Sociedad de Estadística e Investigación Operativa
Test (2001) *Vol.* **10**, *No.* 2, pp. 419 440

Trimmed means for functional data

Ricardo Fraiman*
Departamento de Matemática y Ciencias
Universidad de San Andrés, Argentina
Graciela Muniz
IMERL
Universidad de la República, Uruguay

Abstract

In practice, the use of functional data is often preferable to that of large finite-dimensional vectors obtained by discrete approximations of functions. In this paper a new concept of data depth is introduced for functional data. The aim is to measure the centrality of a given curve within a group of curves. This concept is used to define ranks and trimmed means for functional data. Some theoretical and practical aspects are discussed and a simulation study is given. The results show a good performance of our method, in terms of efficiency and robustness, when compared with the mean. Finally, a real-data example based on the Nasdaq 100 index is discussed.

Key Words: Data depth, functional data, trimmed means estimates.
AMS subject classification: 62G07, 62G05.

1 Introduction

In one dimension, order statistics and ranks are widely used for several applications, such as distribution free tests and some simple robust estimation procedures. In this case, they are easily defined through the natural order on the real line, and there is a vast literature on their applications.

L-estimates, which are defined as linear combinations of order statistics, are a well known class of robust location estimates. In particular trimmed means, which are defined as the average of the most central $(1-\alpha)n$ observations, $(0 \leq \alpha \leq 1)$ constitute a class of estimates that range from the sample mean to the sample median.

In more than one dimension, the concepts of order statistics and ranks are more involved and several definitions have been proposed in Mahalanobis (1936), Tukey (1975), Brown (1983), Oja (1983), Liu (1988,1990),

*Correspondence to: Ricardo Fraiman, Departamento de Matemática y Ciencias, Universidad de San Andrés, Buenos Aires, Argentina. Email: rfraiman@udesa.edu.ar

Received: June 2001; Accepted: September 2001

Small (1990), Gordaliza (1991), Singh(1991), Donoho and Gasko (1992),Liu and Singh (1993), Cuesta-Albertos, Gordaliza and Matrán (1997) and Fraiman and Meloche (1999).

All of them are based on different notions of depth. A data depth is a device introduced to measure the "centrality" of a multivariate data point within a given data cloud. Although these definitions are quite different for multivariate data, they are very similar when we look at them for univariate data. We now briefly describe two of them.

Let $Y_1, \dots, Y_n$ be independent and identically distributed random vectors on $\mathbb{R}^k$ with common distribution F.

Tukey's Depth. Tukey's depth at x is defined as

$$TD(x) = inf_H\{F(H) : x \in H\},$$

where H is a half space. The sampling version TD_n is defined by replacing F by the empirical distribution F_n. In one dimension ($k = 1$)

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

Simplicial Depth. Let $Y_1, \dots, Y_{k+1}$ be $k + 1$ i.i.d. random vectors with distribution F. The simplicial depth at x is defined as

$$SD(x) = P_F(x \in S[Y_1, \dots, Y_{k+1}]),$$

where we denote by $S[Y_1, \dots, Y_{k+1}]$ the closed simplex with vertices on the vectors $Y_1, \dots, Y_{k+1}$; its sample version SD_n is defined by replacing F by the empirical distribution F_n. The simplicial depth at x is the proportion of closed simplices with vertices in the sample to which the point x belongs. On the real line,

$$SD(x) = 2F(x)(1 - F(x^-)), \quad \text{and} \quad SD_n(x) = 2F_n(x)(1 - F_n(x^-)).$$

If there are no ties among the elements of the sample, we get the standard "both sides" order statistics, which can also be defined through the one dimensional depth

$$D(x) = 1 - \left|\frac{1}{2} - F(x)\right|. \tag{1.1}$$

Several applications of the notions of depth have been proposed. A first one is to consider the data point that maximizes a depth as a multivariate median. A second one is to construct multivariate L-estimates using the "order statistics" provided by the depth. This problem will be considered in this paper for functional data.

Other applications of data depth are, for instance, depth-based multivariate classification rules, multivariate two sample "nonparametric" tests, robust quality control, among others.

A different approach has also been followed by Brown and Hettmansperger (1987) who define multivariate quantiles which have both magnitude and direction, based on the gradient of Oja's (1983) measure of scatter.

All the proposed definitions attempt to order the observations according to their "depth" in the data cloud, the deepest observation defining the multivariate median.

Nowdays the real time monitoring of many processes in different fields such as stock markets, audience ratings, medicine, chemometrics, is available, providing large functional data sets. It has also been shown (see for instance Ramsay and Silverman 1997) that in practice, the use of functional data is often preferable to that of large finite-dimensional vectors obtained by discrete approximations of the functions. On the other hand the effective calculus of multivariate depth in high dimensional spaces is almost impossible for computational reasons.

In what follows we will define a natural notion of depth for functional data, i.e. when data are curves (realizations of a stochastic process). The idea is to measure "how long" remains a curve in the middle of a group of them.

In Section 2 we introduce a definition of depth for functional data, and we define α–trimmed means based on this notion. In Section 3 we provide strong consistency results for the proposed estimates. In Section 4, we report on the results of a simulation that compares the performance of a few estimates of location. In Section 5 we provide a real data example. All the proofs can be found in the Appendix.

2 Depth concepts for functional data and α–trimmed means

Let $X_1(t), \ldots, X_n(t)$ be independent and identically distributed stochastic processes with continuous trajectories defined on an interval $[a, b]$. Without loss of generality we assume $[a, b] = [0, 1]$. F_t will stand for the marginal univariate distribution function of $X_1(t)$.

For each realization we get a group of curves, and we want to know, for instance which of them stays more "in the middle of the group" for most of the time. This idea corresponds in a functional setup to the intuitive notion of median. More generally, the purpose of this article is, to define a depth concept for functional data as well as to propose trimmed-mean estimates for functional data.

Let D_n be a depth defined on R. For each fixed $t \in [0, 1]$ we consider

$$D_n(X_i(t)) = Z_i(t),$$

as the univariate depth of $X_i(t)$ at t with respect to $X_1(t), \ldots, X_n(t)$. In this way, at each single point t we have ranked the values $X_1(t), \ldots, X_n(t)$ according to their depths $Z_i(t)$, $1 \leq i \leq n$, which take a finite set of positive values.

We define now

$$I_i = \int_0^1 Z_i(t)dt, \quad 1 \leq i \leq n,$$

we rank the functions $X_1(t), \ldots, X_n(t)$ according to the associated I_i's values obtaining order statistics. Thus, the functional median will correspond to the $X_i(t)$ for which I_i is maximum. Ranks R_i are defined from the relationship $I_i = I^{(R_i)}$ that indicates the I_j's position in the ordered vector I's. L-estimates and trimmed-means estimates are easily derived from these order statistics.

More formally, the population functional depth is defined from a univariate depth D as follows.

Let D_t be the depth associated with the univariate distribution F_t by D, and $x = x(t)$ a continuous function on $[0, 1]$. Set

$$Z(t) = D_t(x(t)).$$

In particular, for the simplicial depth, we have

$$Z(t) = F_t(x(t))[1 - F_t(x(t^-))].$$

Then, the corresponding depth measure turns out to be

$$I(x) = \int_0^1 Z(t)dt = \int_0^1 F_t(x(t))[1 - F_t(x(t^-))]dt.$$

The sampling version $I_n(x)$ is defined by replacing F_t by the empirical distribution $F_{n,t}$, so we have

$$Z_n(t) = F_{n,t}(x(t))[1 - F_{n,t}(x(t^-))],$$

and

$$I_n(x) = \int_0^1 Z_n(t)dt = \int_0^1 F_{n,t}(x(t))[1 - F_{n,t}(x(t^-))]dt.$$

In what follows we will consider the univariate depth

$$D(x) = 1 - \left| \frac{1}{2} - F(x) \right|,$$

so that

$$Z(t) = 1 - \left| \frac{1}{2} - F_t(x(t)) \right| \tag{2.1}$$

and

$$I(x) = 1 - \int_0^1 \left| \frac{1}{2} - F_t(x(t)) \right|. \tag{2.2}$$

If the observations $X_1(t), \ldots, X_n(t)$ are ranked according to decreasing values of $I_n(X_i(t))$ we get order statistics $X^{(1)}(t), \ldots, X^{(n)}(t)$, where $X^{(1)}(t)$ denotes the deepest one (functional median), while the last ones will be on the "outer skin" of the data.

A functional version of the α trimmed mean is defined as the average of the $n - [n\alpha]$ deepest observations.

More precisely, we will consider for $\beta > 0$

$$\hat{\mu}_n = \frac{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I_n(X_i))X_i}{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I_n(X_i))} \tag{2.3}$$

as a functional trimmed mean, where

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}[\beta,+\infty)(I_n(X_i)) \simeq 1 - \alpha \tag{2.4}$$

and $\mathbf{1}_A$ stands for the indicator function of the set A.

Remark 2.1. A formal definition of robustness for continuous time data has not yet been proposed. A first stage would be to define which kind of contamination is allowed. For instance, if we allow for contamination on all data on a small (but possibly different) time interval, the whole data set could be corrupted. In this case, a robust nonparametric smoother (see for instance Härdle and Tsybakov 1988, or Boente and Fraiman 1989) should be applied at each single curve as a first step, and just then apply the methods defined above. In this paper we are mainly interested in a somewhat more realistic type of contamination, where a certain fraction of the data are corrupted on some possibly large time interval.

3 Strong consistency results

In this section we will show the uniform convergence of the empirical functional depth I_n to its population version I over an appropriate set of functions, and derive the strong consistency of the α–trimmed mean estimates.

We will assume the following two assumptions.

H1. For a positive constant A (large enough) let

$$Lip[0,1]=\{x:[0,1]\rightarrow\mathbb{R},x \text{ is Lipschitz with constant less or equal to } A\}$$

be the space of functions where the paths of the stochastic process $X_1(t)$ take values.

H2. There exists a constant $c > 0$ for which

$$E(\lambda(\{t : X_1(t) \in [u(t), u(t) + c\epsilon]\}) < \epsilon/2,$$

where λ stands for the Lebesgue's measure on $\mathbb{R}$ and $u \in Lip[0,1]$.

Theorem 3.1. *Under H1 and H2, if*

$$J_n(x) = \int_0^1 F_{n,t}(x(t))dt, \quad \textit{and} \quad J(x) = \int_0^1 F_t(x(t))dt,$$

we have that

$$\lim_{n\rightarrow+\infty} \sup_{\{x\subset Lip[0,1]\}} \mid J_n(x) - J(x) \mid = 0 \quad a.s. \tag{3.1}$$

and

$$\lim_{n\to+\infty} \sup_{\{x\in Lip[0,1]\}} \mid I_n(x) - I(x) \mid = 0 \quad a.s., \tag{3.2}$$

where $I_n(x)$ and $I(x)$ are respectively the empirical depth and the population depth of $x(t)$.

Let

$$\hat{\mu}_n = \frac{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I_n(X_i))X_i}{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I_n(X_i))} \tag{3.3}$$

be our trimmed mean estimate. Define also

$$\mu_n = \frac{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I(X_i))X_i}{\sum_{i=1}^n \mathbf{1}_{[\beta,+\infty)}(I(X_i))} \tag{3.4}$$

(an artificial unobservable estimate) and

$$\mu = \frac{E(X_1\mathbf{1}_{[\beta,+\infty)}(X_1))}{E(\mathbf{1}_{[\beta,+\infty)}(I(X_1))} \tag{3.5}$$

the population trimmed mean.

Theorem 3.2. *If the stochastic process $X_1(t)$ takes values on an arbitrary space $E[0,1] := E$ where*

$$\lim_{n\to+\infty} \sup_{\{x\subset E\}} \mid J_n(x) - J(x) \mid = 0 \quad a.s., \tag{3.6}$$

then

$$\hat{\mu}_n \to \mu \quad a.s.$$

In particular, under H1 and H2, $\hat{\mu}_n \to \mu$ a.s.

4 Simulation results

In this section, we report on the results of a simulation that compares trimmed mean estimates with the regular mean under four different models.

The basic one, (model M1) consist of p functions satisfying

$$X_i(t) = g(t) + e_i(t) \quad 1 \leq i \leq p,$$

where $e_i(t)$ is a Gaussian stochastic process with zero mean and covariance function

$$E(e_i(t)e_i(s)) = \left(\frac{1}{2}\right)^{|t-s|p}$$

and the function $g(t) = 4t$ and corresponds to the non-contaminated model.

Then we have considered two kinds of contamination of the basic model, a total one and a partial one (on trajectories), and we also consider symmetric and asymmetric contamination.

In the case of symmetric total contamination, model M2 is given by

$$Y_i(t) = X_i(t) + \epsilon_i \sigma_i M \quad 1 \leq i \leq p,$$

where ϵ_i and σ_i are independent sequences of random variables, ϵ_i takes values 1 with probability q and 0 with probability (1-q) the contamination fraction— and σ_i takes values 1 and -1 with probability 1/2. M is the size of the contamination (a constant).

In the case of asymmetric total contamination, model M3 is defined by

$$Y_i(t) = X_i(t) + \epsilon_i M \quad 1 \leq i \leq p,$$

where ϵ_i and M are defined in model M2.

For partial contamination we consider model M4 defined as

$$Y_i(t) = X_i(t) + \epsilon_i \sigma_i M \quad \text{for } t \geq T_i \quad 1 \leq i \leq p$$

and

$$Y_i(t) = X_i(t) \quad \text{for } t < T_i,$$

where T_i is randomly chosen according to a uniform distribution on $(0, 1)$.

In each case we perform $N = 500$ replications for $p = 50$ and $p = 80$ curves, $q = 0.05$ and $q = 0.1$, $M = 5$ and $M = 25$ and $\alpha = 0.2$ and 0.3.

For each model we consider the mean and trimmed mean estimates,

$$\hat{\mu}_n(t) = \sum_{i=1}^{p} X_i(t)/p$$

and

$$\hat{\mu}_{n,\alpha}(t) = \sum_{i=1}^{p-[p\alpha]} X^{(i)}(t)/(p - [p\alpha])$$

for $\alpha = 0.2$ and 0.3.

For each of the 500 replications the estimates were evaluated at $I = 30$ equally spaced points of $[0, 1]$, and we calculate the integrated error for each replication

$$EI(j) = \frac{1}{I}\sum_{k=1}^{I}\left[f\left(k/I\right) - g\left(k/I\right)\right]^2,$$

where f denotes μ_n or $\mu_{n,\alpha}$ respectively.

In the tables we report for each estimate the mean integrated error

$$E = \frac{1}{N}\sum_{j=1}^{N} EI(j)$$

and its standard deviation

$$s = \left(\frac{1}{N}\sum_{j=1}^{N}\left(EI(j) - E\right)^2\right)^{1/2}.$$

We also report robust measures of the performance of the estimators.

$$M = median(EI(j) \quad j = 1, \ldots, N)$$

and

$$MAD(EI) = median(|EI(j) - M| \quad j = 1, \ldots, N)/0.675.$$

We can see that the functional trimmed mean estimates behaves very well under all models —even under asymmetric contamination. This is also reasonable since depth trimming is not necessarily a symmetric procedure. As expected the mean breaks down under all the contamination models. The worst behaviour for the mean is under asymmetric contamination (model M3). The functional trimmed means are calculated for trimming levels α equal to 0.2 and 0.3, different contamination fractions q (0.05 and 0.1), sample sizes p (50 and 80) and for two different contamination constants (5 and 25). For each sample size p the integrated mean square error varies very little for the different contamination models, and the values are very close to the integrated mean square error of the non–contaminated model.

				Models			
				M1	M4	M3	M2
Sample	Contam.	Contam.	Estimator	non	piecewi-	asymme-	symme-
size	probability	constant		cont.	se cont.	tric cont.	tric cont.
$p = 50$	$q = 0.05$	5	Mean	0.033	0.049	0.118	0.057
				(0.015)	(0.036)	(0.101)	(0.050)
				0.030	0.039	0.086	0.043
				(0.014)	(0.022)	(0.069)	(0.025)
			20% Trim-	0.043	0.043	0.043	0.044
			med mean	(0.020)	(0.020)	(0.019)	(0.020)
				0.039	0.039	0.039	0.040
				(0.019)	(0.018)	(0.016)	(0.017)
$p = 80$			Mean	0.020	0.034	0.096	0.035
				(0.009)	(0.029)	(0.078)	(0.030)
				0.019	0.027	0.073	0.027
				(0.008)	(0.014)	(0.058)	(0.016)
			20% Trim-	0.027	0.028	0.029	0.027
			med mean	(0.013)	(0.013)	(0.014)	(0.013)
				0.024	0.025	0.026	0.024
				(0.012)	(0.012)	(0.013)	(0.011)

Table 1: $N = 500$ *replications, trimming level* $= 0.2$.

				Models			
				M1	M4	M3	M2
Sample	Contam.	Contam.	Estimator	non	piecewi-	asymme-	symme-
size	probability	constant		cont.	se cont.	tric cont.	tric cont.
$p = 50$	$q = 0.1$	5	Mean	0.033	0.081	0.338	0.086
				(0.015)	(0.079)	(0.253)	(0.080)
				0.030	0.055	0.262	0.058
				(0.014)	(0.040)	(0.224)	(0.041)
			20% Trim-	0.043	0.042	0.045	0.044
			med mean	(0.020)	(0.019)	(0.023)	(0.021)
				0.039	0.038	0.040	0.040
				(0.019)	(0.016)	(0.019)	(0.017)
$p = 80$			Mean	0.020	0.076	0.308	0.101
				(0.009)	(0.063)	(0.197)	(0.077)
				0.019	0.059	0.266	0.081
				(0.008)	(0.046)	(0.172)	(0.064)
			20% Trim-	0.027	0.027	0.029	0.029
			med mean	(0.013)	(0.014)	(0.014)	(0.014)
				0.024	0.024	0.026	0.026
				(0.012)	(0.011)	(0.013)	(0.013)

Table 2: $N = 500$ *replications, trimming level* $= 0.2$.

				Models			
				M1	M4	M3	M2
Sample	Contam.	Contam.	Estimator	non	piecewi-	asymme-	symme-
size	probability	constant		cont.	se cont.	tric cont.	tric cont.
$p = 50$	$q = 0.05$	25	Mean	0.033	0.410	2.078	0.742
				(0.015)	(0.713)	(2.182)	(1.022)
				0.030	0.200	1.212	0.316
				(0.014)	(0.250)	(1.488)	(0.423)
			20% Trimmed mean	0.043	0.043	0.041	0.044
				(0.020)	(0.020)	(0.018)	(0.021)
				0.039	0.038	0.039	0.040
				(0.019)	(0.017)	(0.018)	(0.019)
$p = 80$			Mean	0.020	0.463	2.079	0.583
				(0.009)	(0.583)	(1.856)	(0.619)
				0.019	0.255	1.590	0.396
				(0.008)	(0.287)	(1.346)	(0.479)
			20% Trimmed mean	0.027	0.028	0.029	0.027
				(0.013)	(0.014)	(0.014)	(0.013)
				0.024	0.025	0.025	0.025
				(0.012)	(0.012)	(0.010)	(0.010)

Table 3: $N = 500$ *replications, trimming level* $= 0.2$.

				Models			
				M1	M4	M3	M2
Sample	Contam.	Contam.	Estimator	non	piecewi-	asymme-	symme-
size	probability	constant		cont.	se cont.	tric cont.	tric cont.
$p = 50$	$q = 0.1$	25	Mean	0.033	1.227	7.387	2.093
				(0.015)	(1.172)	(5.934)	(2.447)
				0.030	0.765	6.057	1.165
				(0.014)	(0.808)	(5.128)	(1.394)
			20% Trimmed mean	0.043	0.044	0.070	0.044
				(0.020)	(0.021)	(0.024)	(0.025)
				0.039	0.040	0.039	0.040
				(0.019)	(0.020)	(0.017)	(0.018)
$p = 80$			Mean	0.020	1.335	6.970	1.819
				(0.009)	(1.254)	(4.325)	(1.565)
				0.019	0.923	6.205	1.507
				(0.008)	(0.877)	(3.912)	(1.363)
			20% Trimmed mean	0.027	0.028	0.030	0.029
				(0.013)	(0.013)	(0.017)	(0.013)
				0.024	0.026	0.027	0.025
				(0.012)	(0.012)	(0.013)	(0.011)

Table 4: $N = 500$ *replications, trimming level* $= 0.2$.

				Models		
Sample size	Contam. probability	Contam. constant	Estimator	M4 piecewise cont.	M3 asymmetric cont.	M2 symmetric cont.
$p = 50$	$q = 0.05$	25	Mean	0.437	2,090	0.709
				(0.749)	(2.433)	(0.981)
				0.212	1.090	0.298
				(0.264)	(1.384)	(0.396)
			30% Trimmed mean	0.048	0.052	0.050
				(0.022)	(0.025)	(0.024)
				0.043	0.046	0.044
				(0.019)	(0.023)	(0.023)
$p = 80$			Mean	0.368	1.929	0.445
				(0.486)	(1.854)	(0.720)
				0.154	1.478	0.163
				(0.203)	(1.412)	(0.218)
			30% Trimmed mean	0.032	0.034	0.031
				(0.015)	(0.016)	(0.014)
				0.029	0.031	0.028
				(0.014)	(0.015)	(0.014)
$p = 80$	$q = 0.1$		Mean	1.323	7.024	1.825
				(1.250)	(4.382)	(1.556)
				0.896	6.123	1.503
				(0.885)	(3.864)	(1.344)
			30% Trimmed mean	0.033	0.040	0.034
				(0.016)	(0.020)	(0.017)
				0.029	0.036	0.030
				(0.014)	(0.018)	(0.015)

Table 5: $N = 500$ *replications, trimming level* $= 0.3$.

5 A real data example

In order to illustrate how the functional depth concept works in practice we consider a real data example.

We analize 100 curves used to build up the Nasdaq 100 Index, taking 63 daily measurements starting on April 12, 2000.

Figure 1 shows the 100 curves all together. Figure 2 is a plot of the 70 deepest curves. Figure 3 is a plot of the remaining 30 curves. Finally, Figure 4 shows the 80 deepest curves, while Figure 5 shows the remaining "more external" 20 curves.

It is apparent that the deepest curves can be inserted between the plots of the outer curves. In this sense the functional depth (3) does its job quite well.

Appendix

Proof of Theorem 3.1.

Let $Y(t)$ be a stochastic process with distribution P. Since

$$E_P(\mathbf{1}_{(-\infty,u]}(Y(t))) = F_t(u),$$

we have

$$I(x) = E_P\left(\int_0^1 \mathbf{1}_{(-\infty,x(t)]}(Y(t))dt\right) = \int_0^1 P(Y(t) \le x(t))dt.$$

Given $x \in D[0,1]$ we define

$$g_x(z) = \int_0^1 \mathbf{1}_{(-\infty,x(t)]}(z(t))dt,$$

then

$$I_n(x) - I(x) = P_n g_x - P g_x,$$

where $Pf = \int f dP$. Let us define a family $\mathcal{F} = \{g_x : x \in D[0,1]\}$ of functions with envelope $\boldsymbol{F} \equiv 1$. In order to prove the uniform consistency, we will show that for any fixed $\epsilon \ge 0$, $\ln \boldsymbol{N}_1(\epsilon, P_n, \mathcal{F}) = o_P(n)$ where

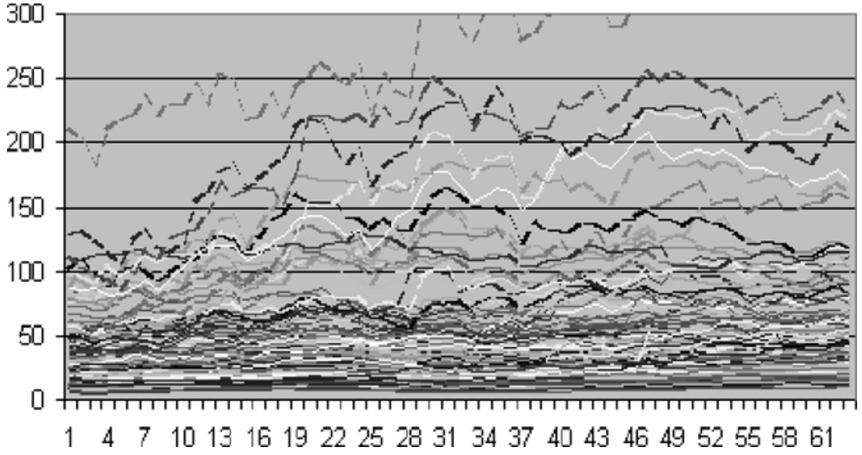

Figure 1: NASDAQ 100

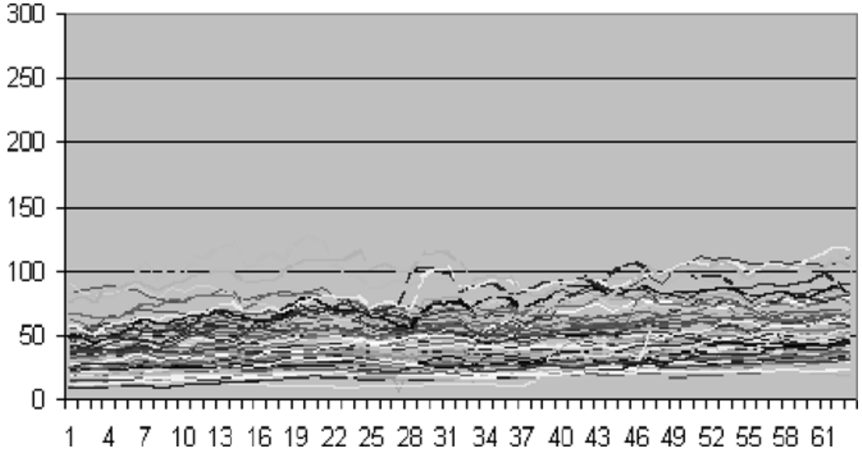

Figure 2: 70% less deepest curves

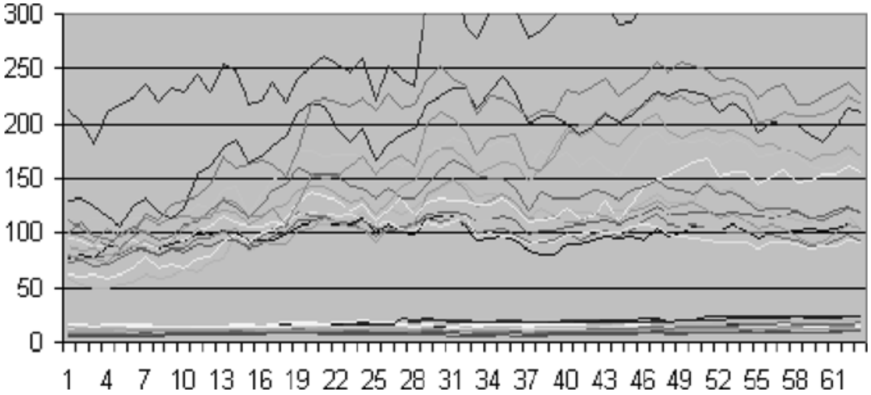

Figure 3: *30% less deepest curves*

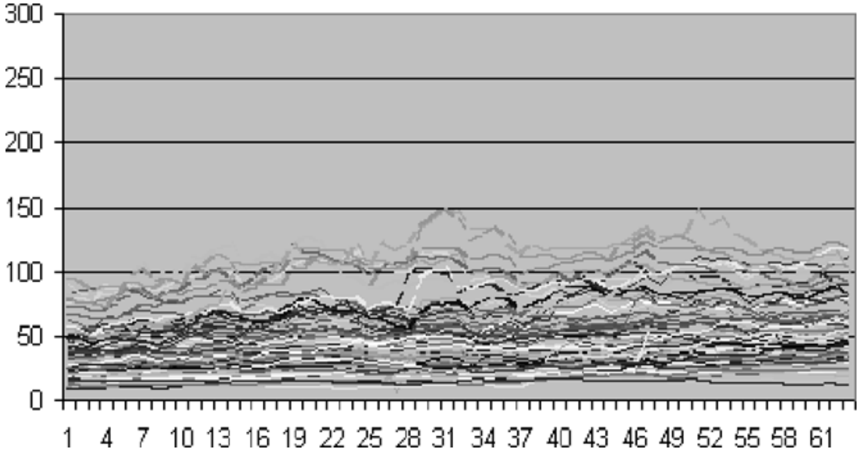

Figure 4: *80% less deepest curves*

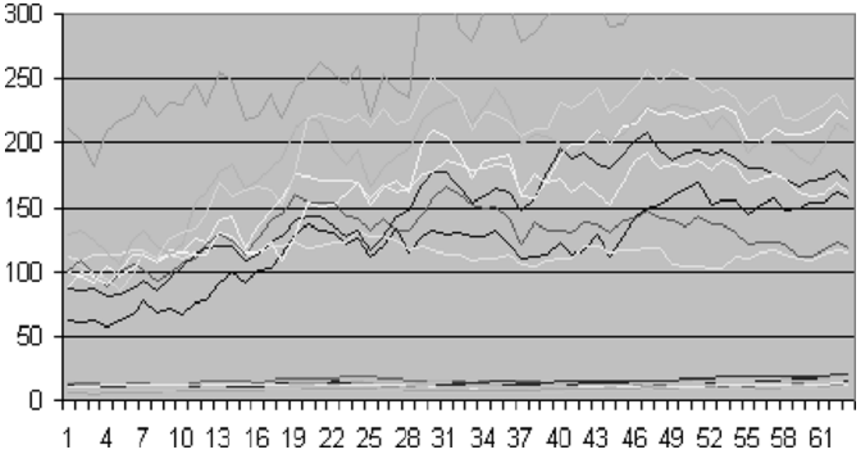

Figure 5: 20% less deepest curves

$\boldsymbol{N}_1(\epsilon, P_n, \mathcal{F})$ is the family's entropy. (see Pollard 1984, Theorem 24 pp. 25). We start finding a bound for

$$\begin{aligned}\|g_x - g_{x'}\|_{L^1(P_n)} &= \frac{1}{n}\sum_{i=1}^{n}|g_x(X_i) - g_{x'}(X_i)| \\ &= \frac{1}{n}\sum_{i=1}^{n}\left|\int_0^1 \mathbf{1}_{(-\infty,x(t)]}(X_i(t))dt\right. \\ &\quad \left. - \int_0^1 \mathbf{1}_{(-\infty,x'(t)]}(X_i(t))dt\right| .\end{aligned}$$

Set $y(t) := \min(x(t), x^{'}(t))$ and $z(t) := \max(x(t), x^{'}(t))$. Then

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left|\int_0^1 \mathbf{1}_{(-\infty,x(t)]}(X_i(t))dt\right. &- \left.\int_0^1 \mathbf{1}_{(-\infty,x'(t)]}(X_i(t))dt\right| \\ &= \frac{1}{n}\sum_{i=1}^{n}\int_0^1 \mathbf{1}_{[y(t),z(t)]}(X_i(t))dt \\ &= \frac{1}{n}\sum_{i=1}^{n}\lambda\left\{t : X_i(t) \in [y(t), z(t)]\right\},\end{aligned}$$

where λ is the Lebesgue measure on $\mathbb{R}$. From the Strong Law of Large Numbers, this expression converges almost surely to

$$E(\lambda\{t : X_1(t) \in [y(t), z(t)]\}).$$

Let us define

$$D_n = D_n(\delta) = \left\{ \omega : \frac{1}{n}\sum_{i=1}^{n} \lambda(\{t : X_i(t) \in [y(t), z(t)]\}) < \delta \right\}.$$

By H2 there exists a constant $c > 0$ for which

$$E(\lambda\{t : X_1(t) \in [u(t), u(t) + c\epsilon]\}) < \epsilon/2$$

and $u(t) \in Lip[0,1]$. Hence, if $\| x - x^{'} \|_\infty < c\epsilon$ we can conclude that $P(D_n^c(\epsilon)) \to 0$ as $n \to \infty$.

Indeed, with probability 1, we have

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} \lambda\{t : X_i(t) \in [y(t), z(t)]\} \\ &\leq \frac{1}{n}\sum_{i=1}^{n} \lambda\{t : X_i(t) \in [y(t), y(t) + c\epsilon]\} \\ &\leq_{\forall n \geq n_o(\omega)} E(\lambda\{t : X_1(t) \in [y(t), y(t) + c\epsilon]\}) + \frac{\epsilon}{2} < \epsilon. \end{aligned}$$

Given $M > A$, for each $\delta > 0$ we consider now a family of polygonals $\mathcal{H} \subset D[0,1]$ whose elements y satisfy:

$$y(0) \text{ takes values on the sequence}$$

$$(-M, -J\delta/3, -(J-1)\delta/3, \ldots, -\delta/3, 0, \delta/3, \ldots, J\delta/3, M) =: (-M, M; \delta/3)\,,$$

where J is the integer part of $3M/\delta$, and for b in the sequence $(-A\delta/3, A\delta/3; \delta/3)$

$$\begin{aligned} y\left(\frac{\delta}{3A}\right) &= y(0) + b \\ y\left(2\frac{\delta}{3A}\right) &= y\left(\frac{\delta}{3A}\right) + b \\ y\left(k\frac{\delta}{3A}\right) &= y\left((k-1)\frac{\delta}{3A}\right) + b. \end{aligned}$$

Given $x \in D[0,1]$ with $|x(0)| < M$, there exists an element of $\mathcal{H}$ for which

$$||x(t) - y(t)||_\infty \leq \delta.$$

To see this, observe that there exists y for which

$$|x(0) - y(0)| \leq \delta/3,$$

and

$$|x(\delta/3M) - y(\delta/3M)| \leq \delta/3.$$

We have that $y(0) - \delta/3 \leq x(0) \leq y(0) + \delta/3$. Besides, since $x(t) \in Lip[0,1]$,

$$x(0) - \frac{\delta}{3} \leq x(0) - \frac{\delta A}{3M} \leq x(t) \leq x(0) + \frac{\delta A}{3M} \leq x(0) + \frac{\delta}{3}, \quad \text{for } t \in \left[0, \frac{\delta}{3A}\right].$$

Then, $y(0) - 2\delta/3 \leq x(0) - \delta/3 \leq x(t) \leq x(0) + \delta/3 \leq y(0) + 2\delta/3$. We can also choose y such that: $y(0) - \delta/3 \leq y(t) \leq y(0) + \delta/3$. Then, $\mid x(t) - y(t) \mid \leq \delta$ for $t \in [0, \delta/3A]$. Since $\mid y(\delta/3A) - x(\delta/3A) \mid \leq \delta/3$ we can repeat the procedure for the next interval.

On the other hand it is easy to see that $\sharp\mathcal{H} \leq [6M/\delta]^{3A/\delta}$, where $\sharp C$ stands for the cardinal of the set C.

Let us consider now the family of functions

$$\{g_y : y \in \mathcal{H}\}$$

For a given $\epsilon > 0$, we have that on $D_n(\epsilon)$

$$\frac{1}{n}\sum_{i=1}^{n} \lambda(\{t : X_i(t) \in [y(t), z(t)]\}) < \epsilon, \quad \text{if } n \geq n_0(\epsilon, \omega)$$

and

$$\| x - x^{'} \|_\infty < \delta =: c\epsilon.$$

On D_n, if $\mid x(0) \mid < M$ there exists $x^{'} \in \mathcal{H}$ such that

$$\| g_x - g_{x'} \|_{L^1(P_n)} < \epsilon.$$

If $|x(0)| > M$ and $X_i(0) < M - A$, for $i = 1, \ldots, n$ we can take $y \in \mathcal{H}$ (on the boundary) such that

$$\| g_x - g_{x'} \|_{L^1(P_n)} = 0.$$

If we define $C_n = \bigcap_{i=1}^n \{|X_i(0)| < n\}$ (with $M - A = n$), then

$$P(C_n^c) \leq nP(|X_1(0)| > n) \to 0$$

as $n \to \infty$, whenever $E(|X_1(0)|) < \infty$. Consider now

$$E_n = C_n \cap D_n.$$

We have that

$$\frac{1}{n} \ln \boldsymbol{N}_1(\epsilon, P_n, \mathcal{F}) \xrightarrow{P} 0.$$

Indeed, for all $\eta > 0$

$$\begin{aligned} P & \left(\frac{1}{n} \ln \boldsymbol{N}_1(\epsilon, P_n, \mathcal{F}) > \eta\right) \\ & \leq P\left(\frac{1}{n} \ln \boldsymbol{N}_1(\epsilon, P_n, \mathcal{F}) > \eta, E_n\right) + P(E_n^c) \\ & = P\left(\frac{c}{n\delta} \ln(6M/\delta) > \eta, E_n\right) + P(E_n^c). \end{aligned}$$

The last term tends to zero, while the first one is zero. Then, $\boldsymbol{N}_1(\epsilon, P_n, \mathcal{F})$ $= o_p(1)$, and we conclude

$$\sup_{x \in Lip[0,1]} \mid P_n g_x - P g_x \mid \longrightarrow_{n\to\infty} 0 \quad a.s.$$

i.e.

$$\sup_{x \subset Lip[0,1]} \mid J_n(x) - J(x) \mid \longrightarrow_{n\to\infty} 0 \quad a.s.$$

which entails

$$\sup_{x \subset Lip[0,1]} \mid I_n(x) - I(x) \mid \longrightarrow_{n\to\infty} 0 \quad a.s.$$

which concludes the proof.

Proof of Theorem 3.2.

Let us define

$$S_n := \sup_{\{x \subset E\}} |I_n(x) - I(x)|.$$

By assumption (3.6), we have

$$\sup_{\{x \subset E\}} |J_n(x) - J(x)| \longrightarrow 0 \quad a.s.$$

which implies

$$S_n \longrightarrow 0 \quad a.s.$$

Let h be one of the two functions $h(t) = t$ or $h(t) = 1$. Define

$$\hat{\rho}_{n,h} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[\beta,+\infty)}(I_n(X_i))h(X_i)$$

and

$$\rho_{n,h} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[\beta,+\infty)}(I(X_i))h(X_i).$$

When $h(t) = t$, $\hat{\rho}_{n,h}$ is the numerator of $\hat{\mu}_n$ and when $h(t) = 1$, then $\hat{\rho}_{n,h}$ is the denominator of $\hat{\mu}_n$ defined in (3.3). The same holds for $\rho_{n,h}$ and μ_n defined in (3.4). Since by the Law of Large Numbers,

$$\rho_{n,h} \longrightarrow E(\mathbf{1}_{[\beta,+\infty)}(I(X_1))h(X_1)) \quad a.s.$$

as $n \to \infty$, it suffices to show that

$$\lim_{n\to\infty} |\hat{\rho}_{n,h} - \rho_{n,h}| = 0 \quad a.s.$$

Now if δ is any positive number,

$$\begin{aligned} |\hat{\rho}_{n,h} - \rho_{n,h}| \quad \leq \quad & \frac{1}{n} \sum_{i=1}^{n} |h(X_i)||\mathbf{1}_{[\beta,+\infty)}(I(X_i) + \delta) \\ & -\mathbf{1}_{[\beta,+\infty)}(I(X_i) - \delta)|\mathbf{1}_{\{S_n \leq \delta\}} \\ & +\frac{1}{n} \sum_{i=1}^{n} |h(X_i)||\mathbf{1}_{[\beta,+\infty)}(I(X_i) + S_n) \\ & -\mathbf{1}_{[\beta,+\infty)}(I(X_i) - S_n)|\mathbf{1}_{\{S_n \geq \delta\}}. \end{aligned}$$

The second summand can be majorized by

$$\frac{1}{n} \sum_{i=1}^{n} |h(X_i)|\mathbf{1}_{\{S_n \geq \delta\}}$$

that converges a.s. to zero as $n \to \infty$. As, for the first summand, we have

$$\left| \frac{1}{n} \sum_{i=1}^{n} |h(X_i)| |\mathbf{1}_{[\beta,+\infty)}(I(X_i)+\delta) - \mathbf{1}_{[\beta,+\infty)}(I(X_i)-\delta)| \mathbf{1}_{\{S_n \leq \delta\}} \right.$$

$$\left. -E\left(|h(X)||\mathbf{1}_{[\beta,+\infty)}(I(X)+\delta) - \mathbf{1}_{[\beta,+\infty)}(I(X)-\delta)|\right) \right|.$$

This difference converges a.s. to 0 by the Law of Large Numbers. Finally since $\mathbf{1}_{[\beta,+\infty)}(\cdot)$ is bounded and monotone and $E(|h(X)|) < \infty$, as a consequence of the dominated convergence theorem, we obtain

$$\lim_{\delta \to 0} E\left(|h(X)||\mathbf{1}_{[\beta,+\infty)}(I(X)+\delta) - \mathbf{1}_{[\beta,+\infty)}(I(X)-\delta)|\right) = 0$$

which concludes the proof.

References

Boente, G. and R. Fraiman (1989). Robust nonparametric regression estimation. *Journal of Multivariate Analysis*, **29**, 180–198.

Brown, B.M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society, B*, **45**, 25–30.

Brown, B.M. and T.P. Hettmansperger (1987). Affine invariant rank methods in the bivariate location model. *Journal of the Royal Statistical Society, B*, **49**, 301–310.

Cuesta-Albertos, J.A., A. Gordaliza and C. Matrán (1998). Trimmed k-means: An attempt to robustify quantizers. *The Annals of Statistics*, **25**, 553 576.

Donoho, D.L. and M. Gasko (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, **20**, 1803–1827.

Fraiman, R. and J. Meloche (1999). Multivariate L-estimation. *Test*, **8**, 255–317.

Gordaliza, A. (1991). Best approximations to random variables based on trimming procedures. *Journal of Approximation Theory*, **64**, 162–180.

Härdle, W. and A.B. Tsybakov (1988). Robust nonparametric regression with simultaneous scale curve estimation . *The Annals of Statistics*, **16**, 120–135.

Liu, R. (1988). On a notion of simplicial depth. *Proceedings of the National Academy of Sciences*, U.S.A., **85**, 1732–1734.

Liu, R. (1990). On a notion of data depth based on random simplices. *The Annals of Statistics*, **18**, 405–414.

Liu, R. and K. Singh (1993). A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, **421**, 252 260.

Mahalanobis, P.C. (1936). On the generalized distance in Statistics. *Proceedings of the National Academy of India*, **12**, 49 55.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, **1**, 327–332.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer Verlag.

Ramsay, J.O. and B.W. Silverman (1997). *Functional Data Analysis.* Springer Verlag.

Singh, K. (1991). A notion of majority depth. Technical Report, Rutgers University, Department of Statistics.

Small, C.G. (1990). A survey of multidimensional medians. *International Statistical Review*, **58**, 263–277.

Tukey, J.W. (1975). Mathematics and picturing data. *Proceedings of the International Congress of Mathematics*, Vancouver, **2**, 523 531.