(p) - V_F^*(p)| \xrightarrow{a.s.} 0, \ n \to \infty. \tag{3.2}$$

3.2. Uniform weak convergence of the modified scale curve. The weak convergence result of Einmahl and Mason (1992) (or see Serfling [**12**]) immediately yields

LEMMA 3.2. *Under Conditions A(i) and A(ii), and assuming that support(F) $= \mathbb{R}^d$, the process*

$$n^{1/2}\, v_F^*(p)^{-1}\, (V_n^*(p) - V_F^*(p)), \ a \leq p \leq b, \tag{3.3}$$

converges weakly to the Brownian bridge $B(\cdot)$ over $[a, b]$.

Let us now studentize by developing for $v_F^*(p)$ a depth-based estimator whose construction does not depend on an assumption of ellipsoidal symmetry for F. We first establish useful representations for $v_F^*(p)$ and $v_{F,D}(p)$.

Denote the squared Mahalanobis distance by

$$R = (\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$$

and its cdf by F_R. Under Conditions A(i) and A(ii), f is constant on the ellipsoids

$$E_c = \{\boldsymbol{x} \in \mathbb{R}^d : (\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = c\}.$$

Let $f_{(c)}$ be the value of f on E_c. It is straightforward to show that

$$V_F^*(p) = \frac{\pi^{d/2} F_R^{-1}(p)^{d/2} |\boldsymbol{\Sigma}|^{1/2}}{\Gamma(d/2 + 1)},$$

whence by differentiation we obtain

LEMMA 3.3. *Under Conditions A(i) and A(ii),*

$$v_F^*(p) = \frac{1}{f_{(F_R^{-1}(p))}}. \tag{3.4}$$

We now establish

LEMMA 3.4. *Let F have a "center" $\boldsymbol{M}_F$ and a density f, and let $D(\cdot, F)$ be differentiable and strictly decreasing along any ray from $\boldsymbol{M}_F$. Then*

$$v_{F,D}(p) = \frac{1}{\textit{average of } f \textit{ over } \partial C_{F,D}(p)}. \tag{3.5}$$

PROOF. For $\boldsymbol{y} = (y_1, \ldots, y_d)' \in \mathbb{R}^d$, let $\alpha = D(\boldsymbol{y}, F)$ and let $(r, \theta_1, \ldots, \theta_{d-1})'$ be the spherical coordinates of $\boldsymbol{y} - \boldsymbol{M}_F$. The mapping $\phi : \boldsymbol{y} \mapsto (\alpha, \theta_1, \ldots, \theta_{d-1})'$ is one-to-one. Denote the Jacobian of the inverse transformation, $\boldsymbol{y} = \phi^{-1}(\alpha, \theta_1, \ldots, \theta_{d-1})$, considered as a function of α, by J_α. Then

$$\begin{aligned} F_D(z) &= P(D(\boldsymbol{X}, F) \leq z) = \int_{\{\boldsymbol{y}: D(\boldsymbol{y},F)\leq z\}} f(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int_0^z \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} f(\phi^{-1}(\alpha, \theta_1, \ldots, \theta_{d-1}))\, |J_\alpha|\, d\theta_{d-1} \cdots d\theta_1 d\alpha, \end{aligned}$$

and

$$f_D(z) = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} f(\phi^{-1}(z, \theta_1, \ldots, \theta_{d-1})) \, |J_z| \, d\theta_{d-1} \cdots d\theta_1.$$

On the other hand,

$$\begin{aligned} \text{volume}(\{\boldsymbol{y} : D(\boldsymbol{y}, F) \geq z\}) &= \int_{\{\boldsymbol{y}: D(\boldsymbol{y},F) \geq z\}} d\boldsymbol{y} \\ &= \int_z^{\max\{D(\cdot,F)\}} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_\alpha| \, d\theta_{d-1} \cdots d\theta_1 d\alpha, \end{aligned}$$

and

$$\frac{d}{dz}\text{volume}(\{\boldsymbol{y} : D(\boldsymbol{y}, F) \geq z\}) = -\int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_z| \, d\theta_{d-1} \cdots d\theta_1.$$

This yields

$$V_{F,D}(p) = \int_{\alpha(p)}^{\max\{D(\cdot,F)\}} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_\alpha| \, d\theta_{d-1} \cdots d\theta_1 d\alpha,$$

where $\alpha(p) = F_D^{-1}(1-p)$, and thus

$$\begin{aligned} v_{F,D}(p) &= -\left(\frac{d\alpha(p)}{dp}\right) \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_{\alpha(p)}| \, d\theta_{d-1} \cdots d\theta_1 \\ &= \frac{1}{f_D(F_D^{-1}(1-p))} \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_{\alpha(p)}| \, d\theta_{d-1} \cdots d\theta_1. \end{aligned}$$

From the above we see that the average of f over $\partial C_{F,D}(p)$ is given by

$$f_D(F_D^{-1}(1-p)) \bigg/ \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \cdots \int_{-\pi/2}^{\pi/2} |J_{\alpha(p)}| \, d\theta_{d-1} \cdots d\theta_1 \; .$$

Thus follows (3.5). □

This result establishes a relationship between the density f and the volume functional $V_{F,D}(p)$: the reciprocal of the average of f over the boundary of $C_{F,D}(p)$ equals the velocity with which $C_{F,D}(p)$ expands toward the tails.

On the basis of Lemma 3.4, a depth-based estimator of $v_{F,D}(p)$ is given by

$$v_{n,D}(p) = \frac{1}{\text{average of } f_n(\boldsymbol{X}_i) \text{ over } \boldsymbol{X}_i \in \partial C_{F_n,D}(\lfloor pn \rfloor/n)},$$

where f_n is a nonparametric density estimator of f. Further, by Lemma 3.3, under Conditions A this is also an estimator of $v_F^*(p)$. Then, under typical consistency conditions on f_n and $D(\boldsymbol{x}, F_n)$, it follows that $v_{n,D}(p)$ is uniformly consistent over $0 \leq p \leq p_0 < 1$, and we obtain

THEOREM 3.3. *Under Conditions A along with consistency conditions on f_n and $D(\boldsymbol{x}, F_n)$, the process*

$$n^{1/2} \, v_{n,D}(p)^{-1} \, (V_n^*(p) - V_F^*(p)), \;\; a \leq p \leq b, \tag{3.6}$$

converges weakly to the Brownian bridge $B(\cdot)$ over $[a, b]$.

This yields, for example, uniform confidence bands for $V_F^*(p)$ in the same fashion as described for the univariate case in Section 2.

4. Complements

Other ways to utilize $V_F(p)$. For each fixed p, the "volume functional" $V_F(p)$ defined over distributions F serves as a measure of multivariate scatter and may be used for ordering multivariate distributions. See Liu, Parelius and Singh (1999) and Zuo and Serfling [**19**]. Further, two multivariate distributions F and G may be compared by a single curve, the graph of $V_G V_F^{-1}$, extending the "spread-spread plot" given in the univariate case by Balanda and MacGillivray (1990). It also serves as the basis for a new measure of multivariate kurtosis (Wang and Serfling [**18**].

Influence function of the volume functional. Influence functions are developed in Wang and Serfling [**17**] for a general class of the generalized quantile functions of Einmahl and Mason (1992). As a special case, the influence function (IF) of $V_{F,D}(p)$, considered as a functional of F, is obtained for D belonging to a general class including the halfspace depth. This IF is a two-valued step function with jump on the boundary of the pth central region, from a negative value inside to a positive value outside. Thus this functional has finite gross error sensitivity and infinite local shift sensitivity, and the influence of contamination at location $\boldsymbol{y}$ causes underestimation or overestimation according as $\boldsymbol{y}$ is within or without the pth central region.

Variants of the scale curve. In some cases it is more convenient to index central regions by a measure of outlyingness other than the probability weight, and then the scale curve is a plot of the volume versus this index. See Serfling [**13**], [**14**] for treatment of a "spatial scale curve" based on the spatial quantile function.

An heuristic paradigm for scale curve asymptotics. Based on *nested* central regions of increasing probability, the scale curve $V_{F,D}(\cdot)$ has an inverse,

$$V_{F,D}^{-1}(y) = F\text{-probability of smallest central region } C_{F,D}(\cdot) \text{ having volume } \geq y,$$

with sample analogue

$$V_{n,D}^{-1}(y) = F_n\text{-probability of smallest central region } C_{F_n,D}(\cdot) \text{ having volume } \geq y,$$

for $y > 0$. Then, using

$$\frac{d}{dy} V_{F,D}^{-1}(y)\,\big|_{y=V_{F,D}(p)} = v_{F,D}(V_{F,D}^{-1}(V_{F,D}(p))^{-1} = v_{F,D}(p)^{-1},$$

we may write

$$\begin{aligned} v_{F,D}(p)^{-1}(V_{n,D}(p) - V_{F,D}(p)) &\doteq V_{F,D}^{-1}(V_{n,D}(p)) - V_{F,D}^{-1}(V_{F,D}(p)) \\ &\doteq V_{n,D}^{-1}(V_{n,D}(p)) - V_{n,D}^{-1}(V_{F,D}(p)) \\ &\doteq p - V_{n,D}^{-1}(V_{F,D}(p)) \\ (&= -(V_{n,D}^{-1}(V_{F,D}(p)) - V_{F,D}^{-1}(V_{F,D}(p)))\ ). \end{aligned}$$

Here the first approximation step is based on Taylor expansion, the second on a modulus of continuity result for $(V_{n,D}^{-1} - V_{F,D}^{-1})(\cdot)$, and the third on the definition of $V_{n,D}^{-1}$. These steps all need to be made precise in order to obtain weak convergence of the scale curve process $v_{F,D}(\cdot)^{-1}(V_{n,D}(\cdot) - V_{F,D}(\cdot))$. This is similar to the treatment of the classical univariate quantile process by reduction to the classical empirical process, except that here we have reduced to the empirical probability weight $V_{n,D}^{-1}(V_{F,D}(p))$ of the central region $C_{F,D}(p)$, a function which increases by jumps of size $1/n$ as $C_{F,D}(p)$ expands with increasing p. A remaining issue is to deal with the difference between

$$V_{n,D}^{-1}(V_{F,D}(p)) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{\boldsymbol{X}_j \in \text{ smallest } C_{F_n,D}(\cdot) \text{ with volume } \geq V_{F,D}(p)\}$$

and its variant with $C_{F_n,D}$ replaced by $C_{F,D}$, which simplifies to

$$\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}\{\boldsymbol{X}_j \in C_{F,D}(p)\}$$

(in which case convergence to the Brownian bridge would be immediate). We will pursue this approach to scale curve asymptotics in future work.

Other symmetry structures for F. The results for F ellipsoidally symmetric can be extended to the case of F having mean $\boldsymbol{\mu}$, covariance matrix $\boldsymbol{\Sigma}$, and density of form

$$f(\boldsymbol{x}) = |\boldsymbol{\Sigma}|^{-1/2}\, h(\|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})\|)$$

for some continuous and strictly increasing nonnegative function h and choice of norm $\|\cdot\|$. The contours of equal density enclose regions of form

$$\{\boldsymbol{x} : \|\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \leq a\}.$$

For the Euclidean norm $\|\cdot\|_2$, F is ellipsoidally symmetric and these regions are ellipsoids. Alternatively, one may consider the L_1 norm, $\|\boldsymbol{x}\| = |x_1| + \cdots + |x_d|$ for $\|\boldsymbol{x}\| = (x_1, \ldots, x_d)'$. In this case the contours enclose hypertetrahedral regions.

With Conditions A modified using the above generalized condition, and defining $V_F^*(\cdot)$ accordingly, Lemmas 3.1 and 3.2 and Theorems 3.2 and 3.3 hold unchanged, and Lemmas 3.3 and 3.4 hold with obvious modifications. The same depth-based estimator of $v_{F,D}(p)$ is used unchanged.

In this case, we can estimate $v_F^*(p)$ directly not using D, via

$$v_F(p) = \frac{1}{f\left(F^{-1}_{\|\boldsymbol{\Sigma}^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\|}(p)\right)}$$

and an empirical analogue based on f_n and

$$G_n(y) = \frac{1}{n}\sum_{1}^{n} \mathbf{1}\{\|\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{X_i} - \hat{\boldsymbol{\mu}}))\| \leq y\}.$$

One advantage of the depth-based estimator is that it does not depend upon such an explicit structural assumption.

References

[1] K. P. Balanda and H. L. MacGillivray, *Kurtosis and spread*, Can. J. Statist. **18** (1990), 17–30.

[2] J. H. J. Einmahl and D. M. Mason, *Generalized quantile processes*, Ann. Statist. **20** (1992), 1062–1078.

[3] R. Grübel, *The length of the shorth*, Ann. Statist. **16** (1988), 619–628.

[4] X. He and G. Wang, *Convergence of depth contours for multivariate datasets*, Ann. Statist. **25** (1997), 495–504.

[5] R. Y. Liu, *On a notion of data depth based on random simplices*, Ann. Statist. **18** (1990), 405–414.

[6] R. Y. Liu, J. M. Parelius and K. Singh, *Multivariate analysis by data depth: Descriptive statistics, graphics and inference* (with discussion), Ann. Statist. **27** (1999), 783–858.
[7] R. Y. Liu and K. Singh, *A quality index based on data depth and multivariate rank tests*, J. Amer. Statist. Assoc. **88** (1993), 252–260.
[8] D. M. Mason, *Some characterizations of almost sure bounds for weighted mutlidimensional empirical distributions and a Glivenko-Cantelli theorem for sample quantiles*, Z. Wahrsch. verw. Geb. **59** (1982), 505–513.
[9] K. Mosler, *Multivariate Dispersion, Central Regions and Depth*, Springer, New York, 2002.
[10] P. J. Rousseeuw and A. M. Leroy, *A robust scale estimator based on the shortest half*, Statist. Neerland. **42** (1988), 103–116.
[11] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York, 1986.
[12] R. Serfling, *Generalized quantile processes based on multivariate depth functions, with applications in nonparametric multivariate analysis*, J. Multiv. Analys. **83** (2002), 232–247.
[13] R. Serfling, *A depth function and a scale curve based on spatial quantiles*, Statistical Data Analysis Based On the L_1-Norm and Related Methods (Y. Dodge, ed.), (2002), 25-38, Birkhaüser.
[14] R. Serfling, *Nonparametric multivariate descriptive measures based on spatial quantiles*, J. Statist. Plann. Inf. **123** (2004), 259–278.
[15] R. Serfling, *Depth functions in nonparametric multivariate inference*, *this volume*, to appear.
[16] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.
[17] J. Wang and R. Serfling, *Influence functions for a general class of depth-based generalized quantile functions*, J. Multiv. Analysis (2005), to appear.
[18] J. Wang and R. Serfling, *Nonparametric multivariate kurtosis and tailweight measures*, J. Nonpar. Statist. (2005), to appear.
[19] Y. Zuo and R. Serfling, *General notions of statistical depth function*, Ann. Statist. **28** (2000), 461–482.
[20] Y. Zuo and R. Serfling, *Structural properties and convergence results for contours of sample statistical depth functions*, Ann. Statist. **28** (2000), 483–499.

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083 USA

Current address: Department of Mathematics and Statistics, Northern Arizona University, Flagstaff, Arizona 86011-5717 USA

E-mail address: `jin.wang@nau.edu`

Department of Mathematical Sciences, University of Texas at Dallas, Richardson, Texas 75083 USA

E-mail address: `serfling@utdallas.edu`

URL: `www.utdallas.edu/∼serfling`