AD-A198 290

DTIC FILE COPY

AFOSR-TR- 88-0795

2

[redacted] Technical Report
Grant No. AFOSR-87-0224
July 1, 1987 - June 30, 1988

ROBUST ALGORITHMS FOR DETECTING A CHANGE IN A STOCHASTIC PROCESS WITH INFINITE MEMORY

Submitted to:

Air Force Office of Scientific Research
Building 410
Bolling Air Force Base
Washington, DC 20332-6448

Attention: Major Brian W. Woodruff, NM

Submitted by:

P. Papantoni-Kazakos
Department of Electrical Engineering
University of Virginia
Charlottesville, VA 22901

Rakesh Kumar Bansal
Postdoctoral Research Associate

Report No. UVA/525682/EE88/105
March 1988

SCHOOL OF ENGINEERING AND APPLIED SCIENCE

DEPARTMENT OF ELECTRICAL ENGINEERING

UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22901

88 8 25 126

[redacted] Technical Report
Grant No. AFOSR-87-0224
July 1, 1987 - June 30, 1988

ROBUST ALGORITHMS FOR DETECTING A CHANGE IN A STOCHASTIC PROCESS WITH INFINITE MEMORY

Submitted to:

Air Force Office of Scientific Research
Building 410
Bolling Air Force Base
Washington, DC 20332-6448

Attention: Major Brian W. Woodruff, NM

Submitted by:

P. Papantoni-Kazakos
Department of Electrical Engineering
University of Virginia
Charlottesville, VA 22901

Rakesh Kumar Bansal
Postdoctoral Research Associate

Department of Electrical Engineering
SCHOOL OF ENGINEERING AND APPLIED SCIENCE
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA

Report No. UVA/525682/EE88/105 Copy No. _____
March 1988

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

1b. RESTRICTIVE MARKINGS: None

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S): UVA/525682/EE88/105

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 88-0795

6a. NAME OF PERFORMING ORGANIZATION: University of Virginia, Dept. of Electrical Engr.

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code): Thornton Hall, Charlottesville, VA 22901

7b. ADDRESS (City, State and ZIP Code): Building 410, Bolling Air Force Base, Washington, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Air Force Office of Scientific Research

8b. OFFICE SYMBOL (If applicable): NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-87-0224

8c. ADDRESS (City, State and ZIP Code): Building 410, Bolling Air Force Base, Washington, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE (Include Security Classification): Robust Algorithms For Detecting A Change In A Stochastic Process With Infinite Memory

12. PERSONAL AUTHOR(S): P. Papantoni-Kazakos and Rakesh Kumar Bansal

13a. TYPE OF REPORT: ~~Annual Technical~~ Journal

13b. TIME COVERED: FROM 7/1/87 TO 6/30/88

14. DATE OF REPORT (Yr., Mo., Day): 1988, March 23

15. PAGE COUNT: 9

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): The

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The authors ~~We~~ present and discuss a class of continuous operations on the family of discrete time stochastic processes, which serves as a guide to construct qualitatively robust operations for a given class of processes, namely the one induced by a nominal process and a substitutive contaminating process. ~~Our~~ The results are general enough to help develop any robust statistical procedure, but we have concentrated ~~our~~ attention on detection of a change from one class of processes to another (disjoint) class of processes, while both classes consist of not necessarily Markov processes and satisfy certain mixing conditions in addition to stationarity and ergodicity. Two quantitative measures of robustness, breakdown point and influence functions are also developed for few examples. (KR)

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Brian W. Woodruff

22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5028

22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

Preprint of a paper that will appear in the Proceedings of the 1988 Annual Conference on Information Sciences and Systems.

ROBUST ALGORITHMS FOR DETECTING A CHANGE IN A STOCHASTIC PROCESS WITH INFINITE MEMORY

by

Rakesh K. Bansal and P. Papantoni-Kazakos
Department of Electrical Engineering
University of Virginia
Charlottesville, VA 22901

Abstract

We present and discuss a class of continuous operations on the family of discrete time stochastic processes, which serves as a guide to construct qualitatively robust operations for a given class of processes, namely the one induced by a nominal process and a substitutive contaminating process. Our results are general enough to help develop any robust statistical procedure, but we have concentrated our attention on detection of a change from one class of processes to another (disjoint) class of processes, while both classes consist of not necessarily Markov processes and satisfy certain mixing conditions in addition to stationarity and ergodicity. Two quantitative measures of robustness, breakdown point and influence functions are also developed for few examples.

0. Introduction

Consider two stationary and ergodic processes $[\mu_0, X_0, R]$ and $[\mu_1, X_2, R]$, where μ_0 and μ_1 are the two distinct probability measures on (R^∞, B_∞) or $(R^\infty_{-\infty}, B^\infty_{-\infty})$ as the case may be, X_1 and X_2 are their names and R is the real line on which both processes take their values. As usual (R^∞, B_∞) and $(R^\infty_{-\infty}, B^\infty_{-\infty})$ denote respectively the one sided and two sided infinite product of real line with itself with their corresponding product σ-algebras. Let W_i^j, $j \geq i$ denote an observed sequence. Suppose we start observing at time instant one and suppose initially the process μ_o is active. Suppose at some time instant $t \geq 1$, process μ_0 becomes inactive and μ_1 becomes active and remains so. Our objective is to formulate a meaningful test to detect this shift from μ_0 to μ_1. To attain this objective a number of algorithms have been proposed, developed and studied in literature. Most widely studied ones are the Page's algorithm, Page [1954] and Shiryayev-Roberts' algorithm, Shiryayev (1963) and Roberts (1966). Page developed the algorithm under i.i.d. set up and Lorden (1971) studies it and proved its asymptotic optimality using a minimax criteria. Bansal and Papantoni-Kazakos (1986) modified it and proved it's optimality under non i.i.d. setup.

But suppose our description of the two measures under consideration is imperfect or perhaps our observations are vulnerable to contamination by another unknown measure then we need to develop robust/outlier resistant algorithms in order to achieve relatively stable performance possibly by sacrificing the efficiency of the algorithm that is achievable at the ideal model (i.e. in the absence of any deviation whatsoever from the assumed structure). Before we proceed further, however, let us state our generalized observation model in a precise manner.

Let Y_j be the observation taken at time instant j where

$$Y_j = (1-A_j)W_j + A_j Z_j, \; j \geq 1 . \tag{1}$$

Here W_j is governed by the nominal measure μ_0 or μ_1 and $\{Z_j\}$ is an i.i.d. process which is arbitrary and $\{A_j\}$ is a binary i.i.d. process with

$$\Pr\{A_j = 1\} = 1 - \Pr\{A_j = 0\} = \varepsilon .$$

Essentially each nominal random variable W_j is replaced by an arbitrary random variable Z_j with frequency (probability) ε before we get to observe it. $\{Z_j\}$ is the contaminating process and $\{A_j\}$ determines the contamination law. In the absence of contaminating process (ε=0), we have perect observations and then we can apply our optimal algorithm Bansal et al (1986). However, under contamination, the optimal algorithm becomes totally unreliable in the sense that just a single bad observation can overwhelm the evidence provided by other good observations and upset the decision. This will become apparent the moment we see the optimal test, which is defined as follows.

Given the (uncontaminated) data sequence $w=(w_1, w_2, ...)$ and letting w_1^n denote the finite sequence $(w_1, w_2, ... w_n)$ stop at

$$N_\delta^0(w) = \inf\{n: T_n^0(w_1^n) \geq \log\delta\} \tag{2}$$

where log δ is the logarithmic threshold chosen and

$$T_n^0(w_1^n) \overset{\Delta}{=} \max_{1 \leq k \leq n+1} \left[\sum_{i=k}^{n} \log \frac{f_1(w_i \mid w_k^{i-1})}{f_0(w_i \mid w_k^{i-1})} \right] \tag{3}$$

is the test statistic with appropriate end conditions and $f_i(\cdot/\cdot)$; i=0,1 denote the conditional densities of μ_i with respect to an appropriate σ-finite measure, whose existence we assume. See Bansal et al (1986) for other necessary regularity conditions and complete details. Notice from the expression of $T_n^0(w_1^n)$, a single term inside the summation can make $T_n^0(w_1^n)$ too large or too small if μ_1 and μ_0 do not have compact support. And therefore $N_\delta^0(w)$ can be too small or too large and in essence the test may became unreliable. This is invariably the problem with all the classical parametric tests or estimators, many of which are optimal in an appropriate sense. Lately (from the last twenty five years or so) we have become "more" aware of our inability to model a phenomena accurately and the vulnerability of our observations to gross errors and have focused our attention to the development of robust procedures by sacrificing efficiency or the optimality. Naturally one would like to quantify robustness in order to evaluate the tradeoff, which opens the new area of optimal robust procedures. However often because of the complexity of the observation model it becomes (or at least appears to be) impossible to design optimal robust procedures. In

This work was supported by the Air Force Office of Scientific Research under the grant: AFOSR-87-0224.

our problem at hand we perceive this handicap. Therefore, we have attempted to look for intuitively meaningful procedures and examine their performance in terms of efficiency (or loss thereof) and the breakdown point and the influence function. The structure of the optimal algorithm is used as the starting point and the guide to the development of such procedures. Before we discuss our approach it is important to consider the special cases of the observation model in (1) and what procedures have already been studied in literature. This will serve two purposes. One, offer insight into what we could reasonably expect from our robust algorithms for the general case and two, the inapplicability of the existing robust algorithms under those special cases which we are about to discuss.

Note from (1), that if all three component processes $\{W_j\}$, $\{A_j\}$ and $\{Z_j\}$ are i.i.d. then the process $\{Y_j\}$ is i.i.d. and then it suffices to develop procedures that are based on one-dimensional marginal distribution alone. In fact then, $\{Y_j\}$ could have this alternative description

$$f_{Y_j}(y) = (1-\epsilon) f_{W_j}(y) + \epsilon h(y) \tag{4}$$

where $h(y)$ is the density function of Z_j. Also $\{W_j\}$ being i.i.d. under μ_1 and μ_0 both, means

$$T_n^0(w_1^n) = \max_{1 \le k \le n+1} \left\{ \sum_{i=k}^{n} \log \frac{f_1(w_i)}{f_0(w_i)} \right\} . \tag{5}$$

Under these stricter conditions on the process $\{Y_j\}$ it seems natural to replace the pair $(f_1(w_i), f_0(w_i))$ by a least favorable pair $(q_1(w_i), q_0(w_i))$ where $q_1(\cdot)$, $q_0(\cdot)$ minimize the Kullback distance. See Huber (1965) for details. Notice that (4) and (1) both describe two distinct classes of processes obtained by two nominal measures μ_1 and μ_0.

Let us denote the new test statistic by $T_n'(w_1^n)$ and the resulting stopping varriable by $N_\delta'(w)$, where

$$T_n'(w_1^n) = \max_{1 \le k \le n+1} \left\{ \sum_{i=k}^{n} \log \frac{q_1(w_i)}{q_0(w_i)} \right\} . \tag{6}$$

Notice that $N_\delta'(w)$ is the stopping variable resulting from Page's test, applied to the least favorable pair of (i.i.d.) processes, which detects the change from μ_{q_0} to μ_{q_1} in an optimal manner as shown by Lorden (1971). And since this test resulted from the minimax robustification of the test for μ_0 to μ_1 shift, it becomes the optimal minimax robust test, which we prefer to call optimal in super minimax sense. This was quite straightforward because of the i.i.d. structure of all the component processes, which in turn induced an i.i.d. structure on the observation process.

But suppose that our nominal measures are not i.i.d., but they are Markovian (Note that the observation process is no more Markovian even though the component processes are). Then again one is tempted to robustify $T_n^0(w_1^n)$ in (3) by applying suitable transformation on $\frac{f_1(wi/w_k^{i-1})}{f_0(w_i/w_k^{i-1})}$. In order to apply the approach used for i.i.d. case we need to obtain two classes of conditional densities similar to the ones given in (4). It turns out that it is impossible to obtain an exact description of the model in (1) in the mixture form given in (4). Note that (4) and (1) are equivalent under i.i.d. setup, but otherwise (4) is a strict enlargement of (1). But (4) by itself does not enable us to obtain suitable replacement of the conditional log likelihood ratios. To overcome this problem we have used two approaches, one leading to approximate description of (1) in mixture form, based on nominal conditional densities of the two measures μ_1 and μ_0 and the other being a strict enlargement of (1), using the variational metric and an additional assumption of ϕ-mixing on the nominal measures. Then Huber's operations were applied and the corresponding pairs of least favorable conditional densities were used to obtain a test stastic, which we denote by $T_n''(w_1^n)$ here. These results were developed and reported in Bansal and Papantoni-Kazakos (1987a) in detail and in Bansal and Papantoni-Kazakos (1987b) in part and in condensed form. Under the Markovian assumption of the nominal measures μ_1 and μ_0, the conditional likelihood ratios in (3) have finite memory. Huber's operation induces uniformly bounded conditional likelihood ratios, where the lower and upper bounds both depend on the finite number of 'past' observations. Moreover, if the nominal conditional densities are continuous functions of the observation block, then the modified densities are also continuous. Boundedness of each term under the summation, it's continuity as a point function and its dependence on only finite number of variables suffices to ensure the qualitative robustness of the test statistic $T_n''(w_1^n)$, the corresponding stopping variable and the sequence of functionals $E_{\mu_i}\{n^{-1} T_n''(w_1)\}$. Here $E_{\mu_i}\{\cdot\}$ denotes the expected value under the measure μ_i. Readers are referred to Boente et al (1987) and Papantoni-Kazakos (1987) for extensive discussion of qualitative robustness for stochastic process. However the moment we relax the Markovian assumption, our nominal conditional densities depend on the entire past which grows to infinity as sample size goes to infinity. Then the quantities under the summation will have unbounded number of arguments and the functional $\lim_{n\to\infty} E_{\mu_i}\{T_n''(w_1^n)/n^{-1}\}$ will depend on the entire process and not just on finite order marginals of the process. Under these circumstances, boundedness of $n^{-1}T_n''(w_1^n)$ is not enough to guarantee qualitative robustness in general. Counter examples to illustrate this phenomenon in case of estimation of the parameters of a moving average process are given in Martin and Yohai (1986) and Boente et al (1987). The phenomena is explained in different ways in the above two works but we will provide our own explanations toward the end of this section.

However, the robustness of the 'Huberized' conditional log likelihood ratios used in $T_n''(w_1^n)$ can be restored if we limit the memory being used in the likelihood ratios in an artifical manner. But we need to make a judicious choice of that because too large a memory will result in weaker robustness (as measured through breakdown point for example) and higher efficiency and too small a memory would result in stronger robustness and lower efficiency. This approach will suffer from another weakness, that is the resulting algorithm will not reduce to the optimal one unless we let the size of the memory being used, depend on the design parameter ϵ such that it goes to infinity as $\epsilon \downarrow 0$. We intend to report the work on this issue elsewhere. The results we are going to report in this paper are obtained by using an alternative approach. But before we discuss these, it is profitable to discuss some of the limiting properties of the algorithm based on $T_n''(w_1^n)$.

By definition,

$$T_n''(w_1^n) \stackrel{\Delta}{=} \max_{1 \le k \le n+1} \left\{ \sum_{i=k}^{n} \log \frac{q_1(w_i/w_k^{i-1})}{q_0(w_i/w_k^{i-1})} \right\} . \tag{7}$$

For convenience, we had proposed the minor modification of $T_n''(w_1^n)$ in Bansal et al (1987 a, 1987b)

$$T_n'''(w_1^n) = \max_{1\le k\le n+1}\left[\sum_{i=k}^{n} \log \frac{q_1(w_i/w_1^{i-1})}{q_0(w_i/w_1^{i-1})}\right]$$

$$= \max_{1\le k\le n+1}\left[\sum_{i=k}^{n} g_i'''(w_1^i)\right] \qquad (8)$$

where

$$g_i(w_1^i) = \begin{cases} c_{1,\varepsilon}(w_1^{i-1})\,;\ \ln\frac{f_1(w_i/w_1^{i-1})}{f_0(w_i/w_1^{i-1})} \le c_{1,\varepsilon}(w_1^{i-1}) \\ \ln\frac{f_1(w_i/w_1^{i-1})}{f_0(w_i/w_1^{i-1})}\,;\ c_{1,\varepsilon}(w_1^{i-1}) < \ln\frac{f_1(w_i/w_1^{i-1})}{f_0(w_i/w_1^{i-1})} \\ \qquad < c_{0,\varepsilon}(w_1^{i-1}) \\ c_{0,\varepsilon}(w_1^{i-1})\,;\ \ln\frac{f_1(w_i/w_1^{i-1})}{f_0(w_i/w_1^{i-1})} \ge c_{0,\varepsilon}(w_1^{i-1}) \end{cases} \qquad (9)$$

and $c_{1,\varepsilon}(w_1^{i-1})$ and $c_{0,\varepsilon}(w_1^{i-1})$ are the lower and upper thresholds determined by applying Huber's operation. These thresholds depend on w_1^{i-1} at stage i, but are uniformly bounded below and above by $-\ln\frac{1-\varepsilon}{\varepsilon}$ and $\ln\frac{1-\varepsilon}{\varepsilon}$ respectively.

Compare T_n', T_n'' and T_n''' in (6), (7) and (8) respectively. Note that if the observation process is i.i.d. T_n'' and T_n''' both reduce to T_n' which was seen to be optimal in super minimax sense. Next, as $\varepsilon\downarrow 0$, our two classes of processes described in (1) reduce to a single pair (μ_0,μ_1) and then $T_n''(w_1^n)$ reduces to $T_n^0(w_1^n)$ which has been proved to be optimal. Seeing these two features of our proposed algorithm in the two extreme (or limiting) cases, we hope to have achieved good robustness without sacrificing a lot of efficiency. Therefore, in whatever we suggest and study next, we would like to retain these two attractive features. This is used as the basic guideline to develop robust algorithms under the most general set of assumptions on the nominal measures.

Here then is the brief outline of the rest of the paper. In section I, we discuss the nature of qualitative robustness for stochastic processes from function analytic approach, describe a convenient class of qualitatively robust operations and provide a simple counter example to illustrate that boundedness and pointwise, coordinatewise continuity is not enough to ensure robustness in general. Then in section II, a simple approach is described to obtain meaningful robust substitutes for the nominal likelihood ratio, which works for a pair of linear processes as nominals. In section III, we investigate the approach in detail on an example of first order moving average processes and evaluate the efficiency, the breakdown point and a version of influence function. Finally in conclusion, other interesting possibilities for further research are pointed out.

I. Qualitative Robustness for Stochastic Processes

Consider the space of measures M defined on (R^∞, B_∞). Let $w=(w_1,..)$ denote a realization of the process. We are interested in robust (continuous) functionals on the space M and in particular continuous linear functionals on M, because most test statistics or estimators induce a linear operation on M itself in general or in special cases on the restrictions of M to an n dimensional Euclidean space, the later denoted by M_n. To ensure robustness, our choice centers then on the functionals which are continuous with respect to the weak star topology on M or its restriction M_n as the case may be. For example, when we form our attention only on those members of M which generate i.i.d. processes, then our statistical operations induce functionals (often linear) on M_1 and for a class of finite order (m say) Markov process class, they induce functionals on M_m alone. Robustness of operations in these two cases is sufficiently well understood. Regarding the justification for the choice of weak star topology on M_n, one can see Hampel (1971), Boente et al (1987) etc. A weak star neighborhood (since it's metrizable by Prohorov metric) captures both the kinds of deviations from the nominal, namely small number of gross errors (outliers) and small error in large number of observations. And weak star topology on M itself will induce weak star topology on M_n. Knowing the fact that weak star topology can be generated by the Prohorov metric, it is important in this case to have a proper choice of the distortion measure itself on the data sequences. In fact, the usual Euclidean metric suffices for R^n, but there is no equivalent of that on R^∞. There are three natural ways however. One, the so called uniform metric

$$\rho_u(x,y) = \max_i \rho(x_i, y_i)\,;\ \bar{\rho} = \lim_{n\to\infty} n^{-1}\sum_{1}^{n}\rho(x_i, y_i)$$

where $x=(x_1,..)$, $y=(y_1,..)\,\varepsilon\, R^\infty$ and ρ is a bounded metric, generating the usual topology on R_1. Incidently ρ_u and $\bar\rho$ both are so strong that they make our usually estimators like mean robust. It's because they induce a strong topology on R^∞ which in turn induces strong topology on M. In fact, the right choice turns out to be product topology which is weak enough to correspond to our desired notion of robustness. And it can be induced by a metric of the following form

$$\rho_p(n,y) = \sum_{1}^{\infty}\alpha_i \frac{\rho(x_i,y_i)}{1+\rho(x_i,y_i)} \qquad (10)$$

where $\alpha_i>0$ and $\sum_1^\infty \alpha_i$ converges. ρ is a metric which induces the usual topology on R_1.

The above justification leads us to concentrate on the class of continuous, linear functionals on M where M is endowed with weak star topology induced by continuous functions (with respect to ρ_p or the product topology). Symbolically we are interested in linear functionals $T{:}M\to R$ of the following form

$$T_\psi(\mu) = \int \psi(x)\,\mu(dx) \quad x\,\varepsilon\, R^\infty,\ \mu\,\varepsilon\, M \qquad (11)$$

where $\psi: R^\infty \to R_1$, ψ is continuous and bounded on R^∞. It's trival to see that for a ψ which is continuous (coordinatewise) and bounded, its restriction to R^n induces continuous functionals on M_n. It's possible to have ψ bounded and continuous (coordinatewise) which are not continuous in product topology on R^∞. An example:

$$\psi(x) = \max_i \min(|x_i|, 1)$$

Therefore, $T_\psi(\mu) = \int \max_i \min(|x_i|, 1)\, \mu(dx)$ is not continuous with respect to μ, even though $\psi(x)|_{R^n}$ is continuous and bounded and therefore, $T_\psi(\mu|_{R^n})$ is continous on M_n.

Now adapting the general discussion above to our observation model in (1), we notice that we are considering the narrow class of stationary and ergodic members of M as described by (1). We will assume that the classes in (1) inherit the subspace topology from M.

II. Linear Processes

Consider the following pair of nominal measures.

$$\left.\begin{aligned} \mu_0 : W_n &= \sum_{i=1}^{n} a_i W_{n-i} + U_n \\ \mu_1 : W_n &= \sum_{i=1}^{n} b_i W_{n-i} + V_n \end{aligned}\right\} \tag{12}$$

where $\{a_i\}, \{b_i\}; i\geq 1$ are distinct sequences of real numbers and $\{U_n\}$ and $\{V_n\}$ are i.i.d. sequences of Gaussian random variables, with identical variance σ^2. Let

$$E\,U_n = m_0 \,;\, E\,V_n = m_1 \ .$$

Note that the above class is general enough to contain ARMA processes.

$$\text{Define } A(w_1^{n-1}) = \sigma^{-1}\left[\sum_{i=1}^{n-1} (b_i-a_i)\, w_{n-i} + (m_1-m_0)\right]$$

$$B(w_1^{n-1}) = \sigma^{-1}\left[\sum_{i=1}^{n-1} (b_i+a_i)\, w_{n-i} + (m_1+m_0)\right] \tag{13}$$

Then the optimal stopping rule $N_\delta^0(w)$ is

$$N_\delta^0(w) = \min\,\{L_\delta^0(w_k^\infty) + k-1 \,/\, k=1,2,3,\ldots\}$$

(Bansal et al (1986)

where

$$L_\delta^0(w) = \inf\,\{n : \sum_{i=1}^{n} g_i^0\,(w_1^i) \geq \log \delta\} \tag{14}$$

and

$$g_i^0(w_1^i) = 2^{-1} A\,(w_1^{i-1}) \left[2\sigma^{-1} w_n - B\,(w_1^{i-1})\right] .$$

Asymptotically as the number of observations increases we have

$$g_i^0\,(w_1^i) \xrightarrow{D} g^0\,(w_{-\infty}^0)$$

$$A\,(w_1^{n-1}) \xrightarrow{D} A\,(w_{-\infty}^0) \tag{15}$$

$$B\,(w_1^{n-1}) \xrightarrow{D} B\,(w_{-\infty}^0) \ .$$

Also the rate at which a shift from μ_0 to μ_1 is detected is determined by $E_{\mu_1}\,\{g^0\,(w_{-\infty}^0)\}$.

$A\,(w_{-\infty}^0)$ and $B(w_{-\infty}^0)$ are not continuous functions of $(w_0, w_{-1},\ldots) \in R_{-\infty}^0$. However, a simple trick will make them continuous and bounded at the same time. If we replace $\{w_i\}$ by say $\{w_i'\}$ where w_i' are uniformly bounded, then under usual conditions on $\{a_i\}$ and $\{b_i\}$, $A(w_{-\infty}'^0)$ and $B(w_{-\infty}'^0)$ will be continuous and bounded functions of the entire one-sided sequence $\{w_{-\infty}^0\}$. And, therefore, $g^0(w_{-\infty}'^0)$, $\{n^{-1} L_\delta^0\, w_n^{-1}\}_{n\geq 1}$ will all be continuous and bounded functions on their respective demains. As a result, the functional

$$T'(\mu) = \int g'(w_{-\infty}^0)\, \mu\, d(w_{-\infty}^0) \tag{16}$$

will be continuous, where

$$g'(w_{-\infty}^0) \triangleq g^0(w_{-\infty}'^0) \tag{17}$$

and the new rate will be determined by

$$T'(\mu_1) = E_{\mu_1}\,\{g'(w_{-\infty}^0)\} \ . \tag{18}$$

A simple approach to construct w_i' from w_i is by using Huber's operations on the marginal loglikelihood ratio which leads to fixed lower and upper bounds on $\ln(f_1(w_i)/f_0(w_i))$. These bounds can be mapped back to w_i space to obtain w_i'. Some ε' can be used as the design parameter. The advantage of this approach is that as $\varepsilon' \downarrow 0$, $w_i' \to w_i$, that is we return to the ideal case and also when μ_1 and μ_0 both are i.i.d., we return to the optimal robust operation.

For the breakdown point and the influence function we will use the following definitions which are same as the ones used in our previous studies Bansal et al (1987 a,b).

Suppose $\mu_{i,\zeta,z}$ denotes the measure induced by the nominal μ_i and an i.i.d. sequence of outliers occuring with frequency (probability ζ) and magnitude z.

Our measure of efficiency and credibility both is the quantity $E_{\mu_i}\{g'(w_{-\infty}^0)\}$ because $E_{\mu_1}\{g'(w_{-\infty}^0)\}$ determines the asymptotic rate with which our algorithm detects the change. In the absense of contamination $E_{\mu_1}\,\{g'(w_{-\infty}^0)\}$ is positive and $E_{\mu_0}\{g'(w_{-\infty}^0)\}$ is negative. In the presence of strong contaminating measure either or both may reverse their sign and then our algorithm becomes useless. So it's of interest to find the largest percentage of outliers (for example) that our algorithm can withstand. Formally,

(Definition) Breakdown point is the largest frequency ζ of outliers such that $E_{\mu_{i,\zeta,z}}\{g'(w_{-\infty}^0)\}$ still retains its nominal algebraic sign for i=0,1. Here z will be chosen such that it leads to worst case or earliest breakdown.

Next, the influence function measures the normalized influence of a single observation at a particular value z on the quantities of interest, which are again $E_{\mu_{i,\zeta,z}}\,\{g'(w_{-\infty}^0)\}$, i=0,1. Formally,

(Definition) The influence function $IF_{\mu_i}(z)$, is

$$I\,F_{\mu_i}(z) = \lim_{\zeta\to 0} \frac{E_{\mu_{i,\zeta,z}}\,\{g'(w_{-\infty}^0)\} - E_{\mu_i}\,\{g'(w_{-\infty}^0)\}}{\zeta} \tag{19}$$

III. An Example

$$\begin{aligned} \text{Let } \mu_0 &: U_n - a\,U_{n-1} \\ \mu_1 &: U_n - a\,U_{n-1} + \theta \quad (\theta>0) \ . \end{aligned} \tag{20}$$

$U_n \sim N\,(0,1)$ and $0 < a$. $N\,(\cdot,\cdot)$ referes to Gaussian distribution. Alternatively,

$$\mu_0 : W_n = -\sum_{1}^{\infty} a^i\, W_{n-i} + U_n$$

$$\mu_1 : W_n = -\sum_1^{\infty} a^i W_{n-i} + V_n \tag{21}$$

where $V_n \sim N(\theta/1-a, 1)$.

Also, $W_n \sim N(0, 1+a^2)$ under μ_o and $W_n \sim N(\theta, 1+a^2)$ under μ_1. Note that μ_0, μ_1 both have infinite memory.

$$\ln \frac{f_1(w_n)}{f_0(w_n)} = \frac{\theta}{1+a^2} [w_n - \frac{\theta}{2}] \tag{22}$$

For a given ε', Huber's operation transforms the likelihood ratio as shown in the picture.

From Fig. 1 we obtain the following description of w'_n

$$w'(w) \begin{cases} = w_n \text{ on } [-d(1+a^2)/\theta + \frac{\theta}{2}, d(1+a^2)/\theta + \frac{\theta}{2} \\ = w'_m = -d(1+a^2)/\theta + \theta/2 \text{ on } w_n < -d(1+a^2)/\theta + \frac{\theta}{2} \\ = w'_m = d(1+a^2)/\theta + \frac{\theta}{2} \text{ on } w_n > d(1+a^2)/\theta + \frac{\theta}{2} \end{cases}$$

Thus

$$g'(w^0_{-\infty}) \stackrel{\Delta}{=} g^0(w^0_{-\infty}) \tag{23}$$

$$= \frac{\theta}{1-a} \left\{ w'_0 - \left[-\sum_1^{\infty} a^i w'_{-i} + \frac{\theta}{2(1-a)} \right] \right\} \tag{24}$$

Therefore

$$E_{\mu_1} g'(w^0_{-\infty}) = \frac{\theta}{(1-a^2)} \left\{ E_{\mu_1} w' - \frac{\theta}{2} \right\} . \tag{25}$$

Recalling that $\mu_{i,\zeta,z}$ denotes the measure induced by the nominal μ_i and i.i.d. sequence of outliers occuring with frequency (probability ζ) and magnitude z, we have

$$E_{\mu_{i,\zeta,z}} \{w'_0\} = (1-\zeta) E_{\mu_i} w'_0 + \zeta w'(z) .$$

Since $E_{\mu_1} g'(w^0_{-\infty}) > 0$, breakdown in (25) will occur due to negative extreme outliers, when

$$E_{\mu_{1,\zeta,-\infty}} \{g'(w'_{-\infty})\} = \frac{\theta}{(1-a)^2} \left[E_{\mu_{1,\zeta,-\infty}} \{w'_0\} - \frac{\theta}{2} \right] \le 0 \tag{26}$$

$$\text{or } (1-\zeta) E_{\mu_1} w'_0 + \zeta w'(-\infty) - \frac{\theta}{2} \le 0$$

$$\text{or } \zeta \ge \zeta^* = 1 / \left[1 + \frac{(1+a^2)}{\theta} \left\{ \left[E_{\mu_1} w'_0 - \frac{\theta}{2} \right] d(\varepsilon')^{-1} \right\} \right] . \tag{27}$$

Similarly breakdown can occur when

$$E_{\mu_{0,\zeta,z}} \{-g'(w^0_{-\infty})\} \ge 0$$

which gives the breakdown at

$$\zeta \ge \zeta^{**} = 1 / \left\{ 1 + d(1+a^2)/\theta (\frac{\theta}{2} - E_{\mu_0} w'_0) \right\} . \tag{28}$$

Therefore, the overall breakdown point

$$\bar{\zeta} = \min(\zeta^*, \zeta^{**}) \tag{29}$$

Comments

(1) As $\varepsilon' \downarrow 0$, $d(\varepsilon') \uparrow \infty$, $w'_n \to w_n$

$$E_{\mu_1} g'(w^0_{-\infty}) \to E_{\mu_1} g^0(w^0_{-\infty}) = \frac{\theta^2}{(1-a)^2} .$$

(2) As $\varepsilon' \uparrow$ (to it's maximum allowable value), which may be less than 1/2,

$$\bar{\zeta} \to \frac{2\Phi(\theta/2)-1}{2\Phi(\frac{\theta}{2})} . \tag{30}$$

And then if $\theta \uparrow \infty$, $\bar{\zeta} \to 1/2$, the maximum achievable breakdown point.

(3) If we use the memoryless robust algorithm, that is in (14) if we replace $g_i^0(w_i^i)$ by $\ln \frac{f_1}{f_0}(w_n)|_{w_n = w'_n}$ then again we have same breakdown point. But our algorithm is more efficient because

$$E_{\mu_1} \left\{ \ln \frac{f_1}{f_0}(w_n)|_{w_n = w'_0} \right\} = \frac{\theta}{(1+a^2)} \left[E_{\mu_1} w'_n - \frac{\theta}{2} \right]$$

$$\le \frac{\theta}{(1-a)^2} \left[E_{\mu_1} w'_n - \frac{\theta}{2} \right]$$

$$= E_{\mu_1} \{g'(w^0_{-\infty})\} \tag{31}$$

and $1+a^2 \ge (1-a)^2 = 1+a^2 - 2a$ for $a > 0$. However if $a < 0$, then one should use the memoryless algorithm for higher efficiency for the same breakdown point.

The influence function

From the definition in (19),

$$IF_{\mu_i}(z) = \lim_{\zeta \to 0} \frac{\theta}{(1-a)^2} \left[(1-\zeta) E_{\mu_i} w'_0 + \zeta w'_0(z) - \frac{\theta}{2} - \left[E_{\mu_i} w'_0 - \frac{\theta}{2} \right] \right] / \zeta$$

$$= \frac{\theta}{(1-a)^2} (w'_0(z) - E_{\mu_i} w'_0) .$$

Since $|w'_0(z)| \le \max(|w'_m|, w'_M)$ and $w'_0(z)$ is continuous function of z, $IF_{\mu_i}(z)$ is continuous and bounded.

Because of space limitations, we can not present another example, the numerical results and a comparison of the suggested algorithm with the one studied for Markovian situation as it would be meaningful to apply the suggested algorithm to Markov processes themselves.

IV. Conclusion

We have discussed the issue of robustness in time series from an abstract point of view and pointed out the general failure of operations designed to be robust under i.i.d. and Markovian set up. A stronger notion of continuity of the point function of $\psi(x)$ was needed to achieve robustness of operations which make use of entire distribution of the process. A simple technique based on Huber's approach was used to obtain the pseudo observations which replaced the true ones. One could benefit by using higher order densities to achieve the same goal. This of course needs further study as one can numerically (if not analytically) optimize with respect to the order itself.

Another approach for designing robust operations would be to put a bound on breakdown point and maximize the efficiency or vice versa.

Ideally one would like to extend the minimax optimality results under i.i.d. set up to non i.i.d. set up. This seems to be a formidable task. One should perhaps start from Markovian set up first.

References

[1] Bansal, R. K., and Papantoni-Kazakos, P. (1986), "An Algorithm for Detecting a Change in a Stochastic Processes," IEEE Trans. Inf. Th., IT-32, 2, 227-235.

[2] Bansal, R. K., and Papantoni-Kazakos, P. (1987a), "Outlier Resistant Algorithms for Detecting a Change in a Stochastic Process," submitted to IEEE Trans. Inf. Th.

[3] Bansal, R. K., and Papantoni-Kazakos, P. (1987b), "Outlier Resistant Algorithms for Detecting a Change in a Stochastic Process," Proceedings of 1987 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore.

[4] Boente Graciela, Fraiman Ricardo and Yohai, Victor J. (1987) "Qualitative Robustness for Stochastic Processes," Annals of Statistics, vol. 15, no. 3, 1293-1312.

[5] Hampel, F. R. (1971), "A General Qualitative Definition of Robustness," Annals of Math. Statistics, vol. 42, 1887-1896.

[6] Huber, P. J. (1965), "A Robust version of the Probability Ratio Test," Annals of Math. Statistics, vol 36, 1753-1758.

[7] Lorden, G. (1971), "Procedures for Reacting to a Change in Distribution," Annals of Math. Statistics, vol. 42, 1897-1908.

[8] Martin, R. Douglas and Yohai Victor J. (1986)," Influence Functionals for Time Series," Annals of Statistics, vol. 14, no. 3, 781-855.

[9] Page, E. S. (1954), "Continuous Inspection Schemes," Biometrika, vol. 41, 100-115.

[10] Papantoni-Kazakos, P. (1987), "Qualitative Robustness in Time Series," Information and Computation, vol 72, 239-269.

[11] Roberts, S. W. (1966), "A Comparison of some Control Chart Procedures," Technometrics, vol. 8, 411-430.

[12] Shiryayev, A. N. (1963), "On Optimal Methods in Earliest Detection Problems," Theory of Probability and its Applications, vol. 8, 26-51.

DISTRIBUTION LIST

Copy No.	
1 - 6	Air Force Office of Scientific Research Building 410 Bolling Air Force Base Washington, DC 20332-6448 Attention: Major Brian W. Woodruff, NM
7 - 8	Dr. P. Papantoni-Kazakos, EE
9 - 10	Ms. E. H. Pancake, Clark Hall
11	SEAS Publications Files

1195:ald:DW211R

UNIVERSITY OF VIRGINIA
School of Engineering and Applied Science

The University of Virginia's School of Engineering and Applied Science has an undergraduate enrollment of approximately 1,500 students with a graduate enrollment of approximately 560. There are 150 faculty members, a majority of whom conduct research in addition to teaching.

Research is a vital part of the educational program and interests parallel academic specialties. These range from the classical engineering disciplines of Chemical, Civil, Electrical, and Mechanical and Aerospace to newer, more specialized fields of Biomedical Engineering, Systems Engineering, Materials Science, Nuclear Engineering and Engineering Physics, Applied Mathematics and Computer Science. Within these disciplines there are well equipped laboratories for conducting highly specialized research. All departments offer the doctorate; Biomedical and Materials Science grant only graduate degrees. In addition, courses in the humanities are offered within the School.

The University of Virginia (which includes approximately 2,000 faculty and a total of full-time student enrollment of about 16,400), also offers professional degrees under the schools of Architecture, Law, Medicine, Nursing, Commerce, Business Administration, and Education. In addition, the College of Arts and Sciences houses departments of Mathematics, Physics, Chemistry and others relevant to the engineering research program. The School of Engineering and Applied Science is an integral part of this University community which provides opportunities for interdisciplinary work in pursuit of the basic goals of education, research, and public service.