A Measure of Directional Outlyingness with Applications to Image Data and Video

Peter J. Rousseeuw, Jakob Raymaekers and Mia Hubert

Presented by Meredith King
1. Introduction

2. Univariate measures of outlyingness

3. Multivariate and functional settings

4. Outlier detection

5. Method performance
Two main approaches for outliers:

1. Apply outlier detection method to data set and remove outliers before analysis
2. Use a data analysis method that is robust to outliers

Rousseeuw et al. propose a measure of directional outlyingness to identify outliers in univariate and multivariate functional data
Motivating data

Univariate functional data:

$Y_i(d)$ are spectra at $d = 1, \ldots, 750$ wavelengths from $n = 180$ archaeological glass samples

**Goal:** Detect outlying curves and identify how/where they are outlying
Motivating data

Multivariate functional data:

Figure: MRI image from a single patient (left) and its derivatives, horizontal (center) and vertical (right).

Figure: MRI image from a single patient (left) and its derivatives, horizontal (center) and vertical (right).
Motivating data

**Multivariate data:**

$Y_i(j, k)$ is a trivariate response:
- $j$ and $k$ reference the pixel $j = 1, \ldots, 176$ and $k = 1, \ldots, 208$
- $Y_i(j, k) = (\text{grayscale value, horizontal derivative, vertical derivative})$ for pixel $j, k$
- Sample of $n = 416$ subjects

**Goal:** Detect outlying images and identify outlying segments of images
Key contributions

- Propose outlier detection method for univariate or multivariate functional data
- Can consider functional data with a multivariate domain (ex: images, videos)
- Method robust to skewness in data
- Computationally efficient - $O(n)$ per direction
Outline

1. Introduction

2. Univariate measures of outlyingness

3. Multivariate and functional settings

4. Outlier detection

5. Method performance
Stahel-Donoho outlyingness (SDO)

For univariate data let $Y = \{y_1, \ldots, y_n\}$

$$SDO(y; Y) = \frac{|y - med(Y)|}{MAD(Y)}$$

where $MAD$ is the median average deviation:

$$MAD(Y) = \frac{med_i(|y_i - med(Y)|)}{\Phi^{-1}(0.75)}$$

**Intuition:** Center and scale each response using robust methods

**Limitation:** If “inliers” have a skewed distribution using MAD will not account for asymmetry
Directional outlyingness (DO)

Rousseeuw et al. propose directional outlyingness (DO):

- Split sample into two subsamples: those above (a) and below (b) the median
- Robustly estimate scale parameter for two subsamples
- Will account for potential skewness in data

\[
DO(y; Y) = \begin{cases} 
\frac{y - \text{med}(Y)}{S_a(Y)} & \text{if } y \geq \text{med}(Y) \\
\frac{\text{med}(Y) - y}{S_b(Y)} & \text{if } y \leq \text{med}(Y)
\end{cases}
\]

\(S_a(Y)\) and \(S_b(Y)\) are M-estimates of the scale parameters for the two subsamples
Properties of DO

- Affine invariant:
  \[ DO(-y; -Y) = DO(y; Y) \text{ OR} \]
  \[ DO(cy; cY) = DO(y, Y) \text{ for } c \neq 0 \]

- Robust to small contaminations: small deviations from parametric distributions
Outline

1. Introduction
2. Univariate measures of outlyingness
3. Multivariate and functional settings
4. Outlier detection
5. Method performance
Let \( y \) be a \( d \)-variate point from a sample \( Y = \{y_1, \ldots, y_n\} \). The multivariate DO is:

\[
DO(y; Y) = \sup_{\nu \in \mathbb{R}^d} DO(y^T \nu; Y^T \nu)
\]

**Intuition:**

- Use projections to yield a univariate measure of outlyingness
- If a point is outlying with respect to the data set it will stand out in at least one direction
Multivariate DO in practice

- Impossible to calculate $DO(y^T \nu; Y^T \nu)$ for all directions $\nu$ in a $d$-dimensional space

Solution:
1. Randomly select $d$ data points
2. Compute hyperplane passing through the points
3. Take direction $\nu$ orthogonal to it
4. Repeat Steps 1-3 $250d$ times to yield set of directions
5. Compute multivariate DO for each $i$

Multivariate DO is also affine invariant:
It doesn’t change when adding constant vector to data or multiplying by non-singular $d \times d$ matrix
Let $Y = \{Y_1, \ldots, Y_n\}$ where each $Y_i$ is a function on a discrete set of time points $\{t_1, \ldots, t_T\}$

The **functional directional outlyingness (fDO)** is:

$$fDO(Y_i : Y) = \sum_{j=1}^{T} DO(Y_i(t_j); Y(t_j))W(t_j)$$

where $W(\cdot)$ is a weight function where $\sum_{j=1}^{T} W(t_j) = 1$

- We can weight the importance of different time points
- If we have multivariate functional data we use the Multivariate DO based on projections
fDO of glass data

Figure: $n = 180$ glass functions.

Figure: fDO of $n = 180$ glass functions.
DO of glass data

Figure: Heatmap of DO at all wavelengths.

Presented by M. King
Outline

1. Introduction
2. Univariate measures of outlyingness
3. Multivariate and functional settings
4. Outlier detection
5. Method performance
Cutoff for DO or fDO

- The $DO_i$ and $fDO_i$ will have a right skewed distribution
- Take a log-transformation to yield a more Gaussian distribution

$$\{LDO_1, \ldots, LDO_n\} = \{\log(DO_1 + 0.1), \ldots, \log(DO_n + 0.1)\}$$

- Then robustly center and normalize
- Flag $y_i$ as an outlier if

$$\frac{LDO_i - \text{med}(LDO)}{\text{MAD}(LDO)} > \Phi^{-1}(0.995)$$

- Same process can be applied to the $fDO_i$
Extend functional outlier map of Hubert et al. 2015 to DO and propose cutoff

The fDO for curve $Y_i$ could be considered its ‘average outlyingness’

We need a measure for the variability of its outlyingness

$$vDO(Y_i; Y) = \frac{SD_j(\text{DO}(Y_i(t_j), Y(t_j)))}{1 + fDO(Y_i; Y)}$$

They divide by $1 + fDO(Y_i; Y)$ to yield relative variability

The FOM is a scatter plot of points

$$(fDO(Y_i; Y), vDO(Y_i; Y)) \quad i = 1, \ldots, n$$
Define the combined functional outlyingness (CFO) of \( Y_i \) as:

\[
CFO_i = CFO(Y_i : Y) = \sqrt{\left(\frac{fDO_i}{\text{med}(fDO)}\right)^2 + \left(\frac{vDO_i}{\text{med}(vDO)}\right)^2}
\]

- CFO quantifies the distance of FOM points from the origin
- Large values of \( CFO_i \) correspond to outliers
- The \( CFO_i \) will also be right skewed so again take the log-transformation to get a cutoff
- \( Y_i \) is an outlier if:

\[
\frac{L\text{CFO}_i - \text{med}(L\text{CFO})}{\text{MAD}(L\text{CFO})} > \Phi^{-1}(0.995)
\]
FOM for glass data

Figure: fDO of $n = 180$ glass functions.

Figure: FOM of $n = 180$ glass functions.
Location of FOM points can indicate the type of outlier:

- Lower left points correspond to most central curves
- Lower right points correspond to high $fDO$ but low variability in $DO$ values
  - Correspond to *shift outliers*
  - Same shape but shifted on domain
- Upper left points have low $fDO$ but high $vDO$
  - Correspond to *local outliers*
  - Only outlying on small part of domain
- Upper right points have high $fDO$ and high $vDO$
  - Strong outliers on a large portion of the domain
Introduction

Univariate measures of outlyingness

Multivariate and functional settings

Outlier detection

Method performance
Image data application

- \( Y_i(j, k) = (\text{grayscale value, horizontal derivative, vertical derivative}) \) for pixel \( j, k \)
- \( j = 1, \ldots, 176, k = 1, \ldots, 208 \), and \( n = 1, \ldots, 416 \)

The fDO of MRI image \( Y_i \) is:

\[
fDO(Y_i; Y) = \frac{1}{176 \times 208} \sum_{j=1}^{176} \sum_{k=1}^{208} DO(Y_i(j, k); Y(j, k)) W_{jk}
\]

- \( DO(Y_i(j, k); Y(j, k)) = \sup_{\nu \in \mathbb{R}^3} DO(Y_i^T \nu; Y^T \nu) \)
- \( W_{jk} = 0 \) if pixels not part of the brain
Image data application
Image data application

- Subject 387 (top) has lowest CFO and is most ‘central’ image in dataset
- Subject 92 (middle) has high \( fDO \) and \( vDO \) - large outlying regions
- Subject 126 (bottom) has high \( fDO \) but low \( vDO \) - shifted outlier
Propose an outlier detection method for univariate or multivariate functional data
Can accommodate potentially skewed data
Simulation studies show their method is more powerful than competing Adjusted Outlyingness (AO) Brys et al. (2005)

Limitations:
Method may not account for/exploit correlation within curves
Dai & Genton (2017) demonstrate cases where the functional outlier map has poor detection power
Dai & Genton propose a similar method and graphical tool

They use multivariate SDO as measure of directional outlyingness:

\[
SDO(Y_i(t); Y(t)) = \sup_{\|u\|=1} \frac{\|u^T Y_i(t) - med(u^T Y(t))\|}{MAD(u^T Y(t))}
\]

Their definition of directional outlyingness is

\[
O(Y_i(t); Y(t)) = SDO(Y_i(t); Y(t)) \cdot v_i(t)
\]

where \(v_i(t) = \frac{Y_i(t) - Z(t)}{\|Y_i(t) - Z(t)\|}\) and \(Z(t)\) is the unique median as defined by the SDO.
Competing Method: Dai & Genton (2017)

Mean directional outlyingness:

\[ MO(Y_i; Y) = \int_{I} O(Y_i(t); Y(t))w(t)dt \]

Variation of directional outlyingness:

\[ VO(Y_i; Y) = \int_{I} \| O(Y_i(t); Y(t)) - MO(Y_i; Y) \|^2w(t)dt \]

They introduce the magnitude-shape plot which plots:

\( (MO^T, VO)^T \)
Simulations - Univariate example
Simulations - Multivariate example

Presented by M. King
Appendix: Robust scale estimates

For univariate data $y_1 \leq y_2 \leq \ldots \leq y_n$ define $Y_a = \{y_{h+1}, \ldots, y_n\}$ and $Y_b = \{y_1, \ldots, y_h\}$ where $h = \left\lfloor \frac{n + 1}{2} \right\rfloor$

$Z_a = Y_a - \text{med}(Y)$ and $Z_b = \text{med}(Y) - Y_b$

They compute initial scale estimates:

$$s_{o,a}(Y) = \frac{\text{med}(Z_b)}{\Phi^{-1}(0.75)} \quad s_{o,b}(Y) = \frac{\text{med}(Z_b)}{\Phi^{-1}(0.75)}$$
Appendix: Robust scale estimates

The M-estimates are then:

\[ s_a(Y) = s_{o,a}(Y) \sqrt{\frac{1}{2\alpha h} \sum_{z_i \in Z_a} \rho_c \left( \frac{z_i}{s_{o,a}(Y)} \right)} \]

\[ s_b(Y) = s_{o,b}(Y) \sqrt{\frac{1}{2\alpha h} \sum_{z_i \in Z_b} \rho_c \left( \frac{z_i}{s_{o,b}(Y)} \right)} \]

where

- \( \rho_c \) is the Huber rho function for scale \( \rho_c(t) = \left( \frac{t}{c} \right)^2 I_{[-c,c]} + I_{(-\infty,c) \cup (c,\infty)} \)
- \( c \) is a tuning parameter regulating bias/efficiency trade-off
- \( \alpha = \int_0^\infty \rho_c(x) d\Phi(x) \)