ROBUSTNESS IN PARAMETER ESTIMATION

by

P. PAPANTONI-KAZAKOS

OCTOBER 1975

TECHNICAL REPORT #7517

ROBUSTNESS IN PARAMETER ESTIMATION

by

P. Papantoni-Kazakos
Department of Electrical Engineering
Rice University
Houston, Texas

ABSTRACT

Due to vagueness in the definition of robustness, there has been no natural transition between robust and nonparametric parameter estimators.

In this work, qualitative ideas first expressed by Hampel [1] are extended in an effort to present a theory that unifies the nonparametric and robust concepts.

Robustness is defined in a precise mathematical way that transists to nonparametricness naturally. As a result, some general constructive characteristics of robust estimators are studied.

1. INTRODUCTION -- BACKGROUND

The subject estimation procedures that are distribution free, or distribution insensitive, have attracted several statisticians since the early years of this century. Tukey [2,3] was the first to systematically survey the work done until then on the subject. Hodges and Lehman [4] noticed that estimates of location could be derived from Wilcoxon and other tests and Huber [5] solved asymptotic minimax problems for set of distributions differing less than ϵ from a normal.

As Huber recognizes in his review paper [8], "robustness" has been a vague concept. The term attached to an estimation scheme means little sensitivity to distribution changes (stability) where the sense of this "sensitivity" is also vague. In addition, the "robustness" of a certain estimator is, in general, a local property, valid only inside a particular contaminated distribution family where the concept of contamination has been kept rather specialized in most cases [5]. Hampel [1] was the first to recognize the lack of a solid theory that brings out the serious aspect of stability in robustness and examined the problem in analogy to the stability of mechanical structures.

In the present work, Hampel's theory is extended, precise definitions of "sensitivity" and "contamination" are given, and analogy of the approach to the stability of real functions is discussed. Theorems that bring out the structural properties of the "robust estimators", as defined here, are stated.

2. STATEMENT OF OBJECTIVES -- ANALOGIES TO THE REAL FUNCTION SPACE

The robust procedures were established after it was recognized that a designer can almost never assume that he has complete knowledge of the true statistics of the disturbances appearing in a system. This statistical ambiguity has been called statistical contamination and it has been modeled linearly by Huber. Specifically, Huber has given the following model for a contaminated distribution $F(x)$

$$F(x) = (1-\epsilon)F_o(x) + \epsilon F_1(x), \ \forall\, x \tag{2.1}$$

where ϵ is a positive small real number, $F_o(x)$ a well-defined distribution, and $F_1(x)$ a distribution belonging to a general family such as the symmetric, finite variance one.

A robust procedure is supposed to be performing well, in a sense to be defined, inside a family of contaminated distributions, as described by (1), or some other general contamination criterion. The sense of good performance has remained unclear for the robust procedures and the sense of contamination may be too restrictive.

Before we state definitions and approaches that will hopefully clarify the ambiguities we mentioned, we will explain the notation to be used.

$X_n^T = [x_1, x_2, \ldots, x_n]$	is the observation vector of dimensionality n .
S	is the p-dimensional vector unknown parameter to be estimated.
$Q_n(X/S) = Q(X_n/S)$	is the n-dimensional distribution of the observation vector X_n , conditioned on the value S of the parameter.
$Q_n(S)$	is the general description of the distribution $Q_n(X/S)$ when no particular observation vector is given.
$\hat{S}(X)$	is the estimate of S when the vector X is observed.

The contamination model of Huber in (1), when applied to distributions $Q(X/S)$, becomes

$$Q_n(X/S) = (1-\epsilon)Q_{on}(X/S) + \epsilon Q_{1n}(X/S), \ \forall\ X,S \qquad (2.2)$$

where the dimensionality n is assumed fixed, the distribution $Q_{on}(X/S)$ is well defined, and $Q_{1n}(X/S)$ belongs to a general family of distributions.

Let us choose the family of $Q_{1n}(X/S)$ to be nonparametric, that is such that it cannot be described by a finite number of parameters for given X and S. Then we may observe that if $\epsilon = 1$ in (2), the contaminated family described by $Q_n(X/S)$ is equivalent to the nonparametric family described by $Q_{1n}(X/S)$. Also, if $\epsilon = 0$, $Q_n(X/S)$ becomes equal to the parametric, well-defined distribution $Q_{on}(X/S)$. These observations become our first initiative toward the formulation of a theory that connects the robust with the nonparametric concepts.

Let us further observe here that if (2) is true, then it is easily derived that

$$|Q_n(X/S) - Q_{on}(X/S)| \leq \epsilon, \; \forall\, X,S \tag{2.3}$$

The inequality in (3) is more general than (2) and the absolute value on the left-hand side of it describes a particular distance measure between the two distributions $Q_n(S)$ and $Q_{on}(S)$ namely the Kolmogorov one. Distance measures will be discussed in more detail in the following section.

In the continuation of the present section, let us just denote by $d\,(Q_n(S),Q_{on}(S))$ the distance between $Q_n(S)$ and $Q_{on}(S)$ when the arbitrary measure $d(\cdot,\cdot)$ has been assigned. The particular choice of the measure $d(\cdot,\cdot)$ will depend on the objectives of the particular problem. Using the arbitrary distance $d(\cdot,\cdot)$ we will now proceed to defining statistical contamination in a general sense.

In this definition the parametric distribution $Q_{on}(S)$ as well as the size ϵ of the contaminated family will be included.

Definition 2.1

Let $Q_{on}(S)$ be a well-defined distribution for every value of the parameter S and for fixed positive integer n. Then, a family $\mathcal{F}_n = \mathcal{F}(Q_n(S))$ of distributions is called Q_{on}-(d) contaminated with degree of contamination ϵ, iff for every value of the parameter S the members of the family cannot be, in general, described (with the exception of $Q_{on}(S)$ itself) by a finite number of parameters and if, in addition,

$$\sup_{Q_n \in \mathcal{F}_n} \sup_{S} d(Q_n(S), Q_{on}(S)) = \epsilon \qquad (2.4)$$

We may observe that the distance measure used is included in the definition of a contaminated distribution family. For given ϵ, the families that are Q_{on}-contaminated depend heavily on the $d(\cdot,\cdot)$ choice. Also, according to definition, the distribution family that contains $Q_n(X/S)$ in (2) is Q_{on} - (d_k) contaminated with degree of contamination ϵ, where

$$d_k(Q_1,Q_2) = \sup_X |Q_1(X) - Q_2(X)|$$

By definition 2.1, a Q_{on} - (d) contaminated family of distributions with degree of contamination one is a typical nonparametric family, while to degree of contamination zero the parametric, well-defined distribution Q_{on} corresponds.

Definition 2.1 can be extended to distribution of continuous observed waveforms (instead of observation vectors) in the following way:

Definition 2.2

Let $Q_o(S)$ describe a well-defined distribution $Q_o(x(t)/S)$ for every value of the parameter S and the continuous waveform $x(t)$. Then a family $\mathcal{F} = \mathcal{F}(Q(S))$ of distributions $Q(x(t)/S)$ is called Q_o - (d) contaminated with degree of contamination ϵ, iff for every sampling choice $t_1, \ldots, t_n$ and every n, the $Q_n(x(t_1), \ldots, x(t_n)/S)$ distributions that result from the distributions $Q(S)$ in $\mathcal{F}$, consist of a Q_{on} - (d) contaminated family with degree of contamination ϵ, where Q_{on} represents $Q_o(x(t_1), \ldots, x(t_n)/S)$.

Now that contamination has been defined in a precise general manner, we want to return to our ultimate objective of describing robustness qualitatively. What we actually require from a robust estimator is as little sensitivity as possible to small statistical variations of the data. This sensitivity was first interpreted by Hampel [1] as stability of the statistics of the estimator. By another interpretation, analogies from the theory of real functions can be used. In particular, let $f(x)$ be a real function of x. Then, if $f(x)$ is continuous at the point x_o, small perturbations from it will result in small perturbations of $f(x)$ from $f(x_o)$. This idea can be extended to the space of the estimators if they are considered as mappings from one distribution space to another. Specifically, let us consider the space of the distributions Q_n of the observation vector X_n. Then a certain estimator $\hat{S}_n(X) = \hat{S}(X_n)$ corresponds to a certain distribution Q_n, a distribution D_{Q_n} that is p-dimensional and describes the statistics of $\hat{S}(X_n)$ when X_n is Q_n-distributed. Therefore, for fixed n, the estimator $\hat{S}_n(X)$ can be seen as a mapping from the distribution space Q_n to the distribution space D_{Q_n}.

$$Q_n \xrightarrow{\hat{S}_n} D_{Q_n}$$

Extending the idea of continuity of real functions to the distribution mapping, one ends up with a requirement that guarantees small variations (in sense described by a distance measure) of the distribution D_{Q_n} for local small perturbations of the distribution Q_n. That idea was used by Hampel [1].

Let us suppose now that we want to require similar statistical stability of the estimator for a whole contaminated family of distributions Q_n. Then the extension of the concept of continuity is not good enough. Indeed, let us use again the analogy with the real functions. A real function $f(x)$ that is continuous inside a whole x-interval may vary abruptly with varying x. That would result in large perturbations of $f(x)$ for quite small x-variations. Such an $f(x)$ function is shown in Fig. 1. However, if the function $f(x)$ were differentiable in the interval $[a,b]$, the abrupt changes of the $f(x)$ value would not appear (Fig. 2) and for $[a,b]$ length small, the values of the function $f(x)$ will be quite close to the value $f(x_o)$ of the central point $x_o = a + (b-a)/2$. We will add here that we may require the function $f(x)$ to satisfy a stronger property, namely the property of absolute continuity. It turns out that the extension of this last property from the space of real functions to the space of the estimators has certain constructive advantages that will become clear in section 4.

So, robustness will be defined as a property of the estimator $\hat{S}_n$ that corresponds to absolute continuity of real functions where $\hat{S}_n$ is considered a mapping from the n-dimensional distribution space Q_n to the p-dimensional distribution space D_{Q_n}. We will observe here that for varying n, $\hat{S}_n$ forms a sequence of functions that have double identity. They are mappings from and onto distribution spaces and, at the same time, real functions from the Euclidean space E^n of the observation vector X_n onto the Euclidean space E^p of the parameter S . This double identity will be very useful in our effort to find constructive properties of the robust estimators where "robustness" will be defined precisely in section 4.

In the following section we will discuss briefly distribution distance measures.

3. DISTRIBUTION DISTANCE MEASURES

In definition 2.1 of the previous section, an arbitrary distribution distance $d(\cdot,\cdot)$ was used. Such a distance measure must satisfy, in general, certain conditions. Specifically, it must be such that: For Q_{1n}, Q_{2n} any n-dimensional discrete distributions

i) $d(Q_{1n},Q_{2n}) = 0$ if and only if $Q_{1n} \equiv Q_{2n}$

ii) $0 \leq d(Q_{1n},Q_{2n}) \leq 1$

iii) $d(Q_{1n},Q_{2n}) = d(Q_{2n},Q_{1n})$

Two particular distance measures that are used widely are the Kolmogorov and the Lèvy ones. The Kolmogorov distance denoted by $d_K(Q_1,Q_2)$ is defined by the following equation

$$d_K(Q_{1n},Q_{2n}) = \sup_X |Q_{1n}(X) - Q_{2n}(X)| \qquad (3.1)$$

The Lèvy distance, denoted by $d_L(Q_1,Q_2)$ is defined as

$$d_L(Q_{1n},Q_{2n}) = \inf\{\epsilon: Q_{1n}(X) \leq Q_{2n}(X_\epsilon) + \epsilon \ , \ \forall X\} \qquad (3.2)$$

where X_ϵ is such an n-dimensional data vector, that

$$(-\infty, X_\epsilon) = \cup\{(-\infty,Y): d_{on}(Y,X) \leq \epsilon\}$$

$$= \{Y: \exists X_1 \in (-\infty,X): d_{on}(X_1,Y) \leq \epsilon\} \qquad (3.3)$$

The measure $d_{on}(\cdot,\cdot)$ in expression (3.3) is a properly chosen distance measure between two n-dimensional vectors and can be given, for example, by either of the following two relationships

$$d_{on}(X,Y) = \sum_{i=1}^{n} |x_i - y_i| \tag{3.4}$$

$$d_{on}(X,Y) = \max_{1 \leq i \leq n} |x_i - y_i| \tag{3.5}$$

where $X = \{x_i\}$, $Y = \{y_i\}$. Therefore, the Lèvy distance depends heavily on the choice of the measure $d_{on}(X,Y)$. We must further observe here that it is always true (for the arbitrary n-dimensional distributions Q_{1n}, Q_{2n}) that $d_L(Q_{1n},Q_{2n}) \leq d_K(Q_{1n},Q_{2n})$.

Convergence in Kolmogorov sense is a stronger property than the convergence in Lèvy. If the Lèvy distance is used, in definition 2.1 for example, a similar definition with Kolmogorov distance instead will be stronger.

For further discussion in this section as well as in the sections that will follow, we will need the following notation

F : denotes one-dimensional distribution defined as $F(x)$, $-\infty \leq x \leq \infty$.

F^n : denotes the n-dimensional distribution that corresponds to n independent experiments, each with outcome distributed according to F.

Let F_1, F_2 be two different one-dimensional distributions. Then, the following lemma can be stated.

Lemma 3.1

Under distance $d_{on}(\cdot,\cdot)$ described by either (3.4) or (3.5), it is true that

$$d_L(F_1,F_2) \leq d_L(F_1^n,F_2^n) \quad , \forall\, n \tag{3.6}$$

It is also true that

$$d_K(F_1,F_2) \leq d_K(F_1^n,F_2^n) \quad , \forall\, n \tag{3.7}$$

<u>Proof</u>

i) If ε_1 is a candidate for $d_L(F_1^n,F_2^n)$, as defined by (3.2), then

$$F_1(x_1)\ldots F_1(x_n) \leq F_2(x_1+\rho_1)\ldots F_2(x_n+\rho_n) + \varepsilon_1 \quad , \forall\, \{x_i\} \tag{3.8}$$

where $X_{\varepsilon_1}^T = [x_1+\rho_1,\ldots,x_n+\rho_n]$, $\rho_i > 0$ and $\begin{cases} \Sigma\rho_i = \varepsilon_1, \text{ if (3.4) true} \\ \max\limits_{1\leq i\leq n} \rho_i = \varepsilon_1, \text{ if (3.5) true} \end{cases}$

Choose $\{x_i\} : x_2 = x_3 = \ldots = x_n = \infty$. Then, it is obtain from (3.8)

$$F_1(x_1) \leq F_2(x_1+\rho_1) + \varepsilon_1 \leq F_2(x_1+\varepsilon_1) + \varepsilon_1 \quad , \forall\, x_1 \Rightarrow \varepsilon_1$$

is a candidate for $d_L(F_1,F_2)$, hence

$$d_L(F_1,F_2) \leq d_L(F_1^n,F_2^n) \quad , \forall\, n$$

ii) $d_K(F_1^n,F_2^n) = \sup\limits_X |F_1^n(X) - F_2^n(X)| \Rightarrow$

$$\Rightarrow d_K(F_1^n,F_2^n) \geq \sup_{x_1} |F_1(x_1)F_1(\infty)\ldots F_1(\infty) - F_2(x_1)F_2(\infty)\ldots F_2(\infty)| =$$

$$= \sup_x |F_1(x) - F_2(x)| = d_K(F_1,F_2)$$

Let F_1,F_2 be two different one-dimensional distributions and let n independent experiments be contacted from each of them. We would like a distance measure between these two multidimensional distributions that is equal to the initial marginal distance $d(F_1,F_2)$. Similarly, for arbitrary waveform distributions $Q_1(x(t)/S)$, $Q_2(x(t)/S)$ a unique distance measure is desirable, that will be independent of particular sampling methods and sample sizes, and will just depend on the assigned S value. These desirable properties of the distribution distance together with the conclusion of lemma 3.1 lead to the following definition of distance between $Q_1(S)$ and $Q_2(S)$.

Definition 3.1

Given two, in general waveform distributions $Q_1(x(t)/S)$ and $Q_2(x(t)/S)$, their distance is defined as

$$d(Q_1(S),Q_2(S)) = \inf_{\substack{\text{sampling methods} \\ \text{complying with } S}} \inf_n d(Q_{1n}(S),Q_{2n}(S)) \qquad (3.9)$$

The distance $d(\cdot,\cdot)$ in (3.9) can be the Kolmogorov or Lèvy one. On the right hand part of (3.9) a certain sampling method that complies with the given S is assumed. If S is only a location parameter, sampling methods and statistics are not influenced by its values. However, that is not true if S includes correlation components.

We will finally observe that if definition 3.1 is taken into consideration definition 2.2 in section 2 can be substituted by the following one.

Definition 3.2

Let $Q_o(S)$ describe a well-defined distribution $Q_o(x(t)/S)$ for every value of the parameter S and the continuous waveform $x(t)$. Then, a family $\mathcal{F} = \mathcal{F}(Q(S))$ of distributions $Q(x(t)/S)$ is called Q_o - (d) contaminated with degree of contamination ϵ , iff for every S the members of the family cannot be in general described by a finite number of parameters, and if, in addition

$$\sup_{Q \in \mathcal{F}} \sup_{S} d(Q_1(S), Q_o(S)) = \epsilon \tag{3.10}$$

where $d(Q_1(S), Q_o(S))$ defined by (3.9).

In the following sections, the distribution distance measure $d(\cdot,\cdot)$, as defined in definition 3.1, will be used.

4. DEFINITION OF ROBUSTNESS

We will start this section by considering a data vector X available to the receiver. The estimate of S that is a function of this vector X will be denoted $\hat{S}_n(X) = \hat{S}(X_n)$ whenever X is equal to X_n that is n-dimensional. If the dimensionality of X increases, a sequence $\{\hat{S}_n\}$ of estimators is formed. This sequence can be looked at either as a sequence of real functions from the E^n Euclidean space to the E^p one (p is the dimensionality of the parameter S) or as a sequence of mappings from the n-dimensional distribution space of the vector X_n to the p-dimensional distribution space of $\hat{S}_n(X)$.

As explained in section 2, robustness will be defined here as a distribution stability property. In other words, a sequence $\{\hat{S}_n\}$ of estimators that has robust properties should be such that if the distribution of the data-vector X changes slightly, no dramatic changes occur in the distribution of the variable $\hat{S}_n(X)$. This idea was first introduced by Hampel [1] who defined robustness in a way analogous to the continuity of real functions. Here we will define robustness as a much stronger property. In particular, in our effort to unify robustness and nonparametric behavior we will extend the concept of absolute continuity as explained at the end of section 2. What

Hampel called robust, we will call weakly-robust. We will discuss the definitions in more detail as they appear.

Definition 4.1

A sequence of estimators $\{\hat{S}_n\}$ is called weakly-(d_1,d_2) robust at $Q_o(S)$ iff:

Given $\epsilon > 0$, there is a $\delta(\epsilon,S) > 0$ for every n , such that:

For every distribution $Q(S)$ satisfying

$$d_1(Q_o(S),Q(S)) < \delta(\epsilon,S)$$

it is implied that

$$d_2(D_{Q_o(S)}(\hat{S}_n),\ D_{Q(S)}(\hat{S}_n)) < \epsilon$$

In definition 4.1, $D_{Q(S)}(\hat{S}_n)$ denotes the p-dimensional distribution of the estimate $\hat{S}_n(X) = \hat{S}_n(X_n)$ that corresponds to X_n data-vector distributed according to $Q(S)$. Also, the distance measures for the distributions $Q(S)$ and $D_{Q(S)}(\hat{S}_n)$ are not necessarily the same. The particular measure choice is included in the definition. It can be observed that weak-robustness has been defined exactly as continuity of a certain real function at a given point where the function is a mapping from distribution space onto distribution space and the given point is the well-defined distribution $Q_o(S)$. Finally, the exact robust behavior of the sequence $\{\hat{S}_n\}$ is in general different for different values of the parameter S because of the dependence of $\delta(\epsilon,S)$ on it. A stronger property that eliminates that dependence is stated by the following definition.

Definition 4.2

A sequence of estimators $\{\hat{S}_n\}$ is called uniformly weakly-(d_1,d_2) robust at $Q_o(\cdot)$ iff it is weakly-(d_1,d_2) robust at $Q_o(S)$ for every S and iff, in addition, $\delta(\epsilon,S) = \delta(\epsilon)$ in definition 4.1.

In definition 4.2, $Q_o(\cdot)$ denotes the class of all $Q_o(S)$ distributions for the whole range of S values.

The weak-robustness, as defined by definition 4.1, is characterized by stability properties that are the same for any number of data that might be observed. For systems such that a large number of data is an accessible goal, a less strong asymptotic property might be satisfactory. This property is expressed by the following definition.

Definition 4.3

A sequence of estimators $\{\hat{S}_n\}$ is called asymptotically weakly-(d_1,d_2) robust at $Q_o(S)$ iff

Given $\epsilon > 0$, there is a positive integer n_o and a $\delta(\epsilon,S) > 0$ such that

For every $Q(S)$ satisfying

$$d_1(Q_o(S),Q(S)) < \delta(\epsilon,S)$$

and for every $n \geq n_o$, it is implied that

$$d_2(D_{Q_o(S)}(\hat{S}_n),D_{Q(S)}(\hat{S}_n)) < \epsilon$$

In analogy to the concept of continuity of a real function inside a whole set of values of its variable, we will define weak-robustness inside a family of distributions. Such a family will be centered around a well-defined distribution $Q(S)$ and its size will be characterized by a degree of contamination. Specifically, the following definition is given.

Definition 4.4

A sequence of estimators $\{\hat{S}_n\}$ is called ζ-weakly-(d_1,d_2) robust at $Q_o(\cdot)$ iff it is weakly-(d_1,d_2) robust at every $Q(S)$ that is a member of a Q_o-(d_1) contaminated family $\mathcal{F}(Q(S))$ with degree of contamination ζ.

As we said in section 2, if a real function $f(x)$ is continuous inside a whole x-interval, there is no guarantee that its values will not take large excursions in it. Similarly, if a sequence of estimators in ζ-weakly robust at a certain distribution, there is no guarantee that the estimator distributions will not vary a lot inside the Q_o-contaminated family. A property that prevents large such variations for real functions is the absolute continuity. In analogy, robustness is defined below as a property of a sequence of estimators that operate on a family of data distributions.

Definition 4.5

A sequence of estimators $\{\hat{S}_n\}$ is called ζ-(d_1,d_2) robust at $Q_o(S)$ iff:

Given $\epsilon > 0$, there is a $\delta(\epsilon,S) > 0$ for every positive integer n, such that

For all distribution pairs $\{Q_i(S),Q_i'(S)\}$ that are finite in number, that cannot be described by a finite number of parameters and are such that

$$d_1(Q_o(S),Q_i(S)) < \zeta$$

$$d_1(Q_o(S),Q_i'(S)) < \zeta$$

and $$\sum_i d_1(Q_i(S),Q_i'(S)) < \delta(\varepsilon,S)$$

it is implied that

$$\sum_i d_2(D_{Q_i(S)}(\hat{S}_n),D_{Q_i'(S)}(\hat{S}_n)) < \varepsilon$$

It is clear from the statement of definition 4.5 that it is defined in direct analogy to absolute robustness of real functions where the estimators are being looked at as mappings from and onto distribution spaces and the data distribution space considered is a $Q_o(S)$-(d_1) contaminated family with degree of contamination ζ and fixed value S .

For systems such that the number of data is not restricted, an asymptotic version of robustness might be useful. This concept is given by the following definition.

<u>Definition 4.6</u>

A sequence $\{\hat{S}_n\}$ of estimators is called asymptotically ζ-(d_1,d_2) robust at $Q_o(S)$ iff:

given $\varepsilon > 0$, there is a $\delta(\varepsilon,S) > 0$ and a positive integer n_o , such that:

for every $n \geq n_o$ and every finite choice of distribution pairs $\{Q_i(S), Q_i'(S)\}$ that satisfy

$$d_1(Q_o(S),Q_i(S)) < \zeta$$

$$d_1(Q_o(S),Q_i'(S)) < \zeta$$

$$\sum_i d_1(Q_i(S), Q_i'(S)) < \delta(\epsilon, S)$$

it is implied that

$$\sum_i d_2(D_{Q_i(S)}(\hat{S}_n), D_{Q_i'(S)}(\hat{S}_n)) < \epsilon$$

With definition 4.6 we have completed the series of precise definitions on robustness. It will be only pointed out here that the same definitions can be extended to the case of waveform observations. Then, instead of having the sequence $\{\hat{S}_n\}$ and the integers n, n_o, we will have $\hat{S}_T$, T and T_o, where T indicates time. In section 5, a study will be offered on some constructive properties of the estimators that lead to robustness.

5. CONSTRUCTIVE ANALYSIS OF ROBUSTNESS

The objective of this section is to investigate methods for breaking up the definitions in section 4 into constructive parts. For that, the double identity of the estimators as functions on Euclidean as well as distribution spaces is used. Also, only data vectors are considered and emphasis is given to the case in which S is a location parameter.

An approach that was used by Hampel [8] is explained in more detail here and is extended to the ζ-robust estimates as defined by definition 4.5. The following theorem by Strassen [9] is essential to that approach.

Theorem 5.1

Given two n-dimensional distributions Q_{1n}, Q_{2n}, then the inequality

$$d_L(Q_{1n}, Q_{2n}) \leq \zeta \tag{5.1}$$

is true if and only if there is a joint density $D_n(\cdot,\cdot)$ defined on the E^{2n} Euclidean space such that it has the distribution Q_{1n}, Q_{2n} as marginals and in addition it satisfies the inequality

$$D_n\{X_n, Y_n : d_{on}(X_n, Y_n) > \zeta\} \leq \zeta \tag{5.2}$$

for every X_n, Y_n vector outputs from the distributions Q_{1n}, Q_{2n} correspondingly. In (5.1) $d_L(\cdot,\cdot)$ is the Lèvy distance as defined by (3.2) and $d_{on}(\cdot,\cdot)$ in (5.2) is the distance in (3.3). This last distance may be defined by (3.4) or (3.5).

Let us now suppose that the parameter S is a parameter that includes marginal statistics only. Such statistics could be the moments of any order of a one-dimensional distribution F. Then, all the information about S is included in such an F and the method of contacting repeated trials from it does not influence the value of S.

Suppose that n independent experiments are done from the distribution F and let us call X_n the n-dimensional outcome. The distribution of X_n is then described, of course, by F^n. Let us now define

$$n_{X_n}(y) = \frac{\#\text{ of } X_n \text{ components that have values inside } (-\infty, y]}{n} \tag{5.3}$$

Then, for given X_n and varying y, $n_{X_n}(y)$ defines a step kind distribution characterized by probabilities that are multiples of 1/n. Let us call A_n the space of all possible $n_{X_n}(y)$ distributions when the vector X_n takes all the possible values. Also, call B_n any subspace of A_n and $F(B_n)$ the probability that, when n independent experiments are done from F, then the outcome X_n will correspond to an $n_{X_n}(y)$ that belongs to B_n.

Based on the distribution $n_{X_n}(y)$ and the definition of $F(B_n)$, the following lemma can be expressed.

Lemma 5.1

Let a sequence of estimators $\{\hat{S}_n\}$ and a one-dimensional distribution $F_o(S)$ be such that:

given $\epsilon > 0$, $\eta > 0$ and $0 < \zeta \leq 1$, then for every one-dimensional distributions $F_i(S)$ that satisfies $d_L(F_o(S), F_i(S)) < \zeta$ and for every n, there is a $B_{n_i} = B_n(F_i) \subset A_n$ such that

$$F_i(B_{n_i}) > 1 - \eta$$

For the same ϵ, ζ, let us also assume that there is also a $\delta(\epsilon, S) > 0$ such that for any $n_{Y_n}(y), n_{X_n}(y) \in B_{ni}$ and $d_L(n_{X_n}(y), n_{Y_n}(y)) < \delta(\epsilon, S)$, it is implied that

$$d_{on}(\hat{S}_n(X_n), \hat{S}_n(Y_n)) < \epsilon, \text{ where } d_{on}(\cdot,\cdot) \text{ is given by either (3.4) or (3.5)}$$

Then, the sequence $\{\hat{S}_n\}$ is ζ-(d_L, d_L) robust at $F_o(S)$.

Proof

Consider definition 4.5 and the $\epsilon > 0$ given in it. Also, call M the upper bound on the number of distributions that satisfy the conditions of definition 4.5. Since a finite number of distribution pairs $\{Q_i(S), Q_i'(S)\}$ is considered there, M exists.

Choose now $\epsilon_1 = \epsilon/M > 0$, $\eta = \epsilon_1/2$. Then, if the conditions in the lemma are satisfied, then given $\epsilon_1 > 0$, there is a $\delta(\epsilon_1,S) > 0$ such that if $n_{X_n}(y) \in B_{ni}$, the inequality $d_L(n_{X_n}(y), n_{Y_n}(y)) < \delta(\epsilon_1,S)$ implies

$$d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_n)) < \epsilon_1 \text{ , for every } n_{Y_n}(y) \tag{5.4}$$

where $d_{on}(\cdot,\cdot)$ is given by either (3.4) or (3.5).

Now define

$$\delta(\epsilon,S) = \min\{\delta(\epsilon_1,S), \tfrac{1}{4}\epsilon_1^2\} \tag{5.5}$$

Then, due to (5.4), we also have that

$$\left.\begin{array}{l} d_L(n_{X_n}(y),\ n_{Y_n}(y)) < \delta(\epsilon,S) \\ n_{X_n}(y) \in B_{ni} \end{array}\right\} \text{imply } d_{on}(\hat{S}_n(X_n),\hat{S}(Y_n)) < \epsilon_1 \tag{5.6}$$

Let us now suppose that $\{F_i(S)\}$ and $\{F_i'(S)\}$ are two arbitrary distribution sets satisfying

$$\sum_i d_L(F_i(S),F_i'(S)) < \delta(\epsilon,S) \tag{5.7}$$

From (5.7) we also have

$$d_L(F_i(S),F_i'(S)) < \delta(\epsilon,S) \tag{5.8}$$

and applying theorem 5.1, we conclude that if (5.8) is true, then there is a two-dimensional distribution Q_i with marginals $F_i(S), F_i'(S)$ and such that

$$Q_i\{x_i,y_i : |x_i-y_i| \geq \delta(\epsilon,S)\} < \delta(\epsilon,S)$$

where x_i, y_i are scalar outcomes with marginal distributions $F_i(S)$ and $F_i'(S)$ correspondingly.

If n independent experiments are contacted from the two-dimensional distribution Q_i, a variant of Chebyshev's inequality (for nonnegative random variables) gives

$$Q_i\{X_n,Y_n: \frac{\#\text{ of } i\text{'s}: |x_i-y_i| \geq \delta(\epsilon,S)}{n} \geq \delta^{\frac{1}{2}}(\epsilon,S)\} < \delta^{\frac{1}{2}}(\epsilon,S) \qquad (5.9)$$

where x_i, y_i the i^{th} component of the vector X_n, Y_n correspondingly.

$$X_n,Y_n: \frac{\#\text{ of } i\text{'s}: |x_i-y_i| > \delta(\epsilon,S)}{n}.$$

describes a joint step kind distribution with marginals $n_{X_n}(y), n_{Y_n}(y)$. Therefore, theorem 5.1 can be applied again on (5.9) which becomes:

$$Q_i\{n_{X_n}(y) \in B_{ni} \text{ and } d_L(n_{X_n}(y), n_{Y_n}(y)) < \delta^{\frac{1}{2}}(\epsilon,S)\} >$$

$$> 1 - \delta^{\frac{1}{2}}(\epsilon,S) - \eta = 1 - \delta^{\frac{1}{2}}(\epsilon,S) - \frac{\epsilon_1}{2} = 1 - \epsilon_1 \qquad (5.10)$$

From (5.10) and (5.6), we finally obtain

$$Q_i\{d_{on}(\hat{S}_n(X_n), \hat{S}_n(Y_n)) < \epsilon_1\} > 1 - \epsilon_1 \qquad (5.11)$$

Application of theorem (5.1) once more in combination with (5.11) leads to the following final conclusion

$$d_L(D_{F_i}(\hat{S}_n), D_{F_i'}(\hat{S}_n)) < \epsilon_1$$

and from it

$$\sum_{i=1}^{K} d_L(D_{F_i}(\hat{S}_n), D_{F_i'}(\hat{S}_n)) < K\epsilon_1 = \frac{K}{M}\epsilon < \epsilon$$

for every finite choice of pairs $\{F_i(S), F_i'(S)\}$. The proof is here complete.

The importance of lemma 5.1 is in the fact that it presents robustness as a property of the sampling distributions and the estimators when the latter are looked upon as functions on Euclidean spaces. In particular, ζ-robustness at $F_o(S)$ is guaranteed if the experimental distributions $n_{X_n}(y)$ and

$n_{Y_n}(y)$ being in Lèvy distance closer than a $\delta(\varepsilon,S) > 0$ implies that the corresponding estimator distance $d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_n))$ is smaller than an $\varepsilon > 0$.

The search for robustness is moved this way to the more tractable problem of studying experimental averages and real vector functions. The conditions in lemma 5.1 will be broken into two simpler parts. Before this is done, a lemma will be stated which is similar to 5.1 and concerns weakly robust estimates.

Lemma 5.2

Let a sequence $\{\hat{S}_n\}$ of estimates and a distribution $F_o(S)$ be such that:

given $\varepsilon > 0$ and $\eta > 0$, there is a $B_n \subset A_n$ and a $\delta(\varepsilon,S) > 0$ such that

i) $F_o(B_n) > 1 - \eta$

ii) For only Y_n, $X_n \in B_n$, and $d_L(n_{X_n}(y),n_{Y_n}(y)) < \delta(\varepsilon,S)$
it is implied that

$$d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_n)) < \varepsilon$$

where $d_{on}(\cdot,\cdot)$ is defined by (3.4) or (3.5). Then, $\{\hat{S}_n\}$ is weakly-(d_L,d_L) robust at $F_o(S)$.

The proof is as in lemma 5.1, only simpler. This lemma was first stated by Hampel [8] and it was used for his definition of robustness.

To proceed with our effort to establish simpler conditions that, if satisfied, guarantee robustness, we present the following definition which was offered by Hampel [8].

Definition 5.1

For given n, the estimator $\hat{S}_n$ is continuous at X_n as a real function from the E^n Euclidean space onto E^p (where X_n given data vector) iff:

given $\varepsilon > 0$, there is a $\delta(\varepsilon,S) > 0$ such that

For every Y_n with $d_{on}(Y_n,X_n) < \delta(\varepsilon,S)$, it is implied that $d_{on}(\hat{S}_n(X_n), \hat{S}_n(Y_n)) < \varepsilon$ where $d_{on}(\cdot,\cdot)$ is again defined by either (3.4) or (3.5).

The concept expressed in definition 5.1 is a simple extension to the E^n space of the continuity of real function on E^1. Similarly, one can express the following definition.

Definition 5.2

For given n, the estimator $\hat{S}_n$ is continuous as a function on E^n if and only if it is continuous at every data vector X_n.

A stronger property that is a simple extension of the concept of absolute continuity of a function defined on E^1 is given by the following definition.

Definition 5.3

Given n, an estimator $\hat{S}_n$ is absolutely continuous as a function on E^n iff:

given $\epsilon > 0$, there is a $\delta(\epsilon,S) > 0$ such that:

For every finite collection of vector pairs $\{Y_n^{(i)},X_n^{(i)}\}$ that satisfy

$$\sum_i d_{on}(Y_n^{(i)},X_n^{(i)}) < \delta(\epsilon,S)$$

it is implied that

$$\sum_i d_{on}(\hat{S}_n(Y_n^{(i)}),\hat{S}_n(X_n^{(i)})) < \epsilon$$

All definitions expressed in this section will be used to break the conditions of each of the lemmas 5.1 and 5.2 into two simpler parts that can be investigated more easily. The two lemmas that will follow now will use definitions 5.1 and 5.3 correspondingly to guarantee satisfactions of these conditions when the number of samples n is bounded.

Lemma 5.3

If a sequence of estimators $\{\hat{S}_n\}$ is such that $\hat{S}_n$ is continuous as a function on E^n for every n , then the conditions of lemma 5.2 are satisfied for every bounded $n \leq n_o$.

Proof

Let us fix n . Then, since $\hat{S}_n$ is continuous as a real function on E^n , given $\epsilon > 0$ and an outcome (n-dimensional data vector) X_n , there is a $\delta(\epsilon/2,X_n) > 0$ such that for every Y_n with $d(Y_n,X_n) < \delta(\epsilon/2,X_n)$ it is implied that

$$d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_n)) < \epsilon/2$$

Now let us choose a sequence $\{c_i\}$ with components positive and monotonically decreasing toward zero. Then define

$$A_i = \{X_n : \delta(\epsilon/2, X_n) > c_i\}$$

Obviously, the sets A_i are such that $A_i \subseteq A_{i+1} \to E^n$, $\cup A_i = E^n$. Define now

$$U_{\delta(\epsilon/2,X_n)}(X_n) = \{Y_n : d_{on}(X_n, Y_n) < \delta(\epsilon/2, X_n)\}$$

$$B_i = \cup\{U_{\frac{1}{2}\delta(\epsilon/2,X_n)}(X_n) : X_n \in A_i\}$$

Then

$$B_i \subseteq B_{i+1}, \cup B_i = E^n$$

Now, given $\eta > 0$, there is a $j(n,\eta)$ such that

$$F_o^n\{\bigcup_{1\leq i\leq j(n,\eta)} B_i\} > 1 - \eta$$

where F_o one-dimensional distribution and F_o^n the distribution of n independent experiments from F_o.

Denote

$$\mathcal{E}'_n = \bigcup_{1\leq i\leq j(n,\eta)} B_i$$

and choose

$$Y_n^{(1)} \in \mathcal{E}'$$

$$Y_n^{(2)} : d_n(F_o) = d_{on}(Y_n^{(1)}, Y_n^{(2)}) < \min[\frac{c_{j(n,\eta)}}{2}, 1/n] = \delta_n$$

Then, there is an X_n with $\delta(\epsilon/2, X_n) > 2\delta_n$ which is such that

$$Y_n^{(1)} \in U_{\frac{1}{2}\delta(\epsilon/2,X_n)}(X_n)$$

So

$$Y_n^{(2)} \in U_{\delta(\epsilon/2,X_n)}(X_n)$$

and

$$d_{on}(\hat{S}_n(Y_n^{(1)},\hat{S}_n(Y_n^{(2)})) \leq d_{on}(\hat{S}_n(Y_n^{(1)}),\hat{S}_n(X_n)) +$$

$$+ d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_n^{(2)})) < \epsilon/2 + \epsilon/2 = \epsilon$$

Call $\mathcal{E}''_n$ the space that evolves from all the possible permutations at the vectors of the space $\mathcal{E}'_n$. Then $B_n(F_o)$ is the corresponding $n_{X_n}(y)$ set. Let now $n_{Y_n^{(1)}}(y) \in B_n(F_o)$, $n_{Y_n^{(2)}}(y)$ any distribution and $d_L(n_{Y_n^{(1)}}(y), n_{Y_n^{(2)}}(y)) < \delta_n$. In this case, there are some $Y_n^{(1)},Y_n^{(2)}$ for which $Y_n^{(1)} \in \mathcal{E}''_n$ and $d_{on}(Y_n^{(1)},Y_n^{(2)}) < \delta_n$. Then, of course, it is implied that

$$\delta_{on}(\hat{S}_n(Y_n^{(1)}),\hat{S}_n(Y_n^{(2)})) < \epsilon$$

Define $\delta_1(\epsilon,S) = \min\{\delta_i,\ 1 \leq i \leq n_o\}$ and the space $B_n(F_o)$ above. Then $\delta_1(\epsilon)$ and $B_n(F_o)$ are the characteristics satisfying the conditions of lemma 5.2 for $n \leq n_o$.

Lemma 5.4

If a sequence of estimators $\{\hat{S}_n\}$ is such that $\hat{S}_n$ is absolutely continuous as a function on E^n for every n , then the conditions of lemma 5.1 are satisfied for every bounded $n \leq n_o$.

Proof

If $\hat{S}_n$ is absolutely condinuous as a function on E^n , then given $\epsilon > 0$ there is a unique $\delta(\epsilon/2,S) > 0$, the same for every X_n , such that:

For every Y_n such that $d_{on}(X_n, Y_n) < \delta(\varepsilon/2, S)$ it is implied that

$d_{on}(\hat{S}_n(X_n), \hat{S}_n(Y_n)) < \varepsilon/2$

In this case, the sequence of real numbers $\{c_i\}$ that was defined in the definition of the previous lemma can be bounded from below by $\delta(\varepsilon/2, S)$ that is the same for every distribution $F_i(S) \cdot j(n, \eta)$ corresponds to the largest c_j that is smaller than $\delta(\varepsilon/2)$, and

$$\delta_n(F_i) = \delta_n = \min\{\frac{c_{j(n,\eta)}}{2}, \frac{1}{n}\}$$

Otherwise, the proof works as in lemma 5.3.

In lemmas 5.3 and 5.4 we showed that continuity and absolute continuity of the estimator $\hat{S}_n$, as defined by definitions 5.2 and 5.3, are sufficient properties to guarantee satisfactions of weakly robust and ζ-robust conditions correspondingly or dimensionality n bounded.

Additional properties that guarantee such conditions for nonbounded n are needed. The following three definitions serve this purpose.

<u>Definition 5.4</u>

A sequence of estimators $\{\hat{S}_n\}$ is continuous at $F_o(S)$ iff:

Given $\varepsilon > 0$, there is a $\delta(\varepsilon, S) > 0$ and a positive integer $n_o > 0$ such that:

For every $n, m \geq n_o$ and data vectors X_n, Y_m with empirical distributions (defined by equation (5.3)) $n_{X_n}(y)$, $n_{Y_m}(y)$ satisfying

$$d_L(n_{X_n}(y), F_o) < \delta(\varepsilon, S) \text{ and } d_L(n_{Y_m}(y), F_o(S)) < \delta(\varepsilon, S)$$

it is implied that

$$d_{on}(\hat{S}_n(X_n),\hat{S}_n(Y_m)) < \epsilon$$

Definition 5.5

A sequence of estimators $\{\hat{S}_n\}$ is ζ-continuous at $F_o(S)$ iff: It is continuous (according to definition 5.4) at every $F_i(S)$ that satisfies

$$d_L(F_o(S),F_i(S)) \leq \zeta$$

It is interesting to observe that definitions 5.4 and 5.5 are concerned with the empirical distributions of the data and the real function identity of the estimators. A last definition expressing a stronger property that will be used in the constructive characterization of ζ-robustness is given by

Definition 5.6

A sequence of estimators $\{\hat{S}_n\}$ is ζ-absolutely continuous at $F_o(S)$ iff: Given $\epsilon > 0$ and $0 < \zeta \leq 1$, there are $\delta(\epsilon,S) > 0$ and positive integer n_o, such that for every finite set of distributions $\{F_i\}$ satisfying

$$d_L(F_o(S),F_i(S)) < \zeta$$

for every $m,n \geq n_o$ and all empirical distributions $n_{X_{ni}}(y),n_{Y_{mi}}(y)$ that satisfy

$$\sum_i d_L(F_i(S),n_{X_{ni}}(y)) < \delta(\epsilon)$$

$$\sum_i d_L(F_i(S),n_{Y_{mi}}(y)) < \delta(\epsilon)$$

it is implied that

$$\sum_i d_{on}(\hat{S}_n(X_{ni}),\hat{S}_m(Y_{mi})) < \epsilon$$

The importance of definitions 5.4-5.6 to the constructive approach to robustness becomes apparent from the following two lemmas.

Lemma 5.5

Let a sequence of estimators $\{\hat{S}_n\}$ be continuous at $F_o(S)$ (def. 5.4). Then, $\{\hat{S}_n\}$ is asymptotically weakly-(d_L,d_L) robust at $F_o(S)$ (def. 4.3).

Proof

Let the sequence $\{\hat{S}_n\}$ be continuous at $F_o(S)$. Then it can be easily shown [Hampel, Lemma 2] that $\hat{S}_n(X_n)$ converges to a single p-dimensional value $S(F_o(S))$ in probability. That means that

given $\epsilon > 0$, $\eta > 0$

there are $\delta_o(\epsilon,S) > 0$ and positive integer n_1 such that for $d_L(F_o(S), n_{X_n}(y)) < 2\delta_o(\epsilon)$ it is implied that

$$d_{on}(\hat{S}_n(X_n),S(F_o(S))) < \epsilon/2, \forall\, n \geq n_1$$

and that there is an integer $n_o \geq n_1$ such that

$$F_o\{d_L(F_o(S),n_{X_n}(y)) \geq \delta_o(\epsilon,S)\} < \eta \text{ , for } n \geq n_o$$

choose

$$B_n(F_o(S)) = \{n_{X_n}(y):d_L(F_o(S),n_{X_n}(y)) < \delta_o(\epsilon,S)\}$$

for all n's that are larger than n_o and pick $\delta_o(\epsilon,S)$. For this $B_n(F_o(S))$ and $\delta_o(\epsilon,S)$ the conditions of lemma 5.2 are satisfied for all $n > n_o$, therefore the sequence $\{\hat{S}_n\}$ is asymptotically weakly-(d_L,d_L) robust.

Lemma 5.6

Let a sequence of estimators $\{\hat{S}_n\}$ be ζ-absolutely continuous at $F_o(S)$

(def. 5.6). Then, $\{\hat{S}_n\}$ is asymptotically ζ-(d_L, d_L) robust at $F_o(S)$ (def. 4.6).

Proof

From the fact that $\{\hat{S}_n\}$ is ζ-absolutely continuous at $F_o(S)$, the existence of an n_o and a $\delta(\epsilon,S)$ that are common for every distribution $F_i(S)$ considered is guaranteed. These n_o and $\delta(\epsilon,S)$ are such that for all $n_{X_{ni}}(y)$ and $n_{Y_{mi}}(y)$ satisfying

$$d_L(F_i(S), n_{X_{ni}}(y)) < \delta(\epsilon,S)$$
$$d_L(F_i(S), n_{Y_{mi}}(y)) < \delta(\epsilon,S).$$

it is implied that

$$d_{on}(\hat{S}_n(X_{ni}), \hat{S}_m(Y_{mi})) < \epsilon$$

when $n,m > n_o$. If the B_{ni} of lemma 5.1 are then chosen in a way similar to the way B_n was chosen in the proof of lemma 5.5, the conditions of lemma 5.1 are satisfied for $n > n_o$ and the sequence $\{\hat{S}_n\}$ is absolutely ζ-(d_L, d_L) robust at $F_o(S)$.

The conclusion from lemmas 5.5 and 5.6 is that whenever asymptotic robust properties are sufficiently good for our purposes, sequences of estimators that are either continuous or ζ-absolutely continuous at a certain one-dimensional distribution $F_o(S)$ satisfy the requirements. However, if robustness for every sample size n is required, the combined properties expressed in definitions 5.3 and 5.6 or 5.2 and 5.4 are needed. Because of that and due to lemmas 5.3, 5.4, 5.5 and 5.6, the following two theorems are directly expressed.

Theorem 5.2

Let a sequence of estimators $\{\hat{S}_n\}$ be such that

i) $\hat{S}_n$ is continuous as a function on E^n, for every n (def. 5.2)

ii) The sequence $\{\hat{S}_n\}$ is continuous at $F_o(S)$ (def. 5.4)

then $\{\hat{S}_n\}$ is weakly-(d_L, d_L) robust at $F_o(S)$.

Theorem 5.3

Let a sequence of estimators $\{\hat{S}_n\}$ be such that

i) $\hat{S}_n$ is absolutely continuous as a function on E^n, for every n (def. 5.3)

ii) The sequence $\{\hat{S}_n\}$ is ζ-absolutely continuous at $F_o(S)$ (def. 5.6)

Then, $\{\hat{S}_n\}$ is ζ-(d_L, d_L) robust at $F_o(S)$.

The satisfaction of one of the lemmas 5.5, 5.6 or theorems 5.2,5.3 is, in general, required for every value of the parameter S. Then, the restrictions on the sequence $\{\hat{S}_n\}$ are stronger, of course.

Section 5 has been dedicated entirely to n-dimensional observation vectors that have independent components identically distributed according to a one-dimensional distribution function F.

REFERENCES

1. F.R. Hampel (1971) "A general qualitative definition of robustness," Ann. Math. Statis. 42: 1887-1896.

2. J.W. Tukey (1960) "A survey of sampling from contaminated distributions," in Contributions to Probability and Statistics, I. Olkin, ed., Stanford Univ. Press.

3. J.W. Tukey (1962) "The future of data analysis," Ann. Math. Statis. 33: 1-67.

4. J.L. Hodges and E.L. Lehman (1963) "Estimates of location based on rank tests," Ann. Math. Statis. 34: 598-611.

5. P.J. Huber (1964) "Robust estimation of a location parameter," Ann. Math. Statis. 36: 1753-1758.

6. P.J. Huber (1967) "The behavior of maximum likelihood estimates under nonstandard conditions," Proc. Fifth Berkeley Symp. Math. Statis. Prob. 1: 221-233.

7. P.J. Huber (1970) "Studentizing robust estimates" in Nonparametric Techniques in Statistical Inference, M.L. Puri, ed., Cambridge Univ. Press, pp. 453-463.

8. P.J. Huber (1972) "Robust statistics. A review," Ann.Math. Statis. 43: 1041-1067.

9. V. Strassen (1965) "The existence of probability measures with given marginals," Ann. Math. Statis. 36: 423-439.

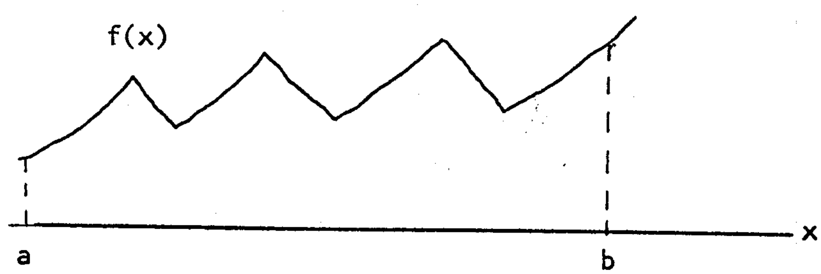

FIG. 1 CONTINUOUS f(x)

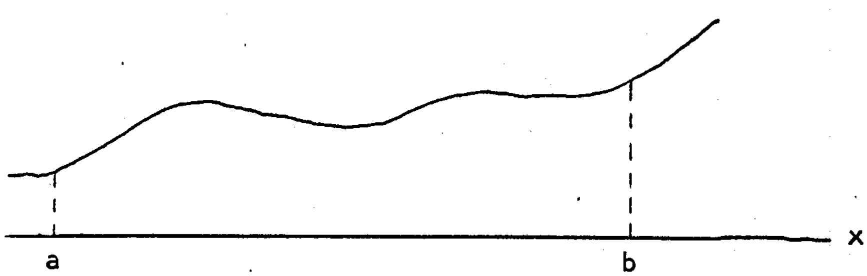

FIG. 2 DIFFERENTIABLE f(x)