Depth, Outlyingness, Quantile, and Rank Functions

Robert Serfling[1]

University of Texas at Dallas

For Mini-Workshop:
Level Sets and Depth Contours in High-Dimensional Data

Mathematisches Forschungsinstitut Oberwolfach
February 28, 2011

[1] www.utdallas.edu/∼serfling

Goal of the Talk

This talk aims to provide an overview of the landscape of *depth functions*, broadly considered,

- As a general framework for the various talks at this Workshop on specific "depth function" topics,
- As a general orientation for the "level sets" community at this Workshop, and
- As a prelude to general discussion of a "level sets – depth functions synergy" at this Workshop.

Depth, Outlyingness, Quantile, and Rank Functions in $\mathbb{R}^d$

Desired Equivariance and Invariance Properties

Further Desired Properties of DOQR Functions

Practical Applications of DOQR Functions

Computation of DOQR Functions

Convergence of Sample DOQR Functions

DOQR Functions in Other Settings

Concluding Remarks

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.
- This becomes clear when we make *the right definitions*.

Outlyingness Functions: Finding "Outliers" in $\mathbb{R}^d$

- "*Outliers*" have been under discussion 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.

Example 1: Projection Pursuit Approach

- Extension of a *univariate* outlyingness function $O(x, F)$ to $\mathbb{R}^d$ by *projection pursuit* yields "*projection outlyingness*"

$$O_P(\mathbf{x}, F) = \sup_{\|\mathbf{u}\| = 1} O(\mathbf{u}'\mathbf{x}, F_{\mathbf{u}'\mathbf{x}}), \ \mathbf{x} \in \mathbb{R}^d$$

- With *univariate "scaled deviation outlyingness"*

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

and, e.g., $(\mu(F), \sigma(F)) = (\text{Med}(F), \text{MAD}(F))$,

  - $O_P(\mathbf{x}, F)$ is *affine invariant*.
  - $O_P(\mathbf{x}, \mathbb{X}_n)$ is *highly masking robust*. [Dang and Serfling, 2010]

- But $O_P(\mathbf{x}, \mathbb{X}_n)$ is *computationally intensive*.

Example 2: Mahalanobis Distance Approach

- Extension of *univariate scaled deviation* $O(\cdot,\cdot)$ by *substitution*, with *multivariate* location and dispersion measures $\boldsymbol{\mu}(F)$ and $\boldsymbol{\Sigma}(F)$ and Euclidean distance $\|\cdot\|$, yields the popular "*Mahalanobis distance outlyingness*"

$$O_{\mathrm{MD}}(\mathbf{x}, F) = \|\boldsymbol{\Sigma}(F)^{-1/2}(\mathbf{x} - \boldsymbol{\mu}(F))\|, \ \mathbf{x} \in \mathbb{R}^d$$

- It is *weakly affine invariant*.
- For suitable choice of $\boldsymbol{\mu}(\mathbb{X}_n)$ and $\boldsymbol{\Sigma}(\mathbb{X}_n)$, $O_{\mathrm{MD}}(\mathbf{x}, \mathbb{X}_n)$ is highly masking robust. [Becker and Gather, 1999]
- However, the contours of $O_{\mathrm{MD}}(\mathbf{x}, F)$ are *ellipsoidal*, even when the density contours of F are not.
- Nevertheless, $O_{\mathrm{MD}}(\mathbf{x}, F)$ is versatile and popular.

Example 3: Spatial Sign Function Approach

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

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

we define the "*spatial outlyingness function*"

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

- Univariate case: $O(x, F) = |ES(x - X)| = |2F(x) - 1|$.
- $O_{\mathrm{S}}(\mathbf{x}, F)$ is *orthogonally invariant*.
- $O_{\mathrm{S}}(\mathbf{x}, \mathbb{X}_n)$ is moderately masking robust. [Dang/Serfling, 2010]

Example 4: "Mahalanobis Spatial" Approach

- We first standardize, using any scatter functional $\mathbf{\Sigma}(F)$, and then apply the spatial outlyingness function, forming

$$O_{\mathrm{MS}}(\mathbf{x}, F_{\mathbf{X}}) = O_{\mathrm{S}}(\mathbf{\Sigma}(F_{\mathbf{X}})^{-1/2}\mathbf{x}, F_{\mathbf{\Sigma}(F_{\mathbf{X}})^{-1/2}\mathbf{X}})$$
$$= \left\| E\,\mathbf{S}(\mathbf{\Sigma}(F_{\mathbf{X}})^{-1/2}(\mathbf{x} - \mathbf{X}))\right\|, \ \mathbf{x} \in \mathbb{R}^d$$

- In homage to Mahalanobis, who promoted the role of *standardization* in multivariate analysis, we call this the "*Mahalanobis spatial outlyingness function*".
- It is *fully affine invariant*.
- It is moderately masking robust.
- Its contours *need not be ellipsoidal*.

Depth Functions: Define Order Statistics in $\mathbb{R}^d$?

Depth function: a center-outward ordering $D(\mathbf{x}, F)$ 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" $\mathbf{M}_F$.
- *Order data points* by their *sample depths*,
- Define *analogues of univariate statistics involving order statistics*.

Further Aspects

- Depth functions yield *nested contours*.
- Depth functions *ignore multimodality*.
- General treatments: Liu, Parelius, and Singh (1999), Zuo and Serfling (2000), Serfling (2006).

Example & Nonexample

- *Halfspace Depth* [Tukey, 1975]

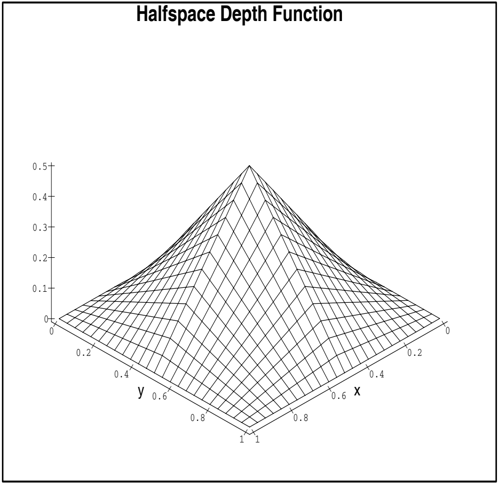
$$D_{\mathrm{H}}(\mathbf{x}, F) = \inf\{P(H) : \mathbf{x} \in H \text{ closed halfspace}\}, \ \mathbf{x} \in \mathbb{R}^d$$

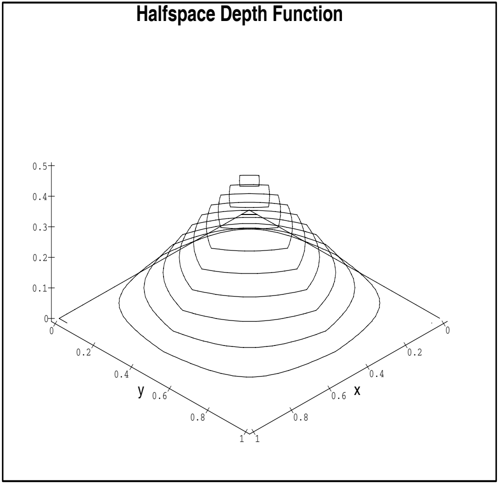
- *Univariate case*:

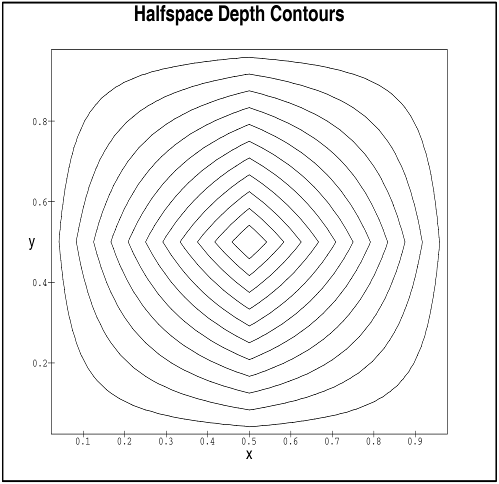
$$D(x, F) = \min\{F(x), 1 - F(x)\}.$$

- 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*.

Halfspace Depth for F Uniform on $[0, 1]^2$

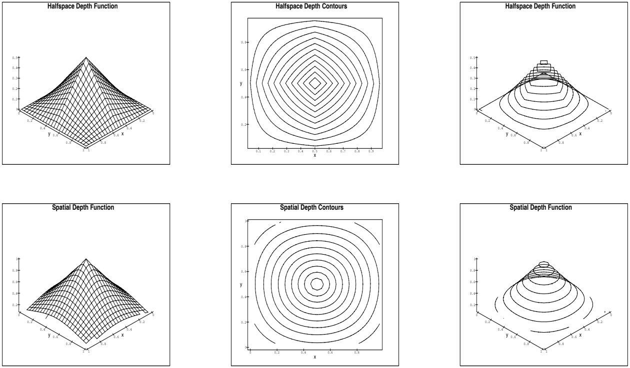



Halfspace Depth Function
0.5
0.4
0.3
0.2
0.1
0
0
0.2
0.4
y
0.6
0.8
1
1
0.8
0.6
x
0.4
0.2
0


Halfspace Depth Contours
0.8
0.6
y
0.4
0.2
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
x

Some Other Depth Functions

- *Simplicial depth* (Liu, 1988)
- *Spatial depth* (Dudley and Koltchinski, 1992; Möttönen and Oja, 1995; Chaudhuri, 1996; Vardi and Zhang, 2000)
- *Majority depth* (Singh, 1991)
- *Projection depth* (Liu, 1992; Zuo, 2003)
- *Simplicial volume depth* (based on Oja, 1983)
- *L^p depth* (Zuo and Serfling, 2000)
- *Mahalanobis depth* (Liu and Singh, 1993)
- *Zonoid depth* (Koshevoy and Mosler, 1997)
- And more ... a growing industry!



Comparative Views, Halfspace and Spatial Depths,

F Uniform on $[0, 1]^2$

Quantile Functions: "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 $\mathbb{B}^d(\mathbf{0})$ with *nested* contours

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

- 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: 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

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

Centered Rank Function: "Signs", "Ranks" in $\mathbb{R}^d$?

Centered rank function $\mathbf{R}(\mathbf{x}, F)$: takes values in $\mathbb{B}^d(\mathbf{0})$, with origin $\mathbf{0}$ assigned to a multivariate median $\mathbf{x} = \mathbf{M}_F$, and for other $\mathbf{x}$ denotes a "directional rank" in $\mathbb{B}^d(\mathbf{0})$.

- 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 multivariate version of the *univariate sign test*.

Example: 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*.

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

Equivalence: the DOQR 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, **Q**, and **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 $\mathbf{u} = \mathbf{R}(\mathbf{x}, F)$ is *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.*

Some History on DOQR Functions

Depth. Hotelling – Chamberlin – Hodges – Hill – Wilk – Gnanadesikan – Tukey – **Liu** – Donoho – Gasko – Dümbgen – Giné – Chen – Singh – Tyler – Nolan – Vardi – Zhang, C.-H. – **Müller** – Rousseeuw – Ruts – Struyf – **Hubert** – Massé – He – Bai – Portnoy – Romanazzi – Koshevoy – **Fraiman** – **Serfling** – **Mosler** – Zhang, J. – Mizera – Zuo – **Einmahl** – **Li, J.** – Lopez-Pintado – Hallin/Paindaveine/Sîman

- *Milestones*. Tukey, 1975 – Donoho/Gasko, 1982, 1992 – **Liu**, 1988 – Rousseeuw/**Hubert**, 1999 – **Liu**/Parelius/Singh, 1999 – Zuo/**Serfling**, 2000 – Mizera, 2002 – **Mosler**, 2002 – DIMACS Workshop (**Liu**/**Serfling**/Souvaine), 2003 – *This Workshop*, 2011

Outlyingness. Mosteller – Tukey – Donoho – Gasko – Rousseeuw – Zhang, J. – **Hubert** – **Dang/Serfling** – Mazumder/**Serfling**

Quantiles. Breckling – Chambers – Dudley – Koltchinskii – **Chaudhuri** – Chakraborty – **Serfling**/Zhou

Ranks. Brown – Hettmansperger – Nyblom – Oja – Möttönen

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*.
- *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.

- Transformation to other coordinates should not affect relative "outyingness" rankings and comparisons. The outlyingness function (as defined with domain the unit ball $\mathbb{B}^d(\mathbf{0})$) should be *affine invariant*: for any nonsingular $d \times d$ $\mathbf{A}$ and d-vector $\mathbf{b}$,

$$O(\mathbf{A}\mathbf{x} + \mathbf{b}, F_{\mathbf{A}\mathbf{X}+\mathbf{b}}) = O(\mathbf{x}, F_{\mathbf{X}}), \ \ \mathbf{x} \in \mathbb{R}^d.$$

- *Unqualified insistence* on equivariance/invariance as a principle is *not justified*, however, for it may lead to undue compromises of robustness or computational efficiency.

Quantile Equivariance and Outlyingness Invariance

Definition. An $\mathbb{R}^d$-valued quantile function $\mathbf{Q}(\mathbf{u}, F)$, $\mathbf{u} \in \mathbb{B}^d(\mathbf{0})$, 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 $\mathbf{b}$,

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

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

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

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

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

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

$\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{x}})$ *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 **A** orthogonal, in which case $\mathbf{v}(\mathbf{u}, \mathbf{A}, \mathbf{b}, F_{\mathbf{X}}) = \pm\mathbf{A}\mathbf{u}$.
- A point **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*.

Equivariance of Multivariate Median

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

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

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 $\mathbf{Y} = \mathbf{AX} + \mathbf{b}$,

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

- 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 $\mathbf{Q}(\cdot, F)$ and invariance of $O(\mathbf{x}, F)$ yields equivariance and invariance properties for the related $\mathbf{R}$ and D functions.
- Affine invariance of $D(\mathbf{x}, F)$ follows through its inverse relationship with $O(\mathbf{x}, F)$.
- 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:

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

With Multivariate Data, First Standardize!

- We can make statistics of interest suitably equivariant or invariant by first transforming the data appropriately.
- *Example*. To make the *spatial quantile function* affine equivariant, first standardize with a *TR functional*.

  Choosing a *SICS 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 *SICS functional*.
- We will define TR and SICS functionals.

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 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.

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 $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$ 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 a Structure A ICS functional 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*.

Construction of ICS and SICS 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 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.
  - These ICS functionals are not in general SICS.
- Ilmonen, Nevalainen, and Oja (2010) show that under some fairly severe restrictions on F (excluding elliptical cases, for example) the above constructions can be SICS.

A Family of Sample SICS Functionals *[Serfling, 2009, 2010]*

- 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

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

- This is a fully nonparametric approach.

- A special case of the preceding is a functional $\mathbf{M}_0(\mathbb{X}_n)$ of Chaudhuri and Sengupta (1993) based on a $\mathbb{Z}_N$ of size $N = d + 1$ derived by extensive computation prohibitive for high d.
- Alternatively, Mazumder and Serfling (2010) take for $\mathbb{Z}_N$ the set of observations selected and used in computing $\mathbf{\Sigma}_{\mathrm{MCD}}$ with, say, $N \approx 0.75n$. 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 (2010) is to compute $O_{\mathrm{MS}}(\mathbf{x}, F_{\mathbf{X}})$ 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 then to take the 75% least outlying points.

Non-Examples of SICS Functionals

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

Further Desired Properties of DOQR Functions

- *Robustness*. $O(\mathbf{x}, \mathbb{X}_n)$ and thus $D(\mathbf{x}, \mathbb{X}_n)$ should not be corrupted by the presence of outliers.
- *Maximality at center*. If P is symmetric about $\boldsymbol{\theta}$ in some sense, then $D(\mathbf{x}, P)$ is maximal at this point.
- *Symmetry*. If P is symmetric about $\boldsymbol{\theta}$ in some sense, then so is $D(\mathbf{x}, P)$.
- *Decreasing along rays*. The depth $D(\mathbf{x}, P)$ decreases along each ray from deepest point.
- *Vanishing at infinity*. $D(\mathbf{x}, P) \to 0$ as $\|\mathbf{x}\| \to \infty$.
- *Continuity of $D(\mathbf{x}, P)$ as a function of* $\mathbf{x}$.
- *Continuity as of $D(\mathbf{x}, P)$ a functional of P*.
- *Quasi-concavity as a function of* $\mathbf{x}$. The *level set* $\{\mathbf{x}: D(\mathbf{x}, P) \geq c\}$ is convex for each real c.

What We Compute from Depth Functions

- *Contours*.
- *Depth-based order statistics*. Ordering of data by depth value, center-outward.
- *Depth-weighted location functionals*.

$$\frac{\int_{\mathbb{R}^d} \mathbf{x}\, W(D(\mathbf{x}, F))\, dF(\mathbf{x})}{\int_{\mathbb{R}^d} W(D(\mathbf{x}), F))\, dF(\mathbf{x})}.$$

- *Depth-weighted scatter matrix functionals*.
- *Scale curves*. Plot of volumes within contours as a function of probability weight.
- *Skewness functionals*. Scaled difference of two location functionals.
- *Kurtosis functionals*. Via transformation of scale curve.
- And more ...

Applications: Depth-Based Statistical Procedures

- *Bagplots, sunburst plots*. Extends boxplot to $d = 2, 3$, using contours and rays to outlying points. Answers question: *where is the "middle half" of the data?*
- *DD, PP, QQ plots*. Compare two samples by a plot of depth values of combined sample, or of the volumes of sample central regions versus each other, or of kurtosis curves against each other, or of depth-based quantiles versus each other.
- *Comparison of several distributions*. Plot scale curves in a single exhibit. Or kurtosis curves.
- *Nonparametric description of multivariate distributions*. Measures of location, spread, asymmetry, kurtosis.

- *Testing multivariate symmetry*. For *spherical symmetry*, plot fraction of data in smallest sphere containing the pth sample central region. For *central symmetry*, plot fraction of data within the intersection of pth sample central region and its reflection.
- *Diagnosis of nonnormality*. Use trimmed depth-weighted scatter matrix. Or kurtosis curve.
- *Outlier identification*.
- *P-values in hypothesis testing via bootstrap and data depth*.
- *One- and multi-sample multivariate rank statistics defined on depth-based ranks*.

- *Statistical process control procedures*. Can use depth-based ranks for monitoring and thresholding with multivariate data using univariate quality control procedures. Compare outlyingness of a new observation relative to in-control reference point cloud.
- *Multivariate density estimation by probing depth*.
- *Depth-based quality indices*.
- *Depth-based classification rules*.
- *Depth-based cluster analysis*.

Contexts of Application

Depth-based methods provide *new competitors* for standard approaches having diverse applications:

- Exploratory data analysis
- Multi-sample inference
- Regression
- Classification, clustering, discrimination
- Directional analysis
- Multivariate density estimation

Further, in some application contexts, depth-based methods are *especially natural* or *advantageous*:

- Monitoring of aviation safety data
- Industrial quality control
- Measuring economic disparity and concentration
- Social choice and voting (Ordering social choices by voters' preferences, and finding "median voter")
- Game-theoretic analysis of competition

Computational Burden ... Involves Computational Geometry!

- *halfspace and simplicial depth*: $O(n^{d-1}\log n)$
- *nested convex hulls of fractions of data*: $O(n^{d-1}\log n)$
- *bivariate halfspace contours, bagplot, and data depths*: $O(n^2)$
- *deepest regression hyperplane*: $O(n^d)$
- *regression depth of a k-flat*: $O(n^{d-2} + n\log n)$, for $1 \le k \le d-2$
- *Stahel-Donoho estimators*: $O(n^{d+1})$
- *volume of sample pth central region.*
- *simulation and bootstrap: many samples needed.*
- *issues of NP-hard, NP-complete, coNP-complete, etc.*

Consistency of Depth Process

We want

$$\|D(\mathbf{x}, \hat{F}_n) - D(\mathbf{x}, F)\|_\infty \to 0, \ \ n \to \infty.$$

- For *halfspace depth*: Donoho and Gasko (1992).
- For *simplicial depth*: Liu (1990), Dümbgen (1990), and Arcones and Giné (1993).
- For *majority and Mahalanobis depths*: Liu and Singh (1993).
- For *projection depth and Type D depths*: Zuo and Serfling (2000).

Weak Convergence of Depth Process

We want

$$n^{1/2}[D(\mathbf{x}, \hat{F}_n) - D(\mathbf{x}, F)] \ \textit{converges weakly.}$$

- For *halfspace depth*: Massé (2004)
- For *simplicial depth*: Dümbgen (1990), and Arcones and Giné (1993). Note: This depth process is a special case of *U-process*.

Convergence of Sample Central Regions

- We desire convergence of contours in *Hausdorff distance*.
- For *halfspace depth*: Eddy (1985), Nolan (1992), Donoho and Gasko (1992), Massé and Theodorescu (1994), and He and Wang (1997).
- For *simplicial, projection, and Type D depths*: Zuo and Serfling (2000).

Asymptotic Normality of Location Estimator $L(\widehat{F}_n)$

$$L(F) =$$

$$\int_{\mathbb{R}^d} \mathbf{x}\, W(D(\mathbf{x}, F))\, dF(\mathbf{x}) \Big/ \int_{\mathbb{R}^d} W(D(\mathbf{x}), F))\, dF(\mathbf{x}) \;.$$

- For *simplicial depth*: Dümbgen (1992).
- For *halfspace depth*: Massé (1999, 2004).
- For *generalized Tukey depth*: Zhang (2002).
- For *projection depth, Mahalanobis distance depth, and others*: Zuo, Cui, and He (2004).

Asymptotic Normality of Scatter Estimator $S(\widehat{F}_n)$

$$S(F) =$$

$$\int_{\mathbb{R}^d} (\mathbf{x} - L(F))(\mathbf{x} - L(F))' \, w(D(\mathbf{x}, F)) \, dF(\mathbf{x}) \bigg/ \int_{\mathbb{R}^d} w(D(\mathbf{x}), F)) \, dF(\mathbf{x}) \, .$$

- Zuo and Cui (2005).

Weak Convergence of Sample Scale Functions and Similar Constructions

- Sample volume and scale functionals are formulated in Liu, Parelius, and Singh (1999).
- The volume functional may be written as a "*generalized (univariate) quantile function*" in the sense of Einmahl and Mason (1992). See Serfling (2002).
- Accordingly, appropriate sample versions may be treated as generalized quantile processes, but the formulation of sample versions that fit into the Einmahl/Mason framework is problematic.
- Partial results are available in Serfling (2002) and Wang and Serfling (2006).

Bahadur-Kiefer Representations for Sample Quantile Functions

- For the *spatial quantile function*, Chaudhuri (1992, 1996) and Koltchinskii (1994).
- For *Mahalanobis spatial quantile functions*, Serfling (2008).
- For *spatial U-quantile functions*, Zhou and Serfling (2008).
- Applications of BK Representations:
  - *Joint asymptotic normality of a vector of sample quantiles*
  - *Linkage between estimators and test statistics.*
  - *Almost sure convergence and law of iterated logarithm for sample quantile function.*

Comments

- The preceding examples on asymptotics are representative illustrations, with the emphasis on *depth functions* merely for convenience.
- The "D" examples imply – via the DOQR paradigm – corresponding examples for O-Q-R.
- The relevant asymptotics must in practice include the significant added technical complication of dealing with sample versions based on standardizing functionals.
- Asymptotic problems in the DOQR landscape are diverse and complex with many *open gaps and needed extensions*.

What Target Inference Problems are Possible?

- The foregoing treatment has been confined to *location depth*, i.e., to depth functions such that the maximal depth point is a *location parameter*. This is the setting of *location inference*.
- To cover a greater range of inference settings, we define depth functions on the *parameter space*.
- Then, for example, the maximal depth point can be a *dispersion matrix* or a *regression hyperplane*.

- *Key Idea*. Let the inference problem using a data set $\mathbb{X}_n$ from a distribution F be defined by a target parameter $\theta(F)$, concerning location, scale, skewness, kurtosis, or a regression line or hyperplane, for example.
- *Approach*. Define a data-based function of θ representing its "outlyingness" in Θ when considered as an estimate of $\theta(F)$. Take as estimator the minimum-outlyingness element.
- *Analogy with maximum likelihood*. This is similar to evaluating a density function $f(x;\theta)$, $x \in \mathcal{X}$, at the observed data $\mathbb{X}_n$ and then considering the data-based function $L(\theta) = f(\mathbb{X}_n;\theta)$, $\theta \in \Theta$, as an objective function of θ to be maximized in the parameter space Θ.

General Formulation

- A "parameter" space Θ.
- A sample space $\mathcal{X}$ for "data" $\mathbb{X}_n = (X_1, \ldots, X_n)$.
- A criterion function describing relevance or "closeness" of θ to $\mathbb{X}_n$.
- *Goal*. Find $\theta \in \Theta$ "*best fitting*" the data set $\mathbb{X}_n$ or (for robustness) a reduced data set after *identification and possible elimination of outliers*.
- *Maximal depth approach*. Represent the *relevant criterion function* as a *data-based* depth function $D(\theta, \mathbb{X}_n)$, $\theta \in \Theta$, defined on the parameter space Θ, with "*best fit*" defined as *maximal depth*.
- *Issue*: Passing from the "*relevant criterion function*" to an *appropriate depth function*.

Example: Location Estimation in $\mathbb{R}^d$

- $\Theta = \mathbb{R}^d = \mathcal{X}$
- $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$
- A depth function $D(\mathbf{x}, F)$, $\mathbf{x} \in \mathbb{R}^d$, is maximized at a "center" interpreted as a location parameter. The sample analogue $D(\mathbf{x}, \mathbb{X}_n)$, $\mathbf{x} \in \mathbb{R}^d$, induces, by reinterpretation, a depth on the possible values for the location parameter, i.e., $\widetilde{D}(\theta, \mathbb{X}_n)$, $\boldsymbol{\theta} \in \Theta$ $(= \mathbb{R}^d)$, with $\widetilde{D}(\cdot, \mathbb{X}_n) = D(\cdot, \mathbb{X}_n)$.
- Thus maximal $\widetilde{D}$-depth is equivalent to maximal D-depth. That is, location estimation by maximizing sample depth $D(\mathbf{x}, \mathbb{X}_n)$ in $\mathbb{R}^d$ is equivalent to maximizing sample depth $\widetilde{D}(\theta, \mathbb{X}_n)$ in the parameter space which also is $\mathbb{R}^d$.

Example: Dispersion Estimation in $\mathbb{R}^d$

- $\Theta = \{\ d \times d$ covariance matrices $\mathbf{\Sigma}\}$.
- For univariate data $\mathbb{Y}_n = (Y_1, \ldots, Y_n)$, and some univariate spread measure $\Delta(\cdot)$, for example the MAD, define outlyingness of a scale parameter σ by

$$O_{1n}(\sigma, \mathbb{Y}_n) = 2 \left| \frac{\Delta(\mathbb{Y}_n)/\sigma}{1 + \Delta(\mathbb{Y}_n)/\sigma} - \frac{1}{2} \right| .$$

- Then, for $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, define outlyingness of $\mathbf{\Sigma}$ via projection-pursuit:

$$O_{dn}(\mathbf{\Sigma}, \mathbb{X}_n) = \sup_{\|\mathbf{u}\|=1} O_{1n}\left(\sqrt{\mathbf{u}'\mathbf{\Sigma}\mathbf{u}},\ \mathbf{u}'\mathbb{X}_n\right) .$$

- *Estimation*. Minimization of $O_{dn}(\mathbf{\Sigma}, \mathbb{X}_n)$ with respect to $\mathbf{\Sigma}$ yields a *maximum "dispersion depth" estimator*.

Example: Regression

- $\Theta = \{\text{hyperplanes}\}$
- For $\mathbb{R}^d$-valued data $\mathbb{X}_n = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$, define the depth of a hyperplane h as the

  minimum fraction of observations in $\mathbb{X}_n$ whose removal makes h a "nonfit".

  (Rousseeuw and Hubert, 1999).
- We make "fit" and "nonfit" precise in the more general setting considered next.

A Broad Example: Shape Fitting

- *Shape-fitting problems* arise not only in regression but also in *computer vision*, *machine learning*, and *data mining*, for example.

- E.g., find the *point*, *line*, *hyperplane*, *hypersphere*, *sphere*, or *cylinder* best fitting a point set $\mathbb{X}_n$.

  I.e., find the *minimum radius sphere*, *minimum width cylinder*, *smallest width slab*, *minimum radius spherical shell*, or *minimum radius cylindrical shell* that encloses the point set $\mathbb{X}_n$.

- This is the realm of *computational geometry*.

- *The goal*: Given
  - a *family of shapes* Θ
  - a set of input points $\mathbb{X}_n$ in $\mathcal{X}$
  - a *fitting criterion* $\beta(\theta, \mathbb{X}_n)$ defined over $\theta \in \Theta$,

  which shape θ best fits $\mathbb{X}_n$ in the sense of minimizing $\beta(\theta, \mathbb{X}_n)$?

- A major open problem in the general case is *identification and handing of outliers*.

  Only *ad hoc* solutions in special cases have been given.

  A *general approach* is needed.

Formulation of Shape Fitting Criteria

Assumption. Let the shape-fitting criterion $\beta(\theta, \mathbb{X}_n)$ be a function of *pointwise* functions

$$\beta(X_i, \theta, \alpha(\mathbb{X}_n)), \ 1 \leq i \leq n,$$

that measure the closeness of the data points X_i to a given θ.

Example A

With the data space given by $\mathbb{R}^d$ and the shape space Θ given by some class of subsets in $\mathbb{R}^d$, the pointwise objective function

$$\beta(\mathbf{x}, \theta) = \min_{\mathbf{y} \in \theta} \|\mathbf{x} - \mathbf{y}\|$$

represents the minimum Euclidean distance from $\mathbf{x}$ to a point in the set θ. Then

$$\beta(\theta, \mathbb{X}_n) = \max_{1 \leq i \leq n} \beta(\mathbf{X}_i, \theta)$$

represents the *maximal distance* from a point in $\mathbb{X}_n$ to the nearest point in θ.

Thus *fitting a line to* $\mathbb{X}_n$ becomes the same as finding *the minimum radius enclosing cylinder* θ.

*Example A**

Alternatively, replace *maximum* by *sum*:

$$\beta(\theta, \mathbb{X}_n) = \sum_{1 \leq i \leq n} g(\beta(\mathbf{X}_i, \theta)).$$

With $g(\cdot)$ the *identity function*, this gives the *total distance* from the points of $\mathbb{X}_n$ to θ as the optimization criterion.

Thus, for example, fitting a line to $\mathbb{X}_n$ yields the *principal components regression line*.

For *location problems*, $\beta(\mathbf{x}, \theta)$ becomes $\|\mathbf{x} - \boldsymbol{\theta}\|$ and yields the *spatial median* as the optimal fit.

With $g(\cdot)$ an *increasing* function and θ now representing a hyperplane, other objective functions arising in regression problems are obtained.

Example B: Clustering

- The *k-center problem* involves the objective function of Example A.
- The *k-median problem* involves the objective function of Example A* with $g(\cdot)$ the identity function. This is a special case of the *facility location problem*.

Example C: Projective Clustering

- The *approximate k-flat problem* involves the objective function of Example A, with θ a k-flat in $\mathbb{R}^d$.
- The *j approximate k-flat problem* involves the maximum-type objective function with the pointwise function
$$\beta(\mathbf{x}, \theta) = \min_{1 \le i \le j} \min_{\mathbf{y} \in \mathcal{F}_i} \|\mathbf{x} - \mathbf{y}\|,$$
where θ is a vector $(\mathcal{F}_1, \ldots, \mathcal{F}_j)$ of j k-flats in $\mathbb{R}^d$.

Data-Based Depth on Fits θ: Approach 1

- Consider any $\mathcal{X}$, Θ, and criterion function $\beta(x, \theta, \alpha(\mathbb{X}_n))$.
- A fit θ is "*inadmissible*" if improvable by some other value $\tilde{\theta}$: $\beta(x, \tilde{\theta}, \alpha(\mathbb{X}_n)) \leq \beta(x, \theta, \alpha(\mathbb{X}_n))$, $x \in \mathbb{X}_n$, with at least one inequality strict.
- Define the *depth* $D_n(\theta)$ of θ as the *minimum fraction of points in* $\mathbb{X}_n$ *whose removal makes* θ *"inadmissible"*.
- In the *location* problem, the points outside the convex hull of the data are "nonfits".
- For a special case of $\beta(\cdot, \cdot)$, this *maximal depth* point generalizes the *halfspace median*.
- Extends Rousseeuw and Hubert (1999), Mizera (2002), and Zhang, J. (2002), and includes our previous examples.

Data-Based Depth on Fits θ: Approach 2

- Again consider any $\mathcal{X}$, Θ, and pointwise criterion function $\beta(x, \theta, \alpha(\mathbb{X}_n))$.
- Define the depth of θ by

$$D_n^*(\theta) = \frac{1}{1 + \max_{1 \leq i \leq n} \beta(X_i, \theta, \alpha(\mathbb{X}_n))}.$$

Data-Based Depth on Fits θ: Approach 3

- Again consider any $\mathcal{X}$, Θ, and pointwise criterion function $\beta(x, \theta, \alpha(\mathbb{X}_n))$.
- Define the depth of θ by

$$D_n^*(\theta) = \frac{1}{1 + \sum_{1 \le i \le n} \beta(X_i, \theta, \alpha(\mathbb{X}_n))}.$$

For a special case of $\beta(\cdot, \cdot)$, the maximal depth point now generalizes the *spatial median*.

The Key Outlier Problem in Shape Fitting

- From depth and outlyingness defined on *fits*, pass to
  1. *identification of "outliers" in the input*, i.e., those points causing the fit to be significantly less than "optimal",
  2. *optional removal* of selected outliers,
  3. *optimal fit* to the remaining points.
- One approach towards a solution:

residuals analysis

A General "Residuals Approach"

1. Obtain an initial fit θ_n^* as a *maximum depth fit*, employing for example a suitable version of depth $D_n(\theta)$ based on a given criterion function $\beta(x, \theta, \alpha(\mathbb{X}_n))$.
2. Define an *outlyingness measure* for each input data point, using the *outlyingness function* induced on $\mathcal{X}$ via $O_n(x) = \beta(x, \theta_n^*, \alpha(\mathbb{X}_n))$. The "residuals" are $O_n(X_i)$, $1 \leq i \leq n$.
3. Set a threshold for *removal of outliers* among $X_1, \ldots, X_n$ and reduce $\mathbb{X}_n$ to $\widetilde{\mathbb{X}}_n$.
4. *Iterate* Steps 2 and 3 to reach a satisfactory "optimum".

NOTE. The ultimate success of this depends on choosing a *robust* version of $D_n(\theta)$ in Step 1. (We do not treat this here.)

Final Comments

- An *apologia*: This presentation is brief and uneven, with much relevant work unmentioned. It provides a sketchy overview of the *landscape* and suppresses technical detail.
- In preparation: *Depth and Quantile Functions in Nonparametric Multivariate Analysis*, Springer
- 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.

DANKE SCHÖN!!