

## A Simple Recursive Algorithm for Diagnosis of Abrupt Changes in Random Signals

Igor V. Nikiforov

**Abstract**—We address the problem of detecting and isolating abrupt changes in random signals. An asymptotic optimal solution to this problem, which has been proposed in previous works, involve the number of computations at time  $t$  which grows to infinity with  $t$ . In this correspondence, we propose another more realistic criterion, establish a new simple recursive change detection/isolation algorithm, and investigate its statistical properties.

**Index Terms**—Kullback–Leibler information, minimax detection, sequential detection and isolation, sequential decision procedures, signal detection.

### I. INTRODUCTION

This correspondence treats the problem of *abrupt change diagnosis* (detection and isolation) in random signals. An optimal solution to this problem was obtained in [5]–[7]. The *novelty* of this work with respect to the previous papers is some *practical* aspects of the proposed theory: a more realistic criterion of optimality and a simple recursive solution. First, we minimize now the supremum of the mean delay for detection/isolation over the change time  $t_0$  instead of minimizing this supremum over the past “trajectory”  $X_1, \dots, X_{t_0-1}$  of stochastic process and  $t_0$  together. Second, in the previous papers we fixed *a priori* the change time  $t_0 = 1$  in the definition of the probability of false isolation to simplify theoretical difficulties. In practice, it is difficult to justify this assumption, for this reason we examine now the supremum of the error probability over  $t_0 \geq 1$ . Next, the algorithms developed in [5]–[7] involve the number of the likelihood ratio (LR) computations at time  $t$  which grows to infinity with  $t$ . Now we design a *simple recursive* algorithm which involves one LR computation at every stage. The correspondence is organized as follows. First, we state the problem in Section II. Next, we discuss the design of the recursive change diagnosis algorithm and its statistical properties in Section III. The main results are established in Theorems 1 and 2. Finally, we compare in this section the theoretical formulas and the results of Monte Carlo simulations.

### II. PROBLEM STATEMENT

#### A. Model with Abrupt Changes

We consider a finite family of distributions

$$\mathcal{P} = \{P_l, l = 0, \dots, K-1\}$$

with densities  $\{p_l, l = 0, \dots, K-1\}$ . Let  $(X_t)_{t \geq 1}$  be an independent random sequence observed *sequentially*

$$\mathcal{L}(X_t) = \begin{cases} P_0, & \text{if } t < t_0 \\ P_l, & \text{if } t \geq t_0 \end{cases} \quad (1)$$

where  $l = 1, \dots, K-1$ , and  $\mathcal{L}(\cdot)$  is the probability law. The change

time  $t_0$  and number  $l$  are unknown (but non random). Let

$$Z_t(l, j) = \log(p_{\theta_l}(X_t)/p_{\theta_j}(X_t))$$

be the log LR between hypotheses  $\mathcal{H}_l : P = P_l$  and  $\mathcal{H}_j : P = P_j$ .

#### B. Criterion of Optimality

The change detection/isolation algorithm has to compute a pair  $(N, \nu)$  based on the observations  $X_1, X_2, \dots$ , where  $N$  is the *alarm time* at which a  $\nu$ -type change is detected/isolated and  $\nu, \nu = 1, \dots, K-1$  is the *final decision*. Let  $P_{t_0}^l$  be the distribution of the observations  $X_1, X_2, \dots, X_{t_0}, X_{t_0+1}, \dots$  when  $t_0 = 1, 2, \dots$  and  $X_{t_0}$  is the first observation with distribution  $P_l$ . In previous papers [5]–[7] we minimized the “worst case” mean delay for detection/isolation

$$\bar{\tau}_l^* = \sup_{t_0 \geq 1} \text{esssup} E_{t_0}^l(N - t_0 + 1 | N \geq t_0, X_1, \dots, X_{t_0-1}).$$

Now we propose to measure the speed of detection/isolation with the aid of the maximum mean delay for detection/isolation

$$\bar{\tau} = \max_{1 \leq l \leq K-1} \bar{\tau}_l, \quad \bar{\tau}_l = \sup_{t_0 \geq 1} E_{t_0}^l(N - t_0 + 1 | N \geq t_0). \quad (2)$$

In the case of change detection this performance index is discussed in [8], [4]. We measure the levels of false alarms and false isolations by using the following equation:

$$\begin{aligned} \min_{1 \leq j \leq K-1} E_0 \left( \inf_{k \geq 1} \{N(k) : \nu(k) = j\} \right) &= \gamma \\ \max_{1 \leq l \leq K-1} \max_{1 \leq j \neq l \leq K-1} \sup_{t_0 \geq 1} \beta_{t_0}(j, l) &= \beta \end{aligned} \quad (3)$$

where  $\beta_{t_0}(j, l) = P_{t_0}^l(\nu = j \neq l | N \geq t_0)$  is the error probability,  $\gamma$  is the minimum of the mean times before a false alarm, and  $\beta$  is the maximum of the probability of a false isolation. In the above criterion we suppose that there exists a sequence  $N(1), N(2), \dots$  of false alarms. The first false alarm of a  $j$ -type is defined by  $\inf_{k \geq 1} \{N(k) : \nu(k) = j\}$ .<sup>1</sup> In brief, we require that the maximum mean detection/isolation delay given by (2) should be *as small as possible* subject to the constraints given by (3). In this correspondence, we will discuss the asymptotic case when  $\gamma \rightarrow \infty, \beta \rightarrow 0$ , such that  $\log \gamma \gtrsim \log \beta^{-1}$ .

### III. RECURSIVE ALGORITHM AND ITS STATISTICAL PROPERTIES

#### A. Recursive Algorithm

We denote a pair *alarm time–final decision* for the recursive algorithm by  $(N_r, \nu_r)$ , where

$$\begin{aligned} N_r &= \min \{N_r^1, \dots, N_r^{K-1}\} \\ \nu_r &= \arg \min \{N_r^1, \dots, N_r^{K-1}\}. \end{aligned} \quad (4)$$

We define the stopping time  $N_r^l$  in the following recursive manner

$$N_r^l = \inf \left\{ t \geq 1 : \min_{0 \leq j \neq l \leq K-1} [g_t(l, 0) - g_t(j, 0) - h_{t,j}] \geq 0 \right\} \quad (5)$$

<sup>1</sup>Naturally, we assume that after a false alarm the observation process restarts immediately from scratch.

Manuscript received March 26, 1998; revised February 22, 2000.

The author is with Université de Technologie de Troyes, LM2S, BP 2060, 10010 Troyes Cedex, France (e-mail: nikiforov@univ-troyes.fr).

Communicated by P. Moulin, Associate Editor for Nonparametric Estimation, Classification, and Neural Networks.

Publisher Item Identifier S 0018-9448(00)09659-0.

where the recursive decision functions  $g_t(l, 0)$  are defined by

$$g_t(l, 0) = (g_{t-1}(l, 0) + Z_t(l, 0))^+, \quad x^+ = \max(0, x),$$

$$g_0(l, 0) = 0, \quad l = 1, 2, \dots, K-1 \quad (6)$$

and  $g_t(0, 0) \equiv 0$ . The thresholds  $h_{l,j}$  are chosen by the following formula:

$$h_{l,j} = \begin{cases} h_d, & \text{if } l = 1, \dots, K-1 \quad \text{and } j = 0 \\ h_i, & \text{if } j, l = 1, \dots, K-1 \quad \text{and } j \neq l \end{cases} \quad (7)$$

where  $h_d$  is the detection threshold and  $h_i$  is the isolation threshold.

### B. Discussion

Let us compare the design of the recursive rule (4)–(6) with the nonrecursive one. We start with the nonrecursive rule  $(\bar{N}_n, \nu_n)$  [5], [6]. If for some  $k \leq t$  the observations  $X_k, \dots, X_t$  are such that all the LR between  $\mathcal{H}_l$  and  $\mathcal{H}_j, 0 \leq j \neq l \leq K-1$  are greater than or equal to the thresholds  $h_{l,j}$

$$S_k^t(l, 0) = \sum_{i=k}^t Z_i(l, 0) \geq h_{l,0}$$

$$S_k^t(l, 1) = \sum_{i=k}^t Z_i(l, 1) \geq h_{l,1}, \dots$$

$$S_k^t(l, K-1) = \sum_{i=k}^t Z_i(l, K-1) \geq h_{l,K-1}$$

(i.e., the observations are *significant* for accepting the hypothesis  $\mathcal{H}_l$  with respect to this set of alternatives) then the nonrecursive rule stops the observation process at time  $t$  ( $N_n = t$ ) and the final decision is  $\nu_n = l$ . In practice, this nonrecursive algorithm has two disadvantages. First, sometimes (it depends on the mutual “geometry” of the hypotheses) the probability of false isolation seriously increases when  $t_0 \rightarrow \infty$ . It occurs due to an uncontrolled growth of some cumulative sums  $S_1^t(m, j)$  when  $X_1, \dots, X_t \sim P_0$ .<sup>2</sup> Second, the nonrecursive algorithm cannot be rewritten directly in a recursive manner and the number of the LR computations at time  $t$  grows to infinity with  $t$ . Unlike this nonrecursive algorithm, the rule  $N_r^l$  (5) is based on the recursive decision functions  $g_t(l, 0)$ . It is easy to see that the recursive algorithm (4)–(6) is nothing but  $K-1$  parallel CUSUM tests (see the decision functions  $g(1, 0), \dots, g(l, 0)$ ) plus a simple logical rule which compares  $g_t(l, 0) - g_t(j, 0)$  with the thresholds  $h_{l,j}$ . Before the change time  $t_0$ , the nonnegative functions  $g_t(l, 0)$  are stochastically small (because  $E_0(Z_t(l, 0)) < 0$ ) and, hence, only an insignificant growth of the probability of false isolation takes place when  $t_0 > 1$ . Let us note also that  $S_{t_0}^t(l, j) \simeq g_t(l, 0) - g_t(j, 0)$  when  $t \gg t_0$  and  $E_l(g_t(j, 0)) \geq 0$ . Therefore, both algorithms extract approximately the same information from the observations  $X_{t_0}, \dots, X_t$ . Nevertheless, if  $E_l(g_t(j, 0)) < 0$  then the recursive algorithm partly loses the information from these observations. In order to fix this gap we solve the detection/isolation problem under the constraint  $h_d \geq h_i$  (see details in Theorems 1 and 2 and Appendix I).

### C. Statistical Properties

Let us consider now the recursive detection/isolation algorithm  $(N_r, \nu_r)$  (4)–(6). We start with the mean detection/isolation delay  $\bar{\tau}$ :

<sup>2</sup>The introduction of the “window-limited scheme” (see [4], [11]), where  $t - n + 1 \leq k \leq t$  and  $n$  is a tuning parameter, only partially improves the situation.

**Theorem 1:** Let  $(N_r, \nu_r)$  be the test (4)–(6). Suppose that  $0 < \rho_{l,j} = E_l(Z_t(l, j)) < \infty$  for all  $0 \leq l \neq j \leq K-1$  and the following regularity condition is fulfilled: the moment-generating function (m.g.f.)  $\varphi(\zeta) = E_l(e^{\zeta Z_t(l, j)}) < \infty$  exists for all real  $\zeta \in ]-\eta; \eta[$ , where  $\eta > 0$ , and for all  $1 \leq l \leq K-1$  and  $0 \leq j \neq l \leq K-1$ . Let  $h_{l,j}$  be given by (7) and  $h_d \geq h_i$ . Then

$$\bar{\tau} \lesssim \max\left(\frac{h_d}{\rho_{l,0}}, \frac{h_i}{\min_{j \neq 0, l} \rho_{l,j}}\right) \quad \bar{\tau} \lesssim \max\left(\frac{h_d}{\rho_d^*}, \frac{h_i}{\rho_i^*}\right) \quad (8)$$

as  $h_i \rightarrow \infty$ , where

$$\rho_d^* = \min_{1 \leq j \leq K-1} \rho_{j,0}, \quad \rho_i^* = \min_{1 \leq l \leq K-1} \min_{1 \leq j \neq l \leq K-1} \rho_{l,j}.$$

*Proof of Theorem 1:* See Appendix I.

Let us discuss now the probability of false isolation. From Theorem 1 it follows that the delay for detection  $\bar{\tau}_l$  is mainly defined by the stopping time  $N_r^l$  when the hypothesis  $\mathcal{H}_l$  is true. The false isolation  $\nu_r = j$  means that due to the noise  $N_r^j < \bar{N}_r^j = \min_{i \neq j} \{N_r^i\}$  given  $N_r \geq t_0$ . Naturally, this is a rare event. Roughly speaking, to estimate  $\beta_{t_0}(j, l)$  we have to compute the conditional probability to stop the observation process by the “false” stopping time  $N_r^j$  before then it will be stopped by the “true” stopping time  $N_r^l$  given  $N_r \geq t_0$ . Therefore, we prove now an asymptotic upper bound for  $\sup_{t_0 \geq 1} \beta_{t_0}(j, l)$ . The result is stated in the following theorem.

**Theorem 2:** Let  $(N_r, \nu_r)$  be the test (4)–(6) and let the conditions of Theorem 1 be satisfied. Then

$$\beta = \max_{1 \leq l \leq K-1} \max_{1 \leq j \neq l \leq K-1} \sup_{t_0 \geq 1} \beta_{t_0}(j, l)$$

$$\lesssim \max_{1 \leq l \leq K-1} \max_{1 \leq j \neq l \leq K-1} e^{-h_i}$$

$$\times \left\{ \max\left(\frac{h_d}{\rho_{l,0}}, \frac{h_i}{\min_{j \neq 0, l} \rho_{l,j}}\right) + h_i \right\}$$

$$\lesssim e^{-h_i} \left\{ \max\left(\frac{h_d}{\rho_d^*}, \frac{h_i}{\rho_i^*}\right) + h_i \right\}, \quad \text{as } h_i \rightarrow \infty. \quad (9)$$

*Proof of Theorem 2:* See Appendix II.

From Theorem 2 it results that if  $t_0 = 1$ , then the probability of false isolation is

$$\beta_1 = \max_{1 \leq l \leq K-1} \max_{1 \leq j \neq l \leq K-1} \beta_1(j, l) \lesssim e^{-h_i} \max\left(\frac{h_d}{\rho_d^*}, \frac{h_i}{\rho_i^*}\right)$$

as  $h_d \geq h_i$  and  $h_i \rightarrow \infty$ . Hence, at least upper bounds for  $\beta_1$  and  $\beta$  are close to each other. As a result, from [5] and (5), the mean time before a false alarm for algorithm (4)–(6) is given by

$$E_0(\inf_{k \geq 1} \{N_r(k) : \nu_r(k) = j\}) \geq e^{h_d}, \quad \text{for } j = 1, \dots, K-1.$$

The analysis of Theorem 2 [5] and Theorem 4 [6] shows that the asymptotic equation

$$n(\gamma, \beta) \sim \max\left\{\frac{\log \gamma}{\rho_d^*}, \frac{\log \beta^{-1}}{\rho_i^*}\right\}$$

for the infimum  $n(\gamma, \beta)$  of  $\bar{\tau}^*$  still hold with the mean delay  $\bar{\tau}$  instead of  $\bar{\tau}^*$  and the new definition of the class of tests (3).

*Corollary 1:* The recursive detection/isolation algorithm  $(N_r, \nu_r)$  (4)–(6) is asymptotically equivalent to the optimal (nonrecursive) test [5], [6] when  $t_0 = 1$

$$\bar{\tau} \sim n(\gamma, \beta) \sim \max \left\{ \frac{\log \gamma}{\rho_d^*}, \frac{\log \beta^{-1}}{\rho_i^*} \right\},$$

$$\text{as } \gamma \rightarrow \infty, \beta \rightarrow 0, \log \gamma \gtrsim \log \beta^{-1}. \quad (10)$$

The above theoretical results have been obtained by ignoring the “excess” of the decision function  $\hat{g}_t(l, j) = g_t(l, 0) - g_t(j, 0)$  over the boundary  $h_{l,j}$ , namely, the quantity

$$\epsilon = (\hat{g}_t(l, j) - h_{l,j} | \hat{g}_t(l, j) \geq h_{l,j}).$$

It is well known that this *Wald's* approximation is not very accurate (see details in [1], [3], [9]). For this reason, we suggest to use the *corrected terms* proposed by Siegmund [9, Ch. X] for a possibly improved approximation of the mean time before a false alarm and the probability of false isolation. The idea of this approximation is to replace the threshold  $h_d$  by  $h_d + \varrho_d$  and  $h_i$  by  $h_i + \varrho_i$ , where  $\varrho_d$  and  $\varrho_i$  are positive constants. Hence, we get the following *corrected* asymptotic equations:<sup>3</sup>

$$E_0 \left( \inf_{k \geq 1} \{N_r(k) : \nu_r(k) = j\} \right) \geq e^{h_d + \varrho_d}$$

$$\sup_{t_0 \geq 1} \beta_{t_0}(j, l) \lesssim e^{-h_i - \varrho_i} \left\{ \max \left( \frac{h_d}{\rho_{l,0}}, \frac{h_i}{\min_{j \neq 0:l} \rho_{l,j}} \right) + h_i \right\}. \quad (11)$$

#### D. Example

The goal of this example is to compare the statistical properties of the recursive (4)–(6) and nonrecursive [5], [6] rules using Monte Carlo simulation and to compare the results of the simulation with (8) and (9), (11). Let  $X \in \mathbb{R}^2$  be a Gaussian vector,  $\mathcal{L}(X) = \mathcal{N}(\theta, I)$ . We consider the following hypotheses:

$$\mathcal{H}_0 : \{\theta_0 = (0, 0)^T\}$$

$$\mathcal{H}_1 : \{\theta_1 = (1, 0)^T\}$$

$$\mathcal{H}_2(i) : \{\theta_2(i) = (\vartheta_{21}(i), \vartheta_{22}(i))^T\}$$

where  $i = \overline{1, 5}$ . The values of  $\theta_2(i)$  are given in Table I. The goal of the first simulation is to detect/isolate the change from  $\mathcal{H}_0$  to  $\mathcal{H}_1$  when  $\mathcal{H}_2(i)$  varies between  $\mathcal{H}_2(1)$  and  $\mathcal{H}_2(5)$ . The second simulation experiment is devoted to the detection/isolation of the change from  $\mathcal{H}_0$  to  $\mathcal{H}_2(i)$  when  $\mathcal{H}_2(i)$  varies between  $\mathcal{H}_2(1)$  and  $\mathcal{H}_2(5)$ . In this manner we consider different combinations of the hypotheses  $\mathcal{H}_0$ ,  $\mathcal{H}_1$ , and  $\mathcal{H}_2(i)$  mutual geometry. The thresholds are  $h_d = h_i = 5$ . The results are given in Tables II and III. Each point in these tables is based on  $10^7$  simulations. The detection of the change from  $\mathcal{H}_0$  to  $\mathcal{H}_1$  (see Tables II) shows that the statistical characteristics of both tests (recursive and nonrecursive) are comparable. Nevertheless, in the case of the changes from  $\mathcal{H}_0$  to  $\mathcal{H}_2(1)$  and from  $\mathcal{H}_0$  to  $\mathcal{H}_2(2)$  (see the fifth and sixth rows in Tables III), the nonrecursive algorithm makes many false isolations when  $t_0 = 10$ , due to an uncontrolled growth of the cumulative sum  $S_1^t(1, 2)$ .

The first simulation (see Table II) shows a relatively good accuracy of the asymptotic mean detection/isolation delay  $\bar{\tau}_1$  (8). In the second case (see Table III) accuracy is lower because the true values of the mean detection/isolation delay are small, and, hence, the approximation cannot be considered as asymptotic. It follows from the

<sup>3</sup>It is worth noting that: i) this suggestion is heuristic; ii) the tightness of the bounds (11) is defined by several factors (not only by the “excess” errors), therefore, only partial improvement of the accuracy of the corrected equations can be expected.

TABLE I  
THE HYPOTHESES  $\mathcal{H}_2(i)$

$i$	1	2	3	4	5
$\vartheta_{21}(i)$	3	2.121	0	-2.121	-3
$\vartheta_{22}(i)$	0	2.121	3	2.121	0

proof of Theorem 2 that the proposed asymptotic upper bound (9) for the probability of false isolation  $\beta$  cannot be fairly tight. The simulation shows that the corrected term  $\varrho_i = \varrho_+ \|\theta_1 - \theta_2\|$ , where  $\varrho_+ \simeq 0.583$  [9, Ch. X], suggested in (11), really improves this approximation of  $\beta$  in this case. The simulation confirms the results of Theorem 2 for  $\sup_{t_0 \geq 1} \beta_{t_0}(j, l)$ . Let us analyze the growth of the probability of false isolation  $\beta_{t_0}$  when  $t_0 \rightarrow \infty$  for the recursive test. It follows from Theorem 2 that this growth should be more significant when  $\max(h_d/\rho_{l,0}, h_i/\min_{j \neq 0:l} \rho_{l,j})$  is small in comparison with  $h_i$ . It happens in the case of the change from  $\mathcal{H}_0$  to  $\mathcal{H}_2(i)$  (see the fifth, sixth, and seventh rows in Table III).

#### IV. CONCLUSION

A simple recursive algorithm for diagnosis of abrupt changes in random signals was proposed and its statistical performance was investigated. An important feature of this algorithm is its ability to warrant an acceptable level of false isolation when  $t_0 > 1$ . Another attractive feature of this algorithm is the fact that it is based on  $K - 1$  parallel CUSUM rules. The CUSUM rules are well known in signal processing, automatic control (fault detection), and industrial quality control. Hence, this fact simplifies the implementation of the proposed algorithm.

#### APPENDIX I PROOF OF THEOREM 1

We suppose that the thresholds  $h_{l,j}$  are given by (7), where  $h_d \geq h_i$ . Let  $0 < \delta < 1$  and let  $n_h$  be the smallest integer

$$\geq \max \left\{ \frac{h_d}{\rho_{1,0}}, \frac{h_i}{\min_{j \neq 0:1} \rho_{1,j}} \right\} (1 - \delta)^{-1}.$$

We assume that  $h_d$  and  $h_i$  are so chosen that  $n_h > 0$ . Let  $l = 1$  (without any loss of generality) and we consider the probability

$$\mathbf{P}_{t_0}^1(N_r^1 - t_0 + 1 > M | N_r \geq t_0)$$

where  $M = mn_h$  and  $m$  is a positive integer number. Taking into account that

$$g_t(j, 0) = \max_{0 \leq k \leq t} \sum_{i=k+1}^t Z_i(j, 0)$$

and  $\sum_{i=k}^n \dots = 0$  when  $n < k$ , we get from (5)

$$\begin{aligned} & \mathbf{P}_{t_0}^1(N_r^1 - t_0 + 1 > M | N_r \geq t_0) \\ & \leq \mathbf{P}_{t_0}^1 \left\{ \min_{j \neq 1} \left( \max_{0 \leq k \leq t_0 - 1 + M} \sum_{i=k+1}^{t_0 - 1 + M} Z_i(1, 0) \right. \right. \\ & \quad \left. \left. - \max_{0 \leq k \leq t_0 - 1 + M} \sum_{i=k+1}^{t_0 - 1 + M} Z_i(j, 0) - h_{1,j} \right) < 0 | N_r \geq t_0 \right\}. \end{aligned} \quad (12)$$

Since the functions  $g_t(j, 0)$  ( $1 \leq j \leq K - 1$ ) are Markov sequences, it then follows that the behavior of the function  $g_t(j, 0)$  when  $t \geq t_0$  depends on  $g_{t_0-1}(j, 0)$  and the observations  $X_{t_0}, X_{t_0+1}, \dots$ . To simplify

TABLE II  
THE CHANGE FROM  $\mathcal{H}_0$  TO  $\mathcal{H}_1$  WHEN THE THIRD HYPOTHESIS IS  $\mathcal{H}_2(i)$ ,  $i = \overline{1, 5}$

$\mathcal{H}_2(i)$	mean delay $\bar{\tau}_1$						probability of false isolation $\beta(2, 1)$							
	$(N_r, \nu_r)$			$(N_n, \nu_n)$			$(N_r, \nu_r)$			$(N_n, \nu_n)$			Eq. (9)	Eq. (11)
	$t_0$			$t_0$			$t_0$			$t_0$				
	1	10	50	1	10	Eq. (8)	1	10	50	1	10			
1	12.9	12.1	12	10.9	9.9	10	$6.7 \cdot 10^{-3}$	$5.8 \cdot 10^{-3}$	$5.7 \cdot 10^{-3}$	$8 \cdot 10^{-3}$	$2 \cdot 10^{-3}$	$10^{-1}$	$3.2 \cdot 10^{-2}$	
2	11.4	10.7	10.6	10.5	9.7	10	$5.3 \cdot 10^{-3}$	$4.6 \cdot 10^{-3}$	$4.4 \cdot 10^{-3}$	$8.9 \cdot 10^{-3}$	$7 \cdot 10^{-3}$	$10^{-1}$	$2.6 \cdot 10^{-2}$	
3	10.5	9.8	9.7	10.4	9.7	10	$2.5 \cdot 10^{-3}$	$2.3 \cdot 10^{-3}$	$2.2 \cdot 10^{-3}$	$6.4 \cdot 10^{-3}$	$6.1 \cdot 10^{-3}$	$10^{-1}$	$1.7 \cdot 10^{-2}$	
4	10.4	9.7	9.6	10.4	9.7	10	$4.1 \cdot 10^{-4}$	$4.1 \cdot 10^{-4}$	$4.4 \cdot 10^{-4}$	$5.7 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$	$10^{-1}$	$1.1 \cdot 10^{-2}$	
5	10.4	9.7	9.6	10.4	9.7	10	$1.1 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$1.4 \cdot 10^{-4}$	$1.3 \cdot 10^{-4}$	$10^{-1}$	$9.7 \cdot 10^{-3}$	

TABLE III  
THE CHANGE FROM  $\mathcal{H}_0$  TO  $\mathcal{H}_2(i)$ ,  $i = \overline{1, 5}$ , WHEN THE THIRD HYPOTHESIS IS  $\mathcal{H}_1$

$\mathcal{H}_2(i)$	mean delay $\bar{\tau}_2$						probability of false isolation $\beta(1, 2)$							
	$(N_r, \nu_r)$			$(N_n, \nu_n)$			$(N_r, \nu_r)$			$(N_n, \nu_n)$			Eq. (9)	Eq. (11)
	$t_0$			$t_0$			$t_0$			$t_0$				
	1	10	50	1	10	Eq. (8)	1	10	50	1	10			
1	3.3	3.5	3.5	3.2	2.7	2.5	$1.7 \cdot 10^{-4}$	$9.7 \cdot 10^{-4}$	$1.1 \cdot 10^{-3}$	$7.1 \cdot 10^{-4}$	0.71	$5 \cdot 10^{-2}$	$1.6 \cdot 10^{-2}$	
2	2.5	2.6	2.6	2.4	2.3	1.7	$6 \cdot 10^{-5}$	$5.7 \cdot 10^{-4}$	$6.4 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$1.7 \cdot 10^{-1}$	$4.5 \cdot 10^{-2}$	$1.3 \cdot 10^{-2}$	
3	1.8	1.9	1.8	1.8	1.8	1.1	$10^{-6}$	$9.8 \cdot 10^{-5}$	$1.1 \cdot 10^{-4}$	$2 \cdot 10^{-6}$	$1.9 \cdot 10^{-3}$	$4 \cdot 10^{-2}$	$8.3 \cdot 10^{-3}$	
4	1.8	1.8	1.7	1.8	1.8	1.1	0	$4 \cdot 10^{-6}$	$5 \cdot 10^{-6}$	0	$5 \cdot 10^{-6}$	$4 \cdot 10^{-2}$	$5.6 \cdot 10^{-3}$	
5	1.8	1.8	1.7	1.8	1.8	1.1	0	$10^{-7}$	$5 \cdot 10^{-7}$	0	$10^{-7}$	$4 \cdot 10^{-2}$	$4.9 \cdot 10^{-3}$	

the notations, we put  $t_0 = 1$  and start with the case of  $g_{t_0-1}(j, 0) = 0$ . Since  $\mathbf{P}(\bigcup_{j \neq 1} A_j) \leq \sum_{j \neq 1} \mathbf{P}(A_j)$ , we get from (12)

$$\begin{aligned}
 \mathbf{P}_1^1(N_r^1 > M) &\leq \sum_{j \neq 1} \mathbf{P}_1^1 \left( \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(1, 0) \right. \\
 &\quad \left. - \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(j, 0) < h_{1,j} \right) \\
 &\leq \sum_{j \neq 1} \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) \right. \\
 &\quad \left. - \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(j, 0) < h_{1,j} \right). \quad (13)
 \end{aligned}$$

Let us assume that  $E_1(Z_i(j, 0)) > 0$  and  $j \geq 2$  (we will discuss the case when  $E_1(Z_i(j, 0)) \leq 0$  and/or  $j = 0$  below). We consider the probability

$$\begin{aligned}
 p_j(m) &= \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) - \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(j, 0) < h_{1,j} \right) \\
 &= \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, j) + \xi_M < h_{1,j} \right) \quad (14)
 \end{aligned}$$

where

$$\xi_M = \min_{0 \leq k \leq M} \sum_{i=1}^k Z_i(j, 0).$$

It follows from the definition of  $n_h$  and (7) that  $h_{1,j} \leq n_h \rho_{1,j} (1 - \delta)$ . Let  $\tilde{Z}_i(1, j) = Z_i(1, j) - \rho_{1,j}$ , this yields

$$p_j(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, j) + \xi_M < n_h \rho_{1,j} (1 - \delta - m) \right). \quad (15)$$

Putting together the inequalities

$$n_h \rho_{1,j} (1 - \delta - m) \leq -(\rho_{1,j} M \delta / 2) - (\rho_{1,j} M \delta / 2)$$

and

$$\mathbf{P}_1^1(x + y < a) \leq \mathbf{P}_1^1(x < a/2) + \mathbf{P}_1^1(y < a/2)$$

with (15), we get

$$p_j(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, j) < -\frac{\rho_{1,j} M \delta}{2} \right) + \mathbf{P}_1^1 \left( \xi_M < -\frac{\rho_{1,j} M \delta}{2} \right). \quad (16)$$

The right side of (16) should be bounded above. Taking into account the regularity conditions of Theorem 1, we get by Chernoff's bound (see details in [10, Ch. IV] and ([3, Ch. 4])

$$\mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, j) < -\frac{\rho_{1,j} M \delta}{2} \right) \leq e^{-MH \tilde{Z}_i(1, j) \left( -\frac{\rho_{1,j} \delta}{2} \right)} \quad (17)$$

where  $H_x(y) = \sup_{\omega < 0} [y\omega - \log E(e^{\omega x})]$  is the Cramér transform of the distribution function of the random value  $x$  and  $H_{\tilde{Z}_i(1, j)} \left( -\frac{\rho_{1,j} \delta}{2} \right)$  is positive. On the other hand, it is easy to see that

$$\mathbf{P}_1^1 \left( \xi_M < -\frac{\rho_{1,j} M \delta}{2} \right) \leq \mathbf{P}_1^1 \left[ \pi - \left( \frac{-\rho_{1,j} M \delta}{2} \right) < \infty \right]$$

where

$$\pi_-(x) = \inf \{ k \geq 1 : \sum_{i=1}^k Z_i(j, 0) \leq x \}$$

$x < 0$  and it is assumed that  $\inf \emptyset = \infty$ . This last is the probability that the barrier  $-\rho_{1,j} M \delta / 2$  is crossed by the cumulative sum  $\sum_{i=1}^k Z_i(j, 0)$  started at the origin when  $E_1(Z_i(j, 0)) > 0$ . Let  $Q$  be the distribution function of  $Z_i(j, 0)$  when  $\mathcal{L}(X_i) = P_1$ . Two situations are possible. First, the random value  $Z_i(j, 0)$  is one-sided, i.e., it takes only nonnegative values with positive probability. In this case, the answer is trivial,  $\mathbf{P}_1^1[\pi_-( -\rho_{1,j} M \delta / 2) < \infty] = 0$ .

Second,  $Z_i(j, 0)$  is two-sided, i.e., it takes both positive and negative values with positive probability.<sup>4</sup> To compute an upper bound for  $\mathbf{P}_1^1[\pi_-(-\rho_{1,j}M\delta/2) < \infty]$ , the function  $Q$  should be imbedded in an exponential family [9, Ch. VIII]. For this reason we assume that the m.g.f.  $\varphi(\zeta) = E_1(e^{\zeta Z_i(j,0)}) < \infty$  exists (the integral  $\int e^{\zeta x} dQ(x)$  converges) in some neighborhood of 0,  $\zeta \in ]-\eta; \eta[$ , where  $\eta > 0$  (the regularity condition of the theorem). It follows from the properties of m.g.f. that  $\zeta \mapsto \varphi(\zeta)$  is twice differentiable in this open interval  $]-\eta; \eta[$ . Because the distribution of  $Z_i(j, 0)$  is not concentrated at zero, the second derivative  $\varphi''(\zeta) = \int x^2 e^{\zeta x} dQ(x)$  is strictly positive in  $]-\eta; \eta[$  and hence the function  $\zeta \mapsto \varphi(\zeta)$  is strictly convex (from below) in its interval of definition  $]-\eta; \eta[$ . We next apply the ‘‘generalized’’ Wald’s inequality [2], [3], [9]:  $\mathbf{P}_1^1[\pi_-(x) < \infty] \leq e^{-\zeta_1 x}$ , where  $x < 0$  and  $\zeta_1 < 0$  is a negative root of the equation  $\varphi(\zeta_1) = 1 (= \varphi(0))$  (the existence of a negative root  $\zeta_1$  is warranted because the random value  $Z_i(j, 0)$  is two-sided,  $\varphi(\zeta)$  is convex, and  $\varphi'(0) = E_1(Z_i(j, 0)) > 0$ ). In our context this leads to

$\mathbf{P}_1^1[\xi_M < -\rho_{1,j}M\delta/2] \leq \mathbf{P}_1^1[\pi_-(-\rho_{1,j}M\delta/2) < \infty] \leq e^{-\alpha_\xi \rho_{1,j}M\delta}$  where  $\alpha_\xi = |\zeta_1/2|$ . If  $h_{1,j} \rightarrow \infty$ , then  $n_h \rightarrow \infty$ . This means that for any chosen  $0 < \delta < 1$ ,  $\exists h_0 = h_0(\delta) : \forall h_{1,j} \geq h_0$

$$p_j(m) \leq e^{-\alpha_z m n_h} + e^{-\alpha_\xi \rho_{1,j} m n_h} \leq e^{-\alpha_j m n_h}$$

for any  $m > 0$ , where  $\alpha_z = H_{\tilde{Z}_i(1,j)}(-\rho_{1,j}\delta/2)$  and the positive constant  $\alpha_j$  is conveniently chosen.

Let us assume now that  $E_1(Z_i(j, 0)) < 0$  and  $j \geq 2$ . It follows from (14) that

$$p_j(m) = \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) + \zeta_M < h_{1,j} \right)$$

where

$$\zeta_M = \min_{0 \leq k \leq M} \sum_{i=k+1}^M -Z_i(j, 0).$$

Taking into account that  $h_{1,j} \leq h_d \leq n_h \rho_{1,0}(1 - \delta)$  and  $\tilde{Z}_i(1, 0) = Z_i(1, 0) - \rho_{1,0}$  we get the analog of (16), namely,

$$p_j(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, 0) < -\frac{\rho_{1,0}M\delta}{2} \right) + \mathbf{P}_1^1 \left( \zeta_M < -\frac{\rho_{1,0}M\delta}{2} \right). \quad (18)$$

As a result of [9, Appendix 2],  $\zeta_M$  has the same distribution as

$$\min_{0 \leq k \leq M} \sum_{i=1}^k -Z_i(j, 0).$$

Since  $E_1(-Z_i(j, 0)) > 0$ , the results of the previous case (when  $E_1(Z_i(j, 0)) > 0$ ) can be applied to the right-hand side of (18), replacing  $\tilde{Z}_i(1, 0)$ ,  $\rho_{1,0}$ , and  $\zeta_M$  by  $\tilde{Z}_i(1, j)$ ,  $\rho_{1,j}$ , and  $\xi_M$ , respectively. Let  $E_1(Z_i(j, 0)) = 0$ ,  $j \geq 2$ , and let  $\varepsilon > 0$  be a positive constant, then

$$p_j(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) - \max_{0 \leq k \leq M} \sum_{i=k+1}^M (Z_i(j, 0) + \varepsilon) < h_{1,j} \right).$$

Hence, by analogy with (16), we get

$$p_j(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, j) < -\frac{\rho_{1,j}M\delta}{2} \right) + \mathbf{P}_1^1 \left( \xi_M < -\frac{\rho_{1,j}M\delta}{2} + M\varepsilon \right) \quad (19)$$

where

$$\xi_M = \min_{0 \leq k \leq M} \sum_{i=1}^k (Z_i(j, 0) + \varepsilon).$$

<sup>4</sup>Taking into account that  $E_1(Z_i(j, 0)) > 0$ , we get by Jensen’s inequality  $E_1(Z_i^2(j, 0)) = \int x^2 dQ(x) \geq [\int x dQ(x)]^2 = [E_1(Z_i(j, 0))]^2 > 0$ . Hence, the distribution of  $Z_i(j, 0)$  is not concentrated at zero.

If  $\varepsilon$  is so chosen that  $0 < \varepsilon < \frac{\rho_{1,j}\delta}{2}$ , then again the results of the case when  $E_1(Z_i(j, 0)) > 0$  can be applied to (19). Finally, if  $j = 0$ , then  $g_t(0, 0) \equiv 0$ , we get

$$p_0(m) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, 0) < n_h \rho_{1,0}(1 - \delta - m) \right) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M \tilde{Z}_i(1, 0) < -\rho_{1,0}M\delta \right)$$

and apply directly Chernoff’s bound as in (17). Therefore, we have proved that  $p_j(m) \leq e^{-\alpha_j m n_h}$  holds for any sign of  $E_1(Z_i(j, 0))$  and  $0 \leq j \neq 1 \leq K - 1$ . It follows from (13) that for any given  $0 < \delta < 1 \exists h_1(\delta) \geq h_0 : \forall h_{1,j} \geq h_1(\delta)$

$$\mathbf{P}_1^1(N_r^1 > m n_h) \leq (K - 2)e^{-\alpha_j m n_h} \leq e^{-\alpha m n_h}$$

where  $\alpha > 0$  is a constant. By using the method proposed by Lai [4] we get

$$E_{t_0}^1(n_h^{-1}(N_r^1 - t_0 + 1) | N_r \geq t_0, g_{t_0-1} = 0) = \sum_{m=0}^{\infty} \mathbf{P}_{t_0}^1(n_h^{-1}(N_r^1 - t_0 + 1) > m | N_r \geq t_0, g_{t_0-1} = 0)$$

where  $g_{t_0-1} = (g_{t_0-1}(2, 0), \dots, g_{t_0-1}(K - 1, 0))$ , and then,

$$E_{t_0}^1(N_r^1 - t_0 + 1 | N_r \geq t_0, g_{t_0-1} = 0) \leq \max \left\{ \frac{h_d}{\rho_{1,0}}, \frac{h_i}{\min_{j \neq 0;1} \rho_{1,j}} \right\} (1 - \delta)^{-1} (1 - e^{-\alpha n_h})^{-1}. \quad (20)$$

Let us discuss now the case when  $g_{t_0-1}(j, 0) = \zeta_j \geq 0$ . From (14) it follows that

$$p_j(m | g_{t_0-1}(j, 0) = \zeta_j) = \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) - \max_{0 \leq k \leq M} \sum_{i=k+1}^M \tilde{Z}_i(j, 0) < h_{1,j} | \zeta_j \right)$$

where  $\tilde{Z}_i(j, 0) = Z_i(j, 0)$  when  $i > 1$  and  $\tilde{Z}_1(j, 0) = Z_1(j, 0) + \zeta_j$ . It is easy to see that

$$p_j(m | g_{t_0-1}(j, 0) = \zeta_j) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) - \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(j, 0) < h_{1,j} + \zeta_j | \zeta_j \right). \quad (21)$$

Let us suppose that  $\zeta_{t_0-1} = \max_{2 \leq j \leq K-1} g_{t_0-1}(j, 0)$ . It follows from (21) that the following inequality is fulfilled for every  $j : 2 \leq j \leq K - 1$ :

$$p_j(m | g_{t_0-1}(j, 0) = \zeta_j) \leq \mathbf{P}_1^1 \left( \sum_{i=1}^M Z_i(1, 0) - \max_{0 \leq k \leq M} \sum_{i=k+1}^M Z_i(j, 0) < h_{1,j} + \zeta_{t_0-1} | \zeta_j \right).$$

By applying the previously obtained results (from (14)–(20)), we get from (20), replacing  $h_d$  and  $h_i$  by  $h_d + \zeta_{t_0-1}$  and  $h_i + \zeta_{t_0-1}$ , respectively,

$$E_{t_0}^1(N_r^1 - t_0 + 1 | N_r \geq t_0, g_{t_0-1}) \leq \max \left\{ \frac{h_d + \zeta_{t_0-1}}{\rho_{1,0}}, \frac{h_i + \zeta_{t_0-1}}{\min_{j \neq 0;1} \rho_{1,j}} \right\} (1 - \delta)^{-1} (1 - e^{-\alpha n_h})^{-1}. \quad (22)$$

Taking into account that  $E_{t_0}^1(\zeta_{t_0-1}) = E_0(\zeta_{t_0-1})$  and in view of the property

$$E(y | A) = E_x[E(y | B, x) | A]$$

of iterated conditional expectations [10], we get from (22)

$$E_{t_0}^1(N_r^1 - t_0 + 1 | N_r \geq t_0) \leq \max \left\{ \frac{h_d + E_0(\zeta_{t_0-1})}{\rho_{1,0}}, \frac{h_i + E_0(\zeta_{t_0-1})}{\min_{j \neq 0;1} \rho_{1,j}} \right\} (1 - \delta)^{-1} (1 - e^{-\alpha n_h})^{-1}. \quad (23)$$

Note here that  $N_r = \min_{1 \leq j \leq K-1} \{N_r^j\}$ . It follows from this fact, (2) and (23) that

$$\bar{\tau}_1 \leq \max \left\{ \frac{h_d + \sup_{t_0 \geq 1} E_0(\zeta_{t_0-1})}{\rho_{1,0}}, \frac{h_i + \sup_{t_0 \geq 1} E_0(\zeta_{t_0-1})}{\min_{j \neq 0:1} \rho_{1,j}} \right\} \cdot (1-\delta)^{-1} (1 - e^{-\alpha n_h})^{-1}. \quad (24)$$

From [9, Appendix 2] it follows that

$$G_t(x) = \mathbf{P}_0(g_t(j, 0) \geq x) = \mathbf{P}_0(\pi_+(x) \leq t)$$

where  $\pi_+(x) = \inf\{k \geq 1 : \sum_{i=1}^k Z_i(j, 0) \geq x\}$ ,  $x > 0$ , and

$$G_\infty(x) = \lim_{t \rightarrow \infty} \mathbf{P}_0(g_t(j, 0) \geq x) = \mathbf{P}_0(\pi_+(x) < \infty).$$

By applying Wald's inequality to the LR  $\sum_{i=1}^k Z_i(j, 0)$  between hypotheses  $\mathcal{H}_j$  and  $\mathcal{H}_0$  [2], [3], [9], we get  $G_\infty(x) \leq e^{-x}$ . It is easy to see that

$$\begin{aligned} \sup_{t_0 \geq 1} E_0(g_{t_0-1}(j, 0)) &= \sup_{t_0 \geq 1} \int_0^\infty G_{t_0-1}(x) dx \\ &\leq \int_0^\infty G_\infty(x) dx \leq 1. \end{aligned}$$

Hence

$$\sup_{t_0 \geq 1} E_0(\zeta_{t_0-1}) \leq \sum_{j=2}^{K-1} \sup_{t_0 \geq 1} E_0(g_{t_0-1}(j, 0)) \leq K-2.$$

Combining this with (24) and taking into account that  $\delta$  is arbitrary,  $\alpha > 0$  and  $n_h \rightarrow \infty$  as  $h_d \geq h_i \rightarrow \infty$ , we get (8).

#### APPENDIX II PROOF OF THEOREM 2

Let  $x_1, \dots, x_n$  and  $y_1, \dots, y_n$  be two sets of real numbers. Since

$$\begin{aligned} \max(0, x_1 - y_1, \dots, x_n - y_n) \\ \geq \max(0, x_1, \dots, x_n) - \max(0, y_1, \dots, y_n) \end{aligned}$$

it then follows that for any  $t \geq t_0$

$$\begin{aligned} g_t(j, l) &= \max_{t_0-1 \leq k \leq t} \sum_{i=k+1}^t \bar{Z}_i(j, l) \\ &\geq \max_{t_0-1 \leq k \leq t} \sum_{i=k+1}^t \bar{Z}_i(j, 0) \\ &\quad - \max_{t_0-1 \leq k \leq t} \sum_{i=k+1}^t \bar{Z}_i(l, 0) = g_t(j, 0) - g_t(l, 0) \quad (25) \end{aligned}$$

where  $\bar{Z}_i(j, l) = Z_i(j, l)$  when  $i > t_0$  and  $\bar{Z}_{t_0}(j, l) = \bar{Z}_{t_0}(j, l) + g_{t_0-1}(j, 0)$  and  $\bar{Z}_i(m, 0) = Z_i(m, 0)$  when  $i > t_0$  and  $\bar{Z}_{t_0}(m, 0) = Z_{t_0}(m, 0) + g_{t_0-1}(m, 0)$  with  $m = l$  or  $m = j$ . Let us consider the "artificial" stopping time  $N_{t_0}^{j,l}$  which is activated at time  $t_0$

$$\begin{aligned} N_{t_0}^{j,l} &= \inf\{t \geq t_0 : g_t(j, l) \geq h_i\} \\ g_t(j, l) &= (g_{t-1}(j, l) + Z_t(j, l))^+ \\ g_{t_0-1}(j, l) &= g_{t_0-1}(j, 0). \quad (26) \end{aligned}$$

From the definition of the rule  $(N_r, \nu_r)$  it follows that  $\{N_r^l = N_r^j\} = \emptyset$ ,  $j \neq l$ . Taking into account this fact, (25), and (26), we obtain

$$\begin{aligned} \beta_{t_0}(j, l) &= \mathbf{P}_{t_0}^l(\nu_r = j | N_r \geq t_0) \\ &\leq \mathbf{P}_{t_0}^l(N_r^j < \bar{N}_r^j | N_r \geq t_0) \\ &\leq \mathbf{P}_{t_0}^l(N_{t_0}^{j,l} < \bar{N}_r^j | N_r \geq t_0) \quad (27) \end{aligned}$$

where  $\bar{N}_r^j = \min_{m \neq j} \{N_r^m\}$ . Let  $M = \min\{N_{t_0}^{j,l}, \bar{N}_r^j\}$ . It is obvious that  $N_{t_0}^{j,l} = M + I_{\{N_{t_0}^{j,l} > M\}}(N_{t_0}^{j,l} - M)$ , where  $I_A$  is the indicator of  $A$ . By computing the conditional expectation of  $N_{t_0}^{j,l} - M$  under  $\mathbf{P}_{t_0}^l$  given  $N_r \geq t_0$ , we get

$$\begin{aligned} E_{t_0}^l(N_{t_0}^{j,l} - M | N_r \geq t_0) \\ = E_{t_0}^l(N_{t_0}^{j,l} - M | \{N_{t_0}^{j,l} > M\} \cap \{N_r \geq t_0\}) \\ \cdot \mathbf{P}_{t_0}^l(N_{t_0}^{j,l} > M | N_r \geq t_0). \quad (28) \end{aligned}$$

Combining (27) with (28) and

$$\mathbf{P}_{t_0}^l(N_{t_0}^{j,l} < \bar{N}_r^j | N_r \geq t_0) \leq 1 - \mathbf{P}_{t_0}^l(N_{t_0}^{j,l} > M | N_r \geq t_0)$$

we obtain

$$\beta_{t_0}(j, l) \leq 1 - \frac{E_{t_0}^l(N_{t_0}^{j,l} - M | N_r \geq t_0)}{E_{t_0}^l(N_{t_0}^{j,l} - M | \{N_{t_0}^{j,l} > M\} \cap \{N_r \geq t_0\})}. \quad (29)$$

Since the function  $g_t(j, l)$  is a Markov sequence, it then follows that the random variable  $N_{t_0}^{j,l} - n$  given  $N_{t_0}^{j,l} > n \geq t_0$  depends on  $g_n(j, l)$  and  $X_{n+1}, X_{n+2}, \dots$ . Let us denote  $E_l(N_{t_0}^{j,l} | g_0(j, l) = x)$  as a function of  $x$  by  $L(x)$  on  $[0; h_i]$ . It follows from [1] that the average run length  $L(x)$  of the stopping time  $N_{t_0}^{j,l}$ , when the decision function  $g_t(j, l)$  starts from  $x \geq 0$ , is a decreasing function of  $x$ . The maximum is obtained for  $x = 0$  and is given by  $L(0)$ . Hence

$$E_{t_0}^l(N_{t_0}^{j,l} - M | \{N_{t_0}^{j,l} > M\} \cap \{N_r \geq t_0\}) \leq L(0).$$

After substitution of the last inequality into (29) and taking into account that for all large  $h_d, h_i$ ,  $E_{t_0}^l(N_{t_0}^{j,l} - M | N_r \geq t_0) > 0$ , we obtain

$$\begin{aligned} \sup_{t_0 \geq 1} \beta_{t_0}(j, l) &\leq 1 - \frac{\inf_{t_0 \geq 1} E_{t_0}^l(N_{t_0}^{j,l} - t_0 + 1 | N_r \geq t_0)}{L(0)} \\ &\quad + \frac{\sup_{t_0 \geq 1} E_{t_0}^l(N_r - t_0 + 1 | N_r \geq t_0)}{L(0)}. \quad (30) \end{aligned}$$

Let  $F_t(x) = \mathbf{P}_0(g_t(j, 0) < x)$  for  $x \geq 0$ . In view of the Markov property of  $g_t(j, l)$ , and denoting the density of the distribution  $F_t(x)$  by  $f_t(x)$ , we get

$$\begin{aligned} E_{t_0}^l(N_{t_0}^{j,l} - t_0 + 1 | N_r \geq t_0) \\ \geq \int_0^{h_i} L(x) f_{t_0-1}(x) dx = L(0) - L(h_i) G_{t_0-1}(h_i) \\ + \int_0^{h_i} \frac{dL(x)}{dx} G_{t_0-1}(x) dx. \quad (31) \end{aligned}$$

It follows from [9] that

$$\mathbf{P}_0(g_t(j, 0) < x) = \mathbf{P}_0\left(\max_{0 \leq k \leq t} \sum_{i=1}^k Z_i(j, 0) < x\right)$$

hence  $G_1(x) \leq G_2(x) \leq \dots \leq G_\infty(x)$  for any  $x \geq 0$ . Since the function  $x \mapsto L(x)$  is decreasing on the interval  $[0; h_i]$ , it results from (31) that

$$\inf_{t_0 \geq 1} E_{t_0}^l(N_{t_0}^{j,l} - t_0 + 1 | N_r \geq t_0) \geq \int_0^{h_i} L(x) f_\infty(x) dx. \quad (32)$$

From [1] it follows that

$$\begin{aligned} L(x) &= E_l[T(x)] + \mathbf{P}_l(S_T \leq -x) L(0) \geq l(x) \\ &= \frac{1}{\rho_{l,j}} \left( x - (h_i + \omega) e^{-(h_i-x)} \right) + \left( 1 - e^{-(h_i-x)} \right) L(0) \quad (33) \end{aligned}$$

where  $\omega = \sup_{r>0} E_l(Z_t(j, l) - r | Z_t(j, l) \geq r > 0)$  is an upper bound for the average "excess" over the boundary  $h_i$  (see details in

[1], [3], [9]),  $E_l[T(x)]$  is the average sample number of the sequential test  $T(x) = \inf\{t \geq 1 : (S_t \geq h_i - x) \cup (S_t \leq -x)\}$  with  $S_t = \sum_{i=1}^t Z_t(j, l)$ , and  $\mathbf{P}_l(S_T \leq -x)$  is the probability that the cumulative sum  $S_t$  of the sequential test  $T(x)$  reaches the lower threshold  $-x$ . By using the lower bound  $l(x)$  from (33) instead of  $L(x)$  in (32) and taking into account (31) we get

$$\begin{aligned} & \inf_{t_0 \geq 1} E_{t_0}^l \left( N_{t_0}^{j_l} - t_0 + 1 \mid N_r \geq t_0 \right) \\ & \geq -\frac{1}{\rho_{l,j}} (h_i + \omega) e^{-h_i} + (1 - e^{-h_i}) L(0) + \int_0^{h_i} \frac{G_\infty(x)}{\rho_{l,j}} dx \\ & \quad - \frac{(h_i + \omega) e^{-h_i}}{\rho_{l,j}} \int_0^{h_i} e^x G_\infty(x) dx - e^{-h_i} h_i L(0). \quad (34) \end{aligned}$$

Taking into account that  $0 \leq G_\infty(x) \leq e^{-x}$  and  $L(0) \geq e^{h_i}$ , we obtain, after integration of the third and fourth terms in the right side of (34), by combining (34) with (30), the following inequality:

$$\sup_{t_0 \geq 1} \beta_{t_0}(j, l) \lesssim e^{-h_i} \left( \sup_{t_0 \geq 1} E_{t_0}^l \left( N_r^l - t_0 + 1 \mid N_r \geq t_0 \right) + h_i \right)$$

as  $h_i \rightarrow \infty$ . By using the results of Theorem 1 we get the following upper bound for the probability of false isolation:

$$\sup_{t_0 \geq 1} \beta_{t_0}(j, l) \lesssim e^{-h_i} \max \left\{ \left( \frac{h_d}{\rho_{l,0}}, \frac{h_i}{\min_{j \neq 0: l} \rho_{l,j}} \right) + h_i \right\} \quad (35)$$

as  $h_i \rightarrow \infty$  and  $h_d \geq h_i$ . Equation (9) follows immediately from (35). The proof of Theorem 2 is finished.

## REFERENCES

- [1] M. Basseville and I. Nikiforov, *Detection of Abrupt Changes. Theory and Applications*, ser. Information and System Sciences. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [2] D. R. Cox and H. D. Miller, *The Theory of Stochastic Processes*. London, U.K.: Chapman & Hall, 1965.
- [3] *Handbook of Sequential Analysis*, B. K. Ghosh and P. K. Sen, Eds., Marcel Dekker, New York, 1991.
- [4] T. L. Lai, "Sequential changepoint detection in quality control and dynamical systems," *J. Roy. Statist. Soc. B*, vol. 57, no. 4, pp. 613–658, 1995.
- [5] I. V. Nikiforov, "A generalized change detection problem," *IEEE Trans. Inform. Theory*, vol. 41, pp. 171–187, Jan. 1995.
- [6] —, "On two new criteria of optimality for the problem of sequential change diagnosis," in *Proc. American Control Conf.*, vol. 1, Seattle, WA, June 21–23, 1995, pp. 97–101.
- [7] —, "Two strategies in the problem of change detection and isolation," *IEEE Trans. Inform. Theory*, vol. 43, pp. 770–776, Mar. 1997.
- [8] M. Pollak and D. Siegmund, "Approximations to the expected sample size of certain sequential tests," *Ann. Statist.*, vol. 3, no. 6, pp. 1267–1282, 1975.
- [9] D. Siegmund, *Sequential Analysis—Tests and Confidence Intervals*, ser. Series in Statistics. New York: Springer, 1985.
- [10] A. N. Shiryaev, *Probability*. New York: Springer, 1984. (Graduate Texts in Mathematics vol. 95).
- [11] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, 1976.

## Validation of Nearest Neighbor Classifiers

Eric Bax, *Member, IEEE*

**Abstract**—This correspondence presents a method to bound the out-of-sample error rate of a nearest neighbor classifier.<sup>1</sup> The bound is based only on the examples that comprise the classifier. Thus all available examples can be used in the classifier; no examples need to be withheld to compute error bounds.

The estimate used in the bound is an extension of the holdout estimate. The difference in error rates between the holdout classifier and the classifier consisting of all available examples is estimated using truncated inclusion and exclusion.

**Index Terms**—Error bounds, machine learning, nearest neighbor classifier, statistics, validation.

## I. FRAMEWORK

Consider the following machine learning framework. There is a joint input–output distribution. For example, the input distribution could consist of typical satellite images of the North Atlantic Ocean, and the output could be 1 if the image contains a large iceberg and 0 otherwise.

We have a set of in-sample data examples

$$S = \{(\mathbf{x}_1, \theta_1), \dots, (\mathbf{x}_n, \theta_n)\}$$

with each example drawn independent and identically distributed (i.i.d.) from the joint input–output distribution. We will use a nearest neighbor classifier, composed of the in-sample examples and a distance metric, to classify the inputs of test examples drawn i.i.d. from the input–output distribution. For each test input, the classifier returns the output corresponding to the closest in-sample input. The test error rate is the fraction of test inputs for which the classifier and the test output disagree. The underlying error rate  $L_n$  is the expected test error rate over the input–output distribution. The average of  $L_n$  over all size  $n$  in-sample data sets drawn from the input–output distribution is  $R_n$ . To make the definitions of  $L_n$  and  $R_n$  explicit, let  $y_{NN}$  be the output of the nearest neighbor classifier, and let  $y$  be the test output. Then

$$L_n = \Pr\{y_{NN} \neq y \mid S\} \quad (1)$$

and

$$R_n = E_S L_n = \Pr\{y_{NN} \neq y\}. \quad (2)$$

(We use subscripts to denote the distributions over which expectations are taken.)

## II. INTRODUCTION

While this correspondence focuses on  $L_n$ , the error rate of the classifier at hand, much work in the past has focused on  $R_n$ , the average error rate over classifiers formed from randomly drawn examples. Cover and Hart [7] proved that under mild continuity assumptions,  $R_\infty$  is no more than twice the Bayes (optimal) error rate. Cover [6] and Psaltis, Snapp, and Venkatesh [17] have investigated the convergence of  $R_n$  to  $R_\infty$ . Cover [6] worked with the case of a one-dimensional input space,

Manuscript received October 5, 1998; revised April 29, 2000.

The author is with Math and Computer Science Department, University of Richmond, Richmond, VA 23173 (e-mail: ebax@richmond.edu).

Communicated by P. A. Chou, Associate Editor for Source Coding. Publisher Item Identifier S 0018-9448(00)09656-5.

<sup>1</sup>Patent pending.