Invariance and Equivariance: Benefits, Costs, and Methods

Robert Serfling[1]

University of Texas at Dallas

May 11, 2011

For the
Fourth Erich L. Lehmann Symposium
Rice University

[1] www.utdallas.edu/∼serfling

A Leading Question

In what ways should estimation and test procedures, or perceived geometric features and structures in a data set, desirably transform when the data undergo transformation to another coordinate system?

Popular Points of View

- "Two estimators of a parameter which agree for given data in one coordinate system should continue to agree after transformation to another coordinate system."
- "A test procedure which accepts or rejects a null hypothesis on the basis of given data should make the same decision about the equivalent null hypothesis after transformation to other coordinates."
- "p-values and other interpretations of the data as evidence for or against the null hypothesis should not change after transformation to other coordinates."

- “In general, a statistical decision procedure should be independent of the particular coordinate system of the data.”
- “If an inference problem exhibits *symmetry* with respect to some group of transformations, then one should restrict to decision procedures which likewise exhibit the given symmetry.”

- "A method of ranking of points in a data cloud by centrality or outlyingness should give the same ranking after transformation to other coordinates."
- "Points branded as *outliers* in one coordinate system should remain so under such transformation to other coordinates."
- "The interpretation of a point as a *quantile* relative to a given probability distribution should carry over to its image under transformation to other coordinates."

- "Striking geometric features or structures perceived in a data set or found by data mining should be invariant under transformation of coordinates or else ignored as mere artifacts of the given coordinate system."

Key Technical Concepts

Such requirements, *which we neither endorse nor reject*, are fulfilled when, for example,

- *test statistics* and *outlyingness functions* are invariant,
- *estimators* and *quantile functions* are equivariant, and/or
- *preprocessing of the data* is carried out using an invariant coordinate system transformation.

More specifically, for a data set $\mathbb{X}_n = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ of observations in $\mathbb{R}^d$, and for affine transformations $\mathbf{X} \mapsto \mathbf{Ax} + \mathbf{b}$ with nonsingular $d \times d$ matrix $\mathbf{A}$ and d-vector $\mathbf{b}$,

- *Location estimators* $\mathbf{L}(\mathbb{X}_n)$ should satisfy

$$\mathbf{L}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{L}(\mathbb{X}_n) + \mathbf{b}.$$

More specifically, for a data set $\mathbb{X}_n = \{\mathbf{X}_1, \ldots, \mathbf{X}_n\}$ of observations in $\mathbb{R}^d$, and for affine transformations $\mathbf{X} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ with nonsingular $d \times d$ matrix $\mathbf{A}$ and d-vector $\mathbf{b}$,

- *Location estimators* $\mathbf{L}(\mathbb{X}_n)$ should satisfy

$$\mathbf{L}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{L}(\mathbb{X}_n) + \mathbf{b}.$$

- *Dispersion estimators* $\mathbf{D}(\mathbb{X}_n)$ should satisfy a version of

$$\mathbf{D}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{D}(\mathbb{X}_n)\mathbf{A}'.$$

- *Multivariate quantile functions* $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, $\mathbf{u} \in \mathbb{B}^d(\mathbf{0})$ (the unit ball in $\mathbb{R}^d$), should satisfy some version of

$$\mathbf{Q}(\boldsymbol{\nu}, \mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) + \mathbf{b},$$

for a suitable $\mathbb{B}^d(\mathbf{0})$-valued reindexing $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{u}, \mathbf{A}, \mathbf{b}, \mathbb{X}_n)$.

- *Multivariate quantile functions* $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, $\mathbf{u} \in \mathbb{B}^d(\mathbf{0})$ (the unit ball in $\mathbb{R}^d$), should satisfy some version of

$$\mathbf{Q}(\boldsymbol{\nu}, \mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) + \mathbf{b},$$

for a suitable $\mathbb{B}^d(\mathbf{0})$-valued reindexing $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{u}, \mathbf{A}, \mathbf{b}, \mathbb{X}_n)$.

In particular, the *median* $\mathbf{Q}(\mathbf{0}, \mathbb{X}_n)$ should satisfy

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

- *Multivariate quantile functions* $\mathbf{Q}(\mathbf{u}, \mathbb{X}_n)$, $\mathbf{u} \in \mathbb{B}^d(\mathbf{0})$ (the unit ball in $\mathbb{R}^d$), should satisfy some version of

$$\mathbf{Q}(\boldsymbol{\nu}, \mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{Q}(\mathbf{u}, \mathbb{X}_n) + \mathbf{b},$$

for a suitable $\mathbb{B}^d(\mathbf{0})$-valued reindexing $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{u}, \mathbf{A}, \mathbf{b}, \mathbb{X}_n)$.

In particular, the *median* $\mathbf{Q}(\mathbf{0}, \mathbb{X}_n)$ should satisfy

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

- *Multivariate outlyingness functions* $O(\mathbf{x}, \mathbb{X}_n)$, $\mathbf{x} \in \mathbb{R}^d$, should satisfy

$$O(\mathbf{A}\mathbf{x} + \mathbf{b}, \mathbf{A}\mathbb{X}_n + \mathbf{b}) = O(\mathbf{x}, \mathbb{X}_n),$$

at least up to a multiplicative constant.

- We want matrix-valued *invariant coordinate system (ICS) transformations* $\mathbf{M}(\mathbb{X}_n)$ such that the data $\mathbb{X}_n$ after transformation to new coordinates $\boxed{\mathbf{M}(\mathbb{X}_n)\mathbb{X}_n}$ agrees with affine counterparts $\boxed{\mathbf{M}(\mathbf{A}\mathbb{X}_n + \mathbf{b})(\mathbf{A}\mathbb{X}_n + \mathbf{b})}$ up to homogeneous scale changes and translations.

- We want matrix-valued *invariant coordinate system (ICS) transformations* $\mathbf{M}(\mathbb{X}_n)$ such that the data $\mathbb{X}_n$ after transformation to new coordinates $\boxed{\mathbf{M}(\mathbb{X}_n)\mathbb{X}_n}$ agrees with affine counterparts $\boxed{\mathbf{M}(\mathbf{A}\mathbb{X}_n + \mathbf{b})(\mathbf{A}\mathbb{X}_n + \mathbf{b})}$ up to homogeneous scale changes and translations.

  That is, $\boxed{\mathbf{M}(\mathbb{X}_n)\mathbb{X}_n}$ captures the *affine invariant* geometric structures, artifacts, and patterns inherent in the original data set $\mathbb{X}_n$.

Goals of This Talk

Invariance ($\mathcal{I}$) and *Equivariance* ($\mathcal{E}$) have intuitive appeal and a certain force of logic.

Practical implementation, however, requires a formal development and a broad perspective.

In the setting of multivariate data in $\mathbb{R}^d$, and with special focus on *matrix transformations* of the data, let us examine

- formulations of $\mathcal{I}$ and $\mathcal{E}$ for various purposes,
- costs of $\mathcal{I}$ and $\mathcal{E}$ as trade-offs against efficiency, robustness, and computational ease,
- methods for acquiring $\mathcal{I}$ and $\mathcal{E}$ via suitable transformations of the data,
- technical issues with approaches to $\mathcal{I}$ and $\mathcal{E}$.

Invariance, Groups, and Symmetry [Lehmann, 1959]

- "The mathematical expression of symmetry is invariance under a suitable group of transformations."

Invariance, Groups, and Symmetry [Lehmann, 1959]

- "The mathematical expression of symmetry is invariance under a suitable group of transformations."
- General Setting (not just $\mathbb{R}^d$):
  - Arbitrary sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$
  - Orbits $\{g(x), g \in \mathcal{G}\}$, $x \in \mathcal{X}$, partition $\mathcal{X}$ into equivalence classes of x, x' related by $x = g(x')$, some g.

Invariance, Groups, and Symmetry [Lehmann, 1959]

- "The mathematical expression of symmetry is invariance under a suitable group of transformations."
- General Setting (not just $\mathbb{R}^d$):
  - Arbitrary sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$
  - Orbits $\{g(x), g \in \mathcal{G}\}$, $x \in \mathcal{X}$, partition $\mathcal{X}$ into equivalence classes of x, x' related by $x = g(x')$, some g.
- A function $T(x)$ is *invariant* if <u>constant</u> on any $\mathcal{G}$-orbit $\mathcal{O}$:

$$T(x) = \text{constant}, \ \ x \in \mathcal{O}.$$

Invariance, Groups, and Symmetry [Lehmann, 1959]

- "The mathematical expression of symmetry is invariance under a suitable group of transformations."
- General Setting (not just $\mathbb{R}^d$):
  - Arbitrary sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$
  - Orbits $\{g(x), g \in \mathcal{G}\}$, $x \in \mathcal{X}$, partition $\mathcal{X}$ into equivalence classes of x, x' related by $x = g(x')$, some g.
- A function $T(x)$ is *invariant* if constant on any $\mathcal{G}$-orbit $\mathcal{O}$:

$$T(x) = \text{constant}, \;\; x \in \mathcal{O}.$$

- A *maximal invariant* function $T_0(x)$ labels the orbits: if $T_0(x) = T_0(x')$, then x and x' belong to the same orbit. Then each invariant $T = h \circ T_0$ for some h

- Extended Setting:
  - Sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$

- Extended Setting:
  - Sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$
  - Family $\{P_\theta, \theta \in \Theta\}$ of distinct distributions on $\mathcal{A}$
  - Induced group $\overline{\mathcal{G}}$ on Θ: for $g \in \mathcal{G}$ define $\overline{g} : \theta \mapsto \overline{g}\theta$ by $P_{\overline{g}\theta}(X \in A) = P_\theta(gX \in A)$, $\theta \in \Theta$.
  - Assumption making $\overline{g}$ 1:1 on Θ: $\overline{g}\Theta = \Theta$

- Extended Setting:
  - Sample space $\mathcal{X}$ and measurable subsets $\mathcal{A}$
  - Group $\mathcal{G}$ of 1:1 transformations of $\mathcal{X}$, $g : x \mapsto gx$, $x \in \mathcal{X}$, such that $g\mathcal{A} = \mathcal{A}$
  - Family $\{P_\theta, \theta \in \Theta\}$ of distinct distributions on $\mathcal{A}$
  - Induced group $\overline{\mathcal{G}}$ on Θ: for $g \in \mathcal{G}$ define $\overline{g} : \theta \mapsto \overline{g}\theta$ by $P_{\overline{g}\theta}(X \in A) = P_\theta(gX \in A)$, $\theta \in \Theta$.
  - Assumption making $\overline{g}$ 1:1 on Θ: $\overline{g}\Theta = \Theta$
  - Induced group $\mathcal{G}^*$ on the decision space $\mathcal{D}$: for $g \in \mathcal{G}$, define $g^* : d \mapsto g^*d$ such that $g \mapsto g^*$ is a homomorphism and loss L is unchanged: $L(\overline{g}\theta, g^*d) = L(\theta, d)$.

- *Invariance of statistical decision problem* $(\mathcal{X}, \{P_\theta, \theta \in \Theta\}, L)$:
  $g\mathcal{A} = \mathcal{A}$, $\overline{g}\Theta = \Theta$, $L(\overline{g}\theta, g^* d) = L(\theta, d)$.
- *Invariance of statistical decision procedure* δ:
  (A) $\quad \delta(gx) = g^* \delta(x)$.
- For an invariant *testing* problem, we want $g^* =$ the identity in (A), which then expresses *Invariance* ($\mathcal{I}$) of the test function δ.
- For an invariant *estimation* problem, we want $g^* \neq$ the identity, and then, in current terminology, (A) expresses *Equivariance* ($\mathcal{E}$) of the estimator δ.

Some Classical Examples of Invariant Tests

- $\mathcal{G} = \{\text{translations } gx = x + c,\ c \in \mathbb{R}\}$. A maximal invariant is $T_0(\mathbb{X}_n) = (X_1 - X_n, \ldots, X_{n-1} - X_n)$.

  The testing problem $H_0 : X \sim f_0(x - \theta)$ versus $H_1 : X \sim f_1(x - \theta)$ with θ unknown is invariant under $\mathcal{G}$ and the induced $\overline{\mathcal{G}}$.

  An invariant test is then a function of $T_0(\mathbb{X}_n)$, whose distribution does not depend on θ.

  The Neyman-Pearson Lemma yields a UMP invariant test, which turns out to be quite reasonable.

- $\mathcal{G} = \{\text{scale changes } gx = cx,\ c > 0\}$. A maximal invariant is $T_0(\mathbb{X}_n) = (X_1/X_n, \ldots, X_{n-1}/X_n)$.
- $\mathcal{G} = \{\text{linear transformations } gx = ax + b,\ a \neq 0\}$. A maximal invariant is $T_0(\mathbb{X}_n) = ((X_1 - X_n)/(X_{n-1} - X_n), \ldots, (X_{n-2} - X_n)/(X_{n-1} - X_n))$.
- $\mathcal{G} = \{\text{continuous and strictly increasing functions } g(x)\}$. A maximal invariant is $T_0(\mathbb{X}_n) =$ the vector of *ranks* of $X_1, \ldots, X_n$.

An Example for Data in $\mathbb{R}^2$

- The data consists of two bivariate observations, $\mathbf{X}_1 \sim N(\mathbf{0}, \mathbf{\Sigma})$ and $\mathbf{X}_2 \sim N(\mathbf{0}, \Delta\mathbf{\Sigma})$. With probability 1, $\mathcal{X}$ is the sample space of nonsingular 2×2 matrices.

  The problem of testing $\Delta = 1$ versus $\Delta > 1$ is invariant with respect to $\mathcal{G} = \{\text{g}: g\mathbf{X} = \mathbf{AX}, \text{ nonsingular } 2 \times 2 \ \mathbf{A}\}$.

  However, there is only one orbit and so the invariant and maximal invariant functions are the constant functions.

  Then the UMP invariant size α test is $\phi \equiv \alpha$, with power α. Yet a good noninvariant test can be developed with power function increasing in Δ.

  This shows that the best invariant procedure can be inadmissable, outperformed by a noninvariant one.

Equivariance versus Other Criteria

- Three desirable properties of a univariate location estimator $\theta(\mathbb{X}_n)$ are
  - *Equivariance*: $\theta(a\mathbb{X}_n + b) = a\theta(\mathbb{X}_n) + b$
  - *Monotonicity*: for all nonnegative $b_1, \ldots, b_n$,

$$\theta(X_1 + b_1, \ldots, X_n + b_n) \geq \theta(X_1, \ldots, X_n)$$

  - *50% breakdown point*: up to 50% of the sample may be corrupted without taking $\theta(\mathbb{X}_n)$ to ∞.

Only one statistic possesses all three of these properties: the sample median (Bassett, 1991). The median, however, gives up some efficiency.

This shows that restriction to equivariant procedures can overly compromise efficiency.

Example: Chaudhuri and Sengupta (1993)

- Chaudhuri and Sengupta (1993) test $H : \boldsymbol{\theta} = \mathbf{0}$ versus $H : \boldsymbol{\theta} \neq \mathbf{0}$ in the location model $F_{\mathbf{X}} = F_0(\mathbf{x} - \boldsymbol{\theta})$ in $\mathbb{R}^d$.
- Since $\mathbf{A}\boldsymbol{\theta} = \mathbf{0}$, all nonsingular $\mathbf{A}$, if and only if $\boldsymbol{\theta} = \mathbf{0}$, they suggest using a test function ϕ satisfying $\phi(\mathbf{A}\mathbf{x}) = \phi(\mathbf{x})$, all nonsingular $\mathbf{A}$, thus making the same decision before and after any nonsingular transformation of the coordinate system.
- This motivates choosing the test procedure to be some function of a maximal invariant statistic relative to the group of nonsingular transformations $\mathbf{A}$.

A Maximal Invariant for This Example

- Based on $\mathbb{X}_n$, and for each fixed choice of d distinct indices $\mathbb{J} = \{i_1, \ldots, i_d\}$ from $\{1, \ldots, n\}$, Chaudhuri and Sengupta define the matrix

$$\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n) = [\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_d}]_{d \times d}$$

and show that for $F_{\mathbf{X}}$ absolutely continuous the transformed observations (i.e, "data-driven coordinates")

$$\mathbb{Y}_n^{(\mathbb{J})} = \mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}\mathbb{X}_n$$

form a maximal invariant statistic with respect to the nonsingular transformations $\mathbf{A}$.

An Important Property of the Matrix $\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)$

- We observe here that, for each $\mathbb{J}$, the matrix $\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)$ satisfies the following structural property:

$$\mathbf{W}_{\mathbb{J}}(\mathbf{A}\mathbb{X}_n) = \mathbf{A}\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n), \qquad (1)$$

  for any $d \times d$ $\mathbf{A}$.

- This is a step in the proof of the maximal invariance of $\mathbb{Y}_n^{(\mathbb{J})}$ by Chaudhuri and Sengupta. However, they do not otherwise comment on this property, nor apply it directly, nor interpret it.

Why is This Property Important?

- Equivalently, putting $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n) = \mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}$, the property may be stated

$$\mathbf{M}_{\mathbb{J}}(\mathbf{A}\mathbb{X}_n) = \mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbf{A}^{-1},$$

for any $d \times d$ $\mathbf{A}$.

- It then follows that

$$\mathbf{M}_{\mathbb{J}}(\mathbf{A}\mathbb{X}_n)(\mathbf{A}\mathbb{X}_n) = \mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbf{A}^{-1}(\mathbf{A}\mathbb{X}_n) = \mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbb{X}_n$$

for any $d \times d$ $\mathbf{A}$.

- That is, these new data-driven coordinates $\boxed{\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbb{X}_n}$ represent an *affine invariant coordinate system*.

A Family of Sign Tests

- Let $\mathcal{C}(d, n)$ denote the class of all sets of d distinct integers from $\{1, \ldots, n\}$. It follows that the statistic

$$\xi_n = \{\mathbb{Y}_n^{(\mathbb{J})}, \mathbb{J} \subset \mathcal{C}(d, n)\}$$

  is also maximal invariant.
- It has the further desirable property of being invariant over permutations of the indices of the observations, i.e., ξ_n is symmetric in the observations, although this latter is obtained at the cost of considerable extra computation.
- Chaudhuri and Sengupta develop *affine invariant multivariate sign tests* in the elliptical location model, based on the multivariate signs of the variables in ξ_n.

Example: Chakraborty and Chaudhuri (1996)

- For *estimating* location rather than testing a specified value, Chakraborty and Chaudhuri (1996) introduce a variation of the Chaudhuri and Sengupta (1993) transformation, namely

$$\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n) = \left[(\mathbf{X}_{i_1} - \mathbf{X}_{i_{d+1}}), \ldots, (\mathbf{X}_{i_d} - \mathbf{X}_{i_{d+1}})\right]_{d \times d},$$

  with the index set $\mathbb{J} = \{i_1, \ldots, i_{d+1}\}$ in $\mathcal{C}(d+1, n)$, thus using $d+1$ sample observations.
- The "data-driven coordinates"

$$\mathbb{Z}_n^{(\mathbb{J})} = \mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}(\mathbb{X}_n - \mathbf{X}_{i_{d+1}})$$

  form a maximal invariant statistic with respect to invertible affine transformations $\mathbf{Ax} + \mathbf{b}$.

Key Structural Property

- Analogous to (1), we have an important structural property:

$$\mathbf{W}_{\mathbb{J}}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{A}\mathbf{W}_{\mathbb{J}}(\mathbb{X}_n). \qquad (2)$$

- With $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n) = \mathbf{W}_{\mathbb{J}}(\mathbb{X}_n)^{-1}$, this may be expressed

$$\mathbf{M}_{\mathbb{J}}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbf{A}^{-1}.$$

- By a simple argument as before, the data-driven coordinates $\boxed{\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)(\mathbb{X}_n - \mathbf{X}_{i_{d+1}})}$ represent an *affine invariant coordinate system*.
- The same is true of the simpler data-driven coordinates $\boxed{\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbb{X}_n}$

TR Coordinatewise Median

- Chakraborty and Chaudhuri use $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)$ to develop a fully affine equivariant version of the sample coordinatewise median, which initially is not affine equivariant.
- This "transformation-retransformation (TR)" coordinatewise median is obtained by computing the usual coordinate-wise median on the transformed observations $\{\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)\mathbf{X}_i, i \notin \mathbb{J}\}$, and then retransforming that result back to the original coordinates via the inverse $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)^{-1}$.
- The key step in the proof is an application of property (2).
- Later we will define "TR functionals" precisely.

Choice of $\mathbb{J}$

- Based on optimality considerations in elliptical models, Chakraborty and Chaudhuri select $\mathbb{J}$ to make the matrix $\mathbf{W}_{\mathbb{J}}'\widehat{\boldsymbol{\Sigma}}^{-1}\mathbf{W}_{\mathbb{J}}$ approximate a matrix of form $\lambda\mathbf{I}_d$, i.e., so as to make the coordinate system $\widehat{\boldsymbol{\Sigma}}^{-1/2}\mathbf{W}_{\mathbb{J}}$ as orthonormal as possible, with $\widehat{\boldsymbol{\Sigma}}$ a consistent (at least proportionally) estimator of the population scatter matrix, for example FAST-MCD.
- However, the computational burden includes more than getting $\widehat{\boldsymbol{\Sigma}}$. The continuing steps to find the "optimal" $\mathbb{J}$ by checking all combinations are of order $O(n^{d+1})$ and become prohibitive quickly as d increases.
- Affine equivariance can cost a lot computationally.

Further Illustrations Involving TR Transformations

- Using $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)$, Chakraborty, Chaudhuri, and Oja (1998) develop a fully affine equivariant TR sample spatial median.
- Again using $\mathbf{M}_{\mathbb{J}}(\mathbb{X}_n)$, Chakraborty (2001) extends to a fully affine equivariant TR version of the spatial quantile function $\mathbf{Q}_S(\mathbf{u}, \mathbb{X}_n)$ of Chaudhuri (1996).

- Randles (2000) develops an affine invariant, computationally easy, multivariate sign test using an affine invariant version of the spatial sign function based on transforming by the well-known Tyler (1987) scatter matrix using the location specified by the null hypothesis.
- Hettmansperger and Randles (2002) develop an affine equivariant, computationally easy, multivariate median based on transforming by the Tyler (1987) matrix as obtained by simultaneously solving equations for the matrix and an associated location.

- Serfling (2010) shows that the TR sample spatial quantile function based on any transformation matrix $\mathbf{M}(\mathbb{X}_n)$ which is the inverse square root of a covariance matrix suffices for affine equivariance.
  - In particular, this establishes that the special property

$$\mathbf{M}(\mathbf{A}\mathbb{X}_n + \mathbf{b}) = \mathbf{M}(\mathbb{X}_n)\mathbf{A}^{-1}$$

    is not needed for equivariance in this particular application.
  - Thus computationally attractive matrices such as the Tyler (1987) matrix may be used for the purpose of affine equivariant spatial quantiles.

Fast Dispersion Matrix Estimation

- A *fast dispersion matrix estimator* based on pairwise robust covariance estimation was first proposed by Gnanadesikan and Kettenring (1972) and later modified by Maronna and Zamar (2002) into an "orthogonalized Gnanadesikan-Kettenring estimate" (OGK).
- This estimator lacks affine equivariance. However, simulations of Maronna and Zamar (2002) show that
  - OGK performs similarly to Fast-MCD at lower computational cost.
  - Certain weighted versions are "more equivariant".
- See also Maronna, Martin, and Yohai (2006).
- Here equivariance is sacrificed for computational gain.

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 $\mathbf{b}$,

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

  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 known as a "*shape functional*".
  [Paindaveine, 2008; Tyler, Critchley, Dümbgen, and Oja, 2009]

Transformation-Retransformation (TR) Functionals

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

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

  with $k_2 = k_2(\mathbf{A}, \mathbf{b}, F_{\mathbf{X}})$ a positive scalar function.

  [Chakraborty and Chaudhuri, 1996; Randles, 2000]

- TR approaches modify *estimation (testing)* procedures to achieve (*hopefully*) full *affine equivariance (invariance)*.
  - Carry out the procedure on transformed data $\mathbf{M}(\mathbb{X}_n)\mathbb{X}_n$.
  - For equivariance, retransform to original coordinates via $\mathbf{M}(\mathbb{X}_n)^{-1}$. For invariance, do not retransform.
  - Verify that the equivariance (invariance) indeed holds.

Connection between TR and WC Functionals

- **Theorem**. *Every TR functional* $\mathbf{M}(F)$ *is equivalent to a WC functional, and conversely.*
  - Given a TR fcnl $\mathbf{M}(F)$, $\mathbf{C}(F) = (\mathbf{M}(F)'\mathbf{M}(F))^{-1}$ is WC.
  - Given a WC fcnl $\mathbf{C}(F)$, any solution $\mathbf{M}(F)$ of $\mathbf{C}(F) = (\mathbf{M}(F)'\mathbf{M}(F))^{-1}$ 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.

Solutions $\mathbf{M}(F)$ *of* $\mathbf{C}(F) = (\mathbf{M}(F)'\mathbf{M}(F))^{-1}$

- In particular, one may choose $\mathbf{M}(F)$ to be the symmetric square root of $\mathbf{C}(F)^{-1}$ or the unique upper triangular matrix in the Cholesky factorization with "1" in the uppermost diagonal cell. Thus the choice of $\mathbf{M}(F)$ is not unique.
- Also, besides these structurally differing cases, for any solution $\mathbf{M}(F)$ we have that additional solutions are given by $\mathbf{O}\,\mathbf{M}(F)$ for any orthogonal matrix $\mathbf{O}$.
- Other solutions, quite different structurally from the above, will be seen below.

Invariant Coordinate System (ICS) Functionals

Definition. A matrix-valued functional $\mathbf{D}(F)$ is an *invariant coordinate system (ICS) functional* if the $\mathbf{D}(\cdot)$-standardizion of $\mathbf{X}$

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

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

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

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

[Tyler, Critchley, Dümbgen, and Oja, 2009]

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

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

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

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

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

- Some applications, however, for example *outlyingness*, require *homogeneity* of scale changes and sign changes.

Strong ICS (SICS) 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 (SICS) functional* is an ICS functional of Structure A with $\mathbf{J} = \mathbf{I}_d$.

  [Serfling, 2010]

- For a *strong* ICS functional, only *homogeneous* scale changes and sign changes are involved.

Connection between ICS and TR Functionals

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

Key Property of SICS Functionals

- A *SICS* 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*.

- Likewise, for sample versions, $\mathbf{D}(\mathbb{X}_n)\mathbb{X}_n$ *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 possibly *a homogeneous scale change* and *a translation*.

Non-Examples of SICS Functionals

- The Tyler (1987) TR functional is not a SICS functional.
- A symmetrized version (DT) of the Tyler functional given by Dümbgen (1998) does not involve a location functional and also is a TR functional (but also not SICS).

Constructions Using Two WC Functionals

- Tyler, Critchley, Dümbgen, and Oja (2009) construct ICS functionals using two WC 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 of corresponding eigenvectors* is ICS.
- Various choices of $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ are considered.

Constructions Using Two WC Functionals

- Tyler, Critchley, Dümbgen, and Oja (2009) construct ICS functionals using two WC 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 of corresponding eigenvectors* is ICS.
- Various choices of $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ are considered.
- These ICS functionals are not in general SICS.

Constructions Using Two WC Functionals

- Tyler, Critchley, Dümbgen, and Oja (2009) construct ICS functionals using two WC 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 of corresponding eigenvectors* is ICS.
- Various choices of $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ are considered.
- These ICS functionals are not in general SICS.
- For $\mathbf{V}_1 = \mathbf{I}_d$ and $\mathbf{V}_2 = \mathbf{\Sigma}(F)$, this gives Principle Components Analysis (PCA).

Constructions Using Two WC Functionals

- Tyler, Critchley, Dümbgen, and Oja (2009) construct ICS functionals using two WC 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 of corresponding eigenvectors* is ICS.
- Various choices of $\mathbf{V}_1(F)$ and $\mathbf{V}_2(F)$ are considered.
- These ICS functionals are not in general SICS.
- For $\mathbf{V}_1 = \mathbf{I}_d$ and $\mathbf{V}_2 = \boldsymbol{\Sigma}(F)$, this gives Principle Components Analysis (PCA).
- For $\mathbf{V}_1 = \boldsymbol{\Sigma}(F)$ and $\mathbf{V}_2$ a matrix-valued kurtosis measure, this gives Indepependent Components Analysis (ICA).

- For $\mathbf{V}_1 = \mathbf{\Sigma}(F)$ and $\mathbf{V}_2$ given by various matrices $\mathbf{V}_2(\mathbf{X}, \mathbf{Y})$ based on the means and covariances of $\mathbf{X}|\mathbf{Y}$, this gives "supervised ICA" and includes
  - Sliced Inverse Regression (SIR),
  - Sliced Average Variance Regression (SAVE),
  - Principle Hessian Directions (PHD),

  for example.

Sample Versions Using Two WC Functionals

- Ilmonen, Nevalainen, and Oja (2010) show that, for F continuous, the sample versions of these constructions are SICS when the solutions are selected in a unique way.
- However, the population versions can be SICS only under some fairly severe restrictions on F (excluding elliptical cases, for example).
- See also Ilmonen, Oja, and Serfling (2011).

Direct Construction of Sample SICS Functionals

[Serfling, 2010, 2011]

- Construction:
  1. Let $\mathbb{Z}_N = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_N\}$ be a subset of $\mathbb{X}_n$ of size N obtained through some permutation-invariant procedure.
  2. Form $d+1$ means $\overline{\mathbf{Z}}_1, \ldots, \overline{\mathbf{Z}}_{d+1}$ based on blocks of size $m = \lfloor N/(d+1) \rfloor$ from $\mathbb{Z}_N$.
  3. Form the matrix

$$\mathbf{W}(\mathbb{X}_n) = \left[(\overline{\mathbf{Z}}_1 - \overline{\mathbf{Z}}_{d+1}) \ \cdots \ (\overline{\mathbf{Z}}_d - \overline{\mathbf{Z}}_{d+1})\right]_{d \times d}.$$

  4. Then a SICS functional is given by

$$\mathbf{D}(\mathbb{X}_n) = \mathbf{W}(\mathbb{X}_n)^{-1}.$$

- A special case of the preceding is the functional $\mathbf{M}(\mathbb{X}_n)$ of Chakraborty and Chaudhuri (1996) based on a $\mathbb{Z}_N$ of size $N = d + 1$ derived by extensive computation.
- Alternatively, Mazumder and Serfling (2010a) take for $\mathbb{Z}_N$ the set of observations selected and used in computing $\mathbf{\Sigma}_{\mathrm{MCD}}$ with, say, $N \approx 0.75n$. This uses all of the data in selecting $\mathbb{Z}_N$ and all of its observations in defining $\mathbf{W}(\mathbb{X}_n)$. Little computation beyond that for $\mathbf{\Sigma}_{\mathrm{MCD}}$ is needed, but the latter becomes computationally prohibitive for higher d.
- Another approach of Mazumder and Serfling (2010a) is to compute the sample TR spatial outlyingness with $\mathbf{M}(F)$ the well-known TR functional of Tyler (1987), which is moderately robust and can be computed quickly in any dimension, and let $\mathbb{Z}_N$ be the 75% least outlying points.

Results for SICS Transformations

- *A SICS functional* $\mathbf{D}(F)$ *is neither symmetric nor triangular*.

  This compares with more typical types of TR functional as some choice of square root of the inverse of the associated WC functional $\mathbf{C}(F) = (\mathbf{D}(F)'\mathbf{D}(F))^{-1}$. Popular choices of such square roots are either symmetric or triangular.

  However, if a TR functional $\mathbf{M}(F)$ is *symmetric or triangular and also SICS*, then $\mathbf{M}(F)\mathbf{A}^{-1}$ must also be symmetric or triangular for arbitrary $\mathbf{A}$. It is easy to find counterexamples to this possibility.

- *Using two SICS functionals successively is equivalent to just using the most recent one in the first place*.

- *Using two SICS functionals successively is equivalent to just using the most recent one in the first place.*
- *If* $\mathbf{D}(F)$ *is SICS, then so is* $c\mathbf{D}(F)$*, for any constant* c.

- *Using two SICS functionals successively is equivalent to just using the most recent one in the first place*.
- *If* $\mathbf{D}(F)$ *is SICS, then so is* $c\mathbf{D}(F)$, *for any constant* c.
- *If* $\mathbf{X}$ *and* $\mathbf{Y}$ *are affinely equivalent in distribution, i.e.,* $\mathbf{Y} \overset{d}{=} \mathbf{A}\mathbf{X} + \mathbf{b}$, *then* $\mathbf{D}(F_{\mathbf{X}})$ *and* $\mathbf{D}(F_{\mathbf{Y}})$ *are proportional*.

- *Let* $\boldsymbol{\theta}(\mathbb{X}_n)$ *be a translation invariant d-vector. If* $\mathbf{D}(\mathbb{X}_n)$ *is SICS with proportionality constant* $k_1 = k_1(\mathbf{A}, \mathbf{b})$ *not depending on* $\mathbb{X}_n$*, then also SICS is*

$$\widetilde{\mathbf{D}}(\mathbb{X}_n) = \mathbf{D}(\widetilde{\mathbb{X}}_n),$$

where $\widetilde{\mathbb{X}}_n = \{\widetilde{\mathbf{X}}_i,\ 1 \leq i \leq n\}$*, with*

$$\widetilde{\mathbf{X}}_i = \|\mathbf{D}(\mathbb{X}_n)(\mathbf{X}_i - \theta(\mathbb{X}_n))\|^{\alpha}(\mathbf{X}_i - \boldsymbol{\theta}(\mathbb{X}_n)),\ 1 \leq i \leq n,$$

for any constant α*,* $-\infty < \alpha < \infty$.

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

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

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

  $$\mathbf{D}(F_{\mathbf{X}})^{-1}\mathbf{T}(\mathbf{u}, F_{\mathbf{D}(F_{\mathbf{X}})\mathbf{X}})$$

  is *affine equivariant*.

Examples Using the Theorem

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

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

Examples Using the Theorem

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

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

- *Spatial quantile functional.*

$$\mathbf{T}(\mathbf{u}, F) = \mathbf{Q}_S(\mathbf{u}, F).$$

Application: Projection Pursuit Outlyingness with Finitely Many Directions

- A *projection pursuit* outlyingness approach defines outlyingness of $\mathbf{x}$ in $\mathbb{R}^d$ as a function of the quantities

$$\{O(\mathbf{u}'\mathbf{x}, F_{\mathbf{u}'\mathbf{x}}), \mathbf{u} \in \Delta\}, \tag{3}$$

for some set Δ of unit vectors $\mathbf{u}$ in $\mathbb{R}^d$, and using the univariate scaled deviation outlyingness

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

- For Δ the set of *all directions*, and taking the *supremum* over (3), this yields $O_{\mathrm{P}}(\mathbf{x}, \mathbb{X}_n)$, which is affine invariant but computationally intensive (e.g., see Zuo, 2003).

- Alternatively, we may consider *finite* $\Delta = \{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$.
- However, with Δ finite, and using the supremum or a quadratic form, *not even orthogonal invariance holds*.
- On the other hand, Peña and Prieto (2001) introduce an *affine invariant* method using the supremum and $2d$ *data-driven* directions.
  - These are selected using univariate measures of kurtosis over candidate directions, choosing the d with local extremes of high kurtosis and the d with local extremes of low kurtosis.
  - Ultimately, in their very complex algorithm, the "outliers" are selected using Mahalanobis distance, thus yielding *ellipsoidal contours*.

- Filzmoser, Maronna, and Werner (2008) incorporate the Peña and Prieto (2001) approach into an even more elaborate one, using also a principal components step, that achieves certain improvements in performance for detection of location outliers, especially in high dimension.
  - However, *this gives up affine invariance* (although a SICS pre-standardization might regain it).
  - See also Maronna, Martin, and Yohai (2006).

- The use of finitely many *deterministic* directions strongly appeals on computational grounds, and it is desirable to take directions *approximately uniformly scattered* on the d-dimensional unit sphere.
- Fang and Wang (1994) provide convenient numerical algorithms for this purpose.

- Using finitely many deterministic directions approximately uniformly scattered, Pan, Fung, and Fang (2000) develop a finite Δ approach calculating a sample quadratic form based on the differences

$$\{O(\mathbf{u}'\mathbf{x}, \mathbf{u}'\mathbb{X}_n) - O(\mathbf{u}'\mathbf{x}, F_{\mathbf{u}'\mathbf{x}}),\ \mathbf{u} \in \Delta\}.$$

  - Since these differences involve the unknown F, a bootstrap step is incorporated.
  - The number of directions is data-driven.
  - The method is *not affine invariant* (although a SICS pre-standardization could correct for this).

A New SICS-Based Approach

- After first standardizing the data with a SICS $\mathbf{D}(F)$, the modified outlyingness function RMSP defined by

$$\widetilde{O}_{\Delta}(\mathbf{x}, F) = \sup_{\mathbf{u} \in \Delta} O(\mathbf{u}'\mathbf{D}(F_{\mathbf{X}})\mathbf{x}, F_{\mathbf{u}'\mathbf{D}(F_{\mathbf{X}})\mathbf{X}})$$

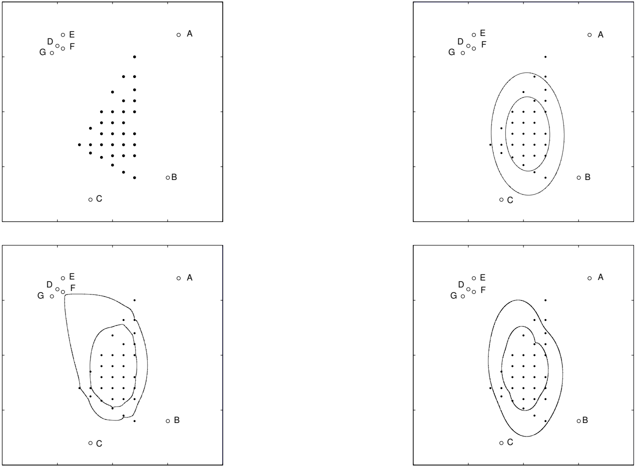
  is now affine invariant for any choice of finite Δ.
- See Serfling (2010) and Mazumder and Serfling (2010b).

Relevant Steps for RMSP:

1. Perform SICS pre-standardization with robust $\mathbf{D}(\mathbb{X}_n)$ developed using observations indexed by $\mathbb{J}_{\text{DT}}$.
2. Choose Δ of size $s = 5d$ uniformly scattered on the d-sphere (e.g., Fang and Wang, 1994).
3. Form s-vector $\boldsymbol{\eta}$ of scaled deviations for directions in Δ.
4. Obtain DT scatter matrix for $\mathbb{J}_{\text{DT}}$-indexed $\boldsymbol{\eta}$ vectors.
5. Apply the Robust Mahalanobis Spatial (RMS) approach to the $\boldsymbol{\eta}$ vectors $(i = 1, \ldots, n)$ instead of data vectors, using above DT standardization. This yields RMSP.

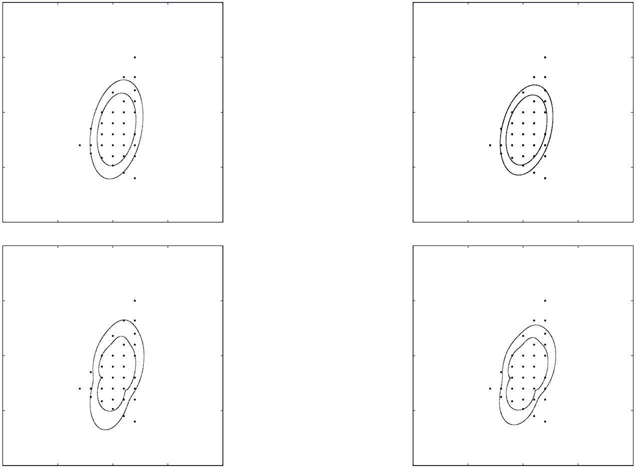
Comments:

- *Affine invariant, due to the SICS transformation*.
- *Robust, due to robustness of the scaled deviations and the $\mathbb{J}_{DT}$-based steps*.
- *Non-ellipsoidal contours in the data space*.



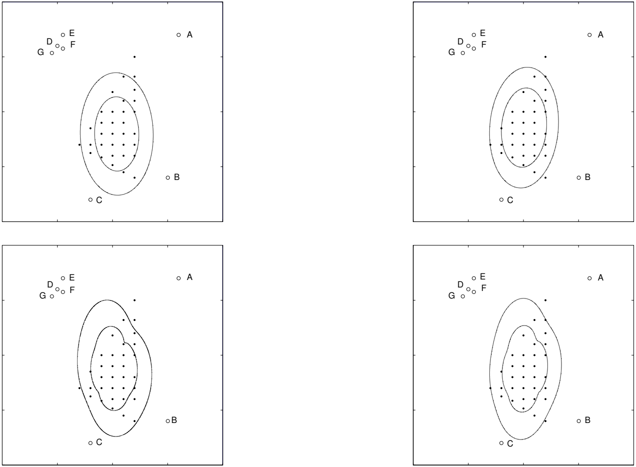
Upper plots: data and MD. Lower plots: MS and RMS.

RMS has robustness comparable to MD.



Upper plots: MD and MDP. Lower plots: RMS and RMSP.

Very similar performance.



Upper plots: MD and MDP. Lower plots: RMS and RMSP.

Again, very similar performance.

An Open Issue with SICS Transformations

- Sample SICS matrices are quite straightforward to construct, as we have seen.
- However, the corresponding population versions are not so straightforward.

- For the versions based on two WC matrices, the population versions are defined only under fairly severe restrictions excluding elliptical distributions.
- However, this is acceptable in ICA modeling.

- For the direct constructions starting with $\mathbf{W}(\mathbb{X}_n)$ based on differences of sample means, and defining $\mathbf{D}(\mathbb{X}_n)$ as $\mathbf{W}(\mathbb{X}_n)^{-1}$, it is tempting to define the corresponding population SICS matrix simply by $\mathbf{D} = (E\{\mathbf{W}\})^{-1}$.
  - However, $E\{\mathbf{W}\}$ is a matrix of zeros.
- Another approach: define

$$\begin{aligned} \mathbf{D}(F) &= E\{\mathbf{W}(\mathbb{X}_{n_0})^{-1}\} \\ \mathbf{W}(F) &= \mathbf{M}(F)^{-1} = (E\{\mathbf{W}(\mathbb{X}_{n_0})^{-1}\})^{-1}, \end{aligned}$$

for some suitable conceptual sample size n_0.
  - This is the analogue of defining the parameters $\theta = E(1/W)$ and $\eta = 1/\theta$ for a univariate random variable W having mean 0.

- We desire better linkage between corresponding sample and population SICS functionals.
- This requires better understanding of the behavior of sample SICS functionals.

Acknowledgments

- The speaker thanks *G. L. Thompson*, Marc Hallin, Pauliina Ilmonen, Hannu Oja, Davy Paindaveine, Ron Randles, and many others including anonymous commentators, for their thoughtful, stimulating, and helpful remarks.
- Support by NSF grants DMS-0103698 and DMS-0805786 and NSA Grant H98230-08-1-0106 is greatly appreciated.

References

Bassett, Jr., G. W. (1991). Equivariant, monotonic, 50% breakdown estimators. *The American Statistician* **45** 135–137.

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Annals of the Institute of Statistical Mathematics* **53** 380–403.

Chakraborty, B., and Chaudhuri, P. (1996). On a transformation and re-transformation technique for constructing an affine equivariant multivariate median. *Proceedings of the American Mathematical Society* **124** 2539–2547.

Chakraborty, B., Chaudhuri, P., and Oja, H. (1998). Operating transformation and retransformation on spatial median and angle test. *Statistica Sinica* **8** 767–784.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association* **91** 862–872.

Chaudhuri, P. and Sengupta, D. (1993). Sign tests in multidimension: Inference based on the geometry of the data cloud. *Journal of the American Statistical Association* **88** 1363–1370.

Dümbgen, L. (1998). On Tyler's M-functional of scatter in high dimension. *Annals of the Institute of Statistical Mathematics* **50** 471–491.

Fang, K. T, and Wang, Y. (1994). *Number Theoretic Methods in Statistics*. Chapman and Hall, London.

Filzmoser, P., Maronna, R., and Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics & Data Analysis* **52** 1694–1711.

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28** 81–124.

Hettmansperger, T. P. and Randles, R. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

Ilmonen, P., Nevalainen, J., and Oja, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. Preprint (2010).

Ilmonen, P., Oja, H., and Serfling, R. (2011). On invariant coordinate system (ICS) functionals. Preprint.

Lehmann, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

Lopuhaä, H. P. and Rousseeuw, J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Annals of Statistics* **19** 229-248.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, Chichester, England.

Maronna, R. A. and Zamar, R. H. (2002). Robust estimation of location and dispersion for high-dimensional data sets. *Technometrics* **44** 307–317.

Mazumder, S. and Serfling, R. (2010a). Spatial trimming, with applications to robustify sample spatial quantile and outlyingness functions, and to construct a new robust scatter estimator. In preparation.

Mazumder, S. and Serfling, R. (2010b). Robust multivariate outlyingness functions based on Mahalanobis standardization and projected scaled deviations. In preparation.

Paindaveine, D. (2008). A canonical definition of shape. *Statistics and Probability Letters* **78** 2240–2247.

Pan, J.-X., Fung, W.-K., and Fang, K.-T. (2000). Multiple outlier detection in multivariate data using projection pursuit techniques. *Journal of Statistical Planning and Inference* **83** 153–167.

Peña, D. and Prieto, F. J. (2001). Robust covariance matrix estimation and multivariate outlier rejection. *Technometrics* **43** 286–310.

Randles, R. H. (2000). A simpler, affine-invariant, multivariate, distribution-free sign test. *Journal of the American Statistical Assocation* **95** 1263–1268.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardization. *Journal of Nonparametric Statistics* **22** 915–936.

Serfling, R. (2011). On strong invariant coordinate system (SICS) functionals. Working paper.

Tyler, D. E. (1987). A distribution-free M-estimator of multivariate scatter. *Annals of Statistics* **15** 234–251.

Tyler, D. E., Critchley, F., Dümbgen, L. and Oja, H. (2009). Invariant co-ordinate selection. *Journal of the Royal Statistical Society, Series B* **71** 1–27.

Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.