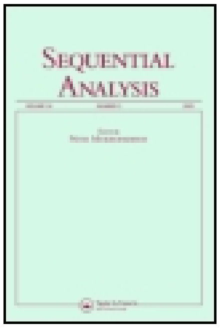

Sequential Analysis: Design Methods and Applications
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsqa20

Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems

Alexander G. Tartakovsky [a] & Venugopal V. Veeravalli [b]

[a] Department of Mathematics, University of Southern California, Los Angeles, California, USA

[b] Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Available online: 31 Oct 2008

Sequential Analysis, 27: 441–475, 2008

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474940802446236

Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems

Alexander G. Tartakovsky[1] **and Venugopal V. Veeravalli**[2]
[1]Department of Mathematics, University of Southern California, Los Angeles, California, USA
[2]Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Abstract: In the standard formulation of the quickest change-point detection problem, a sequence of observations, whose distribution changes at some unknown point in time, is available to a decision maker, and the goal is to detect this change as quickly as possible, subject to false alarm constraints. In this paper, we study the quickest change detection problem in the setting where the information available for decision-making is distributed across a set of geographically separated sensors, and only a compressed version of observations in sensors may be used for final decision-making due to communication bandwidth constraints. We consider the minimax, uniform, and Bayesian versions of the optimization problem, and we present asymptotically optimal decentralized quickest change detection procedures for two scenarios. In the first scenario, the sensors send quantized versions of their observations to a fusion center where the change detection is performed based on all the sensor messages. In the second scenario, the sensors perform local change detection and send their final decisions to the fusion center for combining. We show that our decentralized procedures for the latter scenario have the same first-order asymptotic performance as the corresponding centralized procedures that have access to all of the sensor observations. We also present simulation results for examples involving Gaussian and Poisson observations. These examples show that although the procedures with local decisions are globally asymptotically optimal as the false alarm rate (or probability) goes to zero, they perform worse than the corresponding decentralized procedures with binary quantization at the sensors, unless the false alarm rate (or probability) is unreasonably small.

Keywords: Change-point detection; CUSUM procedure; Distributed decision-making; Optimal fusion; Shiryaev procedure; Shiryaev–Roberts procedure.

Subject Classifications: 62L10; 62L15; 60G40; 62F12; 62F05.

Received October 14, 2007, Revised June 2, 2008, Accepted August 8, 2008
Recommended by Wolfgang Schmid
Address correspondence to Alexander G. Tartakovsky, Department of Mathematics, University of Southern California, KAP-108, Los Angeles, CA 90089-2532, USA; E-mail: tartakov@math.usc.edu and Venugopal V. Veeravalli, DECE and CSL, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA; E-mail: vvv@uiuc.edu

1. INTRODUCTION

An important application area for distributed sensor systems is environment surveillance and monitoring. Specific applications include intrusion detection in computer networks, intrusion detection in security systems, chemical or biological warfare agent detection systems to protect against terrorist attacks, detection of the onset of an epidemic, and failure detection in manufacturing systems and large machines. In all of these applications, the sensors monitoring the environment take observations that undergo a change in statistical distribution in response to the change in the environment. The goal is to detect this change-point as quickly as possible, subject to false alarm constraints.

In the standard formulation of the change-point detection problem, there is a sequence of observations whose distribution changes at some unknown point in time and the goal is to detect this change as soon as possible, subject to false alarm constraints (see, e.g., Basseville and Nikiforov, 1993; Lai, 1995, 1998; Lorden, 1971; Moustakides, 1986; Pollak, 1985, 1987; Shiryaev, 1978; Tartakovsky, 1991). In this paper, we are interested in the generalization of this problem that corresponds to the *multisensor* situation where the information available for decision-making is *distributed*. The observations are taken at a set of N distributed sensors as shown in Figure 1. The statistical properties of the sensors' observations change at the same time. If the sensor observations (or sufficient statistics) can be conveyed directly to a *fusion* center, then we are faced with a centralized change detection problem, which is equivalent to the standard formulation with vector observations. In this case a globally optimal solution with the best possible performance can be obtained. However, in most applications, due to communication constraints, the sensor observations cannot be transmitted directly to the fusion center and therefore the centralized solution is infeasible, in which case we need to consider decentralized approaches where the sensors send either quantized versions of their observations or local decisions to the fusion center. Nevertheless, the centralized solution may be used as a benchmark for comparing decentralized solutions based on compressed data from the sensors.

We consider both the conditional (minimax and uniform) problem formulation, in which the change-point is considered to be deterministic but unknown

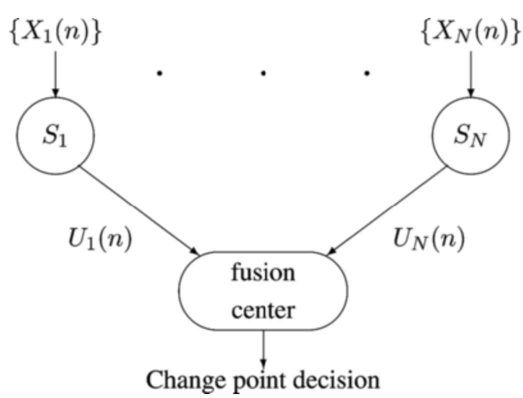

Figure 1. Change detection with distributed sensors.

(see, e.g., Basseville and Nikiforov, 1993; Lai, 1995, 1998; Lorden, 1971; Moustakides, 1986; Page, 1954; Pollak, 1985, 1987; Tartakovsky, 1991), and the Bayesian problem formulation in which the change-point is treated as a random variable with a known prior distribution (see, e.g., Baron and Tartakovsky, 2006; Shiryaev, 1963, 1978; Tartakovsky, 1991; Tartakovsky and Veeravalli, 2005). We begin with a study of centralized detection procedures for these problems, i.e., procedures where all the sensor observations are directly available to the fusion center for decision–making. Because the centralized problem is equivalent to the standard formulation with vector observations corresponding to the concatenation of the sensors' observations at each time, the CUSUM and Shiryaev–Roberts procedures are optimal for the minimax formulation (Lai, 1995, 1998; Lorden, 1971; Moustakides, 1986; Pollak, 1985, 1987) and Shiryaev's test is optimal for the Bayesian formulation (Baron and Tartakovsky, 2006; Shiryaev, 1963, 1978; Tartakovsky, 1991; Tartakovsky and Veeravalli, 2005), when the observations are independent and identically distributed (i.i.d.) in time before and after the change, as well as for more general non-i.i.d. stochastic models. We summarize the asymptotic performance of these centralized procedures in the i.i.d. case in the setting where the false alarm rate (or probability) goes to zero. This asymptotic performance serves as a benchmark in the comparison of decentralized detection procedures.

We then present asymptotically optimal decentralized quickest change detection procedures for two scenarios. In the first scenario, the sensors send quantized versions of their observations to the fusion center where the change detection is performed based on all the sensor messages. In the second scenario, the sensors perform local change detection and send their final decisions to the fusion center for combining. We show that our decentralized procedures for the latter scenario have the same first-order asymptotic performance as the corresponding centralized procedures that have access to all of the sensor observations. We also present Monte Carlo simulation results for examples involving Gaussian and Poisson observations. These examples show that although the procedures with local decisions are globally asymptotically optimal as the false alarm rate (or probability) goes to zero, they perform worse than the corresponding decentralized procedures with binary quantization at the sensors, unless the false alarm rate (or probability) is unreasonably small.

2. PROBLEM FORMULATION AND CENTRALIZED DETECTION PROCEDURES

Consider a distributed N-sensor system in which one observes an N-component stochastic process $\mathbf{X}(n) = (X_1(n), \ldots, X_N(n))$. The component $X_i(n)$, $n = 1, 2, \ldots$ corresponds to observations obtained from the ith sensor S_i, as shown in Figure 1. We begin by studying the centralized fusion problem where the sensors do not quantize their observations or make local decisions, i.e., $U_i(n) = X_i(n)$. At an unknown point in time λ ($\lambda = 1, 2 \ldots$), an event occurs and all of the components change their distribution. The observation sequences $\{X_1(n)\}, \{X_2(n)\}, \ldots, \{X_N(n)\}$ are assumed to be mutually independent, conditioned on the change-point. Moreover, we assume that at a particular sensor, the observations are i.i.d. before and after the change (with different distributions). If the change occurs at $\lambda = k$, then in sensor S_i the data $X_i(1), \ldots, X_i(k-1)$ follow the distribution $F_i^{(0)}$ with a density $f_i^{(0)}(x)$, whereas the data $X_i(k), X_i(k+1), \ldots$ have the common distribution $F_i^{(1)}$, with a density $f_i^{(1)}(x)$ (both with respect to a sigma-finite measure $\mu(x)$).

To be more specific, let $\mathbb{P}_\infty$ (correspondingly $\mathbb{E}_\infty$) stand for the probability measure (correspondingly expectation) when the change does not occur ($\lambda = \infty$), and let $\mathbb{P}_k$ (correspondingly $\mathbb{E}_k$) be the probability measure (correspondingly expectation) when the change occurs at time $\lambda = k$, $k < \infty$. Write $\mathbf{X}_i^n = (X_i(1), \ldots, X_i(n))$ and $\mathbf{X}^n = (\mathbf{X}_1^n, \ldots, \mathbf{X}_N^n)$ for the concatenation of the first n observations from the ith sensor and from all N sensors, respectively. Then, under $\mathbb{P}_\infty$, the density of $\mathbf{X}^n$ is

$$p_0(\mathbf{X}^n) = \prod_{i=1}^{N} \prod_{j=1}^{n} f_i^{(0)}(X_i(j)) \quad \text{for all } n \geq 1$$

and, under $\mathbb{P}_k$, the density of $\mathbf{X}^n$ is

$$p_k(\mathbf{X}^n) = \prod_{i=1}^{N} \left[\prod_{j=1}^{k-1} f_i^{(0)}(X_i(j)) \prod_{j=k}^{n} f_i^{(1)}(X_i(j)) \right] \quad \text{for } k \leq n$$

and $p_k(\mathbf{X}^n) = p_0(\mathbf{X}^n)$ for $k > n$.

Furthermore, let

$$Z_i(n) = \log \frac{f_i^{(1)}(X_i(n))}{f_i^{(0)}(X_i(n))} \tag{2.1}$$

be the log-likelihood ratio (LLR) between the change and no-change hypotheses for the nth observation from the ith sensor, and let

$$\mathcal{I}_i = \mathbb{E}_1 Z_i(1) = \int \log\left(\frac{f_i^{(1)}(x)}{f_i^{(0)}(x)}\right) f_i^{(1)}(x) \mu(dx) \tag{2.2}$$

be the Kullback–Leibler (K–L) information number (divergence between the densities $f_i^{(1)}(x)$ and $f_i^{(0)}(x)$).

A centralized sequential change-point detection procedure is identified with a stopping time τ for an observed sequence $\{\mathbf{X}(n)\}_{n\geq 1}$, i.e., τ is an extended integer-valued random variable, such that the event $\{\tau \leq n\}$ belongs to the sigma-algebra $\mathcal{F}_n = \sigma(\mathbf{X}^n)$ generated by the first n observations from all the sensors. 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 delay $\tau - \lambda$ provided that there is no false alarm (i.e., $\tau \geq \lambda$), whereas the rate of false positives should be low.

2.1. Bayesian Formulation

We first consider the Bayesian formulation of the change-point optimization problem, where the change-point λ is assumed to be random with prior probability distribution

$$\pi_k = \mathbb{P}(\lambda = k), \quad k = 0, 1, 2, \ldots,$$

where π_0 is the probability that the change occurs before observations become available. The goal is to detect the change as soon as possible after it occurs, subject to constraints on the false alarm probability.

In what follows, $\mathbb{P}^\pi$ stands for the average probability measure, which is defined as

$$\mathbb{P}^\pi(\bullet) = \sum_{k=0}^{\infty} \mathbb{P}_k(\bullet)\pi_k$$

and $\mathbb{E}^\pi$ denotes the expectation with respect to $\mathbb{P}^\pi$.

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

$$\mathrm{ADD}^\pi(\tau) = \mathbb{E}^\pi(\tau - \lambda \mid \tau \geq \lambda), \tag{2.3}$$

whereas the false alarm rate can be measured by the probability of false alarm

$$\mathrm{PFA}^\pi(\tau) = \mathbb{P}^\pi(\tau < \lambda) = \sum_{k=0}^{\infty} \pi_k \mathbb{P}_k(\tau < k). \tag{2.4}$$

An optimal Bayesian detection procedure is a procedure for which ADD is minimized while $\mathrm{PFA}(\tau)$ is set at a given level α, $0 < \alpha < 1$. Specifically, define the class of change-point detection procedures

$$\Delta_{\mathrm{B}}(\alpha) = \{\tau : \mathrm{PFA}^\pi(\tau) \leq \alpha\},$$

for which the false alarm probability does not exceed the predefined number α. The optimal Bayesian change-point detection procedure is described by the stopping time

$$\nu_{\mathrm{B}} = \arg \inf_{\tau \in \Delta_{\mathrm{B}}(\alpha)} \mathrm{ADD}^\pi(\tau).$$

2.2. Minimax Formulation

In the minimax formulation of the change-point detection problem, the change-point λ is assumed to be deterministic but unknown. In this case, the false alarm rate (FAR) can be measured by the average run length (ARL) to false alarm

$$\mathrm{ARL}(\tau) = \mathbb{E}_\infty \tau.$$

As a measure of the detection lag, we will use the supremum average detection delay (SADD) proposed by Pollak in mid 1970s

$$\mathrm{SADD}(\tau) = \sup_{1 \leq k < \infty} \mathbb{E}_k(\tau - k \mid \tau \geq k).$$

An optimal minimax detection procedure is a procedure for which $\mathrm{SADD}(\tau)$ is minimized while $\mathrm{ARL}(\tau)$ is set at a given level γ, $\gamma > 1$. Specifically, define the class of minimax change-point detection procedures

$$\Delta_{\mathrm{m}}(\gamma) = \{\tau : \mathrm{ARL}(\tau) \geq \gamma\}$$

for which the ARL exceeds the predefined positive number γ. The optimal minimax change-point detection procedure is described by the stopping time

$$\nu_{\mathrm{m}} = \arg \inf_{\tau \in \Delta_{\mathrm{m}}(\gamma)} \mathrm{SADD}(\tau). \tag{2.5}$$

2.3. Uniform Formulation

Although the Bayesian and minimax formulations are reasonable and can be justified in many applications, it would be most desirable in applications to guarantee small values of the conditional average detection delay $\mathrm{CADD}_k(\tau) = \mathbb{E}_k(\tau - k \mid \tau \geq k)$ uniformly for all $k \geq 1$ when the FAR is fixed at a certain level. However, if the FAR is measured in terms of the ARL to false alarm, i.e., it is required that $\mathrm{ARL}(\tau) \geq \gamma$ for some $\gamma > 1$, then a procedure that minimizes $\mathrm{CADD}_k(\tau)$ for all k does not exist. It is only possible to find minimax detection procedures that minimize $\mathrm{SADD}(\tau) = \sup_k \mathrm{ADD}_k(\tau)$ in the worst case scenario, as we discussed in the previous section. More importantly, the requirement of having large values of the mean time to false alarm $\mathbb{E}_\infty \tau$ generally does not guarantee small values of the local probability of false alarm (PFA) $\mathbb{P}_\infty(k \leq \tau < k + T)$ in a time interval of the fixed length $T \geq 1$ for all $k \geq 1$ or small values of the corresponding conditional local PFA $\mathbb{P}_\infty(\tau < k + T \mid \tau \geq k)$, $k \geq 1$. Indeed, the condition $\mathbb{E}_\infty \tau \geq \gamma$ only guarantees the existence of some k (that possibly depends on γ) for which $\mathbb{P}_\infty(\tau < k + T \mid \tau \geq k) < T/\gamma$ (cf. Tartakovsky, 2005). This means that, for a given $0 < \beta < 1$, the PFA constraint

$$\sup_{k \geq 1} \mathbb{P}_\infty(\tau < k + T \mid \tau \geq k) \leq \beta \quad \text{for a certain } T \geq 1 \tag{2.6}$$

is stronger than the ARL constraint $\mathbb{E}_\infty \tau \geq \gamma$.

At the same time, for many practical applications, including computer intrusion detection and a variety of surveillance applications such as target detection and tracking, it is desirable to control the supremum local PFA, which is given by $\mathrm{PFA}_T(\tau) = \sup_{k \geq 1} \mathbb{P}_\infty(\tau < k + T \mid \tau \geq k)$, at a certain (usually low) level β.

For this reason, we introduce the class of detection procedures that satisfy (2.6):

$$\Delta_{\mathrm{u}}^T(\beta) = \{\tau : \mathrm{PFA}_T(\tau) \leq \beta\},$$

i.e., for which $\mathrm{PFA}_T(\tau)$ does not exceed a predefined value $0 < \beta < 1$ for some $T \geq 1$.

The goal is to find a uniformly optimal change-point detection procedure that is described by the stopping time

$$\eta_{\mathrm{u}} = \arg \inf_{\tau \in \Delta_{\mathrm{u}}^T(\beta)} \mathrm{CADD}_k(\tau) \quad \text{for all } k \geq 1. \tag{2.7}$$

Although we do not address the solution to this optimization problem here, we discuss asymptotically uniformly optimal solutions in the following section.

2.4. CUSUM and Shiryaev–Roberts Centralized Detection Tests: Minimax and Uniform Optimality

It is known that the asymptotic performance of an optimal minimax centralized detection procedure that has access to all data $\mathbf{X}^n$ is given by

$$\inf_{\tau \in \Delta_{\mathrm{m}}(\gamma)} \mathrm{SADD}(\tau) = \frac{\log \gamma}{\mathcal{I}_{\mathrm{tot}}^c}(1 + o(1)), \quad \text{as } \gamma \to \infty, \tag{2.8}$$

where $\mathcal{I}_{\mathrm{tot}}^c = \sum_{i=1}^N \mathcal{I}_i$, with $\mathcal{I}_i$ being defined by (2.2). See, e.g., Basseville and Nikiforov (1993), Lai (1995, 1998), Pollak (1987), Tartakovsky (1991), and Tartakovsky and Veeravalli (2002). This performance is attained for the centralized CUSUM and Shiryaev–Roberts tests that use all available data, which are defined as

$$\tau^c(h) = \min\{n \geq 1 : W^c(n) \geq h\} \quad \text{and} \quad \hat{\tau}^c(h) = \min\{n \geq 1 : \log R^c(n) \geq h\}, \tag{2.9}$$

where the (centralized) CUSUM and Shiryaev–Roberts statistics $W^c(n)$ and $R^c(n)$ are given, respectively, by the recursions

$$W^c(n) = \max\left\{0, W^c(n-1) + \sum_{i=1}^N Z_i(n)\right\}, \quad W^c(0) = 0, \tag{2.10}$$

$$R^c(n) = [1 + R^c(n-1)] \exp\left\{\sum_{i=1}^N Z_i(n)\right\}, \quad R^c(0) = 0, \tag{2.11}$$

where $Z_i(n)$ is defined in (2.1) and threshold h is chosen so that $\mathrm{ARL}(\tau^c(h)) = \gamma$ and $\mathrm{ARL}(\hat{\tau}^c(h)) = \gamma$ (at least approximately). The Shiryaev–Roberts test will be introduced and discussed in more detail in Section 2.5 in connection with the Bayesian problem setting.

It is also known that $\mathrm{ARL}(\tau^c(h)) \geq \mathrm{ARL}(\hat{\tau}^c(h)) \geq e^h$, and hence, $h = \log \gamma$ guarantees $\mathrm{ARL}(\tau^c(h)) \geq \mathrm{ARL}(\hat{\tau}^c(h)) \geq \gamma$. See, e.g., Lorden (1971), Pollak (1987), and Tartakovsky (1991). The latter choice is usually conservative but useful for preliminary estimates and first-order asymptotic analysis. Substantial improvements can be obtained using corrected Brownian motion approximations (Siegmund, 1985) and the renewal theory arguments (Pollak, 1987; Tartakovsky, 2005). In particular, the following asymptotic approximations to the ARL are fairly accurate even for moderately low threshold values (cf. Pollak, 1987; Tartakovsky, 2005):

$$\mathrm{ARL}(\tau^c(h)) = \frac{e^h}{v^2 \mathcal{I}_{\mathrm{tot}}}(1 + o(1)), \quad \mathrm{ARL}(\hat{\tau}^c(h)) = v^{-1} e^h (1 + o(1)) \quad \text{as } h \to \infty, \tag{2.12}$$

where constant v depends on the pre- and postchange distributions and can be computed using renewal-theoretic argument (see, e.g., Siegmund, 1985; Woodroofe, 1982).

If instead of the ARL constraint (i.e., the class $\Delta_{\mathrm{m}}(\gamma)$) and the minimax setting (2.5) we are interested in the uniform optimality setting (2.7), then the asymptotic performance of a uniformly optimal centralized detection procedure that has access to all data $\mathbf{X}^n$ is given by

$$\inf_{\tau \in \Delta_{\mathrm{u}}^T(\beta)} \mathrm{CADD}_k(\tau) = \frac{|\log \beta|}{\mathcal{I}_{\mathrm{tot}}^c}(1 + o(1)), \quad \text{as } \beta \to 0, \tag{2.13}$$

which follows from Tartakovsky (2005).

The asymptotic performance (2.13) is attained again for the centralized CUSUM and Shiryaev–Roberts tests given in (2.9), with the thresholds h chosen so that $\mathrm{PFA}_T(\tau^c(h)) = \beta$ and $\mathrm{PFA}_T(\hat{\tau}^c(h)) = \beta$. To this end, we may use the results of Pollak and Tartakovsky (2008a) and Tartakovsky (2005), which state that the distributions of the suitably standardized stopping times $\tau^c(h)$ and $\hat{\tau}^c(h)$ are asymptotically exponential as $h \to \infty$. More specifically, $\tau^c(h)e^{-h}v^2\mathcal{I}_{\mathrm{tot}}^c$ and $\hat{\tau}^c(h)e^{-h}v$ converge weakly to Exponential(1) and the moment generating functions to that of Exponential(1) when h goes to infinity. Therefore, for sufficiently large h, $\mathrm{PFA}_T(\tau^c(h)) \approx 1 - (1 - e^{-h}v^2\mathcal{I}_{\mathrm{tot}}^c)^T$ and $\mathrm{PFA}_T(\hat{\tau}^c(h)) \approx 1 - (1 - e^{-h}v)^T$, and by selecting $h_\beta = \log(Tv^2\mathcal{I}_{\mathrm{tot}}^c/\beta)$ for the CUSUM test and $h_\beta = \log(Tv/\beta)$ for the Shiryaev–Roberts test, we guarantee $\mathrm{PFA}_T(\tau^c(h_\beta)) \approx \mathrm{PFA}_T(\hat{\tau}^c(h_\beta)) \approx \beta$. The latter approximations are asymptotically accurate. Furthermore, because for every $k \geq 1$,

$$\mathrm{CADD}_k(\tau^c(h)) = h/\mathcal{I}_{\mathrm{tot}}^c(1 + o(1)) \quad \text{and}$$
$$\mathrm{CADD}_k(\hat{\tau}^c(h)) = h/\mathcal{I}_{\mathrm{tot}}^c(1 + o(1)) \quad \text{as } h \to \infty,$$

it follows form (2.13) that with this choice of thresholds, for every $k \geq 1$,

$$\inf_{\tau \in \Delta_{\mathrm{u}}^T(\beta)} \mathrm{CADD}_k(\tau) \sim \mathrm{CADD}_k(\hat{\tau}^c(h_\beta)) \sim \mathrm{CADD}_k(\tau^c(h_\beta)) \sim \frac{|\log \beta|}{\mathcal{I}_{\mathrm{tot}}^c} \quad \text{as } \beta \to 0,$$

i.e., these detection procedures are indeed uniformly asymptotically optimal in the class $\Delta_{\mathrm{u}}^T(\beta)$.

2.5. Shiryaev and Shiryaev–Roberts Centralized Tests: Bayesian and Quasi-Bayesian Optimality

Let $p_n = \mathbb{P}(\lambda \leq n \mid \mathbf{X}^n)$ be the posterior probability that the change occurred before time n. It follows from works of Shiryaev (1961, 1963, 1978) that if the distribution of the change-point is geometric (i.e., $\pi_k = (1 - \pi_0)\rho(1 - \rho)^{k-1}$, $k \geq 1$), then the optimal centralized detection procedure is the one that raises an alarm at the first time such that the posterior probability p_n exceeds a threshold A,

$$v(A) = \min\{n \geq 1 : p_n \geq A\}, \tag{2.14}$$

where the threshold $A = A_\alpha$ should be chosen in such a way that $\mathrm{PFA}^\rho(v(A)) = \alpha$, where

$$\mathrm{PFA}^\rho(v) = (1 - \pi_0) \sum_{k=1}^{\infty} \rho(1 - \rho)^{k-1} \mathbb{P}_k(v < k)$$

is the average probability of false alarm (see (2.4)). However, it is difficult to find a threshold that provides an exact match to the given PFA. Also, until recently there were no results related to the ADD evaluation of this optimal procedure (except for the continuous-time Wiener process).

Although the exact match of the false alarm probability is related to the estimation of the overshoot in the stopping rule (2.14), and for this reason is problematic, a simple upper bound, which ignores overshoot, can be obtained. Indeed, because $\mathbb{P}^{\pi}\{\nu(A) < \lambda\} = \mathbb{E}^{\pi}\{1 - p_{\nu(A)}\}$ and $1 - p_{\nu(A)} \leq 1 - A$ on $\{\nu(A) < \infty\}$, it follows that the PFA obeys the inequality

$$\mathrm{PFA}^{\rho}(\nu(A)) \leq 1 - A. \tag{2.15}$$

Thus, setting $A = A_{\alpha} = 1 - \alpha$ guarantees the inequality $\mathrm{PFA}(\nu(A_{\alpha})) \leq \alpha$. Note that inequality (2.15) holds true for arbitrary (proper), not necessarily geometric, prior distributions and for arbitrary non-i.i.d. models.

In the rest of the paper, we assume that $\mathbb{P}(\lambda = 0) = \pi_0 = 0$ and the prior distribution of the change-point is geometric with the parameter ρ, $0 < \rho < 1$, i.e.,

$$\pi_k = \mathbb{P}(\lambda = k) = \rho(1-\rho)^{k-1} \quad \text{for } k = 1, 2, \ldots. \tag{2.16}$$

For $k \leq n$, introduce the following two statistics

$$\Lambda_n^k := \frac{d\mathbb{P}_k}{d\mathbb{P}_{\infty}}(\mathbf{X}^n) = \prod_{t=k}^{n}\prod_{i=1}^{N}\frac{f_1^{(i)}(X_i(t))}{f_0^{(i)}(X_i(t))} \tag{2.17}$$

and

$$R_{\rho}^{c}(n) = \sum_{k=1}^{n}(1-\rho)^{k-1-n}\Lambda_n^k. \tag{2.18}$$

Note that the statistic $R_{\rho}^{c}(n)$ satisfies the recursion

$$R_{\rho}^{c}(n) = \frac{1}{1-\rho}[1 + R_{\rho}^{c}(n-1)]\Lambda_n^n, \quad R_{\rho}^{c}(0) = 0. \tag{2.19}$$

which is useful in implementation.

Considering that $R_{\rho}^{c}(n) = p_n/[(1-p_n)\rho]$, the centralized Shiryaev stopping rule given in (2.14) can be written in the following form

$$\nu^{c}(B_{\rho}) = \min\{n \geq 1 : R_{\rho}^{c}(n) \geq B_{\rho}\}, \quad B_{\rho} = \frac{A}{(1-A)\rho}. \tag{2.20}$$

Consequently,

$$B_{\rho,\alpha} = (1-\alpha)/(\alpha\rho) \quad \text{implies } \nu^{c}(B_{\rho,\alpha}) \in \Delta_{\mathrm{B}}(\alpha). \tag{2.21}$$

It is worth mentioning that the Shiryaev procedure (2.20) is optimal not only in the i.i.d. case, but also asymptotically optimal in a non-i.i.d. scenario when α approaches zero under fairly general conditions in both discrete and continuous time cases, as has been recently shown by Tartakovsky and Veeravalli (2005) and Baron and Tartakovsky (2006).

In addition to the Bayesian Shiryaev procedure, we will also be interested in a related procedure, namely the Shiryaev–Roberts detection procedure. The Shiryaev–Roberts (SR) procedure is defined by the stopping time

$$\hat{\tau}^c(B) = \min\{n \geq 1 : R^c(n) \geq B\} \tag{2.22}$$

(compare with (2.9)) where the statistic $R^c(n)$ is given by

$$R^c(n) = \sum_{k=1}^{n} \Lambda_n^k. \tag{2.23}$$

Note that this statistic obeys the recursion (2.11).

It can be seen that the SR stopping rule (2.22) represents the limiting form of Shiryaev's stopping rule (2.20) as $\rho \to 0$, i.e., for the improper uniform prior distribution. This useful property has been previously noticed by Pollak (1985) and can be used for establishing an exact optimality property of the SR procedure with respect to the integral average detection delay $\mathrm{IADD}(\tau) = \sum_{k=1}^{\infty} \mathbb{E}_k(\tau - k)^+$. To be specific, let the threshold $B = B_\gamma$ be chosen so that $\mathbb{E}_\infty \hat{\tau}^c(B_\gamma) = \gamma$ (exactly). Then the SR procedure minimizes the integral average detection delay in the class $\Delta_{\mathrm{m}}(\gamma)$ (for every $\gamma > 1$):

$$\inf_{\tau \in \Delta_{\mathrm{m}}(\gamma)} \sum_{k=1}^{\infty} \mathbb{E}_k(\tau - k)^+. \tag{2.24}$$

The intuition behind this statement is as follows. First, taking the limit in (2.18), it is easy to see that $R^c(n) = \lim_{\rho \to 0} R_\rho^c(n)$. Second, the threshold in the Shiryaev rule (normalized by ρ) is chosen as in (2.21), i.e.,

$$B_\rho = \frac{1 - \mathrm{PFA}^\rho}{\rho \mathrm{PFA}^\rho},$$

where $\mathrm{PFA}^\rho \to 1$ as $\rho \to 0$ and

$$\frac{1 - \mathrm{PFA}^\rho}{\rho} = \frac{\sum_{k=1}^{\infty} \rho(1-\rho)^{k-1}[1 - \mathbb{P}_\infty(\nu^c < k)]}{\rho} = \sum_{k=1}^{\infty} (1-\rho)^{k-1} \mathbb{P}_\infty(\nu^c \geq k) \to \mathbb{E}_\infty \hat{\tau}^c,$$

where we used the fact that $\mathbb{P}_k(\nu^c < k) = \mathbb{P}_\infty(\nu^c < k)$, since $\{\nu^c < k\} \in \mathscr{F}_{k-1}$.

Therefore, the PFA constraint in the Shiryaev rule is replaced with the ARL constraint in the SR rule, which shows that the SR detection test is the limit of Shiryaev's detection test, where instead of using the PFA constraint one has to use a constraint on $\mathrm{ARL} = \mathbb{E}_\infty \tau$.

Finally, the Bayesian average detection delay $\mathrm{ADD}^\rho(\tau) = \mathbb{E}^\pi(\tau - \lambda \mid \tau \geq \lambda)$ in the Shiryaev procedure is replaced with the sum $\sum_{k=1}^{\infty} \mathbb{E}_k(\tau - k)^+$ in the SR procedure. To see this, it suffices to note that, as $\rho \to 0$, $\mathrm{ADD}^\rho(\nu^c)$ converges to

$$\mathrm{ADD}^0(\hat{\tau}^c) = \frac{\sum_{k=1}^{\infty} \mathbb{E}_k(\hat{\tau}^c - k \mid \hat{\tau}^c \geq k) \mathbb{P}_\infty(\hat{\tau}^c \geq k)}{\sum_{k=1}^{\infty} \mathbb{P}_\infty(\hat{\tau}^c \geq k)} = \frac{1}{\mathbb{E}_\infty \hat{\tau}^c} \sum_{k=1}^{\infty} \mathbb{E}_k(\hat{\tau}^c - k)^+. \tag{2.25}$$

Because the Shiryaev rule is optimal, this allows us to conjecture that the SR rule is (exactly) optimal with respect to $\mathrm{IADD}(\tau) = \sum_{k=1}^{\infty} \mathbb{E}_k(\tau - k)^+$. A rigorous proof with all details can be found in Pollak and Tartakovsky (2008b).

In a variety of surveillance applications such as intrusion detection in computer networks, target detection in radar and IR systems, etc., it is of utmost importance to detect a real change as quickly as possible after its occurrence, even at the price of raising many false alarms (using a repeated application of the same stopping rule) before the change occurs. This essentially means that the change-point λ is very large compared to the constant γ, which, in this case, defines the mean time between consecutive false alarms. It also means that the cost of making false detections is relatively small to the cost of the detection delay, which of course may be realistic only if an independent mechanism (or algorithm) is available for filtering/rejection false detections.

To be more specific, let $\hat{\tau}_1^c(B_\gamma), \hat{\tau}_2^c(B_\gamma), \ldots$ be sequential independent repetitions of the SR stopping time $\hat{\tau}_1^c(B_\gamma)$ defined in (2.22). Therefore, the SR statistic is renewed from scratch after each alarm. Note that $\mathbb{E}_\infty \hat{\tau}_i^c(B_\gamma) = \gamma$ for $i \geq 1$. Let (for $j \geq 1$) $Q_j = \hat{\tau}_1^c(B_\gamma) + \hat{\tau}_2^c(B_\gamma) + \cdots + \hat{\tau}_j^c(B_\gamma)$ be the time of the jth alarm, and let $J_\lambda = \min\{j \geq 1 : Q_j \geq \lambda\}$, i.e., Q_{J_λ} is the time of detection of a true change that occurs at λ after $J_\lambda - 1$ false alarms have been raised. Figure 2 illustrates this scenario.

The previous optimality result given in (2.24) is useful in showing that the repeated SR procedure defined by Q_{J_λ} is asymptotically (as $\lambda \to \infty$) optimal with respect to the expected delay $\mathbb{E}_\lambda(Q_{J_\lambda} - \lambda)$ in the class of detection procedures $\Delta_{\mathrm{m}}(\gamma)$. See Theorem 2 in Pollak and Tartakovsky (2008b). Note that this result is not asymptotic with respect to the ARL: it holds for every $\gamma > 1$. This result has been first proven by Shiryaev (1961, 1963) in continuous time for the problem of detecting a change in the drift of Brownian motion.

In addition, it follows from the discussion in Section 2.4 that the centralized SR test (along with the CUSUM test) is asymptotically minimax optimal in the class $\Delta_{\mathrm{m}}(\gamma)$ as $\gamma \to \infty$ (i.e., subject to the ARL constraint) and asymptotically uniformly optimal in the class $\Delta_{\mathrm{u}}^T(\beta)$ as $\beta \to 0$ (i.e., subject to the local supremum false alarm probability constraint).

However, both SR and CUSUM change-point detection tests lose their optimality property under the Bayesian criterion (in the class $\Delta_{\mathrm{B}}(\alpha)$) as long as the parameter ρ of the geometric prior distribution is not small. Indeed, as it was established by Tartakovsky and Veeravalli (2005), for small PFA α, the

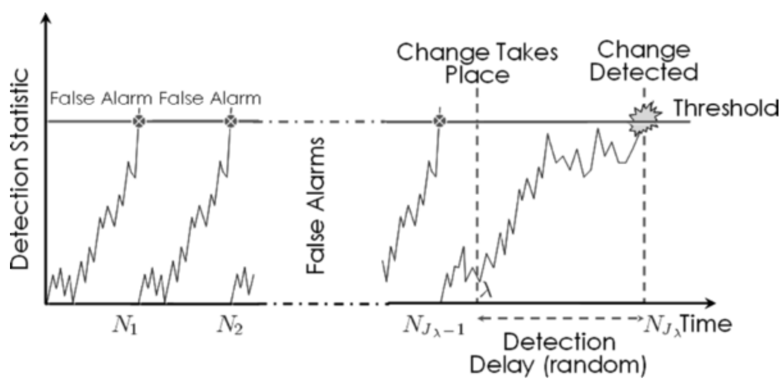

Figure 2. Detection scenario.

average detection delay of the optimal Shiryaev procedure allows for the following asymptotic approximation:

$$\mathrm{ADD}^{\rho}(\nu^c(B_{\rho,\alpha})) \sim \frac{|\log\alpha|}{\mathcal{I}^c_{\mathrm{tot}} + |\log(1-\rho)|} \quad \text{as } \alpha \to 0. \tag{2.26}$$

On the other hand, it follows from Tartakovsky and Veeravalli (2005) that for the SR and CUSUM tests with thresholds $B_\alpha = 1/(\alpha\rho)$ and $h_\alpha = \log(1/\alpha\rho)$, respectively (which guarantee $\mathrm{PFA}^{\rho}(\hat{\tau}^c(B_\alpha)) \le \alpha$, $\mathrm{PFA}^{\rho}(\tau^c(h_\alpha)) \le \alpha$), the Bayesian average detection delays are

$$\mathrm{ADD}^{\rho}(\hat{\tau}^c(B_\alpha)) \sim \mathrm{ADD}^{\rho}(\tau^c(h_\alpha)) \sim \frac{|\log\alpha|}{\mathcal{I}^c_{\mathrm{tot}}} \quad \text{as } \alpha \to 0. \tag{2.27}$$

Comparing (2.26) and (2.27) shows that neither the SR procedure nor the CUSUM procedure are asymptotically optimal unless $\mathcal{I}^c_{\mathrm{tot}} \gg |\log(1-\rho)|$. However, the latter condition often holds, especially if the number of sensors is large enough, so that the data have much more information (which is expressed in the value of the total K–L information number $\mathcal{I}^c_{\mathrm{tot}}$) compared to the prior knowledge (which is expressed via the term $|\log(1-\rho)|$ in (2.26)).

The above analysis allows us to conclude that both the centralized CUSUM and SR detection tests represent reasonable benchmark procedures in all three considered classes of detection tests and all three performance metrics: minimax—SADD in the class $\Delta_{\mathrm{m}}(\gamma)$; uniform—$\mathrm{CADD}_k$ in the class $\Delta^T_{\mathrm{u}}(\beta)$; and Bayesian—$\mathrm{ADD}^{\rho}$ in the class $\Delta_{\mathrm{B}}(\alpha)$.

Remark 2.1. In the present paper, we do not consider Lorden's essential supremum measure $\mathrm{ESADD}(\tau) = \sup_\lambda \operatorname{ess\,sup} \mathbb{E}_\lambda(\tau - \lambda \,|\, \tau \ge \lambda, \mathbf{X}^{\lambda-1})$, which is often considered to be overly pessimistic, because, obviously, $\mathrm{ESADD}(\tau) \ge \mathrm{SADD}(\tau) \ge \mathrm{ADD}^{\rho}(\tau)$. However, as has been proven by Moustakides (1986), the CUSUM detection procedure is not only asymptotically but exactly optimal with respect to $\mathrm{ESADD}(\tau)$ in the class $\Delta_{\mathrm{m}}(\gamma)$ (i.e., for all $\gamma > 1$) whenever threshold $h = h_\gamma$ is selected in such a way that $\mathrm{ARL}(\tau^c(h_\gamma)) = \gamma$. Such a powerful result is not available in the minimax setting with respect to Pollak's supremum measure $\mathrm{SADD}(\tau)$. The detection delay measures $\mathrm{SADD}(\tau)$ and $\mathrm{ESADD}(\tau)$ differ quite fundamentally. As it is apparent from the previous discussion, Pollak's measure $\mathrm{SADD}(\tau)$ is closer to Shiryaev's measure $\mathrm{ADD}^{\rho}(\tau)$ than to $\mathrm{ESADD}(\tau)$. Indeed, because $\mathrm{ADD}^{\rho}(\tau)$ converges to $\mathrm{ADD}^{0}(\tau)$ defined in (2.25) and because the minimax procedure that minimizes $\mathrm{SADD}(\tau)$ should be an equalizer (i.e., $\mathbb{E}_k(\tau - k \,|\, \tau \ge k) = \mathbb{E}_1(\tau - 1)$ for all $k \ge 1$), we obtain from (2.25) that for the minimax test $\mathrm{ADD}^{\rho}(\tau) = \mathrm{ADD}^{0}(\tau) = \mathbb{E}_1(\tau - 1)$ for every $\rho \ge 0$. It is therefore reasonable to assume that the SR test with a curved threshold that increases initially and becomes constant when time goes on should be exactly minimax in the class $\Delta_{\mathrm{m}}(\gamma)$ for all $\gamma > 1$. If, however, the point of change λ is modeled as a stopping time itself, then Lorden's measure may be the most appropriate detection delay metric. A detailed discussion of this issue can be found in the paper by Moustakides (2008).

3. DECENTRALIZED DETECTION WITH QUANTIZATION AT SENSORS

Consider the scenario where based on the information available at sensor S_i at time n a message $U_i(n)$ belonging to a finite alphabet of size M_i (e.g., binary) is formed and sent to the fusion center (see Figure 1). Write $\mathbf{U}(n) = (U_1(n), \dots, U_N(n))$ for the vector of N messages at time n. Based on the sequence of sensor messages, a decision about the change is made at the fusion center. This test is identified with a stopping time on $\{\mathbf{U}(n)\}_{n\geq 1}$ at which it is declared that a change has occurred. The goal is to find tests based on $\{\mathbf{U}(n)\}_{n\geq 1}$ that optimize the trade-off between detection delay and false alarm rate. This problem was introduced in Veeravalli (2001), and subsequently studied in asymptotic settings in other papers, including Tartakovsky and Veeravalli (2002, 2003, 2004), Tartakovsky and Kim (2006), Mei (2005), and Moustakides (2006).

Various information structures are possible for the decentralized configuration depending on how feedback and local information is used at the sensors (a detailed concert of possibilities is described in Veeravalli, 2001). Here we consider the simplest information structure where the message $U_i(n)$ formed by sensor S_i at time n is a function of only its current observation $X_i(n)$, i.e., $U_i(n) = \psi_{i,n}(X_i(n))$. Moreover, because for a particular sensor S_i, the sequence $\{X_i(n)\}_{n\geq 1}$ is assumed to be i.i.d., it is natural to confine ourselves by *stationary* quantizers for which the quantizing functions $\psi_{i,n}$ do not depend on n, i.e., $\psi_{i,n} = \psi_i$ for all $n \geq 1$. The quantizing functions $\boldsymbol{\psi} = \{\psi_i; i = 1, \dots, N\}$, together with the fusion center stopping time τ, form a policy $\phi = (\boldsymbol{\psi}, \tau)$.

Let H_λ be the hypothesis that the change occurs at time $\lambda \in \{1, 2, \dots\}$, and let H_∞ be the hypothesis that the change does not occur at all. Also, let $g_i^{(j)}$ denote the probability mass function (p.m.f.) induced on U_i when the observation $X_i(n)$ is distributed as $f_i^{(j)}$, $j = 0, 1$. Then, for *fixed* sensor quantizers, the LLR between the hypotheses H_k and H_∞ at the fusion center is given by

$$Z^q(k, n) = \sum_{j=k}^{n} \sum_{i=1}^{N} \log \frac{g_i^{(1)}(U_i(j))}{g_i^{(0)}(U_i(j))}. \tag{3.1}$$

Hereafter the superscript index q stands for quantized versions of the corresponding variables to distinguish from the centralized case where we used the superscript c. For fixed sensor quantizers, the fusion center faces a standard change detection problem based on the vector observation sequence $\{\mathbf{U}(n)\}_{n\geq 1}$.

3.1. Minimax and Uniformly Optimal Detection Tests Based on Quantized Data

Here the goal is to choose the policy ϕ that minimizes $\mathrm{SADD}(\phi)$ and/or $\mathrm{CADD}_\lambda(\phi)$ (for all $\lambda \geq 1$) defined by

$$\mathrm{SADD}(\phi) = \sup_{1\leq\lambda<\infty} \mathbb{E}_\lambda(\tau - \lambda \mid \tau \geq \lambda) \quad \text{and} \quad \mathrm{CADD}_\lambda(\phi) = \mathbb{E}_\lambda(\tau - \lambda \mid \tau \geq \lambda) \tag{3.2}$$

while maintaining the ARL to false alarm (false alarm rate) $\mathrm{ARL}(\phi) = \mathbb{E}_\infty \tau$ at a level not less than $\gamma > 1$ and/or the local probability of false alarm $\mathrm{PFA}_T(\phi) = \sup_{k\geq 1} \mathbb{P}_\infty(\tau < k + T \mid \tau \geq k)$ at the level not more than $0 < \beta < 1$.

We can define the CUSUM and SR statistics by $W^q(n) = \max_{0\le\lambda\le n} Z^q(\lambda, n)$ and $R^q(n) = \sum_{\lambda=1}^{n} e^{Z^q(\lambda,n)}$, respectively, which obey the recursions:

$$
\begin{aligned}
&W^q(n) = \max\{0, W^q(n-1) + Z^q(n,n)\}, \quad W^q(0) = 0;\\
&R^q(n) = [1 + R^q(n-1)]\exp\{Z^q(n,n)\}, \quad R^q(0) = 0.
\end{aligned}
\tag{3.3}
$$

Then the CUSUM and SR detection procedures at the fusion center $\tau^q(h)$ and $\hat{\tau}^q(a)$ are, respectively, given by

$$
\tau^q(h) = \min\{n \ge 1 : W^q(n) \ge h\}, \quad \hat{\tau}^q(a) = \min\{n \ge 1 : \log[R^q(n)] \ge a\} \tag{3.4}
$$

where h and a are positive thresholds, which are selected so that $\mathrm{ARL}(\tau^q) \ge \gamma$ and $\mathrm{ARL}(\hat{\tau}^q) \ge \gamma$ in the case of optimizing SADD in the class $\Delta_{\mathrm{m}}(\gamma)$ and $\mathrm{PFA}_T(\tau^q) \le \beta$ and $\mathrm{PFA}_T(\hat{\tau}^q) \le \beta$ in the case of optimizing CADD_λ for all λ in the class $\Delta_{\mathrm{u}}^T(\beta)$.

Let $\mathcal{I}_i^q = \mathbb{E}_1[g_i^{(1)}(U_i(1))/g_i^{(0)}(U_i(1))]$ denote the K–L information number for quantized data in the ith sensor (i.e., divergence between $g_i^{(1)}$ and $g_i^{(0)}$), and let $\mathcal{I}_{\mathrm{tot}}^q = \sum_{i=1}^{N} \mathcal{I}_i^q$ be the total K–L information accumulated from all sensors.

For the sake of brevity we first restrict our attention to the minimax problem in the class $\Delta_{\mathrm{m}}(\gamma)$. Similar to (2.8) we obtain that detection procedures $\tau^q(h)$ and $\hat{\tau}^q(h)$ given in (3.4), with $h = a = \log\gamma$, are asymptotically minimax optimal as $\gamma \to \infty$ among all procedures with ARL to false alarm greater than γ (for fixed quantizers ψ_i). To be specific,

$$
\inf_{\tau\in\Delta_{\mathrm{m}}(\gamma)} \mathrm{SADD}(\tau) \sim \mathrm{SADD}(\tau^q) \sim \mathrm{SADD}(\hat{\tau}^q) \sim \frac{\log\gamma}{\mathcal{I}_{\mathrm{tot}}^q} \quad \text{as } \gamma \to \infty.
$$

This result immediately reveals how to choose the sensor quantizers.

Lemma 3.1. *It is asymptotically optimum (as $\gamma \to \infty$) for sensor S_i at time n to select ψ_i to maximize $\mathcal{I}_i^q$, the K–L information number.*

By Tsitsiklis (1993), an optimal ψ_i that maximizes $\mathcal{I}_i^q$ is a monotone likelihood ratio quantizer (MLRQ), i.e., there exist thresholds $a_1, a_2, \ldots, a_{M_i-1}$ satisfying $-\infty < a_1 \le a_2 \le \cdots \le a_{M_i-1}$ such that

$$
\psi_{i,\mathrm{opt}}(X_i) = b_i \quad \text{only if } a_{b_i} < Z_i(X_i) \le a_{b_i+1}, \tag{3.5}
$$

where $Z_i(X_i) = \log[f_i^{(1)}(X_i)/f_i^{(0)}(X_i)]$ is the LLR at the observation X_i (at sensor S_i). Note that function $\psi_{i,\mathrm{opt}}$ is independent of n.

Thus, the asymptotically optimal policy ϕ_{opt} for a decentralized change detection problem consists of a stationary (in time) set of MLRQs at the sensors followed by CUSUM or SR procedures based on $\{\mathbf{U}(n)\}$ at the fusion center (as described in (3.4)).

For each i, we denote the corresponding pmfs induced on $U_i(n)$ by $g_{i,\mathrm{opt}}^{(1)}$ and $g_{i,\mathrm{opt}}^{(0)}$ (i.e., at the output of the stationary MLRQ $\psi_{i,\mathrm{opt}}$ that maximizes $\mathcal{I}_i^q$). Then the effective total K–L information number between the change and no-change hypotheses at the fusion center is given by

$$
\mathcal{I}_{\mathrm{tot,opt}}^q = \sum_{i=1}^{N} \mathcal{I}(g_{i,\mathrm{opt}}^{(1)}, g_{i,\mathrm{opt}}^{(0)}). \tag{3.6}
$$

Further, we denote by τ^q_{opt} and $\hat{\tau}^q_{\text{opt}}$, respectively, the CUSUM and SR stopping rules at the fusion center for the case where the sensor quantizers are chosen to be $\boldsymbol{\psi}_{\text{opt}} = \{\psi_{i,\text{opt}}\}$. Finally, we denote by $\phi_{\text{opt}} = (\boldsymbol{\psi}_{\text{opt}}, \tau^q_{\text{opt}})$ and $\hat{\phi}_{\text{opt}} = (\boldsymbol{\psi}_{\text{opt}}, \hat{\tau}^q_{\text{opt}})$ the corresponding CUSUM and SR policies, respectively, with optimal quantization.

We also need the following additional notation:

$$S^q(n) = \sum_{k=1}^{n}\sum_{i=1}^{N} \log \frac{g_i^{(1)}(U_i(k))}{g_i^{(0)}(U_i(k))}, \quad S^q(0) = 0; \quad \eta_h = \min\{n \geq 1 : S^q(n) \geq h\};$$
$$v_q = \lim_{h\to\infty} \mathbb{E}_1 \exp\{-(S^q(\eta_h) - h)\}, \tag{3.7}$$

where the constant v_q can be computed using renewal-theoretic arguments (see, e.g., Siegmund, 1985; Woodroofe, 1982).

The asymptotic performance of the asymptotically optimal solutions to the decentralized change detection problem described above is given in the following theorem.

Theorem 3.1. *Suppose $\mathcal{I}_{\text{tot,opt}}$ is positive and finite.*

(i) *Then $h = a = \log\gamma$ implies that $\text{ARL}(\tau^q_{\text{opt}}) \geq \text{ARL}(\hat{\tau}^q_{\text{opt}}) \geq \gamma$, and moreover, the limits $\lim_{\gamma\to\infty}[\text{ARL}(\tau^q_{\text{opt}})/\gamma]$ and $\lim_{\gamma\to\infty}[\text{ARL}(\hat{\tau}^q_{\text{opt}})/\gamma]$ are bounded.*

(ii) *If, in addition, $Z^q(1,1)$ is nonarithmetic, then*

$$\text{ARL}(\tau^q_{\text{opt}}(h)) \sim \frac{e^h}{v_q^2 \mathcal{I}^q_{\text{tot,opt}}}, \quad \text{ARL}(\hat{\tau}^q_{\text{opt}}(a)) \sim e^a/v_q \quad \text{as } h, a \to \infty; \tag{3.8}$$

$$\text{PFA}_T(\tau^q_{\text{opt}}(h)) \sim Te^{-h} v_q^2 \mathcal{I}^q_{\text{tot,opt}}, \quad \text{PFA}_T(\hat{\tau}^q_{\text{opt}}(a)) \sim Te^{-a} v_q \quad \text{as } h, a \to \infty. \tag{3.9}$$

(iii) *If $a = h = \log\gamma$, then*

$$\inf_{\phi\in\Delta_{\text{m}}(\gamma)} \text{SADD}(\phi) \sim \text{SADD}(\phi_{\text{opt}}) \sim \text{SADD}(\hat{\phi}_{\text{opt}}) \sim \frac{\log\gamma}{\mathcal{I}^q_{\text{tot,opt}}} \quad \text{as } \gamma \to \infty. \tag{3.10}$$

If $h = h_\gamma = \log[v_q^2 \mathcal{I}^q_{\text{tot,opt}}\gamma]$ and $a = a_\gamma = \log(v_q\gamma)$, then $\text{ARL}(\tau^q_{\text{opt}}(h_\gamma)) \sim \gamma$ and $\text{ARL}(\hat{\tau}^q_{\text{opt}}(a_\gamma)) \sim \gamma$ as $\gamma \to \infty$ and asymptotic relations (3.10) *hold.*

(iv) *If $h = h_\beta = \log[Tv_q^2 \mathcal{I}^q_{\text{tot,opt}}/\beta]$ and $a = a_\beta = \log(Tv_q/\beta)$, then $\text{PFA}_T(\tau^q_{\text{opt}}(h_\beta)) \sim \beta$ and $\text{PFA}_T(\hat{\tau}^q_{\text{opt}}(a_\beta)) \sim \beta$ as $\beta \to 0$ and, for all $\lambda \geq 1$,*

$$\inf_{\phi\in\Delta^T_{\text{u}}(\beta)} \text{CADD}_\lambda(\phi) \sim \text{CADD}_\lambda(\phi_{\text{opt}}) \sim \text{CADD}_\lambda(\hat{\phi}_{\text{opt}}) \sim \frac{|\log\beta|}{\mathcal{I}^q_{\text{tot,opt}}} \quad \text{as } \beta \to 0. \tag{3.11}$$

Proof. Although the proof is almost apparent from our previous discussion, we present a proof summary and certain important details.

(i) Obviously, $\tau^q_{\text{opt}}(h) \geq \hat{\tau}^q_{\text{opt}}(h)$ for any $h > 0$. It is easily verified that $R^q(n) - n$ is a zero-mean martingale with respect to $\mathbb{P}_\infty$. Because $R^q(\hat{\tau}^q_{\text{opt}}(h)) \geq e^h$, applying optional sampling theorem yields $\mathbb{E}_\infty \hat{\tau}^q_{\text{opt}}(h) = R^q(\hat{\tau}^q_{\text{opt}}(h)) \geq e^h$. Thus, $\text{ARL}(\tau^q_{\text{opt}}(h)) \geq \text{ARL}(\hat{\tau}^q_{\text{opt}}(h)) \geq e^h = \gamma$ for $a = h = \log\gamma$. Further technical details regarding application of the optional stopping theorem may be found in Pollak (1987) and Tartakovsky and Veeravalli (2004).

(ii) It follows from Pollak and Tartakovsky (2008a) (see also Tartakovsky, 2005, for the CUSUM procedure) that the limiting distributions of $\tau^q_{\rm opt}(h)e^{-h}v_q^2\mathcal{I}^q_{\rm tot,opt}$ and $\hat{\tau}^q_{\rm opt}(a)e^{-a}v_q$ (as $h, a \to \infty$) are Exponential(1) and that the moment generating functions converge to that of Exponential(1), which implies that

$$\lim_{h\to\infty} \mathbb{E}_\infty\big(\tau^q_{\rm opt}(h)e^{-h}v_q^2\mathcal{I}^q_{\rm tot,opt}\big) = 1, \quad \lim_{a\to\infty} \mathbb{E}_\infty\big(\hat{\tau}^q_{\rm opt}(a)e^{-a}v_q\big) = 1,$$

i.e., (3.8) follows. The asymptotic approximations for the local PFA (3.9) follow from the above weak convergence in an obvious manner.

(iii) It is known that for fixed quantizers,

$$\mathbb{E}_1\tau^q(h) = \frac{h}{\mathcal{I}^q_{\rm tot}}(1+o(1)), \quad \mathbb{E}_1\hat{\tau}^q(a) = \frac{a}{\mathcal{I}^q_{\rm tot}}(1+o(1)) \quad \text{as } h, a \to \infty$$

whenever the K–L information number $\mathcal{I}^q_{\rm tot}$ is finite. In fact, these first-order asymptotic approximations may be deduced from Lorden (1971) and Pollak (1987) and follow directly from Tartakovsky (1998b). Because $\mathrm{SADD}(\tau^q(h)) = \mathbb{E}_1\tau^q(h) - 1$ and $\mathrm{SADD}(\hat{\tau}^q(a)) = \mathbb{E}_1\hat{\tau}^q(a) - 1$, taking $a = h = \log\gamma$ or $h = h_\gamma = \log(v_q^2\mathcal{I}^q_{\rm tot}\gamma)$ and $a = a_\gamma = \log(v_q\gamma)$, we obtain

$$\mathrm{SADD}(\tau^q) \sim \mathrm{SADD}(\hat{\tau}^q) \sim \frac{\log\gamma}{\mathcal{I}^q_{\rm tot}} \quad \text{as } \gamma \to \infty. \tag{3.12}$$

Next, by Lai (1998) and Tartakovsky (1998b),

$$\inf_{\tau\in\Delta_{\rm m}(\gamma)} \mathrm{SADD}(\tau) \ge \frac{\log\gamma}{\mathcal{I}^q_{\rm tot}}(1+o(1)) \quad \text{as } \gamma \to \infty,$$

which implies that any asymptotically optimal change detection policy $\phi = (\boldsymbol{\psi}, \tau)$ should maximize the total K–L information $\mathcal{I}^q_{\rm tot} = \sum_{i=1}^N \mathcal{I}^q_i$, i.e., each K–L information number $\mathcal{I}^q_i$. As discussed above, such a maximizer exists: it is the MLRQ given by (3.5). If it is being used, then the latter lower bound becomes

$$\inf_{\boldsymbol{\psi},\tau\in\Delta_{\rm m}(\gamma)} \mathrm{SADD}(\phi) \ge \frac{\log\gamma}{\mathcal{I}^q_{\rm tot,opt}}(1+o(1)) \quad \text{as } \gamma \to \infty,$$

and, by (3.12),

$$\mathrm{SADD}(\phi_{\rm opt}) \sim \mathrm{SADD}(\hat{\phi}_{\rm opt}) \sim \frac{\log\gamma}{\mathcal{I}^q_{\rm tot,opt}} \quad \text{as } \gamma \to \infty,$$

which along with the previous inequality yields (3.10).

(iv) It follows from Tartakovsky (2005) that the CUSUM test $\tau^q(h)$ with threshold $h_\beta = \log(Tv_q^2\mathcal{I}^q_{\rm tot}/\beta)$ is asymptotically uniformly optimal in the class $\Delta_{\rm u}^T(\beta)$ as $\beta \to 0$, and in just the same way it can be shown that the same is true for the SR test $\hat{\tau}^q(a)$ with threshold $a = a_\beta = \log(Tv_q/\beta)$, i.e., for all $\lambda \ge 1$,

$$\inf_{\tau\in\Delta_{\rm u}^T(\beta)} \mathrm{CADD}_\lambda(\phi) \sim \mathrm{CADD}_\lambda(\tau^q(h_\beta)) \sim \mathrm{CADD}_\lambda(\hat{\tau}^q(a_\beta)) \sim \frac{|\log\beta|}{\mathcal{I}^q_{\rm tot}} \quad \text{as } \beta \to 0. \tag{3.13}$$

Again, the MLRQ given by (3.5) maximizes the total K–L information number $\mathcal{I}^q_{\mathrm{tot}}$, which shows in just the same way as in the proof of (iii) that the MLRQ followed by the CUSUM and SR procedures are asymptotically uniformly optimal policies, i.e., (3.11) follows from (3.13) and the argument given above. □

Remark 3.1. It can be shown (cf. Tartakovsky, 2005) that if the second moments $\mathbb{E}_1|Z^q(1,1)|^2$ and $\mathbb{E}_\infty|Z^q(1,1)|^2$ are finite and $Z^q(1,1)$ is nonarithmetic, then, as $h \to \infty$, the following higher order approximations hold

$$\begin{aligned} \mathrm{CADD}_1(\tau^q_{\mathrm{opt}}(h)) &= \frac{1}{\mathcal{I}^q_{\mathrm{tot,opt}}}(h + \varkappa_q + \delta_q) - 1 + o(1), \\ \mathrm{CADD}_\lambda(\tau^q_{\mathrm{opt}}(h)) &= \frac{1}{\mathcal{I}^q_{\mathrm{tot,opt}}}(h + \varkappa_q - \mu_q) - 1 + o(1), \quad \lambda \to \infty, \end{aligned} \tag{3.14}$$

where the values of $\delta_q = \mathbb{E}_1[\min_{n\geq 0} S^q(n)]$, $\mu_q = \mathbb{E}_\infty[\max_{n\geq 0} S^q(n)]$, and $\varkappa_q = \lim_{h\to\infty} \mathbb{E}_1[S^q(\eta_h) - h]$ can be computed numerically. Similar approximations hold for the SR detection procedure with different constants δ_q and μ_q that are difficult to compute.

Remark 3.2. The condition that the LLR $Z^q(1,1)$ is nonarithmetic is imposed due to the necessity of considering certain discrete cases separately in renewal theorems. Because the data at the output of quantizers are discrete, it may happen that the LLR does not obey this condition. If $Z^q(1,1)$ is arithmetic with span $d > 0$, the results of Theorem 3.1(ii) and approximations (3.14) hold true as $h \to \infty$ through multiples of d (i.e., $h = kd$, $k \to \infty$). We stress, however, that even in discrete cases where the data are arithmetic, the LLR is usually nonarithmetic. A typical example is a binary (Bernoulli) case where the sum $\sum_{k=1}^n X_k$ is arithmetic but the LLR, which is a weighted sum, is almost always nonarithmetic.

We now continue by considering the simplest case where $U_i(n) = \psi_i(X_i(n))$ are the outputs of binary quantizers and specify previous results for this case. Also, in the rest of this section we will consider only the CUSUM detection procedure with understanding that analogous results hold for the SR procedure. It follows from Theorem 3.1 that the optimal binary quantizer is the MLRQ that is given by

$$U_i = \psi_i(X) = \begin{cases} 1 & \text{if } Z_i(X) = \log[f_i^{(1)}(X)/f_i^{(0)}(X)] \geq t_i, \\ 0 & \text{otherwise,} \end{cases}$$

where t_i is a threshold that maximizes the K–L information in the resulting Bernoulli sequence.

To be precise, for $j = 0, 1$, let $g_i^{(j)}$ denote the probability induced on $U_i(n)$ when the observation $X_i(n)$ is distributed as $f_i^{(j)}$. Let $g_{0,i} = g_i^{(0)}(U_i(n) = 1)$ and $g_i = g_i^{(1)}(U_i(n) = 1)$ denote the corresponding probabilities under the normal and the anomalous conditions, respectively. The resulting binary (Bernoulli) sequences $\{U_i(n), i = 1, \ldots, N\}$, $n \geq 1$ are then used to form the binary CUSUM statistic similar to (3.3) as

$$W^q(n) = \max\left\{0, W^q(n-1) + \sum_{i=1}^N Z_i^q(n)\right\}, \quad W^q(0) = 0 \tag{3.15}$$

where

$$Z_i^q(n) = \log \frac{g_i^{(1)}(U_i(n))}{g_i^{(0)}(U_i(n))}$$

is the partial LLR between the change and no-change hypotheses for the binary sequence, which is given by

$$Z_i^q(n) = c_i U_i(n) + c_{0,i}.$$

Here

$$c_i = \log \frac{g_i(1-g_{0,i})}{g_{0,i}(1-g_i)}, \quad c_{0,i} = \log \frac{1-g_i}{1-g_{0,i}}.$$

Then the CUSUM detection procedure at the fusion center is given by the stopping time $\tau^q(h)$ defined in (3.4). In what follows this detection procedure will be referred to as the binary quantized CUSUM test and the abbreviation BQ-CUSUM will be used throughout the paper.

It follows from Theorem 3.1 that the BQ-CUSUM procedure with $h = \log\gamma$ is asymptotically optimal as $\gamma \to \infty$ in the class of tests with binary quantization in the sense of minimizing the SADD in the class $\Delta_{\mathrm{m}}(\gamma)$. More specifically, $\mathrm{SADD}(\tau^q) = \mathbb{E}_1(\tau^q - 1)$ and the trade-off curve that relates SADD and ARL for the large ARL is

$$\mathrm{SADD}(\tau^q) \sim \frac{\log(\mathrm{ARL})}{\sum_{i=1}^N (g_i c_i + c_{0,i})}. \tag{3.16}$$

Note that probabilities $g_i = g_i(t_i)$ and $g_{0,i} = g_{0,i}(t_i)$ depend on the value of threshold t_i. To optimize the performance, one should choose thresholds $t_1, \ldots, t_N$ so that the denominator in (3.16) is maximized, i.e.,

$$t_i^0 = \arg \max_{-\infty < t_i < +\infty} \mathcal{I}_i^q(t_i), \quad i = 1, \ldots, N, \tag{3.17}$$

where $\mathcal{I}_i^q(t_i) = g_i(t_i)c_i(t_i) + c_{0,i}(t_i)$ is the K–L information number for the binary sequence in the ith sensor. It follows from (3.16) and (3.17) that the tradeoff curve for the optimal binary test is

$$\mathrm{SADD}(\tau^q) \sim \frac{\log\gamma}{\mathcal{I}_{\mathrm{tot}}^q}, \quad \gamma \to \infty, \tag{3.18}$$

where $\mathcal{I}_{\mathrm{tot}}^q = \sum_{i=1}^N \max_{t_i}[g_i(t_i)c_i(t_i) + c_{0,i}(t_i)]$.

The asymptotic relative efficiency (ARE) of a detection procedure τ_γ with respect to a detection procedure η_γ, both of which meet the same lower bound γ for the ARL, will be defined as

$$\mathrm{ARE}(\tau_\gamma; \eta_\gamma) = \lim_{\gamma\to\infty} \frac{\mathrm{SADD}(\tau_\gamma)}{\mathrm{SADD}(\eta_\gamma)}.$$

Using (2.8) and (3.18), we obtain that the ARE of the globally asymptotically optimal test ν (e.g., for the centralized CUSUM and SR tests) with respect to the BQ-CUSUM test τ^q is

$$\mathrm{ARE}(\nu; \tau^q) = \lim_{\gamma\to\infty} \frac{\inf_{\tau\in\Delta_{\mathrm{m}}(\gamma)} \mathrm{SADD}(\tau)}{\mathrm{SADD}(\tau^q(h_\gamma))} = \frac{\mathcal{I}_{\mathrm{tot}}^q}{\mathcal{I}_{\mathrm{tot}}^c}. \tag{3.19}$$

Obviously, the same result holds for CADD_λ in the class $\Delta_{\mathrm{u}}^T(\beta)$ and for the SR detection procedure. Specifically,

$$\begin{aligned}\mathrm{ARE}(v;\tau^q) = \mathrm{ARE}(v;\hat{\tau}^q) &= \lim_{\beta\to 0}\frac{\inf_{\tau\in\Delta_{\mathrm{u}}^T(\beta)}\mathrm{CADD}_\lambda(\tau)}{\mathrm{CADD}_\lambda(\tau^q(h_\beta))}\\ &= \lim_{\beta\to 0}\frac{\inf_{\tau\in\Delta_{\mathrm{u}}^T(\beta)}\mathrm{CADD}_\lambda(\tau)}{\mathrm{CADD}_\lambda(\hat{\tau}^q(a_\beta))} = \frac{\mathcal{I}_{\mathrm{tot}}^q}{\mathcal{I}_{\mathrm{tot}}^c} \quad \text{for all } \lambda \geq 1.\end{aligned} \tag{3.20}$$

Because the centralized K–L information number $\mathcal{I}_{\mathrm{tot}}^c$ is always larger than $\mathcal{I}_{\mathrm{tot}}^q$, it follows from (3.19) and (3.20) that the value of ARE < 1. However, our study presented below shows that certain decentralized asymptotically globally optimal tests may perform worse in practically interesting prelimit situations when the false alarm rate is moderately low but not very low.

Remark 3.3. It can be shown that for the three particular models, namely Gaussian $\mathcal{N}(0,1)\to\mathcal{N}(\theta,1)$, Poisson $\mathcal{P}(1)\to\mathcal{P}(\theta)$, and Exponential $\mathrm{Exp}(1)\to\mathrm{Exp}(\theta)$, the ARE is a monotone function of θ in the interval $[2/\pi, 1]$, and $\lim_{\theta\to 0}\mathrm{ARE}(\theta) = 2/\pi$ for the Gaussian model and $\lim_{\theta\to 1}\mathrm{ARE}(\theta) = 2/\pi$ for the other two models. See Tartakovsky and Polunchenko (2008). Also, $\lim_{\theta\to\infty}\mathrm{ARE}(\theta) = 1$ for all three models. Therefore, we expect that in the worst case scenario (for close hypotheses) the loss due to binary quantization is about 36%, and it is small for far hypotheses. This is confirmed by simulations.

3.2. Bayes-Optimal Detection Tests Based on Quantized Data

In the Bayesian setting, the goal is to choose the policy ϕ that minimizes $\mathrm{ADD}^\rho(\phi) = \mathbb{E}^\pi(\tau-\lambda\,|\,\tau\geq\lambda)$, while maintaining the average probability of false alarm $\mathrm{PFA}^\rho(\phi) = \mathbb{P}^\pi\{\tau<\lambda\} = \sum_{k=1}^\infty \pi_k\mathbb{P}_\infty(\tau<k)$ at a level not greater than α.

Although all the results may be generalized for an arbitrary prior distribution π_k as in Tartakovsky and Veeravalli (2005), for the sake of simplicity, in this paper we are considering only the geometric prior distribution given in (2.16).

Define the Shiryaev statistic based on the quantized data as

$$R_\rho^q(n) = \sum_{k=1}^n (1-\rho)^{k-1-n}\exp\{Z^q(k,n)\}, \tag{3.21}$$

or recursively,

$$R_\rho^q(n) = \frac{1}{1-\rho}[1+R_\rho^q(n-1)]\exp\{Z^q(n,n)\}, \quad R_\rho^q(0)=0. \tag{3.22}$$

(cf. (2.19)), where $Z^q(k,n)$ is the cumulative LLR after quantization from all sensors given in (3.1). Then the Shiryaev procedure at the fusion center has a stopping time

$$v^q(B_\rho) = \min\{n\geq 1 : R_\rho^q(n)\geq B_\rho\}, \tag{3.23}$$

where B_ρ is a positive threshold, which is selected so that $v^q(B_\rho)\in\Delta_{\mathrm{B}}(\alpha)$.

Following the same steps as those used in Section 3.1 above (see also Tartakovsky and Veeravalli, 2003, 2005), we can conclude that the detection procedure $\nu^q(B_\rho)$, with $B_\rho = B_{\rho,\alpha} = (1-\alpha)/(\rho\alpha)$, is asymptotically optimal as $\alpha \to 0$ among all procedures with PFA no greater than α. To be specific, let $\boldsymbol{\psi} = \{\psi_1, \ldots, \psi_N\}$ be a stationary quantizer. Then, as $\alpha \to 0$,

$$\inf_{\tau \in \Delta_{\mathrm{B}}(\alpha)} \mathrm{ADD}^\rho(\boldsymbol{\psi}, \tau) \sim \mathrm{ADD}^\rho(\boldsymbol{\psi}, \nu^q(B_{\rho,\alpha})) \sim \frac{|\log \alpha|}{|\log(1-\rho)| + \mathcal{I}^q_{\mathrm{tot}}}, \tag{3.24}$$

where $\mathrm{ADD}^\rho(\boldsymbol{\psi}, \tau) = \mathbb{E}^\pi(\tau - \lambda \,|\, \tau \geq \lambda)$ and $\mathcal{I}^q_{\mathrm{tot}} = \sum_{i=1}^N \mathcal{I}^q_i$ with $\mathcal{I}^q_i$ being the K–L information number for quantized data (divergence between $g_i^{(1)}$ and $g_i^{(0)}$) as in Section 3.1. From (3.24) it is clear that Lemma 3.1 holds in the Bayesian setting as well, and the optimal sensor quantizers are MLRQs, with MLRQ $\psi_{i,\mathrm{opt}}$ chosen to maximize $\mathcal{I}^q_i$. With this choice of MLRQs, the effective total K–L information between the change and no-change hypotheses at the fusion center is given by $\mathcal{I}^q_{\mathrm{tot,opt}}$ defined in (3.6).

The identical analysis can be performed for the SR and CUSUM fusion procedures of (3.4) when applied in the Bayesian setting with thresholds $h = h_\alpha = \log(1/\alpha\rho)$ and $a = B_\alpha = 1/(\alpha\rho)$, and with sensor MLRQs $\boldsymbol{\psi}_{\mathrm{opt}}$, to yield (see also (2.27)):

$$\mathrm{ADD}^\rho(\boldsymbol{\psi}_{\mathrm{opt}}, \hat{\tau}^q(B_\alpha)) \sim \mathrm{ADD}^\rho(\boldsymbol{\psi}_{\mathrm{opt}}, \tau^q(h_\alpha)) \sim \frac{|\log \alpha|}{\mathcal{I}^q_{\mathrm{tot,opt}}} \quad \text{as } \alpha \to 0. \tag{3.25}$$

As in the centralized setting, the SR and CUSUM procedures suffer an asymptotic loss in efficiency due to the missing term $|\log(1-\rho)|$ in denominator of (3.25). If $\mathcal{I}^q_{\mathrm{tot}} \gg |\log(1-\rho)|$, as is the case when N is large, the loss in efficiency is negligible.

Now we denote by ν^q_{opt}, τ^q_{opt}, and $\hat{\tau}^q_{\mathrm{opt}}$ the Shiryaev, CUSUM, and SR stopping rules, respectively, at the fusion center for the case where the sensor quantizers are chosen to be $\boldsymbol{\psi}_{\mathrm{opt}}$. Further, we denote by $\phi^\rho_{\mathrm{opt}} = (\boldsymbol{\psi}_{\mathrm{opt}}, \nu^q_{\mathrm{opt}})$, $\phi_{\mathrm{opt}} = (\boldsymbol{\psi}_{\mathrm{opt}}, \tau^q_{\mathrm{opt}})$, and $\hat{\phi}_{\mathrm{opt}} = (\boldsymbol{\psi}_{\mathrm{opt}}, \hat{\tau}^q_{\mathrm{opt}})$, respectively, the corresponding Shiryaev, CUSUM, and SR policies with optimal quantization. The asymptotic performance of these procedures is given in the following theorem.

Theorem 3.2. *Suppose $\mathcal{I}^q_{\mathrm{tot,opt}}$ is positive and finite.*

(i) *Then $B_{\rho,\alpha} = (1-\alpha)/(\alpha\rho)$ implies that $\mathrm{PFA}^\rho(\nu^q_{\mathrm{opt}}) \leq \alpha$ and*

$$\inf_{\phi \in \Delta_{\mathrm{B}}(\alpha)} \mathrm{ADD}^\rho(\phi) \sim \mathrm{ADD}^\rho(\phi^\rho_{\mathrm{opt}}) \sim \frac{|\log \alpha|}{\mathcal{I}^q_{\mathrm{tot,opt}} + |\log(1-\rho)|} \quad \textit{as } \alpha \to 0. \tag{3.26}$$

(ii) *Furthermore, $h_\alpha = \log(1/\alpha\rho)$ and $B_\alpha = 1/(\alpha\rho)$ imply that $\mathrm{PFA}^\rho(\tau^q_{\mathrm{opt}}) \leq \alpha$ and $\mathrm{PFA}^\rho(\hat{\tau}^q_{\mathrm{opt}}) \leq \alpha$, and*

$$\mathrm{ADD}^\rho(\phi_{\mathrm{opt}}) \sim \mathrm{ADD}^\rho(\hat{\phi}_{\mathrm{opt}}) \sim \frac{|\log \alpha|}{\mathcal{I}^q_{\mathrm{tot,opt}}} \quad \textit{as } \alpha \to 0. \tag{3.27}$$

Proof. (i) It follows from Theorem 4 in Tartakovsky and Veeravalli (2005) that, for any fixed sensor quantizer, $B_{\rho,\alpha} = (1-\alpha)/(\alpha\rho)$ implies that $\mathrm{PFA}^\rho(\nu^q) \leq \alpha$ and

$$\inf_{\tau \in \Delta_{\mathrm{B}}(\alpha)} \mathrm{ADD}^\rho(\phi) \sim \mathrm{ADD}^\rho(\nu^q) \sim \frac{|\log \alpha|}{\mathcal{I}^q_{\mathrm{tot}} + |\log(1-\rho)|} \quad \text{as } \alpha \to 0.$$

Therefore, an optimal quantizer must maximize the K–L information $\mathcal{I}^q_{\text{tot}} = \sum_{i=1}^N \mathcal{I}^q_i$, i.e., K–L numbers $\mathcal{I}^q_i$ in each sensor. Because such a quantizer exists and is the MLRQ, (3.26) follows.

(ii) We prove (3.27) only for the SR test. For the CUSUM test the proof is essentially the same.

Note first that the event $\{\hat{\tau}^q < n\}$ belongs to the sigma-algebra $\mathcal{F}_{n-1}$, and hence,

$$\text{PFA}(\hat{\tau}^q) = \sum_{n=1}^{\infty} \pi_n \mathbb{P}_n(\hat{\tau}^q < n) = \sum_{n=1}^{\infty} \pi_n \mathbb{P}_\infty(\hat{\tau}^q < n).$$

Because $R^q(n) - n$ is a zero-mean $\mathbb{P}_\infty$-martingale, the statistic $R^q(n)$ is a $\mathbb{P}_\infty$-submartingale with mean $\mathbb{E}_\infty R^q(n) = n$. Applying Doob's submartingale inequality, we obtain that for any $B > 0$

$$\mathbb{P}_\infty(\hat{\tau}^q(B) < n) = \mathbb{P}_\infty\left\{\max_{1\le k\le n} R^q(k) \ge B\right\} \le n/B, \tag{3.28}$$

which yields

$$\text{PFA}(\hat{\tau}^q(B)) = \sum_{n=1}^{\infty} \pi_n \mathbb{P}_\infty(\hat{\tau}^q(B) < n) \le B^{-1} \sum_{n=1}^{\infty} n\rho(1-\rho)^{n-1} = 1/(B\rho).$$

Thus, $B_\alpha = 1/(\alpha\rho)$ implies $\text{PFA}^\rho(\hat{\tau}^q(B_\alpha)) \le \alpha$ for any fixed quantizer, including the MLRQ.

Now, it follows form Theorem 6 in Tartakovsky and Veeravalli (2005) that for any fixed sensor quantizer and any $0 < \rho < 1$

$$\text{ADD}^\rho(\hat{\tau}^q(B_\alpha)) \sim \frac{|\log \alpha|}{\mathcal{I}^q_{\text{tot}}} \quad \text{as } \alpha \to 0$$

whenever $\mathcal{I}^q_{\text{tot}} < \infty$. Therefore, again the optimal quantizer is the MLRQ and (ii) follows. □

Thus, the AREs of the optimal centralized solutions with respect to their corresponding decentralized solutions with quantization are given by:

$$\text{ARE}(v^c(B_{\rho,\alpha}), v^q_{\text{opt}}(B_{\rho,\alpha})) = \lim_{\alpha\to 0} \frac{\text{ADD}^\rho(v^c(B_{\rho,\alpha}))}{\text{ADD}^\rho(v^q_{\text{opt}}(B_{\rho,\alpha}))} = \frac{\mathcal{I}^q_{\text{tot,opt}} + |\log(1-\rho)|}{\mathcal{I}^c_{\text{tot}} + |\log(1-\rho)|}$$

and

$$\text{ARE}(\tau^c(h_\alpha), \tau^q_{\text{opt}}(h_\alpha)) = \text{ARE}(\hat{\tau}^c(B_\alpha), \hat{\tau}^q_{\text{opt}}(B_\alpha)) = \frac{\mathcal{I}^q_{\text{tot,opt}}}{\mathcal{I}_{\text{tot}}}.$$

Quantization results in loss in efficiency due to the fact that $\mathcal{I}^q_{\text{tot,opt}} < \mathcal{I}_{\text{tot}}$, but as stated in Remark 3.3 the loss may not be large in many practical cases. In Section 4.2, we give constructions of two decentralized detection procedures, with sensors performing local change detection, for which the first-order asymptotic performance is the same as that for the centralized SR procedure. We will see in Section 5.2 that the BQ-SR test performs better than these globally first-order asymptotically optimal tests as long as the PFA is not extremely small.

4. DECENTRALIZED DETECTION BASED ON LOCAL DECISIONS AT SENSORS

4.1. Minimax Problem Setting

We now consider three detection schemes that perform local detection in the sensors and then transmit these local binary decisions to the fusion center for optimal combining and final decision-making. The abbreviation LD-CUSUM will be used for procedures that perform CUSUM tests in sensors and use local decisions. The study in this section follows the work by Tartakovsky and Kim (2006).

4.1.1. Asymptotically Optimal Decentralized LD-CUSUM Test

Let

$$W_i(n) = \max\{0, W_i(n-1) + Z_i(n)\}, \quad W_i(0) = 0$$

be the CUSUM statistic in the ith sensor, where, as before, $Z_i(n) = \log[f_1^{(i)}(X_i(n))/f_0^{(i)}(X_i(n))]$ is the LLR for the original sequence, and let

$$U_i(n) = \begin{cases} 1 & \text{if } W_i(n) \geq \omega_i h \\ 0 & \text{otherwise,} \end{cases}$$

where $\omega_i = \mathcal{I}_i/\mathcal{I}_{\text{tot}} = \mathcal{I}_i/\sum_{i=1}^N \mathcal{I}_i$ ($\mathcal{I}_i = \mathbb{E}_1 Z_i(1)$) and h is a positive threshold.

The stopping time is defined as

$$T_{\text{ld}}(h) = \min\left\{n \geq 1 : \min_{1\leq i\leq N} [W_i(n)/\omega_i] \geq h\right\}. \tag{4.1}$$

In other words, binary local decisions (1 or 0) are transmitted to the fusion center, and the change is declared at the first time when $U_i(n) = 1$ for all sensors $i = 1, \ldots, N$.

It follows from Mei (2005) that if $\mathbb{E}_1|Z_i(1)|^3 < \infty$, then

$$\mathbb{E}_\infty T_{\text{ld}}(h) \geq e^h \quad \text{for every } h > 0.$$

Under an additional Cramér-type condition (see Theorem 4.1 below), it follows from Dragalin et al. (2000) that

$$\text{SADD}(T_{\text{ld}}(h)) = \frac{h}{I_{\text{tot}}} + C_N\sqrt{\frac{h}{\mathcal{I}_{\text{tot}}}} + c + o(1) \quad \text{as } h \to \infty, \tag{4.2}$$

where c is a computable constant that depends on the model and

$$C_N = \mathbb{E} \max_{1\leq i\leq N}\left\{\frac{\sigma_i}{\mathcal{I}_i} Y_i\right\}, \tag{4.3}$$

$Y_1, \ldots, Y_N$ are independent standard Gaussian random variables; $\sigma_i = \sqrt{\text{Var}_i(Z_1(i))}$; Var_i is the operator of variance under $f_1^{(i)}$.

Therefore, if $h = \log \gamma$, then

$$\inf_{\tau \in \Delta_{\mathrm{m}}(\gamma)} \mathrm{SADD}(\tau) \sim \mathrm{SADD}(T_{\mathrm{ld}}(h)) \sim \frac{\log \gamma}{\mathcal{I}_{\mathrm{tot}}}, \quad \text{as } \gamma \to \infty, \tag{4.4}$$

and the detection test $T_{\mathrm{ld}}(h)$ is globally asymptotically optimal (AO), i.e., $\mathrm{ARE}(T_{\mathrm{ld}}; \tau_c) = 1$. This result has been first established by Mei (2005). Correspondingly, we will use the abbreviation AO-LD-CUSUM for this test in the rest of the paper.

However, because the second term in the asymptotic approximation (4.2) is on the order of the square root of the threshold, it is expected that the convergence to the optimum is slow. Furthermore, the performance degradation compared to the optimal centralized test is expected to be more and more severe with growth of the number of sensors, because the constant C_N given by (4.3) increases with N. Note that for the optimal centralized CUSUM and SR tests and for the decentralized CUSUM and SR tests with binary quantization, the residual terms are constants (see (3.14) in Remark 3.1). We therefore expect that for moderate false alarm rates typical for practical applications, the procedures with quantization may perform better. This fact is confirmed by MC simulations for Gaussian models in Mei (2005). In Section 5.1, this conjecture is verified for the Poisson model.

Remark 4.1. Results similar to (4.2) and (4.4) are not available for LD-SR detection test (where local voting is done based on the SR statistics $R_i(n)$ in place of the CUSUM statistics $W_i(n)$) in the class $\Delta_{\mathrm{m}}(\gamma)$. It turns out that the renewal property of the CUSUM statistics $W_i(n)$ plays a crucial factor under the ARL to false alarm constraint (as well as under the local PFA constraint). See Mei (2005) for a more detailed discussion. However, as will be made more apparent in Section 4.2, the LD-SR detection test can be effectively constructed in a Bayesian setting (in the class $\Delta_{\mathrm{B}}(\alpha)$).

4.1.2. Decentralized Minimal and Maximal LD-CUSUM and LD-SR Tests

Let $\tau_i(h) = \min\{n : W_i(n) \geq h\}$ denote the stopping time of the CUSUM test in the ith sensor. Introduce the stopping times

$$T_{\min}(h) = \min(\tau_1, \ldots, \tau_N) \quad \text{and} \quad T_{\max}(h) = \max(\tau_1, \ldots, \tau_N),$$

which will be referred to as minimal LD-CUSUM (Min-LD-CUSUM) and maximal LD-CUSUM (Max-LD-CUSUM) tests, respectively. Similarly, we may define Min-LD-SR and Max-LD-SR tests based on the SR stopping times in sensors $\hat{\tau}_i(h) = \min\{n : \log R_i(n) \geq h\}$. Below we focus on CUSUM-based tests keeping in mind that the results hold for SR-based tests as well.

Consider first the false alarm rate for these detection tests. Clearly, $\mathbb{E}_\infty T_{\max} \geq \mathbb{E}_\infty \tau_i$ for every $i = 1, \ldots, N$. Since $\mathbb{E}_\infty \tau_i \geq e^h$, it follows that

$$\mathrm{ARL}(T_{\max}) \geq e^h \quad \text{for every } h > 0.$$

We now show that

$$\mathrm{ARL}(T_{\min}) \geq N^{-1} e^h \quad \text{for every } h > 0.$$

Indeed,

$$T_{\min} = \min\left\{n : \max_i W_i(n) \geq h\right\} \geq \min\left\{n : \max_i R_i(n) \geq e^h\right\}$$
$$\geq \min\{n : G_N(n) \geq e^h/N\} = \eta,$$

where $G_N(n) = N^{-1}\sum_{i=1}^N R_i(n)$. Because $G_N(n) - n$ is a zero-mean $\mathbb{P}_\infty$-martingale and because $G_N(\eta) \geq e^h/N$, it follows from the optional sampling theorem that $\mathbb{E}_\infty T_{\min} \geq \mathbb{E}_\infty \eta = G_N(\eta) \geq e^h/N$.

However, these inequalities are usually very conservative. For large threshold values, asymptotically sharp approximations can be derived as follows. It follows from Pollak and Tartakovsky (2008a) and Tartakovsky (2005) that, as $h \to \infty$, under the no-change hypothesis, the stopping times τ_i, $i = 1, \ldots, N$ are exponentially distributed with mean values $e^h/(v_i^2 \mathcal{I}_i)$, where v_i are constants that are defined by (3.7), replacing $S^q(\eta_h)$ by $\sum_{k=1}^{\eta_h} Z_i(k)$. These constants can be computed numerically for any particular model using renewal arguments. Therefore, for a large threshold, $T_{\min}(h)$ is approximately exponentially distributed with mean

$$\mathrm{ARL}(T_{\min}) \sim e^h/c_N,$$

where $c_N = \sum_{i=1}^N v_i^2 \mathcal{I}_i$, whereas the mean of the stopping time $T_{\max}$ is

$$\mathrm{ARL}(T_{\max}) \sim e^h/c'_N \quad \text{as } h \to \infty,$$

where $c'_N < c_N$ can be easily computed for any N. In particular, for $N = 5$ and in the symmetric case, $c'_5 = (60/137)v^2\mathcal{I} \approx 0.44v^2\mathcal{I}$ and $c_5 = 5v^2\mathcal{I}$.

In order to derive an asymptotic approximation for $\mathrm{SADD}(T_{\min})$, note that $\mathbb{E}_1 T_{\min}(h) \leq \mathbb{E}_1 \tau_i(h)$ for all $i = 1, \ldots, N$ and, hence,

$$\mathbb{E}_1 T_{\min} \leq \frac{h}{\min_{1\leq i\leq N} \mathcal{I}_i}(1 + o(1)), \quad \text{as } h \to \infty,$$

because $\mathbb{E}_1 \tau_i(h) \sim h/\mathcal{I}_i$.

To derive an approximation for $\mathrm{SADD}(T_{\max})$, introduce the stopping time

$$\eta(h) = \min\left\{n \geq 1 : \min_{1\leq i\leq N} W_i(n) \geq h\right\}$$

and note that $\eta(h) \geq T_{\max}(h)$. Because $W_i(n) = \sum_{k=1}^n Z_i(k) - \min_{1\leq k\leq n} Z_i(k)$ and the second term is a slowly changing sequence, applying Theorem 2.3 of Tartakovsky (1998a) yields

$$\mathbb{E}_1 \eta(h) \sim \frac{h}{\min_{1\leq i\leq N} \mathcal{I}_i}, \quad \text{as } h \to \infty,$$

which implies that

$$\mathbb{E}_1 T_{\max} \leq \frac{h}{\min_{1\leq i\leq N} \mathcal{I}_i}(1 + o(1)), \quad \text{as } h \to \infty.$$

It follows that

$$\mathrm{SADD}(T_{\min}) \le \frac{h}{\max_{1\le i\le N} \mathcal{I}_i}(1+o(1)), \quad \mathrm{SADD}(T_{\max}) \le \frac{h}{\min_{1\le i\le N} \mathcal{I}_i}(1+o(1)).$$

Therefore, taking the thresholds $h = \log(\gamma c_N)$ in the Min-LD-CUSUM and $h = \log(\gamma c'_N)$ in the Max-LD-CUSUM, we obtain bounds for the trade-off curves that relate the SADD and the ARL, as $\gamma \to \infty$:

$$\mathrm{SADD}(T_{\min}) \le \frac{\log\gamma}{\max_{1\le i\le N} \mathcal{I}_i}(1+o(1)), \quad \mathrm{SADD}(T_{\max}) \le \frac{\log\gamma}{\min_{1\le i\le N} \mathcal{I}_i}(1+o(1)).$$

It follows that in the symmetric case where $\mathcal{I}_i = \mathcal{I}$, the asymptotic relative efficiency of these detection tests compared to the optimal centralized test is

$$\mathrm{ARE}(T_{\min}; \tau_c) \ge \mathrm{ARE}(T_{\max}; \tau_c) \ge N.$$

Note that although based on the first-order asymptotics it may be expected that in the symmetric case the Max-LD-CUSUM test may perform as well as the Min-LD-CUSUM test, Monte Carlo simulations in Section 5.1 show that the Min-LD-CUSUM test performs better even in the symmetric case. The same conclusion has been reached by Moustakides (2006) based on the analysis of a two-sensor continuous-time Brownian motion model.

4.2. Bayesian Problem Setting

In the Bayesian setting, for the sake of brevity, we only consider LD tests based on SR statistics at the sensors. Similar results can be obtained for LD tests based on CUSUM statistics. To this end, let $R_i(n) = \sum_{k=1}^n \prod_{\lambda=k}^n e^{Z_i(\lambda)}$ be the SR statistic in sensor S_i based on the original, nonquantized data $(X_i(1), \ldots, X_i(n))$. Clearly, $R_i(n)$ obeys the recursion

$$R_i(n) = [1 + R_i(n-1)]\exp\{Z_i(n)\}, \quad R_i(0) = 0.$$

We now introduce the local stopping times in the sensors,

$$\hat{\tau}_i(B_i) = \min\{n \ge 1 : R_i(n) \ge B_i\}, \quad i = 1, \ldots, N \tag{4.5}$$

based on which we can define three fusion rules as in the minimax setting.

For the first fusion rule, at time n, local binary decisions $U_i(n) = 0$ or 1 are transmitted to the fusion center, where $U_i(n) = 1$ if $R_i(n) \ge B_i$, and 0 otherwise. In this case, the stopping time at the fusion center is given by

$$\hat{\tau}_{\text{all}} = \text{first } n \ge 1 \quad \text{such that } U_i(n) = 1 \quad \text{for all } i = 1, \ldots, N.$$

Note that this stopping time can be rewritten as

$$\hat{\tau}_{\text{all}} = \min\Big\{n \ge 1 : \min_{1\le i\le N}(R_i(n)/B_i) \ge 1\Big\}. \tag{4.6}$$

For the second fusion rule, we stop monitoring the ith sensor once the exceedance has occurred and transmit the decision $U_i = 1$ at time $\hat{\tau}_i$ to the fusion center. At the fusion center the decision in favor of the change hypothesis is made once all the sensors "vote" for this hypothesis. This fusion procedure is equivalent to the stopping time

$$\hat{\tau}_{\max} = \max_{1 \le i \le N} \hat{\tau}_i. \tag{4.7}$$

Note that $\hat{\tau}_{\text{all}}$ is greater than $\hat{\tau}_{\max}$ (almost surely).

We can also define the procedure $\hat{\tau}_{\min}$ that corresponds to the fusion center stopping at the first time a sensor decision is in favor of the change. This fusion procedure is equivalent to the stopping time

$$\hat{\tau}_{\min} = \min_{1 \le i \le N} \hat{\tau}_i.$$

Clearly, $\hat{\tau}_{\max}$ is greater than $\hat{\tau}_{\min}$ (almost surely).

Unlike in the minimax setting, in the Bayesian setting we can show that the procedure $\hat{\tau}_{\max}$ has the same first-order asymptotic performance as $\hat{\tau}_{\text{all}}$ within the class $\Delta_{\text{B}}(\alpha)$, which can be shown to be globally asymptotically optimal (see Theorem 4.1 below). This optimality is not shared by $\hat{\tau}_{\min}$, and therefore we focus on $\hat{\tau}_{\text{all}}$ and $\hat{\tau}_{\max}$ in the following.

To bound the PFA of $\hat{\tau}_{\text{all}}$ and $\hat{\tau}_{\max}$, we use the following argument. Because, under $\mathbb{P}_\infty$, the local stopping times $\hat{\tau}_1, \ldots, \hat{\tau}_N$ are independent and, by (3.28),

$$\mathbb{P}_\infty(\hat{\tau}_i < n) \le n/B_i, \quad n \ge 1,$$

setting $B_i = h^{\omega_i}$ with $\omega_i = \mathcal{I}_i/\mathcal{I}_{\text{tot}}$ and $h > 0$, we obtain

$$\text{PFA}^\rho(\hat{\tau}_{\max}) = \sum_{n=1}^{\infty} \pi_n \prod_{i=1}^{N} \mathbb{P}_\infty(\hat{\tau}_i < n) \le \prod_{i=1}^{N} h^{-\omega_i} \sum_{n=1}^{\infty} n^N \rho(1-\rho)^{n-1}.$$

Noting that $\sum_{i=1}^N \omega_i = 1$ and $\sum_{n=1}^\infty n^N \rho(1-\rho)^{n-1} = \mu_N$ yields $\text{PFA}^\rho(\hat{\tau}_{\max}) \le \mu_N/h$, where $\mu_N = \mu_N(\rho)$ is the Nth moment of the geometric random variable with parameter ρ. Therefore, $h_\alpha = \mu_N/\alpha$ guarantees the inequality $\text{PFA}^\rho(\hat{\tau}_{\max}) \le \alpha$. Clearly the same bound holds for $\hat{\tau}_{\text{all}}$ as well, because $\hat{\tau}_{\text{all}}$ is greater than $\hat{\tau}_{\max}$ almost surely.

Furthermore, we prove below that

$$\text{ADD}^\rho(\hat{\tau}_{\max}) \sim \text{ADD}^\rho(\hat{\tau}_{\text{all}}) \sim \max_{1 \le i \le N} \frac{\log B_i}{\mathcal{I}_i}.$$

as $\min_i B_i \to \infty$. Now, because

$$\log h_\alpha = |\log \alpha| + \log \beta_N = |\log \alpha| + O(1),$$

it follows that

$$\text{ADD}^\rho(\hat{\tau}_{\max}) \sim \text{ADD}^\rho(\hat{\tau}_{\text{all}}) \sim \frac{|\log \alpha|}{\mathcal{I}_{\text{tot}}} \quad \text{as } \alpha \to 0. \tag{4.8}$$

The following theorem formalizes the asymptotic performance of the these two LD procedures.

Theorem 4.1. *Let* $\hat{\tau}_{\text{all}}$ *and* $\hat{\tau}_{\max}$ *be as in* (4.6) *and* (4.7), *respectively. If* $B_i = (\mu_N/\alpha)^{\mathcal{I}_i/\mathcal{I}_{\text{tot}}}$, $i = 1, \dots, N$, *then* $\text{PFA}^\rho(\hat{\tau}_{\text{all}}) \le \text{PFA}^\rho(\hat{\tau}_{\max}) \le \alpha$ *and, as* $\alpha \to 0$,

$$\begin{aligned} \text{CADD}_k(\hat{\tau}_{\max}) \sim \text{CADD}_k(\hat{\tau}_{\text{all}}) \sim \frac{|\log \alpha|}{\mathcal{I}_{\text{tot}}} \quad \forall k \ge 1, \\ \text{ADD}^\rho(\hat{\tau}_{\max}) \sim \text{ADD}^\rho(\hat{\tau}_{\text{all}}) \sim \frac{|\log \alpha|}{\mathcal{I}_{\text{tot}}}. \end{aligned} \tag{4.9}$$

Suppose, in addition, that $\mathbb{E}_1|Z_i(1)|^3 < \infty$, *and the Cramér condition* $\limsup_{t\to\infty} \mathbb{E}_1 \exp\{\jmath \cdot t(Z_i(1) - \mathcal{I}_i)\} < 1$ *for the characteristic function of* $Z_i(1) - \mathcal{I}_i$ *is satisfied. Let* $B_i = h^{\omega_i}$. *Then*

$$\text{ADD}^\rho(\hat{\tau}_{\max}) \le \text{ADD}^\rho(\hat{\tau}_{\text{all}}) = \frac{\log h}{I_{\text{tot}}} + C_N\sqrt{\frac{\log h}{\mathcal{I}_{\text{tot}}}} + c + o(1), \quad \text{as } h \to \infty, \tag{4.10}$$

where C_N *is defined in* (4.3) *and* c *is another constant.*

Proof. A detailed proof of this theorem is tedious. We present only the major ideas and a proof sketch.

It can be shown that the stopping time $\hat{\tau}_{\text{all}}$ can be represented in the following form

$$\hat{\tau}_{\text{all}} = \min\left\{n \ge 1 : \min_{1\le i\le N} \frac{\mathcal{I}_{\text{tot}}}{\mathcal{I}_i}\left(\sum_{k=1}^n Z_i(k) + G_i(n)\right) \ge \log h\right\},$$

where $\{G_i(n)\}_{n\ge1}$, $i = 1, \dots, N$ are slowly changing sequences such that $n^{-1}G_i(n)$ converge $\mathbb{P}_1$-a.s. to 0. Because $n^{-1}\sum_{k=1}^n Z_i(k)$ converges to $\mathcal{I}_i$ $\mathbb{P}_1$-a.s., it follows that, as $n \to \infty$,

$$\frac{1}{n}\frac{\mathcal{I}_{\text{tot}}}{\mathcal{I}_i}\left(\sum_{k=1}^n Z_i(k) + G_i(n)\right) \to \mathcal{I}_{\text{tot}} \quad \mathbb{P}_1\text{-a.s.}$$

and, by Theorem 2.3 of Tartakovsky (1998a),

$$\mathbb{E}_1\hat{\tau}_{\text{all}} \sim (\log h)/\mathcal{I}_{\text{tot}} \quad \text{as } h \to \infty.$$

The same asymptotic approximation holds for $\text{CADD}_k(\hat{\tau}_{\text{all}})$ for any k, which implies the first asymptotic formula in (4.9). The proof of the second approximation in (4.9) requires certain additional details that are omitted.

The proof of (4.10) can be constructed based on Theorem 4 in Tartakovsky et al. (2003) making use the fact that $\{G_i(n)\}_{n\ge1}$, $i = 1, \dots, N$ are slowly changing sequences. The details are omitted. □

Comparing (4.9) with (2.27), we can see that the first-order asymptotic operating characteristics of the proposed decentralized procedures are the same as those of the SR centralized procedure in the sense that

$$\lim_{\alpha\to0}\frac{\text{ADD}^\rho(\hat{\tau}_{\max})}{\text{ADD}^\rho(\hat{\tau}^c)} = \lim_{\alpha\to0}\frac{\text{ADD}^\rho(\hat{\tau}_{\text{all}})}{\text{ADD}^\rho(\hat{\tau}^c)} = 1.$$

However, these decentralized procedures have somewhat worse second-order performance, because for the SR centralized procedure the second term of expansion of ADD is a constant, whereas for $\hat{\tau}_{\max}$ (and $\hat{\tau}_{\text{all}}$) it grows as a square root of the threshold (see (4.10)). Thus,

$$\text{ADD}^{\rho}(\hat{\tau}_{\max}) - \text{ADD}^{\rho}(\hat{\tau}^{c}) = O(|\log \alpha|^{1/2}) \to \infty \quad \text{as } \alpha \to 0.$$

We remark that we were unable to obtain similar results if the optimal Bayesian (Shiryaev) procedure is used in place of the SR procedure at the sensors. We also note that although the procedures $\hat{\tau}_{\max}$ and $\hat{\tau}_{\text{all}}$ are asymptotically almost globally optimal, their performance for moderate values of α may be far from optimum, and may even be inferior to the procedure that uses binary quantizers at the sensors (see Figure 4 in Section 5.2). Finally, because $\hat{\tau}_{\text{all}}$ uses more information than $\hat{\tau}_{\max}$, we expect it to perform better.

It is also interesting to note that although CUSUM-based $T_{\max}$ and SR-based $\hat{\tau}_{\max}$ LD procedures are by no means asymptotically optimal in the minimax sense in the class $\Delta_{\text{m}}(\gamma)$ with the ARL to false alarm constraint, they are asymptotically optimal in the class $\Delta_{\text{B}}(\alpha)$ with the constraint on the average PFA.

5. SIMULATION RESULTS

5.1. Example 1: Discrete Poisson Case (Minimax Setting)

In this section, we present the results of MC experiments for the Poisson example where observations $X_i(n)$, $n \geq 1$ in the ith sensor follow the common Poisson distribution $\mathscr{P}(\mu_i)$ in the prechange mode and the common Poisson distribution $\mathscr{P}(\theta_i)$ after the change occurs, i.e., for $m = 0, 1, 2, \ldots$ and $\lambda = k$,

$$\mathbb{P}_k(X_i(n) = m) = \begin{cases} \dfrac{(\mu_i)^m}{m!} e^{-\mu_i} & \text{for } k > n, \\ \dfrac{(\theta_i)^m}{m!} e^{-\theta_i} & \text{for } k \leq n, \end{cases}$$

where without loss of generality we assume that $\theta_i > \mu_i$.

Write $Q_i = \theta_i/\mu_i$. It is easily seen that the LLR statistic in the ith senor has the form

$$Z_n(i) = X_i(n) \log(Q_i) - \mu_i(Q_i - 1), \tag{5.1}$$

and the K–L information numbers

$$\mathscr{I}_i = \theta_i \log Q_i - \mu_i(Q_i - 1), \quad i = 1, \ldots, N. \tag{5.2}$$

It follows from (2.8), (5.2) and the above discussion that the centralized CUSUM and AO-LD-CUSUM tests with the thresholds $h = \log \gamma$ are first-order globally asymptotically optimal and

$$\inf_{\tau \in \Delta_{\text{m}}(\gamma)} \text{SADD}(\tau) \sim \text{SADD}(\tau^c) \sim \text{SADD}(T_{\text{ld}}) \sim \frac{\log \gamma}{\sum_{i=1}^{N} [\theta_i \log Q_i - \mu_i(Q_i - 1)]}. \tag{5.3}$$

This means that the ARE of these detection tests with respect to the globally optimal test is equal to 1.

In order to evaluate the ARE of an optimal test v (e.g., the centralized CUSUM test τ^c or the SR test $\hat{\tau}^c$) with respect to the BQ-CUSUM test τ^q defined in (3.4), we use (3.19), which yields

$$\mathrm{ARE}(v;\tau^q) = \frac{\sum_{i=1}^N \max_{t_i}[g_i(t_i)c_i(t_i) + c_{0,i}(t_i)]}{\sum_{i=1}^N [\theta_i \log Q_i - \mu_i(Q_i - 1)]}, \tag{5.4}$$

where the probabilities $g_{0,i}(t_i)$ and $g_i(t_i)$ are given by:

$$g_{0,i}(t_i) = \sum_{k=\lceil t_i \rceil}^{\infty} \frac{\mu_i^k e^{-\mu_i}}{k!}, \quad g_i(t_i) = \sum_{k=\lceil t_i \rceil}^{\infty} \frac{\theta_i^k\, e^{-\theta_i}}{k!}.$$

Note that because the LLRs are monotone functions of $X_i(n)$, it is equivalent to quantize the observations. Here and in the following the thresholds t_i are set in the space of observations rather than in LLR space.

The optimal values of t_i^0 that maximize the K–L numbers (3.17) are easily found based on these formulas. Consider a symmetric case where $\mu_i = 10$ and $\theta_i = 12$ for all $i = 1, \dots, N$. Then $\mathcal{I}_i = \mathcal{I} = 0.1879$, the optimal threshold is $t_i^0 = 12$, and the corresponding maximum K–L information number for the binary sequence $\mathcal{I}_i^q(t_i^0) = \mathcal{I}^q = 0.119$. Therefore, the loss in efficiency of the BQ-CUSUM test compared to the globally asymptotically optimal detection procedure is $\mathrm{ARE}(v;\tau^q) = 0.119/0.1879 = 0.63$, i.e., for large ARL we expect about a 37% (compare with Remark 3.3) increase in the average detection delay compared to the centralized CUSUM (C-CUSUM). The following MC simulations show that for the practically interesting values of the ARL (up to 13,360) the gain of the optimal C-CUSUM test is even smaller, whereas the AO-LD-CUSUM test performs worse than the BQ-CUSUM test due to the reasons discussed in Section 4.1.1.

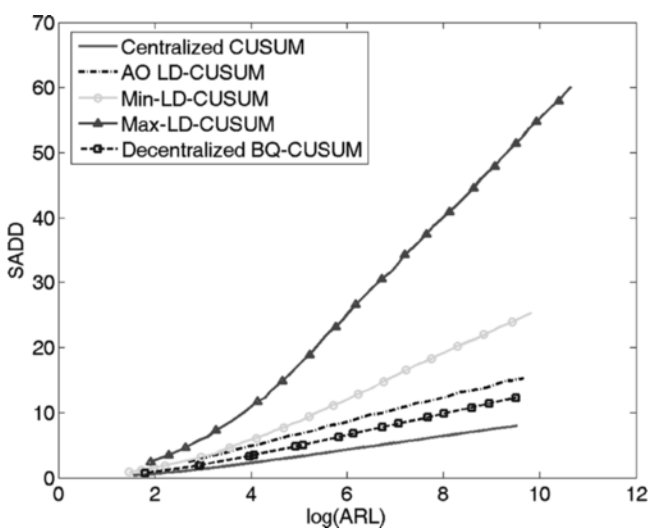

Figure 3. Operating characteristics of detection procedures.

Table 1. Operating characteristics of detection procedures

log(ARL)	3.5	4.5	5.5	6.5	7.5	8.5	9.5
ARL	33	90	245	665	1808	4915	13360
SADD for C-CUSUM	1.82	2.79	3.81	4.85	5.90	6.94	8.00
SADD for AO-LD-CUSUM	3.87	5.79	7.72	9.68	11.52	13.28	15.06
SADD for Min-LD-CUSUM	4.47	7.28	10.46	13.75	17.50	20.84	24.17
SADD for Max-LD-CUSUM	8.30	13.91	21.39	28.95	36.38	43.65	51.37
SADD for BQ-CUSUM	2.75	4.21	5.77	7.40	9.01	10.65	12.28

MC simulations have been performed for the above symmetric situation (i.e., $\mu_i = \mu = 10$ and $\theta_i = \theta = 12$) with $N = 5$ sensors. We used 10^5 MC replications in the experiment. The operating characteristics of the five detection tests (SADD versus log(ARL)) are shown in Figure 3 and Table 1. It is seen that the BQ-CUSUM test substantially outperforms the AO-LD-CUSUM test for the entire false alarm rate range used in simulations. This result confirms our conjecture. It is also seen that both Min-LD-CUSUM and Max-LD-CUSUM perform worse than both BQ-CUSUM and AO-LD-CUSUM tests, especially the Max-LD-CUSUM test.

Table 2 shows the relative efficiency of the BQ-CUSUM procedure with respect to the four other detection procedures, which is defined as the ratio of average detection delays for the same ARL: $\mathrm{SADD}(\tau^q)/\mathrm{SADD}(\nu)$, where ν is a corresponding detection test, i.e., $\nu = \tau^c$, T_{ld}, etc. It follows from the table that for the BQ-CUSUM the increase in the SADD compared to the globally optimal centralized CUSUM is 34% for high false alarm rate, 35% for moderate and low false alarm rate, and 37% for very low false alarm rate. Note that the last column presents the ARE given by (5.4), which according to Remark 3.3 is expected to be approximately $\pi/2 \approx 1.57$ and in reality turns out to be 1.59 for the parameters considered. On the other hand, the BQ-CUSUM outperforms the AO-LD-CUSUM for all range of tested ARL values, from 33 to 13,360. The gain is 30% for high false alarm rate and slowly reduces to 18% for low false alarm rate.

5.2. Example 2: Gaussian Case (Bayesian Setting)

Consider the problem of detecting a nonfluctuating target using N 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 matched filter, matched to the signal corresponding to the target.

Table 2. Relative efficiency of the decentralized BQ-CUSUM test

log(ARL)	3.5	4.5	5.5	6.5	7.5	8.5	9.5	∞
ARL	33	90	245	665	1808	4915	13360	∞
Test	Relative efficiency of the decentralized BQ-CUSUM test							
C-CUSUM	1.51	1.51	1.51	1.53	1.53	1.53	1.54	1.59
AO-LD-CUSUM	0.71	0.73	0.75	0.76	0.78	0.80	0.82	1.59
Min-LD-CUSUM	0.62	0.58	0.55	0.54	0.51	0.51	0.51	0.316
Max-LD-CUSUM	0.33	0.30	0.27	0.26	0.25	0.24	0.24	0.316

The output of the matched filter at sensor S_i at time n (when the time of appearance of the target is λ) is given by $X_i(n) = \xi_i(n)$ if $n < \lambda$ and $X_i(n) = \mu_i + \xi_i(n)$ if $n \geq \lambda$, where $\{\xi_i(n), n = 1, 2, \dots\}$ is a sequence of i.i.d. zero-mean Gaussian random variables with variance σ_i^2. Therefore, the LLR at sensor S_i at time n is given by

$$Z_i(n) = \log \frac{f_i^{(1)}(X_i)}{f_i^{(0)}(X_i)} = \frac{\mu_i X_i(n)}{\sigma_i^2} - \frac{\mu_i^2}{2\sigma_i^2}.$$

From this LLR sequence, we can compute the centralized Shiryaev statistic $R_\rho^c(n)$, the centralized SR statistic $R^c(n)$, and the SR statistics for the individual sensors $R_i(n)$ for the LD tests $\hat{\tau}_{\text{all}}$ and $\hat{\tau}_{\text{max}}$.

Next, consider the tests based on quantization at the sensors. Note that the LLR is a monotonically increasing function of the observation, and hence we can characterize the optimal stationary sensor quantizers in terms of thresholds on the observations. For *binary* quantization (BQ) at the sensors, the quantizers are characterized by a single threshold, i.e., $U_i(n) = 1$ if $X_i(n) \geq t_i$ and 0 otherwise. The distributions induced on $U_i(n)$ by this quantizer are given by

$$g_i^{(j)}(0) = 1 - g_i^{(j)}(1) = \Phi\left(\frac{t_i - j\mu_i}{\sigma_i}\right), \quad j = 0, 1.$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard Gaussian random variable.

The optimal value of t_i, i.e., the one that maximizes $\mathcal{I}_i^q = \mathcal{I}(g_i^{(1)}, g_i^{(0)})$, is easily found based on this formula. Then we can compute the decision statistics $R_\rho^q(n)$ and $R^q(n)$, respectively, for the Shiryaev and SR fusion rules based on the optimal binary quantized sensor outputs.

The operating characteristics in an example with three sensors having identically distributed observations are illustrated in Figure 4. The parameter values are

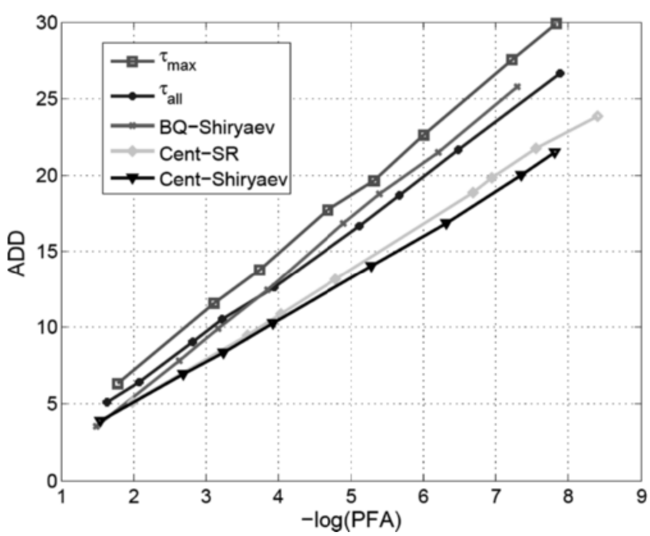

Figure 4. Operating characteristics for an example with three sensors with identically distributed Gaussian observations.

$\rho = 0.1$, $\mu_i = 0.4$, and $\sigma_i^2 = 1$. The K–L information number for the sensor observations is 0.08. The threshold that maximizes the K–L information number at the output of the sensor is $t_i = 0.32$, $i = 1, 2, 3$, and the corresponding maximum K–L information number is 0.0509. Estimates of PFA^ρ and ADD^ρ are obtained using MC methods with the number of trials being $1000/\alpha$. As we expect, for the optimal (Shiryaev) centralized policies, the plot of ADD^ρ versus $-\log(\mathrm{PFA}^\rho)$ is roughly a straight line with slope that is approximately equal to $1/[3\mathcal{I}^c + |\log(1-\rho)|] \approx 2.89$. Interestingly, the SR centralized policy has very similar performance even though the asymptotic slope in this case is $1/(3\mathcal{I}^c) \approx 4.17$. This justifies the use of the SR policy at the sensors in constructing $\hat{\tau}_{\max}$ and $\hat{\tau}_{\mathrm{all}}$. The decentralized policy with binary sensor quantizers and a Shiryaev fusion rule (BQ-Shiryaev) has a trade-off curve with slope that is roughly equal to $1/[\mathcal{I}^q_{\mathrm{tot,opt}} + |\log(1-\rho)|] \approx 3.87$, as expected from Theorem 3.2. 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 decentralized setting, especially with binary quantizers.

Figure 4 also shows the trade-off curves for the procedures $\hat{\tau}_{\max}$ and $\hat{\tau}_{\mathrm{all}}$, where the sensors perform local change detection. It is seen that $\hat{\tau}_{\max}$ performs worse than $\hat{\tau}_{\mathrm{all}}$. Further, it is interesting to see that both of these procedures have performances that are far from that of the centralized SR procedure. Thus, the asymptotic results of Theorem 4.1 appear to hold only for PFA much smaller than 10^{-4} (the smallest value of PFA considered in our simulations).[1] In particular, $\hat{\tau}_{\max}$ performs even worse than the BQ-Shiryaev policy, whereas $\hat{\tau}_{\mathrm{all}}$ is slightly better than the BQ-Shiryaev policy for sufficiently small PFA. These results clearly point to the need for further research on designing procedures that perform local detection at the sensors.

6. CONCLUSIONS

1. We presented a comprehensive asymptotic analysis of centralized Shiryaev, Shiryaev–Roberts, and CUSUM change detection procedures in three problem settings—minimax, uniform, and Bayesian in the classes of detection tests with constraints imposed, respectively, on the ARL to false alarm, on the local PFA in a fixed time interval, and on the average PFA. All three tests serve as benchmarks for a comparative study of decentralized change detection procedures in distributed sensor systems.

2. We proposed two types of decentralized detection procedures that use compressed data. These compressed data are transmitted to the fusion center for making the final decision. The first class uses data quantization at sensors. The second uses local (binary) decisions at sensors, which are then transmitted to a fusion center for combining and making a final decision. For both types of decentralized detection procedures, the required bandwidth for communication with the fusion center is minimal, especially if the quantization is binary. The decentralized procedures with quantized data have the further advantage that they do not require any processing power at the sensors.

[1]Note that Figure 4 shows the natural logarithm of PFA.

3. We have found asymptotically optimal decentralized quickest change detection procedures in both classes. In the case of quantization-based procedures, we have shown that it is optimal for the sensors to use likelihood ratio quantizers that maximize the K–L information divergence between the post- and prechange distributions at the sensor outputs. The fusion center then simply implements a standard change detection procedure based on the quantized data. In the case of local decision-based procedures, we have found procedures that have the same first-order asymptotic performance as the corresponding centralized procedures that have access to all of the sensor observations. In the minimax setting, we found one procedure, T_{ld}, that has this global optimality property, whereas in the Bayesian setting, we found two procedures, $\hat{\tau}_{\mathrm{all}}$ and $\hat{\tau}_{\max}$, that are globally asymptotically optimal.

4. We presented simulation results that compare the performances of the local decision (LD)-based tests and the binary quantization (BQ)-based tests. Our results show that the globally asymptotically optimal LD tests may actually perform worse than the corresponding BQ tests in both the minimax and Bayesian settings at moderate levels of false alarms. Furthermore, the loss in performance due to binary quantization relative to centralized procedures is generally small for Gaussian, Poisson, and Exponential models as we discussed in Remark 3.3. Thus, our results recommend the use of decentralized procedures based on binary quantization at the sensors. Our results also point to the need for further research on designing procedures that perform local detection at the sensors that provide good performance at moderate levels of false alarms.

5. In our study, we assumed complete information about the prechange and the postchange models. However, in a variety of applications, including intrusion detection in computer networks (see, e.g., Kent, 2000; Tartakovsky et al., 2005, 2006), these models are either partially known or unknown. The development of optimal or quasioptimal decentralized detection methods for composite postchange hypotheses and unknown distributions is a challenging problem. In the case of one-parameter postchange families of distributions, an efficient solution to this problem can be obtained by implementing binary quantization to $M \geq 2$ isolated points with a subsequent use of a multichart BQ-CUSUM. In particular, a specially designed 2-CUSUM test performs fairly well (see Tartakovsky and Polunchenko, 2008). However, it is not clear how this problem can be addressed when the distributions are not known. In centralized systems, effective rank-based approaches have been developed in Gordon and Pollak (1994) and Pollak (2007); it is of interest to see if these methods can be adapted to decentralized systems.

ACKNOWLEDGMENTS

The work of Alexander Tartakovsky was supported in part by the U.S. Office of Naval Research grant N00014-06-1-0110 and the U.S. Army Research Office MURI grant W911NF-06-1-0094 at the University of Southern California and by the U.S. ARMY SBIR contract W911QX-04-C-0001 at ADSANTEC. The work of Venugopal Veeravalli was supported in part by the U.S. Army Research Office MURI grant W911NF-06-1-0094, through a subcontract from Brown University at the University of Illinois, and the U.S. National Science Foundation grant CCF-0049089 at the University of Illinois. The authors would like to thank Hongjoong Kim for the help in MC simulations in Example 1.

REFERENCES

Baron, M. and Tartakovsky, A. G. (2006). Asymptotic Bayesian Change-Point Detection Theory for General Continuous-Time Models, *Sequential Analysis* 25: 257–296.

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

Dragalin, V. P., Tartakovsky, A. G., and Veeravalli, V. V. (1999). Multihypothesis Sequential Probability Ratio Tests, Part I: Asymptotic Optimality, *IEEE Transactions on Information Theory* 45: 2448–2461.

Dragalin, V. P., Tartakovsky, A. G., and Veeravalli, V. V. (2000). Multihypothesis Sequential Probability Ratio Tests, Part 2: Accurate Asymptotic Expansions for the Expected Sample Size, *IEEE Transactions on Information Theory* 46: 1366–1383.

Gordon, L. and Pollak, M. (1994). An Efficient Nonparametric Detection Scheme for Detecting a Change in Distribution, *Annals of Statistics* 22: 763–804.

Kent, S. (2000). On the Trial of Intrusions into Information Systems, *IEEE Spectrum* 37(12): 52–56.

Lai, T. L. (1995). Sequential Changepoint Detection in Quality Control and Dynamical Systems, *Journal of Royal Statistal Society, Series B* 57: 613–658.

Lai, T. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917–2929.

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897–1908.

Mei, Y. (2005). Information Bounds and Quickest Change Detection in Decentralized Decision Systems, *IEEE Transactions on Information Theory* 51: 2669–2681.

Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14: 1379–1387.

Moustakides, G. V. (2006). Decentralized CUSUM Change Detection, Proceedings of the 9th International Conference on Information Fusion (Fusion 2006), Florence, Italy, July 2006, CD Rom ISBN 0-9721844-6-5.

Moustakides, G. V. (2008). Sequential Change Detection Revisited, *Annals of Statistics* 36: 787–807.

Page, E. S. (1954). Continuous Inspection Schemes, *Biometrika* 41: 100–115.

Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206–227.

Pollak, M. (1987). Average Run Lengths of an Optimal Method of Detecting a Change in Distribution, *Annals of Statistics* 15: 749–779.

Pollak, M. (2007). Nonparametric Detection of a Change, Proceedings of the 56th Session of the International Statistical Institute (ISI) 22–29 August 2007, Lisbon, Portugal.

Pollak, M. and Tartakovsky, A. G. (2008a). Asymptotic Exponentiality of the Distribution of First Exit Times for a Class of Markov Processes with Applications to Quickest Change Detection, Theory of Probability and Its Applications, in press.

Pollak, M. and Tartakovsky, A. G. (2008b). Optimality of the Shiryaev–Roberts Procedure, *Statistica Sinica*, in press.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer-Verlag.

Shiryaev, A. N. (1961). The Detection of Spontaneous Effects, *Soviet Mathematics—Doklady* 2: 740–743.

Shiryaev, A. N. (1963). On Optimum Methods in Quickest Detection Problems, *Theory of Probability and Its Applications* 8: 22–46.

Shiryaev, A. N. (1978). *Optimal Stopping Rules*, New York: Springer-Verlag.

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

Tartakovsky, A. G. (1994). Asymptotically Minimax Multialternative Sequential Rule for Disorder Detection, *Statistics and Control of Random Processes: Proceedings of Steklov Institute of Mathematics*, vol. 202, pp. 229–236, Providence: American Mathematical Society.

Tartakovsky, A. G. (1998a). Asymptotic Optimality of Certain Multihypothesis Sequential Tests: Non-i.i.d. Case, *Statistical Inference for Stochastic Processes* 1: 265–295.

Tartakovsky, A. G. (1998b). Extended Asymptotic Optimality of Certain Change-Point Detection Procedures, Technical report/preprint, Center for Applied Mathematical Sciences, University of Southern California.

Tartakovsky, A. G. (2005). Asymptotic Performance of a Multichart CUSUM Test Under False Alarm Probability Constraint, Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference (CDC-ECC'05), 12–15 December 2005, pp. 320–325, Seville, Spain, Omnipress CD-ROM, ISBN 0-7803-9568-9.

Tartakovsky, A. G. and Veeravalli, V. V. (2002). An Efficient Sequential Procedure for Detecting Changes in Multichannel and Distributed Systems, Proceedings of the 5th International Conference on Information Fusion (Fusion 2002), Annapolis, MD, 8–11 July 2002, vol. 1, pp. 41–48.

Tartakovsky, A. G. and Veeravalli, V. V. (2003). Quickest Change Detection in Distributed Sensor Systems, Proceedings of the 6th International Conference on Information Fusion (Fusion 2003), Cairns, Australia, 8–10 July 2003, pp. 756–763.

Tartakovsky, A. G. and Veeravalli, V. V. (2004). Change-Point Detection in Multichannel and Distributed Systems with Applications, *Applications of Sequential Methodologies*, N. Mukhopadhyay, S. Datta, and S. Chattopadhyay, eds., pp. 339–370, New York: Marcel Dekker, Inc.

Tartakovsky, A. G. and Veeravalli, V. V. (2005). General Asymptotic Bayesian Theory of Quickest Change Detection, *Theory of Probability and Its Applications* 49: 458–497.

Tartakovsky, A. G. and Kim, H. (2006). Performance of Certain Decentralized Distributed Change Detection Procedures, Proceedings of the 9th International Conference on Information Fusion (Fusion 2006), Florence, Italy, July 2006, CD Rom ISBN 0-9721844-6-5.

Tartakovsky, A. G. and Polunchenko, A. S. (2008). Quickest Changepoint Detection in Distributed Multisensor Systems under Unknown Parameters, Proceedings of the 11th International Conference on Information Fusion (Fusion 2008), Hyatt Regency Hotel, Cologne, Germany, 30 June–3 July 2008, CD Rom.

Tartakovsky, A. G., Li, X. R., and Yaralov, G. (2003). Sequential Detection of Targets in Multichannel Systems, *IEEE Transactions on Information Theory* 49: 425–445.

Tartakovsky, A. G., Shah, K., and Rozovskii, B. L. (2005). A Nonparametric Multichart CUSUM Test for Rapid Intrusion Detection, Proceedings of Joint Statististical Meetings (JSM), Minneapolis, MN, 7–11 August 2005 (CD Rom).

Tartakovsky, A. G., Rozovskii, B. L., Blažek, R., and Kim, H. (2006). Detection of Intrusions in Information Systems by Sequential Change-Point Methods (with Discussion), *Statistical Methodology* 3: 252–340.

Tsitsiklis, J. N. (1993). Extremal Properties of Likelihood-Ratio Quantizers, *IEEE Transactions on Communications* 41: 550–558.

Veeravalli, V. V. (2001). Decentralized Quickest Change Detection, *IEEE Transactions on Information Theory* 47: 1657–1665.

Woodroofe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Philadelphia: SIAM.