An Affine-Invariant Data Depth Based on Random Hyperellipses

Ryan T. Elmore

Ryan.Elmore@anu.edu.au

ANU $\xrightarrow{pr}$ Colorado State University*

8 – 10 June 2005

This is joint work with Thomas P. Hettmansperger, Fengjuan Xuan, and Bruce Brown.

$*$– Note that $\xrightarrow{pr}$ denotes convergence in “profession”.

Outline

- What is data depth?
- Elliptical data depth
- Properties of elliptical depth
- Illustrations
- Applications
- Example(s)

Data Depth

Zuo and Serfling (2000) informal definition:

> "...for a distribution P on $\mathbb{R}^d$, a corresponding depth function is any function $D(x; P)$ which provides a P-based center-outward ordering of points $\boldsymbol{x} \in \mathbb{R}^d$."

- Monotonicity of $D(\cdot; P)$ relative to deepest point.
- Affine invariance of the depth function.
- Maximum depth at the "center" of the distribution.
- $D(\boldsymbol{x}; P) \to 0$ as $\|\boldsymbol{x}\| \to \infty$

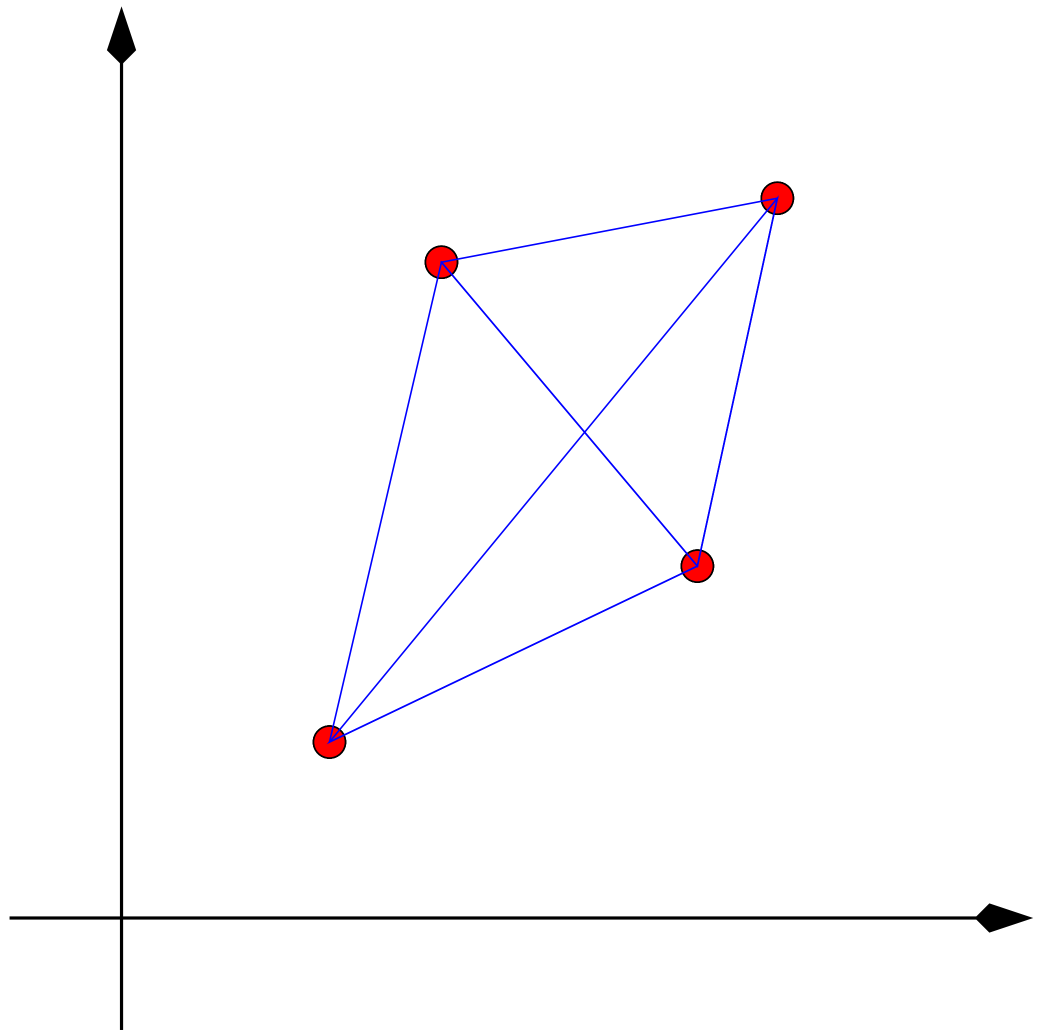
Data Depth (cont.)

- Liu, Parelius, and Singh (1999) provide a comprehensive overview of data depths, their properties, and potential applications.
- Data depth functions are a nonparametric, exploratory data-analytic technique for describing multivariate data sets, e.g. DD-plots or sunburst plots.
- Used to quantify a point or region in high dimensions as a single-dimensional quantity.
- Generally, data depths provide a center-outward ranking of the data; this leads to rank-based inferential procedures.

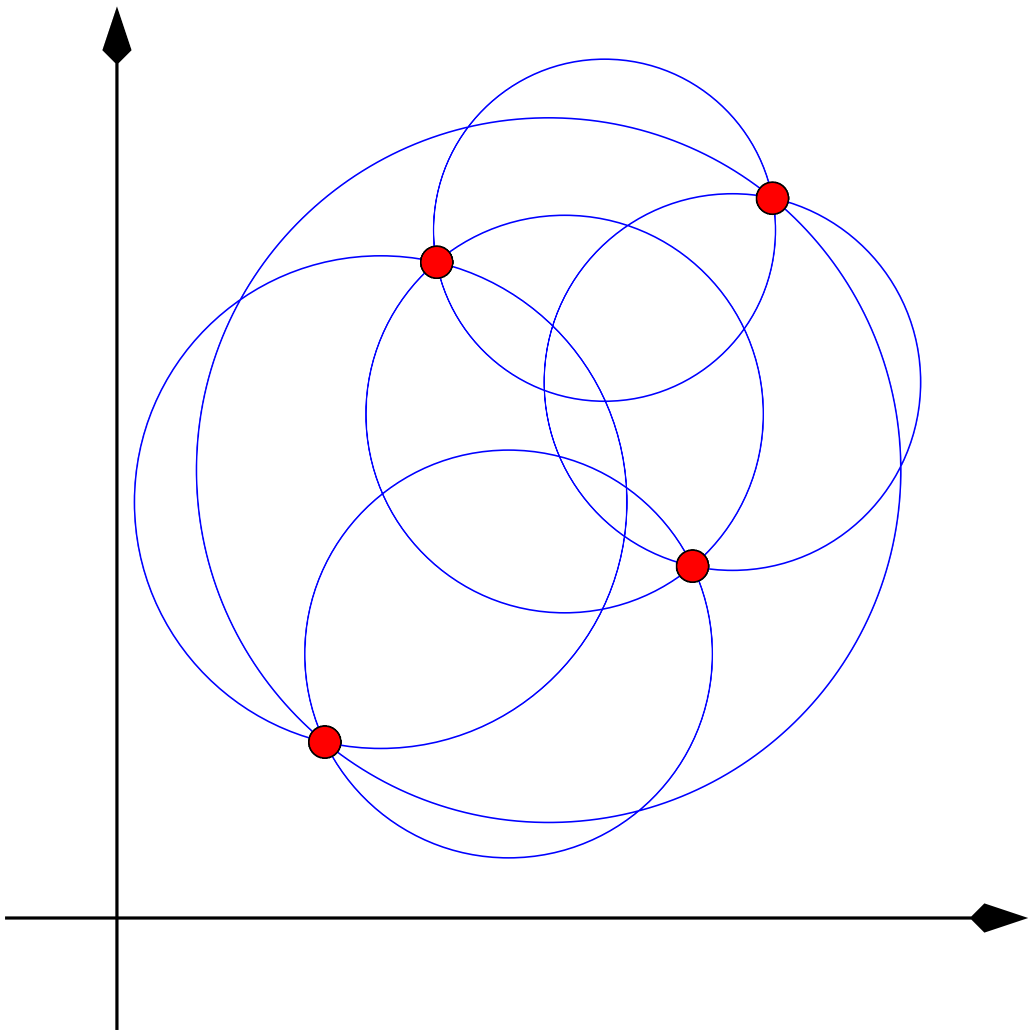
Data Depth (cont.)

- Examples: Mahalanobis depth, Mahalanobis (1936); Halfspace depth, Tukey (1975); Oja's depth, Oja (1983); Simplicial depth, Liu (1990); Spherical depth, Elmore et al. (2004).
- Most of the current depth functions are computationally intractable for high dimensions.
- For example, algorithms for computing the simplicial depth are of order $O(n^{d+1})$; however, Rousseeuw and Ruts (1996) describe an algorithm that reduces the order to $O(n^{d-1}\log n)$.

Sample Simplicial Depth



Sample Spherical Depth



Elliptical Data Depth

- Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two independent random vectors having common probability distribution function F on $\mathbb{R}^d$, $d \geq 1$.
- The elliptical depth function is defined in terms of the distribution F at a point $\boldsymbol{t} \in \mathbb{R}^d$ by

$$D(\boldsymbol{t}; \mathbf{C}_F) = P_F \left[\boldsymbol{t} \in e(\boldsymbol{X}, \boldsymbol{Y})\right]$$

  where the region $e(\boldsymbol{X}, \boldsymbol{Y})$ denotes the unique, closed random *hyperellipse* formed by $\boldsymbol{X}$, $\boldsymbol{Y}$, and the symmetric, positive-definite matrix $\mathbf{C}_F$.
- The elliptical region is defined by

$$e(\boldsymbol{X}, \boldsymbol{Y}) = \left\{\boldsymbol{t} : (\boldsymbol{X} - \boldsymbol{t})^T \mathbf{C}_F^{-1} (\boldsymbol{Y} - \boldsymbol{t}) \leq 0\right\}.$$

Theorems

1. If F is an absolutely continuous distribution on $\mathbb{R}^d$, then $D(\boldsymbol{x}; \mathbf{C}_F)$ is continuous with respect to $\boldsymbol{x}$.

Theorems

1. If F is an absolutely continuous distribution on $\mathbb{R}^d$, then $D(\boldsymbol{x}; \mathbf{C}_F)$ is continuous with respect to $\boldsymbol{x}$.

2. If F is an absolutely continuous distribution on $\mathbb{R}^d$ and it is angularly symmetric about the $\boldsymbol{\theta}$, then $D(\boldsymbol{\theta}; \mathbf{C}_F) = 1/2$.

Theorems

1. If F is an absolutely continuous distribution on $\mathbb{R}^d$, then $D(\boldsymbol{x}; \mathbf{C}_F)$ is continuous with respect to $\boldsymbol{x}$.

2. If F is an absolutely continuous distribution on $\mathbb{R}^d$ and it is angularly symmetric about the $\boldsymbol{\theta}$, then $D(\boldsymbol{\theta}; \mathbf{C}_F) = 1/2$.

3. If F is absolutely continuous and angularly symmetric about the origin, then $D(\alpha\boldsymbol{x}; \mathbf{C}_F)$ is a monotone nonincreasing in $\alpha \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

Theorems

1. If F is an absolutely continuous distribution on $\mathbb{R}^d$, then $D(\boldsymbol{x}; \mathbf{C}_F)$ is continuous with respect to $\boldsymbol{x}$.

2. If F is an absolutely continuous distribution on $\mathbb{R}^d$ and it is angularly symmetric about the $\boldsymbol{\theta}$, then $D(\boldsymbol{\theta}; \mathbf{C}_F) = 1/2$.

3. If F is absolutely continuous and angularly symmetric about the origin, then $D(\alpha\boldsymbol{x}; \mathbf{C}_F)$ is a monotone nonincreasing in $\alpha \geq 0$ for all $\boldsymbol{x} \in \mathbb{R}^d$.

4. For any distribution function F on $\mathbb{R}^d$ and $\boldsymbol{x} \in \mathbb{R}^d$, the elliptical depth function vanishes at infinity, i.e. $\sup_{\|\boldsymbol{x}\| \geq M} D(\boldsymbol{x}; \mathbf{C}_F) \to 0$ as $M \to \infty$.

Empirical Version

- Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ be a random sample from the distribution F.

- We define the sample elliptical depth function at a point $\boldsymbol{t}$ as

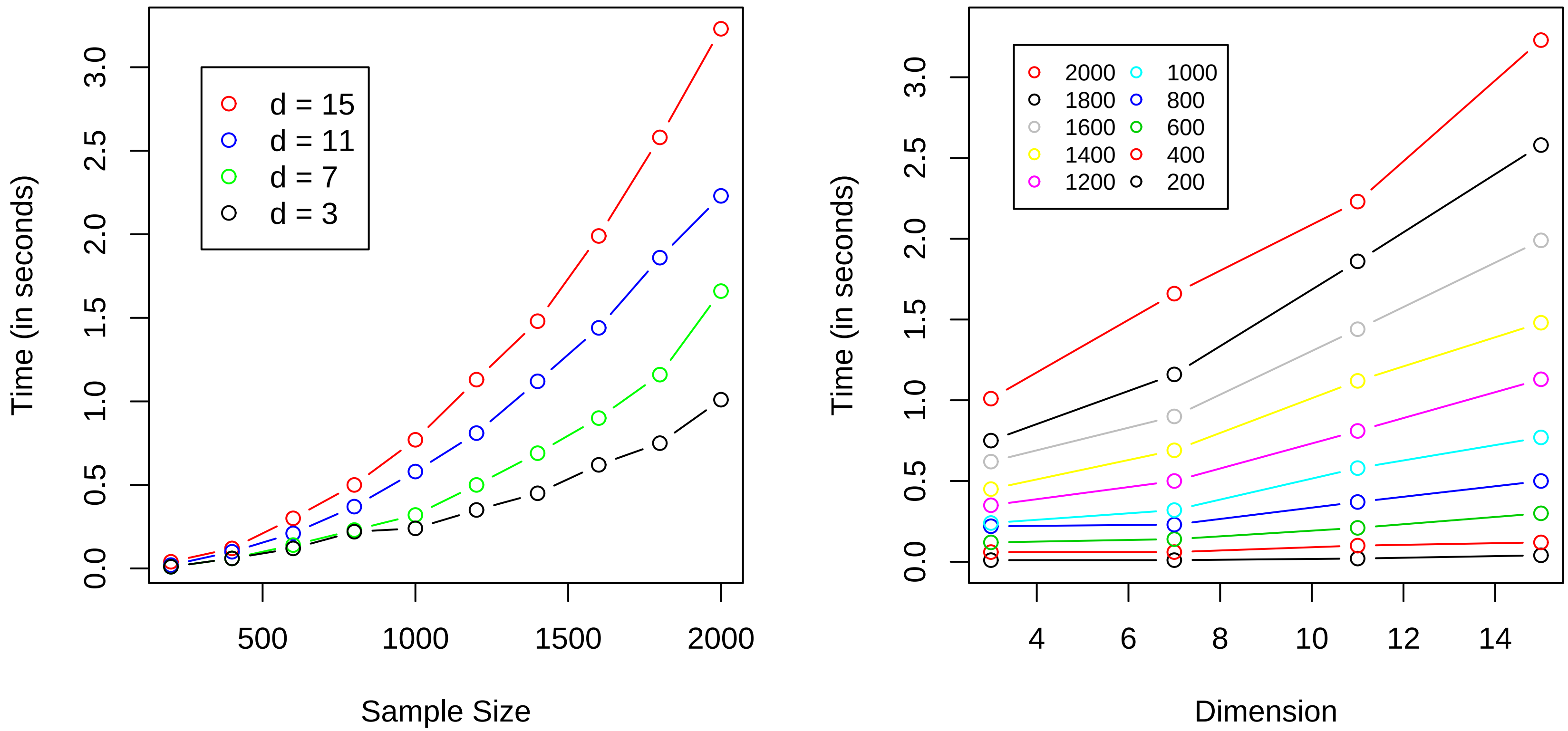
$$D_n(\boldsymbol{t}; \mathbf{C}_F) = \binom{n}{2}^{-1} \sum_{i<j} I\left(\boldsymbol{t} \in e(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)$$

  where $I(A)$ is the indicator function of the event A.

- It is easy to see that even the most naïve algorithm is of order $O(dn^2)$.

- In practice, the matrix $\mathbf{C}_F$ is usually unknown and must be estimated.

Order of Computation



Empirical Version (cont.)

- The more practical estimator given by

$$D_n(\boldsymbol{t};\hat{\mathbf{C}}_x) = \binom{n}{2}^{-1} \sum_{i<j} I\left(\boldsymbol{t} \in e_n(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)$$

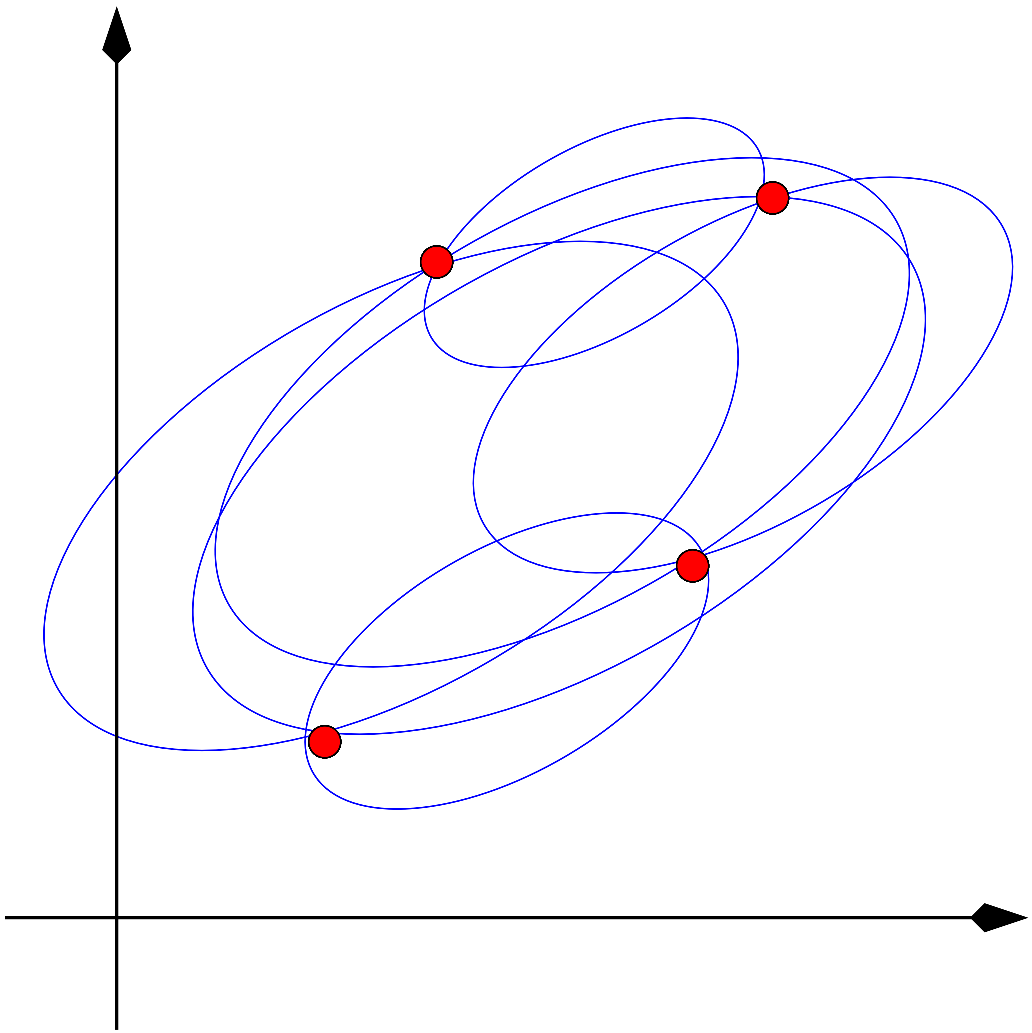
where

$$e_n(\boldsymbol{x}_i, \boldsymbol{x}_j) = \left\{\boldsymbol{t} : (\boldsymbol{x}_i - \boldsymbol{t})^T \hat{\mathbf{C}}_x^{-1} (\boldsymbol{x}_j - \boldsymbol{t}) \le 0\right\}$$

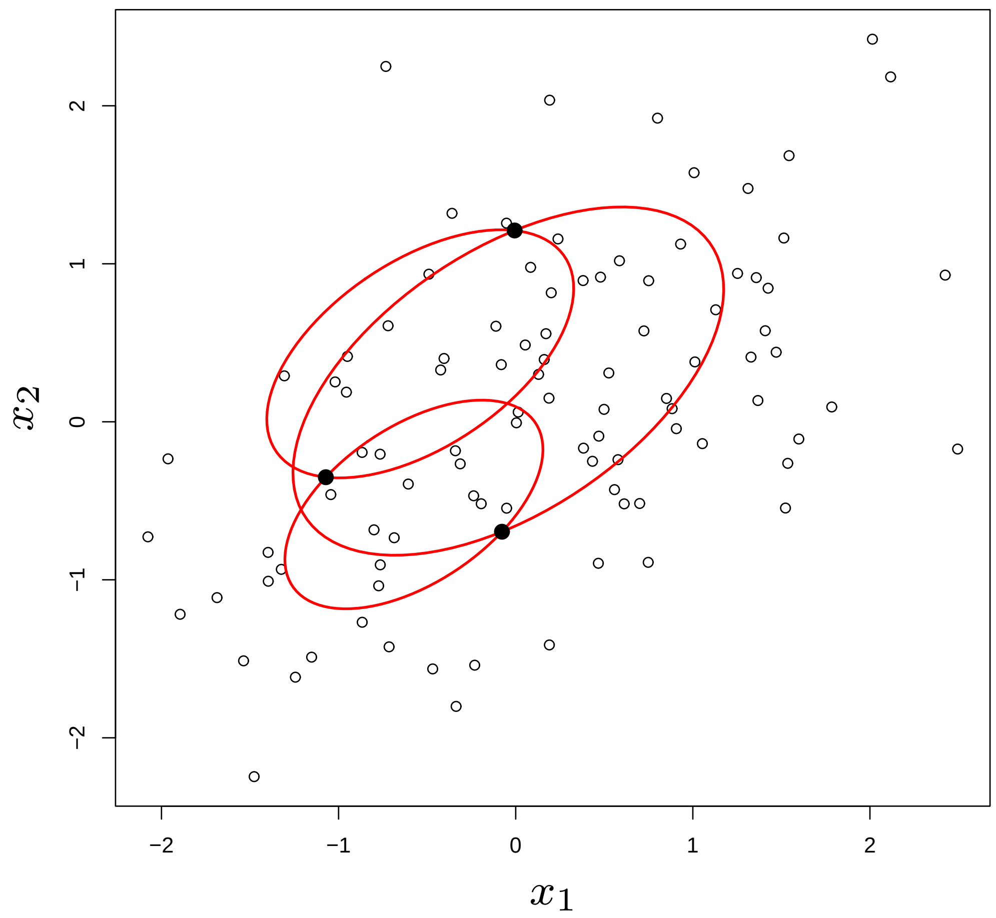
for some affine-equivariant estimator of the scatter matrix, $\mathbf{C}_F$.

- Spherical depth is a special case of elliptical depth with $\mathbf{I}_d$ is used wherever a $\mathbf{C}_F$ is given above.

Sample Elliptical Depth



Sample Elliptical Depth



Scatter Matrices

- A data-determined, symmetric, positive-definite matrix $\hat{\mathbf{B}}_x$ based on $\boldsymbol{x}_i$ for $i = 1, 2, \ldots, n$ is said to be an affine equivariant scatter matrix if and only if whenever each $\boldsymbol{x}_i$ is transformed by a fixed, nonsingular $d \times d$ matrix $\mathbf{D}$ into $\mathbf{D}\boldsymbol{x}_i$, the resulting $\hat{\mathbf{B}}_{\mathbf{D}x}$ matrix satisfies

$$\mathbf{D}^T \hat{\mathbf{B}}_{\mathbf{D}x}^{-1} \mathbf{D} = c_0 \hat{\mathbf{B}}_x^{-1}$$

  where c_0 is a positive scalar that may depend on $\mathbf{D}$ and the $\boldsymbol{x}_i$'s.

- Examples: Sample covariance matrix and Tyler's (1987) M-estimator of scatter.

Tyler's Scatter Matrix

- Tyler's scatter matrix $(\hat{\mathbf{A}}_x^T \hat{\mathbf{A}}_x)^{-1}$ is defined so that $\hat{\mathbf{A}}_x$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\hat{\mathbf{A}}_x(\boldsymbol{x}_i-\boldsymbol{\theta})}{\|\hat{\mathbf{A}}_x(\boldsymbol{x}_i-\boldsymbol{\theta})\|}\right)\left(\frac{\hat{\mathbf{A}}_x(\boldsymbol{x}_i-\boldsymbol{\theta})}{\|\hat{\mathbf{A}}_x(\boldsymbol{x}_i-\boldsymbol{\theta})\|}\right)^T = \frac{1}{d}\mathbf{I}_d$$

  where $\mathbf{I}_d$ is the d-dimensional identity matrix and $\boldsymbol{\theta}$ is a measure of center.

- Tyler argues that his scatter matrix is the "most robust" estimator of $\mathbf{C}$, the scatter matrix of an elliptical distribution.

- He shows that $(\hat{\mathbf{A}}_x^T \hat{\mathbf{A}}_x)^{-1}$ is strongly consistent in estimating $\mathbf{C}$ when sampling from a continuous distribution.

Theorem

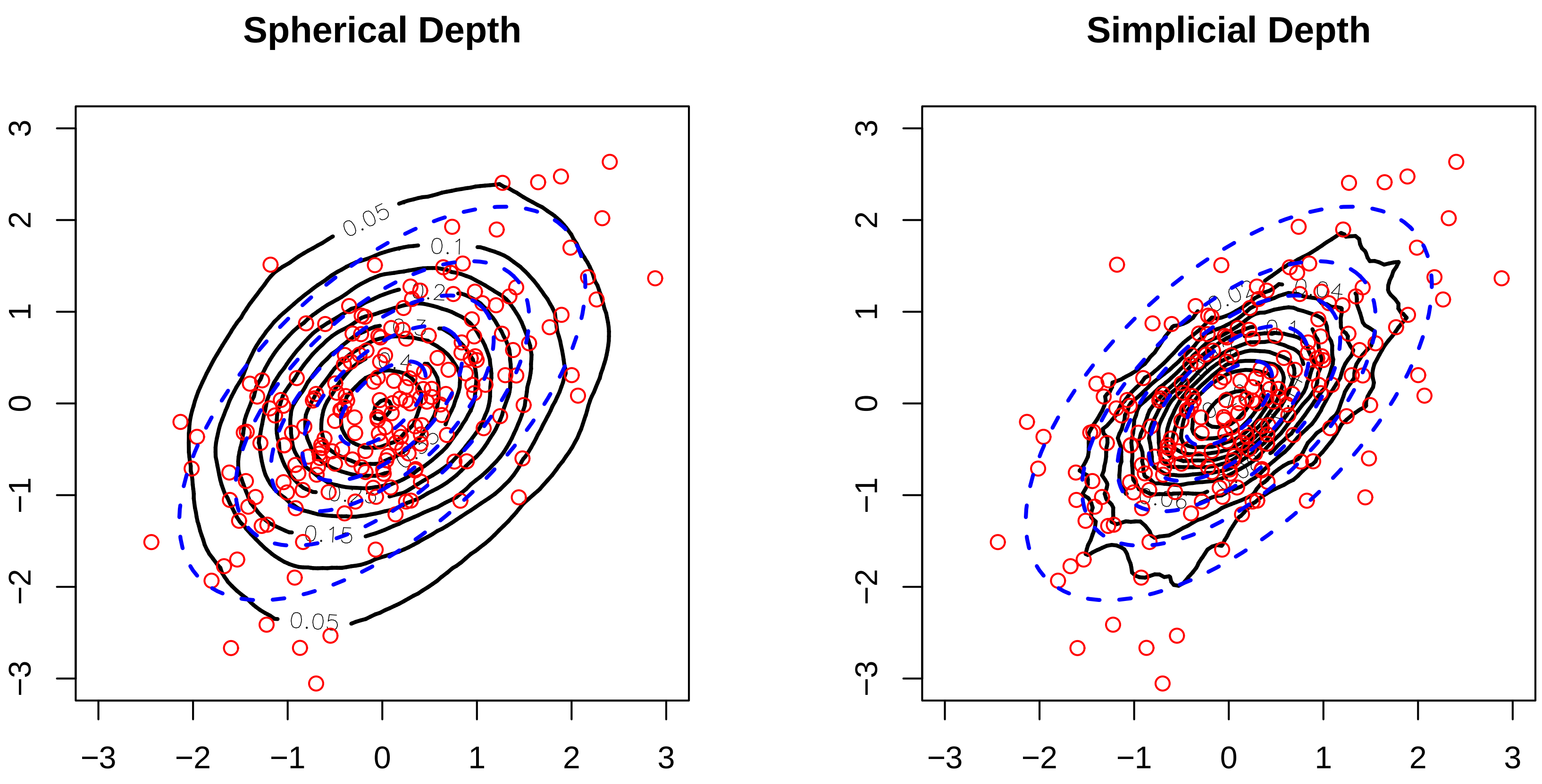
The two elliptical depth measures defined in equations above are affine invariant. That is, for any nonsingular matrix $\mathbf{D}$, we have

$$\begin{aligned} D_n(\boldsymbol{t}; \mathbf{C}_F) &= D_n(\boldsymbol{t}_*; \mathbf{D}\mathbf{C}_F\mathbf{D}^T), \text{ and} \\ D_n(\boldsymbol{t}; \hat{\mathbf{C}}_x) &= D_n(\boldsymbol{t}_*; \hat{\mathbf{C}}_y) \end{aligned}$$

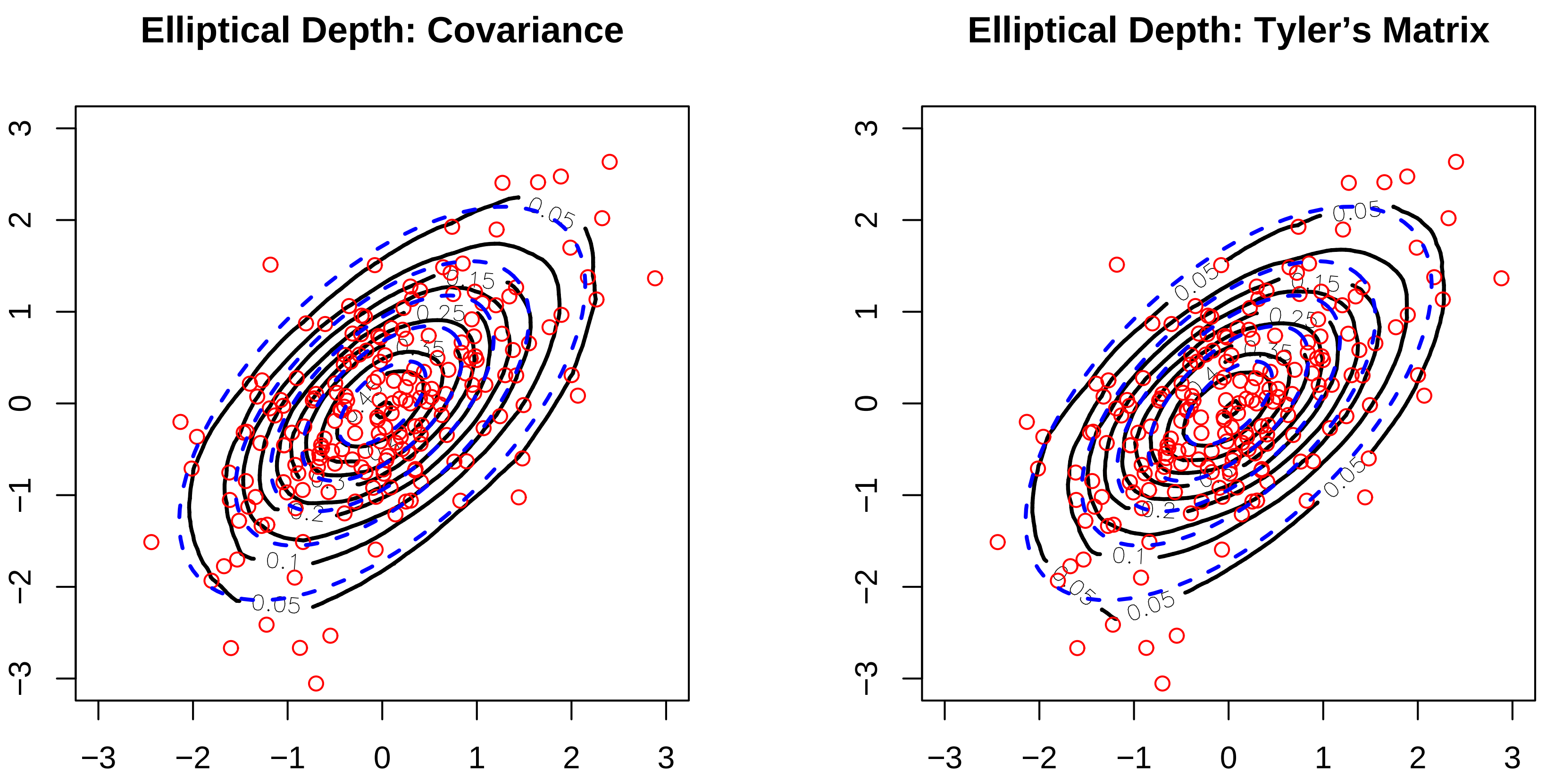
where $\boldsymbol{t}_* = \mathbf{D}\boldsymbol{t}$ and $\hat{\mathbf{C}}_y$ is an affine-equivariant scatter matrix defined by $\boldsymbol{y}_i$, $i = 1, \ldots, n$. To see this, note that for $\boldsymbol{Y} = \mathbf{D}\boldsymbol{X}$

$$\begin{aligned} D_n(\boldsymbol{t}_*; \hat{\mathbf{C}}_y) &= \binom{n}{2}^{-1} \sum_{i<j} I\{(\boldsymbol{y}_i - \boldsymbol{t}_*)^T \hat{\mathbf{C}}_y^{-1} (\boldsymbol{y} - \boldsymbol{t}_*) \le 0\} \\ &= \binom{n}{2}^{-1} \sum_{i<j} I\{(\boldsymbol{x}_i - \boldsymbol{t})^T \mathbf{D}^T \hat{\mathbf{C}}_y^{-1} \mathbf{D} (\boldsymbol{x}_i - \boldsymbol{t}) \le 0\}. \end{aligned}$$

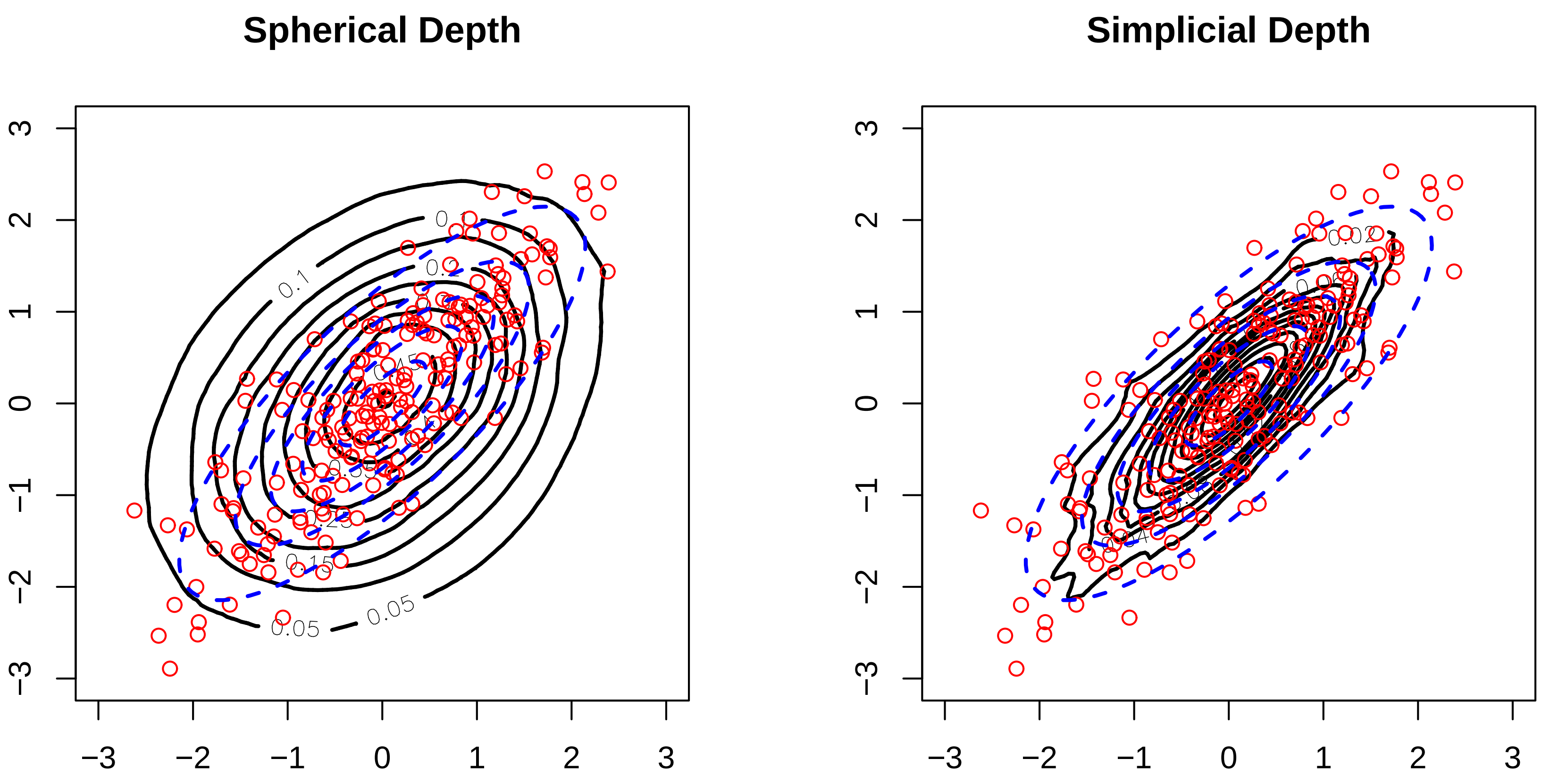
Bivariate Normal, $\rho = 0.6$, $n = 100$



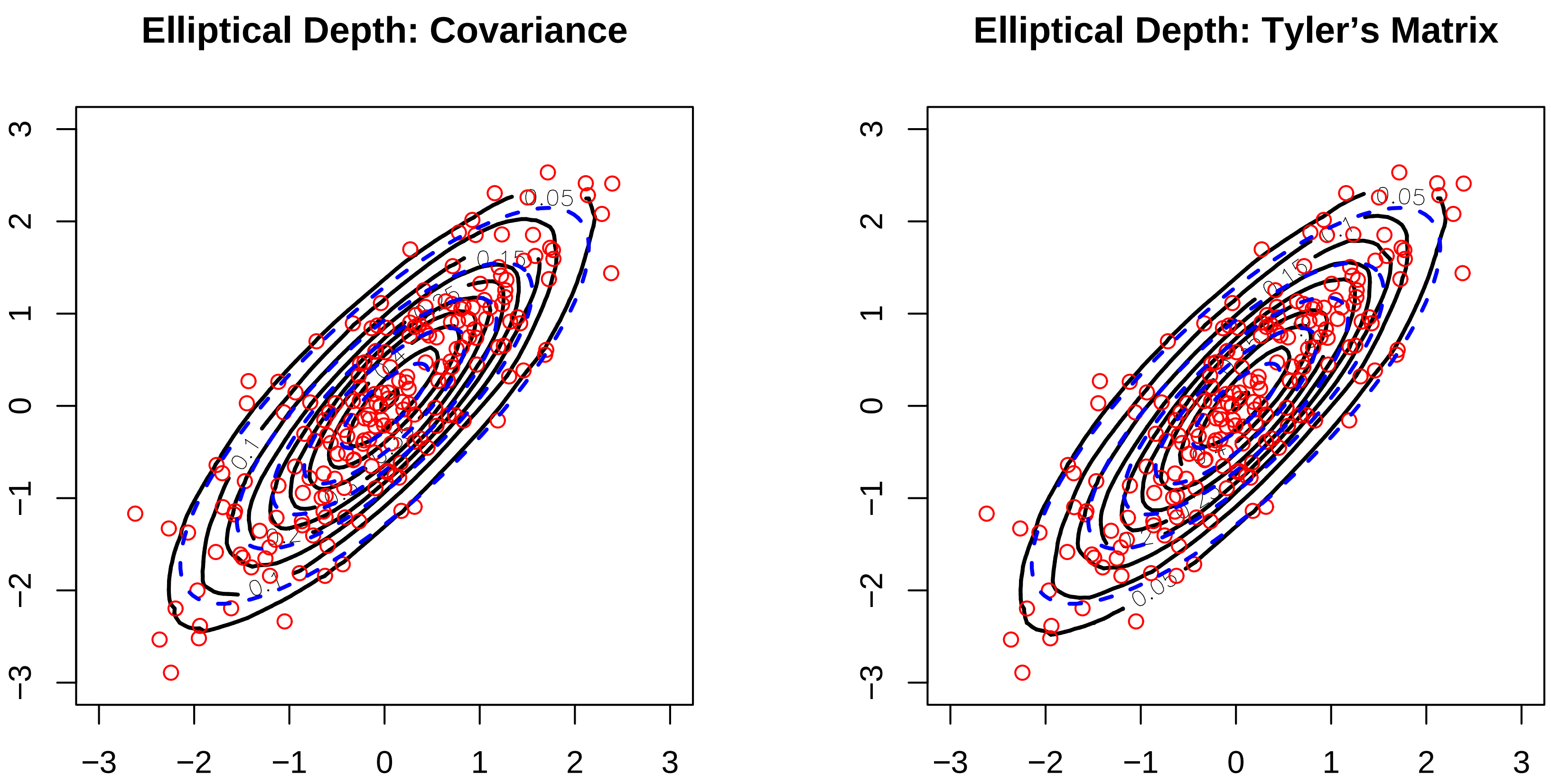
Bivariate Normal, $\rho = 0.6$, $n = 100$



Bivariate Normal, $\rho = 0.8$, $n = 100$



Bivariate Normal, $\rho = 0.8$, $n = 100$



Computation of the Contours

- Power Mac G5, 1.8 GHz, 1GB Memory, 900 MHz Bus Spd
- Each depth function was calculated at each of 10000 equally-spaced points in the grid $[-2.5, 2.5]^2$.

Depth	Time
Spherical	29.5 seconds
Elliptical	86.5 seconds
Simplicial[1]	3.47 hours
Simplicial[2]	16.5 seconds

Multivariate Median

- The elliptical depth median is defined as the point, or region of points, which maximize the elliptical depth function, i.e.

$$\boldsymbol{\theta} = \arg\max_{\boldsymbol{t}} D(\boldsymbol{t}; \mathbf{C}_F).$$

Similarly, the sample spherical median is defined by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{t}} D_n(\boldsymbol{t}; \hat{\mathbf{C}}_x).$$

- The sample elliptical depth median defined above is affine equivariant. This follows from the fact that the depth function is affine invariant.

Consistency Conjecture

Let F be an absolutely continuous distribution on $\mathbb{R}^d$ with bounded density f and scatter matrix $\mathbf{C}_F$. If $\hat{\mathbf{C}}_x$ is an affine-equivariant scatter matrix such that $\hat{\mathbf{C}}_x \to \mathbf{C}_F \ a.s.$, then the following results hold:

1. The sample elliptical depth $D_n(\boldsymbol{t}; \hat{\mathbf{C}}_x)$ is uniformly consistent in estimating $D(\boldsymbol{t}; \mathbf{C})$, i.e.,

$$\sup_{\boldsymbol{t} \in \mathbb{R}^d} |D_n(\boldsymbol{t}; \hat{\mathbf{C}}_x) - D(\boldsymbol{t}; \mathbf{C})| \xrightarrow{a.s.} 0 \quad as \ \ n \to \infty.$$

2. Furthermore, if f does not vanish in a neighborhood of $\boldsymbol{\theta}$ and if $D(\cdot; \mathbf{C}_F)$ is uniquely maximized at $\boldsymbol{\theta}$, then $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}$, as $n \to \infty$.

Notes on the Median

- Note that this objective function $D_n(\boldsymbol{t}; \hat{\mathbf{C}}_x)$ is a step function and traditional gradient-based methods are not feasible.

- Elmore, Hettmansperger, and Xuan (2004) discuss a transformation-retransformation procedure which leads to an affine-invariant spherical depth-based median. The elliptical depth essentially circumvents the need to move between the two spaces, however, the two depth functions are similar.

Example One

- The data set was originally presented in Andrews and Herzberg (1985) and presented again in Hettmansperger and Randles (2002).

- Seven skull measurements were made on a sample ($n = 50$) from the *Macropus giganteus* species of grey kangaroo.

- The measurements include basilar length, occipitonasal length, nasal length, nasal width, crest width, mandible width and mandible length.

- We computed the component sample mean ($\bar{\boldsymbol{X}}$) and median ($\hat{\boldsymbol{\theta}}_c$), an affine-equivariant median ($\hat{\boldsymbol{\theta}}_{HR}$) given in Hettmansperger and Randles (2002), the spherical median ($\hat{\boldsymbol{\theta}}_{1_a}$ and $\hat{\boldsymbol{\theta}}_{1_b}$), and the elliptical median ($\hat{\boldsymbol{\theta}}_2$).

Example One (cont.)

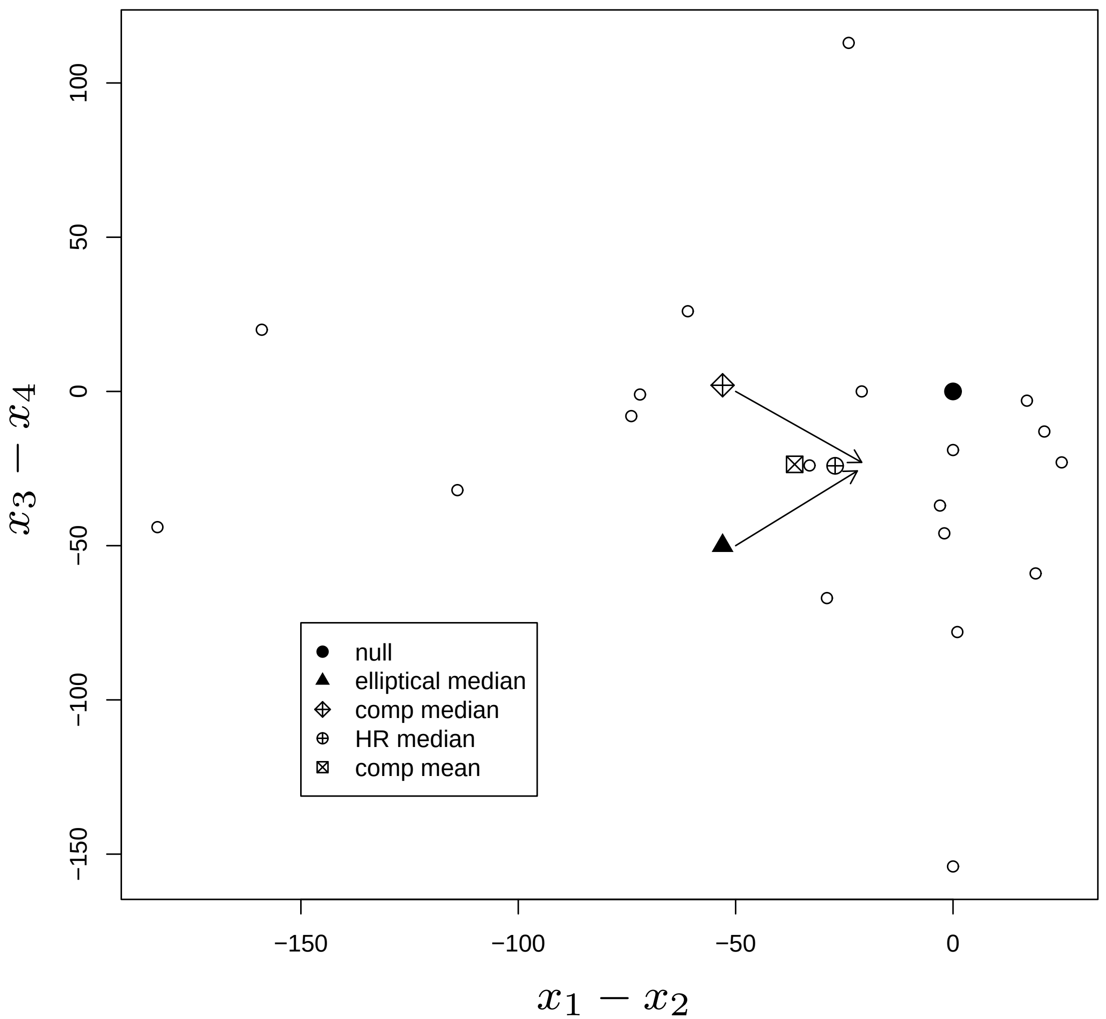
Stat	Dimension						
	I	II	III	IV	V	VI	VII
$\bar{\boldsymbol{X}}$	1491.4	1585.1	702.9	245.4	110.1	135.0	193.8
$\hat{\boldsymbol{\theta}}_c$	1490.5	1570.0	700.5	243.5	113.0	136.0	194.5
$\hat{\boldsymbol{\theta}}_{HR}$	1477.4	1572.3	694.9	243.8	111.7	134.5	192.3
$\hat{\boldsymbol{\theta}}_{1_a}$	1503.6	1578.0	703.3	245.8	104.9	134.7	197.7
$\hat{\boldsymbol{\theta}}_{1_b}$	1478.4	1572.3	693.4	243.3	115.0	134.2	191.8
$\hat{\boldsymbol{\theta}}_2$	1480.5	1575.5	695.8	246.3	110.8	134.8	192.2

Example Two

Treatment	CO_2	Halothane
1	high	N
2	low	N
3	high	Y
4	low	Y

The four treatment combinations for the sleeping-dog dataset as given in Johnson and Wichern (1992). Nineteen dogs were used in the study.

Example Two – Medians



Conclusions and Future Work

- We proposed an new statistical depth function which satisfies all of the desirable properties of a legitimate depth function *and* it is easy to compute in any dimension.
- We develop an affine-equivariant estimator of multivariate locationa based on this test.
- Completing the proofs and finding the asymptotic distribution of the test statistic.
- A multi-sample, multivariate test for location parameter.

Key References

Elmore, R.T., Hettmansperger, T.P., Xuan, F. (2004). Spherical data depth and a multivariate median. *submitted*.

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

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

Liu, R.Y., Parelius, J.M., and Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics, and inference (with discussion). *Ann. Statist.*, **27**, 783 – 858.

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

Zou, Y. and Serfling, R. (2000). General notions of statistical depth functions. *Ann. Statist.*, **28**,461 – 482.