THEORY PROBAB. APPL.
Vol. 49, No. 3, pp. 458–497




# GENERAL ASYMPTOTIC BAYESIAN THEORY OF QUICKEST CHANGE DETECTION*

A. G. TARTAKOVSKY† AND V. V. VEERAVALLI‡

**Abstract.** The optimal detection procedure for detecting changes in independent and identically distributed (i.i.d.) sequences in a Bayesian setting was derived by Shiryaev in the 1960s. However, the analysis of the performance of this procedure in terms of the average detection delay and false alarm probability has been an open problem. In this paper, we develop a general asymptotic change-point detection theory that is not limited to a restrictive i.i.d. assumption. In particular, we investigate the performance of the Shiryaev procedure for general discrete-time stochastic models in the asymptotic setting, where the false alarm probability approaches zero. We show that the Shiryaev procedure is asymptotically optimal in the general non-i.i.d. case under mild conditions. We also show that the two popular non-Bayesian detection procedures, namely the Page and the Shiryaev–Roberts–Pollak procedures, are generally not optimal (even asymptotically) under the Bayesian criterion. The results of this study are shown to be especially important in studying the asymptotics of *decentralized* change detection procedures.

**Key words.** change-point detection, sequential detection, asymptotic optimality, nonlinear renewal theory

**DOI.** 10.1137/S0040585X97981202

## 1. Introduction.

The problem of detecting abrupt changes in stochastic processes arises in a variety of applications including biomedical signal processing, quality control engineering, finance, link failure detection in communication networks, intrusion detection in computer systems, and target detection in surveillance systems [1], [4], [12], [19], [30]. A typical problem is target detection in multisensor systems (radar, infrared, sonar, etc.) [30], [35], [39], where the target appears randomly at an unknown time. The goal is to detect the target as quickly as possible, while maintaining the false alarm rate at a given level. Another application area is intrusion detection in distributed computer networks [4], [12], [39]. Large-scale attacks, such as denial-of-service attacks, occur at unknown points in time and need to be detected in the early stages by observing abrupt changes in the network traffic.

The design of the quickest change detection procedures usually involves optimizing the trade-off between two kinds of performance measures, one being a measure of detection delay and the other being a measure of the frequency of false alarms. There are two standard mathematical formulations for the optimum trade-off problem. The first of these is a minimax formulation proposed by Lorden [17] and Pollak [21], in which the goal is to minimize the worst-case delay subject to a lower bound on the mean time between false alarms. The second is a Bayesian formulation, proposed by Shiryaev [25], [26], [27], in which the change point is assumed to have a geometric prior distribution, and the goal is to minimize the expected delay subject to an upper bound on false alarm probability. The asymptotic performance of various change-

*Received by the editors November 1, 2003. This work was supported in part by the U.S. Office of Naval Research grants N00014-03-1-0027 and N00014-03-1-0823, the U.S. DARPA grant N66001-00-C-8044 at the University of Southern California, and the U.S. National Science Foundation CAREER/PECASE award CCR-0049089 at the University of Illinois at Urbana-Champaign.

http://www.siam.org/journals/tvp/49-3/98120.html

†Center for Applied Mathematical Sciences and Department of Mathematics, University of Southern California, Los Angeles, CA 90089-1113 (tartakov@math.usc.edu).

‡Coordinated Science Laboratory and Department of Electrical & Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 (vvv@uiuc.edu).

point detection procedures is well understood in the minimax context for both the discrete- and continuous-time cases (see [1], [3], [7], [9], [17], [18], [20], [21], [22], [23], [28], [30], [31], [32], [34], [39], [43], [44]). However, there has been little previous work on the asymptotics of Bayesian procedures. An exception is the work by Lai [16] in which the asymptotic properties of Page's cumulative sum (CUSUM) procedure were studied in a Bayesian (as well as minimax) context for general stochastic models. (See also Beibel [3] for continuous-time Brownian motion and Borovkov [5, Lemma 5] for i.i.d. data models.) Our goal is to provide a general Bayesian asymptotic theory for change-point detection.

The paper is organized as follows. In section 2, we formulate the problem. In section 3, we study the behavior of the Shiryaev detection procedure for general, non-i.i.d. data models and prove that it is asymptotically optimal under mild conditions when the false alarm probability goes to zero. We show not only that this procedure is asymptotically optimal with respect to the average detection delay, but also that it is uniformly asymptotically optimal in the sense of minimizing the conditional expected delay for every change point. Moreover, we study the behavior of higher moments of the detection delay and show that under certain general conditions the Shiryaev procedure minimizes moments of the detection delay up to a given order. In section 4, we find the asymptotic operating characteristics of the Shiryaev change detection procedure in the i.i.d. case when the false alarm probability goes to zero. The use of nonlinear renewal theory allows us to obtain sharp asymptotic approximations for the false alarm probability and the average detection delay up to vanishing terms. In section 5, we analyze the asymptotic performance of other well-known change detection procedures (Page's procedure and the Shiryaev–Roberts–Pollak procedure) in the Bayesian framework. The results of this section allow us to conclude that, while being optimal in the minimax context, these procedures may lose their optimality property (even asymptotically) with respect to the Bayesian criterion, depending on the structure of the prior distribution. In section 6, we consider an example of detecting a change in the mean value of an autoregressive process that illustrates general results. In section 7, we consider two additional examples related to detecting changes in distributed multisensor systems. We study the implications of the asymptotic results in decentralized quickest change detection problems, assuming that the sensors send quantized versions of their observations to a *fusion* center (central processor), where the change detection is performed based on all the sensor messages. Finally, in section 8, we conclude the paper by giving several remarks.

The results of this paper were presented in part at the 2002 IEEE International Symposium on Information Theory [36], the 2002 IEEE Information Theory Workshop [37], and at the Sixth International Conference on Information Fusion, 2003 [38].

**2. Problem formulation.** In the conventional setting of the change-point detection problem, one assumes that the observed random variables $X_1, X_2, \dots$ are i.i.d., until a change occurs at an unknown point in time $\lambda$, $\lambda \in \{1, 2, \dots\}$. After the change occurs, the observations are again i.i.d. but with another distribution. In other words, conditioned on $\lambda = k$, the observations $X_1, X_2, \dots$ are independent with $X_n \sim f_0$ for $n < k$ and $X_n \sim f_1$ for $n \geqq k$, where $f_0(x)$ and $f_1(x)$ are, respectively, the prechange and postchange probability density functions (PDFs) with respect to a sigma-finite measure $\mu$. For the sake of brevity, in what follows, this case will be referred to as the "i.i.d. case."

In a Bayesian setting, the change-point $\lambda$ is assumed to be random with prior probability distribution $\pi_k = \mathbf{P}\{\lambda = k\}$, $k = 0, 1, 2, \dots$ . The goal is to detect the change as soon as possible after it occurs, subject to false alarm probability constraints.

In mathematical terms, a sequential detection procedure is identified with a stopping time $\tau$ for an observed sequence $\{X_n\}_{n\geqq 1}$, i.e., $\tau$ is an extended integer-valued random variable, such that the event $\{\tau \leqq n\}$ belongs to the sigma-algebra $\mathcal{F}_n^X = \sigma(X_1,\dots,X_n)$. A false alarm is raised whenever the detection is declared before the change occurs, i.e., when $\tau < \lambda$. A good detection procedure should guarantee a "stochastically small" detection lag $\tau - \lambda$, provided that there is no false alarm (i.e., $\tau \geqq \lambda$), while the rate of false alarms should be low.

Let $\mathbf{P}_k$ and $\mathbf{E}_k$ denote the probability measure and the corresponding expectation when the change occurs at time $\lambda = k$. In what follows, $\mathbf{P}^\pi$ stands for the "average" probability measure, which is defined as $\mathbf{P}^\pi(\Omega) = \sum_{k=0}^\infty \pi_k \mathbf{P}_k(\Omega)$, and $\mathbf{E}^\pi$ denotes the expectation with respect to $\mathbf{P}^\pi$.

In the Bayesian setting, a reasonable measure of the detection lag is the average detection delay (ADD)

$$\begin{aligned} \mathrm{ADD}(\tau) &= \mathbf{E}^\pi(\tau-\lambda \mid \tau \geqq \lambda) = \frac{\mathbf{E}^\pi(\tau-\lambda)^+}{\mathbf{P}^\pi\{\tau \geqq \lambda\}} \\ &= \frac{1}{\mathbf{P}^\pi\{\tau \geqq \lambda\}} \sum_{k=0}^\infty \pi_k \mathbf{P}_k\{\tau \geqq k\}\, \mathbf{E}_k\{\tau - k \mid \tau \geqq k\}, \end{aligned} \tag{2.1}$$

and the false alarm rate can be measured by the probability of false alarm (PFA)

$$\mathrm{PFA}(\tau) = \mathbf{P}^\pi\{\tau < \lambda\} = \sum_{k=1}^\infty \pi_k \mathbf{P}_k\{\tau < k\}. \tag{2.2}$$

Here and henceforth, we use the traditional notation $x^+ = \max(0, x)$ for the positive part of $x$.

An optimal Bayesian detection procedure is a procedure for which ADD is minimized, while PFA is constrained to be below a given (usually small) level $\alpha$, where $\alpha \in (0,1)$. Specifically, define the class of change-point detection procedures $\Delta(\alpha) = \{\tau\colon\ \mathrm{PFA}(\tau) \leqq \alpha\}$ for which the PFA does not exceed the predefined number $\alpha$. The optimal change-point detection procedure is described by the stopping time

$$\nu = \underset{\tau \in \Delta(\alpha)}{\arg\inf}\, \mathrm{ADD}(\tau).$$

Obviously, if $\mathrm{PFA}(\nu) = \alpha$, then $\nu$ also minimizes $\mathbf{E}^\pi(\tau-\lambda)^+$.

Let $\mathbf{X}^n = (X_1,\dots,X_n)$ denote the concatenation of the first $n$ observations, let $\mathcal{F}_n^X = \sigma(\mathbf{X}^n)$ be a sigma-algebra generated by $\mathbf{X}^n$, and let $p_n = \mathbf{P}\{\lambda \leqq n \,|\, \mathcal{F}_n^X\}$ be the posterior probability that the change occurred before time $n$. For the i.i.d. case, Shiryaev [25], [26], [27] proved that if the distribution of the change point is geometric, i.e., $\mathbf{P}\{\lambda = 0\} = \pi_0$, $\pi_k = (1-\pi_0)\,\rho(1-\rho)^{k-1}$, $k \geqq 1$ ($0<\rho<1$, $0 \leqq \pi_0 < 1$), then the optimal detection procedure is the one that raises an alarm at the first time such that the posterior probability $p_n$ exceeds a threshold $A$, i.e.,[1]

$$\nu(A) = \inf\{n \geqq 1\colon\ p_n \geqq A\}, \tag{2.3}$$

where the threshold $A = A_\alpha$ should be chosen so that $\mathrm{PFA}(\nu(A)) = \alpha$. A generalization of this result for an arbitrary prior distribution has been stated in [5], albeit

[1] Hereafter we use the convention that $\inf \varnothing = \infty$, i.e., $\nu(A) = \infty$ if no such $n$ exists.

without proof (see [5, Theorem 8]). However, except for the case of detecting the change in the drift of the Wiener process observed in continuous time, it is difficult to find a threshold that provides an exact match to the given PFA. Also, there are no results related to the ADD evaluation of this optimal procedure, again, except for the continuous-time Wiener process [27] with exponential prior distribution, and for i.i.d. data models with a geometric prior distribution when $\rho \to 0$ (see [5, Lemma 5]).

While the exact match of the PFA is related to the estimation of the overshoot in the stopping rule (2.3), and for this reason is problematic, a simple upper bound, which ignores overshoot, can be obtained. Indeed, since $\mathbf{P}^\pi\{\nu(A) < \lambda\} = \mathbf{E}^\pi\{1 - p_{\nu(A)}\}$ and $1 - p_{\nu(A)} \leqq 1 - A$ on $\{\nu(A) < \infty\}$, we obtain that the PFA defined in (2.2) obeys the inequality

$$\mathrm{PFA}\big(\nu(A)\big) \leqq 1 - A. \tag{2.4}$$

It follows that setting $A = A_\alpha = 1 - \alpha$ guarantees the inequality $\mathrm{PFA}(\nu(A_\alpha)) \leqq \alpha$. Note that inequality (2.4) holds true for arbitrary (proper), not necessarily geometric, prior distributions.

More generally, assume that observations are non-i.i.d. in the prechange and postchange modes. Specifically, let $\mathbf{P}_\infty$ stand for the probability measure under which the conditional density of $X_n$ given $\mathbf{X}^{n-1} = (X_1, \dots, X_{n-1})$ is $f_{0,n}(X_n \,|\, \mathbf{X}^{n-1})$ for every $n \geqq 1$ (i.e., $\lambda = \infty$). Furthermore, for any $\lambda = k$, $1 \leqq k < \infty$, let $\mathbf{P}_k$ stand for the probability measure under which the conditional density of $X_n$ is $f_{0,n}(X_n \,|\, \mathbf{X}^{n-1})$ if $n \leqq k-1$ and is $f_{1,n}(X_n \,|\, \mathbf{X}^{n-1})$ if $n \geqq k$. Therefore, if the change occurs at time $\lambda = k$, then the conditional density of the $k$th observation changes from $f_{0,k}(X_k \,|\, \mathbf{X}^{k-1})$ to $f_{1,k}(X_k \,|\, \mathbf{X}^{k-1})$.

The log-likelihood ratio (LLR) for the hypothesis that the change occurred at the point $\lambda = k$ and at $\lambda = \infty$ (no change at all) is

$$Z_n^k := \log \frac{d\mathbf{P}_k}{d\mathbf{P}_\infty}(\mathbf{X}^n) = \sum_{t=k}^{n} \log \frac{f_{1,t}(X_t \mid \mathbf{X}^{t-1})}{f_{0,t}(X_t \mid \mathbf{X}^{t-1})}, \qquad k \leqq n. \tag{2.5}$$

In what follows, we will use the convention that before the observations become available (i.e., for $n = 0$), $Z_0^0 = \log[f_{1,0}(X_0)/f_{0,0}(X_0)] = 0$ almost everywhere.

In the rest of the paper, we will consider prior distributions $\pi_k = \mathbf{P}\{\lambda = k\}$ concentrated on nonnegative integers. The following two classes of prior distributions will be covered: distributions with exponential tails for which

$$\lim_{k\to\infty} \frac{\log \mathbf{P}\{\lambda \geqq k+1\}}{k} = -d, \qquad d > 0, \tag{2.6}$$

and prior distributions with "heavy" tails for which

$$\lim_{k\to\infty} \frac{\log \mathbf{P}\{\lambda \geqq k+1\}}{k} = 0. \tag{2.7}$$

The first class will be denoted by $\mathcal{E}(d)$ and the second class by $\mathcal{H}$.

Clearly, the geometric prior distribution with the parameter $\rho$,

$$\pi_k = \pi_0 1\!\!1_{\{k=0\}} + (1-\pi_0)\,\rho(1-\rho)^{k-1}\, 1\!\!1_{\{k\geqq 1\}}, \qquad 0 < \rho < 1, \;\; 0 \leqq \pi_0 < 1, \tag{2.8}$$

belongs to the class $\mathcal{E}(d)$ with $d = |\log(1-\rho)|$. Here and henceforth, $1\!\!1_{\{\mathcal{A}\}}$ denotes an indicator of the set $\mathcal{A}$.

Note that a more general case, in which a fixed $d$ is replaced with the value of $d_\alpha$ that depends on $\alpha$ and vanishes when $\alpha \to 0$, can also be handled in a similar way. This more general case has been considered by Lai [16]. However, this slight generalization does not have a substantial impact on practical applications.

For $n \geqq 0$, define the likelihood ratio of the hypotheses "$H_1\colon \lambda \leqq n$" and "$H_0\colon \lambda > n$":

$$\Lambda_n := \frac{d\mathbf{P}\{\mathbf{X}^n \mid H_1\}}{d\mathbf{P}\{\mathbf{X}^n \mid H_0\}} = \frac{\sum_{k=0}^{n} \pi_k \prod_{i=k}^{n} f_{1,i}(X_i \mid \mathbf{X}^{i-1}) \prod_{i=1}^{k-1} f_{0,i}(X_i \mid \mathbf{X}^{i-1})}{\sum_{k=n+1}^{\infty} \pi_k \prod_{i=1}^{n} f_{0,i}(X_i \mid \mathbf{X}^{i-1})},$$

where $f_{1,0}(X_0)/f_{0,0}(X_0) = 1$ almost everywhere by the above convention.

Recall that $p_n = \mathbf{P}\{\lambda \leqq n \,|\, \mathcal{F}_n^X\}$ denotes the posterior probability of the event that the change occurred before time $n$. Write $\Pi_n = \mathbf{P}\{\lambda > n\}$. It is easily verified that

$$\Lambda_n = \Lambda_0 + \Pi_n^{-1} \sum_{k=1}^{n} \pi_k e^{Z_n^k} \quad \text{and} \quad \Lambda_n = \frac{p_n}{1-p_n}, \qquad n \geqq 1, \tag{2.9}$$

where $\Lambda_0 = \pi_0/(1-\pi_0)$. Therefore, the Shiryaev stopping rule given in (2.3) can be written in the following form that is more convenient for asymptotic study:

$$\nu_B = \inf\{n \geqq 1\colon \Lambda_n \geqq B\}, \qquad B = \frac{A}{1-A}, \tag{2.10}$$

where $B > \pi_0/(1-\pi_0)$.

Obviously, inequality (2.4) holds in the general, non-i.i.d. case too. Consequently,

$$B_\alpha = \frac{1-\alpha}{\alpha} \qquad \text{implies} \quad \nu_{B_\alpha} \in \Delta(\alpha), \tag{2.11}$$

provided that $\alpha < 1 - \pi_0$.

It is worth noting that while the Shiryaev procedure (2.10) is optimal in the i.i.d. case (if $B$ is chosen so that $\mathrm{PFA}(\nu_B) = \alpha$), it may not be optimal in the non-i.i.d. scenario even if we can set the threshold to meet the PFA constraint exactly. In fact, the properties of the Shiryaev procedure in the non-i.i.d. scenario have not been investigated previously.

In addition to the Bayesian ADD defined in (2.1), we will also be interested in the behavior of the conditional ADD (CADD) for the fixed change-point $\lambda = k$, which is defined by $\mathrm{CADD}_k(\tau) = \mathbf{E}_k(\tau - k \,|\, \tau \geqq k)$, $k = 1, 2, \ldots$, as well as higher moments of the detection delay $\mathbf{E}_k\{(\tau - k)^m \,|\, \tau \geqq k\}$ and $\mathbf{E}^\pi\{(\tau - \lambda)^m \,|\, \tau \geqq \lambda\}$ for $m > 1$.

In the next section, we study the operating characteristics of the Bayesian procedure (2.10) for small PFA ($\alpha \to 0$) in the general, non-i.i.d. case. In section 4, these results will be specialized to the i.i.d. scenario.

**3. Asymptotic operating characteristics of the detection procedure $\boldsymbol{\nu_B}$ in a non-i.i.d. case.** As we mentioned above, in general, the Shiryaev procedure $\nu_B$ is not optimal even if one is able to chose the threshold $B$ in such a way that $\mathrm{PFA}(B) = \alpha$. However, below we show that this procedure with $B = B_\alpha = (1-\alpha)/\alpha$ is asymptotically optimal for small PFA under some mild conditions. We will show that in the asymptotic setting, $\nu_{B_\alpha}$ minimizes not only the ADD, but also $\mathrm{CADD}_k$ for all $k \geqq 1$. Furthermore, under certain general conditions this procedure minimizes higher moments of the detection delay up to a given order.

**3.1. Asymptotic lower bounds for moments of the detection delay.** We begin by establishing asymptotic lower bounds for moments of the detection delay, in particular for ADD and CADD of any procedure in the class $\Delta(\alpha)$. Later on, these bounds will be used to obtain asymptotic optimality results.

As we will see, the derivation of the lower bounds is based on the application of the Chebyshev inequality that involves certain probabilities, which are shown to go to 0 when $\alpha \to 0$. We start with the study of the behavior of these probabilities.

Let $q$ be a positive finite number and define $L_\alpha = L_\alpha(q_d) = |\log \alpha|/q_d$,

$$\gamma_{\varepsilon,\alpha}(\tau) = \mathbf{P}^{\pi}\big\{\lambda \leqq \tau < \lambda + (1-\varepsilon)\, L_\alpha\big\},$$
$$\gamma^{(k)}_{\varepsilon,\alpha}(\tau) = \mathbf{P}_k\big\{k \leqq \tau < k + (1-\varepsilon)\, L_\alpha\big\}, \qquad 0 < \varepsilon < 1,$$

where $q_d = q+d$ in the case of prior distributions with exponential right tail, $\pi \in \mathcal{E}(d)$, and $q_d = q_0 = q$ in the case of heavy-tailed prior distributions $\pi \in \mathcal{H}$.

The significance of the number $q$ is now explained in more detail. We do not assume any particular model for the observations, and as a result, there is no "structure" on the LLR process. We hence have to impose some conditions on the behavior of the LLR process at least for large $n$. It is natural to assume that there exists a positive finite number $q = q(f_1, f_0)$ such that $Z_n^k/(n-k)$ converges almost surely to $q$, i.e.,

$$\frac{1}{n}\, Z^k_{k+n} \xrightarrow[n\to\infty]{\mathbf{P}_k\text{-a.s.}} q \qquad \text{for every} \quad k < \infty. \tag{3.1}$$

This is always true for i.i.d. data models, in which case $q = D(f_1, f_0) = \mathbf{E}_1 Z_1^1$ is the Kullback–Leibler information number. It turns out that the a.s. convergence condition (3.1) is sufficient for obtaining lower bounds for all positive moments of the detection delay (but is not necessary). In fact, the key condition (3.2) in Lemma 1 and Theorem 1 holds whenever $Z_n^k/(n-k)$ converges almost surely to the number $q$. Therefore, this number is of paramount importance in the general change-point detection theory.

Note that $\mathbf{P}_k\{\tau < k\} = \mathbf{P}_\infty\{\tau < k\}$, since $\{\tau < k\}$ belongs to the sigma-field $\mathcal{F}^X_{k-1}$, and hence,

$$\mathrm{PFA}(\tau) = \sum_{k=1}^{\infty} \pi_k \mathbf{P}_\infty\{\tau < k\}.$$

The following "fundamental" lemma will be used repeatedly to derive lower bounds for the performance indices.

LEMMA 1. *Let $Z_n^k$ be defined as in* (2.5) *and assume that for some* $q > 0$,

$$\mathbf{P}_k\left\{\frac{1}{M} \max_{0 \leqq n < M} Z^k_{k+n} \geqq (1+\varepsilon)\, q\right\} \xrightarrow[M\to\infty]{} 0 \qquad \text{for all } \varepsilon > 0 \text{ and } k \geqq 1. \tag{3.2}$$

*Then, for all* $0 < \varepsilon < 1$ *and* $k \geqq 1$,

$$\lim_{\alpha\to 0} \sup_{\tau \in \Delta(\alpha)} \gamma^{(k)}_{\varepsilon,\alpha}(\tau) = 0, \tag{3.3}$$

*and for all* $0 < \varepsilon < 1$,

$$\lim_{\alpha\to 0} \sup_{\tau \in \Delta(\alpha)} \gamma_{\varepsilon,\alpha}(\tau) = 0. \tag{3.4}$$

*Proof.* Changing the measure $\mathbf{P}_\infty \to \mathbf{P}_k$, we obtain that for any $C > 0$ and $\varepsilon \in (0,1)$,

$$\begin{aligned}
\mathbf{P}_\infty\{k \leqq \tau < k + (1-\varepsilon)\, L_\alpha\} &= \mathbf{E}_k\Big\{ \mathbb{1}_{\{k \leqq \tau < k+(1-\varepsilon)\, L_\alpha\}}\, e^{-Z_\tau^k} \Big\} \\
&\geqq \mathbf{E}_k\Big\{ \mathbb{1}_{\{k \leqq \tau < k+(1-\varepsilon)\, L_\alpha, Z_\tau^k < C\}}\, e^{-Z_\tau^k} \Big\} \\
&\geqq e^{-C}\mathbf{P}_k\Big\{ k \leqq \tau < k+(1-\varepsilon)\, L_\alpha,\ \max_{k \leqq n < k+(1-\varepsilon)\, L_\alpha} Z_n^k < C \Big\} \\
&\geqq e^{-C}\Big[ \mathbf{P}_k\big\{k \leqq \tau < k+(1-\varepsilon)\, L_\alpha\big\} - \mathbf{P}_k\Big\{ \max_{0 \leqq n < (1-\varepsilon)\, L_\alpha} Z_{k+n}^k \geqq C \Big\} \Big], \qquad (3.5)
\end{aligned}$$

where the last inequality follows trivially from the fact that for any events $A$ and $B$ (with $B^c$ being a complement to $B$), $\mathbf{P}(A \cap B) \geqq \mathbf{P}(A) - \mathbf{P}(B^c)$.

Setting $C = (1-\varepsilon^2)\, qL_\alpha$, we obtain

$$\begin{aligned}
\gamma_{\varepsilon,\alpha}^{(k)}(\tau) &\leqq e^{(1-\varepsilon^2) qL_\alpha}\mathbf{P}_\infty\big\{k \leqq \tau < k+(1-\varepsilon)\, L_\alpha\big\} \\
&\quad + \mathbf{P}_k\Big\{ \max_{0 \leqq n < (1-\varepsilon)\, L_\alpha} Z_{k+n}^k \geqq (1-\varepsilon^2)\, qL_\alpha \Big\}. \qquad (3.6)
\end{aligned}$$

Write

$$p_k(\alpha,\varepsilon) = e^{(1-\varepsilon^2)\, qL_\alpha}\mathbf{P}_\infty\big\{k \leqq \tau < k+(1-\varepsilon)\, L_\alpha\big\}$$

and

$$\beta_k(\alpha,\varepsilon) = \mathbf{P}_k\Big\{ \max_{0 \leqq n < (1-\varepsilon)\, L_\alpha} Z_{k+n}^k \geqq (1-\varepsilon^2)\, qL_\alpha \Big\}.$$

By condition (3.2), for every $0 < \varepsilon < 1$ and all $k \geqq 1$,

$$\beta_k(\alpha,\varepsilon) = \mathbf{P}_k\Big\{ \frac{1}{(1-\varepsilon)\, L_\alpha} \max_{0 \leqq n < (1-\varepsilon)\, L_\alpha} Z_{k+n}^k \geqq (1+\varepsilon)\, q \Big\} \xrightarrow[\alpha \to 0]{} 0. \qquad (3.7)$$

Next, for any $\tau \in \Delta(\alpha)$ and $n \geqq 1$,

$$\alpha \geqq \mathrm{PFA}(\tau) \geqq \mathbf{P}\{\tau < n\} \cap \{\lambda > n\} = \mathbf{P}\big\{\tau < n \,|\, \lambda > n\big\}\, \mathbf{P}\{\lambda > n\} = \mathbf{P}_\infty\{\tau < n\}\, \Pi_n,$$

and, therefore,

$$\mathbf{P}_\infty\{\tau < n\} \leqq \alpha(\Pi_n)^{-1}, \qquad n \geqq 1. \qquad (3.8)$$

It follows that the first term in inequality (3.6),

$$p_k(\alpha,\varepsilon) \leqq \alpha e^{(1-\varepsilon^2)\, qL_\alpha}(\Pi_{m_\alpha(k)})^{-1}, \qquad (3.9)$$

where $m_\alpha(k) = \lfloor k + (1-\varepsilon)\, L_\alpha \rfloor$ is the greatest integer number $\leqq k + (1-\varepsilon)\, L_\alpha$. Since $\alpha = e^{-q_d L_\alpha}$ and $(m_\alpha(k) - k - 1)/(1-\varepsilon) \leqq L_\alpha \leqq (m_\alpha(k) - k)/(1-\varepsilon)$, we obtain

$$p_k(\alpha,\varepsilon) \leqq \big(\Pi_{m_\alpha(k)}\big)^{-1} \exp\Big\{ -\frac{d+\varepsilon^2 q}{1-\varepsilon}\, (m_\alpha(k) - k - 1) \Big\}$$

and

$$\frac{\log p_k(\alpha,\varepsilon)}{m_\alpha(k)} \leqq -\frac{\log \Pi_{m_\alpha(k)}}{m_\alpha(k)} - \frac{d+\varepsilon^2 q}{1-\varepsilon}\, \frac{m_\alpha(k) - k - 1}{m_\alpha(k)}.$$

By conditions (2.6) and (2.7),

$$\lim_{\alpha\to 0}\frac{\log p_k(\alpha,\varepsilon)}{m_\alpha(k)} \leqq d - \frac{d+\varepsilon^2 q}{1-\varepsilon} = -\frac{\varepsilon}{1-\varepsilon}\,(d+\varepsilon q),$$

where $d > 0$ for $\pi \in \mathcal{E}(d)$ and $d = 0$ for $\pi \in \mathcal{H}$. It follows that, for any $\pi \in \mathcal{H} \cup \mathcal{E}(d)$,

$$p_k(\alpha,\varepsilon) \to 0 \qquad \text{as } \alpha \to 0 \text{ for all } k \geqq 1 \text{ and } 0 < \varepsilon < 1. \tag{3.10}$$

Therefore, we obtain that for every $\tau \in \Delta(\alpha)$ and $\varepsilon > 0$,

$$\gamma^{(k)}_{\varepsilon,\alpha}(\tau) \leqq p_k(\alpha,\varepsilon) + \beta_k(\alpha,\varepsilon), \tag{3.11}$$

where by (3.7) and (3.10), $\beta_k(\alpha,\varepsilon)$ and $p_k(\alpha,\varepsilon)$ go to zero as $\alpha \to 0$. Since $p_k(\alpha,\varepsilon)$ and $\beta_k(\alpha,\varepsilon)$ do not depend on a particular stopping time $\tau$, (3.3) holds.

Let $N_\alpha = \lfloor \varepsilon L_\alpha \rfloor$ be the greatest integer number less than or equal to $\varepsilon L_\alpha$. Obviously,

$$\gamma_{\varepsilon,\alpha}(\tau) = \sum_{k=1}^{\infty} \pi_k \gamma^{(k)}_{\varepsilon,\alpha}(\tau) \leqq \sum_{k=1}^{N_\alpha} \pi_k \gamma^{(k)}_{\varepsilon,\alpha}(\tau) + \Pi_{N_\alpha}. \tag{3.12}$$

Using (3.11) and (3.12), we obtain

$$\gamma_{\varepsilon,\alpha}(\tau) \leqq \Pi_{N_\alpha} + \sum_{k=1}^{N_\alpha} \pi_k \beta_k(\alpha,\varepsilon) + \sup_{k \leqq N_\alpha} p_k(\alpha,\varepsilon). \tag{3.13}$$

The term $\Pi_{N_\alpha} \to 0$ as $\alpha \to 0$. The second term goes to zero as $\alpha \to 0$ by condition (3.2) (see (3.7)) and Lebesgue's dominated convergence theorem. It suffices to show that the third term vanishes as $\alpha \to 0$.

By (3.9),

$$\sup_{k \leqq N_\alpha} p_k(\alpha,\varepsilon) \leqq \alpha e^{(1-\varepsilon^2)\,qL_\alpha}\left(\Pi_{m_\alpha(N_\alpha)}\right)^{-1} = \left(\Pi_{m_\alpha(N_\alpha)}\right)^{-1}\exp\left\{-(d+\varepsilon^2 q)\,L_\alpha\right\},$$

where $m_\alpha(N_\alpha) = \lfloor N_\alpha + (1-\varepsilon)\,L_\alpha \rfloor$. Obviously, $L_\alpha \leqq m_\alpha(N_\alpha) \leqq L_\alpha + 2$, and hence,

$$\frac{\log \sup_{k \leqq N_\alpha} p_k(\alpha,\varepsilon)}{m_\alpha(N_\alpha)} \leqq -\frac{\log \Pi_{m_\alpha(N_\alpha)}}{m_\alpha(N_\alpha)} - (d+\varepsilon^2 q)\,\frac{L_\alpha}{L_\alpha + 2}.$$

Since

$$-\frac{\log \Pi_{m_\alpha(N_\alpha)}}{m_\alpha(N_\alpha)} \to d \qquad \text{as} \quad \alpha \to 0,$$

we obtain that for any $\pi \in \mathcal{H} \cup \mathcal{E}(d)$,

$$\lim_{\alpha\to 0}\frac{\log \sup_{k \leqq N_\alpha} p_k(\alpha,\varepsilon)}{m_\alpha(N_\alpha)} \leqq -\varepsilon^2 q,$$

showing that $\sup_{k \leqq N_\alpha} p_k(\alpha,\varepsilon) \to 0$ as $\alpha \to 0$.

Since the right-hand side in (3.13) does not depend on $\tau$, (3.4) follows, and the proof is complete.

In the next theorem, we derive the lower bounds for the positive moments of the detection delay of any procedure from the class $\Delta(\alpha)$ under the conditions postulated in Lemma 1. A similar bound has been obtained by Lai [16] for $\mathbf{E}^\pi(\tau-\lambda)^+$ and heavy-tailed prior distributions under a slightly stronger condition with $\sup_k$ in (3.2). Recall that $q_d = q+d$ when $\pi \in \mathcal{E}(d)$ and $q_d = q$ when $\pi \in \mathcal{H}$.

THEOREM 1. *Let condition* (3.2) *hold for some positive finite number* $q$. *Then, for every* $k \geqq 1$ *and* $m > 0$,

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{E}_k\big[(\tau-k)^m \mid \tau \geqq k\big] \geqq \left\{\frac{|\log\alpha|}{q_d}\right\}^m \big(1+o(1)\big) \qquad \text{as} \quad \alpha \to 0, \tag{3.14}$$

*and for all* $m > 0$,

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{E}^\pi\big[(\tau-\lambda)^m \mid \tau \geqq \lambda\big] \geqq \left\{\frac{|\log\alpha|}{q_d}\right\}^m \big(1+o(1)\big) \qquad \text{as} \quad \alpha \to 0, \tag{3.15}$$

*where* $o(1) \to 0$ *as* $\alpha \to 0$.

*Proof.* By the Chebyshev inequality, for any $0<\varepsilon<1$ and $m>0$,

$$\mathbf{E}_k\big[(\tau-k)^+\big]^m \geqq \big[(1-\varepsilon)\,L_\alpha\big]^m \mathbf{P}_k\big\{\tau-k \geqq (1-\varepsilon)\,L_\alpha\big\}.$$

Obviously,

$$\mathbf{P}_k\big\{\tau-k \geqq (1-\varepsilon)\,L_\alpha\big\} = \mathbf{P}_k\{\tau \geqq k\} - \gamma_{\varepsilon,\alpha}^{(k)}(\tau).$$

Therefore,

$$\mathbf{E}_k\big[(\tau-k)^m \mid \tau \geqq k\big] = \frac{\mathbf{E}_k[(\tau-k)^+]^m}{\mathbf{P}_k\{\tau\geqq k\}} \geqq \frac{[(1-\varepsilon)\,L_\alpha]^m}{\mathbf{P}_k\{\tau\geqq k\}} \left[\mathbf{P}_k\{\tau \geqq k\} - \gamma_{\varepsilon,\alpha}^{(k)}(\tau)\right]. \tag{3.16}$$

Next, by (3.8), for any $\tau \in \Delta(\alpha)$ and $\alpha < \Pi_k$,

$$\mathbf{P}_k\{\tau \geqq k\} = 1 - \mathbf{P}_\infty\{\tau < k\} \geqq 1 - \alpha(\Pi_k)^{-1}. \tag{3.17}$$

Using (3.16) and (3.17) yields the inequality

$$\mathbf{E}_k\big[(\tau-k)^m \mid \tau \geqq k\big] \geqq \big[(1-\varepsilon)\,L_\alpha\big]^m \left[1 - \frac{\gamma_{\varepsilon,\alpha}^{(k)}(\tau)}{1-\alpha(\Pi_k)^{-1}}\right], \tag{3.18}$$

which holds for any $\tau \in \Delta(\alpha)$, $0<\varepsilon<1$, and $\{k\colon\ \Pi_k < \alpha\}$. By Lemma 1, $\gamma_{\varepsilon,\alpha}^{(k)}(\tau) \to 0$ as $\alpha \to 0$ uniformly over all $\tau \in \Delta(\alpha)$, which implies

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{E}_k\big[(\tau-k)^m \mid \tau \geqq k\big] \geqq \big[(1-\varepsilon)\,L_\alpha\big]^m \big(1+o(1)\big) \qquad \text{as} \quad \alpha \to 0.$$

Since $\varepsilon$ is arbitrary, the asymptotic lower bound (3.14) follows.

To prove the asymptotic lower bound (3.15) we again use the Chebyshev inequality, according to which, for any $0<\varepsilon<1$ and any $\tau \in \Delta(\alpha)$,

$$\mathbf{E}^\pi\big[(\tau-\lambda)^+\big]^m \geqq \big[(1-\varepsilon)\,L_\alpha\big]^m \mathbf{P}^\pi\big\{\tau-\lambda \geqq (1-\varepsilon)\,L_\alpha\big\},$$

where

$$\mathbf{P}^\pi\big\{\tau-\lambda \geqq (1-\varepsilon)\,L_\alpha\big\} = \mathbf{P}^\pi\{\tau \geqq \lambda\} - \gamma_{\varepsilon,\alpha}(\tau).$$

Since $\mathbf{P}^\pi\{\tau \geqq \lambda\} \geqq 1-\alpha$ for any $\tau \in \Delta(\alpha)$, it follows that

$$\mathbf{E}^\pi\left[(\tau-\lambda)^m \mid \tau \geqq \lambda\right] = \frac{\mathbf{E}^\pi[(\tau-\lambda)^+]^m}{\mathbf{P}^\pi\{\tau \geqq \lambda\}} \geqq \left[(1-\varepsilon)\, L_\alpha\right]^m \left[1 - \frac{\gamma_{\varepsilon,\alpha}(\tau)}{\mathbf{P}^\pi\{\tau \geqq \lambda\}}\right]$$
$$\geqq \left[(1-\varepsilon)\, L_\alpha\right]^m \left[1 - \frac{\gamma_{\varepsilon,\alpha}(\tau)}{1-\alpha}\right]. \tag{3.19}$$

Since $\varepsilon$ can be arbitrarily small and, by Lemma 1, $\sup_{\tau \in \Delta(\alpha)} \gamma_{\varepsilon,\alpha}(\tau) \to 0$ as $\alpha \to 0$, the asymptotic lower bound (3.15) follows, and the proof is complete.

*Remark* 1. It is important to emphasize that the vanishing term $o(1)$ in (3.14) depends on $k$. For this reason, the lower bound (3.15) does not follow directly from inequality (3.14), and an additional effort was needed to prove it.

**3.2. Asymptotic performance of the procedure $\boldsymbol{\nu_B}$ for large $\boldsymbol{B}$.** We begin with the evaluation of the performance of the Shiryaev detection procedure $\nu_B$ for large values of $B$ regardless of the false alarm constraint.

Write

$$Y_t = \log \frac{f_{1,t}(X_t \mid \mathbf{X}^{t-1})}{f_{0,t}(X_t \mid \mathbf{X}^{t-1})},$$

and, for every $k = 1, 2, \dots$ and $\varepsilon > 0$, define the random variable

$$T_\varepsilon^{(k)} = \sup\left\{n \geqq 1\colon \left|\frac{1}{n}\sum_{t=k}^{k+n-1} Y_t - q\right| > \varepsilon\right\}, \tag{3.20}$$

where $\sup \varnothing = 0$. Clearly, in terms of $T_\varepsilon^{(k)}$, the almost sure convergence of (3.1) may be written as $\mathbf{P}_k\{T_\varepsilon^{(k)} < \infty\} = 1$ for all $\varepsilon > 0$ and $k \geqq 1$, which implies condition (3.2).

While these conditions are sufficient for obtaining lower bounds for moments of the detection delay (in particular, for the average detection delay), they need to be strengthened in order to establish asymptotic optimality properties of the detection procedure $\nu_B$, and to obtain asymptotic expansions for moments of the detection delay. Indeed, in general these conditions do not even guarantee finiteness of $\mathrm{CADD}_k(\nu_B)$ and $\mathrm{ADD}(\nu_B)$. In order to study asymptotics for the average detection delay, one may impose the following constraints on the rate of convergence in the strong law for $Z_{k+n}^k/n$:

$$\mathbf{E}_k T_\varepsilon^{(k)} < \infty \qquad \text{for all } \ \varepsilon > 0 \text{ and } k \geqq 1, \quad \text{and} \tag{3.21}$$

$$\sum_{k=1}^\infty \pi_k \mathbf{E}_k T_\varepsilon^{(k)} < \infty \qquad \text{for all } \ \varepsilon > 0. \tag{3.22}$$

Note that (3.21) is closely related to the condition

$$\sum_{n=1}^\infty \mathbf{P}_k\left\{\left|\sum_{t=k}^{k+n-1} Y_t - qn\right| > \varepsilon n\right\} < \infty \qquad \text{for all } \varepsilon > 0 \text{ and } k \geqq 1,$$

which is nothing but the complete convergence of $Z_{k+n}^k/n$ to $q$ under $\mathbf{P}_k$ (cf. [11]). We write this compactly as

$$\frac{1}{n} Z_{k+n}^k \xrightarrow[n\to\infty]{\mathbf{P}_k\text{-completely}} q \qquad \text{for every} \quad k \geqq 1. \tag{3.23}$$

The convergence condition (3.22) is a joint condition on the rates of convergence of $Z^k_{k+n}/n$ for each $\lambda = k$ and the prior distribution. We write this condition compactly as

$$\frac{1}{n} Z^{\lambda}_{\lambda+n} \xrightarrow[n\to\infty]{\mathbf{P}^{\pi}\text{-average-completely}} q. \tag{3.24}$$

To study asymptotics for higher moments of the detection delay, the complete convergence conditions (3.23) and (3.24) should be further strengthened. A natural generalization is to require, for some $r \geqq 1$,

$$\mathbf{E}_k [T^{(k)}_{\varepsilon}]^r < \infty \qquad \text{for all } \ \varepsilon > 0 \text{ and } k \geqq 1 \tag{3.25}$$

and

$$\sum_{k=1}^{\infty} \pi_k \mathbf{E}_k [T^{(k)}_{\varepsilon}]^r < \infty \qquad \text{for all} \quad \varepsilon > 0. \tag{3.26}$$

If (3.25) holds, it is said that $Z^k_{k+n}/n$ converges $r$-quickly to $q$ (cf. [13]). If (3.26) holds, we will say that $Z^{\lambda}_{\lambda+n}/n$ converges average-$r$-quickly to $q$. We will write these modes of convergence compactly as

$$\frac{1}{n} Z^{k}_{k+n} \xrightarrow[n\to\infty]{\mathbf{P}_k\text{-}r\text{-quickly}} q \qquad \text{for every} \quad k \geqq 1 \tag{3.27}$$

and

$$\frac{1}{n} Z^{\lambda}_{\lambda+n} \xrightarrow[n\to\infty]{\mathbf{P}^{\pi}\text{-average-}r\text{-quickly}} q. \tag{3.28}$$

Complete and $r$-quick convergence conditions were previously used by Lai [14], Tartakovsky [33], and Dragalin, Tartakovsky, and Veeravalli [8] to establish the asymptotic optimality of sequential hypothesis tests for general statistical models. Below we take advantage of these results and prove that the conditions (3.23), (3.24), (3.27), and (3.28) are sufficient for asymptotic optimality of the Shiryaev change-point detection procedure.

In the following theorem, we establish the operating characteristics of the detection procedure $\nu_B$ for large values of the threshold $B$ regardless of the false alarm rate constraints for general statistical models when the $r$-quick convergence conditions (3.27) and (3.28) hold. Hereafter $X_B \sim Y_B$ as $B \to \infty$ means that

$$\lim_{B\to\infty} \left( \frac{X_B}{Y_B} \right) = 1.$$

For the sake of compactness, in the rest of the paper we will write $\mathbf{ED}^{\pi}_m(\tau)$ for $\mathbf{E}^{\pi}[(\tau-\lambda)^m \mid \tau \geqq \lambda]$ and $\mathbf{ED}^{(k)}_m(\tau)$ for $\mathbf{E}_k[(\tau-k)^m \mid \tau \geqq k]$, respectively.

THEOREM 2. (i) *Let condition* (3.27) *hold for some positive $q$ and $r \geqq 1$. Then for all $m \leqq r$,*

$$\mathbf{ED}^{(k)}_m(\nu_B) \sim \left( \frac{\log B}{q_d} \right)^m \qquad \textit{as } B \to \infty \textit{ for all } k \geqq 1. \tag{3.29}$$

(ii) *Let condition* (3.28) *hold for some positive $q$ and $r \geqq 1$. Then for all $m \leqq r$,*

$$\mathbf{ED}^{\pi}_m(\nu_B) \sim \left( \frac{\log B}{q_d} \right)^m \qquad \textit{as} \quad B \to \infty. \tag{3.30}$$

To prove this theorem we will need two auxiliary results that we formulate in the form of Lemmas 2 and 3 below. Lemma 2 is similar to, and to some extent is a particular case of, Lemma 1. Lemma 3 is related to convergence of moments of one-sided stopping times that bound the stopping time $\nu_B$ from above. The first lemma will be used to obtain lower bounds for the moments of the detection delay, while the second one will be used for deriving the corresponding upper bounds.

The following notation is used throughout this subsection:

$$\begin{aligned} L_B &= q_d^{-1} \log B, \\ \gamma_\varepsilon^{(k)}(B) &= \mathbf{P}_k\{k \leqq \nu_B < k + (1-\varepsilon)\, L_B\}, \\ \gamma_\varepsilon(B) &= \mathbf{P}^\pi\bigl\{\lambda \leqq \nu_B < \lambda + (1-\varepsilon)\, L_B\bigr\}. \end{aligned}$$

Note that

$$\gamma_\varepsilon(B) = \sum_{k=1}^{\infty} \pi_k \gamma_\varepsilon^{(k)}(B).$$

Lemma 2. *Suppose condition* (3.2) *holds for some* $q > 0$. *Then*

$$\lim_{B\to\infty} \gamma_\varepsilon^{(k)}(B) = 0 \qquad \textit{for all} \quad 0 < \varepsilon < 1 \textit{ and } k \geqq 1; \tag{3.31}$$

$$\lim_{B\to\infty} \gamma_\varepsilon(B) = 0 \qquad \textit{for every} \quad 0 < \varepsilon < 1. \tag{3.32}$$

*Proof.* The proof runs along the lines of the proof of Lemma 1. It suffices to note that $\mathbf{P}^\pi\{\nu_B < \lambda\} \leqq 1/(1+B) \leqq 1/B$. Therefore, replacing $\alpha$ with $1/B$ in Lemma 1 completes the proof.

We now formulate the second important result that will be used to obtain upper bounds for the moments of the stopping time $\nu_B$. Write

$$S_{k+n-1}^k = Z_{k+n-1}^k + n w_{n,k}, \qquad w_{n,k} = n^{-1} \log\left(\pi_k \left(\Pi_{k+n-1}\right)^{-1}\right),$$

and, for $b > 0$, introduce the sequence of one-sided stopping times

$$\eta_b(k) = \inf\{n \geqq 1\colon\ S_{k+n-1}^k \geqq b\}, \qquad k = 1, 2, \dots\ . \tag{3.33}$$

Lemma 3. *Let $m$ be a positive, not necessarily integer number and suppose that for some $r > 0$ condition* (3.25) *is satisfied. Then, for all $m \leqq r$,*

$$\mathbf{E}_k\bigl(\eta_b(k)\, b^{-1}\bigr)^m \to (q_d)^{-m} \qquad \textit{as} \quad b \to \infty. \tag{3.34}$$

*Proof.* For the geometric prior distribution with $\pi_0 = 0$, this lemma can be directly derived from Theorem 4.2 of [8], since in this case $(\log \Pi_n)/n = d = |\log(1-\rho)|$ and $w_{n,k} = n\log(1-\rho)$ for all $n \geqq 1$. In the general case, the proof requires a modification, which is given below.

By (3.25), $Z_{k+n}^k/n \to q$ as $n \to \infty$ $\mathbf{P}_k$-a.s. Since, by assumptions (2.6)–(2.7), $w_{n,k} \to d$ as $n \to \infty$ $(d \geqq 0)$, it follows that $S_{k+n}^k/n \to q_d$ as $n \to \infty$ $\mathbf{P}_k$-a.s. for all $k \geqq 1$. Therefore,

$$\mathbf{P}_k\left\{\frac{1}{M} \max_{0 \leqq n < M} S_{k+n}^k \geqq (1+\varepsilon)\, q_d\right\} \xrightarrow[M\to\infty]{} 0 \qquad \text{for all} \quad \varepsilon > 0 \text{ and } k \geqq 1$$

and the argument identical to that used in the proof of Lemma 1 (with $\alpha$ replaced by $e^{-b}$) yields

$$\mathbf{P}_k\big\{\eta_b(k) \geqq (1-\varepsilon)\, b(q_d)^{-1}\big\} \to 1 \qquad \text{as} \quad b \to \infty.$$

Hence, Chebyshev's inequality (for any $m > 0$)

$$\mathbf{E}_k\big[\eta_b(k)\big]^m \geqq \big[(1-\varepsilon)\, b(q_d)^{-1}\big]^m \mathbf{P}_k\big\{\eta_b(k) \geqq (1-\varepsilon)\, b(q_d)^{-1}\big\}$$

applies to show that

$$\mathbf{E}_k\big[\eta_b(k)\big]^m \geqq (b(q_d)^{-1})^m\big(1+o(1)\big) \qquad \text{as} \quad b \to \infty. \tag{3.35}$$

To obtain the upper bound, define

$$\widetilde{T}_\varepsilon^{(k)} = \sup\Big\{n \geqq 1\colon\ \big|n^{-1} S_{k+n-1}^k - q_d\big| > \varepsilon\Big\}.$$

It is easy to see that $S^k_{k+\eta_b(k)-2} < b$ and

$$S^k_{k+\eta_b(k)-2} \geqq (\eta_b(k)-1)(q_d - \varepsilon) \qquad \text{on} \quad \big\{\eta_b(k) - 1 > \widetilde{T}_\varepsilon^{(k)}\big\}.$$

It follows that for every $0 < \varepsilon < q_d$,

$$\begin{aligned} \eta_b(k) &\leqq 1 + \frac{b}{q_d - \varepsilon}\, \mathbb{1}_{\{\eta_b(k) > \widetilde{T}_\varepsilon^{(k)}+1\}} + (\widetilde{T}_\varepsilon^{(k)} + 1)\, \mathbb{1}_{\{\eta_b(k) \leqq \widetilde{T}_\varepsilon^{(k)}+1\}} \\ &\leqq \widetilde{T}_\varepsilon^{(k)} + 2 + \frac{b}{q_d - \varepsilon}. \end{aligned} \tag{3.36}$$

Since $w_{n,k} \to d$ as $n \to \infty$, the condition (3.25) implies $\mathbf{E}_k[\widetilde{T}_\varepsilon^{(k)}]^r < \infty$. Since $\varepsilon$ can be arbitrarily small, letting $\varepsilon \to 0$ we obtain that for all $m \leqq r$,

$$\mathbf{E}_k[\eta_b(k)]^m \leqq (b(q_d)^{-1})^m\big(1+o(1)\big) \qquad \text{as} \quad b \to \infty,$$

which, along with the reverse inequality (3.35), completes the proof.

We are now ready to prove Theorem 2.

*Proof of Theorem* 2. (i) For $b = \log B$, define the sequence of stopping times $\eta_b(k)$ as in (3.33). It is easily verified that the statistic $\log \Lambda_n$ can be written in the form

$$\log \Lambda_n = Z_n^k + (n-k+1) w_{k+n-1,k} + \ell_{n,k}, \tag{3.37}$$

where the random variable $\ell_{n,k}$ is nonnegative. Thus, for $b = \log B$ and for every $k \geqq 1$,

$$\nu_B - k \leqq \eta_b(k) \qquad \text{on} \quad \{\nu_B \geqq k\} \tag{3.38}$$

and

$$\mathbf{E}_k\big[(\nu_B - k)^+\big]^m \leqq \mathbf{E}_k\big\{[\eta_b(k)]^m;\ \nu_B \geqq k\big\} \leqq \mathbf{E}_k\big[\eta_b(k)\big]^m. \tag{3.39}$$

By (2.4) and (3.8), $\mathbf{P}_k\{\nu_B \geqq k\} \geqq 1 - [(1+B)\,\Pi_k]^{-1}$, and hence, for $B > (1-\Pi_k)/\Pi_k$,

$$\mathbf{ED}_m^{(k)}(\nu_B) = \frac{\mathbf{E}_k[(\nu_B - k)^+]^m}{\mathbf{P}_k\{\nu_B \geqq k\}} \leqq \frac{\mathbf{E}_k[\eta_b(k)]^m}{1 - [(1+B)\,\Pi_k]^{-1}}.$$

Applying Lemma 3, we obtain the following estimate for the upper bound:

$$\mathbf{ED}_m^{(k)}(\nu_B) \leqq \left(\frac{\log B}{q_d}\right)^m \big(1+o(1)\big) \qquad \text{as } B \to \infty \text{ for all } k \geqq 1. \tag{3.40}$$

In order to prove (3.29), it remains to show that the right-hand side of the latter inequality is a lower bound for $\mathbf{ED}_m^{(k)}(\nu_B)$ as $B \to \infty$. To that end, we use Lemma 2 and the Chebyshev inequality.

Indeed, applying the Chebyshev inequality and the inequality $\mathbf{P}_k\{\nu_B \geqq k\} \geqq 1 - [(1+B)\,\Pi_k]^{-1}$ yields, for $B > (1-\Pi_k)/\Pi_k$,

$$\begin{aligned}\mathbf{ED}_m^{(k)}(\nu_B) &\geqq \big[(1-\varepsilon)\,L_B\big]^m \Big[\mathbf{P}_k\{\nu_B \geqq k\} - \gamma_\varepsilon^{(k)}(B)\Big] \\ &\geqq \big[(1-\varepsilon)\,L_B\big]^m \big[1 - (\Pi_k B)^{-1} - \gamma_\varepsilon^{(k)}(B)\big].\end{aligned}$$

Since $\varepsilon$ is arbitrary and, by Lemma 2, $\gamma_\varepsilon^{(k)}(B) \to 0$ as $B \to \infty$ for every $k \geqq 1$, we obtain that the right-hand side in (3.40) is the asymptotic lower bound for $\mathbf{ED}_m^{(k)}(\nu_B)$. The asymptotic formula (3.29) follows.

(ii) Similar to (3.36), for every $0 < \varepsilon < q_d$ and $k \geqq 1$,

$$\nu_B - k \leqq \eta_b(k) \leqq \widetilde{T}_\varepsilon^{(k)} + 2 + \frac{b}{q_d - \varepsilon} \qquad \text{on} \quad \{\nu_B \geqq k\}. \tag{3.41}$$

Applying (3.41) along with the fact that $\mathbf{P}^\pi\{\nu_B \geqq \lambda\} \geqq B/(1+B)$ yields

$$\mathbf{ED}_m^\pi(\nu_B) \leqq \frac{B+1}{B} \sum_{k=1}^\infty \pi_k \mathbf{E}_k \left(\frac{\log B}{q_d - \varepsilon} + \widetilde{T}_\varepsilon^{(k)} + 2\right)^m.$$

Since $\varepsilon$ can be arbitrarily small and, by the assumption of the theorem,

$$\sum_{k=1}^\infty \pi_k \mathbf{E}_k [T_\varepsilon^{(k)}]^m < \infty,$$

which, as before, implies $\sum_{k=1}^\infty \pi_k \mathbf{E}_k[\widetilde{T}_\varepsilon^{(k)}]^m < \infty$, it follows that

$$\mathbf{ED}_m^\pi(\nu_B) \leqq \left(\frac{\log B}{q_d}\right)^m \big(1+o(1)\big) \qquad \text{as} \quad B \to \infty. \tag{3.42}$$

To obtain a lower bound for $\mathbf{ED}_m^\pi(\nu_B)$, we again use the Chebyshev inequality, which yields

$$\begin{aligned}\mathbf{ED}_m^\pi(\nu_B) &\geqq \mathbf{E}^\pi\big[(\nu_B - \lambda)^+\big]^m \geqq \big[(1-\varepsilon)\,L_B\big]^m \left[\mathbf{P}^\pi\{\nu_B \geqq \lambda\} - \gamma_\varepsilon(B)\right] \\ &\geqq \big[(1-\varepsilon)\,L_B\big]^m \left[B(1+B)^{-1} - \gamma_\varepsilon(B)\right],\end{aligned}$$

where $\gamma_\varepsilon(B) \to 0$ as $B \to \infty$ by Lemma 2. Recall that $L_B = (\log B)/q_d$. Since $\varepsilon$ is arbitrary, it follows that

$$\mathbf{ED}_m^\pi(\nu_B) \geqq \left(\frac{\log B}{q_d}\right)^m \big(1+o(1)\big) \qquad \text{as} \quad B \to \infty, \tag{3.43}$$

which, along with the upper bound (3.42), proves (3.30). The proof is complete.

*Remark* 2. The proof suggests that (3.29) holds for all $k < K_\alpha$, where $K_\alpha$ is such an integer number for which $\mathbf{P}\{\lambda > K_\alpha\} < \alpha$. From a theoretical viewpoint this does not cause the problem, since $K_\alpha \to \infty$ as $\alpha \to 0$. However, from the point of view of practical applications it is important to investigate the behavior of $\mathbf{ED}_m^{(k)}(\nu_B)$ for large values of $k$.

**3.3. Asymptotic optimality.** We are now in a position to prove the asymptotic optimality of the detection procedure $\nu_B$ with the threshold $B = (1-\alpha)/\alpha$ in the class $\Delta(\alpha)$ for small values of the PFA $\alpha$. Since everything we need has been prepared, the proof is immediate.

THEOREM 3. *If $B = B_\alpha = (1-\alpha)/\alpha$, then the detection procedure $\nu_{B_\alpha}$ belongs to the class $\Delta(\alpha)$, and the following two assertions hold.*

(i) *Let condition* (3.27) *hold for some positive $q$ and $r \geqq 1$. Then for all $m \leqq r$ and $k \geqq 1$,*

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{ED}_m^{(k)}(\tau) \sim \mathbf{ED}_m^{(k)}(\nu_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{q_d}\right)^m \qquad as \quad \alpha \to 0. \tag{3.44}$$

(ii) *Let condition* (3.28) *hold for some positive $q$ and $r \geqq 1$. Then for all $m \leqq r$,*

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{ED}_m^{\pi}(\tau) \sim \mathbf{ED}_m^{\pi}(\nu_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{q_d}\right)^m \qquad as \quad \alpha \to 0. \tag{3.45}$$

*Proof.* Applying Theorems 1 and 2 yields (3.44) and (3.45).

COROLLARY 1. *Let $B = B_\alpha = (1-\alpha)/\alpha$. If condition* (3.23) *is satisfied for some positive $q$, then*

$$\inf_{\tau\in\Delta(\alpha)} \mathrm{CADD}_k(\tau) \sim \mathrm{CADD}_k(\nu_{B_\alpha}) \sim \frac{|\log\alpha|}{q_d} \qquad as\ \alpha \to 0\ for\ all\ k \geqq 1. \tag{3.46}$$

*If condition* (3.24) *is satisfied for some positive $q$, then*

$$\inf_{\tau\in\Delta(\alpha)} \mathrm{ADD}(\tau) \sim \mathrm{ADD}(\nu_{B_\alpha}) \sim \frac{|\log\alpha|}{q_d} \qquad as\ \alpha \to 0. \tag{3.47}$$

We stress that the detection procedure $\nu_B$ with the threshold $B = B_\alpha = (1-\alpha)/\alpha$ is not only asymptotically optimal relative to the ADD, but also uniformly asymptotically optimal with respect to the conditional ADD for all values of $\lambda = k$, $k = 1, 2, \ldots$. We obtain this strong optimality result primarily because the constraint on false alarms in the Bayesian formulation that we consider is averaged over all possible realizations of the change point. Such a strong optimality result is not available for the minimax formulation of the problem with the constraint on the mean time to false alarm $\mathbf{E}_\infty\tau$ [16], [17], [21], [34].

It is also interesting to observe from Theorem 3 that, for prior distributions with exponential tail ($d > 0$), if $q \gg d$, then the observations contain more information about the change than the prior distribution, and the performance is determined by $q$. On the other hand, if $q \ll d$, then the decision about the change point can be made based solely on the prior distribution to yield an ADD of $|\log\alpha|/d$. For heavy-tailed distributions $\pi \in \mathcal{H}$, prior information does not affect asymptotic performance, as could be expected.

*Remark* 3. The results of Theorem 3 remain true in a more general case, where the $r$-quick convergence condition (3.27) is satisfied with an increasing function $\phi(n)$ in place of $n$, i.e.,

$$\frac{1}{\phi(n)}\, Z_{k+n}^k \xrightarrow[n\to\infty]{\mathbf{P}_k\text{-}r\text{-quickly}} q \qquad \text{for all} \quad k \geqq 1. \tag{3.48}$$

In particular, if (3.48) holds with $\phi(n) = n^p$, $p > 0$, then, similar to (3.44), the following more general result can be proved:

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{ED}_m^{(k)}(\tau) \sim \mathbf{ED}_m^{(k)}(\nu_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{q_d}\right)^{m/p} \qquad \text{as} \quad \alpha \to 0.$$

It is worth noting that the $r$-quick convergence conditions (3.27) and (3.28) are only sufficient and by no means necessary. Indeed, the proof of Lemma 3 suggests that the upper bound holds whenever left-sided versions of $r$-quick conditions are satisfied, i.e., $\mathbf{E}_k[t_\varepsilon^{(k)}]^r < \infty$ and $\sum_{k=1}^\infty \pi_k \mathbf{E}_k[t_\varepsilon^{(k)}]^r < \infty$, where

$$t_\varepsilon^{(k)} = \sup\left\{n \geqq 1:\ n^{-1} Z_{k+n-1}^k - q < -\varepsilon\right\}.$$

Moreover, even these latter conditions can be substantially relaxed into the following condition:

$$\begin{gathered} \lim_{n\to\infty} n^{r-1} \mathbf{P}_k\left\{n^{-1} Z_{k+n-1}^k - q \leqq -\varepsilon)\right\} = 0 \\ \text{for all} \quad \varepsilon > 0,\ k \geqq 1,\ \text{and some } r \geqq 1. \end{gathered} \tag{3.49}$$

A proof can be built upon a generalization of a "trick" exploited recently by Lai [16] for studying the asymptotic optimality of the CUSUM and window-limited CUSUM tests.

However, we find it natural and convenient from the methodological standpoint to formulate conditions in terms of rates of convergence in the strong law of large numbers for the LLR process. In addition, $r$-quick convergence implies the right-tail condition (3.2), which is the key for obtaining the lower bounds.

**4. Asymptotic operating characteristics of the detection procedure $\boldsymbol{\nu_B}$ in the i.i.d. case.** In this section, we will deal with the i.i.d. case, where $f_{0,n}(X_n \mid \mathbf{X}^{n-1}) = f_0(X_n)$ and $f_{1,n}(X_n \mid \mathbf{X}^{n-1}) = f_1(X_n)$. Then $\mathbf{P}_\infty$ is the probability measure under which the PDF of $X_n$ is $f_0(x)$ for every $n \geqq 1$ and for $\lambda = k$, $k \geqq 1$; $\mathbf{P}_k$ is the probability measure under which the PDF of $X_n$ is $f_0(x)$ if $n \leqq k-1$ and is $f_1(x)$ if $n \geqq k$ (with respect to a $\sigma$-finite measure $\mu(x)$). The LLR defined in (2.5) is modified to

$$Z_n^k := \log \frac{d\mathbf{P}_k}{d\mathbf{P}_\infty}(X^n) = \sum_{t=k}^n \log \frac{f_1(X_t)}{f_0(X_t)}, \qquad n \geqq k, \tag{4.1}$$

and the decision statistic $\Lambda_n$ obeys (2.9) with $Z_n^k$ defined in (4.1).

Note also that in the i.i.d. case the statistic $\Lambda_n$ satisfies the recursion

$$\Lambda_n = \left(\frac{\Pi_{n-1}}{\Pi_n}\,\Lambda_{n-1} + \frac{\pi_n}{\Pi_n}\right)\frac{f_1(X_n)}{f_0(X_n)}, \qquad n \geqq 1, \quad \Lambda_0 = \frac{\pi_0}{1-\pi_0}, \tag{4.2}$$

which may be deployed for practical implementation and simulations.

As mentioned in section 2, in the i.i.d. case and for the geometric prior distribution, the Shiryaev procedure (2.10) is optimal when the threshold $B$ can be chosen in such a way that $\mathrm{PFA}(\nu_B) = \alpha$. Since it is difficult to meet this exact requirement, we will study the properties of the detection procedure $\nu_B$ with $B = (1-\alpha)/\alpha$, which guarantees the inequality $\mathrm{PFA}(\nu_B) \leqq \alpha$. Since this choice neglects the overshoot, it is expected that the actual PFA may be substantially smaller than $\alpha$ (see sections 4.2 and 6).

Let

$$D(f_1, f_0) = \mathbf{E}_1 Z_1^1 = \int \log\left(\frac{f_1(x)}{f_0(x)}\right) f_1(x)\,\mu(dx)$$

be the Kullback–Leibler (KL) information number between densities $f_1(x)$ and $f_0(x)$ (also called the KL "distance"). In the i.i.d. case, the KL number $D(f_1, f_0)$ plays the role of the number $q$ that appeared in Theorems 1–3 of the previous section.

By analogy with (3.20) we define the last entry times

$$T_\varepsilon^{(k)} = \sup\left\{ n \geqq 1\colon \left| \frac{1}{n} \sum_{t=k}^{k+n-1} Y_t - D(f_1, f_0) \right| > \varepsilon \right\}. \tag{4.3}$$

Since $Y_t$, $t = 1, 2, \dots$, are i.i.d. random variables, $T_\varepsilon^{(k)}$ have the same statistical properties for all $k \geqq 1$. Therefore, $\mathbf{E}_k[T_\varepsilon^{(k)}]^r = \mathbf{E}_1[T_\varepsilon^{(1)}]^r$ and the condition $\mathbf{E}_1[T_\varepsilon^{(1)}]^r < \infty$ for all $\varepsilon > 0$ (i.e., the $r$-quick convergence condition of $Z_n^1/n$ to $D(f_1, f_0)$ under $\mathbf{P}_1$) implies

$$\sum_{k=1}^{\infty} \pi_k \mathbf{E}_k \left[T_\varepsilon^{(k)}\right]^r = (1-\pi_0)\,\mathbf{E}_1\left[T_\varepsilon^{(1)}\right]^r < \infty \qquad \text{for all} \quad \varepsilon > 0. \tag{4.4}$$

In the i.i.d. case, the condition $\mathbf{E}_1|Z_1^1|^{r+1} < \infty$ is both necessary and sufficient for the $r$-quick convergence of $Z_n^1/n$ to $D(f_1, f_0)$. Indeed, by the Baum–Katz rate of convergence in the strong law [2], the following statements are equivalent for any $r > 0$:

$$\begin{aligned} \mathbf{E}_1 \left|Z_1^1\right|^{r+1} < \infty &\Longleftrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_1\left\{ \left| \sum_{t=1}^{n} \widetilde{Y}_t \right| \geqq \varepsilon n \right\} < \infty \qquad \text{for some} \quad \varepsilon > 0, \\ &\Longleftrightarrow \sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_1\left\{ \left| \sup_{t \geqq n} t^{-1} \widetilde{Y}_t \right| \geqq \varepsilon \right\} < \infty \qquad \text{for all} \quad \varepsilon > 0, \end{aligned} \tag{4.5}$$

where $\widetilde{Y}_t = Y_t - D(f_1, f_0)$. Since $\mathbf{P}_1\{T_\varepsilon^{(1)} > n\} \leqq \mathbf{P}_1\{\sup_{t \geqq n} t^{-1}\widetilde{Y}_t \geqq \varepsilon\}$, the implication $\mathbf{E}_1|Z_1^1|^{r+1} < \infty \Leftrightarrow \mathbf{E}_1[T_\varepsilon^{(1)}]^r < \infty$ for all $\varepsilon > 0$ follows.

Applying Theorem 3, one can conclude that if the KL number $D(f_1, f_0)$ is strictly positive and the $(r+1)$st absolute moment of the LLR $Z_1^1$ is finite, then the Shiryaev detection procedure $\nu_{B_\alpha}$ asymptotically minimizes moments of the detection delay up to the order $r$. However, below we show that the Shiryaev procedure minimizes all positive moments of the detection delay under weaker conditions. As we demonstrate, all that is required is positiveness and finiteness of the KL numbers.

Details are given in the next subsection.

**4.1. Asymptotic optimality.** We will impose the following mild condition on the KL information numbers:

$$0 < D(f_1, f_0) < \infty \quad \text{and} \quad 0 < D(f_0, f_1) < \infty. \tag{4.6}$$

Positiveness of the KL information numbers is not at all restrictive, since it holds whenever the PDFs $f_0(x)$ and $f_1(x)$ do not coincide almost everywhere, i.e., $\mu\{x\colon f_0(x) \neq f_1(x) > 0\}$. If it does not hold, the LLR $Z_1^1$ is equal to zero almost surely, in which case the detection problem is degenerate. The second condition (finiteness) is quite natural and holds in most cases. However, there are reasonable models for which it does not hold. The problem of detecting a change in mean value of a rectangular distribution may serve as a good example. In the latter case, the moments of the LLR are infinite and the detection problem becomes degenerate at least in the asymptotic setting. On the other hand, if the KL numbers are finite, then all the moments of the negative part of the LLR are finite, $\mathbf{E}_1\{-\min(0, Z_1^1)\}^m < \infty$, since the PDFs $f_0(x)$ and $f_1(x)$ are mutually absolutely continuous (i.e., if $f_0(x) = 0$, then so is $f_1(x)$), which implies that

$$\mathbf{E}_{f_1} \frac{f_0(X_1)}{f_1(X_1)} = \mathbf{E}_1 \exp\{-Z_1^1\} = 1. \tag{4.7}$$

In the rest of this section, for the sake of simplicity we will restrict our attention to the geometric prior distribution given in (2.8), in which case $d = |\log(1-\rho)|$ and the statistic $\Lambda_n$ obeys the recursion

$$\Lambda_n = \frac{1}{1-\rho}\,(\Lambda_{n-1} + \rho)\,\frac{f_1(X_n)}{f_0(X_n)}, \qquad n \geqq 1, \quad \Lambda_0 = \frac{\pi_0}{1-\pi_0}. \tag{4.8}$$

The latter recursion follows from recursion (4.2), and we observe that $\pi_n/\Pi_n = \rho/(1-\rho)$ and $\Pi_{n-1}/\Pi_n = 1/(1-\rho)$.

The following theorem establishes asymptotic optimality properties of $\nu_{B_\alpha}$ in the class $\Delta(\alpha)$ with respect to all positive moments of the detection delay.

THEOREM 4. *Let, conditioned on $\lambda = k$, the observations $X_1, \dots, X_{k-1}$ be i.i.d. with the PDF $f_0(x)$ and let $X_k, X_{k+1}$ be i.i.d. with the PDF $f_1(x)$. Further, let the prior distribution of the change-point $\lambda$ be geometric. Suppose that conditions* (4.6) *are satisfied. If $B_\alpha = (1-\alpha)/\alpha$, then* $\mathrm{PFA}(\nu_{B_\alpha}) \leqq \alpha$ *and, as $\alpha \to 0$, for all $m \geqq 1$,*

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{ED}_m^{\pi}(\tau) \sim \mathbf{ED}_m^{\pi}(\nu_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{D(f_1,f_0) + |\log(1-\rho)|}\right)^m, \tag{4.9}$$

$$\inf_{\tau\in\Delta(\alpha)} \mathbf{ED}_m^{(k)}(\tau) \sim \mathbf{ED}_m^{(k)}(\nu_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{D(f_1,f_0) + |\log(1-\rho)|}\right)^m \qquad \forall\ k \geqq 1. \tag{4.10}$$

*Proof.* For $b > 0$, define $\eta_b(k)$ as in (3.33) and note that in the i.i.d. case and for the geometric prior distribution the statistic $S_{k+n-1}^k(\rho) = Z_{k+n-1}^k + n|\log(1-\rho)|$, $n \geqq 1$, is a random walk with mean $\mathbf{E}_k S_k^k(\rho) = D(f_1, f_0) + |\log(1-\rho)|$.

By (4.6), $D(f_1, f_0)$ is positive and finite, and hence

$$\mathbf{E}_k\{-\min(0, Z_k^k)\}^m < \infty \qquad \text{for all} \quad m > 0.$$

Indeed,

$$\mathbf{E}_k \exp\big\{-\min(0, Z_k^k)\big\} = \mathbf{E}_k e^{-Z_k^k}\mathbb{1}_{\{Z_k^k<0\}} + \mathbf{E}_k\mathbb{1}_{\{Z_k^k\geqq 0\}} \leqq \mathbf{E}_k e^{-Z_k^k} + 1 = 2,$$

where the last equality follows from (4.7). Therefore, Theorem III.8.1 of [10] applies to show that for all $m \geqq 1$,

$$\mathbf{E}_k[\eta_b(k)]^m = \left(\frac{b}{D(f_1, f_0) + |\log(1-\rho)|}\right)^m \big(1 + o(1)\big) \qquad \text{as} \quad b \to \infty.$$

Since $\nu_B - k \leqq \eta_{\log B}(k)$ on $\{\nu_B \geqq k\}$ and $\mathbf{P}_k\{\nu_B \geqq k\} \to 1$ as $B \to \infty$, it follows that

$$\mathbf{ED}_m^{(k)}(\nu_{B_\alpha}) \leqq \left(\frac{|\log \alpha|}{D(f_1, f_0) + |\log(1-\rho)|}\right)^m \big(1 + o(1)\big) \qquad \text{as} \quad \alpha \to 0.$$

Note that in the i.i.d. case, condition (3.2) holds trivially with $q = D(f_1, f_0)$, since $Z_{k+n}^k/n$ converges to $D(f_1, f_0)$ ($\mathbf{P}_k$-a.s.) by the strong law of large numbers, and since

$$\mathbf{P}_k\left\{\frac{1}{M} \max_{0 \leqq n < M} Z_{k+n}^k \geqq (1+\varepsilon)\, D(f_1, f_0)\right\}$$

does not depend on $k$. Thus, the lower bound follows from (3.14):

$$\inf_{\tau \in \Delta(\alpha)} \mathbf{ED}_m^{(k)}(\tau) \geqq \left(\frac{|\log \alpha|}{D(f_1, f_0) + |\log(1-\rho)|}\right)^m \big(1 + o(1)\big) \qquad \text{as} \quad \alpha \to 0,$$

which completes the proof of (4.10).

The proof of (4.9) is quite similar and is therefore omitted.

**4.2. Higher-order asymptotic approximations for ADD and PFA.** In this subsection, we use the nonlinear renewal theory developed by Woodroofe [42] (see also [29]) to improve the first-order approximations for the ADD and PFA.

We will also suppose (with minor loss of generality) that $\pi_0 = 0$, i.e., in the rest of this subsection the "pure" geometric prior distribution, $\pi_k = \rho(1-\rho)^{k-1}$, $k \geqq 1$, will be considered. We first observe that, in this case, $\mathrm{CADD}_1(\nu_B) \geqq \mathrm{CADD}_k(\nu_B)$ for all $k \geqq 1$. To understand why, it is sufficient to consider the recursion (4.8) and note that, for $\lambda = k = 1$, the initial condition $\Lambda_0 = 0$ while, for $\lambda = k \geqq 2$, $0 \leqq \Lambda_{k-1} < B$ on $\nu_B \geqq k$. Moreover, for large $B$, the difference between $\mathbf{E}_1(\nu_B - 1)$ and $\mathbf{E}_k(\nu_B - k \,|\, \nu_B \geqq k)$ is a constant that is approximately equal to the mean of the initial condition, $\mathbf{E}_\infty \log \Lambda_{k-1}$ (for $k = 1$ this value is equal to 0). This constant varies for different models and its calculation is usually problematic. For this reason, we will concentrate on the evaluation of the worst-case delay $\mathrm{CADD}_1(\nu_B)$.

In order to apply relevant results from nonlinear renewal theory, we rewrite the stopping time $\nu_B$ in the form of a random walk crossing a constant threshold plus a nonlinear term that is "slowly changing" in the sense defined by [42] and Siegmund [29]. Indeed, the stopping time $\nu_B$ can be written in the following form:

$$\nu_B = \inf\big\{n \geqq 1\colon\ S_n(\rho) + \ell_n \geqq b\big\}, \qquad b = \log(B\rho^{-1}), \tag{4.11}$$

where $S_n(\rho) = Z_n + n|\log(1-\rho)|$ is a random walk with mean $\mathbf{E}_1 S_1(\rho) = D(f_1, f_0) + |\log(1-\rho)|$ and

$$\ell_n = \log\left\{1 + \sum_{i=1}^{n-1} (1-\rho)^i \prod_{s=1}^{i} \frac{f_0(X_s)}{f_1(X_s)}\right\} = \log\left\{1 + \sum_{i=1}^{n-1} (1-\rho)^i e^{-Z_i}\right\}. \tag{4.12}$$

Here and in the rest of this subsection, we write $Z_n$ in place of $Z_n^1$.

For $b > 0$, define $\eta_b(1) = \eta_b$ as in (3.33), i.e.,

$$\eta_b = \inf\big\{n \geqq 1\colon\ S_n(\rho) \geqq b\big\}, \tag{4.13}$$

and let $\varkappa_b = S_{\eta_b}(\rho) - b$ (on $\{\eta_b < \infty\}$) denote the excess (overshoot) of the statistic $S_n(\rho)$ over the threshold $b$ at time $n = \eta_b$. Let

$$G(y, \rho, D) = \lim_{b\to\infty} \mathbf{P}_1\{\varkappa_b \leqq y\} \tag{4.14}$$

be the limiting distribution of the overshoot and let

$$\overline{\varkappa}(\rho, D) = \lim_{b\to\infty} \mathbf{E}_1 \varkappa_b = \int_0^\infty y\, dG(y, \rho, D)$$

denote the related limiting average overshoot. Let us also define

$$\zeta(\rho, D) = \lim_{b\to\infty} \mathbf{E}_1 e^{-\varkappa_b} = \int_0^\infty e^{-y}\, dG(y, \rho, D) \quad \text{and}$$

$$C(\rho, D) = \mathbf{E}_1 \log\left\{01 + \sum_{i=1}^\infty (1-\rho)^i e^{-Z_i}\right\}. \tag{4.15}$$

Note that by (4.11),

$$S_{\nu_B}(\rho) = b - \ell_{\nu_B} + \chi_b \qquad \text{on} \quad \{\nu_B < \infty\},$$

where $\chi_b = S_{\nu_B}(\rho) + \ell_{\nu_B} - b$ is the excess of the process $S_n(\rho) + \ell_n$ over the level $b$ at time $\nu_B$. Taking the expectations on both sides and applying Wald's identity, we obtain

$$\big(D + \big|\log(1-\rho)\big|\big)\, \mathbf{E}_1 \nu_B = b - \mathbf{E}_1 \ell_{\nu_B} + \mathbf{E}_1 \chi_b. \tag{4.16}$$

The crucial observations are that the sequence $\{\ell_n, n \geqq 1\}$ is slowly changing and that $\ell_n$ converges $\mathbf{P}_1$-a.s. as $n \to \infty$ to the random variable

$$\ell = \log\left\{1 + \sum_{i=1}^\infty (1-\rho)^i\, e^{-Z_i}\right\}$$

with finite expectation $\mathbf{E}_1 \ell = C(\rho, D)$. In fact, applying Jensen's inequality yields

$$C(\rho, D) = \mathbf{E}_1 \ell \leqq \log\left(1 + \sum_{k=1}^\infty (1-\rho)^k\right) = \log\left(\frac{1}{\rho}\right). \tag{4.17}$$

Moreover, $\lim_{n\to\infty} \mathbf{E}_1 \ell_n = C(\rho, D)$ due to the fact that $\ell_n \leqq \ell$.

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 [42, Theorem 4.1]). Furthermore, since $\ell_n \to \ell$ and $\mathbf{E}_1 \ell_n \to C(\rho, D)$, using (4.16) we expect that for large $b$,

$$\mathbf{E}_1 \nu_B \approx \frac{1}{D(f_1, f_0) + |\log(1-\rho)|}\, \big(b - C(\rho, D) + \overline{\varkappa}(\rho, D)\big).$$

The mathematical details are given in Theorem 5 below.

More important, nonlinear renewal theory allows us to obtain an asymptotically accurate approximation for $\mathrm{PFA}(\nu_B)$ that takes the overshoot into account. This approximation is important for practical applications, where the value of $D(f_1, f_0)$ is moderate. (For small values of $\rho$ and $D(f_1, f_0)$ the overshoot can be neglected, and formula (2.11) will be reasonably accurate.)

THEOREM 5. *Let the prior distribution of the change-point $\lambda$ be geometric, $\pi_k = \rho(1-\rho)^{k-1}$, $k \geqq 1$, and assume that $Z_n$, $n \geqq 1$, are nonarithmetic with respect to $\mathbf{P}_\infty$ and $\mathbf{P}_1$.*

(i) *If conditions* (4.6) *hold, then*

$$\mathrm{PFA}(\nu_B) = \frac{\zeta(\rho, D)}{B}\big(1+o(1)\big) \qquad \text{as} \quad B \to \infty. \tag{4.18}$$

(ii) *If, in addition, the second moment $\mathbf{E}_1|Z_1|^2$ is finite, then as $B \to \infty$*

$$\mathbf{E}_1\nu_B = \frac{1}{D(f_1, f_0) + |\log(1-\rho)|}\left[\log\frac{B}{\rho} - C(\rho, D) + \overline{\varkappa}(\rho, D)\right] + o(1). \tag{4.19}$$

*Proof.* (i) Obviously,

$$\begin{aligned}\mathrm{PFA}(\nu_B) &= \mathbf{E}^\pi(1-p_{\nu_B}) = \mathbf{E}^\pi(1+\Lambda_{\nu_B})^{-1} = \mathbf{E}^\pi\left[1 + B\left(\frac{\Lambda_{\nu_B}}{B}\right)\right]^{-1} \\ &= \frac{1}{B}\,\mathbf{E}^\pi e^{-\chi_b}\big(1+o(1)\big) \qquad \text{as} \quad B \to \infty,\end{aligned}$$

where $\chi_b = \log\Lambda_{\nu_B} - b$. Since $\chi_b \geqq 0$ and $\mathrm{PFA}(\nu_B) \leqq 1/(1+B) < 1/B$, it follows that

$$\begin{aligned}\mathbf{E}^\pi e^{-\chi_b} &= \mathbf{E}^\pi\{e^{-\chi_b} \mid \nu_B < \lambda\}\mathrm{PFA}(\nu_B) + \mathbf{E}^\pi\{e^{-\chi_b} \mid \nu_B \geqq \lambda\}\big(1 - \mathrm{PFA}(\nu_B)\big) \\ &= \mathbf{E}^\pi\{e^{-\chi_b} \mid \nu_B \geqq \lambda\} + O(B^{-1}) \qquad \text{as} \quad B \to \infty.\end{aligned}$$

Therefore, it suffices to evaluate the value of

$$\mathbf{E}^\pi\{e^{-\chi_b} \mid \nu_B \geqq \lambda\} = \sum_{k=1}^\infty \mathbf{P}\{\lambda = k \mid \nu_B \geqq k\}\,\mathbf{E}_k\{e^{-\chi_b} \mid \nu_B \geqq k\}.$$

To this end, we recall that, by (3.37), for any $1 \leqq k < \infty$,

$$\nu_B = \{n \geqq 1\colon\ S_n^k(\rho) + \ell_{n,k} \geqq b\},$$

where $S_n^k(\rho) = Z_n^k + (n-k+1)|\log(1-\rho)|$, $n \geqq k$, is a random walk with the expectation $\mathbf{E}_k S_k^k(\rho) = D(f_1, f_0) + |\log(1-\rho)|$ and $\ell_{n,k}$, $n \geqq k$, are slowly changing under $\mathbf{P}_k$. Since, by conditions (4.6), $0 < D(f_1, f_0) < \infty$, we can apply Theorem 4.1 of [42] to obtain

$$\lim_{B\to\infty}\mathbf{E}_k\{e^{-\chi_b} \mid \nu_B \geqq k\} = \int_0^\infty e^{-y}\,dG(y, \rho, D) = \zeta(\rho, D).$$

Also,

$$\lim_{B\to\infty}\mathbf{P}\{\lambda = k \mid \nu_B \geqq k\} = \lim_{B\to\infty}\frac{\pi_k\mathbf{P}_\infty\{\nu_B \geqq k\}}{\mathbf{P}^\pi\{\nu_B \geqq \lambda\}} = \pi_k.$$

Consequently,

$$\lim_{B\to\infty} \mathbf{E}^\pi\{e^{-\chi_b} \mid \nu_B \geqq \lambda\} = \lim_{B\to\infty} \mathbf{E}^\pi e^{-\chi_b} = \zeta(\rho, D),$$

which completes the proof of (4.18).

(ii) The proof of (4.19) rests on Woodroofe's Nonlinear Renewal Theorem (see [42, Theorem 4.5]). Indeed, by (4.11), the stopping time $\nu_B$ is based on the thresholding of the sum of the random walk $S_n(\rho)$ and the nonlinear term $\ell_n$. Since

$$\ell_n \xrightarrow[n\to\infty]{\mathbf{P}_1\text{-a.s.}} \ell \quad \text{and} \quad \mathbf{E}_1\ell_n \xrightarrow[n\to\infty]{} \mathbf{E}_1\ell = C(\rho, D),$$

$\ell_n$, $n \geqq 1$, are slowly changing under $\mathbf{P}_1$. In order to apply this theorem we have to check the validity of the following three conditions:

$$\sum_{n=1}^{\infty} \mathbf{P}_1\{\ell_n \leqq -\varepsilon n\} < \infty \qquad \text{for some} \quad 0 < \varepsilon < D; \tag{4.20}$$

$$\max_{0\leqq k\leqq n} |\ell_{n+k}|, \quad n \geqq 1, \quad \text{are } \mathbf{P}_1\text{-uniformly integrable;} \tag{4.21}$$

$$\lim_{B\to\infty} b\,\mathbf{P}_1\big\{\nu_B \leqq \varepsilon b\,(D+\mu)^{-1}\big\} = 0 \qquad \text{for some} \quad 0 < \varepsilon < 1. \tag{4.22}$$

Condition (4.20) holds trivially, since $\ell_n \geqq 0$. The value of $\max_{0\leqq k\leqq n} |\ell_{n+k}|$ is equal to $\ell_{2n}$ because $\ell_n$, $n = 1, 2, \dots$, are nondecreasing, and to prove (4.21) it suffices to show that $\ell_n$, $n \geqq 1$, are $\mathbf{P}_1$-uniformly integrable. Since $\ell_n \leqq \ell$ and, by (4.17), $\mathbf{E}_1\ell < \infty$, the desired uniform integrability follows. Therefore, condition (4.21) is satisfied.

We now turn to checking condition (4.22). Noting that in the notation of subsection 3.2,

$$\mathbf{P}_1\big\{\nu_B < (1-\varepsilon)\, b(D_\rho)^{-1}\big\} = \gamma_\varepsilon^{(1)}(B), \qquad \text{where} \quad D_\rho = D + |\log(1-\rho)|,$$

and using inequalities (3.6) and (3.9) with $\alpha = e^{-b}$, we obtain

$$\mathbf{P}_1\big\{\nu_B < (1-\varepsilon)\, b(D_\rho)^{-1}\big\} \leqq e^{-y_\varepsilon b} + \beta_1(\varepsilon, B),$$

where $y_\varepsilon > 0$ for all $\varepsilon > 0$ and $\beta_1(\varepsilon, B) = \mathbf{P}_1\{\max_{1\leqq n<K_{\varepsilon,B}} Z_n \geqq (1+\varepsilon)\, DK_{\varepsilon,B}\}$, $K_{\varepsilon,B} = (1-\varepsilon)\, b(D_\rho)^{-1}$. The first term in the above inequality is $o(1/b)$ as $B \to \infty$. All that remains to show is that the second term is $o(1/b)$.

To this end, we apply Theorem 1 of [6], according to which for all $\varepsilon > 0$ and $r \geqq 0$,

$$\sum_{n=1}^{\infty} n^{r-1}\mathbf{P}_1\Big\{\max_{1\leqq k\leqq n} (Z_k - Dk) \geqq \varepsilon n\Big\} \leqq C_r\big\{\mathbf{E}_1\big[(Z_1 - D)^+\big]^{r+1} + \big[\mathbf{E}_1(Z_1 - D)^2\big]^r\big\},$$

where $C_r$ is a universal constant. Recall that, by the conditions of the theorem, $\mathbf{E}_1|Z_1|^2 < \infty$. Therefore, the sum on the left-hand side of the previous inequality is finite for $r = 1$ and all $\varepsilon > 0$, which implies that the summand should be $o(1/n)$. Since

$$\beta_1(\varepsilon, B) \leqq \mathbf{P}_1\Big\{\max_{n<K_{\varepsilon,B}} (Z_n - Dn) \geqq \varepsilon D(f_1, f_0)\, K_{\varepsilon,B}\Big\},$$

it follows that $\beta_1(\varepsilon, B) = o(b^{-1})$. Hence condition (4.22) holds for all $0 < \varepsilon < 1$.

Thus, all the conditions of Theorem 4.5 in [42] are satisfied. The use of this theorem yields (4.19) for large $B$.

*Remark* 4. The constants $\overline{\varkappa}(\rho, D)$ and $\zeta(\rho, D)$ are the subject of the renewal theory. The constant $C(\rho, D)$ is not easy to compute in general. For $\rho$ close to 1, the upper bound (4.17) may be useful. Obviously, this bound is asymptotically accurate when $D \to 0$. Monte Carlo experiments may be used to estimate $C$ with reasonable accuracy (see Table 1 in section 6).

The usefulness of Theorem 5 is twofold. First, it provides accurate approximations for both the ADD and the PFA. Second, it allows us to study the important limiting case of $\rho \to 0$.

To analyze the latter case it is convenient to consider the statistic $R_{\rho,n} = \Lambda_n/\rho$. As $\rho \to 0$, this statistic converges to the so-called Shiryaev–Roberts statistic $R_n = \lim_{\rho\to 0} R_{\rho,n}$. Also, as $\rho \to 0$, PFA tends to 1 and $\lim_{\rho\to 0}\{[1-\mathrm{PFA}(\nu_B)]/\rho\} = \mathbf{E}_\infty \nu_B$. Thus, it is natural to consider asymptotics $\alpha \to 1$ and $\rho \to 0$ so that the ratio $(1-\alpha)/\rho$ remains constant at $T$, $0 < T < \infty$, where $T$ may be interpreted as a constraint on the mean time to false alarm $\mathbf{E}_\infty \nu_B$. The requirement $(1-\alpha)/\rho \to T$ as $\alpha \to 1, \rho \to 0$ is similar to that used by Shiryaev [25] for detecting changes in stationary regimes (for the Wiener process).

The following corollary follows directly from Theorem 5 and the above argument.

COROLLARY 2. *Let* $\nu_B = \inf\{n\colon\ R_{\rho,n} \geqq B\}$ *and* $B = (1-\alpha)/(\alpha\rho)$. *Assume that* $\alpha \to 1$ *and* $\rho \to 0$ *in such a way that* $(1-\alpha)/\rho \to T$, *where* $0 < T < \infty$. *Then, under the conditions of Theorem* 5,

$$\mathbf{E}_1 \nu_T = \frac{1}{D(f_1, f_0)} \big[\log T - C(D) + \overline{\varkappa}(D)\big] + o(1) \qquad as \quad T \to \infty, \tag{4.23}$$

*where* $\nu_T$ *represents* $\nu_B$ *under the above limit, and where* $C(D) = C(0, D)$ *and* $\overline{\varkappa}(D) = \overline{\varkappa}(0, D)$.

It can also be shown that under the assumptions of Theorem 5, $\mathbf{E}_\infty \nu_B \sim T$ as $B = T\zeta(D)$ and $T \to \infty$, where $\zeta(D) = \zeta(0, D)$. These results generalize similar results obtained previously by Pollak for exponential families (see [22, Theorem 3]).

**5. Asymptotic performance of other detection procedures.** It is known that in the case where the observations are i.i.d. and the change point is modeled as deterministic but unknown, the CUSUM detection procedure of Page [19] and the randomized Shiryaev–Roberts detection procedure proposed by Pollak [21] are optimal with respect to the minimax expected detection lag, subject to a constraint on the mean time to false alarm.

More specifically, consider the following two detection procedures:

$$\widehat{\tau}_B = \inf\{n \geqq 0\colon\ R_n \geqq B\} \quad \text{and} \quad \tau_B^* = \inf\{n \geqq 1\colon\ U_n \geqq B\},$$

where the statistics $R_n$ and $U_n$ are defined as follows:

$$R_n = \sum_{k=1}^{n} e^{Z_n^k} \quad \text{and} \quad U_n = \max_{1\leqq k\leqq n} e^{Z_n^k}.$$

If $R_0 = 0$, the detection procedure $\widehat{\tau}_B$ is the Shiryaev–Roberts procedure. Its randomized version, when $R_0$ is random with a certain distribution, has been suggested by Pollak [21]. We will refer to this procedure as the Shiryaev–Roberts–Pollak (SRP) test. The second test $\tau_B^*$ is nothing but the CUSUM algorithm.

In 1971, Lorden [17] proposed to measure the loss due to the detection delay by the "worst-worst" minimax risk $\mathrm{ES}(\tau) = \sup_k \operatorname{ess\,sup} \mathbf{E}_k\{(\tau - k + 1)^+ | X^{k-1}\}$ and showed that the CUSUM procedure $\tau_B^*$ with $B = T$ is asymptotically optimal for i.i.d. models, as $T \to \infty$, in the class $\Delta_T = \{\tau\colon \mathbf{E}_\infty \tau \geqq T\}$ of procedures for which the mean time to false alarm $\mathbf{E}_\infty \tau$ exceeds a predefined number $T$. In 1986, Moustakides [18] proved that the CUSUM test is strictly optimal for any $T > 0$ with respect to the risk $\mathrm{ES}(\tau)$ whenever the threshold $B$ is chosen so that $\mathbf{E}_\infty \tau_B^* = T$. In 1996, Shiryaev [28] extended this result for continuous-time processes (detecting a change in the mean of the Brownian motion). In 1998, Lai [16] generalized previous results for a general non-i.i.d. case showing that the CUSUM procedure is asymptotically optimal in the class $\Delta_T$ as $T \to \infty$ with respect to both the "worst-worst" risk $\mathrm{ES}(\tau)$ and the "average-worst" risk (minimax delay) $\mathrm{MD}(\tau) = \sup_k \mathbf{E}_k(\tau - k \,|\, \tau \geqq k)$. The latter measure of the detection speed was introduced earlier by Pollak [21].

In the mid 1980s, Pollak [21] proposed the randomized version of the Shiryaev–Roberts procedure $\widehat{\tau}_B$, where the statistic $R_n$ is randomized at the zero point $n = 0$, i.e., $R_0$ is a random variable ($R_0 = 0$ for the standard Shiryaev–Roberts procedure). Pollak proved that this randomized procedure is nearly optimal with respect to the risk $\mathrm{MD}(\tau)$ for i.i.d. data models. Pollak [22] also presented a comprehensive asymptotic analysis of the procedure $\widehat{\tau}_B$ for exponential families. See [1], [30], [31], [32], [33], [34], [35], [36], [37], [38], and [39] for further extensions and details.

So, both the CUSUM and the SRP detection procedures minimize $\sup_k \mathbf{E}_k(\tau - k \,|\, \tau \geqq k)$ (the expected detection delay in the worst-case scenario) in the class of procedures for which the mean time to false alarm $\mathbf{E}_\infty \tau$ exceeds a predefined number $T$ (at least as $T \to \infty$).

In this section, we study asymptotic properties of these two classical change-point detection procedures in the class $\Delta(\alpha)$. The results of previous sections are used to establish that they lose the asymptotic optimality property under the Bayesian criterion for prior distributions with exponential tail $\mathcal{E}(d)$, but remain optimal for heavy-tailed prior distributions $\mathcal{H}$.

In order to obtain an upper bound for the PFA of the SRP procedure, we note that the statistic $R_n - n$ is a zero-mean $\mathbf{P}_\infty$-martingale (with respect to $\mathcal{F}_n^X$). Therefore, $R_n$ is a submartingale with mean $\mathbf{E}_\infty R_n = n$. Using Doob's submartingale inequality, we get

$$\mathbf{P}_\infty\{\widehat{\tau}_B < n\} = \mathbf{P}_\infty\Big\{\max_{1 \leqq k < n} R_k \geqq B\Big\} \leqq n B^{-1},$$

which yields

$$\widehat{\mathrm{PFA}}(B) = \mathbf{P}^\pi\{\widehat{\tau}_B < \lambda\} = \sum_{n=1}^{\infty} \pi_n \mathbf{P}_\infty\{\widehat{\tau}_B < n\} \leqq \overline{\lambda} B^{-1},$$

where $\overline{\lambda} = \sum_{k=1}^\infty k\pi_k$ is the mean of the prior distribution. Thus, choosing $B = B_\alpha = \overline{\lambda}/\alpha$ guarantees $\widehat{\tau}_{B_\alpha} \in \Delta(\alpha)$.

Since $\tau_B^* \geqq \widehat{\tau}_B$, for the CUSUM procedure, we obtain

$$\mathbf{P}_\infty\{\tau_B^* < n\} \leqq \mathbf{P}_\infty\{\widehat{\tau}_B < n\} \leqq n B^{-1}, \qquad n \geqq 1, \tag{5.1}$$

and hence,

$$\mathrm{PFA}^*(B) = \mathbf{P}^\pi\{\tau_B^* < \lambda\} = \sum_{n=1}^{\infty} \pi_n \mathbf{P}_\infty\{\tau_B^* < n\} \leqq B^{-1} \sum_{n=1}^{\infty} n\pi_n = \overline{\lambda} B^{-1}. \tag{5.2}$$

It follows that $B = B_\alpha = \overline{\lambda}/\alpha$ implies $\tau_{B_\alpha} \in \Delta(\alpha)$.

The following theorem, which is a prototype of Theorem 3, establishes the asymptotic operating characteristics of the CUSUM and SRP procedures in terms of moments of the detection delay.

THEOREM 6. *Let* $\overline{\lambda} = \sum_{k=1}^{\infty} k\pi_k < \infty$ *and* $B = B_\alpha = \overline{\lambda}/\alpha$*. Then both detection procedures* $\tau^*_{B_\alpha}$ *and* $\widehat{\tau}_{B_\alpha}$ *belong to the class* $\Delta(\alpha)$ *and the following two assertions hold.*

(i) *If the condition* (3.25) *is satisfied for some* $r > 0$*, then for all* $m \leqq r$ *and* $k \geqq 1$ *as* $\alpha \to 0$*,*

$$\mathbf{ED}_m^{(k)}(\tau^*_{B_\alpha}) \sim \mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{q}\right)^m. \tag{5.3}$$

(ii) *If the condition* (3.26) *is satisfied for some* $r > 0$*, then for all* $m \leqq r$ *as* $\alpha \to 0$*,*

$$\mathbf{ED}_m^{\pi}(\tau^*_{B_\alpha}) \sim \mathbf{ED}_m^{\pi}(\widehat{\tau}_{B_\alpha}) \sim \left(\frac{|\log\alpha|}{q}\right)^m. \tag{5.4}$$

*Proof.* We provide the proof only for the SRP procedure, since the proof for the CUSUM procedure is essentially the same.

For any $B > 0$, define the one-sided sequential probability ratio test by

$$\eta_B(k) = \inf\{n \geqq 1:\ Z_{k+n-1}^k \geqq \log B\}, \qquad k \geqq 1.$$

Our first observation is that $\widehat{\tau}_B - k \leqq \eta_B(k)$ on $\{\widehat{\tau}_B \geqq k\}$. Consequently, for $B > k$

$$\mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha}) = \frac{\mathbf{E}_k[(\widehat{\tau}_B - k)^+]^m}{\mathbf{P}_k\{\widehat{\tau}_B \geqq k\}} \leqq \frac{\mathbf{E}_k\{[\eta_B(k)]^m;\ \widehat{\tau}_B \geqq k\}}{\mathbf{P}_k\{\widehat{\tau}_B \geqq k\}} \leqq \frac{\mathbf{E}_k[\eta_B(k)]^m}{1 - kB^{-1}},$$

where the latter inequality follows from (5.1). By Lemma 3,

$$\mathbf{E}_k[\eta_B(k)]^m \sim \left(\frac{\log B}{q}\right)^m \qquad \text{as} \quad B \to \infty, \tag{5.5}$$

and hence,

$$\mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha}) \leqq \left(\frac{\log B}{q}\right)^m \big(1 + o(1)\big) \qquad \text{as} \quad B \to \infty. \tag{5.6}$$

Next, similar to (3.41)

$$\widehat{\tau}_B - k \leqq \eta_B(k) \leqq T_\varepsilon^{(k)} + 2 + \frac{\log B}{q - \varepsilon} \qquad \text{on} \quad \{\widehat{\tau}_B \geqq k\}.$$

Applying the latter inequality along with $\mathbf{P}^\pi\{\widehat{\tau}_B \geqq \lambda\} \geqq 1 - \overline{\lambda}/B$ yields (for $B > \overline{\lambda}$)

$$\mathbf{ED}_m^{\pi}(\widehat{\tau}_{B_\alpha}) \leqq \frac{B - \overline{\lambda}}{B} \sum_{k=1}^{\infty} \pi_k \mathbf{E}_k \left(\frac{\log B}{q - \varepsilon} + T_\varepsilon^{(k)} + 2\right)^m.$$

Since by the assumption of the theorem, $\sum_{k=1}^{\infty} \pi_k \mathbf{E}_k[T_\varepsilon^{(k)}]^r < \infty$ and $\varepsilon$ can be arbitrarily small, it follows that for $m \leqq r$,

$$\mathbf{ED}_m^{\pi}(\widehat{\tau}_B) \leqq \left(\frac{\log B}{q}\right)^m \big(1 + o(1)\big) \qquad \text{as} \quad B \to \infty. \tag{5.7}$$

We now proceed with deriving the lower bounds. Write

$$\widehat{\gamma}_\varepsilon^{(k)}(B) = \mathbf{P}_k\big\{k \leqq \widehat{\tau}_B < k + (1-\varepsilon)\, q^{-1}\log B\big\},$$
$$\widehat{\gamma}_\varepsilon(B) = \mathbf{P}^\pi\big\{\lambda \leqq \widehat{\tau}_B < \lambda + (1-\varepsilon)\, q^{-1}\log B\big\}.$$

An argument similar to that used in the proof of Lemma 2 shows that

$$\lim_{B\to\infty} \widehat{\gamma}_\varepsilon^{(k)}(B) = 0 \quad \text{and} \quad \lim_{B\to\infty} \widehat{\gamma}_\varepsilon(B) = 0 \qquad \text{for every} \quad 0<\varepsilon<1. \tag{5.8}$$

Therefore, applying Chebyshev's inequality, we obtain

$$\mathbf{E}^\pi\big[(\widehat{\tau}_B-\lambda)^+\big]^m \geqq \big[(1-\varepsilon)\, q^{-1}\log B\big]^m \big[\mathbf{P}^\pi\{\widehat{\tau}_B \geqq \lambda\} - \widehat{\gamma}_\varepsilon(B)\big].$$

Since $\mathbf{ED}_m^\pi(\widehat{\tau}_B) = \mathbf{E}^\pi[(\widehat{\tau}_B-\lambda)^+]^m/\mathbf{P}^\pi\{\widehat{\tau}_B \geqq \lambda\}$ and $\mathbf{P}^\pi\{\widehat{\tau}_B \geqq \lambda\} \geqq 1-\overline{\lambda}/B$, it follows that, for $B>\overline{\lambda}$,

$$\mathbf{ED}_m^\pi(\widehat{\tau}_B) \geqq \big[(1-\varepsilon)\, q^{-1}\log B\big]^m \left[1 - \frac{\widehat{\gamma}_\varepsilon(B)}{1-\overline{\lambda}B^{-1}}\right],$$

where $\widehat{\gamma}_\varepsilon(B)\to 0$ as $B\to\infty$ by (5.8). Since $\varepsilon$ is arbitrary,

$$\mathbf{ED}_m^\pi(\widehat{\tau}_B) \geqq \left(\frac{\log B}{q}\right)^m \big(1+o(1)\big) \qquad \text{as} \quad B\to\infty,$$

which, along with the upper bound (5.7), yields the asymptotic equality

$$\mathbf{ED}_m^\pi(\widehat{\tau}_B) = \left(\frac{\log B}{q}\right)^m \big(1+o(1)\big) \qquad \text{as} \quad B\to\infty. \tag{5.9}$$

Setting $B = B_\alpha = \log(\overline{\lambda}/\alpha)$ in (5.9) proves (5.4).

Analogously, for $B>k$,

$$\begin{aligned}\mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha}) &\geqq \big[(1-\varepsilon)\, q^{-1}\log B\big]^m \left[1 - \frac{\widehat{\gamma}_\varepsilon^{(k)}(B)}{\mathbf{P}_k\{\widehat{\tau}_B \geqq k\}}\right]\\ &\geqq \big[(1-\varepsilon)\, q^{-1}\log B\big]^m \left[1 - \frac{\widehat{\gamma}_\varepsilon^{(k)}(B)}{1-kB^{-1}}\right].\end{aligned}$$

Since $\varepsilon$ can be arbitrarily small and, by (5.8), $\widehat{\gamma}_\varepsilon^{(k)}(B)\to 0$ as $B\to\infty$ for every $k \geqq 1$, we can conclude that the right-hand side in (5.6) is the asymptotic lower bound for $\mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha})$. Thus, the asymptotic approximation (5.9) holds for $\mathbf{ED}_m^{(k)}(\widehat{\tau}_{B_\alpha})$. Finally, setting $B = B_\alpha = \log(\overline{\lambda}/\alpha)$ proves (5.3).

Comparing Theorems 3 and 6 shows that the CUSUM and SRP procedures are not asymptotically optimal in a Bayesian context for $\pi \in \mathcal{E}(d)$ but remain asymptotically optimal for $\pi \in \mathcal{H}$. In particular, for the geometric prior distribution and i.i.d. data models,

$$\lim_{\alpha\to 0} \frac{\mathrm{ADD}(\widehat{\tau}_{B_\alpha})}{\inf_{\tau\in\Delta(\alpha)} \mathrm{ADD}(\tau)} = \lim_{\alpha\to 0} \frac{\mathrm{ADD}(\tau^*_{B_\alpha})}{\inf_{\tau\in\Delta(\alpha)} \mathrm{ADD}(\tau)} = 1 + \frac{|\log(1-\rho)|}{D(f_1,f_0)}. \tag{5.10}$$

*Remark* 5. While the standard CUSUM algorithm is not optimal, by using the above techniques one can show that the following "Bayesian" modification of the

CUSUM procedure is asymptotically optimal as $\alpha \to 0$. Define the weighted CUSUM detection procedure by

$$\tau_B = \inf\left\{n \geqq 1\colon \max_{0\leqq k\leqq n}\left[(\Pi_n)^{-1}\pi_k \prod_{i=k}^{n} \frac{f_{1,i}(X_i \mid \mathbf{X}^{i-1})}{f_{0,i}(X_i \mid \mathbf{X}^{i-1})}\right] \geqq B\right\},$$

$B > \pi_0/(1-\pi_0)$. If $B = B_\alpha = (1-\alpha)/\alpha$, then $\tau_{B_\alpha} \in \Delta(\alpha)$ and Theorems 3 and 4 hold true for $\tau_{B_\alpha}$.

**6. Example 1: Change detection in the mean of the autoregressive process.** In this section, we consider an example that illustrates the results of the previous sections. This example is focused on single-sensor or centralized detection. In the next section, we will consider two more examples that are related to decentralized (distributed) detection in multisensor systems. In the rest of the paper, the prior distribution (2.8) will be assumed geometric with $\pi_0 = 0$.

Consider the "signal plus noise" model, in which we assume that $X_n = \mathbb{1}_{\{n\geqq\lambda\}}\theta_n + \xi_n$, $n \geqq 1$, where $\theta_n$ is a deterministic signal that appears at an unknown point in time $\lambda$, and $\{\xi_n, n \geqq 1\}$ is a Markov Gaussian sequence (noise), which obeys the recursion

$$\xi_n = \delta\xi_{n-1} + w_n, \qquad n \geqq 1, \quad \xi_0 = 0.$$

Here $w_1, w_2, \dots$ are i.i.d. Gaussian random variables with mean zero and variance $\sigma^2$. The parameters $0 \leqq \delta < 1$ and $\sigma > 0$ are assumed to be known ($\delta$ is the correlation coefficient of noise). Let $\varphi(x) = (2\pi)^{-1/2}\exp\{-x^2/2\}$ denote the PDF of the standard normal distribution.

For this model, the conditional PDFs $f_{0,n}(X_n|\mathbf{X}^{n-1})$ and $f_{1,n}(X_n|\mathbf{X}^{n-1})$ introduced in section 2 are of the form

$$f_{0,n}(X_n \mid \mathbf{X}^{n-1}) = f_0(X_n \mid X_{n-1}) = \frac{1}{\sigma}\,\varphi\left(\frac{X_n - \delta X_{n-1}}{\sigma}\right) \qquad \text{for all} \quad n \geqq 1,$$

$$f_{1,n}(X_n \mid \mathbf{X}^{n-1}) = f_{1,n}(X_n \mid X_{n-1}) = \frac{1}{\sigma}\,\varphi\left(\frac{X_n - \delta X_{n-1} - \theta_n}{\sigma}\right) \qquad \text{for} \quad n = \lambda,$$

$$\begin{aligned} f_{1,n}(X_n \mid \mathbf{X}^{n-1}) &= f_{1,n}(X_n \mid X_{n-1}) \\ &= \frac{1}{\sigma}\,\varphi\left(\frac{X_n - \delta X_{n-1} - (\theta_n - \delta\theta_{n-1})}{\sigma}\right) \qquad \text{for} \quad n \geqq \lambda + 1, \end{aligned}$$

where $X_0 = \theta_0 = 0$.

Write

$$\widetilde{X}_i = X_i - \delta X_{i-1} \quad \text{and} \quad \tilde{\theta}_i = \theta_i - \delta\theta_{i-1}.$$

It is easy to see that

$$Z_n^k = \frac{1}{\sigma^2}\left(\theta_k \widetilde{X}_k + \sum_{i=k+1}^{n} \tilde{\theta}_i \widetilde{X}_i\right) - \frac{1}{2\sigma^2}\left(\theta_k^2 + \sum_{i=k+1}^{n} \tilde{\theta}_i^2\right), \qquad 1 \leqq k \leqq n.$$

Next, conditioned on the change-point $\lambda = k$, the values of $\widetilde{X}_n$, $n = 1, 2, \dots$, are independent normal random variables with variance $\sigma^2$, and $\mathbf{E}_k\widetilde{X}_n = 0$ for $n < k$,

$\mathbf{E}_k \widetilde{X}_n = \theta_n$ for $n = k$, and $\mathbf{E}_k \widetilde{X}_n = \tilde{\theta}_n$ for $n > k$. Therefore, conditioned on $\lambda = k$, $\{Z_n^k, n \geqq k\}$ is a Gaussian process with independent increments and parameters

$$\mathbf{E}_k Z_n^k = -\mathbf{E}_\infty Z_n^k = \frac{1}{2\sigma^2}\left(\theta_k^2 + \sum_{i=k+1}^n \tilde{\theta}_i^2\right), \quad \mathbf{D}_k Z_n^k = \mathbf{D}_\infty Z_n^k = \frac{1}{\sigma^2}\left(\theta_k^2 + \sum_{i=k+1}^n \tilde{\theta}_i^2\right),$$

where $\mathbf{D}$ stands for variance.

Write

$$Q_{k,n} = \frac{1}{\sigma^2 n} \sum_{i=k}^{k+n-1} \tilde{\theta}_i^2$$

and assume that

$$\lim_{n\to\infty} Q_{k,n} = Q \qquad \text{for all} \quad k \geqq 1, \tag{6.1}$$

where $Q$ characterizes the average "signal-to-noise ratio" $0 < Q < \infty$. It is easily verified that

$$\mathbf{P}_k\left\{\left|Z_{k+n-1}^k - \frac{Qn}{2}\right| > \varepsilon n\right\} = 2\Phi\left\{-\frac{(\varepsilon - \Delta_{k,n})\sqrt{n}}{\sqrt{Q_{k,n}}}\right\},$$

where $\Phi(x)$ is the standard normal distribution function and $\Delta_{k,n} = (Q_{k,n} - Q)/2$. By condition (6.1),

$$\sum_{n=1}^\infty n^{r-1} \mathbf{P}_k\left\{\left|Z_{k+n-1}^k - \frac{Qn}{2}\right| > \varepsilon n\right\} < \infty \qquad \text{for all} \quad r > 0,$$

and hence, for all $r > 0$

$$n^{-1} Z_{k+n}^k \xrightarrow[n\to\infty]{\mathbf{P}_k\text{-}r\text{-quickly}} \frac{Q}{2} \quad \text{and} \quad n^{-1} Z_{k+n}^k \xrightarrow[n\to\infty]{\mathbf{P}_\infty\text{-}r\text{-quickly}} -\frac{Q}{2}.$$

Thus, condition (3.25) holds for all positive $r$.

Also, obviously,

$$\sum_{k=1}^\infty \pi_k \left(\sum_{n=1}^\infty n^{r-1} \mathbf{P}_k\left\{\left|Z_{k+n-1}^k - \frac{Qn}{2}\right| > \varepsilon n\right\}\right) < \infty,$$

which implies condition (3.26).

Thus, under condition (6.1) with $0 < Q < \infty$, according to Theorem 3, the Shiryaev detection algorithm $\nu_B$ with $B = (1-\alpha)/\alpha$ asymptotically minimizes all positive moments of the detection delay in the class $\Delta(\alpha)$, and the asymptotic formulas (3.44) and (3.45) hold with $q = Q/2$.

This result can easily be generalized for the problem of detecting a change in the mean of the $p$th order Gaussian autoregressive process

$$\xi_n = \sum_{j=1}^p \delta_j \xi_{n-j} + w_n, \qquad n \geqq 1, \quad \xi_k = 0 \ \text{ for } \ k \leqq 0,$$

where $w_n$, $n \geqq 1$, are i.i.d. $\mathcal{N}(0, \sigma^2)$. Specifically, define $\tilde{\theta}_{p,n} = \theta_n - \sum_{j=1}^p \theta_{n-j}$ and assume that

$$\lim_{n\to\infty} \frac{1}{\sigma^2(n-p)} \sum_{k=p+1}^n \tilde{\theta}_{p,k}^2 = Q, \qquad \text{where} \quad 0 < Q < \infty.$$

Then Theorem 3 and Corollary 1 show that $\nu_B$ is asymptotically optimal, and asymptotic formulas (3.44)–(3.47) hold true with $q = Q/2$.

We now return to the Markov case and assume that the mean value is constant, $\theta_n = \theta \neq 0$. Then condition (6.1) is fulfilled with $Q = \theta^2(1-\delta)^2/\sigma^2$.

In the latter case, the results of subsection 4.2 can be applied. To show this, we first note that the LLR $Z_n^1$ can be written in the form

$$Z_n^1 = \sum_{k=2}^{n} \Delta W_k + \frac{\theta}{\sigma^2}\, X_1 - \frac{\theta^2}{2\sigma^2},$$

where $\Delta W_k = \theta(1-\delta)\,\widetilde{X}_k/\sigma^2 - \theta^2(1-\delta)^2/(2\sigma^2)$, $k \geqq 2$, are i.i.d. Gaussian random variables with parameters

$$\mathbf{E}_1 \Delta W_k = -\mathbf{E}_\infty \Delta W_k = \frac{Q}{2}, \qquad \mathbf{D}_1 \Delta W_k = \mathbf{D}_\infty \Delta W_k = Q. \tag{6.2}$$

Therefore, by adding and subtracting the random variable $\Delta W_1$, which has the same distribution as $\Delta W_2, \Delta W_3, \ldots$, one can represent $Z_n^1$ in the form

$$Z_n^1 = W_n + S$$

with $W_n = \Delta W_1 + \cdots + \Delta W_n$ being a Gaussian random walk with the parameters given by (6.2), and $S = (\theta/\sigma^2)\, X_1 - \theta^2/(2\sigma^2) - \Delta W_1$ being a Gaussian random variable with $\mathbf{E}_1 S = -\mathbf{E}_\infty S = QA_\delta/2$, where $A_\delta = [1-(1-\delta)^2]/(1-\delta)^2$.

In further calculations, including Monte Carlo experiments, the stopping time $\nu_B$ will be defined as $\nu_B = \inf\{n \geqq 1\colon\ R_{\rho,n} \geqq B\}$ with $R_{\rho,n} = \Lambda_n/\rho$.

A slight modification of the proof of Theorem 5 shows that the PFA obeys the asymptotic formula (4.18) and, as $B \to \infty$,

$$\mathbf{E}_1 \nu_B = \frac{1}{Q/2 + |\log(1-\rho)|}\left[\log B - C(\rho, Q) + \overline{\varkappa}(\rho, Q) - \frac{QA_\delta}{2}\right] + o(1). \tag{6.3}$$

As compared to the asymptotic expansion (4.19) for the i.i.d. case, here an additional term $-QA_\delta/2$ appears due to the random variable $S$ in the decomposition of the LLR.

To guarantee the given PFA $\alpha$ in simulations, we used the following threshold value obtained by reverting to (4.18) in Theorem 5:

$$B = \frac{\zeta(\rho, Q)}{\alpha\rho}. \tag{6.4}$$

According to [42, Corollary 2.2.7], the constant $\zeta(\rho, Q)$ is computed from the formula

$$\zeta(\rho, Q) = \frac{2}{Q + 2|\log(1-\rho)|}\, \exp\left\{ -\sum_{k=1}^{\infty} \frac{1}{k}\, F_k(\rho, Q)\right\}, \tag{6.5}$$

where

$$\begin{aligned} F_k(\rho, Q) &= \Phi\left(-\frac{Q + 2|\log(1-\rho)|}{2\sqrt{Q}}\,\sqrt{k}\right) \\ &\quad + (1-\rho)^k \Phi\left(-\frac{Q - 2|\log(1-\rho)|}{2\sqrt{Q}}\,\sqrt{k}\right) \end{aligned} \tag{6.6}$$

and $\Phi(x) = \int_{-\infty}^{x} \varphi(t)\, dt$ is a standard normal distribution function.

To compute the CADD, we used the following higher order (HO) approximation:

$$\text{CADD}_1(\nu_B) \approx \max\left\{0, \frac{2}{Q_\rho}\left(\log B - C(\rho, Q) + \overline{\varkappa}(\rho, Q) - \frac{QA_\delta}{2}\right) - 1\right\}, \tag{6.7}$$

where, according to [42, Corollary 2.2.7],

$$\overline{\varkappa}(\rho, Q) = \frac{Q_\rho^2/4 + Q}{Q_\rho} - \sqrt{Q}\sum_{k=1}^{\infty}\left[k^{-1/2}\varphi\left(\frac{Q_\rho\sqrt{k}}{2\sqrt{Q}}\right) - \frac{Q_\rho}{2\sqrt{Q}}\,\Phi\left(-\frac{Q_\rho\sqrt{k}}{2\sqrt{Q}}\right)\right].$$

Here we used the notation $Q_\rho = Q + 2|\log(1-\rho)|$.

Formula (6.7) follows from the HO asymptotic (6.3). Note that this formula requires the computation of the constant $C(\rho, Q)$ using (4.15). As we observed in Remark 4, we usually have to resort to Monte Carlo methods to estimate $C(\rho, Q)$. Values of $C$ for various choices of $Q$, $\rho$, and $\delta$ are given in Table 1. The number of trials were such that the estimate of the standard deviation of $C$ was within 0.5% of the mean.

For the purpose of comparison, we also used the first order (FO) approximations for CADD (see (3.29))

$$\text{CADD}_1(\nu_B) \approx \max\left\{0,\ \frac{2\log B}{Q + 2|\log(1-\rho)|} - 1\right\}. \tag{6.8}$$

Extensive Monte Carlo simulations have been performed for different values of $Q$, $\rho$, $\delta$, and $\alpha$. The number of trials used for these results is given by $1000/\alpha$. Sample results are shown in Tables 2, 3, and 4. In these tables, we present the Monte Carlo estimates of ADD along with the theoretical values computed according to (6.7) and (6.8). The abbreviations MCADD, $\text{MCCADD}_1$, FOADD, and $\text{HOCADD}_1$ are used for the ADD obtained by the Monte Carlo experiment, $\text{CADD}_1$ obtained by the Monte Carlo experiment, the FO approximation (6.8), and the HO approximation (6.7) for $\text{CADD}_1$, respectively. We also list Monte Carlo estimates for the PFA.

Table 2 contains results of analysis in the i.i.d. case when the threshold $B = (1-\alpha)/(\rho\alpha)$. This threshold value is based on the general upper bound that ignores the overshoot. It can be seen that the Monte Carlo estimates for the PFA in this case are substantially smaller than the design values $\alpha$. This leads to an increase of the true values of the average detection delay, which is undesirable. It can also be seen that FO approximations are inaccurate even for relatively small $\alpha$, while HO approximations are very accurate.

The results in Table 3 correspond to the i.i.d. case, where the threshold $B$ is set using (6.4). It is seen that the Monte Carlo estimates for the PFA match $\alpha$ very closely, especially for values smaller than 0.01. Thus, (6.4) provides an accurate method to design the threshold $B$ to meet the PFA constraint $\alpha$. It is also seen that, as expected, $\text{MCCADD}_1$ exceeds MCADD in all cases. The FOADD values are not good approximations even when PFA is small. On the other hand, the HO approximation for $\text{CADD}_1$ (given by $\text{HOCADD}_1$) is seen to be very accurate even for moderate values of the PFA.

Results for the correlated case with $\delta = 0.5$ are presented in Table 4, with the threshold $B$ being set using (6.4). Here again we see the accuracy of HO order approximations for the PFA and ADD. Also, it is interesting to see that for the same value of effective signal-to-noise ratio, $Q$, the ADD in the correlated case is slightly smaller than in the i.i.d. case. On the other hand, if we fix the value of "actual"

TABLE 1
*Values of the constant $C$ for different $Q, \rho, \delta$.*

| $\delta = 0$ (i.i.d.) | | | $\delta = 0.5$ | | |
|---|---|---|---|---|---|
| $Q$ | $\rho$ | $C$ | $Q$ | $\rho$ | $C$ |
| 1.0 | 0.3 | 0.8366 | 1.0 | 0.3 | 0.5062 |
| 1.0 | 0.1 | 1.2396 | 1.0 | 0.1 | 0.7538 |
| 1.0 | 0.03 | 1.4036 | 1.0 | 0.03 | 0.8681 |
| 1.0 | 0.01 | 1.4647 | 1.0 | 0.01 | 0.9081 |
| 0.5 | 0.3 | 0.9949 | 0.5 | 0.3 | 0.7590 |
| 0.5 | 0.1 | 1.5859 | 0.5 | 0.1 | 1.2211 |
| 0.5 | 0.03 | 1.9001 | 0.5 | 0.03 | 1.5002 |
| 0.5 | 0.01 | 2.0371 | 0.5 | 0.01 | 1.5936 |
| 0.25 | 0.3 | 1.0913 | 0.25 | 0.3 | 0.9479 |
| 0.25 | 0.1 | 1.8694 | 0.25 | 0.1 | 1.6444 |
| 0.25 | 0.03 | 2.3827 | 0.25 | 0.03 | 2.1245 |
| 0.25 | 0.01 | 2.5992 | 0.25 | 0.01 | 2.3196 |
| 0.1 | 0.3 | 1.1630 | 0.1 | 0.3 | 1.0970 |
| 0.1 | 0.1 | 2.1290 | 0.1 | 0.1 | 2.0009 |
| 0.1 | 0.03 | 2.9164 | 0.1 | 0.03 | 2.7597 |
| 0.1 | 0.01 | 3.3528 | 0.1 | 0.01 | 3.1722 |

TABLE 2
*Results for i.i.d. case with $B = (1 - \alpha)/(\rho\alpha)$.*

| $\rho = 0.1$, $Q = 0.25$ | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | MCPFA | MCADD | MCCADD$_1$ | FOADD | HOCADD$_1$ |
| 0.1000 | 0.0768 | 9.2315 | 12.3424 | 18.5338 | 11.9508 |
| 0.0600 | 0.0464 | 11.1187 | 14.4889 | 20.9400 | 14.4484 |
| 0.0300 | 0.0215 | 14.0684 | 17.6605 | 24.0854 | 17.5509 |
| 0.0100 | 0.0070 | 18.7026 | 22.4509 | 28.9431 | 22.3195 |
| 0.0060 | 0.0043 | 20.8690 | 24.6155 | 31.1781 | 24.6034 |
| 0.0030 | 0.0024 | 23.7940 | 27.6285 | 34.2001 | 27.5586 |
| 0.0010 | 0.0007 | 28.5247 | 32.3746 | 38.9779 | 32.3523 |
| $\rho = 0.1$, $Q = 0.1$ | | | | | |
| $\alpha$ | MCPFA | MCADD | MCCADD$_1$ | FOADD | HOCADD$_1$ |
| 0.1000 | 0.0858 | 11.9405 | 16.4280 | 27.9637 | 15.8823 |
| 0.0600 | 0.0460 | 14.7162 | 19.7314 | 31.5316 | 19.4116 |
| 0.0300 | 0.0240 | 18.9581 | 24.2873 | 36.1953 | 24.0955 |
| 0.0100 | 0.0083 | 25.6559 | 31.3594 | 43.3981 | 31.2017 |
| 0.0060 | 0.0048 | 28.8515 | 34.6259 | 46.7120 | 34.5589 |
| 0.0030 | 0.0025 | 33.2216 | 39.0867 | 51.1929 | 39.0017 |
| 0.0010 | 0.0008 | 40.2447 | 46.1591 | 58.2772 | 46.1774 |

signal-to-noise ratio $\theta^2/\sigma^2$, e.g., $Q = 1$ in the i.i.d. case and $Q = 0.25$ in the $\delta = 0.5$ case, then we can see that the correlation slows down the change detection.

**7. More examples: Decentralized quickest change detection.** The results of the previous sections are particularly useful in the analysis of the decentralized version of the change-detection problem described in [41]. We first outline this interesting problem and related asymptotic optimality results for i.i.d. data models. Then we give two examples.

**7.1. A decentralized detection problem.** Assume that the information about the change is available through a set of $L$ separate sensors. At time $n$ an observation $X_{\ell,n}$ is made at sensor $S_\ell$. Conditioned on the change-point $\lambda$, the observation sequences $\{X_{1,n}\}, \dots, \{X_{L,n}\}$ are assumed to be mutually independent. Furthermore,

TABLE 3
*Results for i.i.d. case with $B = \zeta(\rho, Q)/(\rho\alpha)$.*

| $\rho = 0.1$, $Q = 1$ | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0914 | 3.9388 | 4.9192 | 5.6139 | 4.8214 |
| 0.0600 | 0.0554 | 4.6407 | 5.7084 | 6.4577 | 5.6622 |
| 0.0300 | 0.0293 | 5.7191 | 6.8523 | 7.6027 | 6.8195 |
| 0.0100 | 0.0100 | 7.4474 | 8.6344 | 9.4175 | 8.6221 |
| 0.0060 | 0.0059 | 8.2627 | 9.4719 | 10.2614 | 9.4494 |
| 0.0030 | 0.0030 | 9.3973 | 10.6116 | 11.4064 | 10.6225 |
| 0.0010 | 0.0010 | 11.1895 | 12.4177 | 13.2212 | 12.4328 |
| $\rho = 0.01$, $Q = 1$ | | | | | |
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0907 | 8.5173 | 9.9681 | 11.4037 | 9.9424 |
| 0.0600 | 0.0547 | 9.5119 | 11.0090 | 12.4052 | 10.9459 |
| 0.0300 | 0.0290 | 10.7933 | 12.2914 | 13.7642 | 12.3139 |
| 0.0100 | 0.0100 | 12.9459 | 14.4763 | 15.9181 | 14.4788 |
| 0.0060 | 0.0058 | 13.9602 | 15.4800 | 16.9196 | 15.4538 |
| 0.0030 | 0.0030 | 15.2986 | 16.8320 | 18.2786 | 16.8507 |
| 0.0010 | 0.0010 | 17.4523 | 18.9875 | 20.4325 | 18.9818 |
| $\rho = 0.1$, $Q = 0.25$ | | | | | |
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0915 | 8.4385 | 11.4574 | 17.6206 | 11.0010 |
| 0.0600 | 0.0558 | 10.2942 | 13.5159 | 19.8381 | 13.2243 |
| 0.0300 | 0.0284 | 12.9266 | 16.3797 | 22.8471 | 16.2042 |
| 0.0100 | 0.0096 | 17.4060 | 21.0897 | 27.6162 | 21.0265 |
| 0.0060 | 0.0060 | 19.4797 | 23.2396 | 29.8337 | 23.2110 |
| 0.0030 | 0.0029 | 22.4640 | 26.2587 | 32.8426 | 26.2497 |
| 0.0010 | 0.0010 | 27.1694 | 31.0175 | 37.6117 | 31.0254 |
| $\rho = 0.1$, $Q = 0.1$ | | | | | |
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0914 | 11.3236 | 15.7955 | 27.2882 | 15.1322 |
| 0.0600 | 0.0549 | 13.8997 | 18.8115 | 30.5762 | 18.3927 |
| 0.0300 | 0.0298 | 17.8510 | 23.1455 | 35.0377 | 22.8742 |
| 0.0100 | 0.0097 | 24.3888 | 30.0665 | 42.1091 | 29.9915 |
| 0.0060 | 0.0060 | 27.5188 | 33.2905 | 45.3971 | 33.1729 |
| 0.0030 | 0.0030 | 31.9391 | 37.7784 | 49.8586 | 37.7489 |
| 0.0010 | 0.0010 | 38.9244 | 44.8407 | 56.9300 | 44.8114 |

throughout this section, we restrict our attention to the "i.i.d. case" where the observations in a particular sequence, say $\{X_{\ell,n}\}_{n\geqq 1}$, are independently conditioned on $\lambda$, have a common PDF $f_\ell^{(0)}$ before the change, and a common PDF $f_\ell^{(1)}$ from the time of the change. Note that we are assuming that all the sensors change distribution at the change time $\lambda$. As in section 4, we will suppose that the prior distribution is geometric with the parameter $\rho$, $\rho > 0$.

Based on the information available at $S_\ell$ at time $n$, a message $U_{\ell,n}$, belonging to a finite alphabet of size $V_\ell$, is formed and sent to the fusion center. We will use the vector notation $\mathbf{X}_n = (X_{1,n}, \dots, X_{L,n})$ and $\mathbf{U}_n = (U_{1,n}, \dots, U_{L,n})$. Based on the sequence of sensor messages, a decision about the change is made at the fusion center. The fusion center picks a time $\tau$, which is a *stopping time* on $\{\mathbf{U}_n\}_{n\geqq 1}$, at which it is declared that a change has occurred.

Various information structures are possible for the decentralized configuration depending on how feedback and local information is used at the sensors [41]. Consider the simplest information structure, where the message $U_{\ell,n}$ formed by sensor $S_\ell$ at

Table 4
*Results for correlated case with* $\delta = 0.5$.

| $\rho = 0.1$, $Q = 1$ | | | | | |
|---|---|---|---|---|---|
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0839 | 2.9519 | 3.3721 | 5.6139 | 3.1324 |
| 0.0600 | 0.0532 | 3.5588 | 4.0728 | 6.4577 | 3.9739 |
| 0.0300 | 0.0289 | 4.5526 | 5.1437 | 7.6027 | 5.1293 |
| 0.0100 | 0.0100 | 6.2505 | 6.9137 | 9.4175 | 6.9328 |
| 0.0060 | 0.0062 | 7.0655 | 7.7425 | 10.2614 | 7.7790 |
| 0.0030 | 0.0029 | 8.1902 | 8.8830 | 11.4064 | 8.9244 |
| 0.0010 | 0.0010 | 9.9847 | 10.6885 | 13.2212 | 10.7444 |
| $\rho = 0.1$, $Q = 0.25$ | | | | | |
| $\alpha$ | MCPFA | MCADD | $\text{MCCADD}_1$ | FOADD | $\text{HOCADD}_1$ |
| 0.1000 | 0.0895 | 7.8914 | 10.6258 | 17.6206 | 10.3019 |
| 0.0600 | 0.0568 | 9.7496 | 12.7327 | 19.8381 | 12.6246 |
| 0.0300 | 0.0289 | 12.3524 | 15.5596 | 22.8471 | 15.6679 |
| 0.0100 | 0.0098 | 16.7599 | 20.2234 | 27.6162 | 20.3995 |
| 0.0060 | 0.0061 | 18.8899 | 22.4150 | 29.8337 | 22.5276 |
| 0.0030 | 0.0030 | 21.8352 | 25.3999 | 32.8426 | 25.6374 |
| 0.0010 | 0.0010 | 26.5485 | 30.1661 | 37.6117 | 30.4034 |

time $n$ is a function of only its current observation $X_{\ell,n}$, i.e., $U_{\ell,n} = \psi_{\ell,n}(X_{\ell,n})$. Moreover, since for a particular $\ell$ the sequence $\{X_{\ell,n}\}_{n\geqq 1}$ is assumed to be i.i.d., it is natural to confine ourselves to *stationary* quantizers[2] for which the quantizing functions $\psi_{\ell,n}$ do not depend on $n$, i.e., $\psi_{\ell,n} = \psi_\ell$ for all $n \geqq 1$.

The set of quantizing functions $\{\psi_\ell, \ell = 1, \dots, L\} = \Psi$, together with the fusion center stopping time $\tau$, form a policy $\phi = (\tau, \Psi)$. The goal is to choose the policy $\phi$ that minimizes the $\text{ADD}(\phi) = \mathbf{E}^\pi\{\tau - \lambda \,|\, \tau \geqq \lambda\}$, or more generally the moments of the detection delay $\mathbf{ED}_m^\pi(\phi) = \mathbf{E}^\pi\{(\tau-\lambda)^m \,|\, \tau \geqq \lambda\}$ for all $m \geqq 1$, while maintaining the probability of the false alarm $\text{PFA}(\phi) = \mathbf{P}^\pi\{\tau < \lambda\}$ at a level not greater than $\alpha$.

Let $H_k$ be the hypothesis that the change occurs at time $\lambda = k \in \{1, 2, \dots\}$, and let $H_\infty$ be the hypothesis that the change does not occur at all. Since the observations at each sensor $S_\ell$, $\{X_{\ell,n}, n = 1, 2, \dots\}$, are i.i.d., for stationary sensor quantizers, the sensor outputs, $\{U_{\ell,n}, n = 1, 2, \dots\}$, will also be i.i.d. Let $g_\ell^{(j)}$ denote the PMF (probability mass function) induced on $U_{\ell,n}$ when the observation $X_{\ell,n}$ is distributed as $f_\ell^{(j)}$, $j = 0, 1$.

Then, for *fixed* stationary sensor quantizers, the LLRs between the hypotheses $H_k$ and $H_\infty$ at the sensor $S_\ell$ and at the fusion center are given by

$$Z_n^k(\ell) = \sum_{i=k}^{n} \log \frac{g_\ell^{(1)}(U_{\ell,i})}{g_\ell^{(0)}(U_{\ell,i})} \quad \text{and} \quad Z_n^k = \sum_{\ell=1}^{L} Z_n^k(\ell).$$

For fixed sensor quantizers, the fusion center faces a standard change-point detection problem based on the vector observation sequence $\{\mathbf{U}_n\}$. Hence we can define the average likelihood ratio statistic $\Lambda_n^{\text{dc}}$ and the corresponding statistic $R_{\rho,n}^{\text{dc}} = \Lambda_n^{\text{dc}}/\rho$ with $f_1(X_n)/f_0(X_n)$ now replaced by $\prod_{\ell=1}^L [g_\ell^{(1)}(U_{\ell,n})/g_\ell^{(0)}(U_{\ell,n})]$. The index "dc" will be used to denote parameters associated with the decentralized detection problem.

[2]We can prove the optimality of stationary quantizers under some mild conditions on the observations and the quantizers.

The decentralized Shiryaev detection procedure at the fusion center $\nu_B^{\rm dc}$ is given by

$$\nu_B^{\rm dc} = \inf\{n \geqq 1 \colon\ R_{\rho,n}^{\rm dc} \geqq B\}, \tag{7.1}$$

where $B$ is a positive threshold which is selected so that $\mathrm{PFA}(\nu_B) \leqq \alpha$.

If $D(g_\ell^{(1)}, g_\ell^{(0)})$, the KL distances between the $g_\ell^{(1)}$ and $g_\ell^{(0)}$, are positive and finite, then for fixed stationary sensor quantizers, an application of Theorem 4 gives us that the detection procedure $\nu_B^{\rm dc}$ given in (7.1), with $B = B_\alpha = (1-\alpha)/(\alpha\rho)$, is asymptotically optimal as $\alpha \to 0$ among all procedures with the PFA no greater than $\alpha$. To be specific, let $\Psi = \{\psi_1, \dots, \psi_L\}$ be a set of stationary quantizers. Then, as $\alpha \to 0$, for all $m \geqq 1$

$$\inf_{\tau \in \Delta(\alpha)} \mathbf{ED}_m^\pi(\Psi, \tau) \sim \mathbf{ED}_m^\pi(\Psi, \nu_{B_\alpha}^{\rm dc}) \sim \left( \frac{|\log \alpha|}{\sum_{\ell=1}^L D(g_\ell^{(1)}, g_\ell^{(0)}) + |\log(1-\rho)|} \right)^m,$$

where $\mathbf{ED}_m^\pi(\Psi, \tau) = \mathbf{E}^\pi\{(\tau - \lambda)^m \,|\, \tau \geqq \lambda\}$ is the $m$th moment of the detection delay for the policy $(\Psi, \tau)$.

This result immediately reveals how to optimize the sensor quantizers.

Corollary 3. *It is asymptotically optimum* (*as* $\alpha \to 0$) *for sensor* $S_\ell$ *to use the stationary quantizer that maximizes the KL information distance at its output*; *i.e.*, $\psi_{\ell,\rm opt} = \arg\max D(g_\ell^{(1)}, g_\ell^{(0)})$, $\ell = 1, \dots, L$.

Based on the results of Tsitsiklis [40], it is easy to show that the optimum stationary quantizer $\psi_{\ell,\rm opt}$ is a monotone likelihood ratio quantizer; i.e., there exist thresholds $h_{\ell,1}, \dots, h_{\ell,V_\ell - 1}$ satisfying $0 = h_{\ell,0} \leqq h_{\ell,1} \leqq \cdots \leqq h_{\ell,V_\ell-1} \leqq \infty = h_{\ell,V_\ell}$ such that

$$\psi_{\ell,\rm opt}(X) = i \qquad \text{only if} \quad h_{\ell,i-1} < \frac{f_\ell^{(1)}(X)}{f_\ell^{(0)}(X)} \leqq h_{\ell,i}, \qquad i = 1, \dots, V_\ell.$$

Thus, the asymptotically optimal policy $\phi_{\rm opt}$ for the decentralized change detection problem in the class of stationary (in time) quantizers consists of a set of monotone likelihood ratio quantizers at the sensors followed by Shiryaev's procedure based on $\{\mathbf{U}_n\}_{n \geqq 1}$ at the fusion center (as described in (7.1)).

For each $\ell$, let the PMFs induced on $U_{\ell,n}$ by the optimum monotone likelihood ratio quantizer $\psi_{\ell,\rm opt}$ be given by $g_{\ell,\rm opt}^{(1)}$ and $g_{\ell,\rm opt}^{(0)}$. Then the effective KL information distance between the "change" and "no change" hypotheses at the fusion center is given by

$$D_{\rm tot} = \sum_{\ell=1}^L D(g_{\ell,\rm opt}^{(1)}, g_{\ell,\rm opt}^{(0)}).$$

Finally, denote by $\nu_{\rm opt}^{\rm dc}$ Shiryaev's stopping rule at the fusion center for the case where the sensor quantizers are chosen to be $\psi_{\ell,\rm opt}$, and by $\Phi_{\rm st}(\alpha)$ the class of policies $\phi$ with all stationary quantizers and stopping rules at the fusion center such that $\tau \in \Delta(\alpha)$.

The asymptotic performance of the asymptotically optimal solution to the decentralized change detection problem described above is given in the following theorem, which follows directly from Theorem 4 and the argument given above.

Theorem 7. *Suppose that*

$$0 < D\big(g_{\ell,\rm opt}^{(1)}, g_{\ell,\rm opt}^{(0)}\big) < \infty \qquad \textit{for} \quad \ell = 1, \dots, L.$$

*Then* $B_\alpha = (1-\alpha)/(\alpha\rho)$ *implies that* $\mathrm{PFA}(\nu_{\mathrm{opt}}^{\mathrm{dc}}) \leqq \alpha$ *and, for all* $m \geqq 1$,

$$\inf_{\phi\in\Phi_{\mathrm{st}}(\alpha)} \mathbf{ED}_m^\pi(\phi) \sim \mathbf{ED}_m^\pi(\phi_{\mathrm{opt}}) \sim \left(\frac{|\log\alpha|}{D_{\mathrm{tot}} + |\log(1-\rho)|}\right)^m \qquad as \quad \alpha \to 0,$$

*where* $\phi_{\mathrm{opt}} = (\nu_{\mathrm{opt}}^{\mathrm{dc}}, \{\psi_{\ell,\mathrm{opt}}\})$.

Theorem 5 can also be applied to the problem in question. Specifically, if in addition to the conditions of Theorem 7, we assume that the LLR at the fusion center $Z_1^1 = \sum_{\ell=1}^L \log[g_{\ell,\mathrm{opt}}^{(1)}(U_{\ell,1})/g_{\ell,\mathrm{opt}}^{(0)}(U_{\ell,1})]$ is nonarithmetic, then the PFA satisfies asymptotic formula (4.18) with $B$ replaced by $B\rho$ and

$$\mathbf{E}_1\nu_B = \frac{1}{D_{\mathrm{tot}} + |\log(1-\rho)|}\left[\log B - C(\rho, D_{\mathrm{tot}}) + \overline{\varkappa}(\rho, D_{\mathrm{tot}})\right] + o(1) \tag{7.2}$$

with the corresponding modification of the definitions of $\overline{\varkappa}(\rho, D_{\mathrm{tot}})$ and $C(\rho, D_{\mathrm{tot}})$.

**7.2. Example 2: Decentralized detection of a change in the mean of a normal population.** Surveillance systems, such as those used in defense, detect and track moving targets that appear and disappear at unknown points in time. As a result, the target detection problem can be naturally formulated as a multisensor abrupt change detection problem as considered in subsection 7.1. We now consider an example of interest in target detection theory. In the centralized setting, this example is a particular case of Example 1. Here we consider the decentralized problem discussed above.

Consider the problem of detecting a nonfluctuating target using $L$ geographically separated sensors. The observations are corrupted by additive white Gaussian noise that is independent from sensor to sensor. The sensors preprocess the observations using a filter matched to the signal corresponding to the target (see [24]). The output of the matched filter at sensor $S_\ell$ at time $n$ (when the time of appearance of the target is $\lambda$) is given by

$$X_{\ell,n} = \begin{cases} \xi_{\ell,n} & \text{if} \quad n < \lambda, \\ \mu_\ell + \xi_{\ell,n} & \text{if} \quad n \geqq \lambda, \end{cases}$$

where $\{\xi_{\ell,n}, n = 1, 2, \dots\}$ is a sequence of i.i.d. zero-mean Gaussian random variables with variance $\sigma_\ell^2$. Therefore, the likelihood ratio at sensor $S_\ell$ is given by

$$Y_\ell(x) = \frac{f_\ell^{(1)}(x)}{f_\ell^{(0)}(x)} = \exp\left\{\frac{\mu_\ell(x-\mu_\ell/2)}{\sigma_\ell^2}\right\}. \tag{7.3}$$

Since $Y_\ell$ is monotonically increasing, we can characterize the optimum stationary sensor quantizers in terms of thresholds on the observations, rather than on their likelihood ratios. To further simplify the example, we assume that the sensor messages are binary, i.e., $V_\ell = 2$ for all $\ell$. Then the quantizers reduce to binary tests that are characterized by a single threshold, i.e.,

$$U_{\ell,n} = \begin{cases} 1 & \text{if } X_{\ell,n} \geqq h_\ell, \\ 0 & \text{otherwise.} \end{cases}$$

The distributions induced on $U_{\ell,n}$ by this quantizer are given by

$$g_\ell^{(j)}(0) = 1 - g_\ell^{(j)}(1) = \Phi\left(\frac{h_\ell - j\mu_\ell}{\sigma_\ell}\right) = q_\ell^{(j)}, \qquad j = 0, 1, \tag{7.4}$$

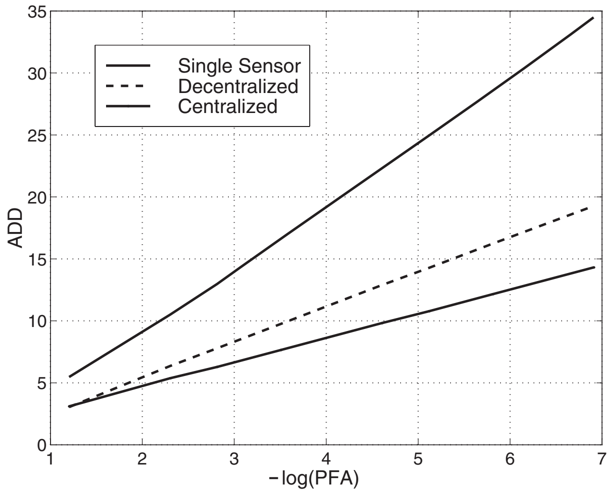


FIG. 1. *Operating characteristics for an example with five sensors with identically distributed Gaussian observations.*

where $\Phi(\cdot)$ is the distribution function of a standard Gaussian random variable. The optimum value of $h_\ell$, i.e., the one that maximizes $D(g_\ell^{(1)}, g_\ell^{(0)})$, is easily found based on (7.4). Then we can compute the decision statistic $R_{\rho,n}$ at the fusion center, which obeys the recursion (assuming that $\pi_0 = 0$)

$$R_{\rho,n} = \frac{1}{1-\rho}\,(1 + R_{\rho,n-1})\, e^{Z_n^n}, \qquad n \geqq 1, \quad R_{\rho,0} = 0, \tag{7.5}$$

with

$$Z_n^n = \sum_{\ell=1}^{L} \left( U_{\ell,n} \log \left[ \frac{1 - q_\ell^{(1)}}{1 - q_\ell^{(0)}} \, \frac{q_\ell^{(0)}}{q_\ell^{(1)}} \right] - \log \left[ \frac{q_\ell^{(0)}}{q_\ell^{(1)}} \right] \right). \tag{7.6}$$

Based on (7.6), we may also compute HO approximations for the PFA and ADD, as given in Theorem 5, using the technique given in [42, section 2.4].

The operating characteristics in an example with five sensors having identically distributed observations are illustrated in Figure 1. The parameter values are $\rho = 0.1$, $\mu_\ell = 0.4$, and $\sigma_\ell^2 = 1$. The KL distance for the sensor observations is 0.08. The threshold that maximizes the KL distance at the output of the sensor is $h = 0.32$, and the corresponding maximum KL distance is 0.0509. The fusion center threshold is set using $B = (1 - \alpha)/(\rho\alpha)$. Estimates for the PFA and ADD were obtained using Monte Carlo methods with the number of trials being $1000/\alpha$. We plot ADD versus $-\log \mathrm{PFA}$ for the optimum decentralized detection policy and compare the performance with that of a centralized policy that has direct access to the observations at the sensors. As we expect, for the centralized policy, the plot of ADD versus $-\log(\mathrm{PFA})$ is a straight line with a slope that is approximately equal to

$$\frac{1}{5D(f^{(1)}, f^{(0)}) + \log(1-\rho)} \approx 1.98.$$

For the optimum decentralized policy, the trade-off curve has a slope that is roughly equal to

$$\frac{1}{D_{\mathrm{tot}} + \log(1-\rho)} \approx 2.78$$

as expected from Theorem 7. The decentralized policy of course suffers a performance degradation relative to the centralized policy. However, the bandwidth requirements for communication with the fusion center are considerably smaller in a decentralized setting, especially with binary quantizers. Figure 1 also shows the trade-off curve for a centralized detection policy with a *single* sensor. As expected, the slope of ADD versus $-\log \mathrm{PFA}$ is five times larger. Furthermore, it can be seen that even if the sensor observations are quantized to one bit, the decentralized policy with five sensors far outperforms the single sensor centralized policy.

**7.3. Example 3: Decentralized detection of a change in a Poisson sequence.** In distributed computer networks, large-scale attacks in their final stages can be readily identified by observing very abrupt changes in the network traffic. However, in the early stage of an attack, these changes are hard to detect and difficult to distinguish from usual traffic patterns. In this subsection, we argue that the Shiryaev detection algorithm can be effectively deployed for an early detection of intrusions from the class of denial-of-service attacks. An efficient nonparametric approach to this problem has been recently proposed by Blažek et al. [4]. Here we consider a parametric approach with a Poisson model for the observables.

Assume that sensor observations are Poisson random variables with different means before and after the disruption. For instance, in the network security applications, $X_{\ell,n}$ may correspond to the number of packets of a particular type (say, TCP-packets) at sensor $S_\ell$ in the $n$th time interval of a certain length. Let the observations at sensor $S_\ell$ have mean $\mu_{0,\ell}$ before the disruption, and mean $\mu_{1,\ell}$ after the disruption. Without loss of generality assume that $\mu_{1,\ell} > \mu_{0,\ell}$. Then the likelihood ratio at $S_\ell$ is given by

$$Y_\ell(X_{\ell,n}) = \left(\frac{\mu_{1,\ell}}{\mu_{0,\ell}}\right)^{X_{\ell,n}} \exp\left\{-(\mu_{1,\ell} - \mu_{0,\ell})\right\}.$$

Note that the likelihood ratio is again monotonically increasing, and hence, we can characterize the optimum stationary sensor quantizers in terms of thresholds on the observations. For binary quantizers,

$$U_{\ell,n} = \begin{cases} 1 & \text{if} \quad X_{\ell,n} \geqq h_\ell, \\ 0 & \text{otherwise.} \end{cases}$$

The distributions induced on $U_{\ell,n}$ by this quantizer are given by

$$g_\ell^{(j)}(0) = 1 - g_\ell^{(j)}(1) = \sum_{k=0}^{\lfloor h_\ell \rfloor} \frac{\mu_{j,\ell}^k \, e^{-\mu_{j,\ell}}}{k!} = q_\ell^{(j)}, \qquad j = 0, 1. \tag{7.7}$$

Here again, the optimum value of $h_\ell$, i.e., the one that maximizes $D(g_\ell^{(1)}, g_\ell^{(0)})$, is easily found based on (7.7). The decision statistic at the fusion center is then given by (7.5) and (7.6).

The operating characteristics in an example with three sensors having identically distributed observations are illustrated in Figure 2. The parameter values are $\rho = 0.1$,

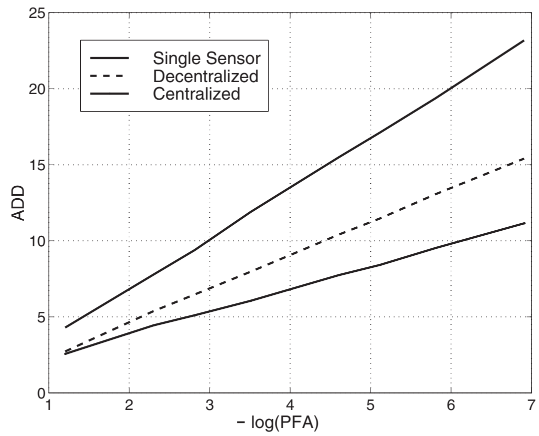


FIG. 2. *Operating characteristics for an example with three sensors with identically distributed Poisson observations.*

$\mu_{0,\ell} = 10$, and $\mu_{1,\ell} = 12$. The KL distance for the sensor observations is 0.1879. The threshold that maximizes the KL distance at the output of the sensor is $h = 11$, and the corresponding maximum KL distance is 0.119. The fusion center threshold is set using $B = (1-\alpha)/(\rho\alpha)$, and estimates for the PFA and ADD were obtained using Monte Carlo methods. As in the previous example, we see that the plots of ADD versus $-\log \mathrm{PFA}$ in the three cases considered have the behavior predicted by the theory.

**8. Conclusions.** We end by giving the following concluding remarks.

1. Most of the asymptotic optimality results remain true for stochastic processes observed in continuous time. However, continuous-time problems have certain special features that should be handled carefully. A general asymptotic detection theory for continuous-time models will be presented elsewhere.

2. The general asymptotic theory that has been developed in this paper covers only simple hypotheses and can be considered as the first step. For most practical applications it is important to consider composite hypotheses, especially in the post-change mode. Mixture-type and adaptive versions of the Shiryaev Bayesian rule are excellent candidates for composite-hypothesis problems. Adaptive Bayesian modifications seem to be especially attractive for on-line implementations.

3. For the decentralized detection problem discussed in section 7, it is of interest to extend the asymptotic analysis to non-i.i.d. observations at the sensors and to possible correlation across sensors (conditioned on the change point). The extension to non-i.i.d. observations is straightforward, whereas the extension to include correlation across sensors appears to be nontrivial.

4. The results of section 7 show that fusion of data in decentralized multisensor systems with quantizers always leads to a certain loss of information which results in the performance degradation of the optimal decentralized policy. Specifically, for the geometric prior distribution the asymptotic relative efficiency of the optimal centralized detection procedure with respect to decentralized is equal to

$$\lim_{\alpha\to 0} \frac{\inf_{\tau\in\Delta(\alpha)} \mathrm{ADD}_{\mathrm{c}}(\tau)}{\inf_{\tau\in\Delta(\alpha)} \mathrm{ADD}_{\mathrm{dc}}(\tau)} = \frac{|\log(1-\rho)| + \sum_{\ell=1}^{L} D(g_\ell^{(1)}, g_\ell^{(0)})}{|\log(1-\rho)| + \sum_{\ell=1}^{L} D(f_\ell^{(1)}, f_\ell^{(0)})} < 1.$$

Interestingly, it is possible to construct decentralized detection procedures with no quantization that are asymptotically equivalent to the optimal centralized procedure (i.e., globally asymptotically optimal) and at the same time have bandwidth requirements for communications between sensors and the fusion center similar to decentralized policies with binary quantization. However, these procedures require significant processing capabilities at the sensors so that they can run individual change detection tests. Such procedures will be discussed in a separate paper.

## REFERENCES


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

[2] L. E. Baum and M. Katz, *Convergence rates in the law of large numbers*, Trans. Amer. Math. Soc., 120 (1965), pp. 108–123.

[3] M. Beibel, *Sequential detection of signals with known shape and unknown magnitude*, Statist. Sinica, 10 (2000), pp. 715–729.

[4] 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 Proceedings of the IEEE Workshop on Information Assurance and Security (West Point, NY), IEEE, Piscataway, NJ, 2001, pp. 220–226.

[5] A. A. Borovkov, *Asymptotically optimal solutions in the change-point problem*, Theory Probab. Appl., 43 (1998), pp. 539–561.

[6] Y. S. Chow and T. L. Lai, *Some one-sided theorems on the tail distribution of sample sums with application to the last time and largest excess of boundary crossings*, Trans. Amer. Math. Soc., 208 (1975), pp. 51–72.

[7] V. P. Dragalin, *Optimality of a generalized CUSUM procedure in quickest detection problem*, in Statistics and Control of Random Processes 202, Proceedings of the Steklov Institute of Mathematics, AMS, Providence, RI, 1994, pp. 107–120.

[8] V. P. Dragalin, A. G. Tartakovsky, and V. V. Veeravalli, *Multihypothesis sequential probability ratio tests.* I. *Asymptotic optimality*, IEEE Trans. Inform. Theory, 45 (1999), pp. 2448–2461.

[9] C. D. Fuh, *SPRT and CUSUM in hidden Markov models*, Ann. Statist., 31 (2003), pp. 942–977.

[10] A. Gut, *Stopped Random Walks: Limit Theorems and Applications*, Springer-Verlag, New York, 1988.

[11] P. L. Hsu and H. Robbins, *Complete convergence and the law of large numbers*, Proc. Natl. Acad. Sci. USA, 33 (1947), pp. 25–31.

[12] S. Kent, *On the trial of intrusions into information systems*, IEEE Spectrum, 37 (2000), pp. 52–56.

[13] T. L. Lai, *On r-quick convergence and a conjecture of Strassen*, Ann. Probab., 4 (1976), pp. 612–627.

[14] T. L. Lai, *Asymptotic optimality of invariant sequential probability ratio tests*, Ann. Statist., 9 (1981), pp. 318–333.

[15] T. L. Lai, *Sequential changepoint detection in quality control and dynamical systems*, J. Roy. Statist. Soc. Ser. B, 57 (1995), pp. 613–658.

[16] T. L. Lai, *Information bounds and quick detection of parameter changes in stochastic systems*, IEEE Trans. Inform. Theory, 44 (1998), pp. 2917–2929.

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

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

[19] E. S. Page, *Continuous inspection schemes*, Biometrika, 41 (1954), pp. 100–115.

[20] G. Peskir and A. N. Shiryaev, *Solving the Poisson disorder problem*, in Advances in Finance and Stochastics, Essays in Honour of Dieter Sondermann, K. Sandmann, and P. J. Schönbucher, eds., Springer, Berlin, 2002, pp. 295–312.

[21] M. Pollak, *Optimal detection of a change in distribution*, Ann. Statist., 13 (1985), pp. 206–227.

[22] M. Pollak, *Average run lengths of an optimal method of detecting a change in distribution*, Ann. Statist., 15 (1987), pp. 749–779.

[23] M. Pollak and D. Siegmund, *A diffusion process and its applications to detecting a change in the drift of Brownian motion*, Biometrika, 72 (1985), pp. 267–280.

[24] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.
[25] A. N. Shiryaev, *The detection of spontaneous effects*, Sov. Math. Dokl., 2 (1961), pp. 740–743.
[26] A. N. Shiryaev, *On optimum methods in quickest detection problems*, Theory Probab. Appl., 8 (1963), pp. 22–46.
[27] A. N. Shiryaev, *Optimal Stopping Rules*, Springer-Verlag, New York, 1978.
[28] A. N. Shiryaev, *Minimax optimality of the method of cumulative sum* (*cusum*) *in the case of continuous time*, Russian Math. Surveys, 51 (1996), pp. 750–751.
[29] D. Siegmund, *Sequential Analysis*: *Tests and Confidence Intervals*, Springer-Verlag, New York, 1985.
[30] A. G. Tartakovsky, *Sequential Methods in the Theory of Information Systems*, Radio i Svyaz Publ. House, Moscow, 1991 (in Russian).
[31] A. G. Tartakovsky, *Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in a nonhomogeneous Gaussian process*, Math. Methods Statist., 4 (1995), pp. 389–404.
[32] A. G. Tartakovsky, *Asymptotically minimax multialternative sequential rule for disorder detection*, in Statistics and Control of Random Processes 202, Proceedings of the Steklov Institute of Mathematics, AMS, Providence, RI, 1994, pp. 229–236.
[33] A. G. Tartakovsky, *Asymptotic optimality of certain multihypothesis sequential tests*: *Non-i.i.d. case*, Statist. Inference Stoch. Process., 1 (1998), pp. 265–295.
[34] A. G. Tartakovsky, *Extended Asymptotic Optimality of Certain Change-Point Detection Procedures*, Preprint, Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, 2003.
[35] A. G. Tartakovsky and V. V. Veeravalli, *An efficient sequential procedure for detecting changes in multichannel and distributed systems*, in Proceeding of the Fifth International Conference on Information Fusion (Annapolis, MD), IEEE, Piscataway, NJ, 2002, pp. 41–48.
[36] A. G. Tartakovsky and V. V. Veeravalli, *Asymptotic analysis of Bayesian quickest change detection procedures*, in Proceedings of the International Symposium on Information Theory (Lausanne), IEEE, Piscataway, NJ, 2002, p. 217.
[37] V. V. Veeravalli and A. G. Tartakovsky, *Asymptotics of quickest change detection procedures under a Bayesian criterion*, in Proceedings of the IEEE Information Theory Workshop (Bangalore), 2002, pp. 100–103.
[38] A. G. Tartakovsky and V. V. Veeravalli, *Quickest change detection in distributed sensor systems*, in Proceedings of the Sixth International Conference on Information Fusion (Cairns), International Society of Information Fusion, Chatillon, France, 2003, pp. 756–763.
[39] 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, New York, pp. 331–363.
[40] J. N. Tsitsiklis, *Extremal properties of likelihood-ratio quantizers*, IEEE Trans. Commun., 41 (1993), pp. 550–558.
[41] V. V. Veeravalli, *Decentralized quickest change detection*, IEEE Trans. Inform. Theory, IT-47 (2001), pp. 1657–1665.
[42] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*, SIAM, Philadelphia, PA, 1982.
[43] B. Yakir, *Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space*, in Change-Point Problems, E. Carlstein, H. Muller, and D. Siegmund, eds., IMS Lecture Notes Monogr. Ser. 23, Inst. Math. Statist., Hayward, CA, 1994, pp. 346–358.
[44] B. Yakir, *A note on optimal detection of a change in distribution*, Ann. Statist., 25 (1997), pp. 2117–2126.