Data-Efficient Quickest Change Detection in Minimax Settings

Taposh Banerjee, *Student Member, IEEE*, and Venugopal V. Veeravalli, *Fellow, IEEE*

***Abstract*—The classical problem of quickest change detection is studied with an additional constraint on the cost of observations used in the detection process. The change point is modeled as an unknown constant, and minimax formulations are proposed for the problem. The objective in these formulations is to find a stopping time and an ON–OFF observation control policy for the observation sequence, to minimize a version of the worst possible average delay, subject to constraints on the false alarm rate and the fraction of time observations are taken before change. An algorithm called DE-CuSum is proposed and is shown to be asymptotically optimal for the proposed formulations, as the false alarm rate goes to zero. Numerical results are used to show that the DE-CuSum algorithm has good tradeoff curves and performs significantly better than the approach of fractional sampling, in which the observations are skipped using the outcome of a sequence of coin tosses, independent of the observation process. This study is guided by the insights gained from an earlier study of a Bayesian version of this problem.**

***Index Terms*—Asymptotic optimality, CuSum, minimax, observation control, quickest change detection.**

I. Introduction

In the problem of quickest change detection, a decision maker observes a sequence of random variables $\{X_n\}$. At some point of time γ, called the change point, the distribution of the random variables changes. The goal of the decision maker is to find a stopping time τ on the $\{X_n\}$, so as to minimize the average value of the delay $\max\{0, \tau - \gamma\}$. The delay is zero on the event $\{\tau < \gamma\}$, but this event is treated as a false alarm and is not desirable. Thus, the average delay has to be minimized subject to a constraint on the false alarm rate. This problem finds application in statistical quality control in industrial processes, surveillance using sensor networks and cognitive radio networks; see [1]–[3].

In the independent and identically distributed (i.i.d.) model of the quickest change detection problem, the random variables $\{X_n\}$ for $n < \gamma$ are i.i.d. with probability density function (p.d.f.) f_0, and $\{X_n\}$ for $n \geq \gamma$ are i.i.d. with p.d.f. f_1. In the Bayesian version of the quickest change detection problem, the change point γ is modeled as a random variable Γ.

In [4] and [5], the i.i.d. model is studied in a Bayesian setting by assuming the change point Γ to be a geometrically distributed random variable. The objective is to minimize the average detection delay with a constraint on the probability of false alarm. It is shown that under very general conditions on f_0 and f_1, the optimal stopping time is the one that stops the first time the *a posteriori* probability $\mathbb{P}(\Gamma \leq n | X_1, \ldots, X_n)$ crosses a predesigned threshold. The threshold is chosen to meet the false alarm constraint with equality. In the following, we refer to this algorithm as the Shiryaev algorithm.

In [6]–[11], no prior knowledge about the distribution on the change point is assumed, and the change point is modeled as an unknown constant. In this non-Bayesian setting, the quickest change detection problem is studied in two different minimax settings introduced in [6] and [7]. The objective in [6]–[11] is to minimize some version of the worst case average delay, subject to a constraint on the mean time to false alarm. The results from these papers show that, variants of the Shiryaev–Roberts algorithm [12], the latter being derived from the Shiryaev algorithm by setting the geometric parameter to zero, and the CuSum algorithm [13], are asymptotically optimal for both the minimax formulations, as the mean time to false alarm goes to infinity.

In many applications of quickest change detection, changes are infrequent and there is a cost associated with acquiring observations (data). As a result, it is of interest to study the classical quickest change detection problem with an additional constraint on the cost of observations used before the change point, with the cost of taking observations after the change point being penalized through the metric on delay. In the following, we refer to this problem as data-efficient quickest change detection. In [14], we studied data-efficient quickest change detection in a Bayesian setting by adding another constraint to the Bayesian formulation of [4]. The objective was to find a stopping time and an ON–OFF observation control policy on the observation sequence, to minimize the average detection delay subject to constraints on the probability of false alarm and the average number of observations used before the change point. Thus, unlike the classical quickest change detection problem, where the decision maker only chooses one of the two controls, to stop and declare change or to continue taking observations, in the data-efficient quickest change detection problem we considered in [14], the decision maker must also decide—when the decision is to *continue*—whether to *take* or *skip* the next observation.

Manuscript received November 14, 2012; revised June 18, 2013; accepted June 21, 2013. Date of publication July 03, 2013; date of current version September 11, 2013. This work was supported in part by the National Science Foundation under Grants CCF 08-30169 and DMS 12-22498, through the University of Illinois at Urbana-Champaign, and in part by the U.S. Defense Threat Reduction Agency through subcontract 147755 at the University of Illinois from Prime Award HDTRA1-10-1-0086. This paper was presented at the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing.

The authors are with the Department of Electrical Communication Engineering and Coordinated Science Laboratory, University of Illinois Urbana Champaign, Urbana, IL 61801 USA (e-mail: banerje5@illinois.edu; vvv@illinois.edu).

Communicated by G. V. Moustakides, Associate Editor for Detection and Estimation.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2013.2272313

0018-9448 © 2013 IEEE

For the i.i.d. model, and for geometrically distributed Γ, we showed in [14] that a two-threshold algorithm is asymptotically optimal, as the probability of false alarm goes to zero. This two-threshold algorithm, that we call the DE-Shiryaev algorithm in the following, is a generalized version of the Shiryaev algorithm from [4]. In the DE-Shiryaev algorithm, the *a posteriori* probability that the change has already happened conditioned on available information, is computed at each time step, and the change is declared the first time this probability crosses a threshold A. When the *a posteriori* probability is below this threshold A, observations are taken only when this probability is above another threshold $B < A$. When an observation is skipped, the *a posteriori* probability is updated using the prior on the change point random variable. We also showed that, for reasonable values of the false alarm constraint and the observation cost constraint, these two thresholds can be selected independent of each other: the upper threshold A can be selected directly from the false alarm constraint and the lower threshold B can be selected directly from the observation cost constraint. Finally, we showed that the DE-Shiryaev algorithm achieves a significant gain in performance over the approach of *fractional sampling*, where the Shiryaev algorithm is used and an observation is skipped based on the outcome of a coin toss.

In this paper, we study the data-efficient quickest change detection problem in a non-Bayesian setting, by introducing an additional constraint on the cost of observations used in the detection process, in the minimax settings of [6] and [7]. We first use the insights from the Bayesian analysis in [14] to propose a metric for data efficiency in the absence of knowledge of the distribution on the change point. This metric is the fraction of time samples taken before change. We then propose extensions of the minimax formulations in [6] and [7] by introducing an additional constraint on data efficiency in these formulations. Thus, the objective is to find a stopping time and an ON–OFF observation control policy to minimize a version of the worst case average delay, subject to constraints on the mean time to false alarm and the fraction of time observations are taken before change. Then, motivated by the structure of the DE-Shiryaev algorithm, we propose an extension of the CuSum algorithm from [13]. We call this extension the DE-CuSum algorithm. We show that the DE-CuSum algorithm inherits the good properties of the DE-Shiryaev algorithm. That is, the DE-CuSum algorithm is asymptotically optimal, is easy to design, and provides substantial performance improvements over the approach of fractional sampling, where the CuSum algorithm is used and observations are skipped based on the outcome of a sequence of coin tosses, independent of the observations process.

The problem of detecting an anomaly in the behavior of an industrial process, under cost considerations, is also considered in the literature of statistical process control. There, it is studied under the heading of sampling rate control or sampling size control; see [15] and [16] for a detailed survey, and the references in [14] for some recent results. However, none of these references study the data-efficient quickest change detection problem under the classical quickest change detection setting, as done by us in [14] and in this paper. For a result similar to our work in [14] in a Bayesian setting, see [17]. See [18] and [19] for other interesting formulations of quickest change detection with observation control.

Since our work in this paper on data-efficient non-Bayesian quickest change detection is motivated by our work on data-efficient Bayesian quickest change detection in [14], in Section II, we provide a detailed overview of the results from [14]. We also comment on the insights provided by the Bayesian analysis, which we use in the development of a theory for the non-Bayesian setting. In Sections III–V, we give details of the minimax formulations, a description of the DE-CuSum algorithm and the analysis of the DE-CuSum algorithm, respectively. In Section VI, we provide some numerical results .

Table I provides a glossary of the terms used in the paper.

TABLE I
GLOSSARY

Symbol	**Definition/Interpretation**
$o(1)$	$x = o(1)$ as $c \to c_0$, if $\forall \epsilon > 0$, $\exists \delta > 0$ s.t., $\lvert x\rvert \leq \epsilon$ if $\lvert c - c_0\rvert < \delta$
$O(1)$	$x = O(1)$ as $c \to c_0$, if $\exists \epsilon > 0, \delta > 0$ s.t., $\lvert x\rvert \leq \epsilon$ if $\lvert c - c_0\rvert < \delta$
$g(c) \sim h(c)$ as $c \to c_0$	$\lim_{c \to c_0} \frac{g(c)}{h(c)} = 1$ or $g(c) = h(c)(1 + o(1))$ as $c \to c_0$
$\mathbb{P}_n$ ($\mathbb{E}_n$)	Probability measure (expectation) when the change occurs at time n
$\mathbb{P}_\infty$ ($\mathbb{E}_\infty$)	Probability measure (expectation) when the change does not occur
ess sup X	$\inf\{K \in \mathbb{R} : \mathbb{P}(X > K) = 0\}$
$L(X)$	$\frac{f_1(X)}{f_0(X)}$
$\ell(X)$	$\log \frac{f_1(X)}{f_0(X)}$
$D(f_1 \parallel f_0)$	K-L Divergence between f_1 and f_0, defined as $\mathbb{E}_1\left(\log \frac{f_1(X)}{f_0(X)}\right)$
$D(f_0 \parallel f_1)$	K-L Divergence between f_0 and f_1, defined as $\mathbb{E}_\infty\left(\log \frac{f_1(X)}{f_0(X)}\right)$
$(x)^+$	$\max\{x, 0\}$
$(x)^{h+}$	$\max\{x, -h\}$
$T(x)$	$\lceil (x)^{h+}/\mu \rceil$
M_n	$M_n = 1$ if X_n is used for decision making
Ψ	Policy for data-efficient quickest change detection $\{\tau, M_1, \cdots, M_\tau\}$
ADD(Ψ)	$\sum_{n=0}^{\infty} \mathbb{P}(\Gamma = n)\ \mathbb{E}_n\left[(\tau - n)^+\right]$
PFA(Ψ)	$\sum_{n=0}^{\infty} \mathbb{P}(\Gamma = n)\ \mathbb{P}_n(\tau < n)$
FAR(Ψ)	$\frac{1}{\mathbb{E}_\infty[\tau]}$
WADD(Ψ)	$\sup_{n \geq 1}$ ess sup $\mathbb{E}_n\left[(\tau - n)^+ \lvert I_{n-1}\right]$
CADD(Ψ)	$\sup_{n \geq 1} \mathbb{E}_n[\tau - n \lvert \tau \geq n]$
PDC(Ψ)	$\limsup_n \frac{1}{n}\mathbb{E}_n\left[\sum_{k=1}^{n-1} M_k \middle\lvert \tau \geq n\right]$
$\mathbb{I}_A$	Indicator function for event A
$\mathbb{E}[X; A]$	$\mathbb{E}[X\mathbb{I}_A]$

II. Data-Efficient Bayesian Quickest Change Detection

In this section, we review the Bayesian version of the data-efficient quickest change detection we studied in [14]. 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 further assume that Γ is geometrically distributed with parameter ρ:

$$\mathbb{P}(\Gamma = n) = (1-\rho)^{n-1}\rho, \quad 0 < \rho < 1.$$

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. Here, $\mathbb{I}_F$ represents the indicator of the event F. Thus, a policy for data-efficient quickest change detection is $\Psi = \{\tau, \phi_0, \ldots, \phi_{\tau-1}\}$.

Define the average detection delay

$$\mathsf{ADD}(\Psi) \triangleq \mathbb{E}\left[(\tau - \Gamma)^+\right],$$

the probability of false alarm

$$\mathsf{PFA}(\Psi) \triangleq \mathbb{P}(\tau < \Gamma),$$

and the metric for data-efficiency in the Bayesian setting we considered in [14], the average number of observations used before the change point,

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

The objective in [14] is to solve the following optimization problem:

Problem 1:

$$\begin{aligned} \underset{\Psi}{\text{minimize}} \quad & \mathsf{ADD}(\Psi), \\ \text{subject to} \quad & \mathsf{PFA}(\Psi) \leq \alpha, \\ \text{and} \quad & \mathsf{ANO}(\Psi) \leq \zeta, \end{aligned} \tag{1}$$

where, α and ζ are given constraints.

When $\zeta \geq \mathbb{E}[\Gamma] - 1$, Problem 1 reduces to the classical Bayesian quickest change detection problem considered by Shiryaev in [4].

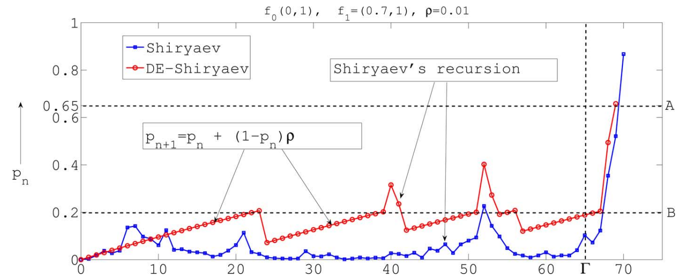

Fig. 1. Typical evolution of the Shiryaev and the DE-Shiryaev algorithms applied to the same set of samples. Parameters used: $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.7,1)$, and $\rho = 0.01$, with thresholds $A = 0.65$ and $B = 0.2$.

A. DE-Shiryaev Algorithm

Define

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

Then, the two-threshold algorithm from [14] is as follows.

Algorithm 1 (DE-Shiryaev: $\Psi(A,B)$): Start with $p_0 = 0$ and use the following control, with $B < A$, for $n \geq 0$:

$$\begin{aligned} M_{n+1} = \phi_n(p_n) &= \begin{cases} 0, & \text{if } p_n < B \\ 1, & \text{if } p_n \geq B \end{cases} \\ \tau_{\mathrm{D}} &= \inf\left\{n \geq 1 : p_n > A\right\}. \end{aligned} \tag{2}$$

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 $L(X_{n+1}) = f_1(X_{n+1})/f_0(X_{n+1})$.

With $B = 0$, the DE-Shiryaev algorithm reduces to the Shiryaev algorithm from [4].

The motivation for this algorithm comes from the fact that p_n is a sufficient statistics for a Lagrangian relaxation of Problem 1. This relaxed problem can be studied using dynamic programming, and numerical studies of the resulting Bellman equation show that the DE-Shiryaev algorithm is optimal for a wide choice of system parameters. For an analytical justification, see Section II-B.

When Algorithm 1 is employed, the probability p_n typically evolves as depicted in Fig. 1. As observed in Fig. 1, the evolution starts with an initial value of $p_0 = 0$. This is because we have implicitly assumed that the probability that the change has already happened even before we start taking observations is zero. Also note that when $p_n < B$, p_n increases monotonically. This is because when an observation is skipped, p_n is updated using the prior on the change point, and as a result, the probability that the change has already happened increases monotonically. The change is declared at time τ_{D}, the first time p_n crosses the threshold A.

B. Asymptotic Optimality and Tradeoff Curves

We note that

$$\mathsf{PFA}(\Psi(A,B)) = \mathbb{E}[1 - p_{\tau_{\mathrm{D}}}] \leq 1 - A.$$

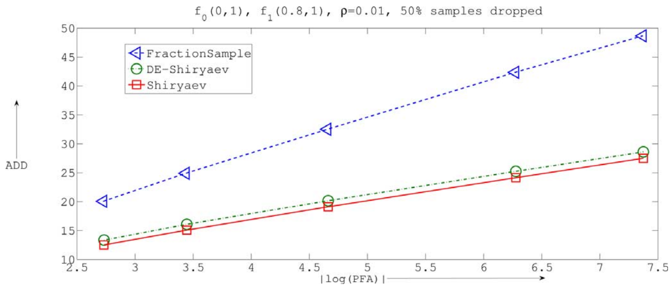

Fig. 2. Comparative performance of schemes for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.8,1)$, and $\rho = 0.01$.

Thus, with $A = 1 - \alpha$,

$$\mathsf{PFA}(\Psi(A,B)) \leq \alpha.$$

It is shown in [14] that the PFA and ADD of the DE-Shiryaev algorithm approach that of the Shiryaev algorithm as $\alpha \to 0$. Specifically, the following theorem is proved.

Theorem 2.1 (see [14]): If

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

and $\ell(X)$ is nonarithmetic (see [20]), then for a fixed ζ, the threshold B can be selected such that for every $A > B$,

$$\mathsf{ANO}(\Psi(A,B)) \leq \zeta,$$

and with B fixed to this value and $A = A_\alpha = 1 - \alpha$,

$$\mathsf{ADD}(\Psi(A_\alpha,B)) \sim \frac{|\log(\alpha)|}{D(f_1 \parallel f_0) + |\log(1-\rho)|} \text{ as } \alpha \to 0, \tag{3}$$

and

$$\mathsf{PFA}(\Psi(A_\alpha,B)) \sim \alpha \left(\int_0^\infty e^{-x} dR(x) \right) \text{ as } \alpha \to 0. \tag{4}$$

Here, $R(x)$ is the asymptotic overshoot distribution of the random walk $\sum_{k=1}^n (\ell(X_k) + |\log(1-\rho)|)$, when it crosses a large positive boundary under f_1. Since (3) and (4) are also the performance of the Shiryaev algorithm as $\alpha \to 0$ [5], the DE-Shiryaev algorithm is asymptotically optimal.

Equation (4) shows that PFA is not a function of the threshold B. In [14], 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.

Although (3) is true for each fixed value of ζ, as ζ becomes smaller, a smaller value of α is needed before the asymptotics "kick in."

Fig. 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 samples are skipped by tossing a biased coin (with probability of success 50/99), without looking at the state of the system. When a sample 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.

More accurate estimates of the delay and that of ANO are available in [14].

C. Insights From the Bayesian Setting

We make the following observations on the evolution of the statistic p_n in Fig. 1.

1) Let

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

Then, after $t(B)$, the number of samples skipped when p_n goes below B is a function of the undershoot of p_n and the geometric parameter ρ. If $L^*(X_n)$ is defined as

$$L^*(X_n) = \begin{cases} L(X_n), & \text{if } M_n = 1 \\ 1, & \text{if } M_n = 0. \end{cases}$$

Then, $\frac{p_n}{1-p_n}$ can be shown to be equal to

$$\frac{p_n}{1-p_n} = \frac{\sum_{k=1}^n (1-\rho)^{k-1} \rho \prod_{i=k}^n L^*(X_i)}{\mathbb{P}(\Gamma > n)}.$$

Thus, $\frac{p_n}{1-p_n}$ is the average likelihood ratio of all the observations taken till time n, and since there is a one-to-one mapping between p_n and $\frac{p_n}{1-p_n}$, we see that the number of samples skipped is a function of the likelihood ratio of the observations taken.

2) When 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 "$H_0 = \text{prechange}$," and "$H_1 = \text{postchange}$." If the decision in the two-sided test is H_0, then samples are skipped depending on the likelihood ratio of the observations, and the two-sided test is repeated on the samples beyond the skipped samples. The change is declared the first time the decision in a two-sided test is H_1.

3) Because of the above interpretation of the evolution of the DE-Shiryaev algorithm as a sequence of roughly independent two-sided tests, we see that the constraint on the observation cost is met by delaying the measurement process on the basis of the prior statistical knowledge of the change point, and then beyond $t(B)$, controlling the fraction of time p_n is above B, i.e., controlling the fraction of time samples are taken.

These insights will be crucial to the development of a theory for data-efficient quickest change detection in the non-Bayesian setting.

III. Data-Efficient Minimax Formulation

In the absence of a prior knowledge on the distribution of the change point, as is standard in the classical quickest change detection literature, we model the change point as an unknown constant γ. As a result, the quantities ADD, PFA, ANO in

Problem 1 are not well defined. Thus, we study the data-efficient quickest change detection problem in minimax settings. In this paper, we consider two most popular minimax formulations: one is due to Pollak [7] and another is due to Lorden [6].

We will use the insights from the Bayesian setting of Section II to study data-efficient minimax quickest change detection. Our development will essentially follow the layout of the Bayesian setting. Specifically, we first propose two minimax formulations for data-efficient quickest change detection. Motivated by the structure of the DE-Shiryaev algorithm, we then propose an algorithm for data-efficient quickest change detection in the minimax settings. This algorithm is a generalized version of the CuSum algorithm [13]. We call this algorithm the DE-CuSum algorithm. We show that the DE-CuSum algorithm is asymptotically optimal under both minimax settings. We also show that in the DE-CuSum algorithm, the constraints on false alarm and observation cost can be met independent of each other. Finally, we show that we can achieve a substantial gain in performance by using the DE-CuSum algorithm as compared to the approach of fractional sampling.

We first propose a metric for data-efficiency in a non-Bayesian setting. In Section II-C, we saw that in the DE-Shiryaev algorithm, observation cost constraint is met using an initial wait, and by controlling the fraction of time observations are taken, after the initial wait. In the absence of prior statistical knowledge on the change point, such an initial wait cannot be justified. This motivates us to seek control policies that can meet a constraint on the fraction of time observations are taken before change. With M_n, I_n, τ, and Ψ as defined earlier in Section II, we propose the following duty cycle based observation cost metric, Prechange Duty Cycle (PDC):

$$\begin{aligned}\mathsf{PDC}(\Psi) &= \limsup_n \frac{1}{n}\mathbb{E}_n\left[\sum_{k=1}^{n-1} M_k \Big| \tau \geq n\right]\\ &= \limsup_n \frac{1}{n}\mathbb{E}_\infty\left[\sum_{k=1}^{n-1} M_k \Big| \tau \geq n\right]. \end{aligned} \tag{5}$$

Clearly, $\mathsf{PDC} \leq 1$.

We now discuss why we use $\limsup$ rather than $\sup$ in defining PDC. In all reasonable policies Ψ, M_1 will typically be set to 1. As mentioned earlier, this is because an initial wait cannot be justified without a prior statistical knowledge of the change point. As a result, in (5), we cannot replace the $\limsup$ by $\sup$, because the latter would give us a PDC value of 1. Even otherwise, without any prior knowledge on the change point, it is reasonable to assume that the value of γ is large corresponding to a rare change, and hence, the PDC metric defined in (5) is a reasonable metric for our problem.

If in a policy all the observations are used for decision making, then the PDC for that policy is 1. If every alternate observation is used, then the $\mathsf{PDC} = 0.5$.

For false alarm, we consider the metric used in [6] and [7], the mean time to false alarm or its reciprocal, the false alarm rate:

$$\mathsf{FAR}(\Psi) = \frac{1}{\mathbb{E}_\infty[\tau]}. \tag{6}$$

For delay, we consider two possibilities: the minimax setting of Pollak [7] where the delay metric is the following supremum over time of the conditional delay[1]

$$\mathsf{CADD}(\Psi) = \sup_n \ \mathbb{E}_n[\tau - n | \tau \geq n], \tag{7}$$

or the minimax setting of Lorden [6], where the delay metric is the supremum over time of the essential supremum of the conditional delay[2]

$$\mathsf{WADD}(\Psi) = \sup_n \operatorname{ess\,sup} \ \mathbb{E}_n\left[(\tau - n)^+ | I_{n-1}\right]. \tag{8}$$

Note that unlike the delay metric in [6], WADD in (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.

Since, $\{\tau \geq n\}$ belongs to the sigma algebra generated by I_{n-1}, we have

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

Our first minimax formulation is the following data-efficient extension of Pollak [7].

Problem 2:

$$\begin{aligned} \underset{\Psi}{\text{minimize}} \quad & \mathsf{CADD}(\Psi),\\ \text{subject to} \quad & \mathsf{FAR}(\Psi) \leq \alpha,\\ \text{and} \quad & \mathsf{PDC}(\Psi) \leq \beta, \end{aligned} \tag{9}$$

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

We are also interested in the data-efficient extension of the minimax formulation of Lorden [6].

Problem 3:

$$\begin{aligned} \underset{\Psi}{\text{minimize}} \quad & \mathsf{WADD}(\Psi),\\ \text{subject to} \quad & \mathsf{FAR}(\Psi) \leq \alpha,\\ \text{and} \quad & \mathsf{PDC}(\Psi) \leq \beta, \end{aligned} \tag{10}$$

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

With $\beta = 1$, Problem 2 reduces to the minimax formulation of Pollak in [7], and Problem 3 reduces to the minimax formulation of Lorden in [6].

In [13], the following algorithm called the CuSum algorithm is proposed:

Algorithm 2 (CuSum: Ψ_{C}): Start with $C_0 = 0$, and update the statistic C_n as

$$C_{n+1} = \left(C_n + \ell(X_{n+1})\right)^+,$$

where $(x)^+ = \max\{0, x\}$ and $\ell(X) = \log \frac{f_1(X)}{f_0(X)}$. Stop at

$$\tau_{\mathrm{C}} = \inf\{n \geq 1 : C_n > A\}.$$

It is shown by Lai in [10] that the CuSum algorithm is asymptotically optimal for both Problems 2 and 3, with $\beta = 1$, as $\alpha \to 0$ (see Section V-B for a precise statement).

[1]We are only interested in those policies for which the CADD is well defined.

[2]The delay metric considered in [6] and [8] is actually $\sup_n \operatorname{ess\,sup} \ \mathbb{E}_n[(\tau - n + 1)^+ | I_{n-1}]$. However, these two metrics are equivalent as the WADD goes to infinity.

In the following, we propose the DE-CuSum algorithm, an extension of the CuSum algorithm for the data-efficient setting, and show that it is asymptotically optimal, for each fixed β, as $\alpha \to 0$; see Section V-E.

IV. DE-CuSum Algorithm

We now present the DE-CuSum algorithm.

Algorithm 3 (DE-CuSum: $\Psi_{\rm DC}(A, \mu, h)$): Start with $D_0 = 0$ and fix $\mu > 0$, $A > 0$ and $h \geq 0$. For $n \geq 0$, use the following control:

$$M_{n+1} = \begin{cases} 0, & \text{if } D_n < 0 \\ 1, & \text{if } D_n \geq 0 \end{cases}$$
$$\tau_{\rm DC} = \inf\{n \geq 1 : D_n > A\}.$$

The statistic D_n is updated using the following recursions:

$$D_{n+1} = \begin{cases} \min\{D_n + \mu, 0\}, & \text{if } M_{n+1} = 0 \\ (D_n + \ell(X_{n+1}))^{h+}, & \text{if } M_{n+1} = 1 \end{cases}$$

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

When $h = \infty$, the DE-CuSum algorithm works as follows. The statistic D_n starts at 0 and evolves according to the CuSum algorithm till it goes below 0. When D_n goes below 0, it does so with an undershoot. Beyond this, D_n is incremented deterministically (by using the recursion $D_{n+1} = D_n + \mu$), and observations are skipped till D_n crosses 0 from below. As a consequence, the number of observations that are skipped is determined by the undershoot (log-likelihood ratio of the observations) as well as the parameter μ. When D_n crosses 0 from below, it is reset to 0 (this is the mathematical version of the statement that beyond the skipped samples, the DE-CuSum statistics is computed using fresh samples). Once $D_n = 0$, the process renews itself and continues to evolve this way until $D_n > A$, at which time a change is declared.

If $h < \infty$, D_n is truncated to $-h$ when D_n goes below 0 from above. In other words, the undershoot is reset to $-h$ if its magnitude is larger than h. A finite value of h guarantees that the number of consecutive samples skipped is bounded by $\frac{h}{\mu} + 1$. The parameter h can be selected based on practical considerations. This feature will also be crucial to the WADD analysis of the DE-CuSum algorithm in Section V-D.

If $h = 0$, the DE-CuSum statistic D_n never becomes negative and hence reduces to the CuSum statistic and evolves as: $D_0 = 0$, and for $n \geq 0$,

$$D_{n+1} = \max\{0, D_n + \ell(X_{n+1})\}.$$

Thus, μ is a substitute for the Bayesian prior ρ that is used in the DE-Shiryaev algorithm described in Section II-A. But unlike ρ which represents a prior statistical knowledge of the change point, μ is a design parameter. An appropriate value of μ is selected to meet the constraint on PDC; see Section V-A for details.

The evolution of the DE-CuSum algorithm is plotted in Fig. 3. 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) [21], with log boundaries A and 0, is used to distinguish between the two hypotheses "$H_0 = \text{prechange}$" and "$H_1 = \text{postchange}$." If the decision in the SPRT is in favor of H_0, then samples are skipped based on the likelihood ratio of all the observations taken in the SPRT. A change is declared the first time the decision in the sequence of SPRTs is in favor of H_1. If $h = 0$, no samples are skipped and the DE-CuSum reduces to the CuSum algorithm, i.e., to a sequence of SPRTs (also see [20]).

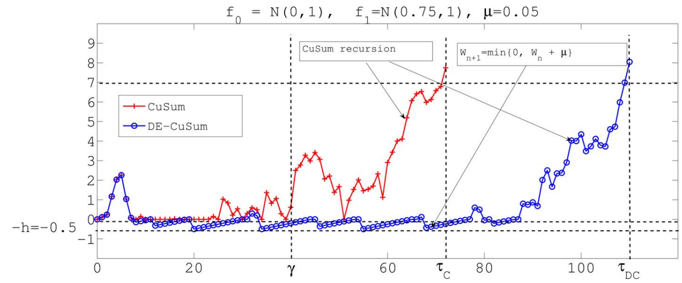

Fig. 3. Typical evolution of the CuSum and the DE-CuSum algorithms applied to the same set of samples. Parameters used: $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75, 1)$, $\gamma = 40$, $A = 7$, and $\mu = 0.05$. With $h = 0.5$, the undershoots are truncated at -0.5.

Unless it is required to have a bound on the maximum number of samples skipped, the DE-CuSum algorithm can be controlled by just two parameters: A and μ. We will show in the following 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 $\mathsf{FAR} \leq \alpha$ independent of the value of μ. Also, it is possible to select a value of μ such that $\mathsf{PDC} \leq \beta$ independent of the choice of A.

Remark 1: With the way the DE-CuSum algorithm is defined, we will see in the following that it may not be possible to meet PDC constraints that are close to 1, with equality. We ignore this issue in the rest of the paper, as in many practical settings, the preferred value of PDC would be closer to 0 than 1. But, we remark that the DE-CuSum algorithm can be easily modified to achieve PDC values that are close to 1 by resetting D_n to zero if the undershoot is smaller than a predesigned threshold.

Remark 2: One can also modify the Shiryaev–Roberts algorithm [12] and obtain a two-threshold version of it, with an upper threshold used for stopping and a lower threshold used for ON–OFF observation control. Also note that the SPRTs of the two-sided tests considered above have a lower threshold of 0. One can also propose variants of the DE-CuSum algorithm with a negative lower threshold for the SPRTs.

Remark 3: For the CuSum algorithm, the supremum in (7) and (8) is achieved when the change is applied at time $n = 1$ [see also (23)]. This is useful from the point of view of simulating the test. However, in the data-efficient setting, since the information vector also contains information about missed samples, the worst case change point in (7) would depend on the observation control and may not be $n = 1$. But note that in the DE-CuSum algorithm, the test statistic evolves as a Markov process. As a result, the worst case usually occurs in the initial slots, before the process hits stationarity. This is useful from the point of view of simulating the algorithm. In the analysis of the DE-CuSum algorithm provided in Section V, we will see that the WADD of the DE-CuSum algorithm is equal to its delay

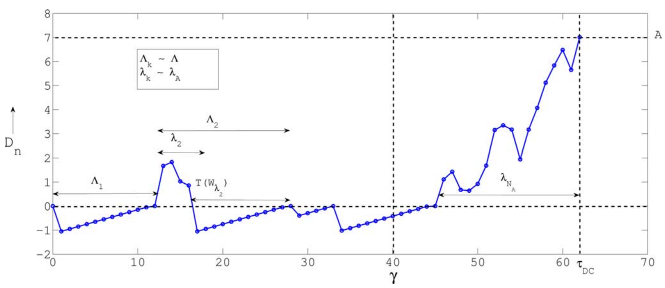

Fig. 4. Evolution of D_n for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$, and $\gamma = 40$, with $A = 7$, $h = \infty$, and $\mu = 0.1$. The two-sided tests with distribution of Λ are shown in the figure. Also shown are the two components of Λ: λ_A and $T(x)$.

when change occurs at $n = 1$, plus a constant. Similarly, even if computing the PDC may be a bit difficult using simulations, we will provide simple numerically computable upper bound on the PDC of the DE-CuSum algorithm that can be used to ensure that the PDC constraint is satisfied. We also provide an approximation using which the PDC can be set approximately.

V. Analysis and Design of the DE-CuSum Algorithm

The identification/interpretation of the DE-CuSum algorithm as a sequence of two-sided tests will now be used in this section to perform its asymptotic analysis. In the rest of the paper, we use $L(X)$ and $\ell(X)$ to denote $\frac{f_1(X)}{f_0(X)}$ and $\log \frac{f_1(X)}{f_0(X)}$, respectively.

Recall that the DE-CuSum algorithm can be seen as a sequence of two sided tests; each two-sided test contains an SPRT and a possible sojourn below zero, the length of the latter being dependent on the likelihood ratio of the observations. To capture these quantities mathematically, we now define some new variables.

Define the stopping time for an SPRT

$$\lambda_A \triangleq \inf\{n \geq 1 : \sum_{k=1}^{n} \ell(X_k) \notin [0, A]\}. \tag{11}$$

To capture the sojourn time below 0, define for $x < 0$

$$T(x) = \lceil (x)^{h+}/\mu \rceil. \tag{12}$$

Note that $T(0) = 0$. We also define the stopping time for the two-sided test

$$\Lambda = \lambda_A + T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A} < 0\}}. \tag{13}$$

Let λ_∞ and Λ_∞ be the variables λ_A and Λ, respectively, when the threshold $A = \infty$.

The variables λ_A, Λ, and $T(x)$ should be interpreted as follows. The DE-CuSum algorithm can be seen as a sequence of two-sided tests, with the stopping time of each two-sided test distributed accordingly to the law of Λ. Each of the above two-sided tests consists of an SPRT with stopping time distributed accordingly to the law of λ_A, and a sojourn of length $T(D_{\lambda_A})$ corresponding to the time for which the statistic D_n is below 0, provided that at the stopping time for the SPRT, the accumulated log likelihood is negative, i.e., the event $\{D_{\lambda_A} < 0\}$ happens. See Fig. 4. In the figure, $\Lambda_1, \Lambda_2, \ldots$ are random variables distributed accordingly to the law of Λ, and $\lambda_1, \lambda_2, \ldots$ are random variables distributed accordingly to the law of λ_A.

The CuSum algorithm can also be seen as a sequence of SPRTs, with the stopping time of each SPRT distributed according to the law of λ_A (see [20]).

We now provide some results on the mean of λ_A and $T(x)$ that will be used in the analysis of the DE-CuSum algorithm in Sections V-A, V-C, and V-D.

If $0 < D(f_0 \parallel f_1) < \infty$, then from [22, Corollary 2.4],

$$\mathbb{E}_\infty[\lambda_\infty] < \infty, \tag{14}$$

and by Wald's lemma

$$\mathbb{E}_\infty[|D_{\lambda_\infty}|] = D(f_0 \parallel f_1)\, \mathbb{E}_\infty[\lambda_\infty] < \infty. \tag{15}$$

Also for $h \geq 0$

$$\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|] \leq \mathbb{E}_\infty[|D_{\lambda_\infty}|] < \infty, \tag{16}$$

where the finiteness follows from (15).

In the following lemma, we show that the quantity $\mathbb{E}_\infty[\lambda_A | D_{\lambda_A} < 0]$ is finite for every A and provide a finite upper bound to it that is not a function of the threshold A. This result will be used in the PDC analysis in Section V-A.

Lemma 1: If $0 < D(f_0 \parallel f_1) < \infty$, then for any A, $\mathbb{E}_\infty[\lambda_A | D_{\lambda_A} < 0]$ is well defined and finite:

$$\mathbb{E}_\infty[\lambda_A | D_{\lambda_A} < 0] \leq \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{P}_\infty(\ell(X_1) < 0)} < \infty.$$

Proof: The proof of the first inequality is provided in the Appendix. The second inequality is true by (14) and because $\mathbb{P}_\infty(\ell(X_1) < 0) > 0$. ■

In the following lemma, we provide upper and lower bounds on $\mathbb{E}_\infty[T(D_{\lambda_A}) | D_{\lambda_A} < 0]$ that are not a function of the threshold A. The upper bound will be useful in the FAR analysis in Section V-C, and the lower bound will be useful in the PDC analysis in Section V-A. Define

$$T_L^{(\infty)} = \frac{\mathbb{E}_\infty[|\ell(X_1)^{h+}|;\ \{\ell(X_1) < 0\}]}{\mu}, \tag{17}$$

and

$$T_U^{(\infty)} = \frac{\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|]}{\mu\, \mathbb{P}_\infty(\ell(X_1) < 0)} + 1. \tag{18}$$

Lemma 2: If $0 < D(f_0 \parallel f_1) < \infty$ and $\mu > 0$, then

$$T_L^{(\infty)} \leq \mathbb{E}_\infty[T(D_{\lambda_A}) \big| D_{\lambda_A} < 0] \leq T_U^{(\infty)}. \tag{19}$$

Moreover, $T_U^{(\infty)} < \infty$, and if $h > 0$, then $T_L^{(\infty)} > 0$.

Proof: The proof is provided in the Appendix. ■

In the next lemma, we show that the mean of $\mathbb{E}_1[T(D_{\lambda_A}) | D_{\lambda_A} < 0]$ is finite under $\mathbb{P}_1$ and obtain a finite upper bound to it that is not a function of A. This result will be used for the CADD and WADD analysis in Section V-D. Let

$$T_U^{(1)} = \frac{\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|]}{\mu\, \mathbb{P}_1(\ell(X_1) < 0)} + 1. \tag{20}$$

Lemma 3: If $0 < D(f_0 \parallel f_1) < \infty$ and $\mu > 0$, then

$$\mathbb{E}_1[T(D_{\lambda_A})|D_{\lambda_A} < 0] \quad \leq \quad T_U^{(1)} \quad < \quad \infty.$$

Proof: The proof is provided in the Appendix. ■

A. *Meeting the* PDC *Constraint*

In this section, we show that the PDC metric is well defined for the DE-CuSum algorithm. In general, $\mathsf{PDC}(\Psi_{\text{DC}})$ will depend on both A and μ (apart from the obvious dependence on f_0 and f_1). But, we show that it is possible to choose a value of μ that ensures that the PDC constraint of β can be met independent of the choice of A. The latter is be crucial to the asymptotic optimality proof of the DE-CuSum algorithm provided later in Section V-E.

Theorem 5.1: For fixed values of A, h, and $\mu > 0$, if $0 < D(f_0 \parallel f_1) < \infty$, then

$$\mathsf{PDC}(\Psi_{\text{DC}}(A,\mu,h)) = \frac{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0]}{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0] + \mathbb{E}_\infty[T(D_{\lambda_A}) \mid D_{\lambda_A} < 0]}. \quad (21)$$

Proof: Consider an alternating renewal process $\{V_n, U_n\}$, i.e., a renewal process with renewal times $\{V_1, V_1 + U_1, V_1 + U_1 + V_2, \ldots\}$, with $\{V_n\}$ i.i.d. with distribution of λ_A conditioned on $\{D_{\lambda_A} < 0\}$, and $\{U_n\}$ i.i.d. with distribution of $T(D_{\lambda_A})$ conditioned on $\{D_{\lambda_A} < 0\}$. Thus,

$$\mathbb{E}_\infty[V_1] = \mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0],$$

and

$$\mathbb{E}_\infty[U_1] = \mathbb{E}_\infty[T(D_{\lambda_A}) \mid D_{\lambda_A} < 0].$$

Both the means are finite by Lemmas 1 and 2.

At time n assign a reward of $R_n = 1$ if the renewal cycle in progress has the law of V_1, set $R_n = 0$ otherwise. Then, by renewal reward theorem,

$$\frac{1}{n}\mathbb{E}_\infty\left[\sum_{k=1}^{n-1} R_k\right] \to \frac{\mathbb{E}_\infty[V_1]}{\mathbb{E}_\infty[V_1] + \mathbb{E}_\infty[U_1]}.$$

On $\{\tau_{\text{DC}} \geq n\}$, the total number of observations taken till time $n-1$ has the same distribution as the total reward for the alternating renewal process defined above. Hence, the expected value of the average reward for both the sequences must have the same limit:

$$\mathsf{PDC}(\Psi_{\text{DC}}(A,\mu,h)) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}_n\left[\sum_{k=1}^{n-1} M_k \middle| \tau_{\text{DC}} \geq n\right]$$
$$= \frac{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0]}{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0] + \mathbb{E}_\infty[T(D_{\lambda_A})|D_{\lambda_A} < 0]}. \quad (22)$$

■

If $h = 0$, then $\mathbb{E}_\infty[T(D_{\lambda_A})|D_{\lambda_A} < 0] = 0$ and we get the PDC of the CuSum algorithm that is equal to 1.

As can be seen from (21), PDC is a function of A as well as that of h and μ. We now show that for any A and $h > 0$, the DE-CuSum algorithm can be designed to meet any PDC constraint of β. Moreover, for a given $h > 0$, a value of μ can always be selected such that the PDC constraint of β is met independent of the choice of A. The latter is convenient not only from a practical point of view, but will also help in the asymptotic optimality proof of the DE-CuSum algorithm in Section V-E.

Theorem 5.2: For the DE-CuSum algorithm, for any choice of A and $h > 0$, if $0 < D(f_0 \parallel f_1) < \infty$, then we can always choose a value of μ to meet any given PDC constraint of β. Moreover, for any fixed value of $h > 0$, there exists a value of μ say $\mu^*(h)$ such that for every A,

$$\mathsf{PDC}(\Psi_{\text{DC}}(A,\mu^*,h)) \leq \beta.$$

In fact any μ that satisfies

$$\mu \leq \frac{\mathbb{E}_\infty[|\ell(X_1)^{h+}| \mid \ell(X_1) < 0]\, \mathbb{P}_\infty(\ell(X_1) < 0)^2}{\mathbb{E}_\infty[\lambda_\infty]} \frac{\beta}{1-\beta},$$

can be used as μ^*.

Proof: Note that $\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0]$ is not affected by the choice of h and μ. Moreover, from Lemma 2

$$\mathbb{E}_\infty[T(D_{\lambda_A}) \mid D_{\lambda_A} < 0] \geq T_L^{(\infty)}.$$

Thus, from (17), for a given A and h,

$$\mathbb{E}_\infty[T(D_{\lambda_A})|D_{\lambda_A} < 0] \to \infty \text{ as } \mu \to 0.$$

Therefore, we can always select a μ small enough so that the PDC is smaller than the given constraint of β.

Next, our aim is to find a μ^* such that for every A

$$\frac{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0]}{\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0] + \mathbb{E}_\infty[T(D_{\lambda_A})) \mid D_{\lambda_A} < 0]} \leq \beta,$$

Since, PDC increases as $\mathbb{E}_\infty[\lambda_A|D_{\lambda_A} < 0]$ increases and $\mathbb{E}_\infty[T(D_{\lambda_A}) \mid D_{\lambda_A} < 0]$ decreases, we have from Lemmas 1 and 2,

$$\mathsf{PDC}(\Psi_{\text{DC}}) \leq \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + T_L^{(\infty)}\ \mathbb{P}_\infty(\ell(X_1) < 0)}.$$

Then, the theorem is proved if we select μ such that the right-hand side of the above equation is less than β or using (17), a μ that satisfies

$$\mu \leq \frac{\mathbb{E}_\infty[|\ell(X_1)^{h+}| \mid \ell(X_1) < 0]\, \mathbb{P}_\infty(\ell(X) < 0)^2}{\mathbb{E}_\infty[\lambda_\infty]} \frac{\beta}{1-\beta}.$$

■

While the existence of μ^* proved by Theorem 5.2 above is critical for asymptotic optimality of the DE-CuSum algorithm, the estimate it provides when substituted for μ in (21) may be a bit conservative. In Section V-F, we provide a good approximation to PDC that can be used to choose the value of μ in practice. In Section VI, we provide numerical results showing the accuracy of the approximation.

Remark 4: By Theorem 5.2, for any value of h, we can select a value of μ small enough, so that any PDC constraint close to zero can be met with equality. However, meeting the PDC constraint with equality may not be possible if β is close to 1. This is because if $h \neq 0$, then

$$\mathsf{PDC}(\Psi_{\text{DC}}) \leq \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + \mathbb{P}_\infty(\ell(X) < 0)} < 1.$$

However, as mentioned earlier, for most practical applications, β will be close to zero than 1, and hence, this issue will not be encountered. If β close to 1 is indeed desired, then the DE-CuSum algorithm can be easily modified to address this issue (by skipping samples only when the undershoot is larger than a predesigned threshold).

B. *Analysis of the CuSum Algorithm*

In the sections to follow, we will express the performance of the DE-CuSum algorithm in terms of the performance of the CuSum algorithm. Therefore, in this section, we summarize the performance of the CuSum algorithm.

It is easy to show (see [3]) that

$$\mathsf{CADD}(\Psi_{\text{C}}) = \mathsf{WADD}(\Psi_{\text{C}}) = \mathbb{E}_1[\tau_{\text{C}} - 1]. \tag{23}$$

From [6], if $0 < D(f_1 \parallel f_0) < \infty$, then $\mathbb{E}_1[\tau_{\text{C}}] < \infty$. Moreover, if $\{\lambda_1, \lambda_2, \ldots\}$ are i.i.d. random variables each with distribution of λ_A, then by Wald's lemma [20]

$$\mathbb{E}_1[\tau_{\text{C}}] = \mathbb{E}_1\left[\sum_{k=1}^{N} \lambda_k\right] = \mathbb{E}_1[N]\,\mathbb{E}_1[\lambda_A], \tag{24}$$

where N is the number of two-sided tests (SPRTs)—each with distribution of λ_A—executed before the change is declared.

It is also shown in [6] that $0 < D(f_1 \parallel f_0) < \infty$ is also sufficient to guarantee $\mathbb{E}_\infty[\tau_{\text{C}}] < \infty$ and $\mathsf{FAR}(\Psi_{\text{C}}) > 0$. Moreover,

$$\mathbb{E}_\infty[\tau_{\text{C}}] = \mathbb{E}_\infty\left[\sum_{k=1}^{N} \lambda_k\right] = \mathbb{E}_\infty[N]\,\mathbb{E}_\infty[\lambda_A]. \tag{25}$$

The proof of the following theorem can be found in [6] and [10].

Theorem 5.3: If $0 < D(f_1 \parallel f_0) < \infty$, then with $A = \log \frac{1}{\alpha}$,

$$\mathsf{FAR}(\Psi_{\text{C}}) \leq \alpha,$$

and as $\alpha \to 0$,

$$\mathsf{CADD}(\Psi_{\text{C}}) = \mathsf{WADD}(\Psi_{\text{C}}) = \mathbb{E}_1[\tau_{\text{C}} - 1] \sim \frac{|\log \alpha|}{D(f_1 \parallel f_0)}.$$

Thus, the CuSum algorithm is asymptotically optimal for both Problems 3 and 2 because for any stopping time τ with $\mathsf{FAR}(\tau) \leq \alpha$,

$$\mathsf{WADD}(\tau) \geq \mathsf{CADD}(\tau) \geq \frac{|\log \alpha|}{D(f_1 \parallel f_0)}\Big(1 + o(1)\Big), \tag{26}$$

as $\alpha \to 0$.

C. FAR *for the DE-CuSum Algorithm*

In this section, we characterize the false alarm rate of the DE-CuSum algorithm. The following lemma shows that for a fixed A, μ, and h, if the DE-CuSum algorithm and the CuSum algorithm are applied to the same sequence of random variables, then sample-pathwise, the DE-CuSum statistic D_n is always below the CuSum statistic C_n. Thus, the DE-CuSum algorithm crosses the threshold A only after the CuSum algorithm has crossed it.

Lemma 4: Under any $\mathbb{P}_n,\ n \geq 1$ and under $\mathbb{P}_\infty$,

$$C_n \geq D_n.$$

Thus

$$\tau_{\text{C}} \leq \tau_{\text{DC}}.$$

Proof: This follows directly from the definition of the DE-CuSum algorithm. If a sequence of samples causes the statistic of the DE-CuSum algorithm to go above A, then since all the samples are utilized in the CuSum algorithm, the same sequence must also cause the CuSum statistic to go above A. ∎

It follows as a corollary of Lemma 4 that

$$\mathbb{E}_\infty[\tau_{\text{C}}] \leq \mathbb{E}_\infty[\tau_{\text{DC}}].$$

The following theorem shows that these quantities are finite and also provides an estimate for $\mathsf{FAR}(\Psi_{\text{DC}})$.

Theorem 5.4: For any fixed h (including $h = \infty$) and $\mu > 0$, if

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

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

$$\mathsf{FAR}(\Psi_{\text{DC}}) \leq \mathsf{FAR}(\Psi_{\text{C}}) \leq \alpha.$$

Moreover, for any A

$$\mathbb{E}_\infty[\tau_{\text{DC}}] = \mathbb{E}_\infty[\tau_{\text{C}}] + \frac{\mathbb{E}_\infty[T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A} < 0\}}]}{\mathbb{P}_\infty(D_{\lambda_A} > 0)} \tag{27}$$

and as $A \to \infty$,

$$\frac{\mathsf{FAR}(\Psi_{\text{DC}})}{\mathsf{FAR}(\Psi_{\text{C}})} \to \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + \mathbb{E}_\infty[T(D_{\lambda_\infty})]}, \tag{28}$$

where λ_∞ is the variable λ_A with $A = \infty$. The limit in (28) is strictly less than 1 if $h > 0$.

Proof: For a fixed A, let N_A be the number of two-sided tests of distribution Λ executed before the change is declared in the DE-CuSum algorithm. Then, if $\{\Lambda_1, \Lambda_2, \ldots\}$ is a sequence of random variables each with distribution of Λ, then

$$\mathbb{E}_\infty[\tau_{\text{DC}}] = \mathbb{E}_\infty\left[\sum_{k=1}^{N_A} \Lambda_k\right].$$

Because of the renewal nature of the DE-CuSum algorithm,

$$\mathbb{E}_\infty[N_A] = \mathbb{E}_\infty[N],$$

where N is the number of SPRTs used in the CuSum algorithm. Thus, from (25),

$$\mathbb{E}_\infty[N_A] = \mathbb{E}_\infty[N] \leq \mathbb{E}_\infty[\tau_{\text{C}}] < \infty.$$

Further, from (13),

$$\mathbb{E}_\infty[\Lambda] = \mathbb{E}_\infty[\lambda_A] + \mathbb{E}_\infty[T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A}<0\}}]. \tag{29}$$

From (25) again

$$\mathbb{E}_\infty[\lambda_A] \le \mathbb{E}_\infty[\tau_{\mathrm{C}}] < \infty.$$

Moreover, from Lemma 2

$$\begin{aligned}\mathbb{E}_\infty[T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A}<0\}}] &\le \mathbb{E}_\infty[T(D_{\lambda_A}) \mid D_{\lambda_A} < 0] \\ &\le T_U^{(\infty)} < \infty.\end{aligned}$$

Thus, $\mathbb{E}_\infty[\Lambda] < \infty$ and

$$\mathbb{E}_\infty[\tau_{\mathrm{DC}}] = \mathbb{E}_\infty\left[\sum_{k=1}^{N_A} \Lambda_k\right] = \mathbb{E}_\infty[N_A]\, \mathbb{E}_\infty[\Lambda] < \infty.$$

It follows as a corollary of Lemma 4 and Theorem 5.3 that for $A = \log \frac{1}{\alpha}$,

$$\mathsf{FAR}(\Psi_{\mathrm{DC}}) \le \mathsf{FAR}(\Psi_{\mathrm{C}}) \le \alpha.$$

Since N_A is $\mathrm{Geom}(\mathbb{P}_\infty(D_{\lambda_A} > 0))$, (27) follows from (29) and (25):

$$\begin{aligned}\mathbb{E}_\infty[\tau_{\mathrm{DC}}] &= \frac{\mathbb{E}_\infty[\Lambda]}{\mathbb{P}_\infty(D_{\lambda_A} > 0)} \\ &= \frac{\mathbb{E}_\infty[\lambda_A]}{\mathbb{P}_\infty(D_{\lambda_A} > 0)} + \frac{\mathbb{E}_\infty[T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A}<0\}}]}{\mathbb{P}_\infty(D_{\lambda_A} > 0)} \\ &= \mathbb{E}_\infty[\tau_{\mathrm{C}}] + \frac{\mathbb{E}_\infty[T(D_{\lambda_A})\, \mathbb{I}_{\{D_{\lambda_A}<0\}}]}{\mathbb{P}_\infty(D_{\lambda_A} > 0)}.\end{aligned} \tag{30}$$

Further, since $\mathbb{E}_\infty[N_A] = \mathbb{E}_\infty[N]$, we have

$$\frac{\mathbb{E}_\infty[\tau_{\mathrm{C}}]}{\mathbb{E}_\infty[\tau_{\mathrm{DC}}]} = \frac{\mathbb{E}_\infty[N]\, \mathbb{E}_\infty[\lambda_A]}{\mathbb{E}_\infty[N_A]\, \mathbb{E}_\infty[\Lambda]} = \frac{\mathbb{E}_\infty[\lambda_A]}{\mathbb{E}_\infty[\Lambda]}.$$

Since $\lambda_A \uparrow \lambda_\infty$ and $\Lambda \uparrow \Lambda_\infty$, we have by monotone convergence theorem, as $A \to \infty$,

$$\frac{\mathbb{E}_\infty[\tau_{\mathrm{C}}]}{\mathbb{E}_\infty[\tau_{\mathrm{DC}}]} \to \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\Lambda_\infty]} = \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + \mathbb{E}_\infty[T(D_{\lambda_\infty})]}.$$

The limit is clearly less than 1 if $h > 0$. ■

Remark 5: Thus, unlike the Bayesian setting where the PFA of the DE-Shiryaev algorithm converges to the PFA of the Shiryaev algorithm, here the FAR of the DE-CuSum algorithm is strictly less than the FAR of the CuSum algorithm. Moreover, for large A, the right-hand side of (28) is approximately the PDC achieved. Thus, (28) shows that, asymptotically as $A \to \infty$, the ratio of the FARs is approximately equal to the PDC. This also shows that one can set the threshold in the DE-CuSum algorithm to a value smaller than $A = \frac{1}{\alpha}$ to meet the FAR constraint with equality. This latter fact will be used in obtaining the numerical results in Section VI.

D. CADD *and* WADD *of the DE-CuSum Algorithm*

We now provide expressions for CADD and WADD of the DE-CuSum algorithm The main content of Theorems 5.5 and 5.6 is that for each value of A, the CADD and WADD of the DE-CuSum algorithm is within a constant of the corresponding performance of the CuSum algorithm. This constant is independent of the choice of A, and as a result, the delay performances of the two algorithms is asymptotically the same. The results depend on the following fundamental lemma. The lemma says that when the change happens at $n = 1$, then the average delay of the DE-CuSum algorithm starting with $D_0 = x > 0$ is upper bounded by the average delay of the algorithm when $D_0 = 0$, plus a constant.

Let

$$\tau_{\mathrm{DC}}(x) = \inf\{n \ge 1 : D_n > A; D_0 = x\}.$$

Here, D_n is the DE-CuSum statistic and evolves according the description of the algorithm in Section IV. Thus, $\tau_{\mathrm{DC}}(x)$ is the first time for the DE-CuSum algorithm to cross A, when starting at $D_0 = x$. Clearly, $\tau_{\mathrm{DC}}(x) = \tau_{\mathrm{DC}}$ if $x = 0$.

Lemma 5: Let $0 < D(f_1 \parallel f_0) < \infty$ and $0 \le x < A$. Then,

$$\mathbb{E}_1[\tau_{\mathrm{DC}}(x)] \le \mathbb{E}_1[\tau_{\mathrm{DC}}] + T_U^{(1)},$$

where $T_U^{(1)}$ is an upper bound to the variable $T(x)$ [see (20)]. Moreover, if $h < \infty$, then

$$\mathbb{E}_1[\tau_{\mathrm{DC}}(x)] \le \mathbb{E}_1[\tau_{\mathrm{DC}}] + \lceil h/\mu \rceil.$$

Proof: The proof is provided in the Appendix. ■

We first provide the result for the CADDs of the two algorithms.

Theorem 5.5: Let

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

Then, for fixed values of $\mu > 0$ and h, and for each A,

$$\mathsf{CADD}(\Psi_{\mathrm{DC}}) \le \mathsf{CADD}(\Psi_{\mathrm{C}}) + K_1,$$

where K_1 is a constant not a function of A. Thus, as $A \to \infty$,

$$\mathsf{CADD}(\Psi_{\mathrm{DC}}) \le \mathsf{CADD}(\Psi_{\mathrm{C}}) + O(1).$$

Proof: If the change happens at $n = 1$, then

$$\mathbb{E}_1[\tau_{\mathrm{DC}} - 1 | \tau_{\mathrm{DC}} \ge 1] = \mathbb{E}_1[\tau_{\mathrm{DC}}] - 1 \le \mathbb{E}_1[\tau_{\mathrm{DC}}].$$

Let the change happen at time $n > 1$. Then, on $\{D_{n-1} \ge 0\}$, by Lemma 5, the average delay $\mathbb{E}_n[\tau_{\mathrm{DC}} - n | \tau_{\mathrm{DC}} \ge n]$ is bounded from above by $\mathbb{E}_1[\tau_{\mathrm{DC}}] + T_U^{(1)}$, and on $\{D_{n-1} < 0\}$, the average delay is bounded from above by $\mathbb{E}_1[\tau_{\mathrm{DC}}]$ plus the maximum possible average time spent by the DE-CuSum statistic below 0 under $\mathbb{P}_\infty$, which is $T_U^{(\infty)}$. Thus, from Lemma 2, for $n > 1$,

$$\begin{aligned}&\mathbb{E}_n[\tau_{\mathrm{DC}} - n | \tau_{\mathrm{DC}} \ge n] \\ &\quad \le \left(\mathbb{E}_1[\tau_{\mathrm{DC}}] + T_U^{(1)}\right) \mathbb{P}_\infty(D_{n-1} \ge 0) \\ &\quad + \left(\mathbb{E}_1[\tau_{\mathrm{DC}}] + T_U^{(\infty)}\right) \mathbb{P}_\infty(D_{n-1} < 0).\end{aligned}$$

Thus, for all $n \ge 1$

$$\mathbb{E}_n[\tau_{\mathrm{DC}} - n | \tau_{\mathrm{DC}} \ge n] \le \mathbb{E}_1[\tau_{\mathrm{DC}}] + T_U^{(1)} + T_U^{(\infty)}.$$

Since the right-hand side of the above equation is not a function of n, we have

$$\mathsf{CADD}(\Psi_{\rm DC}) \leq \mathbb{E}_1[\tau_{\rm DC}] + T_U^{(1)} + T_U^{(\infty)}.$$

Following Theorem 5.4 and its proof, it is easy to see that

$$\mathbb{E}_1[\tau_{\rm DC}] = \mathbb{E}_1[\tau_{\rm C}] + \frac{\mathbb{E}_1[T(D_{\lambda_A})\,\mathbb{I}_{\{D_{\lambda_A}<0\}}]}{\mathbb{P}_1(D_{\lambda_A}>0)}.$$

From Lemma 3 and the fact that $\mathbb{P}_1(D_{\lambda_A}>0) > \mathbb{P}_1(D_{\lambda_\infty}>0)$, we have

$$\frac{\mathbb{E}_1[T(D_{\lambda_A})\,\mathbb{I}_{\{D_{\lambda_A}<0\}}]}{\mathbb{P}_1(D_{\lambda_A}>0)} \leq \frac{T_U^{(1)}}{\mathbb{P}_1(D_{\lambda_\infty}>0)}.$$

Also from (23), we have $\mathsf{CADD}(\Psi_{\rm C}) = \mathbb{E}_1[\tau_{\rm C}-1]$. Thus,

$$\begin{aligned}\mathsf{CADD}(\Psi_{\rm DC}) \leq\ & \mathsf{CADD}(\Psi_{\rm C}) \\ &+ \frac{T_U^{(1)}}{\mathbb{P}_1(D_{\lambda_\infty}>0)} + T_U^{(1)} + T_U^{(\infty)} + 1.\end{aligned}$$

This proves the theorem. ■

Note that the above theorem is valid even if h is not finite. In contrast, it is possible that $\mathsf{WADD}(\Psi_{\rm DC}) = \infty$ if $h=\infty$; e.g., when f_0 and f_1 are Gaussian densities. As a result, we need a bound on the number of consecutive samples skipped for finiteness of worst case delay according to the criterion of Lorden.

We now express the WADD of the DE-CuSum algorithm in terms of the WADD of the CuSum algorithm.

Theorem 5.6: Let

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

Then, for fixed values of $\mu > 0$ and $h < \infty$, and for each A,

$$\mathsf{WADD}(\Psi_{\rm DC}) \leq \mathsf{WADD}(\Psi_{\rm C}) + K_2,$$

where K_2 is a constant not a function of A. Thus, as $A \to \infty$,

$$\mathsf{WADD}(\Psi_{\rm DC}) \leq \mathsf{WADD}(\Psi_{\rm C}) + O(1).$$

Proof: From Lemma 5, it follows that for $n > 1$

$$\text{ess sup}\ \mathbb{E}_n\left[(\tau_{\rm DC}-n)^+ | I_{n-1}\right] \leq \lceil h/\mu \rceil + \mathbb{E}_1[\tau_{\rm DC}].$$

Since the right-hand side is not a function of n and it is greater than $\mathbb{E}_1[\tau_{\rm DC}-1]$, we have

$$\mathsf{WADD}(\Psi_{\rm DC}) \leq \lceil h/\mu \rceil + \mathbb{E}_1[\tau_{\rm DC}].$$

Thus, from the proof of theorem above and (23)

$$\begin{aligned}\mathbb{E}_1[\tau_{\rm DC}] &\leq \mathbb{E}_1[\tau_{\rm C}] + \frac{T_U^{(1)}}{\mathbb{P}_1(D_{\lambda_\infty}>0)} \\ &= \mathsf{WADD}(\Psi_{\rm C}) + \frac{T_U^{(1)}}{\mathbb{P}_1(D_{\lambda_\infty}>0)} + 1,\end{aligned}$$

and we have

$$\mathsf{WADD}(\Psi_{\rm DC}) \leq \mathsf{WADD}(\Psi_{\rm C}) + \frac{T_U^{(1)}}{\mathbb{P}_1(D_{\lambda_\infty}>0)} + \frac{h}{\mu} + 2.$$

This proves the theorem. ■

The following corollary follows easily from Theorems 5.3, 5.5, and 5.6.

Corollary 1: If $0 < D(f_1 \,\|\, f_0) < \infty$ and $0 < D(f_0 \,\|\, f_1) < \infty$, then for fixed values of μ and h, including the case of $h=\infty$ (no truncation), as $A \to \infty$,

$$\mathsf{CADD}(\Psi_{\rm DC}) \sim \frac{A}{D(f_1 \,\|\, f_0)}.$$

Moreover, if $h < \infty$, then as $A \to \infty$,

$$\mathsf{WADD}(\Psi_{\rm DC}) \sim \frac{A}{D(f_1 \,\|\, f_0)}.$$

E. Asymptotic Optimality of the DE-CuSum Algorithm

We now use the results from the previous sections to show that the DE-CuSum algorithm is asymptotically optimal.

The following theorem states that for a given PDC constraint of β, the DE-CuSum algorithm is asymptotically optimal for both Problems 2 and 3, as $\alpha \to 0$, for the following reasons:

1) the PDC of the DE-CuSum algorithm can be designed to meet the constraint independent of the choice of A,
2) the CADD and WADD of the DE-CuSum algorithm approaches the corresponding performances of the CuSum algorithm,
3) the FAR of the DE-CuSum algorithm is always better than that of the CuSum algorithm, and
4) the CuSum algorithm is asymptotically optimal for both Problems 2 and 3, with $\beta = 1$, as $\alpha \to 0$.

Theorem 5.7: Let $0 < D(f_1 \,\|\, f_0) < \infty$ and $0 < D(f_0 \,\|\, f_1) < \infty$. For a given α, set $A = \log\frac{1}{\alpha}$, then for any choice of h and μ,

$$\mathsf{FAR}(\Psi_{\rm DC}) \leq \mathsf{FAR}(\Psi_{\rm C}) \leq \alpha.$$

For a given β, and for any given h, it is possible to select $\mu = \mu^*(h)$ such that $\forall A$, and hence even with $A = \log\frac{1}{\alpha}$,

$$\mathsf{PDC}(\Psi_{\rm DC}) \leq \beta.$$

Moreover, for each fixed β, for any h and with $\mu^*(h)$ selected to meet this PDC constraint of β, as $\alpha \to 0$ (or $A \to \infty$ because $A = \log\frac{1}{\alpha}$),

$$\mathsf{CADD}(\Psi_{\rm DC}(\log\frac{1}{\alpha}, h, \mu^*(h))) \sim \mathsf{CADD}(\Psi_{\rm C}) \sim \frac{|\log\alpha|}{D(f_1 \,\|\, f_0)}.$$

Furthermore, if the h chosen above is finite, then

$$\mathsf{WADD}(\Psi_{\rm DC}(\log\frac{1}{\alpha}, h, \mu^*(h))) \sim \mathsf{WADD}(\Psi_{\rm C}) \sim \frac{|\log\alpha|}{D(f_1 \,\|\, f_0)}.$$

Proof: The result on FAR follows from Theorem 5.4. The fact that one can select a $\mu = \mu^*(h)$ to meet the PDC constraint independent of the choice of A follows from Theorem 5.2. Finally, the delay asymptotics follow from Theorems 5.5 and 5.6 and Corollary 1. ■

Since, by Theorem 5.3, $\frac{|\log \alpha|}{D(f_1 \| f_0)}$ is the best possible asymptotics performance for any given FAR constraint of α, the above statement establishes the asymptotic optimality of the DE-CuSum algorithm for both Problems 2 and 3.

F. Design of the DE-CuSum Algorithm

We now discuss how to set the parameters μ, h, and A so as to meet a given FAR constraint of α and a PDC constraint of β.

Theorem 5.4 provides the guideline for choosing A: for any h, μ,

$$\text{if} \quad A = \log \frac{1}{\alpha}, \quad \text{then} \quad \mathsf{FAR}(\Psi_{\text{DC}}) \leq \alpha.$$

As discussed earlier, Theorem 5.2 provides a conservative estimate of the PDC. For practical purposes, we suggest using the following approximation for PDC (obtained in the limit as $A \to \infty$):

$$\mathsf{PDC} \approx \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + \mathbb{E}_\infty[\lceil \frac{D^{h+}_{\lambda_\infty}}{\mu} \rceil]}. \tag{31}$$

For large values of A, (31) will indeed provide a good estimate of the PDC. We note that $\mathbb{E}_\infty[\lambda_\infty]$ can be computed numerically; see [22, Corollary 2.4]. The quantity $\mathbb{E}_\infty[\lceil \frac{D^{h+}_{\lambda_\infty}}{\mu} \rceil]$ can be computed using Monte Carlo simulations.

If $h = \infty$, then using (15) we can further simplify (31) to

$$\mathsf{PDC} \approx \frac{\mathbb{E}_\infty[\lambda_\infty]}{\mathbb{E}_\infty[\lambda_\infty] + \frac{\mathbb{E}_\infty[|D_{\lambda_\infty}|]}{\mu}} = \frac{\mu}{\mu + D(f_0 \| f_1)}. \tag{32}$$

Thus, to ensure $\mathsf{PDC} \leq \beta$, the approximation above suggests selecting μ such that

$$\mu \leq \frac{\beta}{1-\beta} D(f_0 \| f_1).$$

In Section VI, we provide numerical results that shows that the approximation (32) indeed provides a good estimate of the PDC when $h = \infty$.

VI. Tradeoff Curves

The asymptotic optimality of the DE-CuSum algorithm for all β does not guarantee good performance for moderate values of FAR. In Fig. 5, we plot the tradeoff curves for the CuSum algorithm and the DE-CuSum algorithm, obtained using simulations. We plot the performance of the DE-CuSum algorithm for two different PDC constraints: $\beta = 0.5$ and $\beta = 0.25$. For simplicity, we restrict ourself to the CADD performance for $h = \infty$ in this section. Similar performance comparisons can be obtained for both CADD and WADD with $h < \infty$.

Each of the curves for the DE-CuSum algorithm in Fig. 5 is obtained in the following way. Five different threshold values for A were arbitrarily selected. For each threshold value, a large value of γ was chosen, and the DE-CuSum algorithm was simulated and the fraction of time the observations is taken before change was computed. Specifically, γ was increased in the multiples of 100 and an estimate of the PDC was obtained by

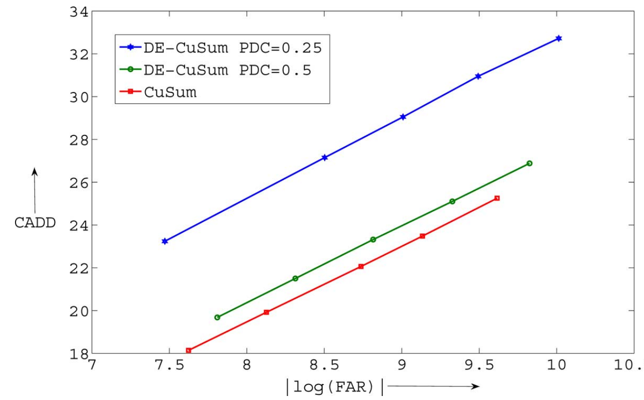

Fig. 5. Tradeoff curves for the DE-CuSum algorithm for $\mathrm{PDC} = 0.25, 0.5$, with $f_0 \sim \mathcal{N}(0,1)$ and $f_1 \sim \mathcal{N}(0.75,1)$.

Monte Carlo simulations. The value of μ was so chosen that the PDC value obtained in simulations was slightly below the constraint $\beta = 0.5$ or 0.25. For this value of μ and for the chosen threshold, the FAR was computed by selecting the change time to be $\gamma = \infty$ (generating random numbers from $f_0 \sim \mathcal{N}(0,1)$). The CADD was then computed for the above choice of μ and A by varying the value of γ from $1, 2, \ldots$ and recording the maximum of the conditional delay. The maximum was achieved in the first five slots.

As can be seen from the figure, a PDC of 0.5 (using only 50% of the samples in the long run) can be achieved using the DE-CuSum algorithm with a small penalty on the delay. If we wish to achieve a PDC of 0.25, then we have to incur a significant penalty (of approximately six slots in Fig. 5). But, note that the difference of delay with the CuSum algorithm remains fixed as FAR $\to 0$. This is due to the result reported in Theorem 5.5 and this is precisely the reason the DE-CuSum algorithm is asymptotic optimal. The tradeoff between CADD and FAR is a function of the K-L divergence between the p.d.f.'s f_1 and f_0: the larger the K-L divergence the more is the fraction of samples that can dropped for a given loss in delay performance.

In Fig. 6, we compare the performance of the DE-CuSum algorithm with the *fraction sampling* scheme, in which, to achieve a PDC of β, the CuSum algorithm is employed, and a sample is chosen with probability β for decision making. Note that this scheme skips samples without exploiting any knowledge about the state of the system. As seen in Fig. 6, the DE-CuSum algorithm performs considerably better than the fractional sampling scheme. Thus, the tradeoff curves show that the DE-CuSum algorithm has good performance even for moderate FAR, when the PDC constraint is moderate.

We now provide numerical results that show that (32) provides a good estimate for the PDC. We use the following parameters: $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$ and set $h = \infty$. In Table II(a), we fix the value of μ and vary A and compare the PDC obtained using simulations and the one obtained using (32), that is using the approximation $\mathsf{PDC} \approx \frac{\mu}{\mu + D(f_0 \| f_1)}$. We see that the approximation becomes more accurate as A increases. We also note that the PDC obtained using simulations does not converge to $\frac{\mu}{\mu + D(f_0 \| f_1)}$, even as A becomes large, because of the effect of the presence of a ceiling function in the PDC expression; see (12) and (21).

TABLE II
COMPARISON OF PDC OBTAINED USING SIMULATIONS WITH THE APPROXIMATION (32) FOR $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$ AND $h = \infty$. (A) FIXED μ. (B) FIXED A

A	μ	PDC Simulations	PDC Approx (32) $\frac{\mu}{\mu+D(f_0\|f_1)}$
1	0.1	0.16	0.26
2	0.1	0.20	0.26
3	0.1	0.22	0.26
4	0.1	0.238	0.26
6	0.1	0.248	0.26

(a)

A	μ	PDC Simulations	PDC Approx (32) $\frac{\mu}{\mu+D(f_0\|f_1)}$
6	0.01	0.033	0.034
6	0.05	0.145	0.151
6	0.2	0.37	0.41
6	0.3	0.46	0.51
6	0.4	0.51	0.58
6	0.6	0.58	0.68

(b)

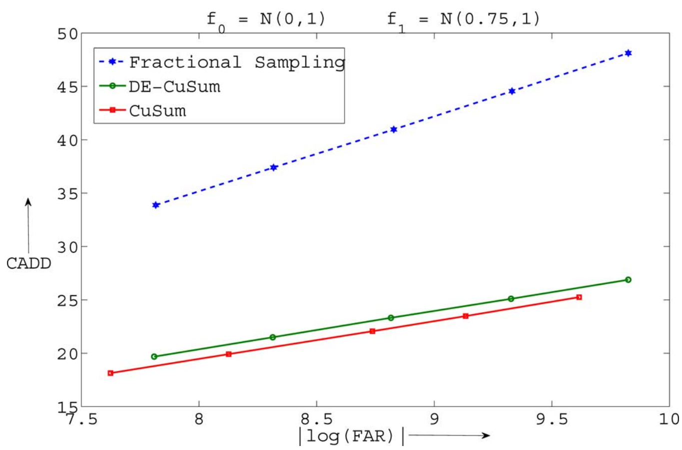

Fig. 6. Comparative performance of the DE-CuSum algorithm with the CuSum algorithm and the fractional-sampling scheme: PDC $= 0.5$, with $f_0 \sim \mathcal{N}(0,1)$ and $f_1 \sim \mathcal{N}(0.75,1)$.

In Table II(b), we next fix a large value of A, specifically $A = 6$, for which the PDC approximation is most accurate in Table II(a), and check the accuracy of the approximation $\frac{\mu}{\mu+D(f_0\|f_1)}$ by varying μ. We see in the table that the approximation is more accurate for small values of μ. This is due to the fact that the effect of the ceiling function in the PDC (12), (21) is negligible when μ is small.

VII. CONCLUSIONS AND FUTURE WORK

We proposed two minimax formulations for data-efficient non-Bayesian quickest change detection, which are extensions of the standard minimax formulations in [6] and [7] to the data-efficient setting. We proposed an algorithm called the DE-CuSum algorithm, that is a modified version of the CuSum algorithm from [13], and showed that it is asymptotically optimal for both the minimax formulations we proposed, as the false alarm rate goes to zero.

We discussed that, like the CuSum algorithm, the DE-CuSum algorithm can also be seen as a sequence of SPRTs, with the difference that each SPRT is now followed by a "sleep" time, the duration of which is a function of the accumulated log likelihood of the observations taken in the SPRT preceding it. This similarity was exploited to analyze the performance of the DE-CuSum algorithm using standard renewal theory tools, and also to show its asymptotic optimality. We also showed in our numerical results that the DE-CuSum algorithm has good tradeoff curves and provides substantial benefits over the approach of fractional sampling. The techniques developed in this paper and the insights obtained can be used to study data-efficient quickest change detection in sensor networks. See [23]–[25] for some preliminary results.

APPENDIX

Proof of Lemma 1: If $0 < D(f_0 \parallel f_1) < \infty$, then $\mathbb{E}_\infty[\lambda_\infty] < \infty$. Thus, $\mathbb{P}_\infty(\lambda_\infty < \infty) = 1$. Also,

$$\mathbb{P}_\infty(D_{\lambda_A} < 0) > \mathbb{P}_\infty(\ell(X_1) < 0) > 0.$$

Thus, $\mathbb{E}_\infty[\lambda_A \big| D_{\lambda_A} < 0]$ is well defined and

$$\begin{aligned}\mathbb{E}_\infty[\lambda_\infty] &\geq \mathbb{E}_\infty[\lambda_A] \\ &\geq \mathbb{E}_\infty[\lambda_A \big| D_{\lambda_A} < 0]\, \mathbb{P}_\infty(D_{\lambda_A} < 0). \\ &\geq \mathbb{E}_\infty[\lambda_A \big| D_{\lambda_A} < 0]\, \mathbb{P}_\infty(\ell(X_1) < 0).\end{aligned}$$

This proves the lemma. ■

Proof of Lemma 2: Again note that

$$\mathbb{P}_\infty(D_{\lambda_A} < 0) > \mathbb{P}_\infty(\ell(X_1) < 0) > 0.$$

Thus, $\mathbb{E}_\infty[T(D_{\lambda_A}) \big| D_{\lambda_A} < 0]$ is well defined.

Since $T(x) = \lceil |x^{h+}|/\mu \rceil$, we have

$$\frac{|x^{h+}|}{\mu} \leq T(x) \leq \frac{|x^{h+}|}{\mu} + 1.$$

We will use this simple inequality to obtain the upper and lower bounds.

We first obtain the upper bound. Clearly,

$$\mathbb{E}_\infty[T(D_{\lambda_A}) \big| D_{\lambda_A} < 0] \leq \frac{\mathbb{E}_\infty[|D_{\lambda_A}^{h+}| \;\big|\; D_{\lambda_A} < 0]}{\mu} + 1.$$

An upper bound for the right-hand side of the above equation is easily obtained. First note that from (16)

$$\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|] \;\le\; \mathbb{E}_\infty[|D_{\lambda_\infty}|] \;<\; \infty,$$

and

$$\begin{aligned}\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|] &\ge \mathbb{E}_\infty[|D_{\lambda_A}^{h+}| \;\Big|\; D_{\lambda_A} < 0]\, \mathbb{P}_\infty(D_{\lambda_A} < 0).\\ &\ge \mathbb{E}_\infty[|D_{\lambda_A}^{h+}| \;\Big|\; D_{\lambda_A} < 0]\, \mathbb{P}_\infty(\ell(X_1) < 0).\end{aligned}$$

This completes the proof for the upper bound.

For the lower bound, we have

$$\begin{aligned}&\mathbb{E}_\infty[T(D_{\lambda_A}) \;\Big|\; D_{\lambda_A} < 0]\\ &\ge \frac{\mathbb{E}_\infty[|D_{\lambda_A}^{h+}| \;\Big|\; D_{\lambda_A} < 0]}{\mu}\\ &\ge \frac{\mathbb{E}_\infty[|D_{\lambda_A}^{h+}|\,;\{\ell(X_1) < 0\} \;\Big|\; D_{\lambda_A} < 0]}{\mu}\\ &= \frac{\mathbb{E}_\infty[|\ell(X_1)^{h+}| \;\Big|\; \ell(X_1) < 0]}{\mu}\, \mathbb{P}_\infty(\ell(X_1) < 0).\end{aligned}$$

In the above equation we have used the fact that the unconditional probability $\mathbb{P}_\infty(\ell(X_1) < 0)$ is smaller than the conditional one $\mathbb{P}_\infty(\ell(X_1) < 0 | D_{\lambda_A} < 0)$. ■

Proof of Lemma 3: First note that

$$\mathbb{P}_1(D_{\lambda_A} < 0) > \mathbb{P}_1(\ell(X_1) < 0) > 0.$$

Thus, $\mathbb{E}_1[T(D_{\lambda_A})|D_{\lambda_A} < 0]$ is well defined. Also using the inequality on $T(x)$ from Lemma 2, we have

$$\mathbb{E}_1[T(D_{\lambda_A})\Big| D_{\lambda_A} < 0] \le \frac{\mathbb{E}_1[|D_{\lambda_A}^{h+}| \;\Big|\; D_{\lambda_A} < 0]}{\mu} + 1. \quad (33)$$

We now get an upper bound on the right-hand side of the above equation. By Wald's likelihood ratio identity [20] and (16),

$$\begin{aligned}&\mathbb{E}_1[|D_{\lambda_\infty}^{h+}|\,;\; D_{\lambda_\infty} < 0]\\ &\quad= \mathbb{E}_1[|D_{\lambda_\infty}^{h+}|\,;\; \lambda_\infty < \infty]\\ &\quad= \mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}| \prod_{k=1}^{\lambda_\infty} L(X_k)\,;\; \lambda_\infty < \infty] \quad (34)\\ &\quad= \mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|\, e^{D_{\lambda_\infty}}\,;\; D_{\lambda_\infty} < 0]\\ &\quad\le \mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|] \;\le\; \mathbb{E}_\infty[|D_{\lambda_\infty}|] \;<\; \infty.\end{aligned}$$

Using again the technique used in the proof of Lemma 1, we have

$$\begin{aligned}&\mathbb{E}_1[|D_{\lambda_\infty}^{h+}|\,;\; D_{\lambda_\infty} < 0]\\ &\quad\ge \mathbb{E}_1[|D_{\lambda_\infty}^{h+}|\,;\; D_{\lambda_A} < 0]\\ &\quad\ge \mathbb{E}_1[|D_{\lambda_\infty}^{h+}| \mid D_{\lambda_A} < 0]\, \mathbb{P}_1(\ell(X_1) < 0) \quad (35)\\ &\quad= \mathbb{E}_1[|D_{\lambda_A}^{h+}| \mid D_{\lambda_A} < 0]\, \mathbb{P}_1(\ell(X_1) < 0).\end{aligned}$$

Thus, from (33)–(35)

$$\begin{aligned}\mathbb{E}_1[T(D_{\lambda_A})|D_{\lambda_A} < 0] &\le \frac{\mathbb{E}_1[|D_{\lambda_A}^{h+}| \;\Big|\; D_{\lambda_A} < 0]}{\mu} + 1\\ &\le \frac{\mathbb{E}_1[|D_{\lambda_\infty}^{h+}|\,;\; D_{\lambda_\infty} < 0]}{\mu\, \mathbb{P}_1(\ell(X_1) < 0)} + 1\\ &\le \frac{\mathbb{E}_\infty[|D_{\lambda_\infty}^{h+}|]}{\mu\, \mathbb{P}_1(\ell(X_1) < 0)} + 1\\ &< \infty.\end{aligned}$$

This proves the lemma. ■

Proof of Lemma 5: Let

$$\tau_{\mathrm{C}}(x) = \inf\{n \ge 1 : C_n > A; C_0 = x\}.$$

Here, C_n is the CuSum statistic and evolves according to the description of the algorithm in Algorithm 2. Thus, $\tau_{\mathrm{C}}(x)$ is the first time for the CuSum algorithm to cross A, when starting at $C_0 = x$. Clearly, $\tau_{\mathrm{C}}(x) = \tau_{\mathrm{C}}$ if $x = 0$. It is easy to see by sample pathwise arguments that

$$\mathbb{E}_1[\tau_{\mathrm{C}}(x)] \le \mathbb{E}_1[\tau_{\mathrm{C}}].$$

The proof depends on the above inequality.

Let $\mathcal{A}_x$ be the event that the CuSum statistic, starting with $C_0 = x$, touches zero before crossing the upper threshold A. Let $q_x = \mathbb{P}_1(\mathcal{A}_x)$. Then,

$$\mathbb{E}_1[\tau_{\mathrm{C}}(x)] = \mathbb{E}_1[\tau_{\mathrm{C}}(x); \mathcal{A}_x] + \mathbb{E}_1[\tau_{\mathrm{C}}(x); \mathcal{A}_x'] \le \mathbb{E}_1[\tau_{\mathrm{C}}].$$

Note that

$$\mathbb{E}_1[\tau_{\mathrm{C}}(x); \mathcal{A}_x'] = \mathbb{E}_1[\tau_{\mathrm{DC}}(x); \mathcal{A}_x'].$$

We call this common constant t_1. Also note that on $\mathcal{A}_x$, the average time taken to reach 0 is the same for both the CuSum and the DE-CuSum algorithm. We call this common average conditional delay by t_2. Thus,

$$\mathbb{E}_1[\tau_{\mathrm{C}}(x)] = (\mathsf{t}_1)(1 - q_x) + q_x(\mathsf{t}_2 + \mathbb{E}_1[\tau_{\mathrm{C}}]) \le \mathbb{E}_1[\tau_{\mathrm{C}}].$$

The equality in the above equation is true because, once the DE-CuSum statistic reaches zero, it is reset to zero and the average delay that point onwards is $\mathbb{E}_1[\tau_{\mathrm{C}}]$.

Then, for any $\mathsf{t}_3 \ge \mathbb{E}_1[\tau_{\mathrm{C}}]$, we have

$$(\mathsf{t}_1)(1 - q_x) + q_x(\mathsf{t}_2 + \mathsf{t}_3) \le \mathsf{t}_3.$$

This is because for $\mathsf{t}_3 \ge \mathbb{E}_1[\tau_{\mathrm{C}}]$

$$\begin{aligned}(\mathsf{t}_1)(1 - q_x)&+ q_x(\mathsf{t}_2 + \mathsf{t}_3)\\ &= (\mathsf{t}_1)(1 - q_x) + q_x(\mathsf{t}_2 + \mathbb{E}_1[\tau_{\mathrm{C}}] + \mathsf{t}_3 - \mathbb{E}_1[\tau_{\mathrm{C}}])\\ &\le \mathbb{E}_1[\tau_{\mathrm{C}}] + q_x(\mathsf{t}_3 - \mathbb{E}_1[\tau_{\mathrm{C}}])\\ &\le \mathsf{t}_3.\end{aligned}$$

It is easy to see that

$$\mathbb{E}_1[\tau_{\mathrm{DC}}(x)] \le (\mathsf{t}_1)(1 - q_x) + q_x(\mathsf{t}_2 + T_U^{(1)} + \mathbb{E}_1[\tau_{\mathrm{DC}}]).$$

This is because on $\mathcal{A}_x$, the average delay of the DE-CuSum algorithm is the average time to reach 0, which is t_2, plus the average time spent below 0 due to the undershoot, which is bounded from above by $T_U^{(1)}$, plus the average delay after the sojourn below 0, which is $\mathbb{E}_1[\tau_{\rm DC}]$. The latter is due to the renewal nature of the DE-CuSum algorithm. Since $T_U^{(1)} + \mathbb{E}_1[\tau_{\rm DC}] \geq \mathbb{E}_1[\tau_{\rm C}]$, the first part of lemma is proved if we set $\mathsf{t}_3 = T_U^{(1)} + \mathbb{E}_1[\tau_{\rm DC}]$.

For the second part, note that $T_U^{(1)} \leq \lceil h/\mu \rceil$. ■

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the associate editor for valuable comments and suggestions.

REFERENCES

[1] H. V. Poor and O. Hadjiliadis, *Quickest Detection*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[2] A. G. Tartakovsky, I. V. Nikiforov, and M. Basseville, *Sequential Analysis: Hypothesis Testing and Change-Point Detection*, ser. Statistics. Boca Raton, FL, USA: CRC Press, 2013.
[3] V. V. Veeravalli and T. Banerjee, Quickest Change Detection Elsevier, 2013, E-reference Signal Processing [Online]. Available: http://arxiv.org/abs/1210.5552
[4] A. N. Shiryaev, "On optimum methods in quickest detection problems," *Theory Probab. Appl.*, vol. 8, pp. 22–46, 1963.
[5] A. G. Tartakovsky and V. V. Veeravalli, "General asymptotic Bayesian theory of quickest change detection," *SIAM Theory Probab. Appl.*, vol. 49, pp. 458–497, Sep. 2005.
[6] G. Lorden, "Procedures for reacting to a change in distribution," *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, Dec. 1971.
[7] M. Pollak, "Optimal detection of a change in distribution," *Ann. Statist.*, vol. 13, pp. 206–227, Mar. 1985.
[8] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions," *Ann. Statist.*, vol. 14, pp. 1379–1387, Dec. 1986.
[9] Y. Ritov, "Decision theoretic optimality of the CUSUM procedure," *Ann. Statist.*, vol. 18, pp. 1464–1469, Nov. 1990.
[10] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *IEEE Trans. Inf. Theory*, vol. 44, no. 7, pp. 2917–2929, Nov. 1998.
[11] A. G. Tartakovsky, M. Pollak, and A. Polunchenko, "Third-order asymptotic optimality of the generalized Shiryaev-Roberts change-point detection procedures," *ArXiv E-Prints*, May 2010.
[12] S. W. Roberts, "A comparison of some control chart procedures," *Technometrics*, vol. 8, pp. 411–430, Aug. 1966.
[13] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, Jun. 1954.
[14] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection with on-off observation control," *Sequential Anal.*, vol. 31, pp. 40–77, Feb. 2012.
[15] G. Tagaras, "A survey of recent developments in the design of adaptive control charts," *J. Quality Technol.*, vol. 30, pp. 212–231, Jul. 1998.
[16] Z. G. Stoumbos, M. R. Reynolds, T. P. Ryan, and W. H. Woodall, "The state of statistical process control as we proceed into the 21st century," *J. Amer. Statist. Assoc.*, vol. 95, pp. 992–998, Sep. 2000.
[17] K. Premkumar and A. Kumar, "Optimal sleep-wake scheduling for quickest intrusion detection using wireless sensor networks," in *Proc. IEEE Conf. Comput. Commun.*, Apr. 2008, pp. 1400–1408.
[18] J. Geng, L. Lai, and E. Bayraktar, "Quickest change point detection with sampling right constraints," presented at the Annu. Allerton Conf. Commun., Control, Comput., Sep. 2012.
[19] V. Krishnamurthy, "When to look at a noisy Markov chain in sequential decision making if measurements are costly?," *ArXiv E-Prints*, Aug. 2012.
[20] D. Siegmund, *Sequential Analysis: Tests and Confidence Intervals*, ser. (Springer series in statistics). New York, NY, USA: Springer-Verlag, 1985.
[21] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Statist.*, vol. 19, no. 3, pp. 326–339, 1948.
[22] M. Woodroofe, *Nonlinear Renewal Theory in Sequential Analysis*, ser. (CBMS-NSF regional conference series in applied mathematics). Philadelphia, PA, USA: SIAM, 1982.
[23] T. Banerjee and V. V. Veeravalli, "Energy-efficient quickest change detection in sensor networks," in *Proc. IEEE Statist. Signal Process. Workshop*, Aug. 2012, pp. 636–639.
[24] T. Banerjee and V. V. Veeravalli, "Data-efficient quickest change detection in distributed and multichannel systems," presented at the IEEE Int. Conf. Acoust., Speech, Signal Process., May 2013.
[25] T. Banerjee, V. V. Veeravalli, and A. Tartakovsky, "Decentralized data-efficient quickest change detection," presented at the IEEE Int. Symp. Inf. Theory, Jul. 2013.

Taposh Banerjee (S'12) received his M.E. in Telecommunications from the ECE Department of the Indian Institute of Science, Bangalore. He is now pursuing his Ph.D. at the Coordinated Science Laboratory and the ECE Department at the University of Illinois at Urbana-Champaign. His research interests are in detection and estimation theory, sequential analysis and wireless communications and networks.

Venugopal V. Veeravalli (M'92–SM'98–F'06) received the B.Tech. degree (Silver Medal Honors) from the Indian Institute of Technology, Bombay, in 1985, the M.S. degree from Carnegie Mellon University, Pittsburgh, PA, in 1987, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1992, all in electrical engineering.

He joined the University of Illinois at Urbana-Champaign in 2000, where he is currently a Professor in the department of Electrical and Computer Engineering and the Coordinated Science Laboratory. He served as a Program Director for communications research at the U.S. National Science Foundation in Arlington, VA from 2003 to 2005. He has previously held academic positions at Harvard University, Rice University, and Cornell University, and has been on sabbatical at MIT, IISc Bangalore, and Qualcomm, Inc. His research interests include distributed sensor systems and networks, wireless communications, detection and estimation theory, including quickest change detection, and information theory.

Prof. Veeravalli was a Distinguished Lecturer for the IEEE Signal Processing Society during 2010–2011. He has been an Associate Editor for Detection and Estimation for the IEEE TRANSACTIONS ON INFORMATION THEORY and for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. Among the awards he has received for research and teaching are the IEEE Browder J. Thompson Best Paper Award, the National Science Foundation CAREER Award, and the Presidential Early Career Award for Scientists and Engineers (PECASE).