Data-Efficient Quickest Change Detection

Taposh Banerjee and Venugopal. V. Veeravalli*

ECE Department and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign, USA.
*Corresponding Author: vvv@illinois.edu

ABSTRACT

In the classical problem of quickest change detection, a decision maker observes a sequence of random variables. At some point in time, the distribution of the random variables changes abruptly. The objective is to detect this change in distribution with minimum possible delay, subject to a constraint on the false alarm rate. In many applications of quickest change detection, e.g., where the changes are infrequent, it is of interest to control the cost of observations or the cost of data acquired before the change point. To this end, in this survey paper, data-efficient versions of the classical quickest change detection problems are studied, and a summary of existing results on these problems is provided. Some extensions to distributed sensor networks are also discussed.

1. Introduction

The problem of detecting an abrupt change in the statistical properties of a phenomenon under observation is encountered in many applications. Examples include quality control, infrastructure, environment or habitat monitoring, and spectrum sensing in cognitive radios. A mathematical formulation in statistics that most accurately models the above detection problem is the problem of quickest change detection (QCD).

In the classical QCD problem, a decision maker observes a sequence of random variables $\{X_n\}$. At some point in time γ, called the change point, the distribution of $\{X_n\}$ changes. The objective is to find a stopping time τ on the sequence $\{X_n\}$ so as to detect this change in distribution as quickly as possi-

ble. Mathematically, the goal is to minimize some metric on the detection delay subject to a constraint on a suitable metric on the false alarms.

The general formulation of the QCD problem allows for arbitrary dependence between the observations, and even allows for observations to be observed in continuous time. However, in this paper we focus on the *i.i.d. model*. In the i.i.d. model, the observations $\{X_n\}$ are independent, conditioned on the change point γ, and identically distributed before and after γ. See Poor and Hadjiliadis (2009), Veeravalli and Banerjee (2013) and Tartakovsky et al. (2014) for a general background on QCD, including references and discussions of applications in engineering and sciences. In Section 2, we provide a brief overview of this classical problem that is relevant to our work.

In many applications, for example surveillance using sensor networks and statistical process control, changes are infrequent, and there is a cost associated with taking observations or acquiring data. However, in the classical problem formulations, although there is a penalty on acquiring data after the change through the metric on delay, there is no penalty on the cost of observations taken before the change point. This motivates the study of a data-efficient version of the classical QCD problem. In Section 3, we discuss our recent results on data-efficient quickest change detection (DE-QCD). In DE-QCD, the objective of the decision maker is to minimize a metric on the detection delay, subject to constraints on a metric on the false alarm and a metric on the cost of observations used before the change point.

The classical QCD problem has been extended in many ways, of which the most practically relevant extension has been to QCD in distributed sensor networks. In Section 4, we discuss ways in which the DE-QCD results of Section 3 can be extended to some sensor network models.

2. Classical Quickest Change Detection: An Overview

In this section we provide a brief overview of the results from the classical QCD literature. A more detailed introduction can be found for example in Veeravalli and Banerjee (2013). We will focus on the i.i.d. model, in which the random variables $\{X_n\}$ are i.i.d. with probability density function (p.d.f.) f_0 before the change point γ, and i.i.d. with p.d.f. f_1 after γ. If the distribution of the change point γ is known, then γ is modelled as a random variable Γ, and the problem is studied in a Bayesian setting. If the distribution of γ is not known, then the change point is modelled as an unknown constant and the

QCD problem is studied in minimax settings. Let $\mathbb{P}_n$ (correspondingly $\mathbb{E}_n$) be the probability measure (correspondingly expectation) when the change occurs at time n. Then, $\mathbb{P}_\infty$ and $\mathbb{E}_\infty$ stand for the probability measure and expectation when the change does not occur.

2.1. Bayesian Quickest Change Detection

In the Bayesian setting studied by Shiryaev (1963), the change point γ is modeled as *geometric* with parameter ρ, i.e., for $0 < \rho < 1$

$$\pi_n = \mathbb{P}\{\Gamma = n\} = \rho(1-\rho)^{n-1}\,\mathbb{I}_{\{n \geq 1\}}, \quad \pi_0 = 0, \tag{2.1}$$

where $\mathbb{I}_F$ represents the indicator of the event F.

Let

$$I_n = [X_1, \ldots, X_n],$$

represents the information at time n. For time $n \geq 1$, based on the information vector I_n, a decision is made whether to *stop and declare change* or *to continue taking observations*. Let τ be a stopping time on the information sequence $\{I_n\}$, that is $\mathbb{I}_{\{\tau=n\}}$ is a measurable function of I_n. Define the average detection delay (ADD) and the probability of false alarm (PFA), as

$$\text{ADD}(\tau) \;=\; \mathbb{E}\left[(\tau - \Gamma)^+\right] = \sum_{n=0}^{\infty} \pi_n \mathbb{E}_n\left[(\tau - n)^+\right] \tag{2.2}$$

$$\text{PFA}(\tau) \;=\; \mathbb{P}(\tau < \Gamma) = \sum_{n=0}^{\infty} \pi_n \mathbb{P}_n(\tau < n) \tag{2.3}$$

Then the Bayesian quickest change detection problem as formulated by Shiryaev is as follows.

Problem 2.1.

$$\begin{aligned} \underset{\tau}{\text{minimize}} \quad & \text{ADD}(\tau), \\ \text{subject to} \quad & \text{PFA}(\tau) \leq \alpha. \end{aligned}$$

Here, $\alpha \leq 1$ is the given constraint on the PFA.

An algorithm that is optimal for the above formulation is provided below. Define,

$$p_n = \mathbb{P}\left(\Gamma \leq n \mid I_n\right),$$

the *a posteriori* probability that change has already happened given the observations. The probability p_n can be computed recursively using the following recursion:

$$p_{n+1} = \frac{\tilde{p}_n L(X_{n+1})}{\tilde{p}_n L(X_{n+1}) + (1 - \tilde{p}_n)} \tag{2.4}$$

with $\tilde{p}_n = p_n + (1 - p_n)\rho$ and $L(X) = f_1(X)/f_0(X)$. The following theorem is proved in Shiryaev (1963):

Theorem 2.1 (Shiryaev (1963)). *If*

$$0 < D(f_0 \parallel f_1) < \infty \quad and \quad 0 < D(f_1 \parallel f_0) < \infty,$$

then for a given PFA *constraint of* α, *it is optimal to start with* $p_0 = 0$, *for* $n \geq 0$, *update* p_n *using (2.4), and stop the first time* p_n *is above the threshold* A_α:

$$\tau_S = \inf \{n \geq 1 : p_n > A_\alpha\} \tag{2.5}$$

provided, $A_\alpha \in (0, 1)$ *can be chosen such that*

$$\mathsf{PFA}(\tau_s) = \alpha.$$

Thus, the *a posteriori* probability or the belief, is computed over time, and a change is declared the first time this belief is larger than a threshold. In the following we refer to the algorithm defined in (2.5) as the Shiryaev algorithm.

In Tartakovsky and Veeravalli (2005) an asymptotic analysis for the Shiryaev algorithm is provided. A simple upper bound on the PFA of the Shiryaev algorithm is given by

$$\mathsf{PFA}(\tau_s) = \mathbb{E}[1 - p_{\tau_s}] \leq 1 - A. \tag{2.6}$$

Thus, setting $A = 1 - \alpha$ ensures that $\mathsf{PFA}(\tau_s) \leq \alpha$. However, the bound for PFA can be quite loose, and hence a more accurate estimate of the PFA is desirable. Also, since the Shiryaev algorithm is optimal, having an estimate of the delay provides a useful lower bound to the ADD of any stopping time.

Let $R(x)$ be the asymptotic distribution of the overshoot when the random walk $\sum_{k=1}^n \log \frac{f_1(X_k)}{f_0(X_k)} + n|\log(1 - \rho)|$ crosses a large positive boundary; see Siegmund (1985). Define,

$$\zeta = \int_0^\infty e^{-x} dR(x). \tag{2.7}$$

Also, let

$$\mathcal{C}_\alpha = \{\tau : \mathsf{PFA}(\tau) \leq \alpha\}. \tag{2.8}$$

Note that $\zeta \leq 1$. The following theorem is proved in Tartakovsky and Veeravalli (2005).

Theorem 2.2 (Tartakovsky and Veeravalli (2005)). *If*

$$0 < D(f_0 \parallel f_1) < \infty \quad and \quad 0 < D(f_1 \parallel f_0) < \infty,$$

then setting $A = 1 - \alpha$ ensures that

$$\mathsf{ADD}(\tau_s) \sim \left(\frac{|\log \alpha|}{D(f_1 \| f_0) + |\log(1 - \rho)|} \right) \ as \ \alpha \to 0. \qquad (2.9)$$

and further if $\log L(X)$ is non-arithmetic (see Siegmund (1985)), then

$$\mathsf{PFA}(\tau_s) \sim \zeta \, \alpha \quad as \ \alpha \to 0. \qquad (2.10)$$

The above theorem implicitly includes an asymptotic lower bound on the performance of any stopping time satisfying a PFA constraint of α:

$$\inf_{\tau \in \mathcal{C}_\alpha} \mathsf{ADD}(\tau) \geq \left(\frac{|\log \alpha|}{D(f_1 \| f_0) + |\log(1 - \rho)|} \right) \ \text{as} \ (1 + o(1)) \ \text{as} \ \alpha \to 0. \tag{2.11}$$

We note that due to (2.6), the non-arithmetic assumption is not needed for the above lower bound to be valid.

2.2. Minimax Quickest Change Detection

In this section we review the classical minimax formulations for quickest change detection. Again, we focus on the i.i.d. model, where the random variables $\{X_n\}$ are i.i.d. with probability density function (p.d.f.) f_0 before the change point γ, and i.i.d. with p.d.f. f_1 after γ.

In the absence of knowledge of the distribution of the change point, ADD and PFA are not well defined. Thus, new performance metrics are needed for the non-Bayesian setting. We note that the Shiryaev algorithm can still be used to detect a change even in a non-Bayesian setting by selecting any arbitrary value of ρ and by using the Shiryaev recursion for p_n. However, p_n will no longer be a probability, and the optimality and performance analysis results of the previous section will no longer be valid. See Veeravalli and Banerjee (2013) for a discussion on Shiryaev-Roberts family of algorithms, which have strong asymptotic optimality property for the minimax formulation in Pollak (1985), and are obtained from the Shiryaev algorithm by setting the geometric parameter ρ to zero.

Without a prior on the change point, a reasonable measure of false alarms is the mean time to false alarm, or its reciprocal, which is the false alarm rate (FAR):

$$\mathsf{FAR}(\tau) = \frac{1}{\mathbb{E}_\infty[\tau]}. \tag{2.12}$$

Finding a uniformly powerful test that minimizes the delay over all possible values of γ subject to a FAR constraint is generally not possible. Therefore it is more appropriate to study the quickest change detection problem in a minimax setting in this case. There are two popular minimax problem formulations in the literature, one due to Lorden (1971) and the other due to Pollak (1985).

In Lorden's formulation, the delay metric used is the supremum of the average delay conditioned on the worst possible realizations. This differs from the metric proposed in Lorden (1971) in that we take the delay to be zero when $\tau = \gamma$.:

$$\mathsf{WADD}(\tau) = \sup_{n \geq 1} \text{ess sup } \mathbb{E}_n \left[(\tau - n)^+ | X_1, \ldots, X_{n-1} \right]. \tag{2.13}$$

The minimax formulation proposed by Lorden (1971) is:

Problem 2.2.

$$\begin{aligned}
\underset{\Psi}{\text{minimize}} \quad & \mathsf{WADD}(\tau), \\
\text{subject to} \quad & \mathsf{FAR}(\tau) \leq \alpha.
\end{aligned}$$

Here, $0 \leq \alpha \leq 1$ is the given constraint.

A less pessimistic way to measure the delay was suggested by Pollak (1985):

$$\mathsf{CADD}(\tau) = \sup_{n \geq 1} \mathbb{E}_n[\tau - n | \tau \geq n]. \tag{2.14}$$

for all stopping times τ for which the expectation is well-defined. The minimax formulation proposed by Pollak is

Problem 2.3.

$$\begin{aligned}
\underset{\Psi}{\text{minimize}} \quad & \mathsf{CADD}(\tau), \\
\text{subject to} \quad & \mathsf{FAR}(\tau) \leq \alpha.
\end{aligned}$$

Here, $0 \leq \alpha \leq 1$ is the given constraint.

Since, τ is a stopping time on $\{X_n\}$, $\{\tau \geq n\} \in \sigma(X_1, X_2, \ldots, X_{n-1})$ and we get

$$\mathsf{WADD}(\tau) \geq \mathsf{CADD}(\tau). \tag{2.15}$$

In Lai (1998), an asymptotic lower bound on the CADD, and hence on the WADD, was obtained for any stopping time τ satisfying $\mathsf{FAR}(\tau) \leq \alpha$. Let

$$\mathcal{D}_\alpha = \{\tau : \mathsf{FAR}(\tau) \leq \alpha\}. \tag{2.16}$$

Then the following theorem is proved in Lai (1998).

Theorem 2.3 (Lai (1998))**.** *If*

$$0 < D(f_0 \,||\, f_1) < \infty \quad and \quad 0 < D(f_1 \,||\, f_0) < \infty,$$

then as $\alpha \to 0$*,*

$$\inf_{\tau \in \mathcal{D}_\alpha} \mathsf{WADD}(\tau) \geq \inf_{\tau \in \mathcal{D}_\alpha} \mathsf{CADD}(\tau) \geq \frac{|\log \alpha|}{D(f_1 \,||\, f_0)}\Big(1 + o(1)\Big). \tag{2.17}$$

Remark 2.1. The fact that $\frac{|\log \alpha|}{D(f_1 \,||\, f_0)}$ is an asymptotic lower bound for the WADD is also shown by Lorden (1971).

It is shown in Lorden (1971) that the CuSum algorithm, proposed by Page in Page (1954), achieves this lower bound. A slightly different version of the CuSum algorithm is actually exactly optimal for the Lorden's formulation; see Moustakides (1986). The CuSum algorithm is described as follows. Update the statistic using the following recursion:

$$C_{n+1} = (C_n + \log L(X_{n+1}))^+, \quad C_0 = 0; \tag{2.18}$$

and stop at

$$\tau_{\mathsf{c}} = \inf\{n \geq 1 : C_n \geq A\}. \tag{2.19}$$

The CuSum statistics is initially at 0. As observations are taken, the CuSum statistic is incremented with the log likelihood ratio of the observations. If the CuSum statistic goes below zero, the statistic is reset to zero. After the change, due to the positive mean of $\log L(X)$, the statistic C_n grows to ∞. A change is declared the first time the CuSum statistic is above the threshold A.

It can be shown that (Veeravalli and Banerjee (2013))

$$\mathsf{CADD}(\tau_{\mathsf{c}}) = \mathsf{WADD}(\tau_{\mathsf{c}}) = \mathbb{E}_1[\tau_{\mathsf{c}} - 1]. \tag{2.20}$$

The proof of the following theorem can be found in Lorden (1971) and Lai (1998).

Theorem 2.4 (Lorden (1971) , Lai (1998))**.** *If*

$$0 < D(f_0 \mid\mid f_1) < \infty \quad and \quad 0 < D(f_1 \mid\mid f_0) < \infty,$$

then with $D = \log \frac{1}{\alpha}$,

$$0 < \mathsf{FAR}(\tau_c) \leq \alpha,$$

and as $\alpha \to 0$,

$$\mathsf{CADD}(\tau_c) = \mathsf{WADD}(\tau_c) = \mathbb{E}_1[\tau_c - 1] \sim \frac{|\log \alpha|}{D(f_1 \mid\mid f_0)}.$$

Thus, the CuSum algorithm is asymptotically optimal for both Problem 2.2 and Problem 2.3 because it achieves the lower bound of Theorem 2.3

3. Data-Efficient Quickest Change Detection

As discussed in the introduction, in many practical application, it is of interest to control the cost of observations used before the change point. In Banerjee and Veeravalli (2012a) and Banerjee and Veeravalli (2013a), we studied Bayesian and minimax versions of the DE-QCD problem, respectively. Recall that in a DE-QCD problem, the objective is to minimize a metric on the delay, subject to constraints on metrics on the false alarms and the cost of observations used before the change point.

In Section 3.1, we consider the DE-QCD problem in the Bayesian setting. We first propose a metric for data-efficiency in the Bayesian setting, that captures the average number of observations used before the change point. We then consider the Bayesian problem of Shiryaev, discussed in 2.1, with an additional constraint on the average number of observations used before the change point. The search is restricted to those policies that employ on-off observation control to limit the observations used before the change. We then discuss a two-threshold extension of the Shiryaev algorithm; which we call the DE-Shiryaev algorithm. In Banerjee and Veeravalli (2012a), we showed that the DE-Shiryaev algorithm is asymptotically optimal for the proposed data-efficient Bayesian formulation.

In Section 3.2, we consider the DE-QCD problem in minimax settings. We first propose a metric for data-efficiency in minimax settings, which captures the fraction of time observations are taken before the change point. We then extend the formulations of Lorden and Pollak by putting an additional constraint on the fraction of time observations are used before the change point. We then

discuss an extension of the CuSum algorithm for DE-QCD; which we call the DE-CuSum algorithm. In Banerjee and Veeravalli (2013a), we showed that the DE-CuSum algorithm is asymptotically optimal for the proposed data-efficent minimax formulations.

The asymptotic optimality of the DE-Shiryaev and the DE-CuSum algorithms is established by showing that their performance achieves the lower bound of (2.11) and (2.17), respectively. Asymptotic optimality follows because the lower bound is valid for any stopping time, whether or not observation control is employed.

In both the classical algorithms, the Shiryaev algorithm and the CuSum algorithm, a sequence of statistics is computed using the likelihood ratio of the observations over time, and a change is declared the first time the sequence of statistics crosses a threshold. A large value of the statistic signifies that the evidence is strong of a change having already happened.

In both the DE-QCD algorithms, the DE-Shiryaev algorithm and the DE-CuSum algorithm, there are two thresholds A, B, with $B < A$. A sequence of statistics is computed over time, and a change is declared if the sequence of computed statistics crosses the threshold A. However, if the computed statistic at time n is below A, then an observation is taken at time $n + 1$ only if the statistic is above the threshold B. Essentially, when a statistic is below the threshold B, it can be viewed as a strong evidence against the change, and it can be treated as an appropriate time to drop an observation. When an observation is skipped the statistic has to be updated without using the observation; the technique for executing this is discussed in the description of the algorithms in the subsections below.

3.1. Data-Efficient Bayesian Quickest Change Detection

We consider the i.i.d. model, i.e., $\{X_n\}$ is a sequence of random variables, $\{X_n\}$ for $n < \Gamma$ are i.i.d. with p.d.f. f_0, and $\{X_n\}$ for $n \geq \Gamma$ are i.i.d. with p.d.f. f_1. We assume that Γ is geometrically distributed with parameter ρ; see (2.1).

For data-efficient quickest change detection we consider the following class of control policies. At each time n, $n \geq 0$, a decision is made as to whether to *take* or *skip* the observation at time $n + 1$. Let M_n be the indicator random variable such that $M_n = 1$ if X_n is used for decision making, and $M_n = 0$ otherwise.

Thus, M_{n+1} is a function of the information available at time n, i.e.,

$$M_{n+1} = \phi_n(I_n),$$

where, ϕ_n is the control law at time n, and

$$I_n = \left[M_1, \ldots, M_n, X_1^{(M_1)}, \ldots, X_n^{(M_n)} \right],$$

represents the information at time n. Here, $X_i^{(M_i)}$ represents X_i if $M_i = 1$, otherwise X_i is absent from the information vector I_n. Also, I_0 is an empty set. For time $n \geq 1$, based on the information vector I_n, a decision is made whether to *stop and declare change* or *to continue taking observations*. Let τ be a stopping time on the information sequence $\{I_n\}$, that is $\mathbb{I}_{\{\tau=n\}}$ is a measurable function of I_n. A policy for data-efficient quickest change detection is $\Psi = \{\tau, \phi_0, \ldots, \phi_{\tau-1}\}$.

To capture the cost of observations used before the change point, we consider the metric the average number of observations (ANO) used before the change point,

$$\mathsf{ANO}(\Psi) \triangleq \mathbb{E}\left[\sum_{n=1}^{\min(\tau, \Gamma-1)} M_n \right].$$

The data-efficient extension of the classical Bayesian problem of Shiryaev (Section 2.1) is:

Problem 3.1.

$$\begin{aligned} \underset{\Psi}{\text{minimize}} \quad & \mathsf{ADD}(\Psi), \\ \text{subject to} \quad & \mathsf{PFA}(\Psi) \leq \alpha, \\ \text{and} \quad & \mathsf{ANO}(\Psi) \leq \beta. \end{aligned} \qquad (3.1)$$

Here, α and β, with $0 \leq \alpha \leq 1$ and $0 \leq \beta$, are given constraints.

When $\beta \geq \mathbb{E}[\Gamma] - 1$, Problem 3.1 reduces to the classical Bayesian quickest change detection problem Problem 2.1. Also note that when α is small, for the Shiryaev algorithm $\mathsf{ANO} \approx \mathbb{E}[\Gamma] - 1$. Hence, the Shiryaev algorithm cannot be a solution to this problem if β is small. We now describe the DE-Shiryaev algorithm, that we studied in Banerjee and Veeravalli (2012a), which is asymptotically optimal for the above formulation, for each fixed β, as $\alpha \to 0$. Define,

$$p_n = \mathbb{P}\left(\Gamma \leq n \mid I_n \right).$$

Algorithm 3.1 (DE-Shiryaev: Ψ_D)**.** Start with $p_0 = 0$ and use the following control, with $0 \leq B < A$, for $n \geq 0$:

$$M_{n+1} = \phi_n(p_n) = \begin{cases} 0 & \text{if } p_n < B \\ 1 & \text{if } p_n \geq B \end{cases} \tag{3.2}$$

$$\tau_\mathrm{D} = \inf\{n \geq 1 : p_n > A\}.$$

The probability p_n is updated using the following recursions:

$$p_{n+1} = \begin{cases} \tilde{p}_n & \text{if } M_{n+1} = 0 \\ \frac{\tilde{p}_n L(X_{n+1})}{\tilde{p}_n L(X_{n+1}) + (1 - \tilde{p}_n)} & \text{if } M_{n+1} = 1 \end{cases}$$

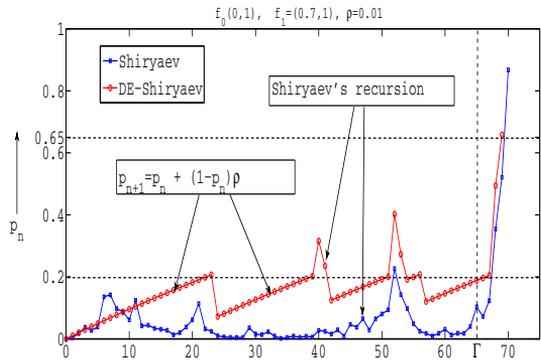with $\tilde{p}_n = p_n + (1 - p_n)\rho$ and $L(X_{n+1}) = f_1(X_{n+1})/f_0(X_{n+1})$.

Figure 3.1: Typical evolution of the Shiryaev and the DE-Shiryaev algorithm, applied to the same set of observations, with thresholds $A = 0.65$ and $B = 0.2$.

With $B = 0$ the DE-Shiryaev algorithm reduces to the Shiryaev algorithm. When $B > 0$, the statistic evolves in a similar manner as the Shiryaev statistic as long as the statistic is above B. Thus, a change is declared when $p_n > A$. However, when p_n goes below B, p_n is updated using the prior on the change point ρ (p_n increases monotonically as a result of this), and observations are skipped as long as p_n is below B. Since $p_0 = 0 < B$, few initial observations are skipped even before the observation process begins. Note that except for these initial skipped observations, the number of consecutive observations skipped at any time is a function of the undershoot of the statistic p_n (which is a function of the likelihood ratio of the observations), when it goes below B, and
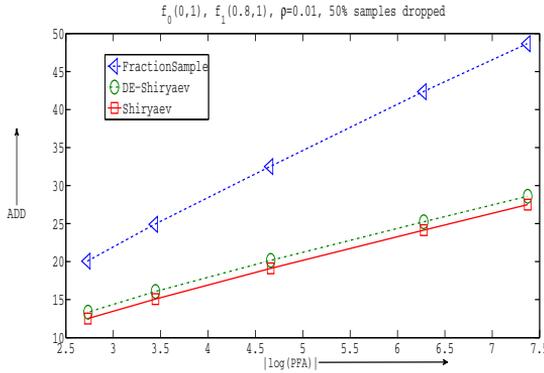
Figure 3.2: Comparative performance of schemes for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.8, 1)$, and $\rho = 0.01$.

the geometric parameter ρ. When Algorithm 3.1 is employed, the probability p_n typically evolves as depicted in Fig. 3.1. Let

$$t(B) = \inf\{n \geq 1 : p_n > B\}.$$

At and beyond $t(B)$, whenever p_n crosses B from below, it does so with an overshoot that is bounded by ρ. This is because

$$p_{n+1} - p_n = (1 - p_n)\rho \leq \rho.$$

For small values of ρ, this overshoot is essentially zero, and the evolution of p_n is roughly statistically independent of its past evolution. Thus, beyond $t(B)$, the evolution of p_n can be seen as a sequence of *two-sided* statistically independent tests, each two-sided test being a test for sequential hypothesis testing between "$H_0 =$ pre-change", and "$H_1 =$ post-change". If the decision in the two-sided test is H_0, then observations are skipped depending on the likelihood ratio of the observations (the undershoot), and the two-sided test is repeated on the observations taken beyond the skipped observations. The change is declared the first time the decision in a two-sided test is H_1.

The following theorem is proved in Banerjee and Veeravalli (2012a).

Theorem 3.1 (Banerjee and Veeravalli (2012a))**.** *If*

$$0 < D(f_0 \mid\mid f_1) < \infty \quad and \quad 0 < D(f_1 \mid\mid f_0) < \infty,$$

the threshold B can be selected such that for every $A > B$,

$$\mathsf{ANO}(\Psi_D) \leq \beta.$$

For any fixed B and $A = 1 - \alpha$ we have

$$\text{PFA}(\Psi_D) \leq \alpha,$$

and

$$\text{ADD}(\Psi_D) \leq \frac{|\log(\alpha)|}{D(f_1 \mid\mid f_0) + |\log(1-\rho)|}(1 + o(1)) \text{ as } \alpha \to 0. \qquad (3.3)$$

If further $\log L(X)$ is non-arithmetic, then for a fixed β,

$$\text{PFA}(\Psi_D) \sim \alpha\zeta \text{ as } \alpha \to 0, \qquad (3.4)$$

where ζ is as defined in (2.7).

Since, (3.3) is also the performance of the Shiryaev algorithm as $\alpha \to 0$ (see Tartakovsky and Veeravalli (2005)), the DE-Shiryaev algorithm is asymptotically optimal for Problem 3.1, for each fixed β, as $\alpha \to 0$. Equation (3.4) shows that the ratio of the PFA of the Shiryaev algorithm to that of the DE-Shiryaev algorithm goes to 1. Equation (3.4) also shows that PFA is not a function of the threshold B. In Banerjee and Veeravalli (2012a), it is shown that as $\alpha \to 0$ and as $\rho \to 0$, ANO is a function of B alone. Thus, for reasonable values of the constraints α and β, the constraints can be met independent of each other. More accurate estimates of the delay and ANO, and numerical results showing the accuracy of these estimates, are available in Banerjee and Veeravalli (2012a).

Fig. 3.2 compares the performance of the Shiryaev algorithm, the DE-Shiryaev algorithm and the fractional sampling scheme, for $\zeta = 50$. In the fractional sampling scheme, the Shiryaev algorithm is used and observations are skipped by tossing a biased coin (with probability of success 50/99), without looking at the state of the system. When an observation is skipped in the fractional sampling scheme, the Shiryaev statistic is updated using the prior on change point. The figure clearly shows a substantial gap in performance between the DE-Shiryaev algorithm and the fractional sampling scheme.

3.2. Data-Efficient Minimax Quickest Change Detection

In this section we discuss the extension of the minimax formulations of Lorden and Pollak for DE-QCD we studied in Banerjee and Veeravalli (2013a). We first describe the new metric introduced by us in Banerjee and Veeravalli (2013a) to capture the cost of observations used before the change point in a

non-Bayesian setting. With M_n, I_n, τ, and Ψ as defined earlier in Section 3.1, the Pre-change Duty Cycle (PDC) metric is defined as:

$$\text{PDC}(\Psi) = \limsup_n \frac{1}{n} \mathbb{E}_n \left[\sum_{k=1}^{n-1} M_k \Big| \tau \geq n \right]. \tag{3.5}$$

Clearly, PDC ≤ 1. Thus, the metric PDC captures the average fraction of time observations are taken before change. If all the observations are used ($M_k = 1$ $\forall k$), then PDC achieved would be equal to 1. Consider a policy in which every second observation is dropped. Then the PDC for that policy will be exactly equal to 0.5.

For false alarm, we consider the metric used in Lorden (1971) and Pollak (1985), the mean time to false alarm or its reciprocal, the false alarm rate:

$$\text{FAR}(\Psi) = \frac{1}{\mathbb{E}_\infty [\tau]}. \tag{3.6}$$

For delay we consider two possibilities: the minimax setting of Pollak (1985) where the delay metric is the following supremum over time of the conditional delay (We are only interested in those policies for which the CADD is well defined.)

$$\text{CADD}(\Psi) = \sup_n \ \mathbb{E}_n \left[\tau - n | \tau \geq n \right], \tag{3.7}$$

or the minimax setting of Lorden (1971), where the delay metric is the supremum over time of the essential supremum of the conditional delay

$$\text{WADD}(\Psi) = \sup_n \text{ess sup} \ \mathbb{E}_n \left[(\tau - n)^+ | I_{n-1} \right]. \tag{3.8}$$

Note that unlike the delay metric in Lorden (1971), WADD in (3.8) is a function of the observation control through $I_{n-1} = \left[M_1, \ldots, M_{n-1}, X_1^{(M_1)}, \ldots, X_{n-1}^{(M_{n-1})} \right]$, which may not contain the entire set of observations. As in (2.15),

$$\text{CADD}(\Psi) \leq \text{WADD}(\Psi).$$

Our first minimax formulation is the following data-efficient extension of the minimax formulation of Lorden (1971).

Problem 3.2.

$$\begin{aligned} \underset{\Psi}{\text{minimize}} \quad & \text{WADD}(\Psi), \\ \text{subject to} \quad & \text{FAR}(\Psi) \leq \alpha, \\ \text{and} \quad & \text{PDC}(\Psi) \leq \beta. \end{aligned} \tag{3.9}$$

Here, $0 \leq \alpha, \beta \leq 1$ are given constraints.

We are also interested in the data-efficient extension of Pollak (1985)

Problem 3.3.

$$
\begin{aligned}
&\underset{\Psi}{\text{minimize}} &&\text{CADD}(\Psi), \\
&\text{subject to} &&\text{FAR}(\Psi) \leq \alpha, &&&(3.10) \\
&\text{and} &&\text{PDC}(\Psi) \leq \beta.
\end{aligned}
$$

Here, $0 \leq \alpha, \beta \leq 1$ are given constraints.

With $\beta = 1$, Problem 3.3 reduces to the minimax formulation of Pollak (1985), and Problem 3.2 reduces to the minimax formulation of Lorden (1971). Note that although the CuSum algorithm trivially achieves the lower bound in the above theorem, it is not asymptotically optimal for Problem 3.2 and Problem 3.3 above for every β. This is because $\text{PDC}(\tau_{\text{C}}) = 1$, and will not satisfy any constraint of $\beta < 1$. Another option is to use the CuSum algorithm and use only the mth observation, and then choose m large enough to meet the PDC constraint of β. However, the delay of such a policy will be approximately $m \text{ CADD}(\tau_{\text{C}})$. Does there exist a policy that achieves any given constraint of β on the PDC, and still achieves the lower bound in (2.17) asymptotically? In Banerjee and Veeravalli (2013a), we have shown that the answer to the above question is in the affirmative. Specifically, the DE-CuSum algorithm, a two-threshold generalization of the CuSum algorithm, can be designed to meet any given constraints of α and β. And for each fixed β, it achieves the lower bound $\left(\frac{|\log \alpha|}{D(f_1 \| f_0)} \right)$, as $\alpha \to 0$.

We now present the DE-CuSum algorithm.

Algorithm 3.2 (DE $-$ CuSum: $\Psi_{\text{w}}(A, \mu, h)$)**.** Start with $W_0 = 0$ and fix $\mu > 0$, $A > 0$ and $h \geq 0$. For $n \geq 0$ use the following control:

$$
M_{n+1} = \begin{cases} 0 & \text{if } W_n < 0 \\ 1 & \text{if } W_n \geq 0 \end{cases}
$$

$$
\tau_{\text{w}} = \inf \left\{ n \geq 1 : W_n > A \right\}.
$$

The statistic W_n is updated using the following recursions:

$$
W_{n+1} = \begin{cases} \min\{W_n + \mu, 0\} & \text{if } M_{n+1} = 0 \\ (W_n + \log L(X_{n+1}))^{h+} & \text{if } M_{n+1} = 1 \end{cases}
$$

where $(x)^{h+} = \max\{x, -h\}$ and $L(X) = \frac{f_1(X)}{f_0(X)}$.

If $h = 0$, the DE-CuSum statistic W_n never becomes negative and hence reduces to the CuSum statistic (see Page (1954)) and evolves as: $W_0 = 0$, and for $n \geq 0$,

$$W_{n+1} = \max\{0, W_n + \log L(X_{n+1})\}.$$

Thus, with $h = 0$, the DE-CuSum algorithm reduces to the CuSum algorithm (2.18).

The evolution of the DE-CuSum algorithm is plotted in Fig. 3.3. If $h = \infty$, the evolution of the DE-CuSum algorithm can be described as follows. As seen in Fig. 3.3, initially the DE-CuSum statistic evolves according to the CuSum statistic untill the statistic W_n goes below 0. Once the statistic goes below 0, observations are skipped depending on the undershoot of W_n (this is also the sum of the log likelihood ratios of the observations) and a pre-designed parameter μ. Specifically, the statistic is incremented by μ at each time step, and observations are skipped untill W_n goes above zero, at which time it is reset to zero. At this point, fresh observations are taken and the process is repeated untill the statistic crosses the threshold A, at which time a change is declared. If $h < \infty$, the maximum number of consecutive observations skipped is bounded by $h/\mu + 1$. This assumption is crucial to the WADD analysis of the DE-CuSum algorithm. This may be also desirable in some applications.

In the DE-Shiryaev algorithm, when an observation is skipped the statistic is updated using the prior on the change point. The parameter μ in the DE-CuSum algorithm is a substitute for the Bayesian prior ρ of the DE-Shiryaev algorithm, and is treated as a design parameter. The parameter μ (and also h if it is chosen to be finite) is chosen to meet the constraint on the PDC, and the threshold A is chosen to meet the constraint on the FAR. The DE-CuSum algorithm is essentially a two-threshold algorithm with an upper threshold of A and a lower threshold of $B = 0$.

In analogy with the evolution of the DE-Shiryaev algorithm, the DE-CuSum algorithm can also be seen as a sequence of independent two-sided tests. In each two-sided test a Sequential Probability Ratio Test (SPRT) (see Wald and Wolfowitz (1948)), with log boundaries A and 0, is used to distinguish between the two hypotheses "H_0 = pre-change" and "H_1 = post-change". If the decision in the SPRT is in favor of H_0, then observations are skipped based on the likelihood ratio of all the observations taken in the SPRT. A change is declared the
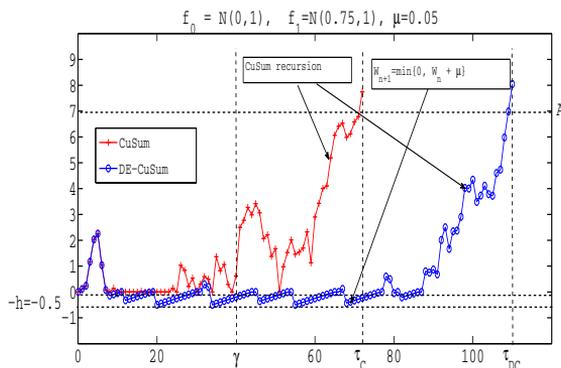
Figure 3.3: Typical evolution of CuSum algorithm and the DE-CuSum algorithm applied to the same set of observations.

first time the decision in the sequence of SPRTs is in favor of H_1. If $h = 0$, no observations are skipped and the DE-CuSum reduces to the CuSum algorithm, i.e., to a sequence of SPRTs (also see Siegmund (1985)). The DE-CuSum algorithm can also be seen as a two-threshold algorithm with the lower threshold $B = 0$.

Unless it is required to have a bound on the maximum number of observations skipped, the DE-CuSum algorithm can be controlled by just two-parameters: A and μ. In the following theorem it is shown that these two parameters can be selected independent of each other directly from the constraints. That is the threshold A can be selected so that FAR $\leq \alpha$ independent of the value of μ. Also, it is possible to select a value of μ such that PDC $\leq \beta$ independent of the choice of A. Moreover, with the parameters selected in this manner, the WADD and hence the CADD, achieves the lower bound in (2.17).

Theorem 3.2 (Banerjee and Veeravalli (2013a)). *Let*

$$0 < D(f_1 \,||\, f_0) < \infty \quad and \quad 0 < D(f_0 \,||\, f_1) < \infty.$$

1. For any μ, h and A

$$C_n \geq W_n \quad \forall n \geq 0, \tag{3.11}$$

where C_n is the CuSum statistic (2.18) and W_n is the DE-CuSum statistic, applied to the same sequence of observations. As a result

$$0 < \mathsf{FAR}(\Psi_w) \leq \mathsf{FAR}(\tau_c) \quad \forall D \geq 0, \tag{3.12}$$
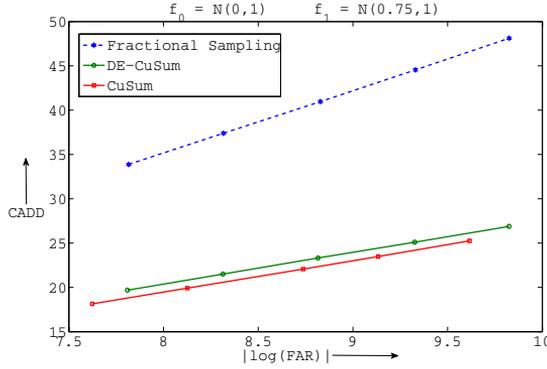
Figure 3.4: Comparative performance of the DE-CuSum algorithm with the CuSum algorithm and the fractional-sampling scheme: PDC $= 0.5$, with $f_0 \sim \mathcal{N}(0,1)$ and $f_1 \sim \mathcal{N}(0.75, 1)$.

Thus, setting $A = A_\alpha = \log \frac{1}{\alpha}$ ensures that

$$\mathsf{FAR}(\Psi_w(A_\alpha, h, \mu)) \leq \alpha.$$

2. *There exists values of μ^* and h^* for which*

$$\mathsf{PDC}(\Psi_w(A, h^*, \mu^*)) \leq \beta \quad \forall A \geq 0,$$

i.e., the PDC *constraint is satisfied independent of the choice of A.*

3. *Also, for each fixed β, as $\alpha \to 0$,*

$$\mathsf{CADD}(\Psi_w(A_\alpha, h^*, \mu^*)) \leq \frac{|\log \alpha|}{D(f_1 \parallel f_0)}(1 + o(1)). \qquad (3.13)$$

And if $h < \infty$, then for each fixed β, as $\alpha \to 0$,

$$\mathsf{WADD}(\Psi_w(A_\alpha, h^*, \mu^*)) \leq \frac{|\log \alpha|}{D(f_1 \parallel f_0)}(1 + o(1)). \qquad (3.14)$$

Since, $\frac{|\log \alpha|}{D(f_1 \parallel f_0)}$ is the best possible asymptotic delay for an FAR constraint of α (see Lai (1998)), the $\mathrm{DE - CuSum}$ algorithm is asymptotically optimal for both Problem 3.2 and Problem 3.3.

In Fig. 3.4, we show that the DE-CuSum algorithm provides a significant gain in performance as compared to the approach of fractional sampling, where the

CuSum algorithm is used and the PDC constraint is met by skipping observations randomly. Fig. 3.4 shows the comparative CADD vs FAR performances of the CuSum algorithm, the DE-CuSum algorithm, and the fractional sampling scheme, for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75, 1)$ and PDC $= 0.5$. As seen in the figure, even for a PDC of 0.5, i.e., even after dropping 50% of the observations before change, there is a small difference in delays of the CuSum and the DE-CuSum algorithm. On the other hand, the difference in performance as compared to the fractional sampling scheme is significant.

4. Data-Efficient Quickest Change Detection in Sensor Networks

In many engineering applications, a sensor network is often deployed to detect a change in the statistical properties of a process or phenomenon under observation.

In the literature, the QCD problem is also studied in the sensor network framework. In the sensor network model we are interested in, there are L sensors and a central decision maker called the fusion center. The sensors are indexed by the index $\ell \in \{1, \cdots, L\}$. In the following we say sensor ℓ to refer to the sensor indexed by ℓ. At sensor ℓ the sequence $\{X_{n,\ell}\}_{n \geq 1}$ is observed, where n is the time index. At some unknown time γ, the distribution of $\{X_{n,\ell}\}$ in a subset

$$\kappa = \{k_1, k_2, \cdots, k_m\} \subset \{1, 2, \cdots, L\}$$

of the streams changes, from $f_{0,\ell}$ to say $f_{1,\ell}$. The observations $\{X_{n,\ell}\}$ are independent across indices n and ℓ, conditioned on γ. The distributions $f_{0,\ell}$ and $f_{1,\ell}$ are known to the decision maker but the affected subset $\{k_1, k_2, \cdots, k_m\}$ may or may not be known. A processed version of the observations is transmitted from each sensor to the fusion center. At each time step, the information transmitted from the sensors to the fusion center could be the raw observations themselves, a quantized version of the observations, or a carefully designed summary message representing all the information available till that time at the sensors. The objective here is to find an efficient way to process and transmit information to the fusion center, and to find a stopping time on the information received at the fusion center to detect the change in distribution as quickly as possible.

The sensor network problem has been studied in both Bayesian and minimax settings in the literature; see Veeravalli (2001), Tartakovsky and Veeravalli (2002), Mei (2005), Tartakovsky and Veeravalli (2008), Zacharias and Sun-

daresan (2007), and Banerjee et al. (2011). However, in this section, we restrict our discussion to the more practically relevant minimax settings. We now describe three well known algorithms in the literature.

1. CuSum-All: This algorithm can be used when the affected subset κ is known. In this case the problem is same as the case when the change affects all the sensors. For this setup, the CuSum-All algorithm is proposed in Mei (2005), and is shown to achieve the lower bound in (2.17). In this algorithm, the CuSum algorithm (2.18) is used at each sensor. A '1' is transmitted each time the CuSum statistics is above a threshold, and a change is declared at the fusion center the first time a '1' is received from all the sensors at the same time.

2. CuSum-Censor-Max: This algorithm can be used when the affected κ is not known, but it is known that $|\kappa| = 1$. Thus, only one of the sensors is affected post-change. In the CuSum-Censor-Max algorithm, the CuSum algorithm is used at each sensor. The CuSum statistics is transmitted from a sensor to the fusion center only if the statistic is larger than a threshold, and a change is declared when the maximum of the transmitted statistics from all the sensors crosses another threshold. This algorithm is proposed in Tartakovsky and Veeravalli (2002) (without the censoring part). It is shown that in Tartakovsky and Veeravalli (2002) that this algorithm is uniformly asymptotically optimal, i.e., it achieves the lower bound in (2.17) for each possible post-change scenario.

3. CuSum-Censor-Sum: For the case when the affected subset κ or even its size is completely unknown, the CuSum-Censor-Sum algorithm is proposed in Mei (2011) (also see Mei (2010)). In this algorithm also the CuSum algorithm is used at each sensor, and the CuSum statistics is transmitted from a sensor to the fusion center only if the statistic is larger than a threshold. A change is declared when the sum of the transmitted statistics from all the sensors crosses another threshold. It is claimed in Mei (2011) that the CuSum-Censor-Sum algorithm is uniformly asymptotically optimal.

The works in Banerjee et al. (2011), Mei (2005), Tartakovsky and Veeravalli (2002), Tartakovsky and Veeravalli (2008), Zacharias and Sundaresan (2007) and Veeravalli (2001), have only focussed on extending the classical problem formulations by putting a constraint on the cost of communication between the

sensors and the fusion center. But the cost of acquiring observations at the sensors itself has not been taken into account. This cost can be controlled by replacing the CuSum algorithm used at the sensors by the DE-CuSum algorithm. This gives us the following algorithms: DE-All, DE-Censor-Max, and DE-Censor-Sum, in which the DE-CuSum algorithm is used in place of the CuSum algorithm in CuSum-All, CuSum-Censor-Max, and CuSum-Censor-Sum, respectively.

Based on the properties of the DE-CuSum algorithm, it can be shown that the above data-efficient algorithms are also asymptotically optimal, under appropriate conditions, i.e, their performance is asymptotically equal to the lower bound of (2.17) (with $D(f_1||f_0)$ replaced by appropriate K-L distance between pre- and post-change distributions). See Banerjee and Veeravalli (2012b), Banerjee and Veeravalli (2013b), and Banerjee and Veeravalli (2013c) for the detailed problem formulations and numerical results. In summary, the DE-All algorithm can be used when there is a severe constraint on amount of information that can be acquired and transmitted and the number of affected subset is known. If the number of affected subset is not known, but is known to be small as compared to L, then the max schemes performs better than the sum scheme, and vice versa.

For the case when the change affects all the sensors, more efficient algorithms have been proposed in Banerjee and Veeravalli (2012b).

1. DE-Dist: When the communication constraint is not so severe, multiple bits approximating the entire statistic can be transmitted from the sensors to the fusion center, instead of transmitting only one bit. In the DE-Dist algorithm, the CuSum algorithm is used at each sensor. Whenever an observation is taken at a sensor, it is transmitted from a sensor to the fusion center. At the fusion center a CuSum is applied to the received observations to detect the change. It is shown in Banerjee and Veeravalli (2013c) that the DE-Dist algorithm performs considerably better than the DE-All algorithm.

2. Serialized DE-CuSum: This is a centralized control based algorithm. In this algorithm, the observations from the sensors are serialized, and then the DE-CuSum algorithm is applied to the serialized observation sequence. This algorithm can be used as a benchmark for data-efficient algorithms in sensor networks.

5. Conclusions

We provided a survey of data-efficient quickest change detection. The DE-QCD problem is the classical QCD problem with an additional constraint on the cost of observations used before the change point. The results surveyed here show that the likelihood ratio of the observations can be utilized to modify the classical QCD algorithms to introduce on-off observation control. A large likelihood ratio shows that the evidence is large of the change having already happened, and a change can be declared. On the other hand, a small likelihood ratio indicates that most likely the change has not happened, and few observations can be dropped. We showed how this intuition is used to modify the single-threshold based Shiryaev algorithm and the CuSum algorithm, to obtain the two-threshold based DE-Shiryaev algorithm and the DE-CuSum algorithm, respectively. We also discussed the optimality properties of these data-efficient algorithms. We then provided a short discussion on how this technique can be used to obtain data-efficient algorithms for sensor networks. Insights obtained from these results can be used to obtain data-efficient algorithms for more general QCD models.

Acknowledgements

References

1. Banerjee, T., Sharma, V., Kavitha, V., and Jayaprakasam, A. (2011). Generalized Analysis of a Distributed Energy Efficient Algorithm for Change Detection, *IEEE Transactions on Wireless Communication* 10: 91–101. DOI:10.1109/TWC.2010.110510.091177

2. Banerjee, T. and Veeravalli, V. V. (2012a). Data-Efficient Quickest Change Detection with On-Off Observation Control, *Sequential Analysis* 31: 4077. DOI:10.1080/07474946.2012.651981

3. Banerjee, T. and Veeravalli, V. V. (2012b). Energy-Efficient Quickest Change Detection in Sensor Networks, *in Proceedings of IEEE Statistical Signal Processing Workshop (SSP)*, AUG 5–8, Ann Arbor, MI, USA. DOI:10.1109/SSP.2012.6319781

4. [4] Banerjee, T. and Veeravalli, V. V. (2013a). Data-Efficient Quickest Change Detection in Minimax Settings, *IEEE Transactions on Information Theory* 59: 69176931. DOI:10.1109/TIT.2013.2272313

5. Banerjee, T. and Veeravalli, V. V. (2013b). Data-efficient Quickest Change Detection in Distributed and Multichannel Systems, *in Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* , MAY 26–31, Vancouver, Canada. DOI:10.1109/ICASSP.2013.6638400

6. Banerjee, T. and Veeravalli, V. V. (2013c). Decentralized Data-Efficient Quickest Change Detection, *in Proceedings of IEEE International Symposium on Information Theory (ISIT)*, JUL 7–12, Istanbul, Turkey. DOI:10.1109/ISIT.2013.6620694

7. Lai, T. L. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917-2929. DOI:10.1109/18.737522

8. Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 18971908. DOI:10.1214/aoms/1177693055

9. Mei, Y. (2005). Information Bounds and Quickest Change Detection in Decentralized Decision Systems, *IEEE Transactions on Information Theory* 51: 2669–2681. DOI:10.1109/TIT.2005.850159

10. Mei, Y. (2010). Efficient Scalable Schemes for Monitoring a Large Number of Data Streams, *Biometrika* 97: 419–433. DOI: 10.1093/biomet/asq010

11. Mei, Y. (2011). Quickest Detection in Censoring Sensor networks, *in Proceedings of IEEE International Symposium on Information Theory (ISIT)*, JUL 31–AUG 5, Saint-Petersburg, Russia. DOI:10.1109/ISIT.2011.6034390

12. Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14: 13791387. DOI:10.1214/aos/1176350164

13. Page, E. S. (1954). Optimal Detection of a Change in Distribution, *Biometrika* 41: 100115. DOI:10.1214/aos/1176346587

14. Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206227. DOI:10.1214/aos/1176346587

15. Poor, H. V. and Hadjiliadis, O. (2009). *Quickest Detection*, Cambridge University Press.

16. Premkumar, K. and Kumar, A. (2008). Optimal Sleep/Wake Scheduling for Quickest Intrusion Detection using Sensor Networks, *in Proceedings of IEEE Conference on Computer Communications (INFOCOM)*, APR 15–17, Phoenix, AZ, USA. DOI:10.1109/INFOCOM.2008.198

17. Shiryaev, A. N. (1963). On Optimal Methods in Quickest Detection Problems, *Theory of Probability and Its Applications* 8: 22–46. DOI:10.1137/1108002

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

19. Tartakovsky, A. G., Nikiforov, I. V., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Change-Point Detection*, Statistics, CRC Press.

20. Tartakovsky, A. G. and Veeravalli, V. V. (2002). An Efficient Sequential Procedure for Detecting Changes in Multichannel and Distributed systems, *in Proceedings of IEEE International Conference on Information Fusion (ICIF)*, JUL 8–11, Annapolis, MD, USA. DOI:10.1109/ICIF.2002.1021129

21. 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. DOI:10.1137/S0040585X97981202

22. Tartakovsky, A. G. and Veeravalli, V. V. (2008). Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems, *Sequential Analysis* 27: 441–475. doi:10.1080/07474940802446236

23. Veeravalli, V. V. (2001). Decentralized Quickest Change Detection, *IEEE Transactions on Information Theory* 47: 1657–1665. DOI:10.1109/18.923755

24. Veeravalli, V. V. and Banerjee, T. (2013). *Quickest Change Detection*, Elsevier: E-reference Signal Processing http://arxiv.org/abs/1210.5552.

25. Wald, A. and Wolfowitz, J. (1948). Optimum Character of the Sequential Probability Ratio Test, *Annals of Mathematical Statistics* 19: 326339. DOI:10.1214/aoms/1177730197

26. Zacharias, L. and Sundaresan, R. (2007). Decentralized Sequential Change Detection using Physical Layer Fusion, *in Proceedings of IEEE International Symposium on Information Theory (ISIT)*, June 24–29, Nice, France. DOI:10.1109/T-WC.2008.070808