On Equivariance/Invariance Properties of Multivariate Quantile and Related Functions, and the Role of Standardization

Robert Serfling[1]

University of Texas at Dallas

Nonparametric Statistics Workshop
April 16, 2009 (updated 06/14/09)

FOCAL TOPICS

Some Questions are ...

- The Goal of Nonparametric Methods is ... ?
- Multivariate Analysis Needs ... ?
- How are Multivariate Depth, Outlyingness, Rank, and Quantile Functions Interrelated?

The Answers are ...

- The Goal of Nonparametric Methods is ...
  Robust Parametric Modeling!
- Multivariate Analysis Needs ...
  Nonparametrics!
- Multivariate Depth, Outlyingness, Rank, and Quantile Functions are ...
  Equivalent!

Nonparametrics & *Robust Parametric Modeling*

- *Parametric models* yield the most complete explanations.
- Fitting of such models is an *ideal endpoint*.
- Due to uncertainty about modeling the uncertainty, the practical goal is *robust parametric modeling*.
- Nonparametric statistics is *an intermediate process* oriented to, and leading toward, this goal.
- Thus nonparametric statistics is *not an end in itself.*
- And *nonparametric approaches should work meaningfully in parametric settings*.

Multivariate Analysis Needs *Nonparametrics*

- For *multivariate data*, there are *relatively few tractable models* (although the situation is improving – e.g., skew-normal and skew-t distributions).
- The wealth of theory available for *multivariate normal* gives it a *central role* (but dimension d may not exceed sample size n).
- To avoid default normal modeling in the multivariate case, *nonparametrics* is even more needed.
- However, in the multivariate case, nonparametric methods need much more development – *the challenge!*

Multivariate *Depth*, *Outlyingness*, *Quantiles*, *Ranks*

- "*Order statistics*", "*outlier identification*", "*quantiles*", "*signs*", and "*ranks*" have their own special roles and their own "practitioners" and "afficionados".
- Along with "*symmetry*", these together comprise the *fundamental elements of nonparametric description*.
- *Intuitively*, they are *interrelated*.
- In fact, D, O, Q, and R are *equivalent* methodologies.
- We "prove" this claim by making *the right definitions*!

How to *Detect Outliers*?

- "*Outliers* have been discussed for centuries:

  Francis Bacon, 1620 ... Daniel Bernoulli, 1777 ...

  Benjamin Pierce, 1852 ... Barnett and Lewis, 1995 ...
- In *higher dimension*, detection by visualization fails.
- Thus we need *algorithmic approaches*:

> Given a cdf F on $\mathbb{R}^d$, an *outlyingness function* $O(\mathbf{x}, F)$ is an associated *center-outward ordering* of points $\mathbf{x}$ in $\mathbb{R}^d$ with *higher* values representing greater "outlyingness".

Finding Outliers via SPSS (For Example)

And SPSS offers algorithms for you!!!

SPSS's advertisement in *Amstat News*, 2007:

> **Quickly find multivariate outliers**
> Prevent outliers from skewing analyses when you use the Anomaly Detection Procedure. This procedure searches for unusual cases based on deviations from similar cases and gives reasons for such deviations. You can flag outliers by creating a new variable. Once you have identified unusual cases, you can further examine them and determine if they should be included in your analyses.

Projection Pursuit Approach

- Given a *univariate* outlyingness function $O_1(\cdot, \cdot)$, define

$$O_d(\mathbf{x}, F) = \sup_{\|\mathbf{u}\| = 1} O_1(\mathbf{u}'\mathbf{x}, F_{\mathbf{u}'\mathbf{X}}), \ \mathbf{x} \in \mathbb{R}^d.$$

- For example, start with *univariate scaled deviation*

$$O_1(x, F) = \left| \frac{x - \mu(F)}{\sigma(F)} \right|,$$

for $\mu(F)$ and $\sigma(F)$ location and spread measures, e.g., Med(F) and MAD(F). Then $O_d(\mathbf{x}, F)$ is *affine invariant*.

Spatial Approach

- Using the d-dimensional, or "*spatial*", sign function (or *unit vector function*),

$$\mathbf{S}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}, \ \mathbf{x} \in \mathbb{R}^d,$$

we obtain the "*spatial outlyingness*"

$$O_{\mathrm{S}}(\mathbf{x}, F) = \|E\mathbf{S}(\mathbf{x} - \mathbf{X})\|, \ \mathbf{x} \in \mathbb{R}^d.$$

- It is (only) *orthogonally invariant*, for $d \geq 2$.
- In the *univariate* case, this is $O_1(x, F) = |2F(x) - 1|$.

Mahalanobis Distance Approach

- Substitution of *multivariate* location and spread measures $\mathbf{m}(F)$ and $\mathbf{S}(F)$ in the *scaled deviation* $O_1(\cdot, \cdot)$ yields the popular "*Mahalanobis distance outlyingness*"

$$O_{\mathrm{MD}}(\mathbf{x}, F) = \|\mathbf{S}(F)^{-1/2}(\mathbf{x} - \mathbf{m}(F))\|.$$

- It is *affine invariant*.
- However, the outlyingness contours are necessarily *ellipsoidal*, regardless of the shape of F.

How to Define *Multidimensional Order Statistics*?

- *Depth function* $D(\mathbf{x}, F)$: a center-outward ordering with higher values representing greater "centrality".

  Tukey '75 – Liu '88 – Donoho & Gasko '92 – Vardi & Zhang '00 – Zuo '03
- Compensates for lack of linear order in $\mathbb{R}^d$, $d \geq 2$, by *orienting to a "center"*.
- *Maximum depth points* define a notion of "center" and a notion of "multidimensional median" $\mathbf{M}_F$.
- Can *order data points* by their *sample depths*.

Example & Nonexample

- *Halfspace Depth* [Tukey, 1975]

$$D_{\mathrm{H}}(\mathbf{x}, F) = \inf\{P(H) : \mathbf{x} \in H \text{ closed halfspace}\}, \quad \mathbf{x} \in \mathbb{R}^d$$

- However, the *density function* is *not* a depth:
  - It does not in general measure centrality or outlyingness.
  - Its interpretation is *local* and has *no global perspective*.
  - The point of maximality is not interpretable as a center.
  - For F uniform on $[0, 1]^d$, for example, the density function yields no contours at all.

How to Define "*Quantiles*" in $\mathbb{R}^d$?

- *Quantile function* $\mathbf{Q}(\mathbf{u}, F)$: attaches to each $\mathbf{x}$ a "quantile representation", indexed by $\mathbf{u}$ in $\mathbb{B}^{d-1}(\mathbf{0})$, with *nested* contours
$$\{\mathbf{Q}(\mathbf{u}, F) : \|\mathbf{u}\| = c\}, \; 0 \leq c < 1.$$
- For *quantile-based* inference in $\mathbb{R}^d$, the "center" $\mathbf{Q}(\mathbf{0}, F)$ should be interpretable as a d-dimensional *median* $\mathbf{M}_F$.
- For $\mathbf{u} \neq \mathbf{0}$, the index $\mathbf{u}$ represents *direction* in some sense: e.g., *direction* to $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$ from $\mathbf{M}_F$, or *expected direction* to $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$ from random $\mathbf{X} \sim F$.
- The *magnitude* $\|\mathbf{u}\|$ represents an *outlyingness parameter*.

Example: the *Spatial Quantile Function*

- The *spatial* quantile function $\mathbf{Q}_S(\mathbf{u}, F)$ gives $\boldsymbol{\theta}$ in $\mathbb{R}^d$ minimizing $E\{\Phi(\mathbf{u}, \mathbf{X} - \boldsymbol{\theta})\}$, where $\Phi(\mathbf{u}, \mathbf{t}) = \|\mathbf{t}\| + \mathbf{u}'\mathbf{t}$

  [Dudley and Koltchinskii, 1992, Chaudhuri, 1996, and Koltchinskii, 1997].
- $\mathbf{Q}_S(\mathbf{u}, F)$ is the solution $\mathbf{x}$ of the equation

$$\mathbf{u} = E\mathbf{S}(\mathbf{x} - \mathbf{X}).$$

- It is (only) *orthogonally equivariant*, for $d \geq 2$.
- $\mathbf{Q}_S(\mathbf{0}, F)$ is the well-known *spatial median*.
- For $\mathbf{x} = \mathbf{Q}_S(\mathbf{u}, F)$, we have

$$\|\mathbf{u}\| = \|E\mathbf{S}(\mathbf{x} - \mathbf{X})\| = O_S(\mathbf{x}, F).$$

How to Define "*Signs*" and "*Ranks*" in $\mathbb{R}^d$?

- *Centered rank function* $\mathbf{R}(\mathbf{x}, F)$: takes values in $\mathbb{B}^{d-1}(\mathbf{0})$, with origin $\mathbf{0}$ assigned to a multivariate median $\mathbf{x} = \mathbf{M}_F$, and for other $\mathbf{x}$ denotes a "directional rank" in $\mathbb{B}^{d-1}(\mathbf{0})$.
- The magnitude $\|\mathbf{R}(\mathbf{x}, F)\|$ measures outlyingness of $\mathbf{x}$.
- *Univariate case.* $R(x, F) = 2F(x) - 1$, with its *sign* giving "direction" (from median $F^{-1}(1/2)$), and its *magnitude* providing the "rank" of x.
- For testing $H_0 : \mathbf{M}_F = \boldsymbol{\theta}_0$, the *sample version* of $\mathbf{R}(\boldsymbol{\theta}_0, F)$ provides a natural *test statistic*, a multivariate version of the *univariate sign test*.

Example: the *Spatial Centered Rank Function*

- The *spatial* centered rank function [Möttönen and Oja, 1995] is

$$\mathbf{R}_S(\mathbf{x}, F) = E\mathbf{S}(\mathbf{x} - \mathbf{X}), \ \mathbf{x} \in \mathbb{R}^d.$$

- For testing H_0 : "spatial median $= \boldsymbol{\theta}_0$", the statistic

$$\sum_{i=1}^{n} \mathbf{S}(\boldsymbol{\theta}_0 - \mathbf{X}_i)$$

provides a *spatial sign test statistic*.

- It is (only) *orthogonally invariant*, for $d \geq 2$

Equivalence: the *D-O-Q-R Paradigm*

Depth, outlyingness, quantiles, and ranks in $\mathbb{R}^d$ are equivalent.

- $D(\mathbf{x}, F)$ and $O(\mathbf{x}, F)$ are equivalent (inversely).
- $\mathbf{Q}(\mathbf{u}, F)$ and $\mathbf{R}(\mathbf{x}, F)$ are equivalent (inversely).
- These couplets are linked by
  - a) $O(\mathbf{x}, F) = \|\mathbf{R}(\mathbf{x}, F)\| \ (= \|\mathbf{u}\|)$,
  - b) $D(\mathbf{x}, F)$ induces a corresponding $\mathbf{Q}(\mathbf{u}, F)$.

Each of D, O, $\mathbf{Q}$, and $\mathbf{R}$ can generate the others, although they are very different in conceptual meaning and appeal.

Example: *Depth-Induced Quantile Functions*

For $D(\mathbf{x}, F)$ having nested contours enclosing "median" $\mathbf{M}_F$ and bounding "central regions" $\{\mathbf{x} : D(\mathbf{x}, F) \geq \alpha\}$, $\alpha > 0$, the depth contours induce a quantile representation for $\mathbf{x} \in \mathbb{R}^d$:

- For $\mathbf{x} = \mathbf{M}_F$, denote it by $\mathbf{Q}(\mathbf{0}, F)$.
- For $\mathbf{x} \neq \mathbf{M}_F$, denote it by $\mathbf{Q}(\mathbf{u}, F)$ with $\mathbf{u} = p\mathbf{w}$, where p is the probability weight of the central region with $\mathbf{x}$ on its boundary and $\mathbf{w}$ is the unit vector toward $\mathbf{x}$ from $\mathbf{M}_F$.

Then $\mathbf{u} = \mathbf{R}(\mathbf{x}, F)$ indicates direction toward $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$ from $\mathbf{M}_F$, and $\|\mathbf{u}\| = \|\mathbf{R}(\mathbf{x}, F)\|$ is the *probability weight* of the central region with $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$ on its boundary.

Desired *Equivariance* and *Invariance* Properties

- *How should estimators and test statistics, or depth, outlyingness, quantile, and rank functions, change when the data are transformed to other coordinates?*
- *Quantile functions* on $\mathbb{R}^d$ are desirably *equivariant*, and *depth and outlyingness functions* should be *invariant*.
- The *new quantile representation* of a point $\mathbf{x}$ should be given by the *same transformation* of the original quantile representation, subject to a possible reindexing,
- The *outlyingness* of $\mathbf{x}$ should remain unchanged.

Popular Points of View

- Different estimators which are equal for a given data point should continue to agree after transformation to another coordinate system.
- Test procedures should make the same decision about the equivalent null hypothesis after transformation.
- Likewise, *p*-values should not change.
- Points branded as "outliers" should remain so after transformation to other coordinates.
- In general, a statistical decision procedure should be independent of the particular coordinate system.

Further Points, and *Our Goal*

- Equivariance/invariance provides a *technical convenience*:
  - For purposes of efficiency comparisons or for setting outlyingness thresholds, for example, one may without loss of generality represent a parametric family by a single member for all computations.
- *Unqualified insistence* on equivariance/invariance as a principle is *not justified*, however, for it may lead to undue compromises of efficiency or robustness, for example.
- We will <u>*formulate equivariance/invariance technically*</u> and see how to <u>*produce it through suitable standardization*</u>.

Equivariance of Multivariate Quantile Functions

Definition. An $\mathbb{R}^d$-valued quantile function $\mathbf{Q}(\mathbf{u}, F)$, $\mathbf{u} \in \mathbb{B}^{d-1}(\mathbf{0})$, is *affine equivariant* if, for $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ with any nonsingular $d \times d$ $\mathbf{A}$ and any d-vector $\mathbf{b}$,

$$\boxed{\mathbf{Q}\left(\mathbf{v}, F_{\mathbf{Y}}\right) = \mathbf{A}\,\mathbf{Q}(\mathbf{u}, F_{\mathbf{X}}) + \mathbf{b},\ \ \mathbf{u} \in \mathbb{B}^{d-1}(\mathbf{0})}$$

with a $\mathbb{B}^{d-1}(\mathbf{0})$-valued re-indexing $\mathbf{v} = \mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ which satisfies

$$\boxed{\|\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}})\| = \|\mathbf{u}\|,\ \ \mathbf{u} \in \mathbb{B}^{d-1}(\mathbf{0})}$$

Equivariance of Multivariate Median

For the *median* $\mathbf{Q}(\mathbf{0}, F_{\mathbf{X}})$, the equivariance property may be stated simply

$$\mathbf{Q}(\mathbf{0}, F_{\mathbf{Y}}) = \mathbf{A}\,\mathbf{Q}(\mathbf{0}, F_{\mathbf{X}}) + \mathbf{b}.$$

Equivariance of Contours

- Denote the *contours* of a quantile function $\mathbf{Q}(\cdot, F)$ by

$$\widetilde{\mathbf{Q}}(c, F) = \{\mathbf{Q}(\mathbf{u}, F) : \|\mathbf{u}\| = c\}, \ 0 < c < 1.$$

- If $\mathbf{Q}(\cdot, F)$ is affine equivariant, then equivalently so are the contours: for $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$,

$$\widetilde{\mathbf{Q}}(c, F_{\mathbf{Y}}) = \mathbf{A}\, \widetilde{\mathbf{Q}}(c, F_{\mathbf{X}}) + \mathbf{b}, \ 0 < c < 1.$$

- Here the mapping $\mathbf{u} \mapsto \mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ is implicit.

Equivariance/Invariance of Related Functions

- Equivariance of $\mathbf{Q}(\cdot, F)$ yields equivariance and invariance properties for the related D, O, and R functions.
- The definition of the centered rank function as the *inverse of the quantile function* immediately yields *equivariance of* $\mathbf{R}(\cdot, F)$, in the following sense:

$$\boxed{\mathbf{R}(\mathbf{y}, F_{\mathbf{Y}}) = \mathbf{v}\left(\mathbf{R}(\mathbf{x}, F_{\mathbf{X}}), \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}\right)}$$

- In turn, the relation $O(\mathbf{x}, F) = \|\mathbf{R}(\mathbf{x}, F)\|$ yields *invariance* of $O(\mathbf{x}, F_{\mathbf{X}})$ and likewise of $D(\mathbf{x}, F_{\mathbf{X}})$. (These also follow from the definition of equivariance of $\mathbf{Q}(\cdot, F)$.)

$\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{x}})$ for *Depth-Induced* $\mathbf{Q}(\mathbf{u}, F)$

- For an *affine equivariant* depth-induced quantile function $\mathbf{Q}(\cdot, F)$, it follows that $\mathbf{M}_{\mathbf{Y}} = \mathbf{A}\,\mathbf{M}_{\mathbf{X}} + \mathbf{b}$ and then the unnormalized direction vector toward $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ from $\mathbf{M}_{\mathbf{Y}}$ is $\mathbf{A}(\mathbf{x} - \mathbf{M}_{\mathbf{X}})$.
- Then $\mathbf{R}(\mathbf{y}, F_{\mathbf{Y}}) = c_0\mathbf{A}\mathbf{R}(\mathbf{x}, F_{\mathbf{X}})$ for some constant c_0, or equivalently $\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = c_0\mathbf{A}\mathbf{u}$, where $\mathbf{x} = \mathbf{Q}(\mathbf{u}, F)$.
- *Outlyingness invariance* then requires $|c_0| = \|\mathbf{u}\|/\|\mathbf{A}\mathbf{u}\|$, yielding, for either choice of sign,

$$\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = \pm\frac{\|\mathbf{u}\|}{\|\mathbf{A}\mathbf{u}\|}\mathbf{A}\mathbf{u}.$$

$\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{x}})$ for the *Spatial* $\mathbf{Q}(\mathbf{u}, F)$

- A well-known limitation of the spatial quantile function is its *orthogonal*, rather than full affine, equivariance.
- That is, the desired equivariance holds only for $\mathbf{A}$ orthogonal, in which case $\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = \pm\mathbf{A}\mathbf{u}$.
- A point $\mathbf{x}$ labeled a (spatial) "outlier" or "nonoutlier" would have the same classification after *orthogonal* transformation to a new coordinate system but not necessarily after transformation by *heterogeneous scale changes*.

Weak Covariance (WC) Functionals

- **Definition**. A matrix-valued functional $\mathbf{C}(F)$ is a *weak covariance (WC) functional* if, for $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ with any nonsingular $\mathbf{A}$ and any $\mathbf{b}$,

$$\boxed{\mathbf{C}(F_{\mathbf{Y}}) = k_1\, \mathbf{A}\, \mathbf{C}(F_{\mathbf{X}})\, \mathbf{A}'}$$

  with $k_1 = k_1(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ a positive scalar function.
- $k_1(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = 1$ gives the usual "*covariance functional*". [e.g., Lopuhaä and Rousseeuw, 1991]
- A WC functional is also sometimes called an "*affine equivariant shape functional*". [Paindaveine, 2008, and Tyler, Critchley, Dümbgen, and Oja, 2009].

Making Spatial Quantiles *Affine* Equivariant

- **Theorem**. For any WC functional $\mathbf{C}(F)$, the quantile function

$$\mathbf{Q}_{\mathrm{MS}}(\mathbf{u}, F_{\mathbf{X}}) = \mathbf{C}(F_{\mathbf{X}})^{1/2}\, \mathbf{Q}_{\mathrm{S}}(\mathbf{u}, F_{\mathbf{C}(F_{\mathbf{X}})^{-1/2}\mathbf{X}})$$

  is *affine equivariant*.

- The relevant re-indexing for $\mathbf{X} \mapsto \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ is

$$\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = \widetilde{\mathbf{A}}(\mathbf{A}, F_{\mathbf{X}})\, \mathbf{u},$$

  where $\widetilde{\mathbf{A}}(\mathbf{A}, F_{\mathbf{X}}) = (\mathbf{A}\, \mathbf{C}(F_{\mathbf{X}})\, \mathbf{A}')^{1/2}(\mathbf{A}')^{-1}\mathbf{C}(F_{\mathbf{X}})^{-1/2}$, which is *orthogonal*.

Comments on the quantile function $\mathbf{Q}_{\text{MS}}(\cdot, F)$

- In homage to Mahalanobis, who promoted the role of *standardization* in multivariate analysis, we may call $\mathbf{Q}_{\text{MS}}(\cdot, F)$ the *"Mahalanobis spatial" quantile function* corresponding to $\mathbf{C}(\cdot)$-standardization.
- The favorable properties of the spatial quantile function $\mathbf{Q}_{\text{S}}(\cdot, F)$ carry over to $\mathbf{Q}_{\text{MS}}(\cdot, F)$, subject to the quality of choice of the WC functional $\mathbf{C}(F)$.
- For example, in contrast to the *"Mahalanobis distance" quantile function* $\mathbf{Q}_{\text{MD}}(\cdot, F)$ corresponding to $O_{\text{MD}}(\cdot, F)$ discussed earlier, the contours of $\mathbf{Q}_{\text{MS}}(\cdot, F)$ *need not be elliptical*.

Transformation-Retransformation (TR) Functionals

- **Definition.** [e.g., Randles, 2000] A matrix-valued functional $\mathbf{M}(F)$ is a *transformation-retransformation (TR) functional* if, for $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ with any nonsingular $\mathbf{A}$ and any $\mathbf{b}$,

$$\boxed{\mathbf{A}'\mathbf{M}(F_{\mathbf{Y}})'\mathbf{M}(F_{\mathbf{Y}})\mathbf{A} = k_2\,\mathbf{M}(F_{\mathbf{X}})'\mathbf{M}(F_{\mathbf{X}})}$$

  with $k_2 = k_2(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ a positive scalar function.
- TR approaches modify *estimation (testing)* procedures to achieve (*hopefully*) full *affine equivariance (invariance)*.
  - Carry out the procedure on transformed data $\mathbf{M}(\mathbb{X}_n)\mathbb{X}_n$, then retransform to original coordinates via $\mathbf{M}(\mathbb{X}_n)^{-1}$.
  - Verify that the equivariance (invariance) holds.

Connection between *TR* and *WC* Functionals

- **Theorem**. Every TR functional $\mathbf{M}(F)$ is equivalent to a WC functional, and conversely.
  - For a TR $\mathbf{M}(F)$, $\mathbf{C}(F) = (\mathbf{M}(F)'\mathbf{M}(F))^{-1}$ is a WC fcnl.
  - For a WC $\mathbf{C}(F)$, $\mathbf{M}(F) = \mathbf{C}(F)^{-1/2}$ is a TR fcnl.
- Selection of a TR functional is merely an indirect but equivalent way to select a WC functional.
- Extensive literature on covariance functionals provides many choices meeting various criteria of robustness and computational efficiency.

TR Connection between $\mathbf{Q}_{\mathrm{MS}}(\cdot, F)$ and $\mathbf{Q}_{\mathrm{S}}(\cdot, F)$

- The defining formula

$$\mathbf{Q}_{\mathrm{MS}}(\mathbf{u}, F_{\mathbf{X}}) = \mathbf{C}(F_{\mathbf{X}})^{1/2}\mathbf{Q}_{\mathrm{S}}(\mathbf{u}, F_{\mathbf{C}(F_{\mathbf{X}})^{-1/2}\mathbf{X}})$$

  exhibits a *"transformation-retransformation" (TR) representation* of $\mathbf{Q}_{\mathrm{MS}}(\cdot, F)$ in terms of $\mathbf{Q}_{\mathrm{S}}(\cdot, F)$.
- The outlyingness function corresponding to $\mathbf{Q}_{\mathrm{MS}}(\cdot, F)$ via the D-O-Q-R paradigm is then given by

$$O_{\mathrm{MS}}(\mathbf{x}, F_{\mathbf{X}}) = O_{\mathrm{S}}\left(\mathbf{C}(F_{\mathbf{X}})^{-1/2}\mathbf{x}, F_{\mathbf{C}(F_{\mathbf{X}})^{-1/2}\mathbf{X}}\right)$$

  and is *affine invariant*.

TR *Sample Version* of $\mathbf{Q}_{\mathrm{S}}(\cdot, F)$

- Advocates of $\mathbf{Q}_{\mathrm{S}}(\cdot, F)$ reject the *idea* of $\mathbf{Q}_{\mathrm{MS}}(\cdot, F)$.
- *In practice*, however, instead of the sample version $\mathbf{Q}_{\mathrm{S}}(\cdot, \mathbb{X}_N)$ they use, for a particular TR fcnl $\mathbf{M}_0(\mathbb{X}_n)$,

$$\mathbf{Q}^{(\mathrm{TR})}(\mathbf{u}, \mathbb{X}_n) = \mathbf{M}_0(\mathbb{X}_n)^{-1}\mathbf{Q}_{\mathrm{S}}(\mathbf{u}, \mathbf{M}_0(\mathbb{X}_n)\mathbb{X}_n),$$

  which achieves *full affine equivariance*.

  [see Chakraborty, Chaudhuri, and Oja, 1998, and Chakraborty, 2001]

- What $\mathbf{Q}^{(\mathrm{TR})}(\mathbf{u}, \mathbb{X}_n)$ actually *estimates*, however, is not $\mathbf{Q}_{\mathrm{S}}(\mathbf{u}, F)$, but rather $\mathbf{Q}_{\mathrm{MS}}(\mathbf{u}, F)$ with $\mathbf{C}(F) = \mathbf{M}_0(F)^{-1/2}$.

Invariant Coordinate System (ICS) Functionals

Definition. [Tyler, Critchley, Dümbgen, and Oja, 2009] A matrix-valued functional $\mathbf{D}(F)$ is an *invariant coordinate system (ICS) functional* if the $\mathbf{D}(\cdot)$-standardizion of $\mathbf{X}$,

$$\mathbf{D}(F_{\mathbf{X}})\mathbf{X},$$

remains unaltered after *affine transformation* to $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ followed by $\mathbf{D}(\cdot)$-*standardization* to

$$\mathbf{D}(F_{\mathbf{Y}})\mathbf{Y},$$

except for *coordinatewise scale changes*, *sign changes* and *translations*.

Practical Interpretation of ICS-Standardization

- With $\mathbf{D}(\cdot)$ an ICS functional, any *geometric structures or patterns* identified in a $\mathbf{D}(\cdot)$-standardized data set
$$\mathbf{D}(\mathbb{X}_n)\mathbb{X}_n$$
remain unaltered after *affine transformation* to $\mathbb{Y}_n = \mathbf{A}\mathbb{X}_n + \mathbf{b}$ followed by $\mathbf{D}(\cdot)$-*standardization* to
$$\mathbf{D}(\mathbb{Y}_n)\mathbb{Y}_n,$$
except for *coordinatewise scale changes*, *sign changes* and *translations*.
- In some applications, e.g., *outlyingness* considerations, both *homogeneity of scale changes* and *homogeneity of sign changes* are needed, however.

Construction of ICS Functionals

Tyler, Critchley, Dümbgen, and Oja (2009) provide an approach for construction of ICS functionals.

- Let $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ be two WC functionals with the eigenvalues of $\mathbf{V}_1(F)^{-1}\mathbf{V}_2(F)$ all distinct. Then the matrix $\mathbf{D}(F)$ of corresponding eigenvectors is an ICS functional.
- Extension for multiplicities among eigenvalues is given.
- Various choices of $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ are considered.

Strong ICS Functionals

Definition. An ICS functional $\mathbf{D}(F)$ has *Structure A* if, for $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$ with any nonsingular $\mathbf{A}$ and any $\mathbf{b}$,

$$\boxed{\mathbf{D}(F_{\mathbf{Y}}) = k_3\,\mathbf{J}\,\mathbf{D}(F_{\mathbf{X}})\,\mathbf{A}^{-1}}$$

with $k_3 = k_3(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ a positive scalar function and $\mathbf{J} = \mathbf{J}(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ a *sign change matrix* (diagonal with ± 1).

Definition. A *strong ICS functional* is satisfying Structure A with $\mathbf{J} = \mathbf{I}_d$.

For a *strong* ICS functional, both scale changes and sign changes are *homogeneous*.

Connection between *ICS* and *TR* Functionals

Theorem. Every ICS functional $\mathbf{D}(F)$ with *Structure A* is a TR functional (and hence $(\mathbf{D}(F)'\mathbf{D}(F))^{-1}$ is a WC functional).

Examples of TR and Strong ICS Functionals

- The TR functional $\mathbf{M}_0$ mentioned earlier with $\mathbf{Q}_S(\cdot, \mathbb{X}_N)$ is a strong ICS functional with $\mathbf{J} = \mathbf{I}_d$. It is constructed by a method quite different from the above.
- The well-known and oft-used Tyler (1987) scatter functional is also a TR functional but not a strong ICS functional.
- A symmetrized version of the Tyler functional given by Dümbgen (1998) does not involve a location functional and also is a TR functional.

Key Property of Strong ICS Functionals

- A *strong* ICS functional $\mathbf{D}(F)$ satisfies, for $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$,

$$\boxed{\mathbf{D}(F_{\mathbf{Y}})\mathbf{Y} = k_3\,\mathbf{D}(F_{\mathbf{X}})\,\mathbf{X} + \mathbf{c}}$$

with $\mathbf{c} = \mathbf{c}(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = k_3\,\mathbf{D}(F_{\mathbf{X}})\mathbf{A}^{-1}\mathbf{b}$, a *constant*.

- Thus the new $\mathbf{D}(\cdot)$-standardized coordinates $\mathbf{D}(F_{\mathbf{Y}})\mathbf{Y}$ agree with the original $\mathbf{D}(\cdot)$-standardized coordinates $\mathbf{D}(F_{\mathbf{X}})\mathbf{X}$, *except* for *a homogeneous scale change* and *a translation*.

Practical Interpretation in the *Strong* ICS Case

With $\mathbf{D}(\cdot)$ a *strong* ICS functional,

$$\mathbf{D}(\mathbb{X}_n)\mathbb{X}_n$$

remains unaltered after *affine transformation* to $\mathbb{Y}_n = \mathbf{A}\mathbb{X}_n + \mathbf{b}$ followed by $\mathbf{D}(\cdot)$-*standardization* to

$$\mathbf{D}(\mathbb{Y}_n)\mathbb{Y}_n,$$

except for *a homogeneous scale change* and *a translation*.

Affine Invariance via Strong ICS Functionals

Theorem. Let $T(\mathbf{x}, F)$ be a *real-valued* functional of $\mathbf{x}$ and F that is *invariant* under homogeneous scale change and translation of $\mathbf{x}$, in the sense that

$$T(c\mathbf{x} + \mathbf{b}, F_{c\mathbf{X} + \mathbf{b}}) = T(\mathbf{x}, F_{\mathbf{X}})$$

for any scalar c and any vector $\mathbf{b}$. Let $\mathbf{D}(F)$ be a *strong ICS functional*. Then the functional

$$T(\mathbf{D}(F_{\mathbf{X}})\mathbf{x}, F_{\mathbf{D}(F_{\mathbf{X}})\mathbf{X}})$$

is *affine invariant*.

Example Needing the Theorem

Scaled-deviation outlyingness for a single projection $\mathbf{u}_0$.

$$T(\mathbf{x}, F_{\mathbf{X}}) = \left| \frac{\mathbf{u}_0'\mathbf{x} - \mu(F_{\mathbf{u}_0'\mathbf{X}})}{\sigma(F_{\mathbf{u}_0'\mathbf{X}})} \right|.$$

With Multivariate Data, *First Standardize*!

- *Example*. To make the *spatial quantile function* affine equivariant, first standardize with a *TR functional*.
  - Choosing a *strong ICS functional* yields added benefits.
- *Example*. In *projection pursuit* methods with multivariate data, univariate standardization after projection does not in general produce desired affine invariance. Standardize first, using a *strong ICS functional*.
  - E.g., to formulate *outlyingness* of $\mathbf{x}$ in $\mathbb{R}^d$ by a quadratic from based on univariate scaled deviations of projections of $\mathbf{x}$ on just finitely many directions $\{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$, use $\mathbf{u}_i'\mathbf{D}(F_\mathbf{X})\mathbf{x}$ instead of $\mathbf{u}_i'\mathbf{x}$, $1 \leq i \leq s$, etc.

Acknowledgments

Stimulating discussions with Marc Hallin and Davy Paindaveine have led to improvements in the presentation. Thanks are due Satyaki Mazumder for finding an error in an earlier version.

Also, support by NSF grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is greatly appreciated.