Proceedings of the
44th IEEE Conference on Decision and Control, and
the European Control Conference 2005
Seville, Spain, December 12-15, 2005

Asymptotic Performance of a Multichart CUSUM Test Under False Alarm Probability Constraint

Alexander G. Tartakovsky, *Senior Member*, IEEE

Abstract— **Traditionally the false alarm rate in change point detection problems is measured by the mean time to false detection (or between false alarms). The large values of the mean time to false alarm, however, do not generally guarantee small values of the false alarm probability in a fixed time interval for any possible location of this interval. In this paper we consider a multichannel (multi-population) change point detection problem under a non-traditional false alarm probability constraint, which is desirable for a variety of applications. It is shown that in the multichart CUSUM test this constraint is easy to control. Furthermore, the proposed multichart CUSUM test is shown to be uniformly asymptotically optimal when the false alarm probability is small: it minimizes an average detection delay, or more generally, any positive moment of the stopping time distribution for any point of change.**

Index Terms— **Change-point detection, sequential detection, multichart CUSUM test, asymptotic optimality, renewal theory, false alarm probability.**

I. INTRODUCTION

The problem of rapid detection of abrupt changes in stochastic systems and processes is of importance for a variety of applications such as signal and image processing, quality control engineering, computer intrusion detection, chemical or biological warfare agent detection systems, failure detection in various systems, target detection in surveillance systems, etc. [1], [2], [5], [15], [20], [22], [23], [24]. In all these applications sensors monitoring the environment take observations that undergo a change in distribution in response to a change in the environment. The change occurs at an unknown point in time, and the practitioners' goal is to detect it as quickly as possible while avoiding frequent false alarms.

The classical change-point detection problem deals with the i.i.d. case where there is a sequence of observations $X_1, X_2, \dots$ that are identically distributed with a probability density function (pdf) $f(x)$ for $n < \lambda$ and with a pdf $g(x)$ for $n \geqslant \lambda$ (with respect to a sigma-finite measure $\mu(x)$), where λ, $\lambda = 1, 2, \dots$ is an unknown point of change. In this paper, we will be interested in the following multi-population (or "multichannel") generalization. Suppose there are N mutually independent populations X_n^i, $i = 1, \dots, N$ which, for $\lambda < n$, are distributed according to the pdfs $f_i(x)$. At an unknown time λ a change occurs, and one of the populations (and only one) changes its statistical properties. If the change occurs in the j-th population, then for $n \geqslant \lambda$ the pdf of X_n^j is $g_j(x)$. In other words, the joint pdf of the vector $\mathbf{Y}_n = (\mathbf{X}_n^1, \dots, \mathbf{X}_n^N)$, $\mathbf{X}_n^i = (X_1^i, \dots, X_n^i)$ conditioned on the hypothesis H_k^j that the change happens in the j-th population at the time $\lambda = k$ has the form

$$p(\mathbf{Y}^n | H_k^j) = \begin{cases} \prod_{s=1}^n \prod_{\substack{i=1 \\ i \neq j}}^N f_i(X_s^i) \cdot \prod_{s=1}^{k-1} f_j(X_s^j) \\ \prod_{s=k}^n g_j(X_s^j), \text{if } k \leqslant n; \\ \prod_{s=1}^n \prod_{i=1}^N f_i(X_s^i), \text{if } k > n. \end{cases} \tag{1}$$

In the sequel we write $\mathbf{P}_\infty$ and $\mathbf{E}_\infty$ for the probability measure and expectation under which there is no change (i.e. $\lambda = \infty$); while for any $\lambda = k < \infty$, $\mathbf{P}_k^j$ and $\mathbf{E}_k^j$ are used to denote the probability measure and expectation when the change occurs in the j-th population at the point $\lambda = k$.

A sequential detection procedure is identified with a random stopping time τ with respect to the sigma-algebra $\mathcal{F}_n = \sigma(\mathbf{X}^n)$, i.e. $\{\tau \leqslant n\} \in \mathcal{F}_n$, $n \geqslant 1$.

The customary performance indices of the sequential procedure τ are the average run lengths (ARLs) $\mathrm{ARL}_0(\tau) = \mathbf{E}_\infty \tau$ and $\mathrm{ARL}_1^j(\tau) = \mathbf{E}_1^j \tau$. The value of ARL_0 measures the false alarm rate (FAR), and the value of ARL_1^j the expected detection delay when the change occurs from the very beginning. It is desirable to make $\mathrm{ARL}_0(\tau)$ large and $\mathrm{ARL}_1^j(\tau)$ small. These performance characteristics are popular in the Statistical Process Control community. More generally, a good detection procedure should guarantee small values of the average (expected) detection delay $\mathrm{ADD}_k^j(\tau) = \mathbf{E}_k^j(\tau - k | \tau \geqslant k)$ for all $k \geqslant 1$ and $j = 1, \dots, N$ when the FAR is fixed at a certain level.

However, if the FAR is measured in terms of the mean time to false alarm, i.e. it is required that $\mathrm{ARL}_0(\tau) \geqslant \gamma$ for some $\gamma > 0$, then a procedure that minimizes the average detection delay $\mathrm{ADD}_k^j(\tau)$ for all k does not exist even for $N = 1$. For $N = 1$, it is only possible to find minimax detection procedures that minimize $\sup_k \mathrm{ADD}_k(\tau)$ in the worst case scenario [1], [8], [9], [10], [12], [13]. For $N > 1$, asymptotically optimal solutions that minimize $\sup_k \mathrm{ADD}_k^j(\tau)$ as $\gamma \to \infty$ for all $j = 1, \dots, N$ have been found in [16], [22]. More importantly, the requirement of having large values of the mean time to false alarm $\mathbf{E}_\infty \tau$ generally does not guarantee small values of the probability of false alarm (PFA) $\mathbf{P}_\infty(k \leqslant \tau \leqslant k + T)$ in a time interval of the fixed length T for all $k \geqslant 1$ or the small values of the corresponding conditional PFA $\mathbf{P}_\infty(\tau \leqslant k + T | \tau \geqslant k)$, $k \geqslant 1$. Indeed, the condition $\mathbf{E}_\infty \tau \geqslant \gamma$ only guarantees that for any $T \geqslant 1$ there exists some $k \geqslant 1$, generally depending on γ, such that

$$\mathbf{P}_\infty \{\tau - k \leqslant T | \tau \geqslant k\} \leqslant T/\gamma. \tag{2}$$

A. Tartakovsky is with the Department of Mathematics, University of Southern California, 3620 S. Vermont Avenue, Los Angeles, CA 90089-2532, USA tartakov@usc.edu

The inequality (2) can be proved by contradiction as follows. Without loss of generality we may assume that $\mathbf{P}_\infty\{\tau \geqslant k\} > 0$, since otherwise $\mathbf{P}_\infty(\tau \geqslant k) = 0$ for all k and $\mathbf{E}_\infty\tau = 0$, which contradicts the inequality $\mathbf{E}_\infty\tau \geqslant \gamma$. Assume that

$$\mathbf{P}_\infty(\tau \geqslant k+T | \tau \geqslant k) < 1 - T/\gamma \quad \text{for all } k \geqslant 1. \quad (3)$$

It suffices to consider only integer T, $T \leqslant \gamma$. Due to (3) and the fact that $\mathbf{P}_\infty(\tau \geqslant i+kT) = \mathbf{P}_\infty(\tau \geqslant i)\mathbf{P}_\infty(\tau \geqslant i+kT|\tau \geqslant i)$ we have

$$\begin{aligned}\mathbf{E}_\infty\tau &= \sum_{i=1}^{\infty} \mathbf{P}_\infty\{\tau \geqslant i\} \\ &= \sum_{i=1}^{T}\sum_{k=0}^{\infty} \mathbf{P}_\infty\{\tau \geqslant i+kT\} \\ &< \sum_{i=1}^{T}\sum_{k=0}^{\infty} \mathbf{P}_\infty\{\tau \geqslant i\}(1-T/\gamma)^k \\ &= (\gamma/T)\sum_{i=1}^{T} \mathbf{P}_\infty\{\tau \geqslant i\} \leqslant \gamma/T,\end{aligned}$$

which contradicts the assumption $\mathbf{E}_\infty\tau \geqslant \gamma$.

Therefore, the condition $\mathbf{E}_\infty\tau \geqslant \gamma$ only guarantees the existence of some k (that possibly depends on γ) for which $\mathbf{P}_\infty(\tau \leqslant k+T|\tau \geqslant k) < T/\gamma$. This means that, for a given $0 < \alpha < 1$, the PFA constraint

$$\sup_{k \geqslant 1} \mathbf{P}_\infty(\tau \leqslant k+T|\tau \geqslant k) \leqslant \alpha \quad \text{for a certain } T \geqslant 1 \quad (4)$$

is stronger than the ARL constraint $\mathbf{E}_\infty\tau \geqslant \gamma$.[1]

At the same time, for many practical applications, including computer intrusion detection and a variety of surveillance applications such as target detection and tracking, it is desirable to control the PFA $\mathbf{P}_\infty(\tau \leqslant k+T|\tau \geqslant k)$ for all $k \geqslant 1$ at a certain (usually low) level α.

Owing to this fact, in what follows we will be interested in the class of detection procedures $\mathbf{\Delta}_T(\alpha)$ for which the PFA $\mathbf{P}_\infty(\tau \leqslant k+T|\tau \geqslant k)$ does not exceed a predefined value $\alpha < 1$ for all $k \geqslant 1$ and some $T \geqslant 1$, i.e. we will focus on detection procedures that satisfy (4).

While in general the large values of the mean time to false alarm $\mathbf{E}_\infty\tau$ do not guarantee small values of $\sup_{k\geqslant 1}\mathbf{P}_\infty(\tau \leqslant k+T|\tau \geqslant k)$, for certain detection procedures the PFA can be easily controlled.

Consider, for example, the following multichart extension of Page's CUSUM (cumulative sum) procedure [11]. For $i = 1, \dots, N$, define the statistics

$$U_i(n) = \max\left\{1, \max_{1\leqslant k\leqslant n}\prod_{s=k}^{n}\frac{g_i(X_s^i)}{f_i(X_s^i)}\right\}, \quad U_i(0) = 1$$

and the Markov stopping times

$$\tau_i(B_i) = \inf\{n \geqslant 1 : U_i(n) \geqslant B_i\},$$

where $B_i > 1$ are finite numbers (thresholds). The stopping time of the multichart CUSUM test is

$$\nu(B_1, \dots, B_N) = \min\{\tau_1(B_1), \dots, \tau_N(B_N)\}.$$

It may be shown that under mild conditions [6] the $\mathbf{P}_\infty$-limiting distribution of $\tau_i(B_i)/\mathbf{E}_\infty\tau_i(B_i)$ is geometric:

$$\lim_{B_i\to\infty}\mathbf{P}_\infty\{\tau_i(B_i)/\mathbf{E}_\infty\tau_i(B_i) > x\} = \exp(-x), \quad x > 0.$$

Since $\mathbf{E}_\infty\tau_i(B_i) \geqslant B_i$ and the stopping times $\tau_1, \dots, \tau_N$ are independent, it follows that for sufficiently large B_i

$$\mathbf{P}_\infty(\nu \leqslant k+T|\nu \geqslant k) \leqslant 1 - \exp\left(-T\sum_{i=1}^{N} B_i^{-1}\right).$$

Consider the particular case where the thresholds $B_i = B$ do not depend on i. Then, for small α,

$$B = B_\alpha(N,T) = TN/|\log(1-\alpha)| \approx TN/\alpha$$

guarantees that $\nu(B_\alpha(N))$ belongs to the class $\mathbf{\Delta}_T(\alpha)$.

More generally, assume that $T = T_\alpha$ depends on α and goes to infinity when $\alpha \to 0$. The main theoretical contribution of this paper is the proof of asymptotic optimality (as $\alpha \to 0$) of the proposed multichart CUSUM test in the sense of minimizing the average detection delay $\mathrm{ADD}_k^j(\tau) = \mathbf{E}_k^j(\tau - k|\tau \geqslant k)$ for all hypotheses H_k^j (i.e. for every change point $k \geqslant 1$ and $j = 1, \dots, N$) whenever $T_\alpha = O(|\log\alpha|)$. (See Theorem 1 in Section II.)

Note that under the classical ARL constraint $\mathrm{ARL}_0 \geqslant \gamma$ the asymptotic optimality (as $\gamma \to \infty$) of this multichart test has been proven in Tartakovsky [16]. An alternative mixture-based multichannel Shiryaev-Roberts detection procedure has been studied in Tartakovsky and Veeravalli [22], also under the ARL constraint.

II. ASYMPTOTIC OPTIMALITY OF THE MULTICHART CUSUM TEST FOR LOW PFA

In what follows we will always assume that $B_1 = \cdots = B_N = B$, in which case the stopping time of the multichart CUSUM test can be written in the form

$$\nu(B) = \inf\left\{n \geqslant 1 : \max_{1\leqslant i\leqslant N} U_i(n) \geqslant B\right\}.$$

Let

$$I_f(i) = \int \log\frac{f_i(x)}{g_i(x)} f_i(x) d\mu(x) \quad \text{and}$$
$$I_g(i) = \int \log\frac{g_i(x)}{f_i(x)} g_i(x) d\mu(x)$$

denote Kullback-Leibler information numbers and let $Z_n(i) = \log[g_i(X_n^i)/f_i(X_n^i)]$, $i = 1, \dots, N$ denote the log-likelihood ratios.

[1]It is easily verified that (4) implies $\mathbf{E}_\infty\tau \geqslant \gamma$ for $\gamma = \gamma(\alpha, T) = \alpha^{-1}\sum_{i=1}^{T}\mathbf{P}_\infty(\tau \geqslant i)$.

A. The Probability of False Alarm

We begin with obtaining an approximation for the PFA $\mathbf{P}_\infty\{\nu(B) \leqslant k+T|\nu(B) \geqslant k\}$ for large values of B.

For $i = 1, \ldots, N$, define

$$\begin{aligned}\eta_i^- &= \inf\{n \geqslant 1 : S_n(i) \leqslant 0\},\\ \eta_i^+ &= \inf\{n \geqslant 1 : S_n(i) > 0\},\\ \eta_i(h) &= \inf\{n \geqslant 1 : S_n(i) \geqslant h\},\\ \gamma_i &= \lim_{h\to\infty} \mathbf{E}_1^i \exp\left\{-(S_{\eta_i(h)} - h)\right\},\end{aligned}$$

where $S_n(i) = \sum_{s=1}^n Z_s(i)$, $S_0(i) = 0$.

The following lemma shows that asymptotically as $B \to \infty$ the distribution of the stopping time $\nu(B)$ is geometric.

Lemma 1: Let $I_f(i)$ and $I_g(i)$ be positive and finite. Assume that $Z_1(i)$, $i = 1, \ldots, N$ are nonarithmetic with respect to $\mathbf{P}_\infty$ and $\mathbf{P}_1^i$. Then, as $B \to \infty$,

$$\mathbf{P}_\infty\left\{\nu(B)N\bar{C}_N/B > x\right\} = e^{-x}(1+o(1)), \quad x > 0, \quad (5)$$

where $\bar{C}_N = N^{-1}\sum_{i=1}^N \gamma_i^2 I_g(i)$.

Proof: By Theorem 3 of Khan [6],

$$\lim_{B\to\infty} \mathbf{P}_\infty\{\tau_i(B)/B > x\} = \exp(-C_i x), \quad x > 0,$$

where $0 < C_i \leqslant 1$ is the constant given by

$$C_i = (1 - \mathbf{E}_\infty \exp\{S_{\eta_i^-}(i)\})^2/[I_g(i)(\mathbf{E}_\infty \eta_i^-)^2].$$

Since $\nu(B) = \min_{1\leqslant i\leqslant N} \tau_i(B)$ and $\tau_1, \ldots, \tau_N$ are independent, we obtain

$$\begin{aligned}\mathbf{P}_\infty\{\nu(B)/B > x\} &= \prod_{i=1}^N \mathbf{P}_\infty\{\tau_i(B)/B > x\}\\ &= \exp\left(-x\sum_{i=1}^N C_i\right)(1+o(1)). \quad (6)\end{aligned}$$

By Wald's likelihood ratio identity,

$$\mathbf{E}_\infty \exp\{S_{\eta_i^-}(i)\} = \mathbf{P}_1^i(\eta_i^- < \infty) = 1 - \mathbf{P}_1^i(\eta_i^- = \infty)$$

and, by Corollary 8.39 (p. 173) in [14], $\mathbf{E}_\infty\eta_i^- = 1/\mathbf{P}_\infty(\eta_i^+ = \infty)$. Thus, C_i can be written in the following form

$$C_i = [\mathbf{P}_1^i(\eta_i^- = \infty)\mathbf{P}_\infty(\eta_i^+ = \infty)]^2/I_g(i). \quad (7)$$

Using Corollary 8.33 (p. 171) in [14], it is easily shown that

$$\gamma_i = (1 - \mathbf{E}_1^i \exp\{S_{\eta_i^+}(i)\})/\mathbf{E}_1^i S_{\eta_i^+}(i). \quad (8)$$

By Wald's identity and Corollary 8.39 (p. 173) in [14], $\mathbf{E}_1^i S_{\eta_i^+}(i) = I_g(i)\mathbf{E}_1^i\eta_i^+ = I_g(i)/\mathbf{P}_1^i(\eta_i^- = \infty)$ and by Wald's likelihood ratio identity

$$1 - \mathbf{E}_1^i \exp\{S_{\eta_i^+}(i)\} = \mathbf{P}_\infty(\eta_i^+ = \infty).$$

Substituting these expressions into (8) yields

$$\gamma_i = \mathbf{P}_1^i(\eta_i^- = \infty)\mathbf{P}_\infty(\eta_i^+ = \infty)/I_g(i).$$

Comparing with (7) shows that $C_i = I_g(i)\gamma_i^2$. Lemma 1 follows by substituting this last expression for C_i into (6). ■

Before we proceed, it is worth noting that the constant γ_i, which is related to a limiting overshoot in a one-sided sequential test $\eta_i(h)$, can be computed by renewal theory reasoning [14], [19], [25].

Lemma 1 allows us to obtain the asymptotic approximation for the PFA: for all $k \geqslant 1$ as $B \to \infty$

$$\mathbf{P}_\infty(\nu(B) \leqslant k+T|\nu(B) \geqslant k) \sim 1 - \exp\left\{-TN\bar{C}_N/B\right\}.$$

Therefore, if $B = B_\alpha(N,T) = TN\bar{C}_N/\alpha$, then for all $k \geqslant 1$

$$\mathbf{P}_\infty(\nu(B_\alpha) \leqslant k+T|\nu(B_\alpha) \geqslant k) \sim \alpha, \quad \text{as } \alpha \to 0.$$

It is worth remarking that Lai [7] proposed to impose constraints on the unconditional PFA

$$\sup_{k\geqslant 1} \mathbf{P}_\infty\{k \leqslant \tau < k+T\}.$$

However, as Lemma 1 suggests, for large B the $\mathbf{P}_\infty$-distribution of the stopping time $\nu(B)$ is approximately geometric and the unconditional PFA

$$\mathbf{P}_\infty\{k \leqslant \nu(B) < k+T\} \approx e^{-(k-1)}TN\bar{C}_N/B, \quad k \geqslant 1.$$

Therefore, despite the fact that

$$\sup_k \mathbf{P}_\infty\{k \leqslant \nu(B) < k+T\} \approx TN\bar{C}_N/B,$$

the unconditional PFA decays exponentially fast when k increases. This nullifies a seemingly natural constraint on $\sup_k \mathbf{P}_\infty\{k \leqslant \tau < k+T\}$.

Assume now that the size of the time window $T = T_\alpha$, where the PFA is confined, is allowed to depend on α and goes to infinity when $\alpha \to 0$. If the threshold $B_\alpha(N)$ is the root of the equation

$$\exp\left\{-T_\alpha N\bar{C}_N/B\right\} = 1 - \alpha, \quad (9)$$

then, as $\alpha \to 0$,

$$\sup_k \mathbf{P}_\infty\{\nu(B_\alpha) < k+T_\alpha|\nu(B_\alpha) \geqslant k\} \sim \alpha. \quad (10)$$

Obviously, we may replace (9) with

$$N\bar{C}_N T_\alpha/B = \alpha. \quad (11)$$

In particular, if T_α is explicitly expressed via α, then $B_\alpha(N) = N\bar{C}_N T_\alpha/\alpha$ implies asymptotic equality (10). However, T_α may implicitly depend on α via the threshold value B. For example, if $T_\alpha = O(\log B_\alpha)$, then in order to find $B_\alpha(N)$ one has to solve the equation (11).

B. Asymptotic Optimality of the Multichart CUSUM test

Let $\mathbf{\Delta}(\alpha) = \{\tau : \sup_{k\geqslant 1}\mathbf{P}_\infty(\tau < k+T_\alpha|\tau \geqslant k) \leqslant \alpha\}$ denote the class $\mathbf{\Delta}_T(\alpha)$ in the case when T_α depends on α. It turns out that the multichart CUSUM procedure with the threshold $B_\alpha(N)$ that obeys the equation (11) is asymptotically optimal, as $\alpha \to 0$, in the class $\mathbf{\Delta}(\alpha)$ when T_α satisfies certain conditions (see (12) below).

To be more specific, let $L_\alpha = |\log\alpha|/\min_i I_g(i)$ and assume that the time interval $T = T_\alpha$ depends on α in such a way that

$$\liminf_{\alpha\to 0}(T_\alpha/L_\alpha) > 1 \quad \text{but} \quad \lim_{\alpha\to 0}[(\log T_\alpha)/L_\alpha] = 0. \quad (12)$$

The following theorem, whose proof is given in the Appendix, establishes the asymptotic optimality result with respect to any positive moment of the detection delay $\mathrm{D}_k^{j,m}(\tau) = \mathbf{E}_k^j\{(\tau-k)^m|\tau \geqslant k\}$, $m>0$.

Theorem 1: Let m be a positive, not necessarily integer number and let T_α satisfy conditions (12). Assume that $0 < I_g(i) < \infty$ and $0 < I_f(i) < \infty$ and that the log-likelihood ratios $Z_1(i)$, $i=1,\dots,N$ are nonarithmetic. If the threshold $B_\alpha(N)$ is chosen from the equation (11), then for all $m>0$, $k \geqslant 1$, and $j=1,\dots,N$,

$$\inf_{\tau\in\mathbf{\Delta}(\alpha)} \mathrm{D}_k^{j,m}(\tau) \sim \mathrm{D}_k^{j,m}(\nu(B_\alpha)) \sim \left(\frac{|\log\alpha|}{I_g(j)}\right)^m \tag{13}$$

as $\alpha \to 0$.

III. HIGHER ORDER APPROXIMATIONS FOR THE AVERAGE DETECTION DELAY

Theorem 1 provides a first-order approximation for moments of the detection delay for small PFA α. As the proof of this theorem shows, regardless of the false alarm constraint, for large threshold values, $\mathrm{D}_k^{i,m}(\nu(B))$ can be upperbounded as in (24). Furthermore, replacing $|\log\alpha|$ with $\log B$ in the proof of Lemma 2 and using an almost identical argument shows that this upper bound is asymptotically sharp, i.e.

$$\mathrm{D}_k^{i,m}(\nu(B)) \sim [(\log B)/I_g(i)]^m \quad \text{as } B\to\infty.$$

The goal of this section is to improve the first-order approximations for the expected detection delay $\mathrm{D}_k^{i,1}(\nu(B)) = \mathrm{ADD}_k^i(\nu(B))$ – to obtain higher-order approximations up to the vanishing term as $B\to\infty$ using Nonlinear Renewal Theory developed by Lai and Siegmund [14] and Woodroofe [25].

Define $S_0^k(i)=0$, $S_n^k(i)=\sum_{l=k}^n Z_l(i)$,

$$\tilde{S}_n(i) = \max\left\{0, \max_{1\leqslant l\leqslant n} S_n^l(i)\right\} = S_n^1(i) - \min_{0\leqslant l\leqslant n} S_l^1(i),$$

$$\bar{\beta}_i = \mathbf{E}_1^i[\min_{n\geqslant 0}\sum_{s=1}^n Z_s(i)], \quad \mu_i = \lim_{n\to\infty}\mathbf{E}_\infty\tilde{S}_n(i),$$

$$\varkappa_i = \lim_{h\to\infty}\mathbf{E}_1^i(S^1_{\eta_i(h)}(i)-h),$$

where $\eta_i(h) = \inf\left\{n\geqslant 1: S_n^1(i)\geqslant h\right\}$ is the one-sided test.

Due to space limitations we are not able to present a complete, mathematically rigorous proof and, for this reason, give only an intuitive argument, which can be regarded as a proof sketch. The complete detailed proof will be presented elsewhere.

Note first that it may be expected that, for large threshold values, the probability $\mathbf{P}_k^i\left\{\nu(B)=\tau_i(B)|\nu(B)\geqslant k\right\}$ is close to 1, so that, as $B\to\infty$,

$$\mathrm{ADD}_k^i(\nu(B)) = \mathbf{E}_k^i\left\{\tau_i(B)-k|\tau_i(B)\geqslant k\right\} + o(1). \tag{14}$$

Therefore, it suffices to evaluate $\mathbf{E}_k^i\left\{\tau_i(B)-k|\tau_i(B)\geqslant k\right\}$ for large B. To this end, we rewrite the stopping time $\tau_i(B)$ in the form of a random walk crossing a threshold plus a nonlinear term that is slowly changing in the sense defined in [14], [25]. This allows us to apply nonlinear renewal theory.

Specifically, observe that $\tau_i(B)$ can be written in the form

$$\tau_i(b) = \inf\left\{n\geqslant 1: \tilde{S}_n(i)\geqslant b\right\}, \quad b=\log B.$$

A simple algebra shows that

$$\tilde{S}_n(i) = \tilde{S}_{k-1}(i) + S_n^k(i) - \beta_i(k,n),$$

where

$$\beta_i(k,n) = \min\left\{0, \min_{k\leqslant l\leqslant n} S_l^1(i) - \min_{1\leqslant l\leqslant k-1} S_l^1(i)\right\}.$$

Further,

$$S_{\tau_i}^k(i) = b - \tilde{S}_{k-1}(i) + \beta_i(k,\tau_i) + \chi_i(b),$$

where $\chi_i(b) = \tilde{S}_{\tau_i}(i) - b$ is the excess of the process $\tilde{S}_n(i)$ over the level b at time τ_i. Observing that the sequence $\{S_n^k(i)\}_{n\geqslant k}$ is a random walk with mean $\mathbf{E}_k^i S_n^k(i) = I_g(i)(n-k+1)$, taking the conditional expectations and applying Wald's identity, we obtain

$$I_g(i)\mathbf{E}_k^i(\tau_i-k+1|\tau_i\geqslant k) = b - \mathbf{E}_k^i[\tilde{S}_{k-1}(i)|\tau_i\geqslant k] + \mathbf{E}_k^i[\beta_i(k,\tau_i)|\tau_i\geqslant k] + \mathbf{E}_k^i(\chi_b|\tau_i\geqslant k).$$

The crucial observation is that the sequence $\{\beta_i(k,n)\}_{n\geqslant k}$ is slowly changing and, moreover, converges $\mathbf{P}_k^i$-a.s. as $n\to\infty$ to the random variable

$$\beta_i(k) = \min\left\{0, \min_{l\geqslant k} S_l^1(i) - \min_{1\leqslant l\leqslant k-1} S_l^1(i)\right\}$$

with finite expectation $\mathbf{E}_k^i\beta_i(k) = \bar{\beta}_i(k)$.

An important consequence of the slowly changing property is that, under mild conditions, the limiting distribution of the excess of a random walk over a fixed threshold does not change by the addition of a slowly changing nonlinear term (see Theorem 4.1 of Woodroofe [25]). Since $\mathbf{P}_k^i(\tau_i\geqslant k) = \mathbf{P}_\infty(\tau_i\geqslant k)\to 1$ as $b\to\infty$, it follows that $\mathbf{E}_k^i[\tilde{S}_{k-1}(i)|\tau_i\geqslant k] \to \mathbf{E}_\infty[\tilde{S}_{k-1}(i)]$, $\mathbf{E}_k^i[\beta_i(k,\tau_i)|\tau_i\geqslant k]\to\bar{\beta}_i(k)$, and $\mathbf{E}_k^i(\chi_b|\tau_i\geqslant k)\to\varkappa_i$. Therefore, we may expect that for a large b,

$$\mathrm{ADD}_k^i(\tau_i(b)) = \frac{1}{I_g(i)}\left(b - \mathbf{E}_\infty\tilde{S}_{k-1}(i) + \bar{\beta}_i(k) + \varkappa_i\right) - 1 + o(1). \tag{15}$$

It is difficult to evaluate $\mathbf{E}_\infty\tilde{S}_{k-1}(i)$ and $\bar{\beta}_i(k)$ for arbitrary values of k. This task seems feasible only for $k=1$ and large k ($k\to\infty$). Indeed, if $k=1$, then $\tilde{S}_0(i)=0$ and $\bar{\beta}_i(1) = \mathbf{E}_1^i \min_{n\geqslant 0} S_n^1(i)$; while when $k\to\infty$, $\bar{\beta}_i(k)\to 0$ and $\mathbf{E}_\infty\tilde{S}_{k-1}(i)\to\mu_i$, where μ_i is the mean of the stationary distribution of the Markov process $\tilde{S}_n(i)$.

The exact result is given in the following theorem.

Theorem 2: Assume that $\mathbf{E}_1^i|Z_1(i)|^2<\infty$ and that $Z_1(i)$ are $\mathbf{P}_1^i$-nonarithmetic, $i=1,\dots,N$. Then, for $k=1$,

$$\mathrm{ADD}_k^i(\nu(b)) = I_g^{-1}(i)\left(b+\varkappa_i+\bar{\beta}_i\right) - 1 + o(1) \tag{16}$$

and, as $k\to\infty$,

$$\mathrm{ADD}_k^i(\nu(b)) = I_g^{-1}(i)\left(b+\varkappa_i-\mu_i\right) - 1 + o(1) \quad \text{as } b\to\infty. \tag{17}$$

Further mathematical details will be presented somewhere else. Here we only remark that crucial points are to prove (14) for all $k \geqslant 1$ and (17) for large k. The proof of (15) for $k = 1$ and exponential families is given by Dragalin [3]. A generalization for an arbitrary distribution with finite second moment requires only proving that, for every $0 < \varepsilon < 1$,

$$\lim_{b\to\infty} b\,\mathbf{P}_1^i\left\{\tau_i(b) \leqslant (1-\varepsilon)b/I_g(i)\right\} = 0.$$

For this purpose, the argument in the proof of Theorem 5 of Tartakovsky and Veeravalli [21] can be used.

We complete this section by giving useful formulas for the constants $\bar{\beta}_i$, μ_i, and $\varkappa_i$. We write $Y^- = -\min(0, Y)$ and $Y^+ = \max(0, Y)$. Under $\mathbf{P}_\infty$, the distribution of the Markov process $\tilde{S}_n(i)$ converges to the stationary distribution, which is equal to the $\mathbf{P}_\infty$-distribution of $\max_{k\geqslant 0} S_k(i)$. Therefore, $\mu_i = \mathbf{E}_\infty(\max_{k\geqslant 0} S_k(i))$. Spitzer's formula and Borovkov's identity apply to yield

$$\mu_i = \sum_{n=1}^{\infty} n^{-1}\mathbf{E}_\infty S_n^+(i). \tag{18}$$

Since $\bar{\beta}_i = -\mathbf{E}_1^i[\max_{n\geqslant 0}\{-S_n(i)\}]$ and $\mathbf{E}[-S_n(i)]^+ = \mathbf{E}[-S_n(i)\mathbb{1}_{\{S_n(i)\leqslant 0\}}] = \mathbf{E}S_n^-(i)$, similar to (18)

$$\beta_i = -\sum_{n=1}^{\infty} n^{-1}\mathbf{E}_1^i S_n^-(i).$$

The constant $\varkappa_i$ is the subject of renewal theory. The following formula is particularly useful:

$$\varkappa_i = \mathbf{E}_1^i Z_1^2(i)/2\mathbf{E}_1^i Z_1(i) - \sum_{n=1}^{\infty} n^{-1}\mathbf{E}_1^i S_n^-(i).$$

IV. CONCLUSIONS

In contrast to the conventional FAR measure $\mathbf{E}_\infty\tau$, in this paper we propose to measure FAR in terms of the conditional probability of false alarm $\sup_{k\geqslant 1}\mathbf{P}_\infty\{\tau \leqslant k+T|\tau \geqslant k\}$, which is shown to be a stronger requirement than $\mathbf{E}_\infty\tau \geqslant \gamma$. It is shown that in the quickest multipopulation change detection problem the multichart CUSUM test easily meets the PFA constraint. Moreover, this detection test is asymptotically optimal in the first-order sense with respect to any positive moment of the detection delay under fairly weak conditions when the PFA is small. We also derived higher order approximations for the average detection delay using nonlinear theory reasoning. These approximations involve several constants that are subject of renewal theory and can be easily computed for specific data models.

Further research is needed to verify accuracy of first-order and higher-order approximations for a number of examples using Monte Carlo simulations.

APPENDIX

The proof of (13) is performed in two steps. The first step is to obtain an asymptotic lower bound for moments of the detection delay $\inf_{\tau\in\mathbf{\Delta}(\alpha)} \mathrm{D}_k^{j,m}(\tau)$ for any procedure from the class $\mathbf{\Delta}(\alpha)$. This is performed in the following lemma whose proof is based on the Chebyshev inequality and the change-of-measure reasoning similar to that used in [4], [7], [18], [21].

Lemma 2: Let T_α satisfy conditions (12). If $0 < I_g(i) < \infty$, $i = 1, \dots, N$, then for all $m > 0$ and $k \geqslant 1$

$$\inf_{\tau\in\mathbf{\Delta}(\alpha)} \mathrm{D}_k^{j,m}(\tau) \geqslant \left(\frac{|\log\alpha|}{I_g(j)}\right)^m (1+o(1)), \tag{19}$$

as $\alpha \to 0$.

Proof: Write $L_\alpha^i = |\log\alpha|/I_g(i)$. By Chebyshev's inequality, for any $0 < \varepsilon < 1$, $m > 0$, and any $\tau \in \mathbf{\Delta}(\alpha)$

$$\mathrm{D}_k^{j,m}(\tau) \geqslant (\varepsilon L_\alpha^j)^m \mathbf{P}_k^j\left\{\tau - k \geqslant \varepsilon L_\alpha^j|\tau \geqslant k\right\}. \tag{20}$$

Therefore, in order to prove (19) it suffices to show that for an arbitrarily small ε and all $k \geqslant 1$, as $\alpha \to 0$,

$$\sup_{\tau\in\mathbf{\Delta}(\alpha)}\sup_{k\geqslant 1}\mathbf{P}_k^j\left\{\tau < k + (1-\varepsilon)L_\alpha^j|\tau \geqslant k\right\} \to 0 \tag{21}$$

whenever $0 < I_g(j) < \infty$, $j = 1, \dots, N$.

Write $S_n^k(i) = \sum_{l=k}^n Z_l(i)$. Changing the measure, we obtain that for any $L \geqslant 1$, $b > 0$ and any stopping time τ

$$\begin{aligned}
&\mathbf{P}_\infty(\tau - k < L|\tau \geqslant k) = \mathbf{E}_k^j\left\{\mathbb{1}_{\{\tau<k+L\}}e^{-S_\tau^k(j)}|\tau \geqslant k\right\}\\
&\geqslant \mathbf{E}_k^j\left\{\mathbb{1}_{\{\tau<k+L,S_\tau^k(j)<b\}}e^{-S_\tau^k(j)}|\tau \geqslant k\right\}\\
&\geqslant e^{-b}\mathbf{P}_k^j\left\{\tau < k+L,\ \max_{k\leqslant n<k+L} S_\tau^k(j) < b|\tau \geqslant k\right\}\\
&\geqslant e^{-b}\Big\{\mathbf{P}_k^j(\tau < k+L|\tau \geqslant k)\\
&\quad - \mathbf{P}_k^j\Big(\max_{k\leqslant n<k+L} S_n^k(j) \geqslant b|\tau \geqslant k\Big)\Big\}.
\end{aligned}$$

Since the event $\{\tau \geqslant k\}$ belongs to the sigma-field $\mathcal{F}_{k-1}$ and the event $\{\max_{k\leqslant n<k+L} S_n^k(j) \geqslant b\}$ does not depend on $\mathbf{Y}_{k-1}$, it follows that

$$\begin{aligned}
&\mathbf{P}_k^j\Big(\max_{k\leqslant n<k+L} S_n^k(j) \geqslant b|\tau \geqslant k\Big)\\
&\qquad = \mathbf{P}_k^j\Big(\max_{k\leqslant n<k+L} S_n^k(j) \geqslant b\Big).
\end{aligned}$$

Furthermore, by the i.i.d. property,

$$\mathbf{P}_k^j\Big(\max_{k\leqslant n<k+L} S_n^k(j) \geqslant b\Big) = \mathbf{P}_1^j\Big(\max_{1\leqslant n\leqslant L}\sum_{l=1}^n Z_l(i) \geqslant b\Big)$$

and, hence,

$$\begin{aligned}
\sup_k \mathbf{P}_k^j(\tau < k+L|\tau \geqslant k) &\leqslant e^b \sup_k \mathbf{P}_\infty(\tau < k+L|\tau \geqslant k)\\
&\quad + \mathbf{P}_1^j\Big(\max_{1\leqslant n\leqslant L}\sum_{l=1}^n Z_l(i) \geqslant b\Big).
\end{aligned}$$

Putting $L = L(\varepsilon,\alpha,j) = (1-\varepsilon)L_\alpha^j = (1-\varepsilon)|\log\alpha|/I_g(j)$ and $b = (1+\varepsilon)I_g(j)L(\varepsilon,\alpha,j) = (1-\varepsilon^2)|\log\alpha|$, we obtain

$$\begin{aligned}
&\sup_k \mathbf{P}_k^j\left\{\tau < k+(1-\varepsilon)L_\alpha^j|\tau \geqslant k\right\} \leqslant\\
&\leqslant e^{(1-\varepsilon^2)|\log\alpha|}\sup_k \mathbf{P}_\infty\left\{\tau < k+(1-\varepsilon)L_\alpha^j|\tau \geqslant k\right\}\\
&\quad + \mathbf{P}_1^j\left\{\max_{1\leqslant n\leqslant(1-\varepsilon)L_\alpha^j}\sum_{l=1}^n Z_l(i) \geqslant (1-\varepsilon^2)|\log\alpha|\right\}.
\end{aligned} \tag{22}$$

Since $0 < I_g(j) < \infty$, by the strong law of large numbers regardless of the choice of τ, as $\alpha \to 0$,

$$\mathbf{P}_1^j \left\{ \max_{1 \leqslant n \leqslant (1-\varepsilon)L_\alpha^j} \sum_{i=1}^n Z_l(i) \geqslant (1-\varepsilon^2)|\log \alpha| \right\} \to 0.$$

Also, by (12),

$$e^{(1-\varepsilon^2)|\log \alpha|} \sup_{\tau \in \mathbf{\Delta}(\alpha)} \sup_{k \geqslant 1} \mathbf{P}_\infty \left\{ \tau < k + (1-\varepsilon)L_\alpha^j | \tau \geqslant k \right\} \leqslant$$
$$\leqslant e^{(1-\varepsilon^2)|\log \alpha|} \alpha = \alpha^{\varepsilon^2},$$

since $T_\alpha > (1-\varepsilon)L_\alpha^j$ for all sufficiently small α.

Thus, both terms on the right-hand side of (22) go to zero uniformly over $\tau \in \mathbf{\Delta}(\alpha)$, which means that (21) follows. The proof of (19) is complete. ■

The second step is to obtain an upper bound for $\mathrm{D}_k^{j,m}(\nu(B))$ and to show that this bound is asymptotically the same as the lower bound when $B = B_\alpha(N)$.

To this end, introduce the sequence of one-sided stopping times

$$\eta_k^j(B) = \min \left\{ n \geqslant k : S_n^k(j) \geqslant \log B \right\}, \ k = 1, 2, \ldots$$

It is easily seen that $\log U_j(n) \geqslant S_n^k(j)$ for any $n \geqslant k$. Thus, $\nu(B) \leqslant \tau_j(B) \leqslant \eta_k^j(B)$ for any $k \geqslant 1$. Moreover, $\nu(B) - k \leqslant \eta_k^j(B) - k$ on $\{\nu(B) \geqslant k\}$. Since $\{\nu(B) \geqslant k\} \in \mathcal{F}_{k-1}$ and $\eta_k^j(B)$ does not depend on $\mathcal{F}_{k-1}$, it follows

$$\begin{aligned} \mathrm{D}_k^{j,m}(\nu(B)) &\leqslant \mathbf{E}_k^j[(\eta_k^j(B) - k)^m | \nu(B) \geqslant k] \\ &= \mathbf{E}_k^j(\eta_k^j(B) - k)^m. \end{aligned}$$

Since the random variables $Z_n(j)$, $n = 1, 2, \ldots$ are i.i.d., the distribution of $\eta_k^j(B) - k + 1$ under $\mathbf{P}_k^j$ is the same as the $\mathbf{P}_1^j$-distribution of the stopping time

$$\eta_1^j(B) = \min \left\{ n \geqslant 1 : \sum_{l=1}^n Z_l(j) \geqslant \log B \right\}.$$

Therefore, using the previous inequality we obtain that

$$\mathrm{D}_k^{j,m}(\nu(B)) \leqslant \mathbf{E}_1^j(\eta_1^j(B) - 1)^m \quad \text{for all } k \geqslant 1, \tag{23}$$

which can be used to obtain the desired upper bound.

Since $I_g(j)$ is assumed positive and finite, Theorem 4.1 of Dragalin, Tartakovsky, and Veeravalli [4] applies to yield, for all positive m,

$$\mathbf{E}_1^j[\eta_1^j(B)]^m \sim \left(\frac{\log B}{I_g(j)} \right)^m \quad \text{as } B \to \infty.$$

Using this last asymptotic relation along with (23), we obtain the asymptotic upper bound

$$\mathrm{D}_k^{j,m}(\nu(B)) \leqslant \left(\frac{\log B}{I_g(j)} \right)^m (1 + o(1)) \quad \text{as } B \to \infty. \tag{24}$$

Finally, note that if $B_\alpha(N)$ obeys the equation (11) with T_α that satisfies conditions (12), then $\log B_\alpha(N) \sim |\log \alpha|$ as $\alpha \to 0$ and, therefore,

$$\mathrm{D}_k^{j,m}(\nu(B_\alpha(N))) \leqslant \left(\frac{|\log \alpha|}{I_g(j)} \right)^m (1 + o(1)) \quad \text{as } \alpha \to 0.$$

Comparing this asymptotic upper bound with the lower bound (19) completes the proof of (13).

References

[1] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes: Theory and Applications*, Prentice Hall, Englewood Cliffs, 1993.

[2] R. Blažek, H. Kim, B. Rozovskii, and A. Tartakovsky, "A novel approach to detection of "denial-of-service" attacks via adaptive sequential and batch-sequential change-point detection methods," In *Proc. IEEE Systems, Man, and Cybernetics Information Assurance Workshop*, West Point, NY, 2001.

[3] V.P. Dragalin, "Optimality of a generalized CUSUM procedure in quickest detection problem," In *Statistics and Control of Random Processes: Proceedings of the Steklov Institute of Mathematics*, Vol. 202, Issue 4, pp. 107–120, 1994. American Mathematical Society: Providence, Rhode Island.

[4] V.P. Dragalin, A.G. Tartakovsky, and V.V. Veeravalli, "Multihypothesis sequential probability ratio tests, part I: asymptotic optimality," *IEEE Trans. Inform. Theory*, Vol. 45, pp. 2448–2461, 1999.

[5] S. Kent, "On the trial of intrusions into information systems," *IEEE Spectrum*, Vol. 37, No. 12, pp. 52–56, 2000.

[6] R.A. Khan, "Detecting changes in probabilities of a multi-component process," *Sequential Analysis*, Vol. 14, No. 4, pp. 375–388, 1995.

[7] T.L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inform. Theory*, Vol. 4, pp. 2917–2929, 1998.

[8] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, Vol. 42, pp. 1897–1908, 1971.

[9] G.V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, Vol. 14, pp. 1379–1387, 1986.

[10] G.V. Moustakides, "Optimality of the CUSUM procedure in continuous time," *Ann. Statist.*, Vol. 32, pp. 302–315, 2004.

[11] E.S. Page, "Continuous inspection schemes," *Biometrika*, Vol. 41, pp. 100–115, 1954.

[12] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, Vol. 13, pp. 206–227, 1985.

[13] A.N. Shiryaev, "Minimax optimality of the method of cumulative sum (cusum) in the case of continuous time," *Russian Math. Surveys*, Vol. 51, No. 4, pp. 750–751, 1996.

[14] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.

[15] A.G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*, Radio i Svyaz', Moscow, 1991 (In Russian).

[16] A.G. Tartakovsky, "Asymptotically minimax multialternative sequential rule for disorder detection," In *Statistics and Control of Random Processes: Proceedings of the Steklov Institute of Mathematics*, Vol. 202, Issue 4, pp. 229–236, 1994. American Mathematical Society: Providence, Rhode Island.

[17] A.G. Tartakovsky, "Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in a nonhomogeneous Gaussian process," *Mathematical Methods of Statistics*, Vol. 4, pp. 389–404, 1995.

[18] A.G. Tartakovsky, "Asymptotic optimality of certain multihypothesis sequential tests: non-i.i.d. case," *Statistical Inference for Stochastic Processes*, Vol. 1, No. 3, pp. 265–295, 1998.

[19] A.G. Tartakovsky and I.A. Ivanova, "Comparison of some sequential rules for detecting changes in distributions," *Problems of Information Transmission*, Vol. 28, pp. 117-124, 1992.

[20] A.G. Tartakovsky and V.V. Veeravalli, "An efficient sequential procedure for detecting changes in multichannel and distributed systems," In *Proceeding of the 5th International Conference on Information Fusion*, Annapolis, Maryland, Vol. 1, pp. 41–48, 2002.

[21] A.G. Tartakovsky and V.V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *Theory Prob. Appl.*, Vol. 49, No. 3, pp. 538–582, 2004.

[22] A.G. Tartakovsky and V.V. Veeravalli, "Change-point detection in multichannel and distributed systems with applications," In *Applications of Sequential Methodologies*; N. Mukhopadhyay, S. Datta and S. Chattopadhyay, Eds.; Marcel Dekker, Inc., New York, pp. 339–370, 2004.

[23] H. Wang, D. Zhang, K.G. Shin, "Detecting SYN flooding attacks," In *Proceedings of INFOCOM2002, 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, Vol. 3, pp. 1530–1539, 2002.

[24] A.S. Willsky, "A survey of design methods for failure detection in dynamical systems," *Automatica*, Vol. 12, pp. 601–611, 1976.

[25] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, Philadelphia, 1982.