Sequential Detection of Transient Changes in Stochastic Systems under a Sampling Constraint

Ehsan Ebrahimzadeh
University of California, Los Angeles
eebrahim@ucla.edu

Aslan Tchamkerten
Telecom ParisTech
aslan.tchamkerten@telecom-paristech.fr

***Abstract*—The problem of detecting a transient change in distribution of a discrete time series is investigated when there is a constraint on the number of observed samples. Under a minimax setting where the change time is unknown, the objective is to design a statistical test that minimizes a measure of worst case delay under a constraint on the average time to false alarm as well as a constraint on the sampling rate. Leveraging the results in the non-transient setting, it is shown that under full sampling there exists an asymptotic threshold on the minimum duration of a change that can be detected *reliably* with such false alarm constrained tests. Next, given a transient change with duration above this asymptotic threshold, the smallest sampling rate for which the change can be detected as *efficiently* as under full sampling is characterized asymptotically.**

I. Introduction

The problem of reliable and quick detection of abrupt temporary changes in stochastic systems finds applications in a variety of fields as diverse as industrial quality control, intrusion detection, on-line fault detection, and monitoring evolution of activities in networks. In the classical quickest change detection problem, a sequence of random variables $\{X_i\}_{i\geq 1}$, monitored sequentially, undergoes a change in distribution at some unknown point ν. It is typically assumed that the random variables X_i are independent with a common probability density function f_0 for $i < \nu$ and with another common density f_1 for $i \geq \nu$. Both f_0 and f_1 are known to the observer. The objective is to design a statistical test to detect the change, if present, based on the sequence observed so far with minimum expected delay and a controlled measure of false alarm. In this setting, Lorden [1] formulated the problem considering minimization a measure of worst case expected delay under the so called *average run length*(ARL) constraint that the mean time to false alarm is bounded from below by a parameter γ. He established an asymptotic lower bound, in the asymptote of γ, on the worst case expected delay for all stopping times satisfying the ARL constraint and showed that the CuSum statistic proposed earlier by Page [2] achieves this lower bound. Moustakides [3] proved the optimality of CuSum rule beyond the asymptotic setting considered by Lorden, casting the problem into an optimal stopping time formulation. Later, Lai in his seminal paper [4] extended the asymptotic results of Lorden to the non i.i.d. setting employing a change of measure argument.

Yet another common formulation of the change point detection problem is to assume a prior distribution on the change time ν and cast the problem into a Bayesian setting. Shiryayev [5] formulated the problem in a Bayesian framework assuming a geometric prior distribution. He showed that a procedure based on threshold comparison of the posteriori probability that a change has occurred is optimal in this setting. Inspired by [4], the asymptotic optimality of this procedure was extended to the non-i.i.d. case [6].

In many practical applications there are multiple data streams to be monitored simultaneously and the changes are quite rare. Moreover, there might be a cost associated with taking observations. We refer to the observations taken from the sequence $\{X_i\}_{i\geq 1}$ as *samples* and the fraction of samples taken in a sampling scheme as its *sampling rate*. In a series of papers, Banerjee and Veeravalli [7], [8] formulated the quickest change point detection within both Bayesian and Minimax frameworks, considering an additional constraint on the sampling rate before the change time. They show in this setting, which they refer to as data-efficient change point detection, that natural variations of the optimal procedures under full sampling achieve the same minimum detection delay under constant rate sparse sampling.

While the problem is well-studied in the non-transient setting, much less is known when the change is transient. In the transient setting, $\{X_i\}_{i\geq 1}$ is a sequence of independent random variables where all random variables have the same density function f_0 except for a possible subsequence of length L starting at an unknown point ν, *i.e.* $\{X_i\}_{\nu}^{\nu+L-1}$, along which the random variables have the common density f_1. Inspired by the alternative criteria proposed in [4], the problem of quickest detection of transient changes is formulated in [9] as minimizing the worst-case probability of missed detection under a constraint on the false alarm rate in a given period. In [10] and [11], the problem is formulated within the framework of partially observed Markov decision processes under several performance criteria. The main challenge in this setting is to design the statistical test in such a way that it reacts to the change before it disappears. In an attempt to address the problem of detectability of a transient change with a given duration, the probability of detection under Page's test is examined in [12] through various approximations. Nevertheless, to authors' best knowledge, no prior work has touched upon the fundamental guarantees on detectability of transient changes beyond approximations and bounds on probability of detection.

In this paper, we look into the problem of quickest detection of transient changes under a Minimax formulation. A fundamental problem of interest is to determine the smallest duration of a change detectable with an ARL constrained sequential test. That is, given the constraint that the expected time to false alarm be at least γ, what is the minimum duration of a change that can be detected *reliably*, when γ tends to infinity. Next, given a transient change with duration greater than the asymptotic minimum duration specified earlier, we seek to determine the smallest sampling rate under which a transient change can be detected as quickly and reliably as in the full sampling regime. We address these two questions by leveraging known results for the non-transient setting. Similar results are shown to be true in the Bayesian setting where a transient change of a given length occurs randomly and uniformly within a time frame of exponential size with respect to the duration of the change [13].

The paper is organized as follows. In Section II, an overview of the classical change point detection problem is presented. Then, in Section III, we formulate the problem of transient change point detection with a sampling constraint followed by characterizing the fundamental asymptotic threshold for detectability of a transient change and the minimum sampling rate. Finally, in Section IV, proofs of the main results are presented.

II. Non-Transient Change Detection under Full Sampling

Let us first review the classical change point detection problem for an independent random process $\{X_i\}_{i\geq 1}$, where all the random variables before an unknown instant ν, so called the change point, have the common density function f_0, while all random variables $\{X_i\}_{i\geq \nu}$ have the common density function f_1, that is

$$X_i \sim \begin{cases} f_0 & \text{if } 1 \leq i < \nu \\ f_1 & \text{if } \nu \leq i \ . \end{cases} \tag{1}$$

The problem of interest is to detect the change point with a possibly small detection delay and a controlled false alarming reaction. Framing the problem as sequential hypothesis testing, a natural approach is to consider a non-randomized stopping time τ with respect to the observed sequence so far. In the setting where no assumption is made on the prior distribution of the change point, Lorden [1] proposed a minimax formulation of the problem as to minimize a measure of worst case expected delay, while constraining expected time to false alarm; that is minimizing

$$\bar{\mathbb{E}}[\tau] \stackrel{\text{def}}{=} \sup_{\nu\geq 1} \operatorname{ess\,sup} \mathbb{E}_\nu[(\tau-\nu+1)^+ | X_1, X_2, \cdots, X_{\nu-1}], \tag{2}$$

over all stopping times τ satisfying

$$\mathbb{E}_\infty[\tau] \geq \gamma, \tag{3}$$

where the essential supremum is taken over the sigma algebra $\sigma(X_1,\cdots,X_{\nu-1})$ and $\mathbb{E}_\nu[\cdot]$ denotes expectation under $\mathbb{P}_\nu$ which is the probability measure when the change occurs at time ν. Lorden showed that for any stopping time τ satisfying (3) and any $\epsilon > 0$, as $\gamma \to \infty$,

$$\bar{\mathbb{E}}[\tau] \geq \sup_{\nu\geq 1} \mathbb{E}_\nu[\tau-\nu | \tau \geq \nu] \geq (1-\epsilon)\frac{\log\gamma}{D(f_1||f_0)}, \tag{4}$$

where $D(f_1||f_0) \stackrel{\text{def}}{=} \mathbb{E}_1 \log\frac{f_1(X_i)}{f_0(X_i)}$. Then, he showed that the stopping time based on the so called CuSum statistic, defined in the following, achieves this lower bound asymptotically.

Definition 1 *The CuSum procedure is defined as*

$$\tau_\gamma^* \stackrel{def}{=} \inf\{n | \max_{k\leq n} \sum_{i=k}^{n} Z_i \geq c\}, \tag{5}$$

where $Z_i \stackrel{def}{=} \log\frac{f_1(X_i)}{f_0(X_i)}$, *and* c *is chosen appropriately such that* $\mathbb{E}_\infty[\tau^*] \geq \gamma$.

Thus,

$$\inf\{\bar{\mathbb{E}}[\tau] | \mathbb{E}_\infty[\tau] \geq \gamma\} \sim \bar{\mathbb{E}}\tau_\gamma^* \sim \frac{\log\gamma}{D(f_1||f_0)}. \tag{6}$$

Later, Lai extended this result for the non independent identically distributed setting using a natural generalization of the CuSum rule in [4], where he introduced new performance criteria which also provide insights on the other variations of the problem. The main ingredient in [4] to prove (6) is the following important observations which stands as the core of asymptotic optimality of the CuSum rule for minimizing (2). For any stopping time τ satisfying (3) and any $\epsilon > 0$, as $\gamma \to \infty$,

$$\mathbb{P}_\nu(\tau - \nu > (1-\epsilon)\frac{\log\gamma}{D(f_1||f_0)} | \tau \geq \nu) = 1 + o(1), \tag{7}$$

where the $o(1)$ term does not depend on ν. Moreover, for the stopping time τ_γ^* defined in (5), as $\gamma \to \infty$,

$$\mathbb{P}_\nu(\tau_\gamma^* - \nu \leq (1+\epsilon)\frac{\log\gamma}{D(f_1||f_0)} | \tau_\gamma^* \geq \nu) = 1 + o(1), \tag{8}$$

for all $\nu \geq 1$.

These observations provide fundamental insight on detectability of transient changes, suggesting an asymptotic threshold on the duration of the smallest piece of change that can be detected reliably by a stopping random variable satisfying (3).

III. Transient Change Point Detection Under Sparse Sampling

Suppose now that the random process $\{X_i\}_{i\geq 1}$ is distributed as follows

$$X_i \sim \begin{cases} f_0 & \text{if } 1 \leq i < \nu \text{ or } i \geq \nu + L \\ f_1 & \text{if } \nu \leq i < \nu + L \ . \end{cases} \tag{9}$$

The sequence $\{X_i\}_{i\geq 1}$ is observed sequentially according to a sampling strategy defined as follows:

Definition 2 *A Sampling Strategy with respect to a random process* $\{X_i\}_{i\geq 1}$ *is an ordered collection of random time indices* $\mathcal{S} = \{S_1, S_2, \cdots\}$, *where* S_i *corresponds to the ith sampling time. In general, the decision as to whether take a sample at a certain time instance or not depends on the past samples. This means that* $S_1 = 1$ *is an arbitrary time index and, for any* $n \geq 2$,

$$S_n \stackrel{def}{=} \Phi_n(\{X_{S_i}\}_{\{i<n\}}), \tag{10}$$

where $\Phi_n : \mathcal{X}^{n-1} \to \{S_{n-1}+1, S_{n-1}+2, \cdots\}$ *is a decision function at time* S_{n-1}.

The objective is to detect the change efficiently, in a suitable sense, with a so called decision policy Ψ which consists of a sampling strategy $\mathcal{S}$ and a stopping time τ with respect to the sampled sequence so far, that is $\Psi = (\mathcal{S}, \tau)$. A decision policy is evaluated with a measure of detection delay and its sampling rate constrained on a measure of false alarm. As in the non-transient setting, we consider the class of all decision policies satisfying the ARL constraint.

Definition 3 (False Alarm Constraint) *An ARL constrained decision policy $\Psi = (\mathcal{S}, \tau)$ is such that*

$$\mathbb{E}_\infty[\tau] \geq \gamma. \tag{11}$$

Definition 4 (Detection Delay) *For a decision policy $\Psi = (\mathcal{S}, \tau)$, and $\epsilon > 0$, the worst case minimum delay in probability is defined as*

$$d(\Psi, \epsilon) \stackrel{def}{=} \inf \left\{ \ell : \sup_{\nu \geq 1} \mathbb{P}_\nu(\tau - \nu > \ell | \tau \geq \nu) \leq \epsilon \right\}. \tag{12}$$

We will later argue that in the transient setting a measure of worst case delay in probability is more appropriate to be adopted compared to the measure of worst case expected delay defined earlier in (2).

Definition 5 (Sampling Rate) *For a decision policy $\Psi = (\mathcal{S}, \tau)$, the pre-change sampling rate is defined as*

$$\rho(\Psi) \stackrel{def}{=} \limsup_{n \to \infty} \mathbb{E}_n \Big[\frac{|\mathcal{S}^{(n)}|}{n} \big| \tau \geq n \Big], \tag{13}$$

where $\mathcal{S}^{(n)} \triangleq \{i \in \mathcal{S} | i \leq n\}$ is the set of indices corresponding to samples taken up to time n.

Note that the use of "limsup" instead of "sup" is to avoid sampling rates close to 1, when the change occurs early in the sequence.

Definition 6 (Achievable Sampling Rate) *Let $\{\rho_\gamma\}_{\gamma>0}$ be an indexed family with $0 \leq \rho_\gamma \leq 1$. Sampling rates $\{\rho_\gamma\}_{\gamma>0}$ are achievable with respect to an indexed family of change durations $\{L_\gamma\}_{\gamma>0}$, if there exists an indexed family of decision rules $\{\Psi_\gamma = (\mathcal{S}_\gamma, \tau_\gamma)\}_{\gamma>0}$, such that for γ large enough,*

(i) The ARL constraint $\mathbb{E}_\infty[\tau_\gamma] \geq \gamma$ is satisfied,

(ii) The sampling rate satisfies $\rho(\Psi_\gamma) \leq \rho_\gamma$,

(iii) The delay satisfies $d(\Psi_\gamma, \epsilon_\gamma) \leq L_\gamma$ for some indexed family $\{\epsilon_\gamma\}_{\gamma>0}$ such that $\lim\limits_{\gamma\to\infty} \epsilon_\gamma = 0$.

Notational Convention

When clear from the context, we represent an indexed family with its representative element, *e.g.*, we denote $\{\rho_\gamma\}_{\gamma>0}$ simply as ρ_γ. Moreover, we will use d_γ instead of $d(\Psi_\gamma, \epsilon_\gamma)$, leaving out any explicit reference to the decision policy Ψ_γ and to the indexed family $\{\epsilon_\gamma\}_{\gamma>0}$ which we assume satisfies $\lim\limits_{\gamma\to\infty} \epsilon_\gamma = 0$, unless it is necessary to make the sampling strategy or the stopping time explicit. Finally, let Ψ^*_γ denote the CuSum decision policy with the stopping time τ^*_γ defined in (5) and full sampling strategy.

Theorem 1 (Transient Change under Full Sampling) *(i) Let $\alpha > 1$ and suppose $L_\gamma \geq \alpha \frac{\log \gamma}{D(F_1||F_0)}$. Then $\rho_\gamma = 1$ is achievable with respect to L_γ. Moreover, we have*

$$d(\Psi^*_\gamma) \sim \frac{\log \gamma}{D(f_1||f_0)}. \tag{14}$$

(ii) Let $0 \leq \alpha < 1$ and suppose that $L_\gamma \leq \alpha \frac{\log \gamma}{D(f_1||f_0)}$. Then $\rho_\gamma = 1$ is not achievable with respect to L_γ. Moreover, in this case, for any decision policy Ψ_γ satisfying the false alarm and the sampling rate constraints in Definition 6, we have

$$\liminf_{\gamma\to\infty} \frac{d_\gamma}{\gamma^{1-\alpha}} \geq 1. \tag{15}$$

Theorem 1 establishes an asymptotic threshold on the minimum duration of a change that can be detected reliably. Specifically, for γ large enough, if the duration of the change is above $\frac{\log \gamma}{D(f_1||f_0)}$, the minimum delay is as short as if the change had infinite duration. Henceforth we call such transient changes *asymptotically detectable*. For transient changes with duration below this threshold, delay grows as a polynomial function of γ, whenever the false alarm constraint $\mathbb{E}_\infty[\tau_\gamma] \geq \gamma$ is satisfied. The lower bound on the asymptotic worst case delay in probability in (15) can be converted to a lower bound on the worst case expected delay defined in (2). Note, however, that the guarantee provided on the asymptotic worst case delay in (14) cannot necessarily be translated to a guarantee on the worst case expected delay. This is because when the event $\{\tau^*_\gamma > \nu + L_\gamma\}$ occurs, although happening with a vanishing probability, the delay can be arbitrarily large, as the rest of the observations are f_0 distributed, which leads the expected delay to grow unbounded.

The following two theorems characterize the minimum achievable sampling rates with respect to duration of detectable transient changes. Theorem 3 is proved using the so called DE-CuSum decision policy proposed in [8], which was used to achieve any constant sampling rate and asymptotically the same worst case expected delay as under full sampling in a non-transient scenario. A brief description of the DE-CuSum rule is provided in the next section right before proof of Theorem 3. For a more details on the description of this algorithm we refer the reader to [8].

Theorem 2 (Minimum Asymptotic Achievable Rate [8]) *Let $\alpha > 1$ and suppose that $L_\gamma \geq \alpha \frac{\log \gamma}{D(f_1||f_0)}$. Then $\rho_\gamma = \omega(\frac{1}{\log \gamma})$ is achievable with respect to L_γ.*[1] *Moreover, for the DE-CuSum decision policy, we have*

$$d(\hat{\Psi}_\gamma) \sim \frac{\log \gamma}{D(f_1||f_0)}. \tag{16}$$

Theorem 2 amounts to a slight tightening of the analysis of the De-CUSUM rule in [8]. Note that in the authors [8] were only interested in constant sampling rates while in the present case, the step size of the DE-CuSum procedure in the idle regime is

[1] We use here the Landau Big O notation, for instance, $f(\gamma) = \omega(g(\gamma))$ if $g(\gamma)/f(\gamma) \to 0$.

not a constant but a function of γ, with the same asymptotic growth rate as ρ_γ.

In the following theorem, we assume that the indexed family L_γ corresponds to durations of some asymptotically detectable changes and show that sampling rates $\rho_\gamma = o(\frac{1}{\log \gamma})$ are not achievable.

Theorem 3 (Converse) *Let $\alpha > 1$ and suppose that $L_\gamma \geq \alpha \frac{\log \gamma}{D(f_1||f_0)}$. Consider sampling rates $\rho_\gamma = o(\frac{1}{L_\gamma})$ with respect to L_γ. Then, for any decision policy Ψ_γ satisfying the false alarm and the sampling rate constraints in Definition 6, and γ sufficiently large, there exists a time period of duration L_γ from which Ψ_γ fails to sample even a point, with probability bounded away from zero. As a consequence,*

$$\liminf_{\gamma \to \infty} d_\gamma \geq \frac{\gamma}{2}. \tag{17}$$

IV. PROOFS

A. Proof of Theorem 1

(i) Suppose that there exists some $\delta > 0$ such that $L_\gamma \geq (1+\delta)\frac{\log \gamma}{D(f_1||f_0)}$ for some $\delta > 0$, and consider the CuSum rule

$$\tau_\gamma^* = \inf\{n | \max_{k \leq n} \sum_{i=k}^{n} Z_i \geq \log \gamma\}. \tag{18}$$

It follows that for any $0 < \epsilon \leq \delta$

$$\begin{aligned} &\mathbb{P}_\nu(\tau_\gamma^* - \nu > (1+\epsilon)\frac{\log \gamma}{D(f_1||f_0)} | \tau_\gamma^* > \nu) \\ \leq\ & \mathbb{P}_\nu\left(\bigcap_{\nu \leq n \leq \nu + (1+\epsilon)\frac{\log \gamma}{D(f_1||f_0)}} \left\{ \max_{k \leq n} \sum_{i=k}^{n} Z_i < \log \gamma \right\} \right) \\ \leq\ & \mathbb{P}_\nu\left(\sum_{i=\nu}^{(1+\epsilon)\frac{\log \gamma}{D(f_1||f_0)}} Z_i < \log \gamma \right) \underset{\gamma \to \infty}{\longrightarrow} 0, \end{aligned} \tag{19}$$

where the first inequality follows from the definition of the CuSum rule in (18) and the last step follows from applying the law of large numbers to the sequence of i.i.d. random variables $\{Z_i\}_{i \geq \nu}$ with mean $D(f_1||f_0)$. Note that (19) establishes an upper bound on the detection delay. Moreover, by causality of the stopping time random variables, (7) gives a lower bound on the detection delay, which combined with the upper bound yields the desired result.

(ii) Now suppose that $L_\gamma \leq (1-\delta)\frac{\log \gamma}{D(f_1||f_0)}$ for some $\delta > 0$. Let $\Psi_\gamma = (\mathcal{S}_\gamma, \tau)$ be any decision policy satisfying the false alarm constraint $\mathbb{E}_\infty[\tau] \geq \gamma$. Since $\mathbb{E}_\infty[\tau] \geq \gamma$, it follows [Proof of Theorem 1 in [4]] that for any integer $m < \gamma$, there is some $\nu \geq 1$ such that

$$\mathbb{P}_\infty(\tau \geq \nu) > 0, \text{ and } \mathbb{P}_\infty(\tau < \nu + m | \tau \geq \nu) \leq \frac{m}{\gamma}. \tag{20}$$

Let m be the largest integer less than $2\gamma^{\delta-\epsilon}$ for some $0 \leq \epsilon < \delta$. Define the events

$$\mathcal{C}_\epsilon = \left\{ 0 \leq \tau - \nu \leq \gamma^{\delta-\epsilon}, \quad \sum_{i=\nu}^{\min\{\tau, \nu+L_\gamma-1\}} Z_i < (1-\epsilon)\log \gamma \right\},$$

and

$$\mathcal{C'}_\epsilon = \left\{ 0 \leq \tau - \nu \leq \gamma^{\delta-\epsilon}, \quad \sum_{i=\nu}^{\min\{\tau, \nu+L_\gamma-1\}} Z_i \geq (1-\epsilon)\log \gamma \right\}.$$

Following the same lines as that of Lai's change of measure argument (Proof of [4, Theorem 1]), we have:

Claim 1 *As long as $\delta > 2\epsilon > 0$,*

$$\mathbb{P}_\nu(\mathcal{C}_\epsilon | \tau \geq \nu) \underset{\gamma \to \infty}{\longrightarrow} 0. \tag{21}$$

Claim 2

$$\mathbb{P}_\nu\left\{\mathcal{C'}_\epsilon \mid \tau \geq \nu\right\} \underset{\gamma \to \infty}{\longrightarrow} 0. \tag{22}$$

Combining Claims 1 and 2, we get $\mathbb{P}_\nu\left\{\tau - \nu \leq \gamma^{\delta-\epsilon} | \tau \geq \nu\right\} \to 0$, as γ tends to infinity, for ν given in (20). Since ϵ can be made arbitrarily small, it follows that

$$\sup_{\nu > 0} \mathbb{P}_\nu\left\{\tau - \nu > \gamma^\delta | \tau \geq \nu\right\} \to 1, \tag{23}$$

which in turn implies

$$\liminf_{\gamma \to \infty} \frac{d_\gamma}{\gamma^\delta} \geq 1,$$

as desired.

B. Description of DE-CuSum decision rule

Let us briefly review the DE-CuSum detection rule $\hat{\Psi}$ proposed in [8]. Define the sampling indicator random variable M_i as being 1 when the time instance i is sampled, and zero otherwise. Start with $D_0 = 0$ and fix $\gamma > 0$, $\mu_\gamma > 0$ and $h > 0$. Also, define the stopping time as

$$\hat{\tau}_\gamma \overset{\text{def}}{=} \inf\{n \geq 1 | D_n > \log \gamma\}. \tag{24}$$

At each step, the statistic D_n is being updated as follows

$$D_{n+1} = \begin{cases} \min\{D_n + \mu, 0\} & \text{if } D_n < 0 \\ (D_n + Z_{n+1})^{h+} & \text{otherwise} \end{cases} \tag{25}$$

where $(x)^{h+} \overset{\text{def}}{=} \max\{x, -h\}$. In fact the algorithm naturally performs a hypothesis test between the distributions f_1 and f_0 and skips the samples while the statistic is below 0. As long as $D_n < 0$, which depends on the last undershoot from 0, samples are skipped and D_n is being updated by adding the deterministic increment μ to D_n.

C. Proof of Theorem 2

We prove this theorem by the DE-CuSum procedure $\hat{\Psi}_\gamma$ described earlier. It is shown in [8] that the sampling constraint $\rho(\hat{\Psi}_\gamma) \leq \rho_\gamma$ is met if

$$\mu_\gamma < K \frac{\rho_\gamma}{1 - \rho_\gamma}, \tag{26}$$

where $K = \frac{\mathbb{E}_\infty[|Z_1^{h+}||Z_1<0]\ \mathbb{P}_\infty(Z_1<0)^2}{\mathbb{E}_\infty[\lambda_\infty]}$ is a constant that does not scale with γ and $\lambda_\gamma = \inf\{n \geq 1 | \sum_{i=k}^{n} Z_i \notin [0, \log \gamma]\}$ [8].

Using standard arguments, in [8] it is shown that the asymptotic worst case delay in probability of the DE-CuSum rule is bounded from above as follows

$$\limsup_{\gamma\to\infty} d_\gamma(\hat{\Psi}) \le \frac{\log\gamma}{D(f_1||f_0)} + K'\frac{1}{\mu_\gamma} + K'', \tag{27}$$

where $K' = \frac{\mathbb{E}_\infty[|D^{h+}_{\lambda_\infty}|]}{\mathbb{P}_1(D_{\lambda_\infty}>0)\mathbb{P}_1(Z_1<0)} + h$ and $K'' = 2 + \frac{1}{\mathbb{P}_1(D_{\lambda_\infty}>0)}$ are well defined constants which do not scale with γ.

Given a sampling rate constraint $\rho_\gamma = \omega(\frac{1}{\log\gamma})$, by setting the step parameter for the sojourn time in the DE-CuSum procedure as $\mu_\gamma = \theta(\rho_\gamma)$ such that (26) is satisfied, the desired result follows by considering (27) and the lower bound (7).

D. Proof of Theorem 3

We show that for any decision policy $\Psi_\gamma = \{S_\gamma, \tau\}$ satisfying the false alarm constraint $\mathbb{E}_\infty[\tau] \ge \gamma$ and the sampling rate constraint $\rho(\Psi_\gamma) \le \rho_\gamma$ with $\rho_\gamma = o(\frac{1}{L_\gamma})$, there exists a time interval of duration L_γ such that, with probability bounded away from zero, no point within this interval is sampled by Ψ_γ, for γ sufficiently large.

First note that since Ψ_γ satisfies the false alarm constraint $\mathbb{E}_\infty[\tau] \ge \gamma$, we get

$$\begin{aligned}\mathbb{P}_\infty(\tau \ge \frac{\gamma}{2}) &\ge \mathbb{P}_\infty(\tau \ge \frac{\gamma}{2} + \nu) \\ &> \frac{1}{2}\mathbb{P}_\infty(\tau \ge \nu) > 0,\end{aligned} \tag{28}$$

where the last inequalities hold for some $\nu \ge 1$ by (20) with $m = \frac{\gamma}{2}$.

The proof is based on the fact that as long as the samples are drawn from the distribution f_0, the sampling rate constraint guarantees existence of sufficiently long gaps among the sampling times. Note that if no change occurs at all, the sampling rate constraint implies

$$\mathbb{E}_\infty\Big[\frac{|\mathcal{S}^{(\gamma/2)}|}{\gamma/2}\big|\tau \ge \frac{\gamma}{2}\Big] = o(\frac{1}{L_\gamma}). \tag{29}$$

Divide the time frame up to time $\nu = \frac{\gamma}{2}$ into consecutive intervals of size L_γ. For each time interval, define the following indicator random variable

$$K_i = \begin{cases} 1 & \text{If at least one point from } i^{th} \text{ interval is sampled} \\ 0 & \text{Otherwise.}\end{cases}$$

We show that there exists an interval from which no sample is taken with probability bounded away from zero. In fact, we show that

$$\limsup_{\gamma\to\infty} \max_j \mathbb{P}_\infty(K_j = 0|\tau \ge \frac{\gamma}{2}) = 1. \tag{30}$$

Otherwise,

$$\liminf_{\gamma\to\infty} \min_i \mathbb{P}_\infty(K_i = 1|\tau \ge \frac{\gamma}{2}) > 0, \tag{31}$$

which implies

$$\begin{aligned}\mathbb{E}_\infty\Big[\frac{|\mathcal{S}^{(\gamma/2)}|}{\gamma/2}\big|\tau \ge \frac{\gamma}{2}\Big] &\overset{(a)}{\ge} \mathbb{E}_\infty\Big[\frac{2}{\gamma}\sum_{j=1}^{\lfloor\gamma/2L_\gamma\rfloor} K_j\big|\tau \ge \frac{\gamma}{2}\Big] \\ &= \frac{2}{\gamma}\sum_{j=1}^{\lfloor\gamma/2L_\gamma\rfloor}\mathbb{P}_\infty(K_j = 1|\tau \ge \frac{\gamma}{2}) \\ &= \theta(\frac{1}{L_\gamma}),\end{aligned} \tag{32}$$

where (a) holds because more than one sample can be taken from a given interval. Note that (32) contradicts (29), establishing that (30) holds.

Combining (28) with (30) yields $\mathbb{P}_\infty(K_j = 0) > 0$ for some $j \ge 1$. Therefore, for any decision policy, there exists some interval of size L_γ from which no sample is taken with probability bounded away from zero.

Suppose now that a change of duration L_γ occurs along the sequence within the jth interval for which $\mathbb{P}_\infty(K_j = 0) > 0$. In such a case, with probability bounded away from zero, no point of the change period is sampled, meaning that only samples from the distribution f_0 are observed. In this case, since $\mathbb{E}_\infty[\tau] \ge \gamma$, it follows using an argument similar to part (ii) of theorem 1 that

$$\liminf_{\gamma\to\infty} d(\Psi_\gamma) \ge \frac{\gamma}{2}, \tag{33}$$

as desired.

References

[1] G. Lorden, “Procedures for reacting to a change in distribution”, *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1897–1908, 1971.

[2] E. S. Page, “A test for a change in parameter occurring at an unknown point”, *Biometrica*, vol. 42, pp. 523–527, 1955.

[3] G. V. Moustakides, “Optimal stopping times for detecting changes in distributions”, *Annals of Statistics*, vol. 14, pp. 1379–1387, Dec. 1986.

[4] T. L. Lai, “Information bounds and quick detection of parameter changes in stochastic systems”, *Information Theory, IEEE Transactions on*, vol. 44, no. 7, pp. 2917–2929, Nov. 1998.

[5] A. N. Shiryayev, “Optimal Stopping Rules”, *NY:Springer*, 1978.

[6] A. G. Tartakovsky and V. V. Veeravalli, “General asymptotic Bayesian theory of quickest change detection”, *SIAM Theory of Prob. and Applications*, vol. 49, pp. 458–497, Sept. 2005.

[7] T. Banerjee, and V. V. Veeravalli, “Data-efficient quickest change detection with on-off observation control”, *Seq. Analysis*, vol. 31, pp. 40–77, 2012.

[8] T. Banerjee, and V. V. Veeravalli, “Data-Efficient Quickest Change Detection in Minimax Settings”, *Information Theory, IEEE Transactions on*, vol. 59, no. 10, pp. 6917–6931, Oct. 2013.

[9] B. K. Gupi, L. Fillatre, and I. Nikiforov, “Sequential Detection of Transient Changes”, *Sequential Analysis*, vol. 31: pp. 528–547, 2012.

[10] K. Premkumar, A. Kumar, and V. V. Veeravalli, “Bayesian quickest transient change detection”, *IWAP*, 2010.

[11] V. Krishnamurthy, “Bayesian Sequential Detection With Phase-Distributed Change Time and Nonlinear Penalty—A POMDP Lattice Programming Approach”, *Information Theory, IEEE Transactions on*, vol. 57, no. 10: pp. 7096–7124, Oct. 2011.

[12] C. Han, P.K. Willett, and D.A. Abraham, “Some methods to evaluate the performance of Page’s test as used to detect transient signals”, *Signal Processing, IEEE Transactions on*, vol. 47, no. 10: pp. 2112–2127, 1999.

[13] V. Chandar, and A. Tchamkerten, “Quickest Transient-Change Detection under a Sampling Constraint”, *CoRR*, vol. abs/1501.05930, 2015.