

DATA-EFFICIENT MINIMAX QUICKEST CHANGE DETECTION

Taposh Banerjee and Venugopal V. Veeravalli

Dept of Electrical and Computer Engineering
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign

ABSTRACT

In [1], a Bayesian two-threshold algorithm was obtained for quickest detection of a change in the distribution of a sequence of random variables, subject to constraints of probability of false alarm and observation cost. This algorithm was shown to be asymptotically optimal and to have good trade-off curves. In this paper, the results in [1] are extended to the more practically relevant minimax setting. Motivated by the structure of the algorithm developed in [1], a CUSUM based algorithm, called DE-CUSUM is proposed, which can be used for on-off observation control and to detect change as quickly as possible subject to a false alarm constraint. It is shown that the DE-CUSUM algorithm inherits the good qualities of the algorithm in [1], i.e., it is also asymptotically optimal and has good trade-off curves. Numerical results show that the DE-CUSUM algorithm provides a substantial savings in the observation cost over the naive approach of fractional sampling.

Index Terms— Change point detection, observation control, energy-efficient sensing, CUSUM.

1 Introduction

In the classical problem of quickest change detection [2], [3], a change in the distribution of a sequence of random variables has to be detected as soon as possible, subject to a constraint on the probability of false alarm. In many engineering applications of quickest change detection, e.g., statistical quality control, sensor networks etc, there is a cost associated with acquiring information or taking observations. In [1], we considered the problem of quickest change detection under constraints on the observation cost and probability of false alarm in the Bayesian setting. We captured the observation cost through the average number of observations used before the change point and designed a two-threshold algorithm for on-off observation control. Observations are taken only if the *a posteriori* probability is above a threshold B , and change is declared the first time the *a posteriori* probability crosses another threshold $A > B$. We showed that this algorithm is asymptotically optimum, i.e., for a fixed constraint on the observation cost, as the probability of false alarm goes to zero, the performance of the two-threshold algorithm approaches that of the Shiryaev algorithm, which is optimum for the case where all the observations are used for detecting the

change [3]. The two-threshold algorithm was also shown to have good delay-observation cost trade-off curves: for moderate values of probability of false alarm, the delay of the algorithm is within 10% of the Shiryaev delay even when the observation cost is reduced by more than 50%.

In most practical applications, prior information about the distribution of the change point is not available. As a result, the Bayesian solution is not directly applicable. Our goal in this paper is to obtain a simple non-Bayesian data-efficient quickest change detection algorithm which has some optimality properties and has good performance.

For the classical quickest change detection problem, an algorithm for the non-Bayesian setting was obtained by taking the geometric parameter of the prior on the change point to zero [4]. Such a technique cannot be used in the data-efficient setting. This is because when an observation is 'skipped' in the two-threshold algorithm in [1], the *a posteriori* probability is updated using the geometric prior. In the absence of prior information about the distribution of the change point, it is by no means obvious what the right substitute for the prior is. But, we note that the duration for which observations are not taken in the algorithm in [1], is also a function of the 'undershoot' of the *a posteriori* probability when it goes below the threshold B . We show in this paper that this fact can be used to design a good test in the non-Bayesian setting.

In the following, we define P_n to be the probability measure when change happens at time n , and E_n to be the corresponding expectation. The notation P_∞ and E_∞ is used when the entire observation sequence is i.i.d. with density f_0 .

2 Bayesian formulation

We begin with a review of the Bayesian setting and the algorithm in [1], which we call the DE-Shiryaev algorithm. Let $\{X_n\}$ be a sequence of random variables whose distribution changes at a random time Γ . Before Γ , the $\{X_n\}$'s are independent and identically distributed (i.i.d.) with density f_0 , after Γ they are i.i.d. with density f_1 and Γ is geometrically distributed with parameter ρ .

In order to minimize the average number of observations used before Γ , at each time instant, a decision is made on whether to use the observation in the next time step, based on all the available information. Let $S_k \in \{0, 1\}$, with $S_k = 1$ if

it is been decided to take the observation at time k , i.e. X_k is available for decision making, and $S_k = 0$ otherwise. Thus, S_k is an on-off (binary) control input based on the information available up to time $k - 1$, i.e.,

$$S_k = \mu_{k-1}(I_{k-1}), \quad k = 1, 2, \dots$$

with μ denoting the control law and I defined as:

$$I_k = [S_1, \dots, S_k, X_1^{(S_1)}, \dots, X_k^{(S_k)}].$$

Here, $X_i^{(S_i)}$ represents X_i if $S_i = 1$, otherwise X_i is absent from the information vector I_k .

Let $\gamma = \{\tau, \mu_0, \dots, \mu_{\tau-1}\}$ represent a policy for data-efficient quickest change detection, where τ is a stopping time on the information sequence $\{I_k\}$. The objective in [1] is to solve the following optimization problem:

$$\begin{aligned} & \underset{\gamma}{\text{minimize}} && \text{ADD}(\gamma) = \mathbb{E}[(\tau - \Gamma)^+], & (1) \\ & \text{subject to} && \text{PFA}(\gamma) = \mathbb{P}(\tau < \Gamma) \leq \alpha, \\ & \text{and} && \text{ANO}(\gamma) = \mathbb{E}\left[\sum_{k=1}^{\min(\tau, \Gamma-1)} S_k\right] \leq \beta. \end{aligned}$$

Here, ADD, PFA and ANO stand for average detection delay, probability of false alarm and average number of observations used, respectively, and α and β are given constraints.

Define,

$$p_k = \mathbb{P}\{\Gamma \leq k \mid I_k\}.$$

Then, the two-threshold algorithm from [1] is:

Algorithm 1 (DE-Shiryayev: $\gamma(A, B)$). *Start with $p_0 = 0$ and use the following control, with $B < A$, for $k \geq 0$:*

$$\begin{aligned} S_{k+1} = \mu_k(p_k) &= \begin{cases} 0 & \text{if } p_k < B \\ 1 & \text{if } p_k \geq B \end{cases} & (2) \\ \tau &= \inf \{k \geq 1 : p_k > A\}. \end{aligned}$$

The probability p_k is updated using the following recursions:

$$p_{k+1} = \begin{cases} \tilde{p}_k = p_k + (1 - p_k)\rho & \text{if } S_{k+1} = 0 \\ \frac{\tilde{p}_k L(X_{k+1})}{\tilde{p}_k L(X_{k+1}) + (1 - \tilde{p}_k)} & \text{if } S_{k+1} = 1 \end{cases}$$

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

With $B = 0$ the DE-Shiryayev algorithm reduces to the Shiryayev algorithm. It is shown in [1] that the PFA and ADD of the DE-Shiryayev algorithm approach that of the Shiryayev algorithm as $\alpha \rightarrow 0$.

When Algorithm 1 is employed, the probability p_k typically evolves as depicted in Fig. 1. As observed in [1], when $p_k < B$, p_k increases monotonically. This is because when an observation is skipped, p_k is updated using the prior ρ . Thus, the duration for which the observations are skipped depends on both the *undershoot* of p_k , when p_k goes below the threshold B , and also on the value of prior ρ . Based on this observation we propose a non-Bayesian algorithm in Section 3.

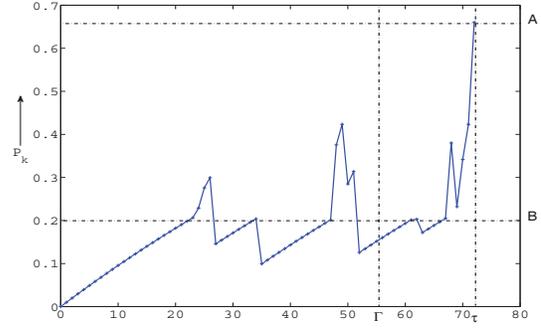


Fig. 1: Evolution of p_k for $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.5, 1)$, and $\rho = 0.01$, with thresholds $A = 0.65$ and $B = 0.2$.

3 Minimax formulation and DE-CUSUM

In order to extend the work in [1] to a minimax setting, we first propose a different way to capture the observation cost. Note that in the Bayesian setting, due to the geometric prior,

$$\text{ANO} \leq \frac{1}{\rho}.$$

Such a bound is not possible in the absence of the prior. But, we observe that the control employed in Algorithm 1, essentially controls the fraction of observations used before change point in the long run. Let I_k , τ , and γ be as defined earlier. We propose the following duty cycle based observation cost function, Pre-change Duty Cycle (PDC):

$$\text{PDC} = \limsup_n \frac{1}{n} \mathbb{E}_n \left[\sum_{k=1}^{n-1} S_k \mid \tau \geq n \right]. \quad (3)$$

For delay and false alarm rate, we use the minimax setting of Pollak [4]: the supremum of the conditional delay

$$\text{CADD}(\gamma) = \sup_n \mathbb{E}_n [\tau - n \mid \tau \geq n],$$

and the false alarm rate

$$\text{FAR} = \frac{1}{\mathbb{E}_\infty[\tau]}.$$

Our objective is to solve the following optimization problem.

$$\begin{aligned} & \underset{\gamma}{\text{minimize}} && \text{CADD}(\gamma), \\ & \text{subject to} && \text{FAR}(\gamma) \leq \zeta, \text{ and } \text{PDC}(\gamma) \leq \eta. \end{aligned} \quad (4)$$

Here, $0 \leq \zeta, \eta \leq 1$. In (4), we have implicitly restricted the search over those control policies for which the CADD, FAR, and PDC are well defined.

It is difficult to solve (4) directly and obtain an exact solution (even with $\eta = 1$). But, it is known that the CUSUM algorithm is asymptotically optimal (as $\zeta \rightarrow 0$) for $\eta = 1$ [5].

Also, the CUSUM algorithm is exactly optimal with respect to another closely related non-Bayesian criterion proposed by Lorden [5]. Moreover, it is well known that the CUSUM algorithm can also be used in a non-parametric setting. For these reasons, we base our data-efficient quickest change detection algorithm on the CUSUM algorithm. We will show in the next section that the DE-CUSUM algorithm is also asymptotically optimal, with the same asymptotic delay as that of the CUSUM algorithm, for each fixed η , as $\zeta \rightarrow 0$.

Algorithm 2 (DE-CUSUM: $\gamma(d, \mu, h)$). Start with $W_0 = 0$ and fix $\mu > 0$, $d > 0$ and $h \geq 0$. For $k \geq 0$ use the following control:

$$S_{k+1} = \begin{cases} 0 & \text{if } W_k < 0 \\ 1 & \text{if } W_k \geq 0 \end{cases}$$

$$\tau_w(d) = \inf \{k \geq 1 : W_k > d\}.$$

The statistic W_k is updated using the following recursions:

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

where $L(X_{k+1}) = f_1(X_{k+1})/f_0(X_{k+1})$, and

$$(x)^{h+} = \begin{cases} \max\{x, 0\} & \text{if } x > -h \\ x & \text{otherwise} \end{cases}$$

If $h = \infty$, the DE-CUSUM algorithm reduces to the CUSUM algorithm [2]: $W_0 = 0$ and for $k \geq 0$,

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

We use $\nu(d)$ for the stopping time of the CUSUM algorithm.

When $h = 0$, the DE-CUSUM algorithm works as follows. The statistic W_k starts at 0, and evolves according to the CUSUM algorithm till it goes below 0. When W_k goes below zero, it does so with an undershoot. Beyond this, W_k is incremented deterministically (by using the recursion $W_{k+1} = W_k + \mu$), and observations are skipped till W_k crosses 0 from below. As a consequence, the number of observations that are skipped is determined by the undershoot as well as the parameter μ . Thus, μ is a substitute for the Bayesian prior ρ . When W_k crosses 0 from below, it is reset to 0. Once $W_k = 0$, the process renews itself and continues to evolve this way until $W_k > d$, at which time a change is declared. If $h = 0$, we note that it may not be possible to achieve PDC values close to 1. If $h \neq 0$, samples are skipped only if the undershoot is greater than h . This ensures achievability of arbitrary PDC values. The evolution of DE-CUSUM is plotted in Fig. 2.

Apart from being a non-Bayesian version of the DE-Shiryayev algorithm, the DE-CUSUM algorithm, like the CUSUM algorithm, has two interesting interpretations. First, it can be seen as a sequence of independent two-sided tests.

In each two-sided test a Sequential Probability Ratio Test (SPRT) is used to distinguish between the two hypotheses 'pre-change' and 'post-change'. If the SPRT stops to decide on 'pre-change', then samples are skipped based on the likelihood ratio of all the observations taken in the SPRT. Second, the DE-CUSUM algorithm has a maximum likelihood change point interpretation. At time n , the CUSUM algorithm is a maximum likelihood test to distinguish between $n + 1$ alternatives, $\{\Gamma = i\}, 1 \leq i \leq n$ and $\{\Gamma > n\}$. The DE-CUSUM algorithm does the same, with the difference that the maximum likelihood statistic is computed using partial set of observations and samples are skipped if the maximum likelihood statistic is below 0.

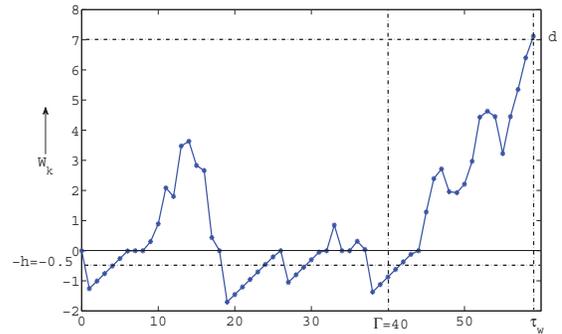


Fig. 2: Evolution of W_k for $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.75, 1)$, and $\Gamma = 40$, with $d = 7$, $h = 0.5$, and $\mu = 0.25$.

4 Asymptotic optimality of DE-CUSUM

In this section we prove the asymptotic optimality of the DE-CUSUM algorithm using techniques introduced in [1]. We first show that we can select a μ and an h so as to meet any PDC constraint. For this define the following random variable

$$\lambda \triangleq \inf \{k \geq 1 : W_k \notin (0, d), W_0 = 0\}.$$

If R is the number of samples skipped if $W_\lambda < d$, then by Wald's lemma and the renewal reward theorem (also see [1]), we have

$$\text{PDC}(\gamma(d, \mu, h)) = \frac{\mathbb{E}_\infty[\lambda | W_\lambda \leq 0]}{\mathbb{E}_\infty[\lambda | W_\lambda \leq 0] + \mathbb{E}_\infty[R]}.$$

Note that $\mathbb{E}_\infty[\lambda | W_\lambda \leq 0]$ is not affected by the choice of h and μ . Also, we can select h such that $\mathbb{E}_\infty[R] \rightarrow \infty$ as $\mu \rightarrow 0$, and for a fixed μ , $\mathbb{E}_\infty[R] \rightarrow 0$ as $h \rightarrow \infty$. Thus, any $\text{PDC} \in [0, 1]$ is achievable.

We now prove the asymptotic optimality of the DE-CUSUM algorithm. Define,

$$\Lambda \triangleq \lambda + R \mathbb{I}\{W_\lambda < d\}. \quad (5)$$

Theorem 1. For any fixed value of d , μ and h ,

$$\text{FAR}(\gamma(d, \mu, h)) \leq \text{FAR}(\nu(d)).$$

Proof. By Wald's lemma [1],

$$E_\infty[\nu(d)] = \frac{E_\infty[\lambda]}{P_\infty[W_\lambda > d]} \leq \frac{E_\infty[\Lambda]}{P_\infty[W_\lambda > d]} = E_\infty[\tau_w(d)].$$

□

We now show that the CADD for the DE-CUSUM algorithm and the CUSUM algorithm are approximately equal as $d \rightarrow \infty$.

Theorem 2. For any fixed value of μ and h ,

$$\text{CADD}(\gamma(d, \mu, h)) = \text{CADD}(\nu(d)) \left(1 + o(1)\right) \text{ as } d \rightarrow \infty.$$

Proof. From Theorem 1 and asymptotic optimality of CUSUM, we have

$$\text{CADD}(\gamma(d, \mu, h)) \geq \text{CADD}(\nu(d)) \left(1 + o(1)\right) \text{ as } d \rightarrow \infty.$$

Based on the 'resetting' arguments given in [1], it can be shown that,

$$E_n[\tau_w(d) - n | \tau_w(d) \geq n] \leq E_n[T_n] + \frac{E_1[\Lambda]}{P_1(W_\lambda > d)}.$$

Here, T_n is a positive random variable which can be bounded by a random variable which is not a function of the threshold d and time n . It follows that

$$\text{CADD}(\gamma(d, \mu, h)) \leq \frac{E_1[\Lambda]}{P_1(W_\lambda > d)} \left(1 + o(1)\right) \text{ as } d \rightarrow \infty.$$

The theorem follows because,

$$\frac{E_1[\Lambda]}{P_1(W_\lambda > d)} = \frac{E_1[\lambda]}{P_1(W_\lambda > d)} \left(1 + o(1)\right) \text{ as } d \rightarrow \infty,$$

and for any d , $\text{CADD}(\nu(d)) = \frac{E_1[\lambda]}{P_1(W_\lambda > d)}$. □

The theorems above taken together imply the asymptotic optimality of the DE-CUSUM algorithm.

5 Trade-off curves

The asymptotic optimality of the DE-CUSUM algorithm for all η does not guarantee good performance for moderate values of FAR. In Fig. 3, we plot the CADD-PDC trade-off curves for the DE-CUSUM and the CUSUM algorithms using simulations, for two different values of PDC constraints: $\eta = 0.5$ and $\eta = 0.25$. For simplicity we restrict ourself to $h = 0$ in this section. Clearly, one can do only better by having additional degree of flexibility, i.e., by using $h \neq 0$. For comparison, we also plot the trade-off curves for the *fraction sampling* scheme, in which, to achieve a PDC of η , the CUSUM algorithm is employed, and a sample is chosen with probability η for decision making. Note that this scheme saves samples without exploiting any knowledge about the state of the system. 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 very 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 6 slots in Fig. 3). But, note that the difference of delay with the CUSUM algorithm remains fixed as $\text{FAR} \rightarrow 0$. This is the result of asymptotic optimality of the DE-CUSUM algorithm. Note also from Fig. 3 the significant reduction in the number of observations used as compared to the fraction sampling scheme. Thus, the trade-off curves show that the DE-CUSUM algorithm has good performance even for moderate FAR, when the PDC constraint is moderate.

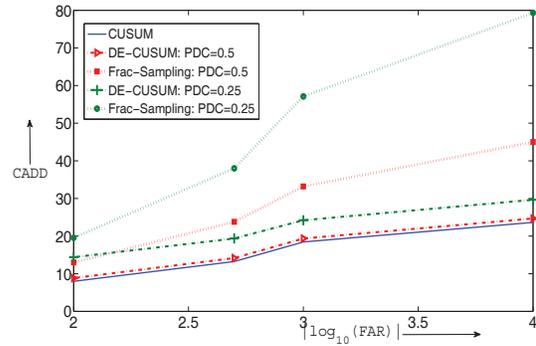


Fig. 3: Trade-off curves for DE-CUSUM for PDC = 0.25, 0.5, with $f_0 \sim \mathcal{N}(0, 1)$ and $f_1 \sim \mathcal{N}(0.75, 1)$.

6 Conclusions and future work

We proposed a data-efficient non-Bayesian quickest change detection algorithm DE-CUSUM, and showed that its performance approaches that of the CUSUM algorithm as the false alarm rate goes to zero. We also showed that the DE-CUSUM algorithm has good trade-off curves and provides substantial benefits over the naive approach of fractional sampling. It is not difficult to see that the DE-CUSUM algorithm can also be applied for data-efficient quickest change detection in a system where the observations arrive at the decision maker in batches, e.g. a sensor network, and the distributions of all the random variables in the batch change at the change point. The DE-CUSUM algorithm can then be applied by serializing the observation sequence in each batch and then treating the batch observation sequence as a single observation sequence.

7 References

- [1] T. Banerjee and V. V. Veeravalli, "Data-Efficient Quickest Change Detection with On-Off Observation Control", Accepted in Sequential Analysis, <http://arxiv.org/abs/1105.1361>.
- [2] E. S. Page, "Continuous inspection schemes," *Biometrika*, 41, pp. 100-115, 1954.
- [3] A. N. Shiryaev, "On optimal methods in quickest detection problems," *Theory Probab. Appl.*, Vol. 8, pp. 22-46, 1963.
- [4] M. Pollak, "Detection of a change in distribution," *The Annals of Statistics*, Vol. 13, No. 1, pp. 206-227, 1985.
- [5] T. L. Lai, "Information bounds and quick detection of parameter changes in stochastic systems," *Information Theory, IEEE Trans. on*, vol.44, no.7, pp.2917-2929, Nov 1998.