ADAPTIVE QUICKEST CHANGE DETECTION WITH UNKNOWN PARAMETER

Chengzhi Li and Huaiyu Dai[*]

Dept. of electrical and computer engineering
North Carolina State University, Raleigh, NC
email: {cli3,hdai}@ncsu.edu

Husheng Li

Dept. of electrical engineering and
computer science
University of Tennessee, Knoxville, TN
email: husheng@ece.utk.edu

ABSTRACT

Quickest detection of an abrupt distribution change with an unknown time varying parameter is considered. A novel adaptive approach is proposed to tackle this problem, which is shown to outperform the celebrated Parallel CUSUM Test. Performance is evaluated through theoretical analysis and numerical simulations.

***Index Terms*—** Quickest detection, CUSUM test, unknown parameter

1. INTRODUCTION

Quickest detection is a technique to detect distribution changes as quickly as possible based on sequential observations [1]. It admits a wide range of applications such as quality control, medical diagnosis and intrusion detection. When both pre-change and post-change distributions are completely specified, many detection procedures have been proposed under different criterias. One well-known procedure is the Cumulative Sum (CUSUM) test proposed by Page in [2]. Lorden showed that Page's CUSUM Test is asymptotically optimal for independent observations [3], and Lai extended this study to dependent observations [4]. In many practical situations, however, the post-change distribution involves unknown parameters. The Generalized Likelihood Ratio (GLR) Test [3] is an optimal procedure to tackle such problems, but unbounded memory requirement makes it infeasible in practice. To improve efficiency in storage and computation, Nikiforov proposed the Parallel CUSUM Test [5], in which multiple Page's CUSUM Tests are carried out simultaneously on some specifically chosen values of the unknown parameters.

In this paper, we propose an adaptive CUSUM algorithm for quickest detection when there is an unknown parameter in the post-change distribution and this parameter may be varying during the detection process. Our new approach can favorably narrow down the range of the unknown parameter and track its change adaptively, thus achieves significant performance improvement over the Parallel CUSUM Test.

*This work was supported in part by the US National Science Foundation under Grant CCF-0515164, CNS-0721815, CCF-0830451 and CCF-0830462.

The remainder of this paper is organized as follows. The system model is given in Section 2. After discussing existing approaches to the unknown-parameter problem in Section 3, we propose the adaptive CUSUM algorithm in Section 4, together with some performance analysis. The simulation results and conclusions are provided in Section 5 and Section 6, respectively.

2. SYSTEM MODEL

Suppose a sensor is monitoring some property in the environment. Denote by $s(t)(t = 1, 2, ...)$ its (independent) observation at time slot t, whose probability density belongs to a single-parameter exponential family $\{p_\theta\}_{\theta\in\Theta}$ with natural parameter space Θ, defined by:

$$p_\theta(x) = h(x)\exp(\theta T(x) - A(\theta)), \tag{1}$$

where $T(x)$ is a sufficient statistic and $A(\theta)$ is a normalization factor. We assume the distribution of $s(t)$ is changed from p_λ to p_φ at some unknown time instant v, where φ is unknown but lies within a given range $\Phi \triangleq [\varphi_{min}, \varphi_{max}]$, and λ is a known value outside Φ. Therefore, two hypotheses of interest are:

$$\begin{cases} H_0 : \theta = \lambda \notin \Phi \\ H_1 : \theta = \varphi \in \Phi. \end{cases} \tag{2}$$

Correspondingly, the log likelihood ratio (LLR) is defined as:

$$l_\varphi(t) = \log\left(\frac{p_\varphi(s(t))}{p_\lambda(s(t))}\right). \tag{3}$$

For performance measurement, we consider detection delay and mean time between false alarms [3], which are defined respectively as[1]

$$\begin{aligned} \bar{T}_1 &= \sup_{v\geqslant 1} \operatorname{esssup} E_v\left[(T^* - v + 1)^+ | s(1), ..., s(v-1)\right], && (4)\\ \bar{T}_0 &= E_\infty\left[T^*\right], && (5) \end{aligned}$$

where T^* is the stopping time determined by the detection algorithms. E_v denotes the expectation under the assumption that the change happens at time slot v ($t = \infty$ means that the change never happens).

[1] *esssup* refers to the worst case of all the pre-change distributions and $[x]^+ = \max(x, 0)$.

3. EXISTING ALGORITHMS

3.1. Page's CUSUM Test

With full knowledge about the pre-change and post-change distributions, the Page's CUSUM Test provides an optimal scheme minimizing the worst-case detection delay in (4) [3]. Specifically, the stopping time T^* in the Page's CUSUM Test is given by

$$T^* = \inf \left\{ t \middle| \max_{1 \leqslant k \leqslant t} \sum_{r=k}^{t} l_\varphi(r) \geq h \right\}, \tag{6}$$

where h is a predetermined threshold and the metric $l_\varphi(r)$ is defined in (3). An alternative expression of T^* is:

$$T^* = \inf \left\{ t \middle| s(t) \geq h \right\}, \tag{7}$$

with $s(t) = \max(s(t-1) + l_\varphi(t), 0)$ (and $s(0) = 0$). It is equivalent to the above one while allows for more efficient computation and memory usage due to recursion.

3.2. GLR Test and Parallel CUSUM Test

The unknown parameter in the post-change distribution makes the detection more challenging. Two well known algorithms in literature are the GLR Test and Parallel CUSUM Test. In the GLR Test the stopping time is given by:

$$T^* = \inf \left\{ t \middle| \max_{1 \leqslant k \leqslant t} \sup_{\varphi \in \Phi} \sum_{r=k}^{t} l_\varphi(r) \geq h \right\},$$

where (c.f. (6)) the $\sup$ operation implicates an optimal estimation for the parameter. An obvious drawback of the GLR Test is its demanding computation cost and memory requirement (as no recursive expression is available).

In contrast to the GLR Test the Parallel CUSUM Test [5] is suboptimal but more efficient. Instead of online estimation of the unknown parameter, the Parallel CUSUM Test carries out a collection of CUSUM Tests over L specially chosen values of φ, denoted by $\varphi_1, \varphi_2, ..., \varphi_L$. The stopping time is given by:

$$T^* = \inf \{T_i, i = 1, 2, ..., L\},$$

where T_i is computed by (7) with $\varphi = \varphi_i$. Its suboptimality lies in the fact that the candidates remain unchanged during the detection process so that the inaccuracy of the parameters affects the performance throughout the process. Fairly large L may be needed to achieve satisfatory detection delay $\bar{T}_1$, which still incurs substantial compuation burden and impacts negatively the mean time between false alarms $\bar{T}_0$.

4. ADAPTIVE CUSUM ALGORITHM

Much remains to be done for quickest detection with an unknown parameter. Especially study is still lacking on the scenarios where the unknown parameter varies over time, e.g., due to channel fading. In such situations, all algorithms originally designed with the assumption of fixed parameter (such as those discussed in Section 3) will degrade. In this section, we propose a new quickest detection algorithm, Adaptive CUSUM Test, which achieves a better tradeoff between performance and complexity and performs stably in changing enviroments.

Our algorithm is recursive in nature, with each recursion comprising two main *interleaved* steps: parameter tracking and CUSUM test. Our parameter tracking approach can in principle be applied to a generic parameterized distribution. For rigorous presentation, we restrict our attention to the exponential family below.

4.1. Parameter Tracking

We denote an estimate of φ by $\hat{\varphi}$ and define[2] $F(\hat{\varphi}) = E[l_{\hat{\varphi}}(t)]$, the mismatched Kullback-Leibler (KL) divergence between p_φ and p_λ. We obtain the following result:

Proposition 1 *For distributions in the exponential family, $F(\hat{\varphi})$ is a strictly concave function and achieves the global maximum at $\hat{\varphi} = \varphi$.*

Proof:

$$\begin{aligned} F(\hat{\varphi}) &= E\left[\log \frac{p_{\hat{\varphi}}(s(t))}{p_\lambda(s(t))}\right] \\ &= E\left[\log \frac{p_{\varphi}(s(t))}{p_\lambda(s(t))}\right] - E\left[\log \frac{p_{\varphi}(s(t))}{p_{\hat{\varphi}}(s(t))}\right] \\ &= E[l_\varphi(t)] - D(p_\varphi \parallel p_{\hat{\varphi}}), \end{aligned} \tag{8}$$

where $D(p_\varphi \parallel p_{\hat{\varphi}})$ is the KL divergence between p_φ and $p_{\hat{\varphi}}$. Since $D(p_\varphi \parallel p_{\hat{\varphi}})$ is nonnegative [6], $F(\hat{\varphi})$ achieves the global maximum when $\hat{\varphi} = \varphi$.

Substituting (1) in (8) and taking derivation, we have

$$\begin{aligned} \frac{dF(\hat{\varphi})}{d\hat{\varphi}} &= -\frac{dD(p_\varphi \parallel p_{\hat{\varphi}})}{d\hat{\varphi}} \\ &= -E\left[\frac{d}{d\hat{\varphi}}\Big((\varphi - \hat{\varphi})T(x) - (A(\varphi) - A(\hat{\varphi}))\Big)\right] \\ &= E[T(x)] - \frac{dA(\hat{\varphi})}{d\hat{\varphi}}, \end{aligned}$$

and $\frac{d^2 F(\hat{\varphi})}{d^2 \hat{\varphi}} = -\frac{d^2 A(\hat{\varphi})}{d^2 \hat{\varphi}}$. According to the differential identities of $A(\varphi)$ [7], $\frac{d^2 A(\hat{\varphi})}{d^2 \hat{\varphi}} = Var(T(x)) > 0$, therefore $F(\hat{\varphi})$ is a strictly concave function.

□

According to Prop. 1 we can draw the following conclusions:

- given a small value δ, we can always find two estimates $\hat{\varphi}_a, \hat{\varphi}_b$ such that $\hat{\varphi}_b = \hat{\varphi}_a + \delta$ and $F(\hat{\varphi}_a) = F(\hat{\varphi}_b)$;
- φ lies within the interval $\hat{\Phi} \triangleq (\hat{\varphi}_a, \hat{\varphi}_b)$.

Intuitively, we can narrow down the range of φ from its original range Φ to $\hat{\Phi}$; this also allows parameter tracking in time-varying environments. An iterative procedure can be used to find $\hat{\varphi}_a$ and $\hat{\varphi}_b$. Given δ, one begins by arbitrarily choosing $\hat{\varphi}_a$ and $\hat{\varphi}_b$ within Φ, say $\hat{\varphi}_a^0$ and $\hat{\varphi}_b^0$. And then the succeeding values of $\hat{\varphi}_a$ and $\hat{\varphi}_b$ are obtained according to the recursion:

[2] Here the expectation is with respect to p_φ.

$$\hat{\varphi}_a^{k+1} = \hat{\varphi}_a^k + \xi \mathcal{D}^k, \quad (9)$$
$$\hat{\varphi}_b^{k+1} = \hat{\varphi}_a^{k+1} + \delta, \quad (10)$$

where $\hat{\varphi}_a^k$ and $\hat{\varphi}_b^k$ represent the values of $\hat{\varphi}_a$ and $\hat{\varphi}_b$ at the kth iteration, $\mathcal{D}^k$ is the difference between $F(\hat{\varphi}_b^k)$ and $F(\hat{\varphi}_a^k)$ at the kth iteration, given by:

$$\mathcal{D}^k = F(\hat{\varphi}_b^k) - F(\hat{\varphi}_a^k) = E\left[\log \frac{p_{\hat{\varphi}_b^k}}{p_{\hat{\varphi}_a^k}}\right],$$

and ξ is a step size controlling the rate of adjustment. In practice, it is a common approach to replace the ensemble average $\mathcal{D}^k$ by the time average

$$\hat{\mathcal{D}}^k = l_{\hat{\varphi}_b^k}(k) - l_{\hat{\varphi}_a^k}(k) = \log \frac{p_{\hat{\varphi}_b^k}(s(k))}{p_{\hat{\varphi}_a^k}(s(k))},$$

or a block updating version of it. Convergence analysis is straightforward. If $\mathcal{D}^k > 0$, φ_a^{k+1} and φ_b^{k+1} will grow according to (9) and (10), so that $\mathcal{D}^{k+1}$ will decrease due to the concavity of the function. Similarly, if $\mathcal{D}^k < 0$, φ_a^{k+1} and φ_b^{k+1} will move back so that $\mathcal{D}^{k+1}$ will increase. In both cases, $\mathcal{D}^k$ converges to zero surely. δ is a key factor in our algorithm. On the one hand larger δ leads to faster convergence. On the other hand, δ is the range for the unknown parameter, which is desired to be small. We give an approach to set δ below.

4.2. CUSUM test

In the process of tracking the unknown parameter, we can conduct change detection through an appropriate CUSUM test simultaneously. What we need is to determine an estimate within the new range of each iteration. Define a non-optimality coefficient ε_φ as

$$\varepsilon_\varphi = 1 - \frac{\bar{T}_1^{opt}}{\bar{T}_1^{\varphi}},$$

where $\bar{T}_1^{opt}$ is the optimal detection delay given by the Page's CUSUM Test when φ is known, and $\bar{T}_1^{\varphi}$ is the detection delay achieved by some detection procedure when the true value φ is unknown (such as the Parallel CUSUM Test or our Adaptive CUSUM Test). In the Parallel CUSUM Test [5], given a threshold on the maximum non-optimality coefficient, $\varepsilon_m \triangleq \sup_{\varphi \in [\varphi_{min}, \varphi_{max}]} \varepsilon_\varphi$, we can predetermine L candidates $\{a_j\}$ and confidence intervals[3] $[\underline{a_j}, \bar{a_j}]$ associated with these candidates . δ can be chosen as the minimal confidence interval, i.e.,

$$\delta = \min_{1 \le j \le L} (\bar{a_j} - \underline{a_j}).$$

After (9) and (10), the value used in the CUSUM test is decided as:

$$\bar{\varphi}^k = \varphi_a^k + \alpha\delta,$$

where $\alpha = 1/2$ when $F(\hat{\varphi})$ is symmetric with respect to φ, or $\alpha = \min_{1 \le j \le L}(a_j - \underline{a_j})/\delta$ when $F(\hat{\varphi})$ is asymmetric. If $\bar{\varphi}^k$ is out of the range $[\varphi_{min}, \varphi_{max}]$, it is set to φ_{min} or φ_{max}, whichever is closer. Our algorithm continues by substituting $\bar{\varphi}^k$ as the true value of φ in the CUSUM test. $\bar{\varphi}^k$ converges to some value $\bar{\varphi}$ when the parameter tracking procedure converges.

Under H_1, $\bar{\varphi} = \varphi$ for the symmetric $F(\hat{\varphi})$ since when the algorithm converges, φ lies right in the middle of $\hat{\varphi}_a$ and $\hat{\varphi}_b$. In such a case this procedure provides an adaptive estimation for the unknown parameter. For the asymmetric $F(\hat{\varphi})$, $\bar{\varphi} \neq \varphi$ but it can be proved[4] that $\varepsilon_\varphi < \varepsilon_m$; and in practice ε_φ is often much smaller than ε_m, as shown by the simulation result in the next section. The detection delay can be approximated by [3]

$$\bar{T}_1 \approx \frac{h}{E_1[l_{\bar{\varphi}}(t)]} \quad as \quad h \to \infty,$$

Under H_0, the mean time between false alarms admits [3]:

$$\bar{T}_0 \ge e^h,$$

which means there is no performance loss in terms of $\bar{T}_0$, unlike the Parallel CUSUM Test for which $\bar{T}_0 \ge \frac{1}{L} e^h$.

5. NUMERICAL RESULTS

5.1. Symmetric $F(\hat{\varphi})$

First we consider detecting a sinusoid wave with an unknown amplitude as an example to demonstrate the performance improvement of our Adaptive CUSUM Test over the Parallel CUSUM Test. The following hypotheses are assumed:

$$\begin{cases} H_0 : s(t) = n(t) \\ H_1 : s(t) = A\sin(wtT_s) + n(t), \end{cases}$$

where A is an unknown amplitude within the range $[2, 36]$; w is the carrier frequency and T_s is the sampling period, both of which are known. $n(t)$ is Gaussian noise with zero mean and unit variance.

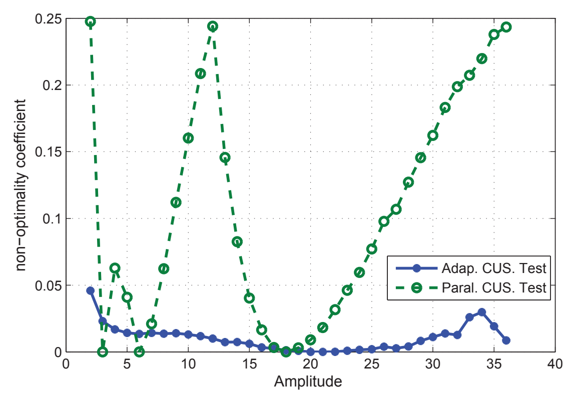

Fig. 1: Performance comparison : fixed amplitude.

Denote by $\hat{A}$ an estimate of amplitude A. It is easy to check $F(\hat{A}) = E[l_{\hat{A}}(t)]$ is symmetric about A. We choose three candidates for the Parallel CUSUM Test so that its computation complexity is comparable with the Adaptive

[3] Confidence intervals are disjoint and their union covers the whole range of the unknown parameter. If the true value lies within a candidate's confidence interval, the condition $\epsilon_\varphi \le \epsilon_m$ can be satisfied by using this candidate in the CUSUM test.

[4] The proof is omitted in the interest of space.

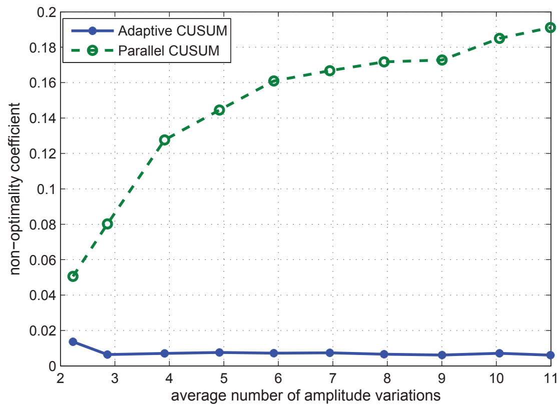

Fig. 2: Performance comparison : time-varying amplitude

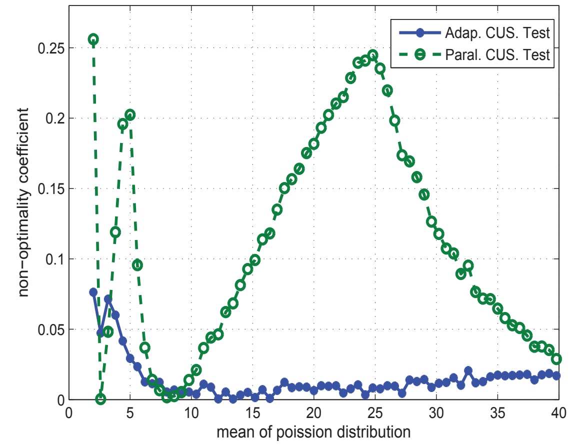

Fig. 3: Performance comparison for the asymmetric case.

CUSUM Test. We set ε_m to 0.25 and the corresponding three candidates of the Parallel CUSUM Test are determined as $3, 6, 18$.

Fig. 1 compares non-optimality coefficients of our Adaptive CUSUM Test and the Parallel CUSUM Test for different true amplitudes, staying fixed during the detection process. We can see non-optimality coefficients of the Parallel CUSUM Test are substantially larger than those of our Adaptive CUSUM Test at almost all possible amplitudes (except for the three candidates chosen by the Parallel CUSUM Test). The average non-optimality coefficient of the Parallel CUSUM Test is 0.11, while for our Adaptive CUSUM Test, it is 0.01, indicating little degradation in optimality.

In Fig. 2, it is assumed that the sinusoid wave goes though a block fading channel, where the amplitude changes every 600 time slots randomly within the range rather than fixed. We compare the average[5] non-optimality coefficients of these two tests under different average number of variations. We can see that the performance of our Adaptive CUSUM Test is stable and close to the optimal detection since it can track the amplitude change, while the Parallel CUSUM Test performs worse as more variations are involved.

5.2. Asymmetric $F(\hat{\varphi})$

To demonstrate the performance of our algorithm when the objective function is asymmetric, we consider detecting the change of the mean in a poisson distribution. The hypotheses are as (2) with $\lambda = 1$ and $\Phi = [2, 40]$. We set the non-optimality coefficient $\epsilon_m = 0.25$, which results in three candidates 2.6, 8.1 and 56.9 for the Parallel CUSUM Test.

Fig. 3 compares non-optimality coefficients of these two tests. We observe that the Adaptive CUSUM, which may not estimate the true parameter in this scenario, still significantly outperforms the Parallel CUSUM (unless some candidate of the Parallel CUSUM coincides with the true value).

[5] For each point of Fig. 2, 250 independent experiments are conducted to guarantee its fidelity.

6. CONCLUSIONS

In this paper, we have studied quickest change detection with an unknown parameter for a single observer. An Adaptive CUSUM Algorithm is proposed to narrow down the range of the unknown parameter and track its possible change during the detection process. Analysis and numerical results show that the new algorithm achieves better performance than the Parallel CUSUM Test. Interesting future directions include quickest change detection with multiple unknown parameters, and collaborative quickest detection in an ad-hoc network. We will also consider applications of our techniques to the spectrum sensing problem in Cognitive Radio.

7. REFERENCES

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

[2] E. S. Page, “Continuous inspection schemes,” *Biometrika*, vol. 41, pp. 100–115, 1954.

[3] G. Lorden, “Procedures for reacting to a change in distributions,” *Ann. Math. Statist.*, vol. 42, pp. 1897–1908, 1971.

[4] T. L. Lai, “Information bounds and quick detection of parameter changes in stochastic systems,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2917–2929, Nov. 1998.

[5] I. V. Nikiforov, “Suboptimal quadratic change detection scheme,” *IEEE Trans. on Inform. Theory*, vol. 46, pp. 2095–2107, Sept. 2000.

[6] T. Cover and J. A. Thomas, *Elements of Information Theory*, 1st edition New York: John Wiley & Sons, 1991.

[7] J. Mei, *Asymptotically Optimal Methods for Sequential*, PhD dissertation, California Institute of Technology, 2003.