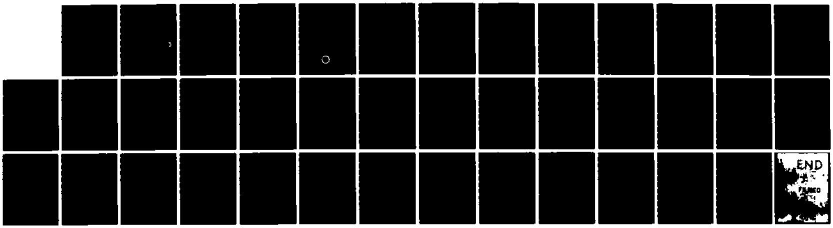
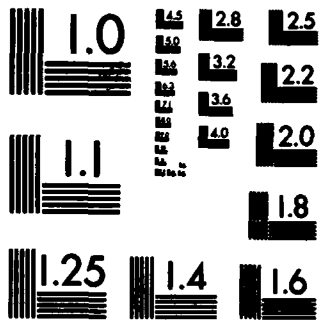
AD-A145 703 QUALITATIVE ROBUSTNESS FOR GENERAL STOCHASTIC PROCESSES 1/1
(U) WASHINGTON UNIV SEATTLE DEPT OF STATISTICS
G BOENTE ET AL. OCT 82 TR-26 N00014-82-K-0062

UNCLASSIFIED F/G 12/1 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 26
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:
4. TITLE (and Subtitle): Qualitative robustness for general stochastic processes
5. TYPE OF REPORT & PERIOD COVERED: Tech. Report 1/1/82-11/30/83
6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Graciela Boente, Ricardo Fraiman and Victor Yohai
8. CONTRACT OR GRANT NUMBER(s): ~~N00014-84-C-0169~~ N00014-82-K-0062
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, GN-22, University of Washington, Seattle, WA 98195
10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: NR 608-002
11. CONTROLLING OFFICE NAME AND ADDRESS: ONR Code N63374, 1107 NE 45th Street / Seattle, WA 98105
12. REPORT DATE: October 1982
13. NUMBER OF PAGES: 35
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
15. SECURITY CLASS. (of this report): Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:
16. DISTRIBUTION STATEMENT (of this Report): APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:
19. KEY WORDS (Continue on reverse side if necessary and identify by block number): Qualitative robustness
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this paper we generalize Hampel's definition of robustness and Π-robustness of a sequence of estimators to the case of non-i.i.d. stochastic processes, using appropriate metrics on the space of finite and infinite dimensional samples. We also present a different approach to qualitative robustness based on uniform insensitivity of the sequence of estimators when the sample is affected by round-off errors or by a small fraction of outliers. We give two definitions based on this approach: strong and weak pointwise robustness. We show that for estimating a finite dimensional real parameter,

20. **ABSTRACT** (continued)

Π-robustness is equivalent to weak pointwise robustness and at least in the i.i.d. case is also equivalent to strong pointwise robustness. Finally we show that the continuity condition given by Papantoni-Kazakos and Gray is sufficient for strong pointwise robustness. This implies the strong pointwise robustness of GM-estimates for autoregressive models.

QUALITATIVE ROBUSTNESS FOR GENERAL STOCHASTIC PROCESSES

BY

GRACIELA BOENTE
RICARDO FRAIMAN
AND
VICTOR YOHAI

TECHNICAL REPORT NO. 26

OCTOBER, 1982

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

Qualitative Robustness for General Stochastic Processes

by Graciela Boente, Ricardo Fraiman and Victor Yohai.[1]

Universidad de Buenos Aires and CONICET, Universidad de Buenos Aires, Universidad de Buenos Aires and CEMA.

Abreviated Title: Qualitative Robustness

SUMMARY

In this paper we generalize Hampel's definition of robustness and Π-robustness of a sequence of estimators to the case of non i.i.d. stochastic processes, using appropriate metrics on the space of finite and infinite dimensional samples. We also present a different approach to qualitative robustness based on uniform insensitivity of the sequence of estimators when the sample is affected by round-off errors or by a small fraction of outliers. We give two definitions based on this approach: strong and weak pointwise robustness. We show that for estimating a finite dimensional real parameter, Π-robustness is equivalent to weak pointwise robustness and at least in the the i.i.d. case is also equivalent to strong pointwise robustness. Finally we show that the continuity condition given by Papantoni-Kazakos and Gray is sufficient for strong pointwise robustness. This implies the strong pointwise robustness of GM-estimates for autoregressive models.

1. This research was supported in part by National Science Foundation Grant SES80-15570, and by the Office of Naval Research under Contract N00014-82-K-0062, while this author was visiting the Department of Statistics of the University of Washington, Seattle, Washington.

2. A.M.S. 1970 subject classifications. Primary 62F35; Secondary 62M10. Key words and phrases: qualitative robustness, robust estimation, GM-estimators, stochastic processes and autoregressive models.

Qualitative robustness for general stochastic processes

by Graciela Boente, Ricardo Fraiman and Victor Yohai.
Universidad de Buenos Aires and CONICET, Universidad de Buenos Aires,
Universidad de Buenos Aires and CEMA.

1. Introduction: Hampel (1971) introduced a definition of qualitative robustness of a sequence of estimators for the case of independent and indentically distributed, i.i.d., observations. This definition states that a sequence of estimators, T_n is robust at a given distribution μ on the sample space X if for any distribution ν close to μ in the Prohorov metric, the laws of T_n under μ and ν are close in the Prohorov metric uniformly for all n.

The use of the Prohorov distance reflects the intuitive meaning of robustness as insensitivity of the estimator to:

(a) small errors in all the observations (e.g. round-off errors)

(b) a small fraction of the observations with large errors (outliers).

Hampel also defines the more restrictive concept of Π-robustness which also requires insensitivity to non i.i.d. deviations of the model.

The generalization of these definitions to the case of stochastic processes with dependent observations requires defining appropiate distances between distributions on X^n in the case of Π-robustness and between distributions on X^∞ in the case of robustness. It turns out that there is not a unique natural way of doing this. In fact several definitions of qualitative robustness based on different probability measures on X^∞ have been given, see Papantoni-Kazakos and Gray (1979) and Cox (1978).

Cox's (1978) proposal is not completely general since it only makes sense for estimators which depend only on a finite marginal empirical distribution. This is not the case on the usual least squares estimate for the parameter of a moving average process of order 1.

A shortcoming of the metric proposed by Papantoni-Kazakos and Gray (1979) calledhere ρ_{2d}) which is mentioned by Cox (1978) is that this metric is not invariant with respect to equivalent metrics d on the sample space X. Moreover the concept of robustness based on this metric only reflects its intuitive meaning when d is bounded. In fact Cox (1978) shows that when d is the usual metric on R, the sample mean is robust with respect to ρ_{2d}.

In Section 2 we propose new metrics $\Pi_{d_{1n}}$ on X^n and ρ_{1d} on X^∞. We compare these metrics with those used by Hampel (1971) and Papantoni-Kazakos and Gray (1979).

In Section 3 we give general definitions of robustness and Π-robustness.

These definitions will be invariant with respect to the underlying metric d. We show that in the i.i.d. case the general definitions of robustness and Π-robustness using respectively the metrics ρ_{1d} and $\Pi_{d_{1n}}$, are equivalent to Hampel's definitions. We also show that for bounded metrics d, the general definitions based on ρ_{1d} and ρ_{2d} are equivalent. However, we will argue that for non i.i.d. processes the meaningful concept is Π--robustness and not robustness.

In Section 4 we propose a different approach to qualitative robustness which, in our opinion more fully capture the intuitive meaning of the concept. Instead of considering the insensitivity of the estimates with respect to small changes in the distribution of the process, we look at how insensitive they are when the sample is affected by errors of types (a) and (b) mentioned above. The advantage of this approach is that we may require as a condition for robustness that the estimator itself and not only its distribution be insensitive to thesekind of errors. We give two definitions of robustness following this approach: weak and strong pointwise robustness. We show that weak pointwise robustness, which may be considered a generalization of the B condition given by Hampel (1971), is equivalent to Π-robustness based on the proposed metrics $\Pi_{d_{1n}}$ if the parameter space is a subset of R^k. We also show that strong and weak pointwise robustness are equivalent at an i.i.d. model. We conjecture that this equivalence should hold even for more general stationary and ergodic processes. Finally we show that the continuity condition given in Papantoni--Kazakos and Gray (1979) is sufficient for strong pointwise robustness. This implies that the conditions given by Lemma 5 of Cox (1978) are sufficient for the strong pointwise robustness of the GM-estimators for autoregressive models.

These estimators are studied by Denby and Martin (1979) and Bustos (1981).

In section 5 we prove some auxiliary lemmas.

2. Distances between probabilities

Let X be the sample space, and d be a distance on X. We shall assume throughout all this paper that (X,d) is a complete and separable metric space (polish space). Let X^n and X^∞ be the cartesian product of n and a denumerable set of copies of X respectively. $\mathcal{F}$ will denote the Borel σ-field on X, and $\mathcal{F}^n$, $\mathcal{F}^\infty$ the corresponding product σ-fields on X^n and X^∞. For any measurable space $(\Omega,\mathcal{A})$, let $\mathcal{P}(\Omega)$ be the class of all probabilities on $\mathcal{A}$. If μ and ν are in $\mathcal{P}(\Omega)$, $\mathcal{P}(\mu,\nu)$ denotes the class of all the probabilities P on $(\Omega\times\Omega, \mathcal{A}\times\mathcal{A})$ with marginals μ and ν.

If (X,d) is a metric space, then the Prohorov distance Π_d between μ and ν, $\mu, \nu \in \mathcal{P}(X)$ is defined by:

$$\Pi_d(\mu,\nu) = \inf \{ \varepsilon: \mu(A) \leq \nu(V(A,\varepsilon,d)) + \varepsilon \quad \forall A \in \mathcal{F}\},$$

where $V(A,\varepsilon,d) = \{x \in X : d(x,A) \dot{<} \varepsilon\}$.

Strassen (1968), establishes that if (X,d) is a polish space, then Π_d is given by:

$$\Pi_d(\mu,\nu) = \inf \{\varepsilon: \exists\ P \in \mathcal{P}(\mu,\nu) \text{ satisfying } P(\{(x,x'):d(x,x') \geq \varepsilon\}) \leq \varepsilon$$

Given $x^n=(x_1,\ldots,x_n) \in X^n$, and $k \leq n$, the k-th empirical marginal distribution induced by x^n is denoted by $\mu_k[x^n]$, and is defined as the element of $\mathcal{P}(X^n)$ which assigns mass $1/(n-k+1)$ to each sample $(x_{j+1}, x_{j+2},\ldots,x_{j+k})$ $0 \leq j \leq n-k$.

Metrics on X^n and $\tilde{X}^n$. Given (X,d) we will consider the following metrics on X^n

$$(2.1) \qquad d_{1n}(x^n,y^n) = \inf \{\varepsilon: \# \{i: d(x_i,y_i) > \varepsilon\}/n \leq \varepsilon\}.$$

$$(2.2) \qquad d_{2n}(x^n,y^n) = \sum_{i=1}^{n} d(x_i,y_i)/n.$$

Let now $\tilde{X}^n$ be the space X^n modulo the permutation of coordinates. Hampel (1971) defines the following distance on $\tilde{X}^n$ which we denote by d_{3n}

$$(2.3) \qquad d_{3n}(x^n,y^n) = \Pi_d(\mu_1[x^n]\ ,\ \mu_1[y^n]).$$

Remark. Two points of X^n are close in the metric d_{1n} if all the coordinates, except a small fraction are close. Therefore this notion of closeness corresponds to the type of errors which are considered in the intuitive notion of robustness. We will show the relation with d_{3n} in Lemma 2.3.

Metrics on $\mathcal{P}(X^n)$. Given (X,d) and $\mu_n,\nu_n \in \mathcal{P}(X^n)$ we consider the Prohorov metric associated to d_{1n}, $\Pi_{d_{1n}}(\mu_n,\nu_n)$, and the Vasershtein distance (Vasershtein (1969)) defined by

$$(2.4) \qquad \rho_{d_{2n}}(\mu_n,\nu_n) = \inf_{R \in \mathcal{P}(\mu_n,\nu_n)} E_R(d_{2n}(x^n,y^n)),$$

where E_R denotes the expectation under R. We also define the pseudo-metric $\Pi_{d_{3n}}(\mu_n,\nu_n)$ given by $\Pi_{d_{3n}}(\mu_n,\nu_n) = \Pi_{d_{3n}}(\tilde{\mu}_n,\tilde{\nu}_n)$, where $\tilde{\mu}_n$ and $\tilde{\nu}_n$ are the probabilities induced by μ_n and ν_n on $\tilde{X}^n$ respectively.

<u>Metrics on $P(X^\infty)$</u>. Given (X,d) and $\mu, \nu \in P(X^\infty)$ we shall denote by

$$(2.5) \qquad \rho_{1d}(\mu,\nu) = \sup_n \Pi_{d_{1n}}(\mu_n,\nu_n)$$

and

$$(2.6) \qquad \rho_{2d}(\mu,\nu) = \sup_n \rho_{d_{2n}}(\mu_n,\nu_n),$$

where μ_n and ν_n are the n-dimensional marginal probabilities of μ and ν respectively. The metric ρ_{2d} was introduced by Gray, Neuhoff and Shields (1975) and used by Papantoni-Kazakos and Gray (1979) to give a general definition of robustness. A shortcoming of ρ_{2d}, is that it is not invariant with respect to equivalent metrics, i.e., equivalent metrics d and d^* on X may not induce equivalent distances ρ_{2d} and ρ_{2d^*} on $P(X^\infty)$. For example, if $X = R$ the two equivalent metrics $d(x,y) = |x-y|$ and $d^*(x,y) = |x-y|/(1+|x-y|)$ induce non equivalent metrics ρ_{2d} and ρ_{2d^*}.

Lemma 2.1 shows that ρ_{1d} is invariant with respect to d, Lemma 2.2 shows that if d is bounded, ρ_{1d} is equivalent to ρ_{2d}, and Lemma 2.3 and its Corollary give the relationships between d_{1n} and d_{3n} and between $\Pi_{d_{1n}}$ and $\Pi_{d_{3n}}$.

<u>Lemma 2.1</u>. <u>Suppose that</u> d <u>and</u> d^* <u>are metrics on</u> X, <u>and let</u> $\delta > 0$

and $\delta^*>0$ *be real numbers such that* $d(x,y)<\delta$ *implies* $d^*(x,y)<\delta^*$. *Then for any* μ_n,ν_n *in* $P(X^n)$ *we have*

$$\Pi_{d_{1n}}(\mu_n,\nu_n)<\delta \Rightarrow \Pi_{d^*_{1n}}(\mu_n,\nu_n)<\delta^* .$$

Then if d *and* d^* *are equivalent* $\Pi_{d_{1n}}$ *and* $\Pi_{d^*_{1n}}$ *are equivalent too*.

Proof. Follows immediately from the definition of Prohorov distance

Lemma 2.2. Given μ_n *and* ν_n *in* $P(X^n)$, *we have*

(2.7) $$\rho_{d_{2n}}(\mu_n,\nu_n)<\delta \Rightarrow \Pi_{d_{1n}}(\mu_n,\nu_n)<\delta^{\frac{1}{2}} ,$$

and

(2.8) $$d<M \text{ and } \Pi_{d_{1n}}(\mu_n,\nu_n)<\delta \Rightarrow \rho_{d_{2n}}(\mu_n,\nu_n)<\delta\,(1+2M).$$

Proof. It is immediate that,

(2.9) $$d_{2n}(x^n,y^n)<\delta \Rightarrow d_{1n}(x^n,y^n)<\delta$$

and

(2.10) $$d<M \text{ and } d_{1n}(x^n,y^n)<\delta \Rightarrow d_{2n}(x^n,y^n)<\delta\,(1+M).$$

Let now μ_n and ν_n in $\mathcal{P}(X^n)$ be such that $\rho_{d_{2n}}(\mu_n,\nu_n) < \delta$. Then, there exists $R \in \mathcal{P}(\mu_n,\nu_n)$ such that $E_R(d_{2n}(x^n,y^n)) < \delta$. Therefore by the Markov inequality, we have

$$R(d_{2n}(x^n,y^n) < \delta^{\frac{1}{2}}) > 1-E_R(d_{2n}(x^n,y^n))/\delta^{\frac{1}{2}} > 1-\delta^{\frac{1}{2}} ,$$

then, by (2.9) $R(d_{1n}(x^n,y^n) < \delta^{\frac{1}{2}}) > 1-\delta^{\frac{1}{2}}$ and therefore $\Pi_{d_{1n}}(\mu_n,\nu_n) < \delta^{\frac{1}{2}}$.

Suppose now that $\Pi_{d_{1n}}(\mu_n,\nu_n) < \delta$, then there exists R in $\mathcal{P}(\mu_n,\nu_n)$ such that $R(d_{1n}(x^n,y^n) < \delta) > 1-\delta$ and by (2.10) we have that $R(d_{2n}(x^n,y^n) < \delta(1+M)) > 1-\delta$. Finally we have that

$$E_R(d_{2n}(X^n,Y^n)) < \delta(1+M) + MR(d_{2n}(x^n,y^n) > \delta(1+M)) < \delta(1+2M),$$

then (2.8) holds.

Lemma 2.3. Let $\mathcal{P}_n$ be the set of all permutations of the first n positive integers. Given x^n and y^n in X^n, if p is in $\mathcal{P}_n$, we denote by $y^n_p = (y_{p(1)},\dots, y_{p(n)})$. Then, we have

$$\text{(2.11)} \qquad \Pi_d(\mu_1[x^n],\mu_1[y^n]) = \min_{p\in\mathcal{P}_n} d_{1n}(x^n,y^n_p) .$$

Proof. It is enough to show that for any $\delta > 0$

$$\text{(2.12)} \qquad d_{1n}(x^n, y^n) < \delta \Rightarrow \Pi_d(\mu_1[x^n],\mu_1[y^n]) < \delta$$

and

(2.13) $\Pi_d(\mu_1[x^n],\mu_1[y^n]) < \delta \Rightarrow \exists\, p \in P_n : d_{1n}(x^n,y^n_p) < \delta$.

Suppose that $d_{1n}(x^n,y^n) < \delta$, then if $S = \{i : d(x_i,y_i) < \delta\}$ we have that $\# S/n > 1-\delta$. Let R be the distribution on $X\times X$ which assigns probalility $1/n$ to each pair (x_i,y_i) $1 \le i \le n$. Then $R \in P(\mu_1[x^n], \mu_1[y^n])$. We also have that $R(d(x,y)<\delta) = \# S/n > 1-\delta$. Then $\Pi_d(\mu_1[x^n],\mu_1[y^n]) < \delta$, and (2.12) holds.

In proving (2.13), we will find a set such that the Prohorov distance is attained.

Assume $\Pi_d(\mu_1[x^n],\mu_1[y^n]) < \delta$. Put $A = \{x_1,\ldots,x_n\}$, $B = \{y_1,\ldots,y_n\}$ and $D_n = \{1,2,\ldots,n\}$. Given p in P_n , let $h(p) = \# \{i : d(x_i,y_{p(i)})<\delta\}$ and t defined by

(2.14) $$t = \max_{p \in P_n} h(p) = h(p^*).$$

We have to prove that $t > n(1-\delta)$.

Without loss of generality, reordering the elements of B if necessary, we may assume that p^* is the identity.

Define $S_1=\{i: d(x_i,y_i)< \delta\}$, $I_1 = \{i: \exists\, j \;\; d(x_i,y_j)<\delta\}$ and $J_1 = \{j : \exists\, i \;\; d(x_i,y_j)<\delta\}$. Clearly we have $I_1 \supset S_1$ and $J_1 \supset S_1$. Let $I_2 = I_1 - S_1$ and $J_2 = I_2 - S_1$. We have

(2.15) $$\# S_1 = t \;, \quad I_1 = S_1 + I_2 \;, \quad J_1 = S_1 + J_2 \;.$$

Given $I \subset D_n$, define the sets $A(I) = \{x_i, i \in I\}$ and $B(I) = \{y_j, j \in I\}$ and the function $f(I) = \{j : y_j \in V(A(I),\delta,d)\}$

Since p^* maximizes $h(p)$, $V(A(I_2),\delta,d) \cap B \subset B(S_1)$, therefore, $f(I_2) \subset S_1$. We denote by $f^{(\ell)}$ the application defined by: $f^{(o)}$ is the identity and $f^{(\ell)} = f^{(\ell-1)} \circ f \quad \ell \geq 1$. We will show that there exists k_o such that

$$(2.16) \qquad f^{(k_o+1)}(I_2) = f^{(k_o)}(I_2) \subset S_1 .$$

Since $I \cap S_1 \subset f(I)$, in order to prove (2.16) it is enough to show that

$$(2.17) \qquad f^{(k)}(I_2) \subset S_1 \quad \forall k.$$

We will prove (2.17) by induction. We already know that (2.17) holds for $k = 1$. Suppose that it holds for all $k < m$, we will show that it also holds for $k = m+1$.

Clearly, $f^{(m+1)}(I_2) \subset J_1$. Since by the inductive hipothesis $f^{(m)}(I_2) \subset S_1$, we have that $f^{(m)}(I_2) \subset f^{(m+1)}(I_2)$. Put $R = f^{(m+1)}(I_2) - f^{(m)}(I_2)$. Then we have $f^{(m+1)}(I_2) = f^{(m)}(I_2) + R$.

Suppose that $R \not\subset S_1$, we will show that there exists a permutation p such that $h(p) = t+1$. Since $R \subset J_1$, (2.15) implies that there exists $r \in J_2 \cap R$.

We will find a finite sequence of numbers $q_o, q_1, \ldots, q_n$ in D_n such that (i) $q_o = r$, (ii) if $m_i = \min \{k : q_i \in f^{(k)}(I_2)\}$ then $m+1 = m_o > m_1 > \ldots > m_n = 0$ and therefore $i \neq j$ implies $q_i \neq q_j$, (iii) $d(x_{q_{i+1}}, y_{q_i}) < \delta$ $0 \leq i \leq n-1$.

In order to find this sequence we proceed as follows: Since $q_o = r \in f^{(m+1)}(I_2)$, there exists $q_1 \in f^{(m)}(I_2) \subset S_1$ such that $d(y_{q_o}, x_{q_1}) < \delta$. Then $m_1 \leq m$. If $m_1 = 0$, $n = 1$ and we have already completed the sequence. If $m_1 > 0$, as $q_1 \in f^{(m_1)}(I_2)$, there exists $q_2 \in f^{(m_1-1)}(I_2)$ such that $d(x_{q_2}, y_{q_1}) < \delta$ and $m_2 \leq m_1 - 1 < m_1$. If $m_2 = 0$, then $n = 2$ and we stop, otherwise we continue in the same way.

Let p the permutation defined by:

$$p(q_j) = q_{j-1} \quad 1 \leq j \leq n, \quad p(q_o) = q_n \in I_2 ,$$

$$p(i) = i \quad \text{if} \quad i \notin \{q_o, \ldots, q_n\} .$$

Clearly $h(p) = t+1$ and this contradicts the definition of t. Then there exists k_o satisfying (2.16). Let $C = A(f^{(k_o)}(I_2) \cup I_2 \cup (D_n - I_1))$. Therefore as (2.16) holds and $V(A(D_n - I_1), \delta, d) \cap B = \phi$, we have $B \cap V(C, \delta, d) = V(A(f^{(k_o)}(I_2) \cup I_2), \delta, d) \cap B$. Then by (2.16) we have $\#(B \cap V(C, \delta, d)) = \#(f^{(k_o)}(I_2))$.

Since $\Pi_d(\mu_1[x^n], \mu_1[y^n]) < \delta$ we also have

$$\mu_1[x^n](C) \leq \mu_1[y^n](V(C, \delta, d)) + \delta$$

or equivalently

(2.18) $\# C/n < \#(f^{(k_o)}(I_2))/n + \delta$.

Since $\# C = \#(f^{(k_o)}(I_2)) + \# I_2 + n - \# I_1 = n - \# S_1 + \#(f^{(k_o)}(I_2))$, (2.18) becomes $\# S_1 > n(1-\delta)$, and then by (2.15), we have that (2.13) holds.

Corollary 2.1. Let μ_n and ν_n in $P(X^n)$. Then,

(2.19) $\Pi_{d_{1n}}(\mu_n,\nu_n) < \delta \Rightarrow \Pi_{d_{3n}}(\mu_n,\nu_n) < \delta$

and

(2.20) $\Pi_{d_{3n}}(\mu_n,\nu_n) < \delta \Rightarrow \Pi_{d_{1n}}(\mu_n,\nu_n^*) < \delta$,

where ν_n^* is the probability induced by the transformation $T(y^n) = y^n_{p^*}$ and p^* is defined by $d_{3n}(x^n,y^n) = d_{1n}(x^n,y^n_{p^*})$.

3. Generalization of Hampel's definitions of robustness. Let $T_n: X^n \to \Lambda$ for $n \geq n_o$, be a sequence of estimators taking values in a polish space (Λ, λ). Given $\mu \in P(X^\infty)$, we denote by $\mathcal{L}(T_n,\mu)$ the distribution of T_n under μ . Cox (1978) gives the following generalization of Hampel's definition of qualitative robustness.

Definition: Let a pseudometric ρ on $P(X^\infty)$, a subset $Z \subset P(X^\infty)$ and

$\mu \in P(X^{\infty})$ be given, then the sequence of estimators $(T_n)_{n>n_0}$ is <u>ρ-robust at μ for Z</u> if given $\varepsilon > 0$, there exists $\delta > 0$, such that for all $\nu \in Z$ we have

$$\rho(\mu,\nu) < \delta \Rightarrow \Pi_\lambda(\mathcal{L}(T_n,\mu), \mathcal{L}(T_n,\nu)) < \varepsilon \quad \forall n > n_0 .$$

This definition specializes to that of Hampel by taking $\mu \in Z =$ = {all i.i.d. processes} and ρ the pseudometric on $P(X^{\infty})$ given by

$$(3.1) \qquad \rho_{Hd}(\mu,\nu) = \Pi_d(\mu_1,\nu_1) \quad ,$$

where μ_1 and ν_1 are the first order marginals of μ and ν respectively.

Papantoni-Kazakos and Gray (1979) definition of robustness corresponds to $\rho = \rho_{2d}$.

The following definition given by Bustos (1980) generalizes Hampel's concept of Π-robustness.

<u>Definition</u>. Let $\mu \in P(X^{\infty})$ and let ρ_n be a pseudometric on $P(X^n)$ for all $n > n_0$, then the sequence $(T_n)_{n>n_0}$ is <u>ρ_n-Π-robust at μ</u> if given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\nu_n \in P(X^n) \wedge n > n_0 \wedge \rho_n(\mu_n,\nu_n) < \delta \Rightarrow \Pi_\lambda(\mathcal{L}(T_n,\mu), \mathcal{L}(T_n,\nu_n)) < \varepsilon .$$

where μ_n is the n-th order marginal of μ

Notice that $\mathcal{L}(T_n,\mu) = \mathcal{L}(T_n,\mu_n)$.

In the case of μ an i.i.d. process and $(T_n)_{n>n_o}$ invariant by permutation of the coordinates, we get Hampel's definition taking as ρ_n the pseudometric $\Pi_{d_{3n}}$.

The following theorem show that ρ_{1d}-robustness, and $\Pi_{d_{1n}}$-Π-robustness are natural generalizations of Hampel's definitions.

<u>Theorem 3.1</u>. a) <u>Let</u> $Z = \{$ i.i.d. processes$\}$ <u>and</u> $\mu \in Z$, <u>then Hampel's definition of robustness</u> (ρ_{Hd}-<u>robustness</u>) <u>is equivalent to</u> ρ_{1d}-<u>robustness</u>.

(b) <u>Let</u> μ <u>be an i.i.d. process and</u> $(T_n)_{n>1}$ <u>a sequence of estimates invariant by permutations of the coordinates, then Hampel's definition of</u> Π-<u>robustness</u> ($\Pi_{d_{3n}}$-Π-<u>robustness</u>) <u>is equivalent to</u> $\Pi_{d_{1n}}$-Π-<u>robustness</u>.

<u>Proof</u>. Let μ and ν be two i.i.d. processes and μ_1,ν_1 their first order marginals. Then, in order to prove (a), it is enough to show:

$$(3.2) \qquad \Pi_d(\mu_1,\nu_1) \leq \rho_{1d}(\mu,\nu) \leq [\Pi_d(\mu_1,\nu_1)]^{\frac{1}{2}} .$$

Suppose that $\rho_{1d}(\mu,\nu)=\delta$ then $\Pi_{d_{11}}(\mu_1,\nu_1)\le\delta$, so there exists $R\in P(\mu_1,\nu_1)$ such that $R(d_{11}(x,y)<\delta)\geq 1-\delta$. But $d_{11}(x,y)=$ $=\min(1,d(x,y))$; then $\{d_{11}(x,y)<\delta\}=\{d(x,y)<\delta\}$ and therefore $R(d(x,y)<\delta)\geq 1-\delta$; so $\Pi_d(\mu_1,\nu_1)\le\delta$.

In order to prove that $\rho_{1d}(\mu,\nu)\le[\Pi_d(\mu_1,\nu_1)]^{\frac{1}{2}}$ it is enough to show:

(3.3) $\qquad \Pi_{d_{1n}}(\mu_1^n,\nu_1^n)\le[\Pi_d(\mu_1,\nu_1)]^{\frac{1}{2}} \quad \forall\, n.$

Suppose that $\Pi_d(\mu_1,\nu_1)=\delta<1$. Therefore there exists $R\in P(\mu_1,\nu_1)$ such that $R(d(x,y)<\delta)\geq 1-\delta$. Let R^n be the product measure of n copies of R. The Markov inequality yields

$$R^n(d_{1n}(x^n,y^n)<\delta^{\frac{1}{2}}) = R^n\Big(\frac{1}{n}\sum_{i=1}^{n} I_{[\delta^{\frac{1}{2}},+\infty)}(d(x_i,y_i))<\delta^{\frac{1}{2}}\Big)$$

$$\geq 1-E(I_{[\delta,+\infty)}(d(x,y)))/\delta^{\frac{1}{2}} = 1-R(d(x,y)\geq\delta)/\delta^{\frac{1}{2}} \geq 1-\delta^{\frac{1}{2}},$$

therefore (3.3) holds and part (a) is proved.

Part (b) follows from Lemma 2.3 and Corollary 2.1.

The following theorem, which is an immediate consequence of Lemma 2.2 establishes the relationship between ρ_{1d}-robustness ($\Pi_{d_{1n}}$-Π-robustness) and ρ_{2d}-robustness ($\rho_{d_{2n}}$-Π-robustness).

Theorem 3.2. (a) For any $\mu \in P(X^\infty)$, and $Z \subset P(X^\infty)$ we have

(i) ρ_{d_1}-robustness at μ for Z implies ρ_{d_2}-robustness at μ for Z.

(ii) If d is bounded, ρ_{d_2}-robustness at μ for Z implies ρ_{d_1}-robustness at μ for Z.

(b) For any $\mu \in P(X^\infty)$ we have

(i) $\Pi_{d_{1n}}$-Π-robustness at μ implies $\rho_{d_{2n}}$-Π-robustness at μ.

(ii) If d is bounded, then $\rho_{d_{2n}}$-Π-robustness at μ implies $\Pi_{d_{1n}}$-Π-robustness at μ.

Therefore if d is bounded both concepts are equivalents. If d is not bounded, it is not true that ρ_{2d}-robustness ($\rho_{d_{2n}}$-Π-robustness) implies ρ_{1d}-robustness ($\Pi_{d_{1n}}$-Π-robustness). Cox (1978) shows that if $R=X$ and $d(x,y) = |x-y|$, then the sample mean $\bar{X}_n = \sum_{i=1} X_i/n$ is ρ_{2d}-robust. However, $\bar{X}_n$ is not ρ_{1d}-robust.

The following theorem, which is a generalization of theorem 3 of Hampel (1971) is immediate.

Theorem 3.3. Suppose ρ_n is a metric on $P(X^n)$ and ρ is the metric on $P(X^\infty)$ defined by $\rho(\mu,\nu) = \sup_{n\geq 1} \rho_n(\mu_n,\nu_n)$, where μ and ν are in $P(X^\infty)$ and μ_n,ν_n are the corresponding n-th order marginals. Then given $\mu \in P(X^\infty)$ and $Z \subset P(X^\infty)$ we have that if $(T_n)_{n\geq n_0}$ is ρ_n-Π-robust at μ, then it is ρ-robust at μ for Z.

Therefore $\Pi_{d_{1n}}$-Π-robustness implies ρ_{1d}-robustness and $\rho_{d_{2n}}$-Π-robustness implies ρ_{2d}-robustness.

However, we consider that the relevant concept for non i.i.d. processes is Π-robustness and not robustness. The reason for this, is that continuous estimators which always depend on a fixed finite set of coordinates may result robust, and this contradicts the intuitive notion of robustness: a small proportion of observations should not affect the estimator too much. Consider the following example: let $T_n : X^n \to X$ be defined by $T_n(x_1,\ldots, x_n) = x_1$. This estimator is ρ_{1d} and ρ_{2d}-robust at any $\mu \in \mathcal{P}(X^\infty)$, because if $\rho_{1d}(\mu,\nu) \leq \varepsilon$, or $\rho_{2d}(\mu,\nu) \leq \varepsilon$, then $\Pi_{d_{11}}(\mu_1,\nu_1) < \varepsilon$ or $\rho_{d_{21}}(\mu_1,\nu_1) < \varepsilon$ respectively, and either of this inequalities implies $\Pi_d(\mu_1,\nu_1) < \varepsilon$ or equivalently $\Pi_d(\mathcal{L}(T_n,\mu),\mathcal{L}(T_n,\nu)) < \varepsilon$. However it is clear that if n is large, changing a small proportion of observations, just the first, the estimate will suffer a large variation.

In next section we will give more evidence that the meaningful concept is Π-robustness.

4. Pointwise robustness. Here, we propose a different approach to qualitative robustness which seems to capture better its intuitive meaning. Instead of considering the insensitivity of the estimates with respect to small changes in the distribution of the process, we look at how insensitive they are at a given infinite sample point $x \in X^\infty$, when:

(a) all the observations have small changes and (b) a small fraction of observations have large changes.

Consider $x^n \in X^n$ and let $V(x^n,\delta,d_{1n})$ be the open sphere of center x^n and radius δ corresponding to the metric d_{1n}. Define

$$S_{n\delta}(x^n) = \sup \{ \lambda(T_n(y^n), T_n(z^n)) : y^n, z^n \in V(x^n, \delta, d_{1n}) \} .$$

We show in Lemma 4.1 (i) that $S_{n\delta}$ is lower semicontinuous and the refore measurable.

Definition. Let $x \in X^\infty$ and x^n its projection in X^n, then $(T_n)_{n > n_o}$ is robust at x if given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$S_{n\delta}(x^n) < \varepsilon \quad \forall\, n > n_o .$$

We will now give two definitions of pointwise-robustness at a given probability $\mu \in P(X^\infty)$.

Definition: Let $\mu \in P(X^\infty)$, then $(T_n)_{n > n_o}$ is strongly pointwise robust at μ if

$$\mu(\{ x \in X^\infty : T_n \text{ is robust at } x\}) = 1.$$

Definition: Let $\mu \in P(X^\infty)$, then $(T_n)_{n > n_o}$ is weakly pointwise robust at μ if given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\mu(\{ x^n \in X^n : S_{n\delta}(x^n) < \varepsilon\}) > 1 - \varepsilon \quad \forall\, n > n_o .$$

We consider that these definitions reflect better the intuitive meaning of the word robustness, since they require that the estimator itself be insensitive to errors of type (a) and (b) mentioned above, while the definitions of ρ-robustness and ρ_n-Π-robustness only require insensitivity of the law of the estimators. However we will show later the equivalence of weak-pointwise robustness with $\Pi_{d_{1n}}$-Π-robustness. We also show that at least in the i.i.d. case strong pointwise robustness and weak pointwise robustness are equivalent.

The following proposition gives a necessary and sufficient condition for strong pointwise robustness.

Proposition 4.1. $(T_n)_{n>n_o}$ is strongly pointwise robust at μ if given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$(4.1) \qquad \mu\Big(\bigcap_{n>n_o} \{ x \in X^\infty : S_{n\delta}(x^n) < \varepsilon \}\Big) > 1 - \varepsilon.$$

Proof. Let $A = \{x \in X^\infty$ such that T_n is not robust at $x\}$. Then it is clear that

$$A = \bigcup_{\varepsilon} \bigcap_{\delta} \bigcup_{n>n_o} \{x : S_{n\delta}(x^n) > \varepsilon\} = \bigcup_{i\geq 1} \bigcap_{j\geq 1} \bigcup_{n>n_o} \{x : S_n(x^n) > 1/i\},$$

then

$$P(A) = 0 \iff P\Big(\bigcap_{j\geq 1} \bigcup_{n>n_o} \{x : S_{n1/j}(x^n) > 1/i\}\Big) = 0 \quad \forall\, i$$

$$\iff \lim_{j\to\infty} P\Big(\bigcup_{n>n_o} \{x : S_{n1/j}(x^n) > 1/i\}\Big) = 0 \quad \forall\, i$$

$$\Longleftrightarrow \forall \varepsilon>0 \quad \lim_{\delta\to 0} \; P\Big(\bigcup_{n\geq n_o} S_{n\delta}(x^n) > \varepsilon\Big) = 0 \; .$$

Corollary. Let $\mu \in P(X^\infty)$, then strong pointwise robustness of $(T_n)_{n\geq n_o}$ at μ implies weak pointwise robustness at μ.

Theorem 4.1. Let $\mu \in P(X^\infty)$ correspond to an i.i.d. process. Suppose that there exists a sequence of compact sets K_n, $1 \leq n < \infty$, such that $X = \bigcup_{n=1}^{\infty} K_n$. Let $(T_n)_{n\geq n_o}$ be invariant by permutations of coordinates and weakly pointwise robust at μ, then $(T_n)_{n\geq n_o}$ is also strongly robust at μ.

Proof. Let $\varepsilon>0$. We are going to show that there exists $\delta>0$ such that (4.1) holds. Since $(T_n)_{n\geq n_o}$ is weakly robust we can find δ^* such that

(4.2) $$\mu(S_{n\delta^*}(x^n) \leq \varepsilon) \geq 1 - \varepsilon \; .$$

It is clear that

(4.3) $$\{S_{n\delta^*/2}(x^n) \leq \varepsilon\} \supset V(\{S_{n\delta^*}(x^n) \leq \varepsilon\} \; , \; \delta^*/2, \; d_{1n}) \; .$$

By hypothesis there exists a compact K such that

(4.4) $$\mu_1(K) \geq 1 - \delta/8 \; .$$

We can find a partition $K_1,\dots,K_h$ of K such that each K_i $1\le i\le h$ has diameter $< \delta^*/2$. Put $K_o = (\bigcup_{i=1}^{h} K_i)'$ the complement of $\bigcup_{i=1}^{h} K_i$ and $m_i = \mu_1(K_i)$ $0 \le i \le h$.

Given $x^n = (x_1,\dots,x_n)$ define $S_{in}(x^n) = \sum_{j=1}^{n} I_{K_i}(x_j)/n$, $i = 0,1,\dots, h$. By a well known form of the strong law of the large numbers for Bernouilli variables, there exist $a>0$ and $0<b<1$ such that if $R_n = \bigcap_{i=0}^{h} \{x^n : |S_{in}(x^n)-m_i| < \delta^*/(8h)\}$, then

(4.5) $\mu(R_n) \ge 1 - ab^n \quad \forall n.$

From (4.2) and (4.5) we can find $n_1 > n_o$ and x^{*n} in $R_n \cap \{S_{n\delta^*}(x^n) < \varepsilon\}$ $\forall n > n_1$. Let P_n be the set of all the points obtai ned by permutation of the coordinates of x^{*n}.

Since $\{S_{n\delta^*}(x^n) < \varepsilon\}$ are invariant by permutation of coordinates, we have

(4.6) $P_n \subset \{S_{n\delta^*}(x^n) < \varepsilon\}$.

We will show now that

(4.7) $R_n \subset V(P_n, \delta^*/2, d_{1n})$.

Let $y^n \in R_n$, then $|S_{in}(x^{*n}) - S_{in}(y^n)| < \delta^*/(4h)$. Define $Q_i = \{j : y_j \in K_i\}$,

$0 \leq i \leq h$. Then there exist sets Q_i^*, $0 \leq i \leq h$, such that:

(4.8) $\quad \# Q_i^* = n\, S_{in}(x^{*n}) \qquad 0 \leq i \leq h$

and

(4.9) $\quad \#(Q_i^* - Q_i) < n\, \delta^*/(4h).$

By (4.8) there exists a point $\bar{x}^n = (\bar{x}_1, \ldots, \bar{x}_n) \in P_n$, such that if $j \in Q_i^*$, then $\bar{x}_j \in K_i$ $\quad 0 \leq i \leq h$. Since the diameter of K_i, $1 \leq i \leq h$, is smaller than $\delta^*/2$ we have that $\{i : |\bar{x}_i - y_i| \geq \delta^*/2\} \subset (\bigcup_{i=1}^{h} (Q_i^* - Q_i)) \cup Q_0^*$. Then (4.4), (4.8) and (4.9) imply that $\#\{i : |\bar{x}_i - y_i| \geq \delta^*/2\} < n\delta^*/4 + \# Q_0^* < n\delta^*/4 + n\delta^*/4 \leq n\delta^*/2$. Therefore $d_{1n}(\bar{x}^n, y^n) < \delta^*/2$, and (4.7) is true. Then by (4.3), (4.5) and (4.6) we have

(4.10) $\quad \mu(S_{n\delta^*/2}(x^n) < \varepsilon) \geq 1 - a\, b^n \qquad \forall\, n \geq n_1 .$

Then there exists n_2 such that:

(4.11) $\quad \mu\left(\bigcap_{n=n_2}^{\infty} S_{n\delta^*/2}(x^n) < \varepsilon\right) \geq 1 - \varepsilon/2.$

Finally, we can find $\delta < \delta^*/2$ such that

(4.12) $\quad \mu(S_{n\delta}(x^n) < \varepsilon) \geq 1 - \varepsilon/(2(n_2 - n_0)).$

Then (4.1) may be derived from (4.11) and (4.12).

The following theorem establishes the equivalence between weak pointwise robustness and $\Pi_{d_{1n}}$-Π-robustness when $\Lambda = R^k$.

Theorem 4.2. Let $\mu \in P(X^\infty)$, then

(i) If $(T_n)_{n \geq n_o}$ is weakly pointwise robust at μ, then it is $\Pi_{d_{1n}}$-Π-robust at μ.

(ii) Suppose that $\Lambda = R^k$ and $\lambda(u,v) = \max_{1 \leq i \leq k} |u_i - v_i|$, or any other equivalent metric, where $u = (u_1,\ldots,u_k)$ and $v = (v_1,\ldots,v_k)$. Let $\mu \in P(X^\infty)$ then if $(T_n)_{n \geq n_o}$ is $\Pi_{d_{1n}}$-Π-robust at μ, it is also weakly pointwise robust at μ.

To prove theorem 4.2 we need Lemmas 4.1 and 4.2 which are proved in the Appendix.

Lemma 4.1. Let (A,ρ) and (Λ,λ) be two polish spaces, $T : A \to \Lambda$ a measurable function with respect to the Borel σ-field.
For any $a \in A$, $\delta > 0$ let $S_\delta(a) = \sup \{ \lambda(T(b),T(c)) : b,c \in V(a,\delta,\rho)\}$.
Then we have

(i) S_δ is lower semicontinuous and therefore measurable.

(ii) For any $\delta > 0$, $\varepsilon > 0$ and $\eta > 0$, there exists measurable functions $U^{(j)} : A \to A$, $j = 1,2$ such that

(a) $U^{(j)} \in V(a,\delta + \eta,\rho)$ $j = 1,2$,

(b) $\lambda(T(U^{(1)}(a)), T(U^{(2)}(a))) > S_\delta(a) - \varepsilon$.

iii) Given $\delta > 0$, there exist measurable functions $U^{(j)}: A \to A$, $j = 1,2$ such that:

(a) $U^{(j)} \in V(a, 2\delta, \rho)$

(b) $\lambda(T(U^{(1)}(a)), T(U^{(2)}(a)) > \dfrac{S_\delta(a)}{2}$.

Lemma 4.2.: Let $F: R \to [0,1]$ be a distribution function and d the usual metric on R $d(x,y) = |x-y|$. Then, given $\varepsilon > 0$ there exists $\delta > 0$ such that if X and Y are random variables satisfying:

(a) $Y > X$

(b) $P(Y > X + \varepsilon) > \varepsilon$

then $\Pi_d(\mathcal{L}(X), \mathcal{L}(Y)) > \delta$.

Proof of Theorem 4.2. i) Given $\varepsilon > 0$ we have to find $\delta > 0$ such that for any $\nu_n \in P(X^n)$ we have

$$(4.13) \quad \Pi_{d_{1n}}(\mu_n, \nu_n) < \delta \Rightarrow \Pi_\lambda(\mathcal{L}(T_n, \mu_n), \mathcal{L}(T_n, \nu_n)) < \varepsilon .$$

By assumption we can choose δ_1 such that $\mu_n(B) > 1 - \varepsilon/2$ where

$B = \{x : S_{n\delta_1}(x) < \varepsilon\}$.

Take $\delta = \min\{\delta_1, \varepsilon/2\}$ and suppose that $\Pi_{d_{1n}}(\mu_n, \nu_n) < \delta$, therefore by the Strassen theorem, there exists $R \in P(\mu_n, \nu_n)$ such that $R(\Delta) > 1 - \delta$ where $\Delta = \{(x^n, y^n) : d_{1n}(x^n, y^n) < \delta\}$. Therefore we have $R(\Delta \cap B) > 1 - \varepsilon$ and this implies (4.13).

(ii) Suppose now that $(T_n)_{n > n_o}$ is not weakly pointwise robust at μ, therefore there exists $\varepsilon > 0$ such that for all $\delta > 0$ there exists $n(\delta)$ such that

$$(4.14) \qquad \mu_{n(\delta)}(x^{n(\delta)} : S_{n(\delta)\delta}(x^{n(\delta)}) > \varepsilon) > \varepsilon .$$

By Lemma 4.1 we can find for any $\delta > 0$ and $n > n_o$ a pair of measurable function $U_n^{(j)}(x^n) : X^n \to X^n$, $j = 1,2$ such that

$$(4.15) \qquad d_{1n}(U_n^{(j)}(x^n), x^n) < 2\delta \qquad j = 1,2$$

and

$$\lambda(T_n(U_n^{(1)}(x^n)), T_n(U_n^{(2)}(x^n))) > S_{n\delta}(x^n)/2.$$

Therefore, we have by (4.14)

$$\mu_{n(\delta)}(\lambda(T_{n(\delta)}(U_{n(\delta)}^{(1)}(x^n)), T_{n(\delta)}(U_{n(\delta)}^{(2)}(x^n))) > \varepsilon/2) > \varepsilon .$$

Suppose that $T_n = (T_{n,1}, \ldots, T_{n,k}) \in R^K$. Then, since λ is the maximum distance between coordinates, there exists $i_o(\delta)$ such that for all δ

$$\mu_{n(\delta)}(|T_{n(\delta),i_o(\delta)}(U^{(1)}_{n(\delta)}(x^n)) - T_{n(\delta),i_o(\delta)}(U^{(2)}_{n(\delta)}(x^n))| \geq \varepsilon/2) > \varepsilon/k.$$

It is clear that we may suppose that $T_{n(\delta),i_o(\delta)}(U^{(1)}_{n(\delta)}(x^n)) < T_{n(\delta),i_o(\delta)}(U^{(2)}_{n(\delta)}(X^n))$ interchanging the functions if necessary. Therefore by Lemma 4.2 there exists ε_o such that

$$\Pi_d(\mathcal{L}(T_{n(\delta),i_o(\delta)}(U^{(1)}_{n(\delta)}(x^n)),\mu_n),\mathcal{L}(T_{n(\delta),i_o(\delta)}(U^{(2)}_{n(\delta)}(x^n)),\mu_n)) > \varepsilon_o .$$

This implies

$$(4.16)\qquad \Pi_\lambda(\mathcal{L}(T_{n(\delta)}(U^{(1)}_{n(\delta)}(x^n),\mu_n),\mathcal{L}(T_{n(\delta)}(U^{(2)}_{n(\delta)}(x^n)),\ \mu_n)) > \varepsilon_o.$$

Choose now δ_o such that for any $\nu_n \in P(X^n)$ we have

$$(4.17)\qquad \Pi_{d_{1n}}(\mu_n,\nu_n) < \delta_o \Rightarrow \Pi_\lambda(\mathcal{L}(T_n,\mu_n),\mathcal{L}(T_n,\nu_n)) < \varepsilon_o/4 \quad \forall\, n \geq n_o .$$

Let $\nu^{(j)}_n = \mathcal{L}(U^{(j)}_{n(\delta)}(x^n),\mu_n)$.

Then, we have that (4.16) implies that for some j

(4.18) $\quad \Pi_\lambda(\mathcal{L}(T_{n(\delta)},\nu_n^{(j)}),\ \mathcal{L}(T_{n(\delta)},\mu_n)) > \varepsilon_o/2$,

and if we choose $\delta = \delta_o/2$ in (4.15) we have

(4.19) $\quad \Pi_{d_{1n}}(\mu_n,\nu_n^{(j)}) < \delta_o \qquad j = 1,2$,

but (4.18) and (4.19) contradicts (4.17).

The following continuity condition, which is a generalization of Hampel's continuity condition (1971) was given by Papantoni-Kazakos and Gray (1979).

<u>Definition</u>. A sequence of estimators $(T_n)_{n \geq n_o}$ <u>is continuous at $\mu \in P(X^\infty)$</u> if given $\varepsilon > 0$ there exist positive integers k and n_1 and $\delta > 0$ such that if $n,m > n_1$ $x^n \in X^n$, $y^m \in X^m$ and

(4.20) $\quad \Pi_{d_{2k}}(\mu_k[x^n],\mu_k) < \delta,\ \ \Pi_{d_{2k}}(\mu_k\,[y^m],\mu_k) < \delta$

imply

(4.21) $\quad \lambda(T_n(x^n),\ T_m(y^m)) < \varepsilon$.

In particular if a single k works for all ε , T_n will be said <u>continuous of order k at μ</u> .

The following theorem gives a sufficient condition for strong pointwise robustness.

Theorem 4.3: Let $\mu \in P(X^\infty)$ and suppose that (a) $(T_n)_{n \geq n_o}$ is continuous at μ, (b) Each T_n is continuous as a function of x^n and (c) μ is stationary and ergodic. Then $(T_n)_{n \geq n_o}$ is strongly pointwise robust at μ.

Proof. By Proposition 4.1 is enough to prove that given $\varepsilon > 0$, there exists $\delta > 0$ such that (4.1) holds. Take k, δ and n_1 such that (4.20) implies (4.21).

Since μ is ergodic the Glivenko-Cantelli theorem holds, then

$$\mu \, (\sup_{A \in F^k} | \mu_k[x^n](A) - \mu_k(A) | \to 0) = 1.$$

Therefore there exists n_2 such that if $\delta_1 = \delta/k$ then

$$\mu \, (\bigcap_{n \geq n_2} \{ x \in X^\infty \; \sup_{A \in F^k} | \mu_k[x^n](A) - \mu_k(A) | < \delta_1/2 \}) > 1 - \varepsilon/2 \, ,$$

and since for any μ_k and ν_k in $P(X^k)$, it is true that

$\Pi_{d_{2k}}(\mu_k, \nu_k) < \sup_{A \in F^k} | \mu_k(A) - \nu_k(A) |$, we have

$$\mu\left(\bigcap_{n>n_2}\{x\in X^\infty : \Pi_{d_{2k}}(\mu_k[x^n],\mu_k)<\delta_1/2\}\right)> 1-\varepsilon/2.$$

Put

(4.22) $A_n = \{x^n : \Pi_{d_{2k}}(\mu_k[x^n],\mu_k)<\delta_1/2\}$ and $B_n = \{S_{n\delta_1/4}(x^n)<\varepsilon\}$.

We will show that there exists n_3 such that

(4.23) $A_n \subset B_n \quad \forall\, n>n_3$.

Take $x^n \in A_n$ and $d_{1n}(x^n,y^n)<\delta_1/4$, then it is easy to see, by Lemma 2.3 that $\Pi_{d_{2k}}(\mu_k[x^n],\mu_k[y^n])<(\delta_1/4)(nk/(n-k))$. Therefore there exists n_3 such that $\Pi_{d_{2k}}(\mu_k[x^n],\mu_k[y^n])<\delta_1/2$ for all $n \geq n_3$. Hence since $x^n \in A_n$ $\Pi_{d_{2k}}(\mu_k[y^n],\mu_k)<\delta_1$ therefore by the continuity condition, if $n_4 = \max(n_1,n_2,n_3)$ we have $\lambda(T_n(x^n),T_n(y^n))<\varepsilon$, therefore $S_{n\delta_1}(x^n)<\varepsilon$ and (4.23) is true, therefore

$$\mu\left(\bigcap_{n>n_4} B_n\right)> \mu\left(\bigcap_{n>n_4} A_n\right)>1-\varepsilon/2 .$$

Finally, since T_n is continuous as a function of x^n , there exists

δ_2 such that $\mu(S_{n\delta_2}(x^n) < \varepsilon) > 1 - \varepsilon/(2(n_4 - n_o)) \forall \; n_o < n < n_4$.

Hence, if $\delta^* = \min(\delta_1/4, \delta_2)$ we have

$$\mu(\bigcap_{n \geq n_o} S_{n\delta^*}(x^n) < \varepsilon) > 1 - \varepsilon$$

and then the theorem is proved.

Corollary 4.1.: Let $T_n(x^n) = T(\mu_k[x^n])$ $n > k$, where $T : Q \to \Lambda$ and $Q \subset P(X^k)$ containing the empirical distributions. Suppose that $\mu \in P(X^\infty)$ is stationary and ergodic, T is continuous at μ_k with respect to $\Pi_{d_{2k}}$ and $T_n(x^n)$ is continuous as a function of x^n, then $(T_n)_{n \geq k}$ is strongly pointwise robust at μ .

A sufficient condition for the continuity of a general class of mappings is given in Lemma 5 of Cox (1979). This lemma entails the strong pointwise robustness of the GM-estimators for autoregressive models presented in Denby and Martin (1979) and Bustos (1982).

Sufficient conditions for the continuity of M, L and R-estimates for location are given by Huber (1981) in Theorems 2.6, 3.1 and 4.1 of Chapter 3. This implies the strong pointwise robustness of these estimators even in the case of dependent stationary and ergodic observations.

5. Appendix.

Proof of Lemma 4.1.: i) Let $(a_i)_{i>1}$ be a sequence such that $\lim \rho(a_i,a) = 0$ we must show that $S_\delta(a) = \liminf S_\delta(a_i)$.

Let b and c be two points in $V_\delta(a) = V(a,\delta,\rho)$ and $\alpha = \min(\rho(a,b),\rho(a,c))$, then $\alpha < \delta$. Since $\lim \rho(a_i,a) = 0$ there exists i_o such that $\rho(a_i,a) < \delta - \alpha$, $i > i_o$. Then $b \in V_\delta(a_i)$, $c \in V_\delta(a_i)$ $i > i_o$ and therefore $\lambda(T(b),T(c)) < S_\delta(a_i)$ which proves (i).

ii) Let $A_o = (a_n)_{n>1}$ be a denumerable dense set on A. Let $a_i^{(j)}$ $1 < i < \infty$, $j = 1,2$ be in $V_\delta(a_i)$ and

(5.1) $\lambda(T(a_i^{(1)}, T(a_i^{(2)})) > S_\delta(a_i) - \varepsilon/3$.

Define $z_i^{(j)} : A \to A$ $1 < i$ $j = 1,2$ by

$$(5.2) \qquad z_i^{(j)} = \begin{cases} a_i^{(j)} & \text{if } a \in V_\delta(a_i) \\ a & \text{if } a \notin V_\delta(a_i) \end{cases}$$

Clearly, $z_i^{(j)}$ are measurable functions. We will show that for any a there exists i such that:

(5.3) $z_i^{(j)} \in V(a, \eta + \delta, \rho)$ $j = 1,2$

(5.4) $\lambda(T(Z_i^{(1)}(a), T(Z_i^{(2)}(a))) > S_\delta(a) - 2\varepsilon/3.$

As A_o is a dense set there exist $(a_i)_{i \geq 1}$ such that $\lim \rho(a_i, a) = 0$ and by the lower semicontinuity of S_δ we obtain $S_\delta(a) < \liminf S_\delta(a_i)$, then there exists i such that $S_\delta(a) < S_\delta(a_i) + \varepsilon/3$ and $a_i \in V_{\eta_1}(a)$ where $\eta_1 = \min(\eta, \delta)$.

Then by (5.2) $\lambda(T(a_i^1), T(a_i^2)) = \lambda(Z_i^1(a), Z_i^2(a)) > S_\delta(a_i) - \varepsilon/3 > S_\delta(a) - 2\varepsilon/3$ and $\rho(a_i^{(j)}, a) < \rho(a_i^{(j)}, a_i) + \rho(a_i, a) < \delta + \eta$, which proves (5.3) and (5.4).

Let $S^*_{\delta,\eta,\varepsilon} = \sup_{i > 1} \lambda(T(Z_i^{(1)}(a), T(Z_i^{(2)}(a))) I_{V_{\delta+\eta}(a_i^{(1)})}(a) I_{V_{\delta+\eta}(a_i^{(2)})}(a)$,

where $I_A(a)$ denotes the indicator function of A at the point a, then $S^*_{\delta,\eta,\varepsilon}(a)$ is measurable and we have

(5.5) $S_{\delta+\eta}(a) > S^*_{\delta,\eta,\varepsilon}(a) > S_\delta(a) - 2\varepsilon/3$.

Now define $i(a)$ as the first i such that $\lambda(T(Z_i^{(1)}(a)), T(Z_i^{(2)}(a)) > S^*_{\delta,\eta,\varepsilon}(a) - \varepsilon/3$ and $Z_i^{(j)} \in V(a, \delta+\eta, \rho)$. Define $U^{(j)}(a) = Z_{i(a)}^{(j)}(a)$ since $S^*_{\delta,\eta,\varepsilon}(a)$ is measurable, $U^{(j)}(a)$ $j = 1,2$ are measurable and satisfy (ii) (a) and (b).

iii) Let $\varepsilon_i \to 0$ as $i \to \infty$ and choose $U_i^{(j)}$ $j = 1,2$ $1 < i$ as in part ii) with $\eta = \delta$ and $\varepsilon = \varepsilon_i$. Define $i(a)$ as the first i such that $\lambda(T(U_i^{(1)}(a)), T(U_i^{(2)}(a))) > S_\delta(a)/2$.

Define now $U^{(j)}(a) = U_{i(a)}^{(j)}(a)$ $j = 1,2$. Clearly since S_δ and

$U_i^{(j)}$ are measurable, these functions are measurable and satisfy the conditions (iii) (a) and (b).

Proof of Lemma 4.2. Choose $x_o < x_1 < \ldots < x_m$ points of continuity of F such that $F(x_m) - F(x_o) > 1 - \varepsilon/3$ and $x_{i+1} - x_i < \varepsilon/3$. Since $x_o, \ldots$ $\ldots, x_m$ are points of continuity of F we can choose λ such that

$$(5.6) \quad F(x_i - \lambda) > F(x_i) - \varepsilon/(3m) \quad 0 \leq i \leq m.$$

We can also choose i_o such that $F(Y > X + \varepsilon \quad x_{i_o} \leq X < x_{i_o+1}) >$ $> 2\varepsilon/(3m)$. It is clear that $\{(Y > x_{i_o} + \varepsilon/3\} \supset \{X > x_{i_o+1}\} \cup$ $\{Y \geq X + \varepsilon, x_{i_o} < X < x_{i_o+1}\}$, therefore $P(Y > x_{i_o} + \varepsilon/3) > P(X > x_{i_o+1}) + 2\varepsilon/(3m)$. Therefore by (5.6) $P(Y > x_{i_o} + \varepsilon/3) > P(X > x_{i_o+1} - \lambda) + \varepsilon/(3m) > P(X > x_{i_o} + \varepsilon/3 - \lambda) + \varepsilon/3m)$. Then if $\delta = \min(\varepsilon/3m, \lambda)$ we have $\Pi_d(\mathcal{L}(X), \mathcal{L}(Y)) > \delta$.

Acknowledgement: Grateful thanks for helpful discussions are due to D. Martin and D. Cox.

References.

Bustos, O. (1980). Qualitative robustness for general processes. Unpublished manuscript. Instituto de Matematica Pura e Aplicada, Rio de Janeiro.

Bustos, O. (1982). General M-estimates for Contaminated pth-Order Autoregressive Processes: Consistency and Asymptotic Normality. Robustness in Autoregressive Processes. Z. Wahrscheinlich-keitstheorie verw. Gebiete, **59**, 491-504.

Cox, D. (1978). Metrics on stochastic processes and qualitative robustness. Technical Report No. 23. Department of Statistics. University of Washington.

Hampel, R.F. (1971). A general qualitative definition of robustness. Ann. Math. Statist., **42**, 1887-1896.

Huber, P.J. (1981). Robust Statistics - Wiley - New York.

Denby, L. and Martin, D.R. (1979). Robust estimation on the first order autoregressive parameter. Jour. Amer. Stat. Assoc., **74**, 140-146.

Papantoni-Kazakos, P. and Gray, R. (1979). Robustness of estimators on stationary observations. Ann. of Probability, 6, 989-1002.

END

FILMED

10-84

DTIC