Impartial Trimmed Means for Functional Data

Juan Antonio Cuesta-Albertos and Ricardo Fraiman

Abstract. In this paper we extend the notion of impartial trimming to a functional data framework, and we obtain resistant estimates of the center of a functional distribution. We give mild conditions for the existence and uniqueness of the functional trimmed means. We show the continuity of the population parameter with respect to the weak convergence of probability measures. As a consequence we obtain consistency and qualitative resistance of the data based estimates. A simplified approximate computational method is also given. Some real data examples are finally analyzed.

1. Introduction

Real time monitoring of many processes are available for applications and modeling in different fields such as medicine, neuroscience, chemometrics, signal transmission, stock markets, meteorology and TV audience ratings. In this context the individual observed responses are rather curves than finite dimensional vectors, and may be modeled as sampling paths $X(t,\omega)$, $\omega \in \Omega$, of independent realizations of a stochastic process centered at a function $\mu(t)$, i.e., functional data.

In practice, the use of functional data is often preferable to that of large finite dimensional vectors obtained by discrete approximations of the functions (see for instance the books by Ramsay and Silverman [**38, 39**]). In [**13, 36, 1, 24, 10, 4, 20, 37, 40, 26, 25, 6, 15, 16, 30, 2, 19, 12**], and the references therein, we have several case–studies and/or theoretical developments for functional data.

Robustness has been an almost not explored area in this context of functional data, never the less there is no reason why we should not expect the presence of outliers in there. For one dimensional data, the simplest robust estimates of a location parameter are the well known trimmed-mean estimates, a family that goes from the sample mean to the sample median as increasing the trimming level. However, there is not a standard way to extend it to higher dimensions. The concept of trimmed-means and medians has been a topic of active research in the last decade, even for two dimensional data.

Usually, trimming is associated to ranks, but in more than one dimension, the concepts of order statistics and ranks are more involved and several definitions have

1991 *Mathematics Subject Classification.* Primary 62H30, 62F15, 62G07; Secondary 62F35.

Key words and phrases. Impartial trimming, robustness, mean, functional data, consistency, uniqueness.

The first author was supported in part by the Spanish Ministerio de Ciencia y Tecnología, grant BFM2002-04430-C02-02.

been proposed for the finite dimensional case. See, for instance, [**31, 43, 5, 34, 27, 28, 41, 22, 42, 14, 29, 18**] and the book by Mosler [**33**]. All of them are based on different notions of depth concepts, a device introduced to measure the centrality of a vector within a given data cloud.

A different approach to trimming, which is not based on ranks, is the "impartial trimming" (self–determined by the data) which was introduced by Gordaliza in [**22**] for the multivariate location model and will be described in detail later on. Roughly speaking, impartial–trimming means that the trimming is not necessarily symmetric, and the data which are to be trimmed are self–determined by the estimate based on the whole data cloud. Moreover, even in the one-dimensional case, has the main advantage over the usual trimming that it is not required to fix a special zone or direction in advance in which the data will be trimmed. This property has shown to be important when dealing with asymmetric contamination.

For a related problem, Cuesta–Albertos, Gordaliza and Matrán [**8, 9**] introduced a class of procedures based on "impartial trimming" to robustify k-means for clustering methods in the finite dimensional case. See also [**32, 11**].

We will employ this approach to define trimmed means for functional data, and study their asymptotic properties under mild conditions.

A robust estimate of the center of a functional distribution has been proposed by Fraiman and Muniz in [**19**]. They define trimmed means for functional data based on a functional depth concept, which is defined as an integral of the univariate depths at each single point t. It is useful to point here that two main differences with the present work, besides the different approach to the problem, can be stated:

(a) the approach introduced here can be extended to the case of trimmed k-means in a similar way as in [**8, 9**] where they generalize the work [**22**].
(b) here the results are obtained under mild assumptions on the underlying stochastic process. This is not the case in [**19**] where strong regularity conditions are required on the trajectories of the process.

From here on we proceed as follow. In Section 2 we borrow the definition of qualitative robustness given in [**3**] and adapt it to the case of functional data considering a new distance between trajectories. In Section 3 we define a robust estimate based on the concept of impartial trimming. We show the existence of an optimal trimming, and, based on this optimal trimming, we define our estimate. We prove the strong consistency and robustness of the estimate.

In Section 4 we give conditions for the uniqueness of the parameter to be estimated. We introduce the concept of symmetric unimodality for stochastic processes which turns out to be a sufficient condition for the uniqueness of the impartial trimmed–mean parameter. The interest of the uniqueness of the target parameter relies on the fact that uniqueness is required in order to obtain robustness. (This is also the case for one dimensional distributions. For instance, the median is not qualitative robust if the underlying distribution of the random variable X satisfies $P[X = -1] = P[X = 1] = 0.5$.) Sharp conditions for the uniqueness of the impartial trimmed mean parameter, even in a finite dimensional setup, remains still an open problem. In this Section we provide an example of a symmetrical, strictly unimodal two-dimensional distribution with strictly positive differentiable density function for which not only the uniqueness fails but the parameter is not the symmetry center. Moreover, at this time we are only aware of the existence of two uniqueness results. The first one, in [**22**], is for one dimensional distributions and requires that the distribution be unimodal, with a differentiable and strictly positive density function. The second one, in [**21**], applies to elliptical, unimodal, differentiable and strictly positive densities.

Computational problems are studied in Section 5 where we introduce a simpler consistent alternative estimate, which can be also used as a starting point for an algorithm searching for the proposed estimate. In Section 6 we analyze some real data examples from TV ratings. Most of the proofs and technical results are deferred to the Appendix.

In what follows we restrict to consider $L^2[0,1]$-valued random elements. However, most of the present results can be extended in a straightforward way to the case of random elements (r.e.'s) with values in an uniformly convex Banach space.

2. Qualitative Robustness

Hampel, in [**23**], introduced the concept of qualitative robustness for a sequence of estimates of a finite dimensional parameter in the case of independent and identically distributed (i.i.d.) observations. His definition stated that a sequence of estimates T_n is robust at a given distribution μ if for any other distribution $\tilde{\mu}$ close to μ in the Prohorov metric, the distribution of T_n under μ and under $\tilde{\mu}$ are close in the Prohorov metric, uniformly in the sample size.

The use of the Prohorov distance reflects the intuitive meaning of robustness as insensitivity of the estimate to round-off errors, and to a small fraction of outliers.

Boente, Fraiman and Yohai in [**3**] extended these notions to the case of stochastic processes with dependent variables, and introduced a new definition of robustness based on the concept of resistance (see also [**35, 7**]). In a functional data setup robustness should reflect insensitivity of the estimate to two kinds of different contamination. Firstly, following Hampel, the notion of robustness should take care of:

(a) small errors in all the realizations of the process (e.g. round-off errors).
(b) a small fraction of far away curves (outliers).

However, it may also happen that in a functional data framework, a much wilder kind of contamination be present: each curve can be "out of control" during a small fraction of time. Thus, the second notion of robustness should also take into consideration that:

(c) each curve may be disturbed in a small time interval.

This last type of contamination is much wilder than the first one, since it allows, in principle, all the data to be contaminated, each curve at a different small interval of time. If the data have sharp peaks at different intervals a robust procedure should mainly ignore (or, in some way, delete) those peaks. However, if we have some kind of data such that, most of it, presents a sharp spike in the same interval, then the method should not delete it. This effect can be seen in the following example.

In [**17**] the behavior of electric power consumers at Buenos Aires, Argentina, is analyzed. For every individual (household) in the sample, measurements were taken at each of the 96 intervals of 15 minutes in every weekday (Monday to Friday) during January 2001. Monthly averages over days for each individual were analyzed. The main conclusion, was that with respect to the peaks, two typical consumer's behavior were found. One with only a peak around 9:00 pm, while another group

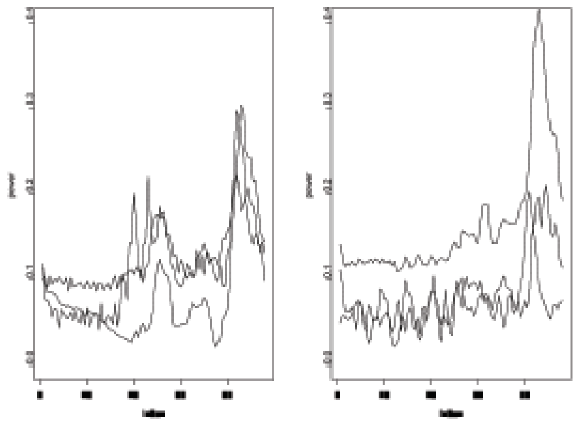

FIGURE 1. Electrical power typical consumers at Buenos Aires, Argentina. Curves represent the mean along week days in January 2001 of instantaneous consume, measured every 15 minutes along the day. Right: Three typical members of the group with only a peak at night. Left: Three typical members of the group with a second peak around noon.

had also a peak around noon. In Figure 1 we plot the curves corresponding to some of the typical consumers in each group.

In this work, the important feature of the data was precisely the hours at which maximum consume was attained. On the other hand, if we use a naive robust estimate that starts by a simple "cut-off" of sharp peaks, we could loose this characteristic.

In this paper, we propose here an estimate which is able to handle the more classical contamination models that takes care of (a) and (b). Another type of estimates are necessary to deal with contamination of type (c). At this time we have some idea on how to handle this contamination and it is the subject of some research in progress.

To finalize this section, we present some definitions which should be fulfilled by estimates robust against those kind of contaminations. Let $E = L^2[0,1]$, and, for $x, y \in L^2[0,1]$ we will denote,

$$d(x,y) = \|x - y\| = \left(\int_0^1 |x(t) - y(t)|^2 dt\right)^{1/2}.$$

In order to define robustness we borrow the ideas in [**3**]. First, we introduce a metric d_n and a pseudo-metric $d_{n,c}$, that will be used for the two different robustness definitions. Let $\overline{x}_n := (x_1, x_2, ..., x_n)$ and $\overline{y}_n := (y_1, y_2, ..., y_n)$ with $x_i, y_i \in E, i \geq 1$, and define

$$d_n(\overline{x}_n, \overline{y}_n) = \inf\{\epsilon : (\#\{i : d(x_i, y_i) \geq \epsilon\})/n \leq \epsilon\},$$

where $\#A$ denotes the cardinal of the set A. If each $y_i,\ i=1,...,n$ is obtained contaminating the corresponding $x_i,\ i=1,...,n$ with contamination of type (a) or (b), this definition makes $\overline{x}_n$ and $\overline{y}_n$ be close.

On the other hand, if we define

$$d_{n,c}(\overline{x}_n,\overline{y}_n)=\inf\{\epsilon:(\#\{i:d_{c,\epsilon}(x_i,y_i)\geq\epsilon\})/n\leq\epsilon\}$$

with

$$d_{c,\epsilon}(x_i,y_i)=\inf_{\{A\subset[0,1]\}}\left(\int_{\{[0,1]-A\}}|x_i(t)-y_i(t)|^2dt\right)^{1/2},\ i=1,...,n,$$

where $\lambda(A)<\epsilon$ and $\lambda(.)$ stands for the Lebesgue measure, we obtain a distance which makes $\overline{x}_n$and $\overline{y}_n$ to be close also under contamination of type (c).

Let (F,D) be a metric space and $T_n: E^n\to F$ be a sequence of estimates taking values in F. Let $\overline{x}=(x_n:n\geq 1)\in E^\infty$, and $\Pi_n(\overline{x})=(x_1,x_2,...,x_n)\in E^n$ be the canonical projection on the first n coordinates. Define

$$\Delta_n(\delta,\Pi_n(\overline{x}))=\sup\{D[T_n(y_n),T_n(\Pi_n(\overline{x}))]:y_n\in B(\Pi_n(\overline{x}),\delta,d_n)\},$$

where $B(\Pi_n(\overline{x}),\delta,d_n)$ stands for the open ball in E^n centered at $\Pi_n(\overline{x})$ and radius δ with respect to the metric d_n.

Definition 2.1. Let $\overline{x}\in E^\infty$. T_n is resistant at $\overline{x}$ if, given $\epsilon>0$, there exist $\delta>0, n_0\in I\!N$ such that $\Delta_n(\delta,\Pi_n(\overline{x}))\leq\epsilon$, fore very $n\geq n_0$.

Definition 2.2. T_n is robust at $\overline{\mu}$, where $\overline{\mu}$ is a probability measure on E^∞, if

$$\overline{\mu}\left(\{\overline{x}\in E^\infty:T_n \text{ is resistant at } \overline{x}\}\right)=1.$$

Changing d_n by $d_{n,c}$ in the previous definitions we get a notion of robustness, that also deals with (c)–type contamination.

3. Definition and properties.

In this paper we will be concerned only with $L^2[0,1]$-valued r.e.'s which, unless otherwise stated, will be defined on the same rich enough probability space (Ω,σ,μ). P will be a fixed probability distribution on the Borel σ algebra on E.

Our robust estimate is based on the idea of impartial trimming introduced in [**22**] which can be extended to the infinite dimensional case in the following way.

Let $\alpha\in(0,1)$. We will say that a measurable function $\tau:E\to[0,1]$ is a *trim at level α for P* if

$$\int\tau(y)dP(y)\geq 1-\alpha.$$

The function τ tells us which part of every point must be trimmed taking into account that a maximum trim of α is allowed.

Let us denote by $\mathcal{P}_\alpha$ the family of all trims at level α for P.

We divide this section in three subsections. In the first one we prove the existence of an optimal trimming in the sense of (3.1) below. In the second one we prove the continuity of the trimmed mean population parameter (see definition below) with respect to the convergence in distribution. In the third one, we introduce the estimate and, as a consequence of the results in previous subsection, we obtain the strong consistency and qualitative robustness of the sequence of empirical trimmed mean estimates.

3.1. Definition of population parameter and existence of an optimal trimming. Given $\alpha \in (0,1)$, we will show that there exist a point $m_P \in E$ and $\tau_P \in \mathcal{P}_\alpha$, which are optimal in the following sense

$$I_\alpha(P) := \inf_{m\in E, \tau\in\mathcal{P}_\alpha} \int ||y-m||^2 \tau(y) dP(y) = \int ||y-m_P||^2 \tau_P(y) dP(y). \tag{3.1}$$

We will call *α-trimmed mean* or, simply, trimmed mean, to every m_P such that there exists a trimming function $\tau_P \in \mathcal{P}_\alpha$ which satisfy (3.1). τ_P will be called *optimal trimming function.* If we apply equation (3.1) to the empirical probability measure P_n, we obtain the empirical α-trimmed mean estimate.

It is possible to extend the previous definition to obtain (Φ, α)-trimmed means (and estimates), as in [**22**], introducing a continuous and non decreasing weight function Φ such that

$$\lim_{t\to\infty} \Phi(t) > \Phi(x) \quad \text{for every} \quad x \in R^+,$$

replacing equation (3.1) by

$$I_{\alpha,\Phi}(P) := \inf_{m\in E, \tau\in\mathcal{P}_\alpha} \int \Phi\left[||y-m||\right] \tau(y) dP(y) = \int \Phi\left[||y-m_P||\right] \tau_P(y) dP(y).$$

This family extends the so-called Z-estimates (see, for instance, [**44**]). However, for the sake of notation, we will assume throughout that $\Phi(t) = t^2$.

We start with some previous results. The following lemma states a very well known property of the mean and justifies the name of trimmed mean we have given to m_P.

LEMMA 3.1. *Let $\alpha > 0$ and let $\tau \in \mathcal{P}_\alpha$. If we define*

$$x_\tau := \frac{\int y\tau(y)dP(y)}{\int \tau(y)dP(y)},$$

then

$$\int ||y-x_\tau||^2 \tau(y) dP(y) \le \int ||y-x||^2 \tau(y) dP(y), \text{ for every } x \in E.$$

We introduce now some additional notation. Given $m \in E$ and $r > 0$, let $B(m,r)$ (resp. $\overline{B}(m,r)$) denotes the open (resp. closed) ball centered at m with radius r. $S(m,r)$ will stand for the associated sphere. Let us also define

$$r_\alpha(m) := \inf\{r > 0 : P[B(m,r)] \ge 1-\alpha\}.$$

It follows that if $r < r_\alpha(m)$ then $P[B(m,r)] < 1-\alpha$ and

$$P[B(m, r_\alpha(m))] \le 1-\alpha \le P[\overline{B}(m, r_\alpha(m))].$$

PROPOSITION 3.2. *Let $\alpha > 0$, $m \in E$ and let $\tau_m \in \mathcal{P}_\alpha$ be such that*

$$\int \tau_m(y) dP(y) = 1-\alpha \text{ and } I_{B(m,r_\alpha(m))} \le \tau_m \le I_{\overline{B}(m,r_\alpha(m))}. \tag{3.2}$$

Then, we have that for every $\tau \in \mathcal{P}_\alpha$,

$$\int ||y-m||^2 \tau_m(y) dP(y) \le \int ||y-m||^2 \tau(y) dP(y). \tag{3.3}$$

PROOF. Let $\tau \in \mathcal{P}_\alpha$. We have that

$$\begin{aligned}
\int ||y-m||^2\tau(y)dP(y) - \int ||y-m||^2\tau_m(y)dP(y) \\
&= \int_{B(m,r_\alpha(m))} ||y-m||^2(\tau(y)-1)dP(y) + \int_{\overline{B}^c(m,r_\alpha(m))} ||y-m||^2\tau(y)dP(y) \\
&\quad + \int_{S(m,r_\alpha(m))} ||y-m||^2(\tau(y)-\tau_m(y))dP(y) \\
&\geq r_\alpha^2(m)\left(\int_{B(m,r_\alpha(m))}(\tau(y)-1)dP(y) + \int_{\overline{B}^c(m,r_\alpha(m))}\tau(y)dP(y)\right. \\
&\quad \left. + \int_{S(m,r_\alpha(m))}(\tau(y)-\tau_m(y))dP(y)\right) \geq 0,
\end{aligned}$$

where the first inequality follows from the fact that $\tau \in [0,1]$ and the second one from the definition of $\mathcal{P}_\alpha$ and (3.2). □

From equation (3.3), if we denote

$$D_\alpha(m,P) := \int ||y-m||^2\tau_m(y)dP(y), \tag{3.4}$$

we obtain easily the following corollary.

COROLLARY 3.3. *Let $\alpha > 0$ and let P be a probability measure. Then, $m_P \in E$ is a trimmed mean parameter of P if and only if*

$$D_\alpha(m_P,P) \leq D_\alpha(m,P), \quad \textit{for every } m \in E.$$

REMARK 3.4. Equality in (3.3) is only possible if τ satisfies (3.2). Therefore, according to Proposition 3.2, if an optimal trimming function, τ_P, does exist, its support is a ball with center at the trimmed mean m_P. Its radius is $r_\alpha(m_P)$. If the trimmed mean is unique, we will often denote it $r_\alpha(P)$ and we will call it *trimming radius.* In what follows, in order to simplify the notation, we will often suppress the symbol α in the subindices, since it is fixed.

REMARK 3.5. If we fix m, the value

$$\int ||y-m||^2\tau(y)dP(y)$$

does not depend on τ as long as τ satisfies (3.2). This justify the notation introduced in (3.4) and allows us to represent by τ_m to every trim at level α for P which satisfies (3.2).

Now we are able to establish the existence result. The proof is given in the Appendix.

THEOREM 3.6. (Existence of optimal trimming) *Let $\alpha \in (0,1)$ and let P be a probability measure on E. Then there exists $m_P \in E$ such that*

$$I_\alpha(P) = D_\alpha(m_P,P).$$

Obviously, uniqueness of of m_P is not guaranteed by this result.

3.2. Continuity. In order to prove consistency and qualitative robustness, we will show the continuity of the trimmed mean parameter with respect to the convergence in distribution. The proof relies in a technique introduced in [**10**], and is given in the Appendix.

THEOREM 3.7. (Continuity) *Let $\{P_n\}$ be a sequence of probability measures which converges in distribution to the probability measure P. Let $\alpha \in (0,1)$ and let us assume that the trimmed mean of P is unique.*

Let $\{m_n\}$ be a sequence of trimmed means of $\{P_n\}$ and denote by $\{r_n\}$ the sequence of the associated trimming radius. Then

$$\lim ||m_n - m_P|| = 0 \text{ and } \lim r_n = r(P).$$

REMARK 3.8. Without the uniqueness assumption in Theorem 3.7, it can be still proved, with the same proof, that the sequence $\{m_n\}$ is sequentially compact in norm and that every accumulation point is an α-trimmed mean of P.

From this remark and Theorem 3.7 we can obtain the following corollary. Its proof (which is deferred to the Appendix) does not require the uniqueness assumption.

COROLLARY 3.9. *Let $\{P_n\}$ be a sequence of probability measures which converges in distribution to the probability measure P. If $\alpha \in (0,1)$, then*

$$\lim I_\alpha(P_n) = I_\alpha(P).$$

3.3. Estimates. Consistency and robustness. Let $\{X_n\}$ be a sequence of i.i.d. r.e.'s with distribution P. For every $n \in I\!N$, we will consider the empirical probability measure P_n defined by

$$P_n := \frac{1}{n} \sum_{i \leq n} \delta_{X_i(\omega)}, \, \omega \in \Omega,$$

where δ_x denotes for Dirac's delta measure on x.

Given $\alpha \in (0,1)$, we will denote by m_n^ω any *empirical trimmed mean* and by τ_n^ω the associated *empirical trimming function.* The radius of the empirical trimming function will be denote by r_n^ω.

We estimate m_P by the sequence $\{m_n^\omega\}$ and $r(P)$ by $\{r_n^\omega\}$. The consistency result is the following.

THEOREM 3.10. (Strong Consistency) *Let $\alpha \in (0,1)$ and let us assume that the probability P has a unique trimmed mean parameter. Let $\{X_n\}$ be a sequence of i.i.d. r.e.'s with distribution P. Then, every sequence of the empirical trimmed means and empirical trimming radius satisfy that*

$$\lim ||m_n^\omega - m_P|| = 0 \text{ and } \lim r_n^\omega = r(P), \text{ for } \mu\text{-a.e. } \omega \in \Omega.$$

PROOF. This result is, in fact, a corollary of the extension to Banach spaces of Glivenko-Cantelli's Theorem and Theorem 3.7. □

The resistance of our procedure can be deduced from the results in [**3**] although, we include an independent proof in the Appendix.

THEOREM 3.11. (Robustness) *Let us assume that the probability P satisfies that its trimmed mean is unique. Let $\{X_n\}$ be a sequence of i.i.d. r.e.'s with distribution P. Then any sequence of empirical trimming means is robust at P^∞.*

4. Uniqueness of the trimmed mean parameter. Unimodality

In this section we will give a sufficient condition (given in (4.1) below) for the uniqueness of the trimmed mean parameter of a distribution. In the finite dimensional case it is closely related to symmetry and unimodality.

DEFINITION 4.1. Let X be an E-valued r.e. We will say that its distribution is *symmetrical and unimodal* with mode at m_0 if the distribution of X satisfies

$$P[\overline{B}(m_0, r)] > P[\overline{B}(m, r)], \text{ for every } m \in E \text{ and } r > 0. \tag{4.1}$$

In the one dimensional case, it is easy to see that if X is symmetrical around m and unimodal, then (4.1) holds; moreover, if P satisfies (4.1), then it can be shown that P is continuous, symmetrical around m_0 and has a unique mode at m_0 in the sense that, for every $m > m_0$ and $\delta > 0$ it follows that

$$P[m, m+\delta] > P[m+\delta, m+2\delta].$$

On the other hand, condition (4.1) holds for every finite-dimensional distribution which admits a representation similar to the one given in Theorem 4.4 below. However, we want to remark that unimodality plus symmetry are not enough to guarantee the uniqueness of the trimmed mean as shown in the following example.

EXAMPLE 4.2. Let us consider the two dimensional and bounded set

$$A := B(\overline{0}, 100) \quad \bigcap \Big[\Big([-10, 10] \times [-1, 1] \Big) \bigcup \Big\{ (x_1, x_2) : |x_1| \leq 10|x_2| \Big\} \Big],$$

where $\overline{0} = (0, 0)$. Let P_A be the uniform distribution on A, $m_0 = (20, 0)$ and let us assume we want to compute the α-trimmed mean of P_A with $\alpha = 1 - P_A[B(m_0, 5)]$. Obviously

$$D_\alpha(\overline{0}, P_A) > D_\alpha(m_0, P_A). \tag{4.2}$$

Given $\lambda > 0$, let P_λ be the probability distribution supported on A with density function given by

$$f_\lambda(x) := K[1 + \lambda(100 - ||x||)I_A(x)],$$

where the constant K is chosen in order that f_λ be a density function. Therefore, the level curves of P_λ are the intersection of spheres with A, and the parameter λ describes how step is f_λ. Moreover, f_λ is strictly unimodal, for every $x \in I\!R^2$:

$$f_\lambda(x) = f_\lambda(-x)$$

and the function $t :\longrightarrow f_\lambda(tx)$, $t \in I\!R$, is strictly decreasing on $\{t > 0 : tx \in A\}$. We also have that

$$\lim_{\lambda \to 0+} P_\lambda = P_A \text{ and } \lim_{\lambda \to 0+} D_\alpha(z, P_\lambda) = D_\alpha(z, P_A),\ z = \overline{0}, m_0.$$

Thus, by (4.2) there exists λ_0 such that $D_\alpha(\overline{0}, P_{\lambda_0}) > D_\alpha(m_0, P_{\lambda_0})$ and $\overline{0}$ is not the trimmed mean of P_{λ_0} and, by the symmetry of this distribution, the trimmed mean is not unique.

Notice that a slight modification of this example allows to choose P_{λ_0} supported by $I\!R^2$ and f_{λ_0} satisfying any desired regularity condition.

THEOREM 4.3. *Let P be a probability measure such that there exists $m_0 \in E$ which satisfies (4.1).*

Then, for every $\alpha \in (0,1)$, the trimmed mean of P is unique and coincides with m_0.

PROOF. Let $\alpha > 0$. Given $m \in E$ and τ_m a trim at level α for P which satisfies (3.2), the function

$$r \to Q_m(r) := \frac{1}{1-\alpha} \int_{\overline{B}(m,r)} \tau_m(y) dP(y),\ r \geq 0,$$

is a distribution function and satisfies that

$$D_\alpha(m, P) = (1-\alpha) \int r^2 dQ_m(r).$$

Moreover, (4.1) implies that the distribution Q_{m_0} is strictly stochastically smaller than Q_m and, in consequence,

$$D_\alpha(m_0, P) < D_\alpha(m, P).$$

□

The next theorem and corollary state that the class of symmetrical and unimodal distributions is rich enough in the infinite dimensional case. The proof is given in the Appendix

THEOREM 4.4. *Let $\{e_1, e_2, ...\}$ be a fixed orthonormal basis of E. Let X be an E-valued r.e. with distribution P which admits the representation*

$$X = \sum_n X_n e_n,$$

where the random variables $\{X_1, X_2, ...\}$ are independent with continuous density functions $\{f_1, f_2, ...\}$ with respect to the Lebesgue measure. Assume also that, for every $n \in I\!N$, f_n is symmetric with respect to 0 and strictly decreasing on $[0, \infty)$.

Then, the distribution of X satisfy (4.1) with $m_0 = 0$.

COROLLARY 4.5. *The assumptions of Theorem 4.4 include all Gaussian Processes with zero mean.*

PROOF. It is easy to verify that we can obtain any covariance linear operator from a stochastic process satisfying the assumptions of Theorem 4.4. □

5. Computational problems

Given a sample $\{X_1, ..., X_n\}$, let P_n stands for the empirical probability measure. The search for the point m_n^ω which minimizes (3.1) is, in general, computationally too expensive. By this reason we propose an alternative much simpler consistent estimate, that we denote $\hat{m}_{k_n}^\omega$. This estimate just consists on the resulting value if we restrict the search of the minimum in (3.1) to the support of P_n. In other words, using the equivalence in Corollary 3.3,we can see that the estimate $\hat{m}_{k_n}^\omega$, is determined by the relationship

$$\hat{m}_{k_n}^\omega \in \{X_1(\omega), ..., X_n(\omega)\} \ \text{ and } \ D_\alpha(\hat{m}_{k_n}^\omega, P) = \inf_{i=1,...n} D_\alpha(X_i(\omega), P_n).$$

It is clear that $\hat{m}_{k_n}^\omega$ is much easier to compute because in this task it is only required to compute and handle the set of distances

$$\{||X_i - X_j|| : i, j = 1, 2, ..., n\}.$$

An algorithm to compute $\hat{m}^{\omega}_{k_n}$ is the following. Let m stands for $n(1-\alpha)$, if $n(1-\alpha)$ is an integer number, or for to the integer part of $n(1-\alpha)+1$, otherwise. The algorithm consist of the following steps:

(1) Compute $D(i,j) = \|X_i(\omega) - X_j(\omega)\|^2, \ i, j = 1, ..., n$
(2) Set $\mathrm{Obj} = \infty$
(3) Repeat for $i = 1 : n$
    $DD = \mathrm{sort}(D(i,:))$
    $DDD = DD(1:m,:)$
    If $\|DDD\|^2 < \mathrm{Obj}$, then set $\mathrm{Obj} = \|DDD\|^2$ and $\hat{m}^{\omega}_{k_n} = X_i(\omega)$

Let $\tilde{I}_\alpha(P_n) := D_\alpha(\hat{m}^{\omega}_{k_n}, P_n)$. Theorem 5.1 states that the sequence $\{\hat{m}^{\omega}_{k_n}\}$ is consistent if the value m_P belongs to the support of P. The proof is given in the Appendix.

THEOREM 5.1. *Let us assume that the hypotheses in Theorem 3.10 hold. Let us also assume that m_P belongs to the support of P. Then*

$$\lim ||\hat{m}^{\omega}_{k_n} - m_P|| = 0, \textit{ for } \mu\textit{-a.e. } \omega \in \Omega.$$

The estimate m^{ω}_n is a mean of some curves in the sample and, in consequence, this curve does not exist in the sample. However, the estimate $\hat{m}^{\omega}_{k_n}$ is one of the curves in the sample, which allows to answer some additional questions. For instance, let us assume that you are interested into select a representative company of those in Wall Street shares market. The data consists of the curves representing the value of every share in the market. If you select the representative employing this procedure, once the company has been selected, you can analyze it to figure out the characteristics of "the typical company in Wall Street" (size, business sector,...).

On the other hand, even in the case that one is interested in computing the estimate m^{ω}_n, it may worth to use the value $\hat{m}^{\omega}_{k_n}$ as an starting point for an iterative algorithm.

6. Some real data example from TV ratings

In this section we apply our method to some real-data examples. Their paths are all sampled at the same time points. If this were not the case we suggest to use a linear interpolation or any smoothing procedure to fill in the possible gaps.

Let us assume that we are trying to start a television advertisement campaign. With some simplification, we can say that the price we have to pay for it is proportional to the sum of the numbers of individuals watching each individual spot. Moreover, as an additional simplification, let us assume that we have decided to insert our spots during the broadcasting of a program whose length is 30 minutes, and that we are allowed to choose the exact time in which our spots are going to appear.

In this setup, we have to pay in advance a price which is proportional to an estimation of the total number of individuals watching this program at the exact time in which the spots will appear. The problem is to make this estimation as accurate as possible.

In practice, quite often, the data to carry out this estimation are obtained from every broadcasting of the program along an earlier month and, essentially, consist

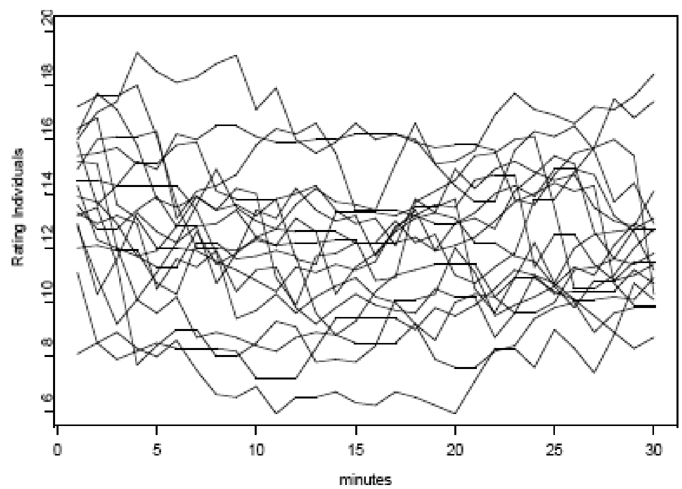

Figure 2. Curves showing the rate of individuals watching Big Brother II at prime time at Montevideo, Uruguay. Each curve corresponds to a day of the 20 first days of emission of the game. Rates were measured every minute during the 30 minutes of broadcasting.

of the number of people watching the program every minute. Those measurements are the so-called rates of audience and the curves which they form in every emission are the audience rating curves.

As an example, we have chosen the first month of the emission at prime time of Big Brother II at Montevideo, Uruguay. Figure 2 is a plot of the minute by minute audience rating curves during the 20 broadcastings in this month.

The usual estimate consists of, simply, the mean of the audience rating curves during the selected month.

A problem with this method is that if, during the month under consideration, there are some days in which the audience of the selected program is atypical, with higher or lower ratings, the estimate will be far from the future real data and we will pay more (or less) than we should. We will illustrate with two examples that this problem can be avoided if we use impartial trimmed means instead of regular ones.

We have computed the 0.15 trimmed mean curve of the data in Figure 2 (which means that we are allowed to trim exactly three curves). In Figure 3 we give a plot of the mean curve and the trimmed mean curve. The trimmed mean is about 2 rating points above the mean, which is around 15 percent more. Therefore, if we assume that the trimmed days are really anomalous days, we can conclude that the audience we can expect in standard days is 2 rating points higher than this given by the usual mean.

As a second example that goes in the opposite direction, let us consider the last 16 days of Big Brother I in the prime time emission also at Montevideo. Take

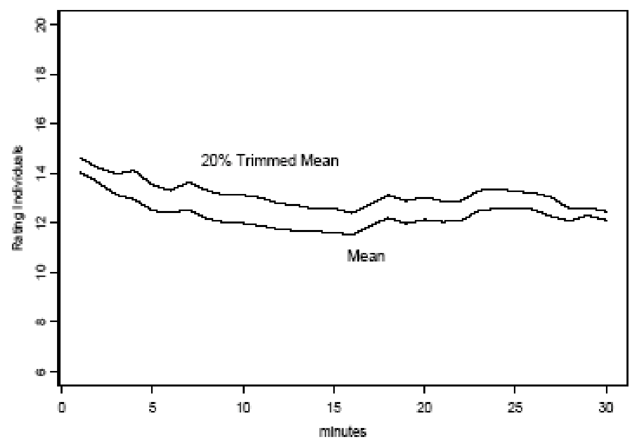

FIGURE 3. Mean and 15 % Trimmed Mean of the curves shown in Figure 2

.

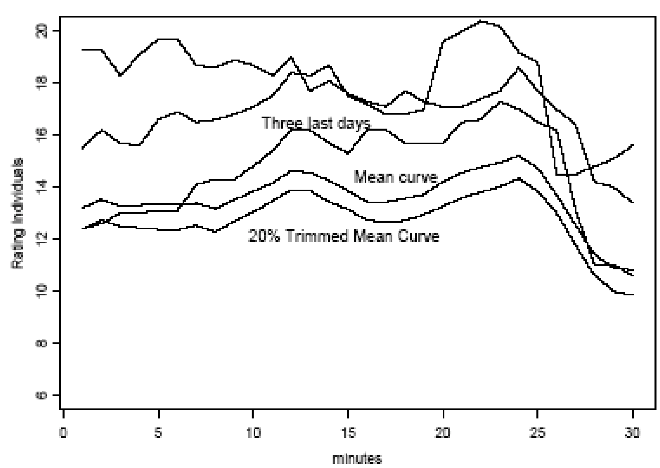

FIGURE 4. Mean and trimmed mean of the 16 curves containing the rates of the individuals watching Big Brother I at Montevideo, Uruguay, during the last 16 days of the game. The trimming proportion was 3/16.

$\alpha = 3/16$ (in order to trim exactly three curves) and let us see the difference between the mean and the α trimmed mean. Those curves appear in Figure 4. This time, the average curve is about 2 points above the trimmed mean.

Now, the question is: Did something unusual happened on the trimmed days? The answer for the second example is rather obvious. The three trimmed curves

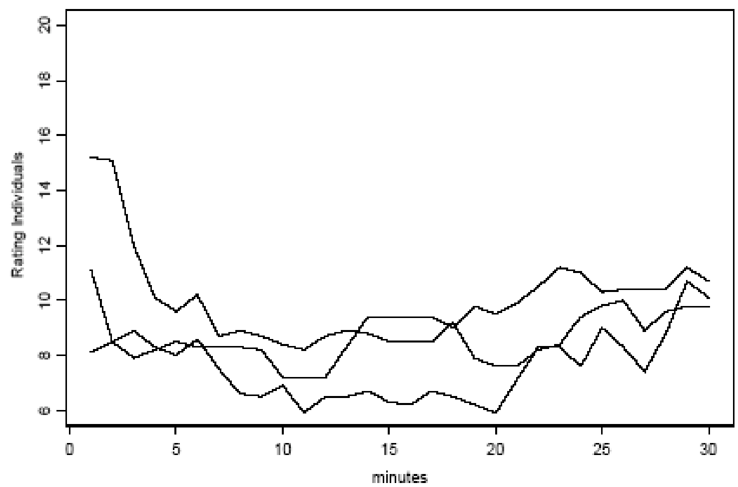

FIGURE 5. Mean and trimmed mean of the 16 curves containing the rates of individuals watching Big Brother I at Montevideo, Uruguay, during the last 16 days of the game plotted jointly with the curves corresponding to the last three days which were, precisely, the trimmed curves. The trimming proportion was 3/16.

correspond to the last three emissions, in which the winner of the game was selected by the audience, and the ratings have increased considerably. Figure 5 is a plot the three trimmed curves with the mean curve and the trimmed mean curve for comparison.

Concerning the three trimmed curves for the first example, once we looked carefully at the data, we found that two of them correspond to days where there were special programs at a sports cable program (soccer games) that got a considerable rating, quite above its average. Figure 6 is a plot of the ratings curves of those events. On a regular day, the ratings for that sports cable channel are below 2 rating points. The last trimmed curve is from a Monday of a long week end holiday. It seems obvious that the trimmed days can not be considered as standard ones and that on those days the audience of the selected program is below than usual.

Finally, we want to remark that, in spite of the fact that a careful designer of the experiment would have deleted from the sample those anomalous days (for instance, if you know the mechanics of Big Brother you will never choose the last week as a typical week or, if you know something about the interest on soccer in Uruguay, you will never choose a day in which the Uruguayan team plays), the interest of the proposed estimate is that it did his job automatically in the right way and discovered those final days in the second example and those days in which the Uruguayan soccer team was in screen[1].

[1]By the way, it was hard, but with those results the Uruguayan soccer team qualified for the World Championship. Regrettably, once there, it was eliminated in the first round

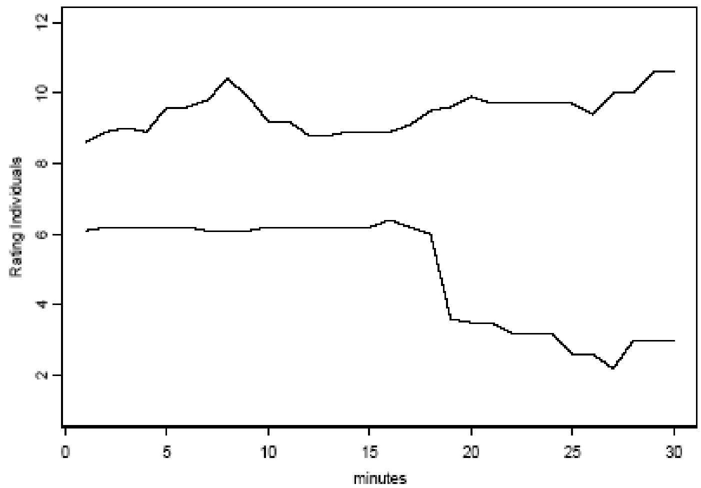

FIGURE 6. Curves showing the rate of individuals watching Tenfield channel at Montevideo, Uruguay, on the 14/08/2001 and the 23/08/2001 during the time at which Big Brother II was on the screen.

7. Appendix

The first lemma is a property of uniformly convex Banach spaces and it is stated here for further reference.

LEMMA 7.1. *Let* $\{x_n\} \subset E$ *be a sequence which converges weakly to* x_0. *Then*
(P.1) $\liminf ||x_n|| \geq ||x_0||$.
(P.2) *If* $\lim ||x_n - x_0|| \neq 0$*, then* $\liminf ||x_n|| > ||x_0||$.

PROPOSITION 7.2. *Let* $\alpha \in (0,1)$ *and let* P *be a probability measure on* E. *Let* $\{x_n, n = 0, 1, ...\} \subset E$ *be such that*

$$\lim ||x_n - x_0|| = 0 \text{ and } \lim r_\alpha(x_n) = r_\alpha(x_0).$$

Then

$$\lim \int ||y - x_n||^2 \tau_{x_n}(y) dP(y) = \int ||y - x_0||^2 \tau_{x_0}(y) dP(y).$$

PROOF. To simplify, let us denote $\tau_n = \tau_{x_n}$. It follows easily that

$$\lim \tau_n(y) = \begin{cases} 1 & \text{if } y \in B(x_0, r_\alpha(x_0)) \\ 0 & \text{if } y \notin \overline{B}(x_0, r_\alpha(x_0)). \end{cases}$$

On the other hand

$$\begin{aligned} 1 - \alpha &= \lim \int \tau_n(y) dP(y) \\ &= \lim \int \left(I_{B(x_0, r_\alpha(x_0))} + I_{\overline{B}^c(x_0, r_\alpha(x_0))} + I_{S(x_0, r_\alpha(x_0))} \right) \tau_n(y) dP(y), \end{aligned}$$

and, the bounded convergence theorem implies that

$$\gamma := 1-\alpha - P[B(x_0, r_\alpha(x_0)] = \lim \int_{S(x_0,r_\alpha(x_0))} \tau_n(y)dP(y). \tag{7.1}$$

Now, since

$$||y-x_n||^2 \left(I_{B(x_0,r_\alpha(x_0))}(y) + I_{\overline{B}^c(x_0,r_\alpha(x_0)}(y) \right) \tau_n(y) \to ||y-x_0||^2 I_{B(x_0,r_\alpha(x_0))}(y),$$

again the bounded convergence theorem implies that

$$\begin{aligned}
\lim & \left[\int ||y-x_n||^2 \tau_n(y)dP(y) - \int ||y-x_0||^2 \tau_0(y)dP(y) \right] \\
&= \lim \int_{S(x_0,r_\alpha(x_0))} \left(||y-x_n||^2 \tau_n(y) - ||y-x_0||^2 \tau_0(y) \right) dP(y) \\
&= r_\alpha^2(x_0) \lim \int_{S(x_0,r_\alpha(x_0))} \left(\tau_n(y) - \tau_0(y) \right) dP(y) = 0
\end{aligned}$$

by (7.1). □

Proof of Theorem 3.6.

PROOF. Let $\{x_n\} \subset E$ be a sequence such that

$$D_\alpha(x_n, P) \to I_\alpha(P). \tag{7.2}$$

Let us denote, to simplify, $\tau_n = \tau_{x_n}$. We start proving that both sequences $\{x_n\}$ and $\{r_\alpha(x_n)\}$ are bounded. Take $H > 0$ such that

$$P[B(0,H)] > \alpha,$$

and let $\delta = P[B(0,H)] - \alpha$. Thus, $\int_{B(0,H)} \tau_n(y)dP(y) \geq \delta$ for every $n \in I\!N$. If the sequence is not bounded, there exists a subsequence $\{x_{n_k}\}$ such that $||x_{n_k}|| \to \infty$, and we have that

$$\begin{aligned}
\lim D_\alpha(x_{n_k}, P) &= \lim \int ||y-x_{n_k}||^2 \tau_{n_k}(y)dP(y) \\
&\geq \lim \int_{B(0,H)} ||y-x_{n_k}||^2 \tau_{n_k}(y)dP(y) \\
&\geq \lim (||x_{n_k}|| - H)^2 \delta = \infty,
\end{aligned}$$

which contradicts (7.2).

Boundness of $\{r_\alpha(x_n)\}$ follows from the boundness of $\{x_n\}$ and the fact that if H^* satisfies that $P[B(0,H^*)] > 1-\alpha$, thus, $r_\alpha(x_n) \leq \sup_k ||x_k|| + H^*$, since

$$B(0,H^*) \subset B(x_n, ||x_n|| + H^*).$$

In consequence, without loss of generality, we can assume that there exists $x_0 \in E$ and r_0 such that the sequence $\{x_n\}$ converges weakly to x_0 and $\lim r_\alpha(x_n) = r_0$. Notice that $r_\alpha(x_0) \leq r_0$. Effectively, if we denote $A = \limsup \overline{B}(x_n, r_\alpha(x_n))$, we have

$$1-\alpha \leq \limsup P[\overline{B}(x_n, r_\alpha(x_n))] \leq P(A).$$

On the other hand, if $y \in A$, there exists a subsequence $\{x_{n_k}\}$ such that $y \in \overline{B}[x_{n_k}, r_\alpha(x_{n_k})]$ for every k and therefore

$$r_0 \geq \limsup ||y - x_{n_k}|| \geq ||y-x_0||,$$

where last inequality follows from (P.1) in Lemma 7.1. Thus, A is contained in the closed ball $\overline{B}(x_0, r_0)$ and, in consequence, $r_\alpha(x_0) \leq r_0$.

Let us, at first, assume that $\{||x_n - x_0||\}$ converges to zero. In this case

$$r_0 = r_\alpha(x_0). \tag{7.3}$$

Effectively, otherwise, since

$$1 - \alpha \geq \liminf P[B(x_n, r_\alpha(x_n)] \geq P[B(x_0, r_0)],$$

the only remaining possibility is that $P[B(x_0, r_0)] = 1 - \alpha$. Therefore there exists $r < r_0$ such that $P[B(x_0, r)] = 1 - \alpha$ and, from the hypothesis, we have that from an index onward

$$B(x_0, r) \subset B(x_n, r_\alpha(x_n)),$$

being this inclusion strict. Thus, there exists $s \in (r, r_\alpha(x_n))$ such that

$$B(x_0, r) \subset B(x_n, s),$$

from where $r_\alpha(x_n) \leq s$, which is not possible. In consequence, the assumptions of Proposition 7.2 are satisfied and we have that

$$I_\alpha(P) = \lim_n D_\alpha(x_n, P) = \int ||y - x_0||^2 \tau_{x_0}(y) dP(y),$$

and the theorem is proved in this case.

The proof will be complete if we show that $||x_n - x_0|| \to 0$ as $n \to \infty$. Define for $\delta > 0$ and a natural number n, the sets

$$\begin{aligned} A_\delta &:= \{y : \liminf ||y - x_n||^2 > ||y - x_0||^2 + \delta\} \\ A_\delta^n &:= \{y : ||y - x_k||^2 > ||y - x_0||^2 + \delta, \text{ for every } k \geq n\}. \end{aligned}$$

It happens that $\lim_n P(A_\delta^n) = P(A_\delta)$ and, from (P.2) in Lemma 7.1, that $\lim_{\delta \to 0+} P(A_\delta) = 1$. Fix $\delta_0 > 0$ such that $P(A_{\delta_0}) = \alpha + \eta_0$ with $\eta_0 > 0$. Let $\epsilon > 0$. There exists $\delta < \delta_0$ such that $P(A_\delta) > 1 - \epsilon$. On the other hand, from the fact that $||x_0|| < \liminf ||x_n||$, we have that, for all $y \in E$,

$$||y - x_0||^2 \tau_n(y) \leq [2 \sup ||x_n|| + \sup r_\alpha(x_n)]^2 =: R.$$

Therefore

$$\begin{aligned} &\lim \int \left(||y - x_n||^2 - ||y - x_0||^2\right) \tau_n(y) dP(y) \\ &= \lim \Bigg[\int_{A_{\delta_0}^n} \left(||y - x_n||^2 - ||y - x_0||^2\right) \tau_n(y) dP(y) \\ &\qquad + \int_{\left(A_{\delta_0}^n\right)^c \bigcap A_\delta^n} \left(||y - x_n||^2 - ||y - x_0||^2\right) \tau_n(y) dP(y) \\ &\qquad + \int_{\left(A_{\delta_0}\right)^c \bigcap (A_\delta)^c} \left(||y - x_n||^2 - ||y - x_0||^2\right) \tau_n(y) dP(y) \Bigg] \\ &\geq \lim_n \left[\delta_0 \eta_0 - R\epsilon\right], \end{aligned}$$

and taking limits on ϵ , we finally have that

$$\begin{aligned} I_\alpha(P) &= \lim \int ||y - x_n||^2 \tau_n(y) dP(y) \\ &\geq \delta_0 \eta_0 + \lim \int ||y - x_0||^2 \tau_n(y) dP(y) \\ &\geq \delta_0 \eta_0 + \int ||y - x_0||^2 \tau_{x_0}(y) dP(y), \end{aligned}$$

which contradicts the definition of $I_\alpha(P)$. □

In our proof of Theorem 3.7 we will employ the very well known Skorohod representation theorem which we state here for the sake of completeness.

LEMMA 7.3. *Let $\{P_n\}$ be a sequence of probability measures which converges in distribution to the probability measure P. Then, there exist a probability space $(\mathcal{X}, \mathcal{S}, \nu)$ and a sequence of E-valued random elements $\{Y_n : n = 0, 1, ...\}$ defined on it such that*

(1) *The distribution of Y_0 is P and for $n \geq 1$ the distribution of Y_n is P_n.*
(2) *The sequence $\{Y_n\}$ converges ν-almost surely to Y_0.*

PROPOSITION 7.4. *Let $\alpha \in (0,1)$ and let $\{P_n\}$ be a sequence of probability measures which converges in distribution to the probability measure P. Then it is satisfied that*

$$\limsup I_\alpha(P_n) \leq I_\alpha(P).$$

PROOF. Given $r > 0$, the function

$$y :\longrightarrow ||y - m_P||^2 I_{B(m_P, r)}(y)$$

is bounded and, if $P[S(m_P, r)] = 0$ it is also P-a.e. continuous.

Let us assume that $P[\overline{B}(m_P, r_\alpha(m_P))] = 1 - \alpha$. Therefore, we can take $\tau_{m_P} = I_{\overline{B}(m_P, r_\alpha(m_P))}$. Let $r > r_\alpha(m_P)$ be such that $P[S(m_P, r)] = 0$ and that $P[B(m_P, r)] > 1 - \alpha$. We have that

$$\begin{aligned} \limsup I_\alpha(P_n) &\leq \limsup \int_{B(m_P, r)} ||y - m_P||^2 dP_n(y) \\ &= \int_{B(m_P, r)} ||y - m_P||^2 dP(y), \end{aligned}$$

and, if we take limits on r, we obtain that

$$\limsup I_\alpha(P_n) \leq I_\alpha(P).$$

Now, let us assume that $P[\overline{B}(m_P, r_\alpha(m_P))] > 1 - \alpha$. By definition of $r_\alpha(m_P)$, if $s < r_\alpha(m_P)$, then $P[\overline{B}(m_P, s)] < 1 - \alpha$.

Let $s < r_\alpha(m_P) < r$ be such that $P[S(m_P, s)] = P[S(m_P, r)] = 0$. Therefore, there exists $n_0 \in I\!N$ such that if $n \geq n_0$, then

$$P_n[\overline{B}(m_P, s)] < 1 - \alpha < P_n[B(m_P, r)].$$

From an index onward, there exists $\tau^n_{r,s}$ a trim at level α for P_n such that

$$I_{\overline{B}(m_P, s)} \leq \tau^n_{r,s} \leq I_{B(m_P, r)}.$$

We have that

$$\begin{aligned}\limsup I_\alpha(P_n) &\leq \limsup \int ||y-m_P||^2 \tau^n_{r,s}(y)dP(y)\\ &\leq \lim \int_{\overline{B}(m_p,s)} ||y-m_P||^2\tau^n_{r,s}(y)dP_n(y)\\ &\quad + r^2\left(1-\alpha-P_n[\overline{B}(m_p,s)]\right)\\ &= \int_{\overline{B}(m_p,s)} ||y-m_P||^2 dP(y) + r^2\left(1-\alpha-P[\overline{B}(m_p,s)]\right).\end{aligned}$$

From here, if we take limits, simultaneously, on $s \to r_\alpha(m_P)$ and $r \to r_\alpha(m_P)$, we have that

$$\begin{aligned}\limsup I_\alpha(P_n) &\leq \int_{B(m_p,r_\alpha(m_P))} ||y-m_P||^2 dP(y)\\ &\quad + r^2_\alpha(m_P)\left(1-\alpha-P[B(m_P,r_\alpha(m_P))]\right)\\ &= \int ||y-m_P||^2\tau(y)dP(y),\end{aligned}$$

where, if we take

$$h=\frac{1-\alpha-P[B(m_P,r_\alpha(m_P))]}{P[\overline{B}(m_P,r_\alpha(m_P))]-P[B(m_P,r_\alpha(m_P))]}$$

the function τ is just

$$\tau(y)=\begin{cases}1 & \text{if } y\in B(m_P,r_\alpha(m_P)),\\ 0 & \text{if } y\notin \overline{B}(m_P,r_\alpha(m_P)),\\ h & \text{if } y\in S(m_P,r_\alpha(m_P)),\end{cases}$$

i.e. τ is a trim at level α for P and the proposition is proved. □

Proof of Theorem 3.7.

Proof. Let $H>0$ be such that $P[B(m_P,H)]>\alpha$ and that $[S(m_P,H)]=0$. By hypothesis, from an index onward,

$$P_n[B(m_P,H)]>\alpha.$$

Thus, taking into account Proposition 7.4, we can apply a similar argument to the one developed in Theorem 3.6 to show that both sequences $\{m_n\}$ and $\{r_n\}$ are bounded. Therefore, the theorem will be proved if we show that every weakly convergent subsequence of $\{m_n\}$ converges in norm and its limit is m_P, and that every convergent subsequence of $\{r_n\}$ converges to $r_\alpha(m_P)$.

Our first step is to show that if a subsequence $\{m_{n_k}\}$ converges weakly to m, then $\lim ||m_{n_k}-m||=0$ is satisfied. To prove this, let us assume that $\{m_{n_k}\}$ is a subsequence which does not satisfy this property and let $\{Y_n, n=0,1,2,...\}$ be the sequence of r.e.'s obtained applying Lemma 7.3 to $\{P_n\}$ and P. We would have that

$$\liminf ||Y_{n_k}(t)-m_{n_k}|| > ||Y_0(t)-m|| = \lim ||Y_{n_k}(t)-m||,$$

for ν-a.e. $t\in\mathcal{X}$. Thus, if we consider the sets

$$\begin{aligned}A_\delta &:= \{y : \liminf ||Y_{n_k}-m_{n_k}|| > \lim ||Y_{n_k}-m||+\delta\}\\ A^h_\delta &:= \{y : ||Y_{n_k}-m_{n_k}|| > ||Y_{n_k}-m||+\delta, \text{ for every } k\geq h\}.\end{aligned}$$

we can use a similar argument to the one in Theorem 3.6 to obtain a contradiction.

Let now $\{m_{n_k}\}$ be a subsequence which converges in norm to $m \in E$. Take a subsequence $\{m_{n'_k}\}$ such that there exists $r_0 = \lim_{k\to\infty} r_{n'_k}$. With the same argument as in (7.3) we get that $r_0 = r_\alpha(m)$. On the other hand, if we take $s < r_0$ and $t \in \mathcal{X}$ such that $Y_0(t) \in B(m,s)$, then

$$||Y_{n'_k}(t) - m_{n'_k}||^2 \tau_{n'_k}(Y_{n'_k}(t)) \to ||Y_0(t) - m||^2.$$

Therefore with a similar argument to the one used at the final part of the proof of Proposition 7.4 we obtain that there exists a trim at level α for P, τ, such that

$$\begin{aligned}
\liminf I_\alpha(P_n) &= \liminf \int ||y - m_{n'_k}||^2 \tau_{n'_k}(y) dP_n(y) \\
&= \liminf \int ||Y_{n'_k}(t) - m_{n'_k}||^2 \tau_{n'_k}(Y_{n'_k}(t)) d\nu(t) \\
&= \int ||Y_0(t) - m||^2 \tau(Y_0(t)) d\nu(t) \\
&= \int ||y - m||^2 \tau(y) dP(y) \\
&\geq I_\alpha(P), \qquad (7.4)
\end{aligned}$$

where τ is defined as in Proposition 7.4. Therefore, Proposition 7.4 implies that $m = m_P$ and that $\tau = \tau_P$.

It remains to show that $\lim r_n = r(P)$. But this has been already proved, because, in fact, we have shown that the whole sequence $\{r_n\}$ is bounded and that every convergent subsequence of it converges to $r_\alpha(m_P, P)$. □

Proof of Corollary 3.9.

Proof. We will employ the same notation as in Theorem 3.7. Let us consider a strictly increasing sequence of natural numbers $\{n_k\}$. According to Remark 3.8, there exists a subsequence $\{n'_k\}$ such that the sequence $\{m_{n'_k}\}$ converges in norm to an α-trimmed mean of P, m_P.

Therefore, if we apply (7.4) to this subsequence, we have that

$$\liminf I_\alpha(P_{n'_k}) \geq I_\alpha(P).$$

By Proposition 7.4, we have that every subsequence of $\{I_\alpha(P_n)\}$ contains a further subsequence which converges to $I_\alpha(P)$. But this is impossible unless $\lim I_\alpha(P_n) = I_\alpha(P)$. □

Proof of Theorem 3.11.

Proof. Given $z = (z_1, ..., z_n) \in E^n$ or $z = (z_1, z_2, ...) \in E^\infty$, let us denote by $m_n(z)$ the α-trimmed mean of the probability measure $P_n^z = n^{-1} \sum_{i\leq n} \delta_{z_i}$.

Let us suppose that the theorem does not hold. If so, there exists $A \subset E^\infty$ with $P^\infty(A) > 0$ and such that if $x \in A$, then m_n is not resistant at x.

According to Theorem 3.10, the set

$$B := \{x \in E^\infty : m_n(x) \to m_P\}$$

satisfies that $P^\infty(B) = 1$. Without loss of generality we can also assume that for every $x \in B$, the sequence $\{P_n^x\}$ converges in distribution to P.

Now, let $x \in A \cap B$. According to the definition, there exists $\epsilon > 0$ and three sequences $\{\delta_k\}$, $\{n_k\}$ and $\{y_{n_k}\}$ such that $\delta_k > 0$ and $\lim \delta_k = 0$; $n_k \in I\!N$ and $\lim n_k = \infty$ and $d_{n_k}(x_{n_k}, y_{n_k}) < \delta_k$ and $||m_{n_k}(x_{n_k}) - m_{n_k}(y_{n_k})|| > \epsilon$.

By Lemma 2.1 in [**3**], we have that if ρ denotes for the Prohorov metric, then

$$\lim_k \rho(P^x_{n_k}, P^{y_{n_k}}_{n_k}) = 0,$$

and in consequence, we have that the sequence of probability measures $\{P^{y_{n_k}}_{n_k}\}$ converges weakly to P. Thus, according to Theorem 3.7, we have also that

$$\lim_k m_{n_k}(y_{n_k}) = m_P,$$

what gives a contradiction with the fact that $||m_{n_k}(x_{n_k}) - m_{n_k}(y_{n_k})|| > \epsilon$ for every k. □

In order to prove Theorem 4.4, we need to introduce some additional notation. Given $A \subset E$ and $n \in I\!N$, A_n will be the projection of A on the subspace generated by $\{e_1, ..., e_n\}$, $A_{n,1}$ its projection on the subspace generated by $\{e_2, ..., e_n\}$, and $A_{\infty,n}$ the projection on the subspace generated by $\{e_{n+1}, e_{n+2}, ...\}$. $-A$ will be the symmetrical set of A with respect to 0 and the subspace $\{e_2, e_3, ...\}$; while, given $\delta \in I\!R$,

$$A^\delta := \{x + \delta e_1 : x \in A\}.$$

Given $x \in E$ we will denote, for instance $x_n = \{x\}_n$ and so on. We will abuse of notation and $A_{n,1}$ will also stand for the coordinates set (in R^n) with respect to the fixed orthonormal basis of E of the set A.

The proof of Theorem 4.4 will be based on the following two propositions.

Proposition 7.5. *Under the hypotheses in Theorem 4.4, if $A \subset E$ is a closed and bounded set such that there exists $\delta > 0$ satisfying that $A_1 + \delta \subset I\!R^-$, then $P[A^\delta] > P[A]$.*

Proof. Notice that the hypothesis implies that $A_1 \subset I\!R^-$. Therefore, the assumptions imply that if $a \in A_1$, then $f_1(a + \delta) > f_1(a)$. Moreover, since A is bounded, we have that A_1 is compact and there exists $\eta > 1$ such that

$$\inf_{a \in A_1} f_1(a + \delta)/f_1(a) \geq \eta,$$

which implies that

$$\begin{aligned} P[(A_n)^\delta] &= \int_{A_n} f_1(x_1 + \delta) f_2(x_2) ... f_n(x_n) dx_1 ... dx_n \\ &\geq \eta \int_{A_n} f_1(x_1) f_2(x_2) ... f_n(x_n) dx_1 ... dx_n = \eta P(A_n). \end{aligned}$$

Having in mind that η depends only on A_1 but not on n, we obtain the result just by taking limits on n. □

Proposition 7.6. *Let us assume the hypotheses in Theorem 4.4. Let $r > 0$. If $m \in E$ satisfies that $m_1 \neq 0$, then*

$$P[\overline{B}(m_{\infty,1}, r)] > P[\overline{B}(m_1, r)]. \tag{7.5}$$

Proof. Let $r > 0$ and $m \in E$ be such that $m_1 > 0$ (the other case is analogous). Let us consider the half-balls

$$B^- = \overline{B}(m,r) \cap \{x : x_1 \leq m_1\} \text{ and } B^+ = \overline{B}(m,r) \cap \{x : x_1 > m_1\}.$$

The proof consists on performing some transformations (symmetries and translations) of B^- and B^+ or some subsets of them. Those transformations depend on r. Thus, we need to split the proof according to the possible values of r. Let us assume at first that $r \leq m_1/2$.

By the symmetry assumption, $P[B^+] = P[-B^+]$. Moreover, $(-B^+)_1 + m_1 \subset I\!R^-$ and, by Proposition 7.5, we have that

$$P[(-B^+)^{m_1}] \geq P[B^+].$$

On the other hand, let us call $S(B^-)$ the image of the half-ball B^- by the symmetry determined by the point $m - re_1$ and the subspace generated by $\{e_2, e_3, ...\}$. It follows that, $(S(B^-))_1 \subset I\!R^+$ and $P[S(B^-)] \geq P(B^-)$. On the other hand, $S(B^-)_1^{2r-m_1} \subset I\!R^+$ and, by Proposition 7.5, $P[S(B^-)] \leq P[S(B^-)^{2r-m_1}]$.

Therefore (7.5) is shown in this case since

$$\overline{B}(m_\infty, r) = (S(B^-))^{2r-m_1} \cup (-B^+)^{m_1},$$

and the sets in the right hand side are disjoint.

The next case to be considered is $m_1/2 < r \leq m_1$. In this case we handle B^+ in the same way as in the previous case. The difference is related to B^- because now $S(B^-)_1 \cap I\!R^- \neq \emptyset$. Thus, let us consider the point

$$m^* := m - \frac{m_1}{2} e_1,$$

and let S^* denote the symmetry determined by m^* and the subspace generated by $\{e_2, e_3, ...\}$. Let

$$C := B^- - [B^- \cap S^*(B^-)].$$

Notice that $P[S^*(C)] > P[C]$. Then the proof is also complete in this case since now

$$\overline{B}(m_{\infty,1}, r) = S^*(C) \cup (B^- \cap S^*(B^-)) \cup (-B^+)^{m_1}.$$

Finally let's consider the case $m_1 < r$. Here we need to decompose

$$B^- = B^{-,1} \cup B^{-,2} := \{x \in B^- x_1 \geq 0\} \cup \{x \in B^- x_1 < 0\}.$$

We apply to $B^{-,1}$ the same transformation as we did to B^- in the second case. i.e. we consider

$$C^1 := B^{-,1} - B^{-,1} \cap S^*(B^{-,1}),$$

and handle C^1 as C in previous case, while $B^{-,1} \cap S^*(B^{-,1})$ remains fixed.

Concerning B^+, let us denote $C^2 = S^*(B^{-,2})$. Obviously $C^2 \subset B^+$. C^2 will also remain fixed. The only remaining part is $B^+ \cap (C^2)^c$. Since

$$P[B^+ \cap (C^2)^c] < P[-(B^+ \cap (C^2)^c)^{m_1}],$$

the proof follows since in this case, we have the decomposition

$$\overline{B}(m_{\infty,1}, r) = [(-(B^+ \cap (C^2)^c))^{m_1}] \cup C^2 \cup S^*(C^1) \cup (B^{-,1} \cap (C^1)^c).$$

$\square$

Proof of Theorem 4.4.

Proof. Let $r > 0$ and let $m \in E$, $m \neq 0$. According to Proposition 7.6, if $m_n \neq 0$ then

$$P[\overline{B}(m_{\infty,n-1}, r)] < P[\overline{B}(m_{\infty,n}, r)],$$

while, if $m_n = 0$, $m_{\infty,n-1} = m_{\infty,n}$.

Thus, we have the inequalities

$$P[\overline{B}(m, r)] \leq P[\overline{B}(m_{\infty,1}, r)] \leq P[\overline{B}(m_{\infty,2}, r)] \leq \dots$$

where at least one of them is strict. On the other hand $\lim ||m_{\infty,n}|| = 0$ and we have that

$$P[\overline{B}(m, r)] < \lim P[\overline{B}(m_{\infty,n}, r)] \leq P[\limsup \overline{B}(m_{\infty,n}, r)] = P[\overline{B}(0, r)].$$

□

The proof of the Theorem 5.1 will be based on the following proposition.

Proposition 7.7. *Let us assume that the hypothesis in Theorem 3.7 hold. Let* $\{x_n\} \subset E$ *be a sequence such that* $\lim ||x_n - m_P|| = 0$. *Then*

$$\limsup D_\alpha(x_n, P_n) \leq I_\alpha(P).$$

Proof. Given $n \in I\!N$, let m_n be a trimmed mean of P_n. By Theorem 3.7 we have that $\lim ||m_n - m_P|| = 0$ and that $\lim_n r_n = r(P)$. Thus $\lim ||m_n - x_n|| = 0$ and, if we denote $R := \sup(r_n)$, it follows that

$$||y - m_n||\tau_n(y) \leq R\tau_n(y).$$

Therefore we have that

$$\begin{aligned} D_\alpha(x_n, P_n) &\leq \int ||y - x_n||^2 \tau_n(y) dP(y) \\ &\leq D_\alpha(m_n, P_n) + \left[||m_n - x_n||^2 + 2||m_n - x_n||R\right](1 - \alpha) \\ &= I_\alpha(P_n) + \left[||m_n - x_n||^2 + 2||m_n - x_n||R\right](1 - \alpha), \end{aligned}$$

which, together with Corollary 3.9, gives the result. □

Proof of Theorem 5.1.

Proof. By the extension of the Glivenko-Cantelli Theorem there exists a μ-probability one set Ω_0 such that if $\omega \in \Omega_0$, then m_P belongs to the closure of the set $\{X_1(\omega), X_2(\omega), ...\}$ and the sequence $\{P_n\}$ converges in distribution to P.

Let us fix $\omega \in \Omega_0$. Given $n \in I\!N$, there exists $X_{i_n}(\omega) \in \{X_1(\omega), ..., X_n(\omega)\}$ such that $||X_{i_n}(\omega) - m_P|| \to 0$. Therefore we have that

$$\limsup I_\alpha(P_n) \leq \limsup D_\alpha(X_{i_n}(\omega), P_n) \leq I_\alpha(P), \tag{7.6}$$

where last inequality is a consequence of Proposition 7.7.

Now (7.6) is a starting point analogous to the conclusion of Proposition 7.4 which, if we make here the same reasoning as in Theorem 3.7, allows to obtain the desired result. □

ACKNOWLEDGMENTS. This research was started during a visit of the first author to the Universidad de la República del Uruguay. He wants to thank for the warm hospitality received there. This stay was supported by a grant from the Universidad de Cantabria.

The present version have been completed during a visit of the second author to the Universidad de Cantabria. This stay has been supported by the Instituto de Cooperación Iberoamericana, Programa de Cooperación Interuniversitaria AL-E 2003.

We would also want to thank to Mediametría TV, Corporación Combex, for providing the TV rating data.

References

[1] P. BESSE and J.O. RAMSAY, *Principal component analysis of sampled functions.* Psychometrika **51** (1986), 285-311.
[2] G. BOENTE and R. FRAIMAN, *Kernel-based functional principal components.* Statist. Probab. Lett. **48** (2000), 335-345 .
[3] G. BOENTE, FRAIMAN, R. and V.J. YOHAI, *Qualitative Robustness for Stochastic Processes.* Ann. Statist. **15** (1987), 1293-1312.
[4] D. BOSQ, *Modelization, non-parametric estimation and prediction for continuous time processes.* Nonparametric Functional Estimation and Related Topics, G.G. Roussas et al. eds., Kluwer Academic Publishers, 1991, 509-529.
[5] B.M. BROWN. *Statistical uses of the spatial median.* J. Roy. Statist. Soc. Ser. B, **45** (1983), 25-30.
[6] B.A. BRUMBACK and J.A. RICE, *Smoothing spline models for the analysis of nested and crossed samples of curves (with discussion)* J. Amer. Statist. Assoc., **93** (1998), 961-994.
[7] D. COX, *Metrics on stochastic processes and qualitative robustness* Technical Report 3. Dept. Statistics, Univ. Washington, 1981.
[8] J.A. CUESTA-ALBERTOS, A. GORDALIZA and C. MATRÁN. *Trimmed k-Means: An Attempt to Robustify Quantizers.* Ann. Statist., **25** (1997), 553-576.
[9] J.A. CUESTA-ALBERTOS, A. GORDALIZA and C. MATRÁN, *Trimmed best k-nets: A robustified version of an* L^∞*–based clustering method.* Statist. Probab. Lett., **36** (1998), 401-413.
[10] J. A. CUESTA and C. MATRÁN. *The Strong Law of Large Numbers for k-Means and Best Possible Nets of Banach Valued Random Variables.* Probab. Theory Related Fields, **78** (1988), 523-534.
[11] J.A. CUESTA-ALBERTOS, C. MATRÁN, C. and A. MAYO-ISCAR. *Stem-cell based estimators in the mixture model.* Technical report, 2003.
[12] A. CUEVAS, M. FEBRERO and R. FRAIMAN. *Linear functional regression: the case of fixed design and functional response.* Unpublised manuscript, 2001.
[13] J. DAUXOIS, A. POUSSE and Y. ROMAIN. *Asymptotic theory for principal components analysis of a vector random function: some applications to statistical inference.* J. Multivariate Anal. **12** (1982), 136-154.
[14] D.L. DONOHO and M. GASKO. *Breakdown properties of location estimates based on halfspace depth and projected outlyingness.* Ann. Statist. **20** (1992), 1803-1827.
[15] J. FAN and S.K. LIN. *Test of significance when the data are curves.* J. Amer. Statist. Assoc. **93** (1998), 1007-1021.
[16] F. FERRATY. and P. VIEU. *Functional nonparametric model for scalar response.* Manuscript, 1999.
[17] D. FRAIMAN. *Un análisis de la estructura de consumo eléctrico en Buenos Aires.*Technical Report, 2001.
[18] R. FRAIMAN and J. MELOCHE. *Multivariate L-estimation (with discussion).* Test **8** (1999), 255-317.
[19] R. FRAIMAN and G. MUNIZ. *Trimmed means for functional data.* Test, **10** (2001), 419-440.
[20] R. FRAIMAN and G. PÉREZ IRIBARREN. *Nonparametric regression estimation in models with weak error's structure.* J. Multivariate Anal. **21** (1991), 180-196.
[21] L.A. GARCIA-ESCUDERO, A. GORDALIZA and C. MATRÁN. *A central limit theorem for multivariate generalized trimmed k-means.* Ann. Statist. **27** (1999), 1061-1079.
[22] A. GORDALIZA. *Best approximations to random variables based on trimming procedures.* J. Approx. Theory **64** (1991), 162-180.
[23] F.R. HAMPEL. *A general qualitative definition of robustness.* Ann. Math. Statist. **42** (1971), 1887-1896.

[24] J.D. HART and T.E. WEHRLY. *Kernel regression estimation using repeated measurement data.* J. Amer. Statist. Assoc. **81** (1986), 1080-1088.
[25] T. HASTIE and R. TIBSHIRANI. *Varying-coefficient models.* J. Royal Statist. Soc., Series B **55** (1993), 757-796.
[26] A. KNEIP and T. GASSER. *Statistical tools to analyze data representing a sample of curves.* Ann. Statist., **20** (1992), 1266-1305.
[27] R. LIU. *On the notion of simplicial depth.* Proceedings of the National Academy of Sciences, U.S.A., **85** (1988), 1732-1734.
[28] R. LIU. *On a notion of data depth based on random simplices.* Ann. Statist. **18** (1990), 405-414.
[29] R. LIU and K. SINGH. *A quality index based on data depth and multivariate rank tests.* J. Amer. Statist. Assoc., **421** (1993), 252-260.
[30] N. LOCANTORE, J.S. MARRON, D.G. SIMPSON, N. TRIPOLI, J. T. ZHANG and K.L. COHEN. *Robust principal component analysis for functional data (with discussion).* Test **8** (1999), 1-74.
[31] P.C. MAHALANOBIS. *On the generalized distance in statistics.* Proceedings of the National Academy of India, **12** (1936), 49-55.
[32] A. MAYO. *Estimación de Parámetros en Mezclas de Poblaciones Normales no Homogéneas.* Ph. D. Thesis. Universidad de Valladolid. Spain, 2001.
[33] K. MOSLER. *Multivariate Dispersion, Central Regions and Depth. The Lift Zonoid Approach.* Lecture Notes in Statistics, 165. Springer, New York, 2002.
[34] H. OJA. *Descriptive statistics for multivariate distributions.* Statist. Probab. Lett. **1** (1983), 327-332.
[35] P. PAPANTONI-KAZAKOS and R.M. GRAY. *Robustness of estimates on stationary observations.* Ann. Probab., **7** (1979), 989-1002.
[36] J.O. RAMSAY. *When the data are functions.* Psychometrika , **47** (1982), 379-396.
[37] J.O. RAMSAY and C.J. DALZELL. *Some tools for functional data analysis (with discussion).* J. Roy. Statist. Soc. Ser. B, **52** (1991), 539-572.
[38] J.O. RAMSAY and B.W. SILVERMAN. *Functional Data Analysis.* Springer Series in Statistics, Springer, New York, 1997.
[39] J.O. RAMSAY and B.W. SILVERMAN. *Applied Functional Data Analysis.* Springer Series in Statistics, Springer, New York, 2002.
[40] J. RICE and B.W. SILVERMAN. *Estimating the mean and covariance structure nonparametrically when data are curves.* J. Roy. Statist. Soc. Ser. B **53** (1991), 233-243.
[41] C.G. SMALL. *A survey of multidimensional medians.* Internat. Statist. Rev. **58** (1990), 263-277.
[42] K. SINGH. *A notion of majority depth.* Technical Report. Rutgers University, Department of Statistics, 1991.
[43] J.W. TUKEY. *Mathematics and picturing data.* Proceedings of the International Congress of Mathematics, Vancouver, **2** (1975), 523-531.
[44] A.W. van der VAART and J. A. WELLNER *Weak Convergence and Empirical Processes.* Springer Series in Statistics. Springer, New York, 1996.

Departamento de Matemáticas, Estadística y Computación, Facultad de Ciencias. Avda. los Castros s.n., 39005 SANTANDER, Spain
E-mail address: cuestaj@unican.es

Departamento de Matemática y Ciencias, Universidad de San Andrés, Argentina[2]
E-mail address: rfraiman@cmat.edu.uy

[2]On leave Centro de Matemática, Universidad de la República, Uruguay.