Review of Robust Location and Scatter Estimators

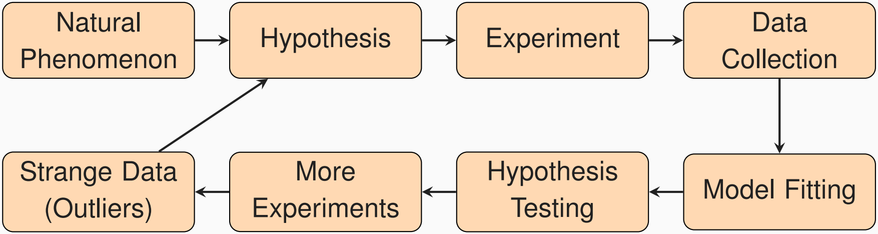
Subhrajyoty Roy †
Supervisors: Prof. Ayanendranath Basu and Dr. Abhik Ghosh
February, 2024

† External Research Fellow
Indian Statistical Institute, Kolkata

OUTLINE

Introduction

Science built of uncertainty modelling



An **outlier** is an observation that deviates from the fit suggested by the majority of the data.

WHY NEED ROBUST ESTIMATORS?

1. Statistical models are built on assumptions.
2. Assumptions are often approximations of reality.
   2.1 Fuzzy knowledge.
   2.2 Outliers, part of data that differs significantly from the majority of it.
3. Robust Statistical Inference builds methods that are resistant.
4. Achieves enough statistical guarantee even when assumptions fail to be met.

Robustness in Multivariate Context - 1

Consider a data matrix

$$\boldsymbol{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \dots & \dots & \dots \\ 13 & 14 & 15 \end{bmatrix}$$

The singular values of $\boldsymbol{X}$ are $35.18, 1.47$ and 0.

Robustness in Multivariate Context - 1

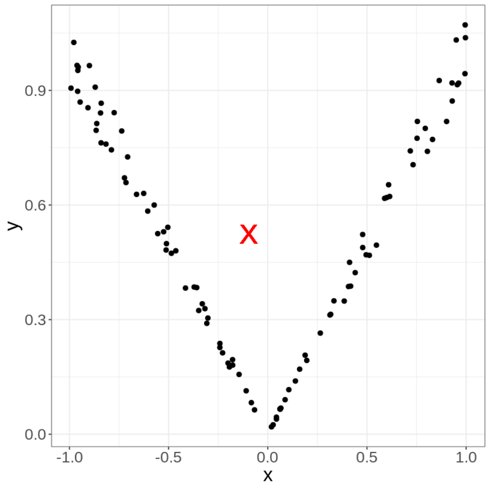
Consider a data matrix

$$X = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \dots & \dots & \dots \\ 13 & 14 & 15 \end{bmatrix}$$

The singular values of X are $35.18, 1.47$ and 0.
Modification of a single entry $X_{11} = 100$ makes the singular values $102.07, 28.62$ and 0.46.

Robustness in Multivariate Context - 2

Coordinatewise **median** is not a good solution!



How?

Huber defines three desirable features that every robust procedure should achieve.

1. **Stability** of the estimator under small deviations from the assumed model.

How?

Huber defines three desirable features that every robust procedure should achieve.

1. **Stability** of the estimator under small deviations from the assumed model.
2. **Efficiency** under the assumed model.

How?

Huber defines three desirable features that every robust procedure should achieve.

1. **Stability** of the estimator under small deviations from the assumed model.
2. **Efficiency** under the assumed model.
3. **Breakdown**-resistance under large amount of contamination.

General Model

Let, $X_1, X_2, \ldots X_n$ be an i.i.d sample from F_θ, for some unknown $\theta \in \Theta$.
We wish to make inference about θ.
and let, $T(X_1, \ldots X_n)$ be the proposed estimator of θ.
We denote $T : \mathcal{F} \to \mathbb{R}$ as the corresponding functional, i.e.,

$$T(F_n) = T(X_1, \ldots X_n),$$

General Model

Let, $X_1, X_2, \ldots X_n$ be an i.i.d sample from F_θ, for some unknown $\theta \in \Theta$.
We wish to make inference about θ.
and let, $T(X_1, \ldots X_n)$ be the proposed estimator of θ.
We denote $T : \mathcal{F} \to \mathbb{R}$ as the corresponding functional, i.e.,

$$T(F_n) = T(X_1, \ldots X_n),$$

Examples:

1. For sample mean, $T_1(F) = \int x dF$.
2. For sample median, $T_2(F) = F^{-1}(1/2)$.

Fisher Consistency

An estimator is **Fisher consistent** if

$$T(F_\theta) = \theta,$$

for all $\theta \in \Theta$.

FISHER CONSISTENCY

An estimator is **Fisher consistent** if

$$T(F_\theta) = \theta,$$

for all $\theta \in \Theta$.

Examples:

1. $T(x) = c$, constant **is not** Fisher consistent.
2. $T(x) = \left(\int (x - \int x dF)^2 dF\right)^{1/2}$ **is** Fisher consistent for standard deviation, **but not** unbiased.

Influence Function

Let, $F_\epsilon = (1-\epsilon)F_\theta + \epsilon\delta_x$.
Let, $Y_1, \ldots Y_n \sim F_\epsilon$.

$$\text{Empirical } IF(x; T, F_\theta) = \lim_{\epsilon \to 0+} \frac{T(Y_1, \ldots Y_n) - T(X_1, \ldots X_n)}{\epsilon}$$

In terms of functional,

$$IF(x; T, F) = \lim_{\epsilon \to 0+} \frac{T((1-\epsilon)F + \epsilon\delta_x) - T(F)}{\epsilon}$$

Asymptotic Variance

Using Von Mises expansion, one can write

$$T(F_n) = T(F) + \int IF(x; T, F) dF_n(x) + \text{remainder}$$

$$\Rightarrow \quad \sqrt{n}(T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i, T, F) + \text{remainder}$$

Under standard regularity assumptions,

$$\Rightarrow \quad \sqrt{n}(T(F_n) - T(F)) \xrightarrow{d} \mathcal{N}\left(0, \int IF(x; T, F)^2 dF(x)\right)$$

Asymptotic Variance

Using Von Mises expansion, one can write

$$T(F_n) = T(F) + \int IF(x; T, F)dF_n(x) + \text{remainder}$$

$$\Rightarrow \quad \sqrt{n}(T(F_n) - T(F)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(X_i, T, F) + \text{remainder}$$

Under standard regularity assumptions,

$$\Rightarrow \quad \sqrt{n}(T(F_n) - T(F)) \xrightarrow{d} \mathcal{N}\left(0, \int IF(x; T, F)^2 dF(x)\right)$$

Therefore, the asymptotic variance of centered and normalized estimator $T(F_n)$ is

$$V(T, F) = \int IF(x; T, F)^2 dF(x)$$

Breakdown Point

Example: Given a sample $X_1, \ldots X_n$ and the estimator sample mean $\bar{X}$, one can modify X_1 to make $\bar{X}$ as large (or as small) as required. In other words, one can break its reliability by contaminating only one sample.

Breakdown Point

Example: Given a sample $X_1, \ldots X_n$ and the estimator sample mean $\bar{X}$, one can modify X_1 to make $\bar{X}$ as large (or as small) as required. In other words, one can break its reliability by contaminating only one sample.

$$\epsilon_n^*(T; x_1, x_2, \ldots x_n) = \frac{1}{n} \max\{m : \max_{i_1,\ldots i_m} \sup_{Y_1,\ldots Y_m} T(Z_1, Z_2, \ldots Z_n) < \infty\}$$

where $Z_l = X_l$ if $l \notin \{i_1, \ldots i_m\}$ and $Z_l = Y_j$ if $l = i_j$.

Breakdown Point

Example: Given a sample $X_1, \ldots X_n$ and the estimator sample mean $\bar{X}$, one can modify X_1 to make $\bar{X}$ as large (or as small) as required. In other words, one can break its reliability by contaminating only one sample.

$$\epsilon_n^*(T; x_1, x_2, \ldots x_n) = \frac{1}{n} \max\{m : \max_{i_1, \ldots i_m} \sup_{Y_1, \ldots Y_m} T(Z_1, Z_2, \ldots Z_n) < \infty\}$$

where $Z_l = X_l$ if $l \notin \{i_1, \ldots i_m\}$ and $Z_l = Y_j$ if $l = i_j$.
Asymptotic Breakdown point is the limit of ϵ_n^* as $n \to \infty$.

Existing Robust Location and Covariance Estimators

MODEL

Let, $X_1, X_2, \ldots X_n$ be an i.i.d sample $\sim F(\mu, \Sigma)$, each $X_i \in \mathbb{R}^p$.
Also, assume $\mathbb{E}(X_i) = \mu$ and Var$(X_i) = \Sigma$.
Both μ and Σ are unknown.
We want to estimate these robustly, so that even if some of the X_is come from G, such that $G \approx F$, our estimate won't change rapidly.

Trimmed and Winsorized estimator

Let, $z_1, z_2, \ldots z_n$ be univariate samples.
Since, usual mean is nonrobust, as it has unbounded influence function.
We wish to have a bounded influence for each sample z_i.
Let, $z_{(1)} < z_{(2)} < \ldots z_{(n)}$ be the order statistics.

1. α-Trimmed Estimator:

$$\bar{z}_\alpha = \frac{1}{(n - 2[n\alpha])} \sum_{i=(1+[n\alpha])}^{n-[n\alpha]} z_{(i)}$$

Trimmed and Winsorized estimator

Let, $z_1, z_2, \ldots z_n$ be univariate samples.
Since, usual mean is nonrobust, as it has unbounded influence function.
We wish to have a bounded influence for each sample z_i.
Let, $z_{(1)} < z_{(2)} < \ldots z_{(n)}$ be the order statistics.

1. α-Trimmed Estimator:

$$\bar{z}_\alpha = \frac{1}{(n - 2[n\alpha])} \sum_{i=(1+[n\alpha])}^{n-[n\alpha]} z_{(i)}$$

2. α-Winsorized Estimator:

$$\bar{z}_\alpha^* = \frac{1}{n}\left([n\alpha] z_{(1+[n\alpha])} + \sum_{i=(1+[n\alpha])}^{n-[n\alpha]} z_{(i)} + [n\alpha] z_{(n-[n\alpha])}\right)$$

Multivariate Trimmed and Winsorized estimator

Concept of order statistic is complicated! Requires data depth! Bickel (1965) introducted two multivariate estimators in similar direction.

1. λ-Trimmed Estimator:

$$\bar{X}_\lambda(\widehat{\theta}) : \widehat{\theta} = \frac{1}{\sum_{i=1}^n \mathbf{1}(\|X_i - \widehat{\theta}\| < \lambda)} \sum_{i=1}^n X_i \mathbf{1}(\|X_i - \widehat{\theta}\| < \lambda)$$

Multivariate Trimmed and Winsorized estimator

Concept of order statistic is complicated! Requires data depth! Bickel (1965) introducted two multivariate estimators in similar direction.

1. λ-Trimmed Estimator:

$$\bar{X}_\lambda(\widehat{\theta}) : \widehat{\theta} = \frac{1}{\sum_{i=1}^{n} \mathbf{1}(\|X_i - \widehat{\theta}\| < \lambda)} \sum_{i=1}^{n} X_i \mathbf{1}(\|X_i - \widehat{\theta}\| < \lambda)$$

2. λ-Winsorized Estimator:

$$\bar{X}^*_\lambda(\widehat{\theta}) : \widehat{\theta} = \frac{1}{n}\left(\sum_{i=1}^{n}\left(\widehat{\theta} + \lambda \frac{(X_i - \widehat{\theta})}{\|X_i - \widehat{\theta}\|}\right)\mathbf{1}(\|X_i - \widehat{\theta}\| \geq \lambda) + \sum_{i=1}^{n} X_i \mathbf{1}(\|X_i - \widehat{\theta}\| < \lambda)\right)$$

Scatter estimates follow from sample covariance matrix of trimmed / winsorized samples.

M-ESTIMATOR

Consider MLE, we wish to maximize the log likelihood $\max_{\mu,\Sigma} n^{-1} \sum_{i=1}^{n} \ell(\mu, \Sigma; X_i)$.

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \mu}(X_i) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \ell}{\partial \Sigma}(X_i) = 0$$

Here, each sample X_i has the same contribution, irrespective of its deviation from the model.

M-ESTIMATOR

Score $\ell'(X)$ is unbounded in X, hence any one sample X_i can make things problematic!
We need to introduce bounded functions!

M-ESTIMATOR

Score $\ell'(X)$ is unbounded in X, hence any one sample X_i can make things problematic!

We need to introduce bounded functions!

General estimating equation,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \rho}{\partial \mu}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \rho}{\partial \Sigma}(X_i) = 0$$

If ρ' are bounded, then effect of every sample X_i is bounded, hence no one point can influence the estimator to behave erratically.

M-ESTIMATOR

Maronna (1976) proposed M-estimators of location and scatter from the same idea, as the solution of the estimating equations

$$\frac{1}{n}\sum_{i=1}^{n}\Psi_1\left((X_i-\widehat{\mu})^\intercal\widehat{\Sigma}^{-1}(X_i-\widehat{\mu})\right)(X_i-\widehat{\mu})=0$$

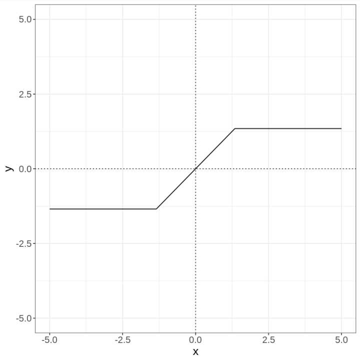
$$\frac{1}{n}\sum_{i=1}^{n}\Psi_2\left((X_i-\widehat{\mu})^\intercal\widehat{\Sigma}^{-1}(X_i-\widehat{\mu})\right)(X_i-\widehat{\mu})(X_i-\widehat{\mu})^\intercal=\widehat{\Sigma}$$

where Ψ_1, Ψ_2 are two bounded functions with both decreasing in absolute value of its argument.

Choice of Ψ functions

- Huber's Psi,

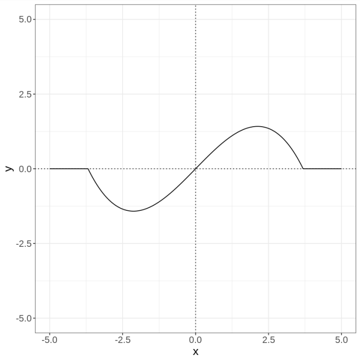
$$\psi(x) = \begin{cases} x & |x| \leq k \\ k\ \text{sign}(x) & |x| > k \end{cases}$$



Choice of Ψ functions

- Tukey's bisquare

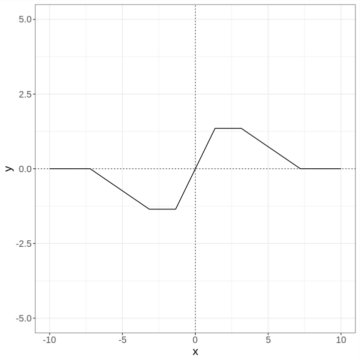
$$\psi(x) = x\left(1 - (x/k)^2\right)^2 \mathbf{1}_{\{|x| \leq k\}}$$



CHOICE OF Ψ FUNCTIONS

- Hampel's piecewise linear

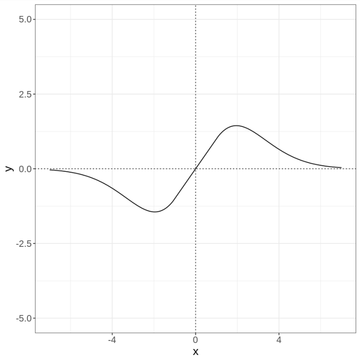
$$\psi(x) = \begin{cases} x & |x| \leq a \\ a \,\text{sign}(x) & a < |x| \leq b \\ a \,\text{sign}(x) \frac{r-|x|}{r-b} & b < |x| \leq r \\ 0 & |x| > r \end{cases}$$



CHOICE OF Ψ FUNCTIONS

- Generalized Gauss Weight

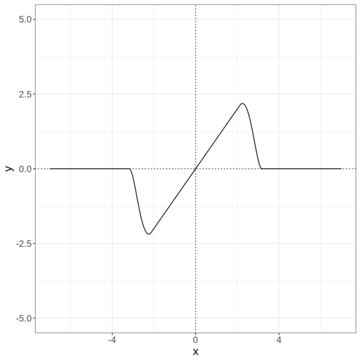
$$\psi(x) = \begin{cases} x & |x| \leq c \\ \exp\left(-\frac{1}{2}\frac{(|x|-c)^b}{a}\right) & |x| > c \end{cases}$$



CHOICE OF Ψ FUNCTIONS

- Maronna's psi

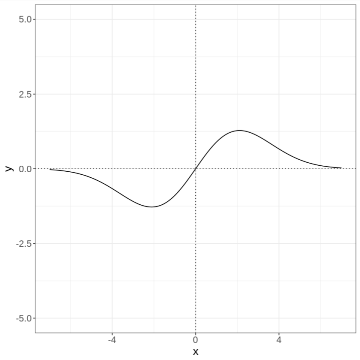
$$\psi(x) = \text{sign}(x)\max\left\{0, -\frac{\phi'(|x|)+c}{\phi(|x|)}\right\}$$



Choice of Ψ functions

- Welsh's psi

$$\psi(x) = xe^{-(x/k)^2/2}$$



Desirable Properties of Robust Location Estimate

Rousseeuw (1985) described two desirable properties for estimator of location for a multivariate sample.
Let, $T(X_1, \ldots X_n)$ be the estimator of location of the samples $X_1, \ldots X_n$, each $X_i \in \mathbb{R}^p$.

1. Breakdown point $\epsilon(T, X)$ should be large, ”close” to $1/2$.

Desirable Properties of Robust Location Estimate

Rousseeuw (1985) described two desirable properties for estimator of location for a multivariate sample.
Let, $T(X_1, \ldots X_n)$ be the estimator of location of the samples $X_1, \ldots X_n$, each $X_i \in \mathbb{R}^p$.

1. Breakdown point $\epsilon(T, X)$ should be large, ”close” to $1/2$.
2. It should be affine equivariant, i.e., for any $b \in \mathbb{R}^p$ and nonsingular matrix A,

$$T(AX_1 + b, \ldots AX_n + b) = AT(X_1, \ldots X_n) + b$$

THREE COMPETING ESTIMATORS?

Sample mean $\bar{X}$

1. Breakdown at 0.
2. Affine equivariant.

THREE COMPETING ESTIMATORS?

Sample mean $\bar{X}$

1. Breakdown at 0.
2. Affine equivariant.

L_1 median

$$\min_{a \in \mathbb{R}^p} \sum_{i=1}^{n} \|X_i - a\|_{L_1}$$

1. Breakdown at $1/2$.
2. Not affine equivariant.

THREE COMPETING ESTIMATORS?

Sample mean $\bar{X}$

1. Breakdown at 0.
2. Affine equivariant.

L_1 median

$$\min_{a\in\mathbb{R}^p}\sum_{i=1}^{n}\|X_i - a\|_{L_1}$$

1. Breakdown at $1/2$.
2. Not affine equivariant.

M-estimator

1. Breakdown at $1/(p+1)$.
2. Affine equivariant.

MVE Estimator

No! we are not trading affine equivariance and breakdown.
Define,

$$T(X_1, \ldots X_n) = \text{Center of the minimum volume ellipsoid}$$
$$\text{containing at least } h \text{ points of } X_1, \ldots X_n$$

Usually, $h = [n/2] + 1$.

MVE Estimator

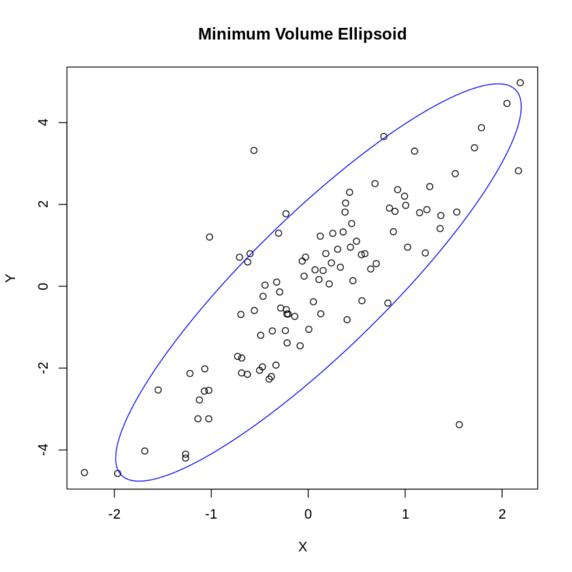
No! we are not trading affine equivariance and breakdown.
Define,

$$T(X_1, \ldots X_n) = \text{Center of the minimum volume ellipsoid containing at least } h \text{ points of } X_1, \ldots X_n$$

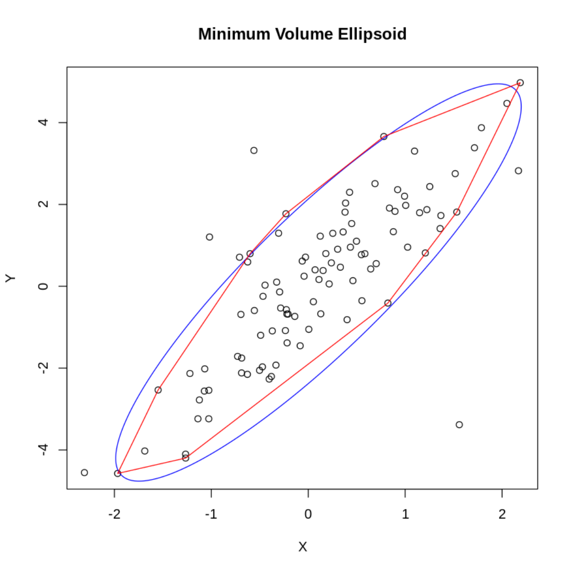
Usually, $h = [n/2] + 1$.

1. Clearly, affine equivariant.
2. Breakdown at $([n/2] - p + 1)/n \approx 1/2$ for large n.
3. Cannot be computed if $p > [n/2] + 1$.

MVE Estimator is NP-hard to solve



MCD ESTIMATOR



MCD Estimator

Instead, we consider the confidence ellipsoid's volume, for Gaussian distribution. Define,

$T(X_1, \ldots X_n) =$ Mean of the h points among $X_1, \ldots X_n$

such that determinant of covariance matrix is minimal.

Usually, $h = [n/2] + 1$.

MCD ESTIMATOR

Instead, we consider the confidence ellipsoid's volume, for Gaussian distribution. Define,

$T(X_1, \ldots X_n) =$ Mean of the h points among $X_1, \ldots X_n$
such that determinant of covariance matrix is minimal.

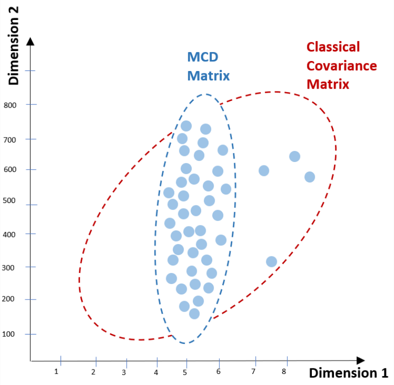
Usually, $h = [n/2] + 1$.

1. Clearly, affine equivariant.
2. Breakdown at $([n/2] - p + 1)/n \approx 1/2$ for large n.
3. Cannot be computed if $p > [n/2] + 1$.
4. Easier to solve by taking convex hull.

LOCATION AND SCATTER ESTIMATOR

Once we identify the best h points through MVE or MCD estimator,

1. Robust estimate of location is sample mean of those best h points.
2. Robust estimate of scatter is sample covariance of those best h points.



S-ESTIMATOR

Introduced by Rousseeuw (1984), develops from a regression problem.
Assume that the data comes from a distribution F, $X_1, \ldots X_n \sim F$.

$$\frac{1}{n}\sum_{i=1}^{n} E_F\left[\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = 1$$

For a robust estimate, we might want

$$\frac{1}{n}\sum_{i=1}^{n} E_F\left[\left|\frac{X_i-\mu}{\sigma}\right|\right] = K'$$

In general, we have scale estimator.

$$\frac{1}{n}\sum_{i=1}^{n} E_F\left[\rho\left((X_i-\mu)^{\intercal}\Sigma^{-1}(X_i-\mu)\right)\right] = K$$

S-ESTIMATOR

The function $\rho(\cdot)$ satisfy,

1. $\rho(0) = 0$.
2. $\rho(\cdot)$ is symmetric about 0 and is increasing in magnitude.
3. It is bounded.

Davies (1987) improvised it to an optimization problem

$$\text{Minimize } \det(\Sigma) \text{ subject to } \sum_{i=1}^{n} \rho\left((X_i - \mu)^\intercal \Sigma^{-1} (X_i - \mu)\right) \leq K$$

it has much better properties than Rousseeuw's version of S-estimator.

STAHEL-DOHONO ESTIMATOR

Usual nonrobust location and scatter estimators,

$$\widehat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

and

$$\widehat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu}_1)(X_i - \widehat{\mu}_1)^\intercal$$

The problem is each X_i has same influence, irrespective of its deviation from the model.

Stahel-Dohono Estimator

Idea is to introduce weights!

$$\widehat{\mu}_2 = \frac{\sum_{i=1}^{n} w(X_i) X_i}{\sum_{i=1}^{n} w(X_i)}$$

and

$$\widehat{\Sigma}_2 = \frac{\sum_{i=1}^{n} w(X_i)(X_i - \widehat{\mu}_2)(X_i - \widehat{\mu}_2)^\intercal}{\sum_{i=1}^{n} w(X_i)}$$

Here, $w(X_i)$s are such that for points with high deviation from the assumed model, they are small.

Stahel-Dohono Estimator

How to choose these weights?

Look for some one-dimensional direction in which X_i is most outside from the data cloud.

$$w(X_i) \propto \text{Maximum value of } \boldsymbol{u}^\intercal X_i \text{ subject to } \|\boldsymbol{u}\| = 1$$

Usually, we should center and scale this before projecting onto $\boldsymbol{u}$,

$$w(X_i) = \sup_{\|u\|=1} \frac{\boldsymbol{u}^\intercal X_i - \text{med}_{1\le j\le n}\boldsymbol{u}^\intercal X_j}{\text{med}_k|\boldsymbol{u}^\intercal X_k - \text{med}_{1\le j\le n}\boldsymbol{u}^\intercal X_j|}$$

Comparison of Existing Estimators

Comparison of Estimators

Method	Affine Equivariance	Asymptotic Breakdown Point	Asymptotic Property	Computational Complexity	Assumption on dimensionality
W-estimator	Yes	α	$\sqrt{n}$-consistent Asymptotic Normal Less efficient for high α	O(np^3) / iteration	None
MVE	Yes	1/2	$n^{1/3}$-consistent Not asymptotic normal	NP-hard Not known	$p \leq [n/2]+1$
MCD	Yes	1/2	$\sqrt{n}$-consistent Asymptotic Normal Not very efficient	O($n^{\min(p^2,h)} \log(n)$)	$p \leq [n/2]+1$
M-estimator	Depends on $\Psi(\cdot)$	$\leq \frac{1}{(p+1)}$, if AE Depends on $\Psi(\cdot)$ Could be 1/2 for some Ψ	$\sqrt{n}$-consistent Asymptotic Normal Very efficient	O($np^3 f(n,p)$) / iteration	Usually $p < n$
S-estimator	Depends on $\rho(\cdot)$	Depends on $\rho(\cdot)$ Higher than similar M-estimator.	$\sqrt{n}$-consistent Asymptotic Normal Very efficient, but less than M-estimator	O($np^3 f(n,p)$) / iteration	$p < n$
Stahel Donoho Estimator	Yes	1/2	$\sqrt{n}$-consistent Asymptotic distribution not known	O($n^2 p$)	None

Application

Example with Real Dataset

Diabetes Dataset:

1. Five measurement of 145 adult patients.
2. Variables are
   2.1 Relative weight.
   2.2 Fasting plasma glucose.
   2.3 Oral Glucose.
   2.4 Insulin Resistance.
   2.5 Steady State Plasma Glucose (SSPG).
3. Three groups: Normal, Chemical Diabetic, Obese.
4. Reference: Reaven, G. M. and Miller, R. G. (1979); An attempt to define the nature of chemical diabetes using a multidimensional analysis.

Example with Real Dataset

For univariate samples $X_1, \ldots X_n$, one simple measure of outlyingness is the z-score.

$$Z_i = \frac{(X_i - \widehat{\mu})}{\widehat{\sigma}}, \; i = 1, 2, \ldots n$$

Example with Real Dataset

For univariate samples $X_1, \ldots X_n$, one simple measure of outlyingness is the z-score.

$$Z_i = \frac{(X_i - \widehat{\mu})}{\widehat{\sigma}},\ i = 1, 2, \ldots n$$

For multivariate samples $X_1, \ldots X_n$ with each $X_i \in \mathbb{R}^p$, we can look at squared Mahalanobis distance

$$Z_i = (X_i - \widehat{\mu})^{\mathsf{T}} \widehat{\Sigma}^{-1} (X_i - \widehat{\mu}),\ i = 1, 2, \ldots n$$

Example with Real Dataset

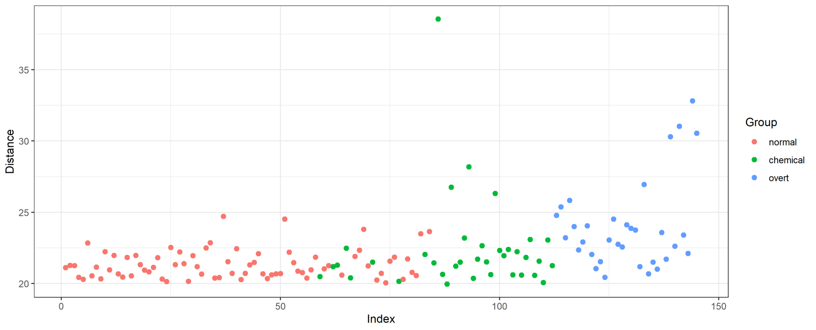


Figure 1: Classical Mahalanobis distance

Example with Real Dataset

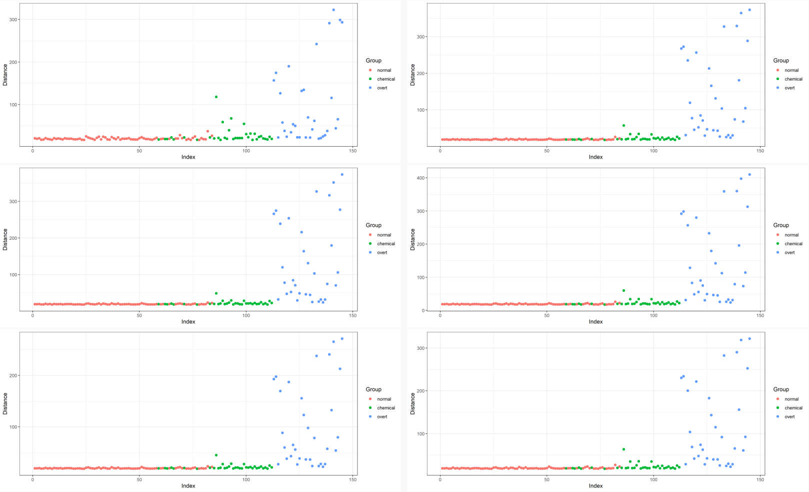


Figure 2: From Top-Left to Bottom-Right: Mahalanobis distance with W-estimator, MCD, MVE, M-estimator, S-estimator, Stahel Donoho estimator.

Background Modelling Problem

Background Modelling Problem

=

+

Background Modelling Problem

=

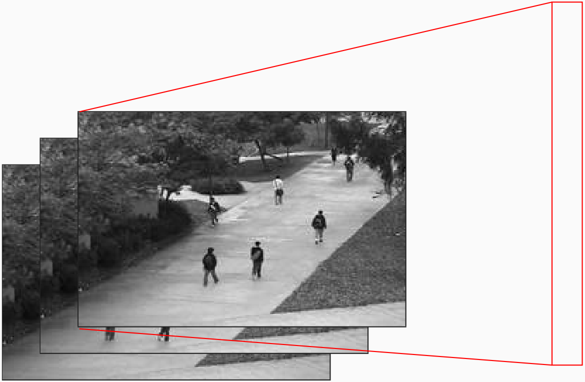
+

Applications ranging security, defence, object tracking, motion detection, video filters, etc.

Background Modelling as Low Rank Decomposition

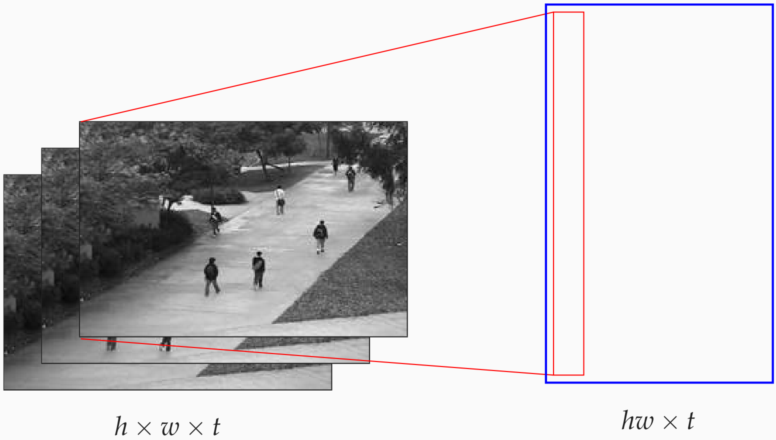
$h \times w \times t$

Background Modelling as Low Rank Decomposition

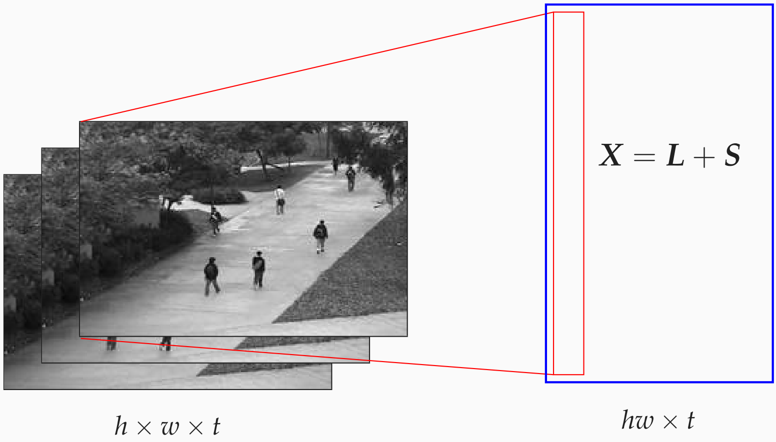


$h \times w \times t$

Background Modelling as Low Rank Decomposition



Background Modelling as Low Rank Decomposition



RESULTS

Truth OP ADMM GoDec rSVDdpd

CONCLUSION

1. Robust Statistical Inference can help with the identification of outliers.
2. It is a great way to deal with current high-dimensional datasets without worrying about outliers, where identifying outliers would be challenging.
3. For $n < p$, Stahel Donoho estimator should be the first choice as robust location and scatter estimates.
4. For $n > p$, M-estimators can be a quick robust estimator, whereas MVE and MCD estimators should be used if high breakdown is required.

References

1. Maronna, R. A., Martin, R. D., Yohai, V. J., & Salibián-Barrera, M. (2019). Robust statistics: theory and methods (with R). John Wiley & Sons.
2. Huber, P. J. (2004). Robust statistics (Vol. 523). John Wiley & Sons.
3. Basu, A., Shioya, H., & Park, C. (2011). Statistical inference: the minimum distance approach. CRC press.
4. Roy, S., Basu, A., & Ghosh, A. (2021). rSVDdpd: A Robust Scalable Video Surveillance Background Modelling Algorithm. arXiv preprint arXiv:2109.10680.

THANK YOU