Data-Efficient Quickest Change Detection with On-Off Observation Control

Taposh Banerjee and Venugopal V. Veeravalli
ECE Department and Coordinated Science Laboratory
University of Illinois at Urbana-Champaign, Urbana, IL
Email: banerje5,vvv@illinois.edu

Abstract

In the classical version of the Bayesian quickest change detection problem proposed by Shiryaev, there is a sequence of observations whose distribution changes at a random time, and the goal is to minimize the average delay in detecting the change, subject to a constraint on the probability of false alarm. We consider this quickest change detection problem with an additional constraint on the average number of observations used in detecting the change, where we have the option to choose whether or not to take a given observation. The objective is to select the observation control policy along with the stopping time at which the change is declared, so as to minimize the average detection delay, subject to constraints on both the probability of false alarm and the average number of observations used. In contrast to the single threshold test that is optimal for the Shiryaev problem, the optimal algorithm for our problem belongs to a class of randomized three-threshold policies. As in the Shiryaev test, the statistic being thresholded is the a posteriori probability of the occurrence of the change, given the observation sequence. Towards characterizing the thresholds for the optimal algorithm, we provide an asymptotic analysis of deterministic three-threshold policies for the case where the probability of false alarm is small, the average number of observations used is large, and the change event is rare. The asymptotic analysis reveals that any three-threshold policy can be well approximated by a two-threshold policy. An advantage of the two-threshold policy is that there exists a unique pair of thresholds that achieves the constraints on the probability of false alarm and the average number of observations used. Therefore, using our analysis, the thresholds can be set directly using the given constraints. We provide extensive simulation results that corroborate our analytical findings.

This research is partially supported by the National Science Foundation under grant CCF 08-30169, through the University of Illinois at Urbana-Champaign. This research was also supported in part by the U.S. Army Research Office MURI grant W911NF-06-1-0094, through a subcontract from Brown University at the University of Illinois.

I. INTRODUCTION

In the Bayesian quickest change detection problem proposed by Shiryaev [1], there is a sequence of random variables, $\{X_n\}$, whose distribution changes at a random time Γ. It is assumed that before Γ, $\{X_n\}$ are independent and identically distributed (i.i.d.) with density f_0, and after Γ they are i.i.d. with density f_1. The distribution of Γ is assumed to be known and modeled as a geometric random variable. The objective is to find a stopping time τ, at which time the change is declared, such that the average detection delay is minimized subject to a constraint on the probability of false alarm.

In this paper we extend Shiryaev's formulation by explicitly accounting for the cost of the observations used in the detection process. We capture the observation penalty through the average number of observations used in the detection process, and allow for a dynamic control policy that determines whether or not a given observation is taken. The objective is to choose the observation control policy along with the stopping time τ, so that the average detection delay is minimized subject to constraints on the probability of false alarm and the average number of observations used. The motivation for this model comes from the consideration of the following engineering applications.

In statistical process control, economic-statistical control chart schemes are designed to detect an abrupt change in an industrial process that can affect the quality of the output, and at the same time minimize the average operational cost in some sense [2], [3], [4]. The traditional approach has been to use simple algorithms from the change detection literature, such as Shewhart, EWMA and CUSUM, as control charts, and optimize over the choice of sample size, sampling interval and control limits [4]. The reasons for choosing the change detection algorithms mentioned above have been simplicity of design and ease of implementation [2], [4]. However, it has been demonstrated, mostly through numerical results, that Bayesian control charts, which choose the parameters of the detection algorithms based on the posterior probability that the process is out of control, perform better than the traditional control charts based on Shewhart, EWMA or CUSUM; see [3], [4], [5], and the references therein. But, Bayesian control rules are not preferred due to the complexity of the rules and also due to the difficulty in designing such a scheme, e.g., choosing thresholds [4]. The process control problem is fundamentally a quickest change detection problem, and it is therefore appropriate that the cost-efficient schemes for process control are developed in this framework. In this paper, we derive a simple two-threshold Bayesian test, show conditions under which it is approximately optimal, and obtain analytical approximations of its performance using which the thresholds can be set directly from the constraints.

In many monitoring applications, for example infrastructure monitoring, environment monitoring, or habitat monitoring, especially of endangered species, surveillance is only possible through the use of inexpensive battery operated sensor nodes. This could be due to the high cost of employing a wired sensor network or a human observer, or the infeasibility of having a human intervention. For example in habitat monitoring of certain sea-birds as reported in [6], the very reason the birds chose the habitat was because of the absence of humans and predators around it. In these applications the sensors are typically deployed for long durations, possibility over months, and due the constraint on energy, the most effective way to save energy at the sensors is to switch the sensor between on and off states. An energy-efficient quickest change detection algorithm can be employed here that can operate over months and trigger other more sophisticated and costly sensors, which are possibly power hungry, or more generally, trigger a larger part of the sensor network [7], when some change is detected in the process under observation. This change could be a fault in the structures in infrastructure monitoring [7], arrival of the species to the habitat [6], etc.

There have been other formulations of the Bayesian quickest change detection problem that are relevant to sensor networks: see [8]-[12]. The change detection problem studied here was earlier considered in a more general set-up for sensor networks in [13]. But owing to the complexity of the problem, the structure of the optimal policy was studied only numerically, and for the same reason, no analytical expressions were developed for the performance.

The goal of this paper is to develop a deeper understanding of the trade-off between delay, false alarm probability, and the cost of observation or information, and to identify a control policy for data-efficient quickest change detection that has some optimality property and is easy to design. As mentioned earlier, we extend Shiryaev's quickest change detection formulation by applying an additional constraint on the number of observations used in the detection process. We characterize the optimal control policy for this problem by solving a Lagrangian relaxation using dynamic programming. Following an approach similar to the one in [13], we show that the *a posteriori* probability p_k of the change having happened before time k, given all the information up to time k, serves as a sufficient statistic for both the stopping rule as well as the observation control law. Furthermore, we show that the optimal control policy that minimizes the Lagrangian cost has a three-threshold structure. However, finding the optimal control for the original constrained optimization problem would require optimizing over the three thresholds and possible randomization over such optimized three-threshold policies. To this end, we analyze the three-threshold policy using nonlinear renewal theory [14], [15], and characterize the optimal choice of thresholds. Specifically, we identify an asymptotic regime in which any three-threshold policy, and hence also the optimal policy for the constrained optimization problem when the optimal policy is deterministic, can be well approximated by a two-threshold policy. The advantage of this simpler two-threshold policy is that there exists a unique pair of thresholds that achieves the constraints on the probability of false alarm and the average number of observations used. Also, using our analytical results, the thresholds can be set directly using the constraints.

In the following section, we set up the data-efficient quickest change detection problem with on-off observation control, and summarize the dynamic programming solution. The layout of the paper is provided at the end of the section.

II. Problem Formulation and Dynamic Programming Solution

As in the model for the classical Bayesian quickest change detection problem described in Section I, we have a sequence of random variables $\{X_n\}$, which are i.i.d. with density f_0 before the random change point Γ, and i.i.d. with density f_1 after Γ. The change point Γ is modeled as geometric with parameter ρ, i.e., for $0 < \rho < 1,\ 0 \leq \pi_0 < 1$,

$$\pi_k = \mathrm{P}\{\Gamma = k\} = \pi_0 \, \mathbb{I}_{\{k=0\}} + (1-\pi_0)\rho(1-\rho)^{k-1} \, \mathbb{I}_{\{k\geq 1\}},$$

where $\mathbb{I}$ is the indicator function, and π_0 represents the probability of the change having happened before the observations are taken. Typically π_0 is set to 0.

In order to minimize the average number of observations used, 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.,

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

with μ denoting the control law and I defined as:

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

Here, $X_i^{(S_i)}$ represents X_i if $S_i = 1$, otherwise X_i is absent from the information vector I_k. The choice of S_1 is based on the prior π_0.

As in the classical change detection problem, the end goal is to choose a stopping time on the observation sequence at which time the change is declared. Denoting the stopping time by τ, we can define the average detection delay (ADD) as

$$\mathrm{ADD} = \mathrm{E}\left[(\tau - \Gamma)^+\right].$$

Further, we can define the probability of false alarm (PFA) as

$$\mathrm{PFA} = \mathrm{P}(\tau < \Gamma).$$

The new performance metric for our problem is the average number of observations (ANO) used in detecting the change:

$$\mathrm{ANO} = \mathrm{E}\left[\sum_{k=1}^{\tau} S_k\right].$$

Let $\gamma = \{\tau, \mu_0, \ldots, \mu_{\tau-1}\}$ represent a policy for cost-efficient quickest change detection. We wish to solve the following optimization problem:

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

where α and β are given constraints. This problem is difficult to solve directly and hence we consider a Lagrangian relaxation of this problem which can be solved using dynamic programming.

The Lagrangian relaxation of the optimization problem in (1) is,

$$R(\gamma) = \min_{\gamma} \mathrm{ADD}(\gamma) + \lambda_f \ \mathrm{PFA}(\gamma) + \lambda_c \ \mathrm{ANO}(\gamma), \tag{2}$$

where λ_f and λ_c are Lagrange multipliers. It is easy to see that if λ_f and λ_c can be found such that the solution to (2) achieves the PFA and ANO constraints with equality, then the solution to (2) is also the solution to (1). In general it can be argued that the solution to the constrained problem (1) can be obtained by randomizing over solutions to (2). Due to the result in [16], it can be further argued that the number of policies over which the randomization has to be done will be finite.

The solution to the optimization problem in (2) can be obtained using dynamic programming and following steps similar to those given in [13]. In the following we summarize these steps.

Let Θ_k denote the state of the system at time k. After the stopping time τ it is assumed that the system enters a terminal state $\mathcal{T}$ and stays there. For $k < \tau$, we have $\Theta_k = 0$ for $k < \Gamma$, and $\Theta_k = 1$ otherwise. Then we can write

$$\mathrm{ADD} = \mathrm{E}\left[\sum_{k=0}^{\tau-1} \mathbb{I}_{\{\Theta_k = 1\}}\right]$$

and $\mathrm{PFA} = \mathrm{E}[\mathbb{I}_{\{\Theta_\tau = 0\}}]$.

Furthermore, let D_k denote the stopping decision variable at time k, i.e., $D_k = 0$ if $k < \tau$ and $D_k = 1$ otherwise. Then the optimization problem in (2) can be written as a minimization of an additive cost over time:

$$R(\gamma) = \min_{\gamma} \mathrm{E}\left[\sum_{k=0}^{\tau} g_k(\Theta_k, D_k, S_k)\right]$$

with

$$g_k(\theta, d, s) = \mathbb{I}_{\{\theta \neq \mathcal{T}\}} \ \left[\mathbb{I}_{\{\theta=1\}}\mathbb{I}_{\{d=0\}} + \lambda_f \ \mathbb{I}_{\{\theta=0\}}\mathbb{I}_{\{d=1\}} \ + \lambda_c \ \mathbb{I}_{\{s=1\}}\mathbb{I}_{\{d=0\}}\right].$$

Using standard arguments [18] it can be seen that this optimization problem can be solved using infinite horizon dynamic programming with sufficient statistic (belief state) given by:

$$p_k = \mathrm{P}\{\Theta_k = 1 \mid I_k\} = \mathrm{P}\{\Gamma \leq k \mid I_k\}.$$

Using Bayes' rule, p_k can be shown to satisfy the recursion

$$p_{k+1} = \begin{cases} \Phi^{(0)}(p_k) & \text{if } S_{k+1} = 0 \\ \Phi^{(1)}(X_{k+1}, p_k) & \text{if } S_{k+1} = 1 \end{cases}$$

where

$$\Phi^{(0)}(p_k) = p_k + (1 - p_k)\rho \tag{3}$$

and

$$\Phi^{(1)}(X_{k+1}, p_k) = \frac{\Phi^{(0)}(p_k)L(X_{k+1})}{\Phi^{(0)}(p_k)L(X_{k+1}) + (1 - \Phi^{(0)}(p_k))} \tag{4}$$

with $L(X_{k+1}) = f_1(X_{k+1})/f_0(X_{k+1})$ being the likelihood ratio, and $p_0 = \pi_0$. Note that the structure of recursion for p_k is independent of time k.

The optimal policy for the problem given in (2) can be obtained from the solution to the Bellman equation:

$$J(p_k) = \min_{d_k, s_{k+1}} \lambda_f \ (1 - p_k)\mathbb{I}_{\{d_k=1\}} + \mathbb{I}_{\{d_k=0\}} \left[p_k + B_J^{(0)}(p_k)\mathbb{I}_{\{s_{k+1}=0\}} + (\lambda_c + B_J^{(1)}(p_k))\mathbb{I}_{\{s_{k+1}=1\}}\right] \tag{5}$$

with

$$B_J^{(0)}(p_k) = J(\Phi^{(0)}(p_k))$$

and

$$B_J^{(1)}(p_k) = \mathrm{E}[J(\Phi^{(1)}(X_{k+1}, p_k))].$$

It is easy to show by an induction argument (see, e.g., [13]) that J, $B_J^{(0)}$ and $B_J^{(1)}$ are all non-negative concave functions on the interval $[0, 1]$, and that $J(1) = B_J^{(0)}(1) = B_J^{(1)}(1) = 0$. Also, by Jensen's inequality

$$B_J^{(1)}(p) \leq J(\mathrm{E}[\Phi^{(1)}(X, p)]) = B_J^{(0)}(p), \quad p \in [0, 1].$$

From these properties of J, $B_J^{(0)}$ and $B_J^{(1)}$, it is easy to show that the optimal policy $\gamma^* = (\tau^*, \mu_0^*, \mu_1^*, \ldots, \mu_{\tau-1}^*)$ for the problem given in (2) has the following three threshold structure:

$$S_{k+1}^* = \mu_k^*(p_k) = \begin{cases} 0 & \text{if } p_k \in [0, B^*) \cup [C^*, A^*] \\ 1 & \text{if } p_k \in [B^*, C^*) \end{cases} \tag{6}$$

$$\tau^* = \inf\{k \geq 1 : p_k > A^*\}$$

with $0 \leq B^* \leq C^* \leq A^* \leq 1$. The value of A^*, B^*, and C^* depend on the choice of λ_f and λ_c.

Remark 1. *With $B^* = 0$ and $C^* = A^*$, the algorithm in* (6) *reduces to the classical Shiryaev test [1].*

The optimal policy described in (6) has the following interpretation. Recall that p_k is the *a posteriori* probability of the change having happened before time k, given all the information up to time k. If p_k is either close to 0 or close to 1, the test has high confidence about the change, and therefore does not take a new observation and relies on the prior. If p_k is large enough, we stop and declare a change.

We next characterize the solution to the constrained optimization problem given in (1) using the dynamic programming solution structure (6) that we derived for the Lagrangian relaxation given in (2). Now, for given (A^*, B^*, C^*), γ^* in (6) has some PFA and ANO performance. For a given α and β, if λ_f and λ_c can be chosen such that PFA$(\gamma^*) = \alpha$ and ANO$(\gamma^*) = \beta$, then it is clear that the solution to (2) is also a solution to (1). In the following we refer to this case by saying that a deterministic policy is optimal. If no such λ_f and λ_c exists, then one has to randomize over three-threshold policies that are solutions to (2), to meet the constraints α and β with equality. Since our objective is to find easy-to-design data-efficient quickest change detection algorithms, we restrict our attention to the characterization of deterministic optimal policies: finding (A^*, B^*, C^*) in (6) for a given α and β.

Towards characterizing deterministic optimal policies, we make the important observation that it is not enough to find a three-threshold policy that meets the constraints α and β with equality. In Table I we show the performance obtained via simulations of a three-threshold policy for various values of thresholds, which we denote by (A, B, C), and for the parameter set: $\rho = 0.01$, $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.75, 1)$. The table clearly shows that multiple values of thresholds (A, B, C) satisfy the same constraints $\alpha = 10^{-3}$ and $\beta = 40$, and one can optimize over the choice of thresholds to get the best ADD. Thus, the dynamic programming argument only tells us that the optimal policy, i.e., the one that minimizes ADD subject to constraints on PFA and ANO, can be found within the class of three-threshold policies. Further, not all three-threshold policies are optimal solutions to the constrained optimization problem described in (1).

We will therefore denote a generic (deterministic) three-threshold policy by $\gamma(A, B, C)$, with the

TABLE I

$\rho = 0.01$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$

A	B	C	ANO	PFA	ADD
0.998993	0.182426	0.991837	40	10^{-3}	189
0.998952	0.222700	0.997268	40	10^{-3}	88
0.998876	0.231475	0.997975	40	10^{-3}	62
0.998708	0.236855	0.998341	40	10^{-3}	46
0.998499	0.239577	0.998438	40	10^{-3}	42.3
0.998449	0.240489	0.998449	40	10^{-3}	42.1

understanding that the policy may not be optimal. In order to obtain the optimal thresholds for given constraints PFA$\leq \alpha$ and ANO$\leq \beta$, i.e. to find (A^*, B^*, C^*) in (6), it is still necessary to obtain analytical expressions for ADD, PFA and ANO, and to solve the optimization problem (1) on the space of thresholds, (A, B, C). To this end, in the following section, we derive analytical approximations for ADD, PFA and ANO in terms of (A, B, C) for any three-threshold policy $\gamma(A, B, C)$.

In Section IV-A we use our analytical approximations to characterize this optimal choice of thresholds. Specifically, observe in Table I that the ADD was minimum for $A = C$. We will show in Section IV-A that this is a typical choice, i.e., a two-threshold policy is optimal among deterministic policies, unless the ANO constraint β is very severe. Even when a two-threshold policy is not optimal, we will show, in Section IV-B via simulations, that using a two-threshold in those cases results in marginal loss in performance.

In Section IV-C, we provide a possible justification for the role of the third threshold for cases where using three thresholds is optimal. In Section IV-D, we provide design guidelines for the two-threshold algorithm, and in Section IV-E we comment on the trade-off between ANO and ADD. Finally, we summarize the results and comment on future work in Section V.

III. Asymptotic analysis of $\gamma(A, B, C)$

In this section we derive asymptotic approximations for ADD, PFA and ANO for a three-threshold policy $\gamma(A, B, C)$. To that end, we first convert the recursion for p_k (see (3) and (4)) to a form that is amenable to asymptotic analysis.

Define, $Z_k = \log \frac{p_k}{1-p_k}$ for $k \geq 0$. This new variable Z_k has a one-to-one mapping with p_k. By defining

$$a = \log \frac{A}{1-A}, \quad b = \log \frac{B}{1-B}, \quad c = \log \frac{C}{1-C},$$

we can write the recursions (3) and (4) in terms of Z_k.
For $k \geq 1$,

$$Z_{k+1} = Z_k + \log L(X_{k+1}) + |\log(1-\rho)| + \log\left(1 + \rho\, e^{-Z_k}\right), \text{ if } Z_k \in [b, c) \tag{7}$$

and

$$Z_{k+1} = Z_k + |\log(1-\rho)| + \log\left(1 + \rho\, e^{-Z_k}\right), \text{ if } Z_k \notin [b, c) \tag{8}$$

with

$$Z_1 = \log\left(e^{Z_0} + \rho\right) + |\log(1-\rho)| + \log\left(L(X_1)\right) \mathbb{I}_{\{Z_0 \in [b,c)\}}.$$

Here we have used the fact that $S_{k+1} = 1$ if $p_k \in [B, C)$, and $S_{k+1} = 0$ otherwise (see (6)). The crossing of thresholds A, B, C by p_k is equivalent to the crossing of thresholds a, b, c by Z_k. Thus the stopping time for $\gamma(A, B, C)$ (equivalently $\gamma(a, b, c)$ with some abuse of notation) is

$$\tau = \inf\left\{k \geq 1 : Z_k > a\right\}.$$

In this section we study the asymptotic behavior of $\gamma(a, b, c)$ in terms of Z_k, under various limits of a, b, c and ρ. Specifically, we provide an asymptotic expression for ADD, for fixed b and ρ, as $c, a \to \infty$. We also provide, as $c, a \to \infty$ and $\rho \to 0$, asymptotic expression for PFA for fixed b, and for ANO with $b \to -\infty$. Note that the limit of $c, a \to \infty$ corresponds to PFA going to zero (and ADD, ANO $\to \infty$), and the limit of $\rho \to 0$ corresponds to a rare change event.

Fig. 1 shows a typical evolution of $\gamma(a, b, c)$, i.e., of Z_k using (7) and (8), starting at time 0. Note that for $Z_k \in [b, c)$, recursion (7) is used, while outside that interval, recursion (8), which only uses the prior ρ, is used. As a result Z_k increases monotonically outside $[b, c)$.

Define,

$$\tau_c = \inf\left\{k \geq 1 : Z_k > c\right\}.$$

From Fig. 1 again, each time Z_k crosses b from below, it can either increase to c (point τ_c), and then monotonically increase to stop at a (point τ), or it can go below b and approach b monotonically again from below, at which time it faces a similar set of alternatives. Thus the passage to threshold c possibly involves multiple cycles of the evolution of Z_k below b. We will show in Section III-B that after the change point Γ, following a finite number of cycles below b, Z_k grows up to cross c, and the time spent on the cycles below b is insignificant as compared to $\tau_c - \Gamma$, as $c, a \to \infty$. In fact we show that, asymptotically, the time to reach c is equal to the time taken by the classical Shiryaev algorithm to move from b to c. (Note that for the classical Shiryaev algorithm the evolution of Z_k would be based on only (7)).

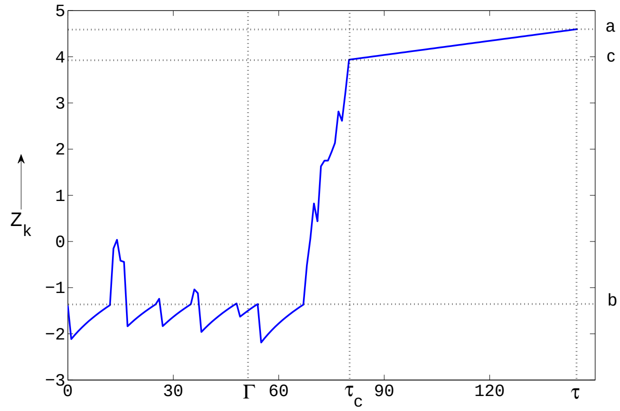

Fig. 1. Evolution of Z_k for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.5,1)$, and $\rho = 0.01$, with thresholds $a = 4.59$, $c = 3.89$, and $b = -1.38$, corresponding to the p_k thresholds $A = 0.99$, $C = 0.98$ and $B = 0.2$, respectively. Also $Z_0 = b$.

When Z_k crosses c from below, it does so with an overshoot. Overshoots play a significant role in the performance of many sequential algorithms (see [14], [17]) and they are central to the performance of $\gamma(a,b,c)$ as well. In Section III-C, we show that PFA depends on the overshoot $(Z_{\tau_c} - c)$ as $c \to \infty$, and on thresholds c and a, but is *not* a function of the threshold b. The overshoot distribution is also used to approximate the time for Z_k to move from c to a.

The number of observations taken during the detection process is the total time spent by Z_k between b and c. As $c, a \to \infty$, Z_k crosses c only after change point Γ, with high probability. The total number of observation taken can thus be divided in to two parts: one taken before Γ, which is the fraction of time Z_k is above b (and hence depends only on b), and the part consumed after Γ. In Section III-D we show that, asymptotically, the average number of observations used after Γ is approximately equal to the delay itself.

In Section III-E we provide numerical results to demonstrate that under various scenarios, for limiting as well as moderate values of a, b, c and ρ, our asymptotic expressions for ADD, PFA and ANO provide good approximations.

In Section IV we use the asymptotic expressions for ADD, PFA and ANO to argue that the optimal three-threshold policy $\gamma^*(a^*, b^*, c^*)$ for given constraints PFA$\leq \alpha$ and ANO$\leq \beta$, can be well approximated by a two-threshold policy. We also provide numerical and simulation results to support the claim.

We begin our analysis by first obtaining the asymptotic overshoot distribution for $(Z_{\tau_c} - c)$ using nonlinear renewal theory [14], [15]. As mentioned above, this will be useful for the ADD analysis and will be critical to the PFA analysis. For convenience of reference, in Table II, we provide a glossary of important terms used in this paper.

TABLE II
GLOSSARY

Symbol	Definition/Interpretation	Symbol	Definition/Interpretation
ADD	Average detection delay	λ	Starting at b, first time Z_k is outside $[b, c)$
PFA	Probability of false alarm	Λ	Starting at b, first time Z_k crosses c
ANO_0	Average # observations used before change		or crosses b from below
ANO_1	Average # observations used after change	ADD^s	Starting at b, time for Z_k to reach c under P_1, when
$\{X_k\}$	Observation sequence		Z_k is reset to b each time it crosses b from below
p_k	*a posteriori* probability of change	$\lambda(x)$	Starting at $x \geq b$, first time Z_k is outside $[b, c)$
Z_k	$\log \frac{p_k}{1-p_k} = \sum_{i=1}^k Y_i + \eta_k$,	$\Lambda(x)$	Starting at $x \geq b$, first time Z_k crosses c
τ (τ_c)	First time for p_k to cross A (C) or		or crosses b from below
	first time for Z_k to cross $a = \log \frac{A}{1-A}$ (c)	$\hat{\lambda}$	Starting at b, first time $Z_k < b$ with $c = \infty$
$\{\eta_k\}$	Slowly changing sequence	$\hat{\lambda}(x)$	Starting at $x \geq b$, first time $Z_k < b$ with $c = \infty$
$R(x)$, $\bar{r}$	Asymptotic distribution and mean of overshoot	T_b	Time spent by Z_k below b, after Γ, when $\tau \geq \Gamma$
	when $\sum_{i=1}^k Y_i$ crosses a large threshold	$\tilde{\Lambda}^x$	Starting at $x \geq b$, first time $Z_k > c$, or crosses b from
$t(x, y)$	Time for Z_k to reach y starting at x using (8)		below, or is stopped by occurrence of change
$\nu(x, y)$	Time for Z_k to reach y starting at x using (7)	δ^x	The fraction of time Z_k is above b, when stopped by $\tilde{\Lambda}^x$
	also, time for Shiryaev test to reach y starting at x	$\tilde{\nu}_b$ ($\hat{\nu}_b$)	Starting at b, time for Z_k to reach c, when Z_k is
ν_b, ν_0	$\nu(b, c)$ and $\nu(-\infty, c)$		reflected at b (reset to b when it crosses b from below)

In what follows, we use E_ℓ and P_ℓ to denote, respectively, the expectation and probability measure when change happens at time ℓ. We use E_∞ and P_∞ to denote, respectively, the expectation and probability measure when the entire sequence $\{X_n\}$ is i.i.d. with density f_0. Also, $g(x) = o(1)$ as $x \to x_0$ is used to denote that $g(x) \to 0$ in the specified limit.

A. *Asymptotic overshoot*

In this section we characterize the overshoot distribution of Z_k as it crosses c as $c \to \infty$. For this analysis, we can therefore assume that $Z_k < c$. Also, in analyzing the trajectory of Z_k, it useful to allow for arbitrary starting point Z_0 (shifting the time axis). We first combine the recursions in (7) and (8) to get:

$$Z_{k+1} = Z_k + \mathbb{I}_{\{Z_k \geq b\}} \log L(X_{k+1}) + |\log(1-\rho)| + \log\left(1 + e^{-Z_k}\rho\right).$$

By defining $Y_k = \log L(X_k) + |\log(1-\rho)|$ and expanding the above recursion, we can write an expression for Z_n:

$$\begin{aligned} Z_n &= \sum_{k=1}^{n} Y_k + \log\left(e^{Z_0} + \rho\right) + \sum_{k=1}^{n-1} \log\left(1 + e^{-Z_k}\rho\right) - \sum_{k=1}^{n} \mathbb{I}_{\{Z_k < b\}} \log L(X_k) \\ &= \sum_{k=1}^{n} Y_k + \eta_n. \end{aligned} \tag{9}$$

Here η_n is used to represent all terms other than the first in the equation above:

$$\eta_n = \log\left(e^{Z_0} + \rho\right) + \sum_{k=1}^{n-1} \log\left(1 + e^{-Z_k}\rho\right) - \sum_{k=1}^{n} \mathbb{I}_{\{Z_k < b\}} \log L(X_k). \tag{10}$$

As defined in [14], η_n is a *slowly changing* sequence if

$$n^{-1} \max\{|\eta_1|, \dots, |\eta_n|\} \xrightarrow[i.p.]{n\to\infty} 0, \tag{11}$$

and for every $\epsilon > 0$, there exists n^* and $\delta > 0$ such that for all $n \geq n^*$

$$\mathrm{P}\{\max_{1\leq k\leq n\delta} |\eta_{n+k} - \eta_n| > \epsilon\} < \epsilon. \tag{12}$$

If indeed $\{\eta_n\}$ is a slowly changing sequence, then the distribution of $Z_{\tau_c} - c$, as $c \to \infty$, is equal to the asymptotic distribution of the overshoot when the random walk $\sum_{k=1}^{n} Y_k$ crosses a large positive boundary. We have the following result.

Theorem 1. *Let $R(x)$ be the asymptotic distribution of the overshoot when the random walk $\sum_{k=1}^{n} Y_k$ crosses a large positive boundary under P_1. Then for fixed ρ and b, under P_1, we have the following:*

1) *$\{\eta_n\}$ is a slowly changing sequence.*
2) *$R(x)$ is the distribution of $Z_{\tau_c} - c$ as $c \to \infty$, i.e.,*

$$\lim_{c\to\infty} \mathrm{P}\left[Z_{\tau_c} - c \leq x | \tau_c \geq \Gamma\right] = R(x). \tag{13}$$

Proof: When $b = -\infty$, Z_k evolves as in the classical Shiryaev test statistic, and it is easy to see that in this case:

$$\begin{aligned} \eta_n &= \left[\log\left(e^{Z_0} + \rho\right) + \sum_{k=1}^{n-1} \log\left(1 + e^{-Z_k}\rho\right)\right] \\ &= \log\left[e^{Z_0} + \sum_{k=0}^{n-1} \rho(1-\rho)^k \prod_{i=1}^{k} \frac{f_0(X_i)}{f_1(X_i)}\right]. \end{aligned}$$

It was shown in [17] that this $\{\eta_n\}$ sequence (for $b = -\infty$), with $Z_0 = -\infty$, is a slowly changing sequence. It is easy to show that $\{\eta_n\}$ is a slowly changing sequence even if Z_0 is a random variable. Also, if L_Z is the last time Z_k crosses b from below, then note that, after L_Z, the last term $\sum_{k=1}^{n} \mathbb{I}_{\{Z_k<b\}} \log L(X_k)$ in (10) vanishes, and η_n in (10) behaves like the η_n for $b = -\infty$. We prove the theorem using these observations. The detailed proof is given in the appendix to this section. ■

B. Delay Analysis

The PFA for $\gamma(a,b,c)$ can be shown to have the following expression and bound [17]:

$$\text{PFA} = \mathrm{E}[1 - p_\tau] \leq 1 - A = \frac{1}{1+e^a} \leq e^{-a}. \tag{14}$$

We will later show that this upper bound is tight if the gap between c and a is significant enough. Using this upper bound we can show that the ADD of $\gamma(a,b,c)$ is given by:

$$\begin{aligned} \text{ADD} &= \mathrm{E}\left[(\tau-\Gamma)^+\right] \\ &= \mathrm{E}[\tau - \Gamma | \tau \geq \Gamma](1 + o(1)) \quad \text{as } c, a \to \infty. \end{aligned} \tag{15}$$

As $c, a \to \infty$, the conditional delay $\mathrm{E}[\tau - \Gamma|\tau \geq \Gamma]$ will be due to the sample paths in which Z_k crosses c after the change point Γ, i.e. $\tau_c \geq \Gamma$. The following lemma establishes that the conditional delay can be written as a sum of two other conditional incremental delay terms. We need the following definition. Let $t(x,y)$ be the constant time taken by Z_k to move from $Z_0 = x$ to y using the recursion (8), i.e.

$$t(x,y) \stackrel{\Delta}{=} \inf\{k \geq 0 : Z_k > y, Z_0 = x,\ x,y \notin [b,c)\}. \tag{16}$$

Lemma 1. *For fixed ρ and b, if $t(c,a)$ is bounded as $c,a \to \infty$, then as $c,a \to \infty$,*

$$\mathrm{E}[\tau - \Gamma|\tau \geq \Gamma] = \left(\mathrm{E}[\tau - \tau_c|\tau_c \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma]\right)(1 + o(1)) \tag{17}$$

Proof: See the appendix to this section for the proof. ■

In the following, we provide asymptotic expressions for $\mathrm{E}[\tau - \tau_c|\tau_c \geq \Gamma]$ and $\mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma]$.

1) Asymptotic expression for $\mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma]$: Let Ψ represent the Shiryaev recursion, i.e., updating Z_k using only (7). Define

$$\nu(x,y) = \inf\left\{k \geq 1 : \Psi(Z_{k-1}) > y, \quad Z_0 = x\right\}. \tag{18}$$

Thus, $\nu(x, y)$ is the time for the Shiryaev test to reach y starting at x. Also, define the stopping times:

$$\nu_b = \nu(b, c), \tag{19}$$

and

$$\nu_0 = \nu(-\infty, c). \tag{20}$$

Note that, ν_0 is the stopping time for the classical Shiryaev test [1] and ν_b is its modified form which starts at b.

From Theorem 1 in [17],

$$\mathrm{E}[\nu_0 - \Gamma | \nu_0 \geq \Gamma] \geq \frac{c}{D(f_1, f_0) + |\log(1-\rho)|}(1 + o(1)) \text{ as } c \to \infty,$$

where, $D(f_1, f_0)$ is the K-L divergence between f_0 and f_1. Also based on the second order approximation for $\mathrm{E}_1[\nu_0]$ developed in [17], we have obtained the following approximation for $\mathrm{E}_1[\nu_b]$:

$$\mathrm{E}_1[\nu_b] = \frac{c - \mathrm{E}[\eta(b)] + \bar{r}}{D(f_1, f_0) + |\log(1-\rho)|} + o(1) \text{ as } c \to \infty, \tag{21}$$

where, $\eta(b)$ is the a.s. limit of the slowly changing sequence η_n with $Z_0 = b$ under f_1, (see (10) and (46)), and

$$\bar{r} = \int_0^\infty x dR(x), \tag{22}$$

with $R(x)$ as in Theorem 1. Since $\eta(b)$ is not a function of the threshold c, we have

$$\mathrm{E}[\nu_0 - \Gamma | \nu_0 \geq \Gamma] = \mathrm{E}_1[\nu_b](1 + o(1)) \quad \text{as} \quad c \to \infty. \tag{23}$$

Using (23) we prove the following lemma.

Lemma 2. *For fixed b and ρ,*

$$\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] \geq \mathrm{E}_1[\nu_b](1 + o(1)) \textit{ as } c \to \infty.$$

Proof: We have for any b and c,

$$\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] \geq \mathrm{E}[\nu_0 - \Gamma | \nu_0 \geq \Gamma].$$

This is true because skipping observations can only lead to larger delay. The result then follows from (23). ■

In the following we show that $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$ is also asymptotically upper bounded by $\mathrm{E}_1[\nu_b]$.

It was discussed in reference to Fig. 1 that each time Z_k crosses b from below, it faces two alternatives, to cross c without ever coming back to b or to go below b and cross it again from below. It was mentioned that the passage to the threshold c is through multiple such cycles. Motivated by this we define the following stopping times λ and Λ:

$$\lambda \triangleq \inf\{k \geq 1 : Z_k \notin [b, c), Z_0 = b\}, \tag{24}$$

and

$$\Lambda \triangleq \inf\{k \geq 1 : Z_k > c \;\text{ or } \exists\; k \; s.t. \; Z_{k-1} < b \text{ and } Z_k \geq b \;, Z_0 = b\}. \tag{25}$$

We can write Λ as a function of λ using (16):

$$\Lambda = (\lambda + t(Z_\lambda, b))\mathbb{I}_{\{Z_\lambda < b\}} + \lambda\,\mathbb{I}_{\{Z_\lambda > c\}} = \lambda + t(Z_\lambda, b)\mathbb{I}_{\{Z_\lambda < b\}}.$$

The significance of these stopping times is as follows. If we start the process at $Z_0 = b$ and *reset* Z_k *to* b *each time it crosses* b *from below*, then the time taken by Z_k to move from b to c is the sum of a finite but random number of random variables with distribution of Λ, say $\Lambda_1, \Lambda_2, \ldots, \Lambda_N$. For $i = 1, \ldots, N-1$, $Z_{\Lambda_i} < b$, and $Z_{\Lambda_N} > c$. Thus the time to reach c in this case is $\mathrm{E}_1\left[\sum_{k=1}^N \Lambda_k\right]$.

Define $\mathrm{ADD}^s = \mathrm{E}_1\left[\sum_{k=1}^N \Lambda_k\right]$. Lemma 3 shows that $\mathrm{E}_1\left[\sum_{k=1}^N \Lambda_k\right]$ is the dominant term in an upper bound to $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$ as $c, a \to \infty$.

Lemma 3. *For a fixed* b *and* ρ*, we have as* $c, a \to \infty, c < a$

$$\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] \leq \mathrm{ADD}^s\,(1 + o(1))\,. \tag{26}$$

Proof: The proof is provided in the appendix. ■

We next show that $\mathrm{E}_1[\nu_b]$ is asymptotically equivalent to ADD^s, and is hence an asymptotic upper bound for $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$.

Lemma 4. *For a fixed* b*,* ADD^s*, the average time for* Z_k *to cross* c *starting at* b*, under* P_1*, with* Z_k *reset to* b *each time it crosses* b *from below, is asymptotically equal to the corresponding time taken by the Shiryaev recursion, i.e.,*

$$\mathrm{ADD}^s = \mathrm{E}_1[\nu_b](1 + o(1)) \text{ as } c, a \to \infty.$$

Hence,

$$\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] \leq \mathrm{E}_1[\nu_b]\,(1 + o(1)) \;\text{ as } c, a \to \infty.$$

Proof: We provide a sketch of the proof here. The details are provided in the appendix. Note that

$$\begin{aligned}
\mathrm{ADD}^s &= \mathrm{E}_1\left[\sum_{k=1}^{N}\Lambda_k\right] \\
&\overset{(i)}{=} \mathrm{E}_1[N]\mathrm{E}_1[\Lambda] \\
&\overset{(ii)}{=} \frac{\mathrm{E}_1[\Lambda]}{\mathrm{P}_1(Z_\lambda > c)} \\
&= \frac{\mathrm{E}_1[\lambda] + \mathrm{E}_1[t(Z_\lambda, b)|\{Z_\lambda < b\}]\mathrm{P}_1(Z_\lambda < b)}{\mathrm{P}_1(Z_\lambda > c)}.
\end{aligned}$$

In the above equation, equality (i) follows from Wald's lemma [14], and equality (ii) follows because $N \sim \mathrm{Geom}(P(Z_\lambda > c))$. The main idea of the proof is to find stopping times which upper and lower bound the Shiryaev time on average and have delay equal to $\frac{\mathrm{E}_1[\lambda]}{\mathrm{P}_1(Z_\lambda > c)}$ as $c \to \infty$. Finally, we use Lemma 3. ■

We have thus proved the following theorem.

Theorem 2. *For a fixed b and ρ, we have as $c, a \to \infty$,*

$$\mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma] = \mathrm{E}_1[\nu_b]\left(1 + o(1)\right). \tag{27}$$

2) Asymptotic expression for $\mathrm{E}[\tau - \tau_c|\tau_c \geq \Gamma]$: The time for Z_k to reach a after it has crossed c is non-zero only if the overshoot $Z_{\tau_c} - c < a - c$. If the overshoot is $x < a - c$, then the time taken is $t(c+x, a)$. Since, with $c \to \infty$, the distribution of $Z_{\tau_c} - c$ is $R(x)$, one can approximate $\mathrm{E}[\tau - \tau_c|\tau_c \geq \Gamma]$ for large c, by averaging $t(c + x, a)$ over $R(x)$.

We first prove a lemma in which we obtain asymptotic upper and lower bounds on $t(x, y)$.

Lemma 5. *For a fixed value of ρ,*

$$\left(\frac{y - x}{|\log(1-\rho)|}\right)(1 + o(1)) \leq t(x, y) \leq \left(\frac{y - x}{|\log(1-\rho)|} + 1\right)(1 + o(1)) \text{ as } x, y \to \infty. \tag{28}$$

Also, for fixed values of x and y, we have

$$t(x, y) = \left(\frac{\log(1 + e^y) - \log(1 + e^x)}{|\log(1-\rho)|}\right)(1 + o(1)) \text{ as } \rho \to 0. \tag{29}$$

Proof: The proof is provided in the appendix. ■

We use Lemma 5 to prove the following theorem.

Theorem 3. *As* $c, a \to \infty$,

$$\left(\int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|} dR(x)\right)(1+o(1)) \leq \mathrm{E}[\tau - \tau_c | \tau_c \geq \Gamma] \leq \left(\int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|} dR(x) + R(a-c)\right)(1+o(1)). \tag{30}$$

The asymptotic upper and lower bounds differ by $R(a-c)$*, which being a distribution satisfies* $0 \leq R(a-c) \leq 1$*. Further if* $a - c \to \infty$ *or* $a - c \to$ *constant* > 0 *with* $\rho \to 0$*, then we have*

$$\mathrm{E}[\tau - \tau_c | \tau_c \geq \Gamma] = \left(\int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|} dR(x)\right)(1+o(1)). \tag{31}$$

Proof: The proof is provided in the appendix. ■

C. PFA *Analysis*

We know from equation (14) that PFA can be written as $\mathrm{E}[1-p_\tau]$. We first obtain an expression for PFA as a function of the overshoot when Z_k crosses a.

Lemma 6. *For fixed* ρ *and* b*, as* $c, a \to \infty$

$$\mathrm{PFA} = \mathrm{E}[1-p_\tau] = e^{-a}\mathrm{E}[e^{-(Z_\tau - a)}](1+o(1)).$$

Proof: See the appendix for the proof. ■

From Lemma 6, it is evident that PFA depends on the overshoot when Z_k crosses a as $a \to \infty$. This overshoot in turn depends on the overshoot of Z_k when it crosses c. Since the latter has an asymptotic distribution (Theorem 1) that depends only on densities f_0, f_1 and prior ρ, and is independent of b, it is natural to expect that as $c \to \infty$, PFA is completely characterized by the asymptotic distribution $R(x)$ and is not a function of the threshold b. This is indeed true and is established using the following argument.

When Z_k crosses c it can either directly jump above a with an overshoot greater than $a - c$, i.e., $\{Z_{\tau_c} > a\}$, or cross c with an overshoot less than $a - c$, i.e., $\{Z_{\tau_c} \leq a\}$. In the former case, the false alarm is then a function of the asymptotic distribution $R(x)$ as $c \to \infty$. In the latter case, because Z_k crosses a with the help of only the prior, the overshoot is small and goes to zero as $\rho \to 0$ (8). As $a \to \infty$, $Z_{\tau-1} \gg 0$. Hence,

$$Z_\tau - Z_{\tau-1} = \log\left(\frac{1+e^{-Z_{\tau-1}}\rho}{1-\rho}\right) \leq \log\left(\frac{1+\rho}{1-\rho}\right) \text{ as } a \to \infty.$$

Thus on the set $\{Z_\tau \leq a\}$, $Z_\tau - a \approx 0$ for ρ small enough, and in this case $\mathrm{PFA} \approx e^{-a}$. Based on this idea we have obtained an asymptotic expression for PFA.

Theorem 4. *For b fixed, as $c, a \to \infty,$ and $\rho \to 0$,*

$$\text{PFA} = \left(e^{-a} R(a-c) + e^{-c} \int_{a-c}^{\infty} e^{-x} dR(x) \right) (1 + o(1)). \tag{32}$$

Proof: The proof is provided in the appendix. ■

D. Computation of ANO

As $c \to \infty$, $\tau_c \geq \Gamma$ with high probability. As a result, the total number of observations used can be separated in two parts, one used before Γ and the other used after Γ. The part used before Γ is the fraction of time the process Z_k is above b. The part used after Γ, for large c, is approximately the time taken by Z_k to reach c. We obtain an asymptotic expression for ANO based on the above ideas.

First note that,

$$\begin{aligned}
\text{ANO} &= \mathrm{E}\left[\sum_{k=1}^{\tau} S_k\right] = \mathrm{E}\left[\sum_{k=1}^{\tau_c} S_k\right] \\
&= \mathrm{E}\left[\sum_{k=1}^{\tau_c} S_k \middle| \tau_c \geq \Gamma\right] \mathrm{P}(\tau_c \geq \Gamma) + \mathrm{E}\left[\sum_{k=1}^{\tau_c} S_k \middle| \tau_c < \Gamma\right] \mathrm{P}(\tau_c < \Gamma) \\
&= \mathrm{E}\left[\sum_{k=1}^{\tau_c} S_k \middle| \tau_c \geq \Gamma\right] (1 + o(1)) \quad \text{as} \quad c \to \infty.
\end{aligned}$$

The last equality follows because $\sum_{k=1}^{\tau_c} S_k \leq \Gamma$ on $\{\tau_c < \Gamma\}$, and $\mathrm{P}(\tau_c < \Gamma) < e^{-c} \to 0$ as $c \to \infty$.

Define ANO_0 as the average number of observations used before Γ, and ANO_1 as the average number of observations used after Γ, conditioned on the event $\{\tau_c \geq \Gamma\}$. We can then write ANO as

$$\begin{aligned}
\text{ANO} &= \left(\mathrm{E}\left[\sum_{k=1}^{\tau_c} S_k \middle| \tau_c \geq \Gamma\right] \right) (1 + o(1)) \\
&= \left(\mathrm{E}\left[\sum_{k=1}^{\Gamma-1} S_k \middle| \tau_c \geq \Gamma\right] + \mathrm{E}\left[\sum_{k=\Gamma}^{\tau_c} S_k \middle| \tau_c \geq \Gamma\right] \right) (1 + o(1)) \\
&= (\text{ANO}_0 + \text{ANO}_1)(1 + o(1)) \quad \text{as} \quad c \to \infty.
\end{aligned}$$

Following (24), we define

$$\hat{\lambda} = \inf\{k \geq 1 : Z_k < b, Z_0 = b, c = \infty\}. \tag{33}$$

The theorem below gives asymptotic expressions for ANO_0 and ANO_1.

Theorem 5. *For fixed threshold b, we have as $c, a \to \infty$,*

$$\text{ANO}_1 = \mathrm{E}_1[\nu_b](1 + o(1)),$$

and as $c, a \to \infty, \rho \to 0$*, and* $b \to -\infty$*, with* ρ *taken to 0 before* b *is taken to* $-\infty$*,*

$$\mathrm{ANO}_0 = \frac{\rho^{-1}\ \mathrm{E}_\infty[\hat{\lambda}]}{\mathrm{E}_\infty[\hat{\lambda}] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]} \frac{1}{1+e^b}(1+o(1)),$$

where, $\hat{\lambda}$ *is as defined in* (33).

Proof: The number of observations used after Γ can be written as the difference between the time for Z_k to reach c and the time spend by it below b. For this we define the variable

$$T_b \triangleq \mathrm{E}\left[\sum_{k=\Gamma}^{\tau_c} 1_{\{Z_k < b\}} \middle| \tau_c \geq \Gamma\right].$$

Thus

$$\mathrm{ANO}_1 = \mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right] - T_b + 1.$$

We know from Theorem 2 that $\mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right] \approx \mathrm{E}_1[\nu_b]$. The following lemma shows that as $c \to \infty$, T_b converges, and even $\mathrm{ANO}_1 \approx \mathrm{E}_1[\nu_b]$ for large c.

Lemma 7. *For a given* ρ *and* b*,*

$$\mathrm{ANO}_1 = \mathrm{E}_1[\nu_b](1+o(1)) \quad \textit{as} \quad c, a \to \infty.$$

Proof: The proof is provided in the appendix. ■

For computation of ANO_0, we allow for the possibility that the process $\{Z_k\}$ started with $Z_0 = z_0 \neq -\infty$, $z_0 < b$. Let $t(b)$ be the first time Z_k crossed b from below, i.e., $t(b) = t(z_0, b)$. Using the fact that observations are used only after $t(b)$, we can write the following:

$$\begin{aligned}\mathrm{ANO}_0 &= \mathrm{E}\left[\sum_{k=1}^{\Gamma-1} S_k \middle| \tau_c \geq \Gamma\right] \\ &= \mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), \tau_c \geq \Gamma\right] \mathrm{P}(\Gamma > t(b) | \tau_c \geq \Gamma).\end{aligned} \tag{34}$$

We now compute each of the two terms in (34). For the first term in (34), we have the following lemma.

Lemma 8. *For a fixed* b*, as* $c, a \to \infty$*,* $\rho \to 0$*, and* $b \to -\infty$*, with* ρ *taken to 0 before* b *is taken to* $-\infty$*,*

$$\mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), \tau_c \geq \Gamma\right] = \frac{\rho^{-1}\ \mathrm{E}_\infty[\hat{\lambda}]}{\mathrm{E}_\infty[\hat{\lambda}] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]}(1+o(1)).$$

Proof: Note that

$$\lim_{c,a\to\infty} \mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), \tau_c \geq \Gamma\right] = \mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), c = \infty\right].$$

To compute the right hand side of the above equation, note that conditioned on $\{\Gamma > t(b)\}$, $\sum_{k=t(b)}^{\Gamma-1} S_k$ is approximately the number of observations used when the process Z_k starts at $Z_0 = b$, goes through multiple cycles below b, with each cycle length having distribution of $\hat{\lambda}$, and the sequence of cycles is interrupted by occurrence of change. See the appendix for the detailed proof. ∎

For the second term in (34), we show that $\mathrm{P}(\Gamma > t(b)|\tau_c \geq \Gamma)$ is equal to $\frac{1}{1+e^b}$ in the limit and is independent of z_0.

Lemma 9.

$$\mathrm{P}(\Gamma > t(b)|\tau_c \geq \Gamma) = \frac{1}{1+e^b} + o(1) \quad \textit{as} \quad c, a \to \infty, \rho \to 0.$$

Proof: The proof is provided in the appendix. ∎

The Lemmas 7-9 taken together completes the proof of Theorem 5. ∎

Simpler Approximation for ANO_0*:* Invoking Wald's lemma [14], we write $\mathrm{E}_\infty[\hat{\lambda}]$ as,

$$\mathrm{E}_\infty[\hat{\lambda}] = \frac{\mathrm{E}_\infty[Z_{\hat{\lambda}}] - \mathrm{E}_\infty[\eta_{\hat{\lambda}}]}{-D(f_1, f_0) + |\log(1-\rho)|}.$$

We have developed the following approximation for $\mathrm{E}_\infty[\hat{\lambda}]$:

$$\mathrm{E}_\infty[\hat{\lambda}] \approx \frac{\bar{r} + \log(1+\rho e^{-b})}{D(f_1, f_0) - |\log(1-\rho)|}. \tag{35}$$

Here, $\log(1+\rho e^{-b})$ is an approximation to $\mathrm{E}_\infty[\eta_{\hat{\lambda}}]$ by ignoring all the random terms after b is factored out of it. This extra b will cancel with the b in $\mathrm{E}_\infty[Z_{\hat{\lambda}}] = b + \mathrm{E}_\infty[Z_{\hat{\lambda}} - b]$. We approximate $\mathrm{E}_\infty[b - Z_{\hat{\lambda}}]$ by $\bar{r}$, the mean overshoot of the random walk $\sum_{i=1}^{k} Y_k$, with mean $D(f_1, f_0) - |\log(1-\rho)|$, when it crosses a large boundary (see (9)).

For the term $\mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]$, we use (29) and the steps followed in the proof of Theorem 3 to get the following approximation:

$$\mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)] \approx \int_0^\infty \frac{\log(1+e^b) - \log(1+e^{b-x})}{|\log(1-\rho)|} dR(x). \tag{36}$$

Thus, we approximate the distribution of $(b - Z_{\hat{\lambda}})$ by $R(x)$. As we will see in the next section, both of these approximations work well for Gaussian observations.

E. Numerical Results

In Sections III-B-III-D, we have obtained asymptotic expressions for ADD, PFA, and ANO as a function of the system parameters: the thresholds a, b, c, the densities f_0 and f_1, and the prior ρ. We summarize the results below for convenience of reference. We write ν_b as $\nu(b, c)$ to show its dependence on b and c. Also, we write $\hat{\lambda}$ (33) as $\hat{\lambda}(b)$ to indicate its dependence on b. The asymptotic approximations are:

$$\begin{aligned}
\text{ADD} &\approx \mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] + \mathrm{E}[t(Z_{\tau_c}, a) | \tau_c \geq \Gamma] \\
&\approx \mathrm{E}_1[\nu(b,c)] + \int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|} dR(x), \qquad (37)\\
\text{PFA} &\approx e^{-a} R(a-c) + e^{-c} \int_{a-c}^{\infty} e^{-x} dR(x), \qquad (38)\\
\text{ANO} &\approx \text{ANO}_0 + \text{ANO}_1 \\
&\approx \frac{\rho^{-1}}{1+e^b} \frac{\mathrm{E}_\infty[\hat{\lambda}(b)]}{\mathrm{E}_\infty[\hat{\lambda}(b)] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}(b)}, b)]} + \ \mathrm{E}_1[\nu(b,c)]. \qquad (39)
\end{aligned}$$

Recall that an approximation for $\mathrm{E}_1[\nu(b,c)]$ was obtained in (21), based on the result from [17], and a simpler approximation for ANO_0 was developed using (35) and (36).

In this section, we compare (37)-(39) with simulation results to demonstrate the accuracy of these approximations. We assume that the observations are Gaussian with $f_0 \sim \mathcal{N}(0,1)$, and $f_1 \sim \mathcal{N}(\theta,1)$, $\theta > 0$, for the simulations and analysis. In the simulations, the PFA values are computed using the expression $\mathrm{E}[1-p_\tau]$ given in (14). This guarantees a faster convergence for small values of PFA. Also, we define

$$\begin{aligned}
\text{ANO\%} = \text{ANO}\ & \textit{expressed as the percentage of the average number of} \\
& \textit{observations used by the Shiryaev test that achieves the same}\ \text{PFA}.
\end{aligned} \qquad (40)$$

Thus, a small ANO% corresponds to a large saving in the average number of observations used for detection by $\gamma(a,b,c)$. In Section IV we will show that the optimal choice of thresholds for $\gamma(a,b,c)$ depends on the ANO%.

In Sections III-B, we identified limits under which $\mathrm{E}_1[\nu(b,c)]$ is a good approximation for $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$: for fixed b, as $c, a \to \infty$. Clearly, in this limit ANO% increases to 100. In Table III, we fix $b = 1.0$ and increase a, and compare $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$, obtained using simulations, to $\mathrm{E}_1[\nu(b,c)]$ from (37). We see in Table III that the approximation improves as the ANO% increases. In general $\mathrm{E}_1[\nu(b,c)]$ is a good approximation for $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$ when ANO% is large, but may not be a good approximation when ANO% is small.

In Table IV we compare $\mathrm{E}_1[\nu(b,c)]$ and $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$ for various values of ρ, thresholds a, b, c, with $c = a$, and post change mean θ. The table demonstrates that the analytical approximation is quite accurate even for these moderate values of the parameters chosen and for 20-50% of savings in the average number of observations used (note the ANO% in the table). We also tabulate the corresponding

TABLE III

$\rho = 0.05$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$, $c = a$

a	b	Simulations $\mathrm{E}[\tau_c - \Gamma \mid \tau_c \geq \Gamma]$	Analysis $\mathrm{E}_1[\nu(b,c)]$	ANO%
1.2	1.0	16	1.7	9%
5.0	1.0	29.6	13.1	37%
9.0	1.0	41.7	25.2	52%
18.0	1.0	68.8	52.2	70%
50.0	1.0	165	149	86%
100.0	1.0	315	299	93%

values of PFA achieved using analysis (38) and using simulations. A comparison of PFA values shows that (38) is quite accurate for the parameters chosen.

TABLE IV

$f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(\theta,1)$, $c = a$

				ADD		PFA		ANO%
θ	ρ	a	b	Simulations $\mathrm{E}[\tau_c - \Gamma \mid \tau_c \geq \Gamma]$	Analysis $\mathrm{E}_1[\nu(b,c)]$	Simulations	Analysis	
0.4	0.01	8.5	-2.2	104.9	111.7	1.608×10^{-4}	1.608×10^{-4}	83%
0.75	0.01	6.467	-2.2	32.3	29.5	1.002×10^{-3}	1.004×10^{-3}	49%
2.0	0.01	7.5	-4.0	6.1	6.23	1.77×10^{-4}	1.768×10^{-4}	47%
0.75	0.005	8.7	-3.0	42.6	40.4	1.076×10^{-4}	1.076×10^{-4}	48%
0.75	0.1	8.5	0.0	23.9	22.18	1.286×10^{-4}	1.285×10^{-4}	75%

To further show the accuracy of (38) as an approximation for PFA, in Table V we compare (38) with the PFA obtained using simulations of $\gamma(a,b,c)$ for the same choice of thresholds (a,b,c). Note that in comparison with Table IV, here $c < a$. From the table we see that (38) gives a very good estimate of PFA.

In Table VI, we show that PFA is not a function of b for large values of c and a. We fix $a = 4.6$ and $c = 3.89$, and increase b from -2.2 to 0.85. We notice that PFA is unchanged in simulations when b is changed this way. This is also captured by the analysis and it is quite accurate.

In Table VII we demonstrate the accuracy of ANO approximations, ANO_0 and ANO_1 (39), for the same set of parameters as in Table IV. The table shows that the approximations in (39) are quite accurate

TABLE V

PFA: FOR $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(\theta,1)$, $c < a$

θ	ρ	a	b	c	PFA Simulations	PFA Analysis
0.4	0.01	3.0	0	2.5	4.63×10^{-2}	4.87×10^{-2}
0.4	0.01	6.0	2.0	5.8	2.239×10^{-3}	2.253×10^{-2}
0.75	0.01	9.0	-2.0	9.0	7.968×10^{-5}	7.964×10^{-5}
2.0	0.01	5.0	-4.0	-1.0	6.649×10^{-3}	6.72×10^{-3}
0.75	0.005	7.6	3.0	7.5	3.531×10^{-4}	3.535×10^{-4}
0.75	0.1	4.0	-3.0	2.0	1.71×10^{-2}	1.83×10^{-2}

TABLE VI

PFA for $\rho = 0.01$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$

a	b	c	Simulations	Analysis
4.6	-2.2	3.89	9.2×10^{-3}	9.24×10^{-3}
4.6	-1.5	3.89	9.2×10^{-3}	9.24×10^{-3}
4.6	-0.85	3.89	9.2×10^{-3}	9.24×10^{-3}
4.6	0	3.89	9.2×10^{-3}	9.24×10^{-3}
4.6	0.85	3.89	9.2×10^{-3}	9.24×10^{-3}

for the parameters chosen.

Table VIII shows the comparison of simulations and analysis for $\mathrm{E}[t(Z_{\tau_c}, a)|\tau_c \geq \Gamma]$, as provided in equation (37). We tabulate the result for various values of ρ and thresholds a and c. The values indicate that the approximation is quite accurate.

TABLE VII

$f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(\theta,1)$, $c = a$

				ANO_0		ANO_1	
θ	ρ	a	b	Simulations	Analysis	Simulations	Analysis
0.4	0.01	8.5	-2.2	66.3	62.88	102.9	111.7
0.75	0.01	6.467	-2.2	34.92	34.24	27.86	29.46
2.0	0.01	7.5	-4.0	42.94	46.4	6.08	6.23
0.75	0.005	8.7	-3.0	77.18	75.09	38.73	40.38
0.75	0.1	8.5	0.0	2.64	3.2	21.17	22.18

TABLE VIII

$\mathrm{E}[t(Z_{\tau_c}, a)|\tau_c \geq \Gamma]$ for $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$

ρ	a	c	Simulations	Analysis
0.01	6.9	4.6	179	179
0.01	4.6	3.9	29.2	29.2
0.2	4.6	3.9	1.5	1.2
0.2	9.2	3.89	21.7	21.2
0.001	2.31	2.2	9.46	9.8

Remark 2. *In Section IV-A we show that the optimal solution to the problem in (1), with* ADD, PFA, *and* ANO *given by the expressions in (37), (38) and (39), has a two-threshold structure. The optimality arguments there depend on* $\mathrm{E}_1[\nu(b,c)]$ *being a good approximation for both* ADD *(with* $c = a$*) and* ANO_1, *and also on the fact that* PFA *does not depend on the threshold* b. *Hence, based on the numerical results shown in Table III-VII, we may surmise that a two-threshold policy is approximately optimal when* ANO% *is large.*

IV. TWO-THRESHOLD STRUCTURE

In this section, using the analytical results developed so far, i.e., using (37)-(39), we argue that the three-threshold policy $\gamma(a,b,c)$ can be well-approximated by a two-threshold policy by showing the sense in which a two-threshold policy is optimal (Section IV-A). By two-threshold policy we mean $\gamma(a,b,c)$ with $c = a$. This two-threshold policy offers uniqueness of operating point and simplicity of design (Section IV-D).

A. Optimality of the two-threshold structure

If the asymptotic expressions (37)-(39) are taken to be the actual system performance, we can solve the optimization problem in (1) by finding the thresholds a, b and c which minimize ADD for given constraints on PFA and ANO. We will now prove that the solution to this constrained optimization problem can be found within the class of two-threshold policies obtained by setting $c = a$.

If $c = a$ then the performance of $\gamma(a,b,c)$ can be obtained by simply substituting $c = a$ in various asymptotic expressions. We identify this as a separate policy $\gamma(a', b')$, where we use a' and b' to name the two thresholds. Then the performance of this two-threshold policy would be:

$$\begin{aligned}
\mathrm{ADD}' &\approx \mathrm{E}_1[\nu(b',a')] && (41)\\
\mathrm{PFA}' &\approx e^{-a'}\int_0^\infty e^{-x}dR(x). && (42)\\
\mathrm{ANO}' &\approx \mathrm{ANO}_0' + \mathrm{ANO}_1' \\
&\approx \frac{1}{\rho}\frac{1}{1+e^{b'}}\frac{\mathrm{E}_\infty[\hat{\lambda}(b')]}{\mathrm{E}_\infty[\hat{\lambda}(b')] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}},b')]} + \ \mathrm{E}_1[\nu(b',a')] && (43)
\end{aligned}$$

Theorem 6. *The optimal solution to the problem in (1), in the class of deterministic three-threshold policies, with* ADD, PFA *and* ANO *given by (37)-(39) is a two-threshold policy.*

Proof: We claim that given any a, b and c for the three-threshold policy $\gamma(a,b,c)$, we can select some a' and b' for the two-threshold policy $\gamma(a',b')$, and get at least as good of an operating point. The key to this argument is the independence of PFA from b and the fact that ANO_1 and $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$, being equal, can be controlled simultaneously.

First we select a' such that the false alarm probabilities are same for $\gamma(a,b,c)$ and $\gamma(a',b')$:

$$e^{-a}R(a-c) + e^{-c}\int_{a-c}^\infty e^{-x}dR(x) = e^{-a'}\int_0^\infty e^{-x}dR(x).$$

It is easy to see that $a' \geq c$. Now select b' such that the following Shiryaev delays are equal:

$$\mathrm{E}_1[\nu(b,c)] = \mathrm{E}_1[\nu(b',a')].$$

Since $a' \geq c$, we have $b' \geq b$. Since the Shiryaev delays are exactly the respective ANO_1s for the two algorithm, we see that: the two systems have the same PFA, and have the same post change ANO, i.e., $\mathrm{ANO}_1 = \mathrm{ANO}_1'$. Also,

$$\mathrm{ANO}_0 \geq \mathrm{ANO}_0' \text{ since } b \leq b'.$$

Using this a' and b' we also get a smaller delay because

$$\mathrm{E}_1[\nu(b',a')] \leq \mathrm{E}_1[\nu(b,c)] + \int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|}dR(x). \tag{44}$$

Thus we have found a' and b' which gives the same PFA performance but at least as good ADD and ANO. Moreover, note the optimal a' and b' can be obtained directly based on the constraints on PFA and ANO using (42) and (43), and no further optimization is required. ■

Thus, based on Theorem 6 and the accuracy of the asymptotic expressions demonstrated in Section III-E, we see that as long as the asymptotic expressions reflect the true system behavior, a three-threshold policy can always be well approximated by a two-threshold policy.

Remark 3. *This means that, if the performance of the optimal deterministic policy is given by* (37)-(39), *then it can be well approximated by a two-threshold policy.*

Remark 4. *Within the class of two-threshold policies, there is a unique policy that meets a given set of constraints* α *and* β *with equality. Thus, no further optimization is needed.*

B. Comparative Performance

The dynamic programming solution suggested that the optimal algorithm has three thresholds. In Section IV-A we showed that if (37)-(39) reflect the true system performance then the threshold c is not required. However, there are cases in which having a third threshold helps. In those cases, it is interesting to know how much one loses in performance by using the two-threshold algorithm. It is not easy to analytically quantify the loss. We therefore study this via simulations.

In Table IX, for various system parameters, we compare the performance of the two-threshold algorithm with the best that can be achieved using three thresholds. We use the simulation set up: $\rho = 0.01$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(\theta,1)$, $\theta > 0$, and $\alpha = 10^{-3}$. For various values of θ and the ANO constraint β, we perform extensive simulations to search for the best three-threshold performance. In Table IX we refer to the best point by (a^*, b^*, c^*). We then compare this best three-threshold point with the performance of the two-threshold algorithm. Although we have chosen $\rho = 0.01$ and $\alpha = 10^{-3}$, this is typical of how the two algorithms compare. The table clearly shows that for ANO savings of up to 90% (ANO% up to 10), there is almost no loss in performance by using a two-threshold policy over the three-threshold policy. For ANO% of 5-10, there is less than 1% loss in performance. However for all the three values of θ considered, it is evident that if we seek 99% of ANO savings, then by using the third threshold we may get a better delay.

Remark 5. *In Section III-E, in Table IV, we showed that for* $\mathrm{ANO\%}$ *of* $50 - 100$, $\mathrm{E}_1[\nu(b,c)]$ *is a good approximation for* $\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]$. *In Theorem 6 we showed that when this happens, a three-threshold policy can be well approximated by a two-threshold policy. Table IX shows that even for* $\mathrm{ANO\%}$ *of* $10 - 50$, *where Theorem 6 may not be applicable, a two-threshold policy is optimal. Also, for* $\mathrm{ANO\%} < 10$, *where a three-threshold policy is optimal, their is a marginal loss in performance by using a two-threshold policy.*

TABLE IX

$\rho = 0.01$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(\theta,1)$, $\alpha = 10^{-3}$

			Three-threshold				Two-Threshold		
θ	β	ANO%	a^*	b^*	c^*	ADD	a	b	ADD
0.2	127	40%	6.787	2.585	6.787	345.83	6.787	2.585	345.83
0.2	50	16%	6.89	4.75	6.65	489.5	6.787	4.89	490.2
0.2	15	5%	6.901	5.6	6.326	557.88	6.787	6.06	558.35
0.2	3.5	1.1%	6.901	5.2	5.38	580.4	6.782	6.6	580.86
0.75	40	30%	6.467	-1.15	6.467	42.1	6.467	-1.15	42.1
0.75	15	11%	6.467	2.18	6.467	206.4	6.467	2.18	206.4
0.75	5	4%	6.88	4.25	5.7	450.55	6.47	5.0	455.3
0.75	1.7	1.3%	6.898	4.8	4.93	540.5	6.345	6.2	549.5
2.0	40	40%	5.768	-3.68	5.768	5.58	5.768	-3.68	5.58
2.0	10	10%	5.768	-1.39	5.768	16.2	5.768	-1.39	16.2
2.0	5	5%	5.768	0.05	5.768	47.44	5.768	0.05	47.44
2.0	2	2%	6.48	2.9	5.2	257.14	5.74	3.4	273.6

C. *Role of the third threshold*

In the last section we saw that two-threshold policies are approximately optimal, unless the ANO constraint is very severe. Based on our analytical study of $\gamma(a,b,c)$, we now provide a possible justification for this behavior, for low values of PFA. To meet the low ANO constraint using two thresholds b and a, we might need to choose a large b. Here is a reason why we may not want a large b. For $\gamma(a,b,c)$, choosing (a^*,b^*,c^*) is equivalent to choosing b^* first, then using c to meet the ANO constraint, and then using a to meet the PFA constraint. Choosing a large b may not be optimal, because for large b, it is possible that $\mathrm{E}[\Gamma] \ll t(-\infty, b)$, and the algorithm may wait for a long time before taking the first sample, even after the change has already occurred. We may get a better trade-off by choosing a smaller b and use $c < a$ to meet the constraint on ANO.

The third threshold is required in one more scenario. For the delay analysis, we used the fact that the passage of $\gamma(a,b,c)$ to c is through multiple cycles below b. However there are cases, for example for $\rho = 0.2$, $f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.1,1)$, for which as soon as Z_k crosses b from below, it grows up to c without ever coming back to b. In such a case it may be possible that we may not be able to meet the ANO constraint exactly by using only two thresholds. However, we can meet a constraint smaller than the one required with a small loss in performance.

D. Design of two-threshold policy and improved ADD *approximation*

In this section we comment on how to choose the thresholds a and b, equivalently A and B, for the two-threshold algorithm. In the previous sections we have found expressions for the PFA and ANO performance of the algorithm.

$$\begin{aligned}\text{PFA} &\approx e^{-a}\int_0^\infty e^{-x} dR(x) \leq e^{-a},\\ \text{ANO} &\approx \frac{1}{\rho}\frac{1}{1+e^b}\frac{\mathrm{E}_\infty[\hat{\lambda}]}{\mathrm{E}_\infty[\hat{\lambda}]+\mathrm{E}_\infty[t(Z_{\hat{\lambda}},b)]} + \mathrm{E}_1[\nu_b],\end{aligned}$$

where,

$$\begin{aligned}\mathrm{E}_\infty[\hat{\lambda}] &\approx \frac{\bar{r}+\log(1+\rho e^{-b})}{D(f_1,f_0)-|\log(1-\rho)|},\\ \mathrm{E}_\infty[t(Z_{\hat{\lambda}},b)] &\approx \int_0^\infty \frac{\log\left(\frac{1+e^b}{1+e^{b-x}}\right)}{|\log(1-\rho)|} dR(x),\\ \mathrm{E}_1[\nu_b] &\approx \frac{a-\mathrm{E}[\eta(b)]+\bar{r}}{D(f_1,f_0)+|\log(1-\rho)|}.\end{aligned}$$

Note that $\int_0^\infty e^{-x} dR(x)$ and $\bar{r}$ can be computed numerically, at least for Gaussian observations [14]. Also, $\mathrm{E}[\eta(b)]$ and $\mathrm{E}_\infty[t(Z_{\hat{\lambda}},b)]$ can be computed using Monte Carlo simulations. Since, PFA is not a function of b, given α and β, we can set,

$$a = \log\frac{\int_0^\infty e^{-x} dR(x)}{\alpha},$$

and use this value of a and given constraint β to select the value of b using the above expressions. As mentioned earlier, this choice of a and b would give approximately the minimum possible ADD.

For data in Table X, we start with the constraints α and β and use the analytical expressions above to choose a and b that meet these constraints. We then simulate the algorithm using these thresholds to check if the performance meets the desired constraints we started with. We also compare the ADD values obtained in simulations and analysis. We see that the analytical expressions provide us with the means to design the two-threshold algorithm.

For computing ANO, if one wants to avoid Monte Carlo simulations in the computation of $\mathrm{E}_\infty[t(Z_{\hat{\lambda}},b)]$ or $\mathrm{E}_1[\nu_b]$, then the following approximations also works well:

$$\begin{aligned}\mathrm{E}_\infty[t(Z_{\hat{\lambda}},b)] &\approx \frac{\log\left(\frac{1+e^b}{1+e^{b-\bar{r}}}\right)}{|\log(1-\rho)|}.\\ \mathrm{E}_1[\nu_b] &\approx \frac{a-b+\bar{r}}{D(f_1,f_0)+|\log(1-\rho)|}.\end{aligned}$$

TABLE X

$f_0 \sim \mathcal{N}(0,1)$, $f_1 \sim \mathcal{N}(0.75,1)$. THRESHOLDS a AND b OBTAINED USING PFA AND ANO EXPRESSIONS.

	Constraints		Two Thresholds		Simulations				Analysis
ρ	α	β	a	b	ANO	ANO%	PFA	ADD	$\mathrm{E}_1[\nu_b]$
0.05	3×10^{-6}	45	12.27	-0.62	43.5	77%	3.002×10^{-6}	41.1	38.7
0.05	5×10^{-4}	30	7.15	-0.62	28.1	71%	5.017×10^{-4}	25.7	23.3
0.01	6.5×10^{-5}	70	9.2	-2.1	69.5	51%	6.523×10^{-5}	42.25	38.6
0.01	1×10^{-3}	60	6.46	-2.06	59.1	47%	1.01×10^{-3}	33	29.13

Although, $\mathrm{E}_1[\nu_b]$ is a good approximation for ANO_1 for almost all values of a and b, unfortunately, as was mentioned earlier, it is not necessarily a very good approximation for ADD. Recall that $\mathrm{E}_1[\nu_b]$ is a good approximation for ADD only when the gap between a and b is large, which corresponds to large ANO%. For moderate gap between a and b, or for smaller ANO%, the quality of approximation depends on other systems parameters. In Section III-E, Table IV, and in this section in Table X, we saw some of these cases where the approximation was good (note the ANO% in Table X). Although, the two-threshold algorithm can be designed by selecting a and b as mentioned above, a better ADD approximation can be obtained as follows.

The technique for this new approximation comes from the proof of Lemma 3. Analogous to the steps in the proof of Lemma 3 we identify three events:

$$\begin{aligned}\mathcal{A} &= \{Z_\Gamma < b\},\\ \mathcal{B} &= \{Z_\Gamma \geq b; Z_k \nearrow b\},\\ \mathcal{C} &= \{Z_\Gamma \geq b; Z_k \nearrow a\},\end{aligned}$$

where, we have replaced threshold c by threshold a. We can write the following expression for $\mathrm{E}[\tau-\Gamma|\tau \geq \Gamma]$,

$$\mathrm{E}[\tau-\Gamma|\tau\geq\Gamma] = \mathrm{E}[\tau-\Gamma;\mathcal{A}|\tau\geq\Gamma] + \mathrm{E}[\tau-\Gamma;\mathcal{B}|\tau\geq\Gamma] + \mathrm{E}[\tau-\Gamma;\mathcal{C}|\tau\geq\Gamma].$$

We then assume that the event $\mathcal{B}\cup\mathcal{C}$ is dominated by $\mathcal{C}$. That is, we assume that if $Z_\Gamma > b$, then Z_k climbs to a. Define, $P_b = \mathrm{P}(Z_\Gamma \geq b)$. Then,

$$P_b = \mathrm{P}(\mathcal{B}\cup\mathcal{C}) \approx \mathrm{P}(\mathcal{C}).$$

Thus,

$$\mathrm{ADD} \approx P_b\ \mathrm{E}[\lambda(Z_\Gamma)|\mathcal{C}] + (1-P_b)(\mathrm{E}[t(Z_\Gamma,b)|\mathcal{A}] + \mathrm{ADD}^s). \tag{45}$$

From proof of Lemma 4, it is easy to show the following:

$$\mathrm{ADD}^s = \mathrm{E}_1[\lambda|\{Z_\lambda > a\}] + (\mathrm{E}_1[\lambda|\{Z_\lambda < b\}] + \mathrm{E}_1[t(Z_\lambda, b)|\{Z_\lambda < b\}])\,\frac{\mathrm{P}_1(Z_\lambda < b)}{1 - \mathrm{P}_1(Z_\lambda < b)}.$$

We now use the following approximations:

$$\begin{aligned}
\mathrm{E}_1[\lambda|\{Z_\lambda > a\}] &\approx \mathrm{E}[\lambda(Z_\Gamma)|\mathcal{C}] \approx \mathrm{E}_1[\nu_b],\\
\mathrm{E}_1[\lambda|\{Z_\lambda < b\}] &\approx \frac{\bar{r} + \log(1+\rho e^{-b})}{D(f_1, f_0) - |\log(1-\rho)|},\\
\mathrm{E}_1[t(Z_\lambda, b)|\{Z_\lambda < b\}] &\approx t(b - \bar{r}, b) \approx \frac{\log(1+e^b) - \log(1+e^{b-\bar{r}})}{|\log(1-\rho)|}.
\end{aligned}$$

Remark 6. *Note that with* $\mathrm{E}[\lambda(Z_\Gamma)|\mathcal{C}] \approx \mathrm{E}_1[\nu_b]$, *and* $\mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}]$ *being independent of threshold* a, *this new approximation* (45) *reduces to the existing one, i.e. to* $\mathrm{E}_1[\nu_b]$ *for a fixed* b *as* $a \to \infty$*: in this limit* ADD^s *approaches* $\mathrm{E}_1[\nu_b]$ *(Lemma 4).*

To compute (45), we also need approximations for $\mathrm{P}_1(Z_\lambda < b)$, P_b and $\mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}]$. Those are provided below. Setting $a = \infty$ we have, by Wald's likelihood identity, Proposition 2.24, Pg 13, [14],

$$\mathrm{P}_1(Z_\lambda < b) = \mathrm{E}_\infty\left[\frac{f_1(X_1)\dots f_1(X_\lambda)}{f_0(X_1)\dots f_0(X_\lambda)}\right].$$

Under P_∞, λ a.s. ends in b and with high probability it takes very small values. Hence, this expressions can be computed using Monte Carlo simulations. Further,

$$\begin{aligned}
P_b &= \mathrm{P}(\Gamma > t(-\infty, b))\mathrm{P}(Z_\Gamma > b|\Gamma > t(-\infty, b))\\
&\approx \frac{1}{1+e^b}\frac{\mathrm{E}_\infty[\hat{\lambda}]}{\mathrm{E}_\infty[\hat{\lambda}] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]}.
\end{aligned}$$

We already have the approximations for $\mathrm{E}_\infty[\hat{\lambda}]$ and $\mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]$. The approximation for $\mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}]$ can be obtained as follows:

$$\begin{aligned}
(1-P_b)\mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}] &= (1-P_b)\mathrm{E}[t(Z_\Gamma, b)|\{Z_\Gamma < b\}]\\
&= \mathrm{E}[t(Z_\Gamma, b)|\{Z_\Gamma < b\} \cap \{\Gamma > t(-\infty, b)\}]\mathrm{P}(\{\Gamma > t(-\infty, b)\} \cap \{Z_\Gamma < b\})\\
&\quad + \mathrm{E}[t(Z_\Gamma, b)|\{Z_\Gamma < b\} \cap \{\Gamma \le t(-\infty, b)\}]\mathrm{P}(\{\Gamma \le t(-\infty, b)\} \cap \{Z_\Gamma < b\}).
\end{aligned}$$

This can be computed using

$$\mathrm{P}(\{\Gamma > t(-\infty, b)\} \cap \{Z_\Gamma < b\}) \approx \frac{1}{1+e^b}\frac{\mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]}{\mathrm{E}_\infty[\hat{\lambda}] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)]},$$

and

$$\mathrm{P}(\{\Gamma \le t(-\infty, b)\} \cap \{Z_\Gamma < b\}) = \mathrm{P}(\{\Gamma \le t(-\infty, b)\}) \approx \frac{e^b}{1+e^b}.$$

To compute conditional expectation of $t(Z_\Gamma, b)$, we need to subtract from $t(x, b)$, the mean of Γ conditioned on $\{\Gamma \leq t(x, b)\}$. Specifically,

$$\mathrm{E}[t(Z_\Gamma, b)|\{Z_\Gamma < b\} \cap \{\Gamma > t(-\infty, b)\}] = t(b - \bar{r}, b) - \frac{1}{\mathrm{P}(\Gamma \leq t(b - \bar{r}, b))} \sum_{k=1}^{t(b-\bar{r},b)} k(1-\rho)^{k-1}\rho,$$

and,

$$\mathrm{E}[t(Z_\Gamma, b)|\{Z_\Gamma < b\} \cap \{\Gamma \leq t(-\infty, b)\}] = t(-\infty, b) - \frac{1}{\mathrm{P}(\Gamma \leq t(-\infty, b))} \sum_{k=1}^{t(-\infty,b)} k(1-\rho)^{k-1}\rho.$$

Thus we have obtained approximations for all the terms for the new approximation for ADD in (45).

We now provide, in Table XI, numerical results to show the accuracy of the new ADD approximation (45), by comparing it with simulations and also with $\mathrm{E}_1[\nu_b]$. All the points here correspond to a low value of ANO%: $\mathrm{ANO} = 10\%$ of the Shiryaev ANO. We also set PFA around 1×10^{-3}. The table clearly demonstrates that the new ADD approximation predicts ADD with less than 5% error.

TABLE XI

$f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.75, 1)$, PFA $\approx 10^{-3}$, ANO=10% OF SHIRYAEV ANO

			ADD		
ρ	a	b	Simulations	Analysis New (45)	Analysis $\mathrm{E}_1[\nu_b]$
0.01	6.4	2.7	250	260	14.42
0.005	6.45	0.6	181	190	22.09
0.001	6.47	-2.7	75	80	33.68
0.0005	6.47	-3.49	74	79	36.49
0.0001	6.47	-5.2	76	80	42.56

E. Trade-off curves

In Fig. 2 and 3 we plot the ANO-ADD trade-off for the two-threshold algorithm. Specifically, we compare the two-threshold algorithm with the classical Shiryaev test and study how much ANO can be reduced without significantly loosing in terms of ADD. Fig. 2 shows that we can reduce ANO by up to 25% as compared to the Shiryaev test, while getting approximately the same ADD performance. Moreover, if we allow for a 10% increase in ADD, then we can reduce ANO by up to 50%. If the change is rarer ($\rho = 0.001$), then Fig. 3 shows that we can reduce ANO by 70% by allowing for 10% excess ADD.

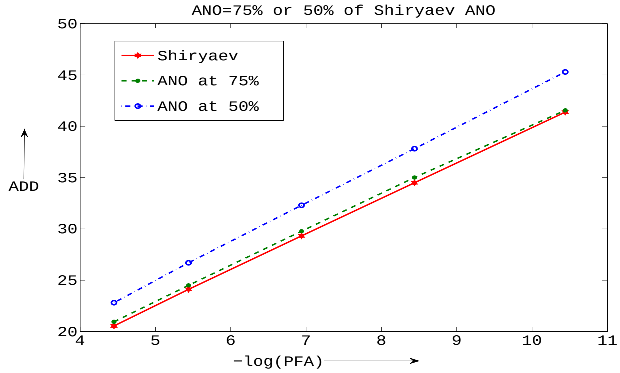

Fig. 2. Trade-off curves comparing performance of two-threshold algorithm with the Shiryaev test for ANO=75% and 50% of Shiryaev ANO. $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.75, 1)$, and $\rho = 0.01$.

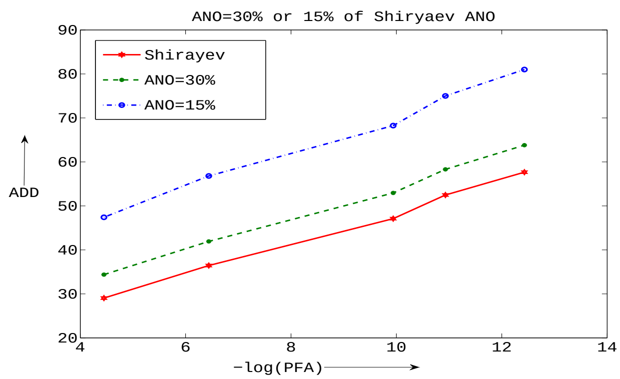

Fig. 3. Trade-off curves comparing performance of two-threshold algorithm with the Shiryaev test for ANO=30% and 15% of Shiryaev ANO. $f_0 \sim \mathcal{N}(0, 1)$, $f_1 \sim \mathcal{N}(0.75, 1)$, and $\rho = 0.001$.

V. Conclusions

We posed a data-efficient version of the classical Bayesian quickest change detection problem, where we control the number of observations taken before the change is declared. We obtained a two-threshold Bayesian test that has some optimality properties and is easy to design. Specifically, we identified an asymptotic regime – when the false alarm probability is small, the change is rare and the ANO constraint is not very severe – in which either a deterministic two-threshold policy is optimal, or the optimal policy can be obtained by randomizing over two-threshold policies. We supported our claim via analytical and simulation results. We derived analytical approximations for the ADD, PFA and ANO performance of the two-threshold policy using which we can design the test by choosing the thresholds. Further, the two-threshold policy which meets a given set of constraints with equality is unique among the class of two-threshold policies. This result has implications in many engineering applications where an abrupt change has to be detected in a process under observation, but there is a cost associated with acquiring the data needed to make accurate decisions.

In the absence of knowledge of the prior on Γ, an important problem for future research is to see if two-threshold policies are optimal in non-Bayesian (e.g., minimax) settings. More importantly, it is of interest to understand how to update the test metric in a non-Bayesian setting when we skip an observation. From an application point of view, one can design a two-threshold test based on the Shiryaev-Roberts or CUSUM approaches [19], and use the undershoot of the metric when it goes below the threshold 'b', to design the off times. Furthermore, if we are able to find useful lower bounds on delay for given false alarm and ANO constraints, we may be able to use these to prove asymptotic optimality of such heuristic tests, as is done for the standard quickest change detection problem [17], [20]. Also, such lower bounds can possibly help in obtaining insights for cases where the observations are not i.i.d. [17], [20]. Other interesting problems in this area include the design of data-efficient optimal algorithms for robust change detection or nonparametric change detection.

Appendix to Section III-A

Proof of Theorem 1: We first show that η_n with $b=-\infty$, and Z_0 a random variable, is a slowly changing sequence. Let Z_0 takes value z_0, then

$$\eta_n = \log\left[e^{z_0} + \sum_{k=0}^{n-1}\rho(1-\rho)^k \prod_{i=1}^{k}\frac{f_0(X_i)}{f_1(X_i)}\right] \xrightarrow[n\to\infty]{\mathrm{P}_1-a.s.} \log\left[e^{z_0} + \sum_{k=0}^{\infty}\rho(1-\rho)^k \prod_{i=1}^{k}\frac{f_0(X_i)}{f_1(X_i)}\right].$$

Define

$$\eta(Z_0) \stackrel{\triangle}{=} \log\left[e^{Z_0} + \sum_{k=0}^{\infty}\rho(1-\rho)^k \prod_{i=1}^{k}\frac{f_0(X_i)}{f_1(X_i)}\right].$$

Note that $\eta(Z_0)$ as a function of Z_0 is well defined and finite under P_1. This is because by Jensen's inequality, for $Z_0 = z_0$,

$$\begin{aligned} \mathrm{E}[\eta(z_0)] &\leq \log\left[e^{z_0} + \sum_{k=0}^{\infty}\rho(1-\rho)^k \mathrm{E}_1\left(\prod_{i=1}^{k}\frac{f_0(X_i)}{f_1(X_i)}\right)\right] \\ &= \log\left[e^{z_0} + \sum_{k=0}^{\infty}\rho(1-\rho)^k\right] = \log\left(e^{z_0}+1\right). \end{aligned}$$

Thus

$$\eta_n \xrightarrow[b=-\infty]{\mathrm{P}_1-a.s.} \eta(Z_0) = \log\left(e^{Z_0}+\rho\right) + \sum_{k=1}^{\infty}\log\left(1+e^{-Z_k}\rho\right). \tag{46}$$

This implies $\sum_{k=1}^{\infty} \log\left(1 + e^{-Z_k}\rho\right)$ converges a.s. for i.i.d. $\{X_k\}$ and $b = -\infty$. This series will also converge with probability 1 if we condition on a set with positive probability.

Let change happen at $\Gamma = l$. We set $Z_0 = Z_\Gamma = Z_l$ and assume that $\{X_k\}$, $k \geq 1$ have density f_1, which would happen after Γ. We first show that starting with the above Z_0, the sequence η_n generated in (10) is slowly changing.

To verify the first condition (11), from (10) note that,

$$n^{-1} \max\{|\eta_1|, \ldots, |\eta_n|\} \leq n^{-1} \left[|\log\left(e^{Z_0} + \rho\right)| + \sum_{k=1}^{n-1} \log\left(1 + e^{-Z_k}\rho\right) + \sum_{k=1}^{n} \left(|\log L(X_k)|\right) \mathbb{I}_{\{Z_k < b\}} \right].$$

Since, $Z_k \to \infty$ a.s., $\log\left(1 + e^{-Z_k}\rho\right) \to 0$, also, $\mathbb{I}_{\{Z_k < b\}} \to 0$ a.s. Thus both the sequences $\{\log\left(1 + e^{-Z_k}\rho\right)\}$ and $\{\left(|\log L(X_k)|\right) \mathbb{I}_{\{Z_k < b\}}\}$ are Cesaro summable and have Cesaro sum of zero. Thus the term inside the square bracket above, when divided by n, goes to zero a.s. and hence also in probability. Thus the first condition is verified.

To verify the second condition (12), we first obtain a bound on $|\eta_{n+k} - \eta_n|$.

$$|\eta_{n+k} - \eta_n| \leq \sum_{i=n}^{n+k-1} \log\left(1 + e^{-Z_i}\rho\right) + \sum_{i=n+1}^{n+k} \left(|\log L(X_i)|\right) \mathbb{I}_{\{Z_k < b\}}.$$

Thus,

$$\max_{1 \leq k \leq n\delta} |\eta_{n+k} - \eta_n| \leq \sum_{i=n}^{n+n\delta-1} \log\left(1 + e^{-Z_i}\rho\right) + \sum_{i=n+1}^{n+n\delta} \left(|\log L(X_i)|\right) \mathbb{I}_{\{Z_k < b\}} \stackrel{\triangle}{=} d_n^1 + d_n^2.$$

Here, for convenience of computation, we use d_n^1 and d_n^2 to represent the first and second partial sums respectively. Now,

$$\mathrm{P}\{\max_{1 \leq k \leq n\delta} |\eta_{n+k} - \eta_n| > \epsilon\} \leq \mathrm{P}(d_n^1 + d_n^2 > \epsilon),$$

and we bound the probability $\mathrm{P}(d_n^1 + d_n^2 > \epsilon)$ as follows.

On the event that $E \stackrel{\triangle}{=} \{Z_k \geq b, \forall k \geq 0\}$, d_n^2 is identically zero, thus for n large enough,

$$\mathrm{P}(d_n^1 + d_n^2 > \epsilon | E) = \mathrm{P}(d_n^1 > \epsilon | E) < \epsilon.$$

This is because d_n^1 behaves like a partial sum of a series of type in (46). Since the series in (46) converges if random variables are generated i.i.d. f_1, it will also converge if conditioned on the event E. Thus, the partial sum d_n^1 converges to 0 almost surely, and hence converges to 0 in probability, i.e., $\mathrm{P}(d_n^1 > \epsilon | E) \to 0$. Select, $n = n_1^*$ such that $\forall n > n_1^*$, $\mathrm{P}(d_n^1 > \epsilon | E) < \epsilon$.

Define

$$L_Z = \sup\{k \geq 1 : Z_{k-1} < b, Z_k \geq b\},$$

with $L_Z = \infty$ if no such k exists. On the event E', which is the compliment of E, L_Z is a.s. finite. Then, by noting that $d_n^2 = 0$ for $L_Z < n$, we get for n large enough,

$$\begin{aligned}
\mathrm{P}(d_n^1 + d_n^2 > \epsilon | E') \stackrel{\triangle}{=} \mathrm{P}_{E'}(d_n^1 + d_n^2 > \epsilon) &\leq \mathrm{P}_{E'}(d_n^1 + d_n^2 > \epsilon; L_Z \geq n) + \mathrm{P}_{E'}(d_n^1 + d_n^2 > \epsilon; L_Z < n) \\
&\leq \mathrm{P}_{E'}(L_Z \geq n) + \mathrm{P}_{E'}(d_n^1 + d_n^2 > \epsilon; L_Z < n) \\
&= \mathrm{P}_{E'}(L_Z \geq n) + \mathrm{P}_{E'}(d_n^1 > \epsilon; L_Z < n) \\
&\leq \mathrm{P}_{E'}(L_Z \geq n) + \mathrm{P}_{E'}(d_n^1 > \epsilon | L_Z < n) \\
&< \epsilon/2 + \epsilon/2 = \epsilon.
\end{aligned}$$

Since, L_Z is almost surely finite, $\mathrm{P}_{E'}(L_Z \geq n) \to 0$ as $n \to \infty$. Thus we can select $n = n_2^*$ such that $\forall n > n_2^*$, $\mathrm{P}_{E'}(L_Z \geq n) < \epsilon/2$. For the second term, note that conditioned on $L_Z < n$, d_n^1 behaves like a partial sum of a series of type in (46), with Z_0 replaced by Z_{L_Z}. Since the series in (46) converges if random variables are generated i.i.d. f_1 beyond L_Z, it will also converge if conditioned on the event $\{L_Z < n\}$. Thus, the partial sum d_n^1 converges to 0 almost surely, and hence converges to 0 in probability, i.e., $\mathrm{P}_{E'}(d_n^1 > \epsilon | L_Z < n) \to 0$. Select, $n = n_3^*$ such that $\forall n > n_3^*$, $\mathrm{P}(d_n^1 > \epsilon | L_Z < n) < \epsilon/2$. Then $n^* = \max\{n_1^*, n_2^*, n_3^*\}$, is the desired n^* and pick any $\delta > 0$. Then for $n > n^*$,

$$\begin{aligned}
\mathrm{P}(d_n^1 + d_n^2 > \epsilon) &= \mathrm{P}(d_n^1 + d_n^2 > \epsilon | E)\mathrm{P}(E) + \mathrm{P}(d_n^1 + d_n^2 > \epsilon | E')\mathrm{P}(E') \\
&< \epsilon \mathrm{P}(E) + \epsilon \mathrm{P}(E') < \epsilon.
\end{aligned}$$

Since the sequence η_n is slowly changing, according to [14], the asymptotic distribution of the overshoot when Z_k crosses a large boundary under f_1 is $R(x)$. Thus we have the following result,

$$\lim_{c \to \infty} \mathrm{P}_\ell \left[Z_{\tau_c} - c \leq x | \tau_c \geq l \right] = R(x),$$

where P_ℓ is the probability measure with change happening at l. Now,

$$\mathrm{P}\left[Z_{\tau_c} - c \leq x | \tau_c \geq \Gamma \right] = \sum_{l=1}^{\infty} \mathrm{P}_l \left[Z_{\tau_c} - c \leq x | \tau_c \geq l \right] \mathrm{P}(\Gamma = l | \tau_c \geq \Gamma),$$

and

$$\lim_{c \to \infty} \mathrm{P}_l \left[Z_{\tau_c} - c \leq x | \tau_c \geq l \right] \mathrm{P}(\Gamma = l | \tau_c \geq \Gamma) = R(x) \mathrm{P}(\Gamma = l) \leq 1.$$

Hence we have the desired result by dominated convergence theorem. ■

APPENDIX TO SECTION III-B

Proof of Lemma 1: Clearly,

$$\mathrm{E}[\tau - \Gamma|\tau \geq \Gamma] = \mathrm{E}[\tau - \tau_c|\tau \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma|\tau \geq \Gamma].$$

Using $\mathrm{E}[Y; D]$ to represent integration of the random variable Y over the set D, we write $\mathrm{E}[\tau_c - \Gamma|\tau \geq \Gamma]$ as follows,

$$\begin{aligned}
\mathrm{E}[\tau_c - \Gamma|\tau \geq \Gamma] &= \mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma, \tau \geq \Gamma] \;\; \mathrm{P}(\tau_c \geq \Gamma|\tau \geq \Gamma) + \mathrm{E}[\tau_c - \Gamma; \tau_c < \Gamma|\tau \geq \Gamma] \\
&\overset{(i)}{=} \mathrm{E}[(\tau_c - \Gamma)|\tau_c \geq \Gamma](1 + o(1)) + \frac{\mathrm{E}[\tau_c - \Gamma; \tau_c < \Gamma \leq \tau]}{\mathrm{P}(\tau \geq \Gamma)} \qquad \text{as} \quad c, a \to \infty. \\
&\overset{(ii)}{=} \mathrm{E}[(\tau_c - \Gamma)|\tau_c \geq \Gamma](1 + o(1)) + o(1) \qquad \text{as} \quad c, a \to \infty.
\end{aligned}$$

Here, (i) follows because $\{\tau_c \geq \Gamma\} \subset \{\tau \geq \Gamma\}$, and $\mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma) \to 0$ as $c, a \to \infty$. To show the latter, we obtain an upper bound on $\mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma)$. Using an argument identical to the one given in (14) we get

$$\begin{aligned}
e^{-c} \geq \mathrm{P}(\tau_c < \Gamma) &= \mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma)(1 - \mathrm{PFA}) + \mathrm{P}(\tau_c < \Gamma|\tau < \Gamma)\mathrm{PFA} \\
&= \mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma)(1 - \mathrm{PFA}) + \mathrm{PFA}.
\end{aligned}$$

This implies,

$$\mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma) \leq \frac{e^{-c} - \mathrm{PFA}}{1 - \mathrm{PFA}} \leq \frac{e^{-c}}{1 - \mathrm{PFA}} \leq \frac{e^{-c}}{1 - e^{-a}} \to 0 \text{ as } c, a \to \infty.$$

For (ii) note that, over $\{\tau_c < \Gamma \leq \tau\}$, $-\Gamma \leq \tau_c - \Gamma \leq 0$, and hence integrable. Thus, $\mathrm{E}[\tau_c - \Gamma; \tau_c < \Gamma \leq \tau] \to 0$ as $c \to \infty$ because

$$\mathrm{P}(\tau_c < \Gamma \leq \tau) \leq \mathrm{P}(\tau_c < \Gamma) \leq e^{-c} \to 0 \text{ as } c \to \infty.$$

Now we want to show $\mathrm{E}[\tau - \tau_c|\tau \geq \Gamma] = \mathrm{E}[\tau - \tau_c|\tau_c \geq \Gamma](1 + o(1))$. Conditioning on $\{\tau_c \geq \Gamma\}$ and its compliment we get,

$$\begin{aligned}
\mathrm{E}[(\tau - \tau_c)|\tau \geq \Gamma] &= \mathrm{E}[(\tau - \tau_c)|\tau_c \geq \Gamma, \tau \geq \Gamma] \;\; \mathrm{P}(\tau_c \geq \Gamma|\tau \geq \Gamma) \\
&\quad + \mathrm{E}[(\tau - \tau_c)|\tau_c < \Gamma, \tau \geq \Gamma] \;\; \mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma) \\
&= \mathrm{E}[(\tau - \tau_c)|\tau_c \geq \Gamma](1 + o(1)) \\
&\quad + \mathrm{E}[(\tau - \tau_c)|\tau_c < \Gamma, \tau \geq \Gamma]\mathrm{P}(\tau_c < \Gamma|\tau \geq \Gamma) \qquad \text{as} \quad c, a \to \infty \\
&= \mathrm{E}[(\tau - \tau_c)|\tau_c \geq \Gamma](1 + o(1)) + o(1) \qquad \text{as} \quad c, a \to \infty.
\end{aligned}$$

We get the above equalities because $\mathrm{P}(\tau_c < \Gamma | \tau \geq \Gamma) \to 0$ as $c, a \to \infty$. Also, from the hypothesis in the lemma,

$$\mathrm{E}[(\tau - \tau_c) | \tau_c < \Gamma, \tau \geq \Gamma] \leq t(c,a) < \infty \text{ as } c, a \to \infty,$$

and thus it follows that

$$\mathrm{E}[(\tau - \tau_c) | \tau_c < \Gamma, \tau \geq \Gamma] \mathrm{P}(\tau_c < \Gamma | \tau \geq \Gamma) \to 0 \qquad \text{as } c, a \to \infty.$$

■

Proof of Lemma 3: We use $\{Z_k \nearrow b\}$ to indicate that Z_k approaches b from below for some $k > \Gamma$, i.e. $\exists k > \Gamma, s.t., Z_{k-1} < b, Z_k \geq b$. and use $\{Z_k \nearrow c\}$ to represent the event that Z_k crossed c without ever coming back to b, i.e., $Z_k \geq b, \forall k > \Gamma$. Define,

$$\lambda(x) = \inf\{k \geq 1 : Z_k \notin [b, c), Z_0 = x, b \leq x < c\}. \tag{47}$$

Also let $\Lambda(x)$ be defined with $Z_0 = x$ similar to (25). Thus, λ and $\lambda(b)$ have the same distribution. Similarly, Λ and $\Lambda(b)$ are identically distributed. The behavior of the delay path depends on Z_Γ, the value of Z_k at the change point Γ, and how Z_k evolves after that point. We thus define the following three disjoint events:

$$\begin{aligned} \mathcal{A} &= \{Z_\Gamma < b\} \\ \mathcal{B} &= \{Z_\Gamma \geq b; Z_k \nearrow b\} \\ \mathcal{C} &= \{Z_\Gamma \geq b; Z_k \nearrow c\}. \end{aligned}$$

We can write,

$$\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma] \;=\; \mathrm{E}[\tau_c - \Gamma; \mathcal{A} | \tau_c \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma; \mathcal{B} | \tau_c \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma; \mathcal{C} | \tau_c \geq \Gamma]. \tag{48}$$

Now consider each of the three terms on the right hand side of the above equation.

Under the event $\mathcal{A}$, the process Z_k starts below b and reaches c after multiple up-crossings of the threshold b. Then,

$$\mathrm{E}[\tau_c - \Gamma; \mathcal{A} | \tau_c \geq \Gamma] \leq \mathrm{E}[t(Z_\Gamma, b) | \mathcal{A}] + \mathrm{E}_1 \left[\sum_{k=1}^{N} \Lambda_k \right] \mathrm{P}(\mathcal{A} | \tau_c \geq \Gamma). \tag{49}$$

This upper bound was obtained as follows. Let t_1 be the first time Z_k crosses b from below. Then the time to reach c, t_1 onwards, is upper bounded by the time to reach c if we reset Z_k to b; this is because $c - Z_{t_1} \leq c - b$. Arguing this way each time Z_k crosses b from below, we have the desired upper bound over $\mathcal{A}$.

Under the event $\mathcal{B}$, the process Z_k starts above b and crosses b before c. It then has multiple up-crossings of b, similar to the case of event $\mathcal{A}$. Arguing in a similar manner, we get

$$\mathrm{E}[\tau_c - \Gamma; \mathcal{B}|\tau_c \geq \Gamma] \leq \mathrm{E}[\Lambda(Z_\Gamma)|\mathcal{B}, Z_{\Lambda(Z_\Gamma)} < b] + \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] \mathrm{P}(\mathcal{B}|\tau_c \geq \Gamma)$$

Similarly,

$$\begin{aligned} \mathrm{E}[\tau_c - \Gamma; \mathcal{C}|\tau_c \geq \Gamma] &= \mathrm{E}[\Lambda(Z_\Gamma)|\mathcal{C}, Z_{\Lambda(Z_\Gamma)} > c]\mathrm{P}(\mathcal{C}|\tau_c \geq \Gamma) \\ &\leq \mathrm{E}_1[\Lambda(b)|Z_{\Lambda(b)} > c]\mathrm{P}(\mathcal{C}|\tau_c \geq \Gamma) \\ &\leq \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] \mathrm{P}(\mathcal{C}|\tau_c \geq \Gamma). \end{aligned}$$

Substituting we get,

$$\begin{aligned} \mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma] &= \mathrm{E}[\tau_c - \Gamma; \mathcal{A}|\tau_c \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma; \mathcal{B}|\tau_c \geq \Gamma] + \mathrm{E}[\tau_c - \Gamma; \mathcal{C}|\tau_c \geq \Gamma]. \\ &\leq \mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}] + \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] \mathrm{P}(\mathcal{A}|\tau_c \geq \Gamma) \\ &\quad + \mathrm{E}[\Lambda(Z_\Gamma)|\mathcal{B}, Z_{\Lambda(Z_\Gamma)} < b] + \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] \mathrm{P}(\mathcal{B}|\tau_c \geq \Gamma) \\ &\quad + \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] \mathrm{P}(\mathcal{C}|\tau_c \geq \Gamma) \\ &= \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right] + \mathrm{E}[t(Z_\Gamma, b)|\mathcal{A}] + \mathrm{E}[\Lambda(Z_\Gamma)|\mathcal{B}, Z_{\Lambda(Z_\Gamma)} < b]. \end{aligned} \tag{50}$$

In equation (50), we observe that except for $\mathrm{ADD}^s = \mathrm{E}_1\left[\sum_{k=1}^{N} \Lambda_k\right]$, other terms are not a function of threshold c. Thus we have

$$\mathrm{E}[\tau_c - \Gamma|\tau_c \geq \Gamma] \leq \mathrm{ADD}^s\,(1 + o(1)) \quad \text{ as } c, a \; \to \infty, c < a.$$

■

Proof of Lemma 4: Based on Ψ, we define two new recursions, one in which the evolution of Z_k is truncated at b,

$$\tilde{\Psi}(Z_k) = \begin{cases} \Psi(Z_k) & \text{if } \Psi(Z_k) \geq b \\ b & \text{if } \Psi(Z_k) < b, \end{cases}$$

and, another in which the overshoot is ignored each time the Shiryaev recursion crosses b from below,

$$\hat{\Psi}(Z_k) = \begin{cases} b & \text{if } Z_k < b \text{ and } \Psi(Z_k) \geq b \\ \Psi(Z_k) & \text{otherwise .} \end{cases}$$

Based on these two recursions we define two new stopping times:

$$\begin{aligned}\tilde{\nu}_b &= \inf\{k \geq 1 : \tilde{\Psi}(Z_{k-1}) > c, Z_0 = b\},\\ \hat{\nu}_b &= \inf\{k \geq 1 : \hat{\Psi}(Z_{k-1}) > c, Z_0 = b\}.\end{aligned}$$

These two stopping times stochastically upper and lower bound the Shiryaev stopping time ν_b defined in (19), i.e.,

$$\mathrm{E}_1[\tilde{\nu}_b] \leq \mathrm{E}_1[\nu_b] \leq \mathrm{E}_1[\hat{\nu}_b]. \tag{51}$$

Recall from (18) that

$$\nu(x, y) = \inf\{k \geq 1 : \Psi(Z_{k-1}) > y, Z_0 = x\}.$$

Using Wald's lemma [14], we can get the following expressions:

$$\mathrm{E}_1[\tilde{\nu}_b] = \frac{\mathrm{E}_1[\lambda]}{\mathrm{P}_1(Z_\lambda > c)}, \qquad \mathrm{E}_1[\hat{\nu}_b] = \frac{\mathrm{E}_1[\lambda] + \mathrm{E}_1[\nu(Z_\lambda, b); \{Z_\lambda < b\}]}{\mathrm{P}_1(Z_\lambda > c)}. \tag{52}$$

Multiplying and dividing ADD^s by $\mathrm{E}_1[\lambda]$ we get

$$\begin{aligned}\mathrm{ADD}^s &= \frac{\mathrm{E}_1[\lambda] + \mathrm{E}_1[t(Z_\lambda, b); \{Z_\lambda < b\}]}{\mathrm{E}_1[\lambda]} \frac{\mathrm{E}_1[\lambda]}{\mathrm{P}_1(Z_\lambda > c)}\\ &= \mathrm{E}_1[\tilde{\nu}_b] \frac{\mathrm{E}_1[\lambda] + \mathrm{E}_1[t(Z_\lambda, b); \{Z_\lambda < b\}]}{\mathrm{E}_1[\lambda]}\\ &= \mathrm{E}_1[\tilde{\nu}_b](1 + o(1)) \qquad \text{as } c, a \to \infty.\end{aligned}$$

The last equality follows because $\mathrm{E}_1[\lambda] \to \infty$ as $c \to \infty$, while $\mathrm{E}_1[t(Z_\lambda, b); \{Z_\lambda < b\}]$ is not a function of c. Similarly, multiplying and dividing ADD^s by $\mathrm{E}_1[\lambda] + \mathrm{E}_1[\nu(Z_\lambda, b); \{Z_\lambda < b\}]$ we get

$$\mathrm{ADD}^s = \mathrm{E}_1[\hat{\nu}_b]\,(1 + o(1)) \qquad \text{as} \quad c \to \infty.$$

Using these two expressions for ADD^s and the relationship that $\mathrm{E}_1[\tilde{\nu}_b] \leq \mathrm{E}_1[\nu_b] \leq \mathrm{E}_1[\hat{\nu}_b]$, we have,

$$\mathrm{ADD}^s = \mathrm{E}_1[\nu_b](1 + o(1)) \text{ as } c \to \infty.$$

■

Proof of Lemma 5: First note that by definition (16), $Z_{t(x,y)} > y \geq Z_{t(x,y)-1}$. Also, from (8)

$$\begin{aligned}Z_{t(x,y)} &= Z_{t(x,y)-1} + \log \frac{1}{1-\rho} + \log(1 + e^{-Z_{t(x,y)-1}}\rho)\\ &\leq y + \log \frac{1}{1-\rho} + \log(1 + e^{-y}\rho).\end{aligned}$$

Thus

$$y < Z_{t(x,y)} \leq y + \log \frac{1}{1-\rho} + \log(1 + e^{-y}\rho),$$

equivalently

$$e^y < e^{Z_{t(x,y)}} \leq e^y \frac{1}{1-\rho}(1+e^{-y}\rho).$$

Further, the recursion (8) can be written in terms of e^{Z_k} for $k \geq 0$:

$$e^{Z_{k+1}} = \frac{\rho + e^{Z_k}}{1-\rho}.$$

Using this we can write an expression for $e^{Z_{t(x,y)}}$:

$$e^{Z_{t(x,y)}} = \frac{e^x}{(1-\rho)^t} + \sum_{k=1}^{t(x,y)} \frac{\rho}{(1-\rho)^k} = \frac{e^x+1}{(1-\rho)^{t(x,y)}} - (1-\rho).$$

Using the bounds for $Z_{t(x,y)}$ obtained above, we get

$$e^y < \frac{e^x+1}{(1-\rho)^{t(x,y)}} - (1-\rho) \leq e^y \frac{1}{1-\rho}(1+e^{-y}\rho).$$

This gives us bounds for $t(x,y)$:

$$\frac{\log(1+e^y-\rho) - \log(1+e^x)}{|\log(1-\rho)|} \leq t(x,y) \leq \frac{\log\left(1+e^y\frac{(1+e^{-y}\rho)}{(1-\rho)} - \rho\right) - \log(1+e^x)}{|\log(1-\rho)|}. \tag{53}$$

By keeping ρ fixed and taking $x, y \to \infty$ we get (28), and by keeping x, y fixed and taking $\rho \to 0$ we get (29). ■

Proof of Theorem 3: Let

$$R_c(x) = \mathrm{P}(Z_{\tau_c} - c \leq x | \tau_c \geq \Gamma).$$

Then

$$\mathrm{E}[\tau - \tau_c | \tau_c \geq \Gamma] = \int_0^{a-c} t(c+x, a) dR_c(x).$$

By using (28) from Lemma 5, and noting that $\lim_{c\to\infty} R_c(x) = R(x)$, we get (30). In (30), as $c, a \to \infty$, if $a - c \to \infty$, or $a - c \to$ constant > 0 and $\rho \to 0$, then $\int_0^{a-c} \frac{a-c-x}{|\log(1-\rho)|} dR(x)$ dominates $R(a-c)$ in the limit and we get (31). ■

Appendix to Section III-C

Proof of Lemma 6: Since, $p_\tau > A$ imply $Z_\tau > a$, we have,

$$\frac{1}{1+e^{-Z_\tau}} \geq \frac{1}{1+e^{-a}}.$$

The required result is obtained by obtaining upper and lower bounds on PFA as follows.

$$\mathrm{PFA} = \mathrm{E}[1-p_\tau] = \mathrm{E}\left[\frac{1}{1+e^{Z_\tau}}\right] \leq \mathrm{E}\left[e^{-Z_\tau}\right].$$

Also,

$$\begin{aligned}\mathrm{PFA} = \mathrm{E}[1-p_\tau] = \mathrm{E}\left[\frac{1}{1+e^{Z_\tau}}\right] &= \mathrm{E}\left[\frac{1}{e^{Z_\tau}}\frac{1}{1+e^{-Z_\tau}}\right] \\ &\geq \mathrm{E}\left[\frac{1}{e^{Z_\tau}}\frac{1}{1+e^{-a}}\right] = \mathrm{E}\left[e^{-Z_\tau}\right](1+o(1)) \text{ as } a\to\infty.\end{aligned}$$

Thus,

$$\mathrm{PFA} = \mathrm{E}[e^{-Z_\tau}](1+o(1)) = e^{-a}\mathrm{E}[e^{-(Z_\tau-a)}](1+o(1)) \text{ as } c,a \ \to\infty, c<a.$$

■

Proof of Theorem 4: First note that,

$$\mathrm{E}[e^{-(Z_\tau-a)}] = \mathrm{E}[e^{-(Z_\tau-a)}|\tau_c\geq\Gamma](1-\mathrm{P}(\tau_c<\Gamma)) + \mathrm{E}[e^{-(Z_\tau-a)}|\tau_c<\Gamma]\mathrm{P}(\tau_c<\Gamma).$$

Since, $\mathrm{P}(\tau_c<\Gamma) = \mathrm{E}[1-p_{\tau_c}] \leq 1-C \leq e^{-c}$, we can write,

$$\mathrm{PFA} = e^{-a}\mathrm{E}[e^{-(Z_\tau-a)}|\tau_c\geq\Gamma](1+o(1)) \qquad \text{as } c,a \ \to\infty, c<a.$$

Further, we evaluate $\mathrm{E}[e^{-(Z_\tau-a)}|\tau_c\geq\Gamma]$ as follows.

$$\begin{aligned}\mathrm{E}[e^{-(Z_\tau-a)}|\tau_c\geq\Gamma] &= \mathrm{E}[e^{-(Z_\tau-a)}|Z_{\tau_c}\leq a,\tau_c\geq\Gamma]\mathrm{P}(Z_{\tau_c}\leq a|\tau_c\geq\Gamma) \\ &\quad +\mathrm{E}[e^{-(Z_\tau-a)};Z_{\tau_c}>a|\tau_c\geq\Gamma]. \\ &= \mathrm{E}[e^{-(Z_\tau-a)}|Z_{\tau_c}\leq a,\tau_c\geq\Gamma]\mathrm{P}(Z_{\tau_c}-c\leq a-c|\tau_c\geq\Gamma) \\ &\quad +e^{a-c}\mathrm{E}[e^{-(Z_\tau-c)};Z_{\tau_c}>a|\tau_c\geq\Gamma] \\ e^{-a}\mathrm{E}[e^{-(Z_\tau-a)}|\tau_c\geq\Gamma] &= e^{-a}\mathrm{E}[e^{-(Z_\tau-a)}|Z_{\tau_c}\leq a,\tau_c\geq\Gamma]\mathrm{P}(Z_{\tau_c}-c\leq a-c|\tau_c\geq\Gamma) \\ &\quad +e^{-c}\mathrm{E}[e^{-(Z_\tau-c)};Z_{\tau_c}>a|\tau_c\geq\Gamma].\end{aligned}$$

From Theorem 1 it follows that:

$$\begin{aligned}\lim_{c\to\infty}\frac{\mathrm{P}(Z_{\tau_c}-c\leq a-c|\tau_c\geq\Gamma)}{R(a-c)} &= 1, \qquad \text{and} \\ \lim_{c\to\infty}\frac{\mathrm{E}[e^{-(Z_\tau-c)};Z_{\tau_c}>a|\tau_c\geq\Gamma]}{\int_{a-c}^{\infty}e^{-x}dR(x)} &= 1.\end{aligned}$$

Further we can show that,

$$\frac{e^a}{1+e^a}(1-\rho) \leq \mathrm{E}[e^{-(Z_\tau-a)}|Z_{\tau_c}\leq a,\tau_c\geq\Gamma] \leq 1,$$

and goes to 1 as $c,a\to\infty$, and $\rho\to 0$. This proves the theorem. ■

APPENDIX TO SECTION III-D

Proof of Lemma 7: Using Theorem 2 we write ANO_1 as

$$\begin{aligned} \mathrm{ANO}_1 &= \mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right] \left(1 - \frac{T_b - 1}{\mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right]}\right) \\ &= \mathrm{E}_1[\nu_b] \left(1 - \frac{T_b - 1}{\mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right]}\right)(1 + o(1)) \quad \text{as} \quad c, a \to \infty. \end{aligned}$$

We now obtain an upper bound on $\frac{T_b - 1}{\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]}$ which goes to zero as $c, a \to \infty$.

Recall from Lemma 3 that $\mathcal{A}$ and $\mathcal{B}$ are the events under which excursions below b are possible. The passage to c is through multiple cycles below b, and the time spend below b in each cycle can be bounded by $t(-\infty, b)$. Define $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$ as one plus the number of cycles below b, under events $\mathcal{A}$ and $\mathcal{B}$ respectively. Then,

$$T_b - 1 \leq T_b \leq \mathrm{P}_1(\mathcal{A}) t(-\infty, b) \mathrm{E}[N_{\mathcal{A}}] + \mathrm{P}_1(\mathcal{B}) t(-\infty, b) \mathrm{E}[N_{\mathcal{B}}].$$

The averages $\mathrm{E}[N_{\mathcal{A}}]$ and $\mathrm{E}[N_{\mathcal{B}}]$ can be written as a series of probabilities, where each term correspond to the event that Z_k goes below b, and not above c, each time it crosses b from below. Each of these probabilities can be maximized by setting Z_k to b, each time it crosses b from below. Hence, $\mathrm{E}[N_{\mathcal{A}}] \leq \mathrm{E}[N]$ and $\mathrm{E}[N_{\mathcal{B}}] \leq \mathrm{E}[N]$. This gives a bound on $T_b - 1$.

$$T_b - 1 \leq t(-\infty, b) \mathrm{E}[N].$$

By using (51) we get as $c, a \to \infty$,

$$\frac{T_b - 1}{\mathrm{E}\left[\tau_c - \Gamma | \tau_c \geq \Gamma\right]} \leq \frac{t(-\infty, b)\mathrm{E}[N]}{\mathrm{E}_1[\nu_b]}(1 + o(1)) \leq \frac{t(-\infty, b)\mathrm{E}[N]}{\mathrm{E}_1[\tilde{\nu}_b]}(1 + o(1)).$$

From (52) we know that $\mathrm{E}_1[\tilde{\nu}_b] = \mathrm{E}_1[\lambda]\mathrm{E}[N]$. Thus the upper bound on $\frac{T_b - 1}{\mathrm{E}[\tau_c - \Gamma | \tau_c \geq \Gamma]}$ goes to 0 as $c, a \to \infty$. This proves the lemma. ■

Proof of Lemma 9: Since $\mathrm{P}\{\tau_c \geq \Gamma\} \to 1$ as $c \to \infty$,

$$\begin{aligned} \mathrm{P}(\Gamma > t(b) | \tau_c \geq \Gamma) &= \mathrm{P}(\Gamma > t(b)) + o(1) \quad \text{as} \quad c, a \to \infty \\ &= \frac{1}{1 + z_0}(1 - \rho)^{t(b)} + o(1) \quad \text{as} \quad c, a \to \infty. \end{aligned}$$

From (29) in Lemma 5, with $y = b$ and $x = z_0$, we have

$$t(z_0, b) = \left(\frac{\log(1 + e^b) - \log(1 + e^{z_0})}{|\log(1 - \rho)|}\right)(1 + o(1)) \text{ as } \rho \to 0.$$

From this, it is easy to show that

$$(1 - \rho)^{t(b)} \to \left(\frac{1 + e^{z_0}}{1 + e^b}\right) \quad \text{as } \rho \to 0.$$

By substituting this in the expression for $\mathrm{P}(\Gamma > t(b)|\tau_c \geq \Gamma)$ we get the desired result. ■

Proof of Lemma 8: Each time Z_k crosses b from below, is satisfies:

$$b \;<\; Z_k \;\leq\; b + \log\frac{1}{1-\rho} + \log(1 + e^{-b}\rho).$$

Define, $b_1 \stackrel{\triangle}{=} b + \log\frac{1}{1-\rho} + \log(1 + e^{-b}\rho)$. Then $b_1 \to b$ as $\rho \to 0$. Also, each time Z_k crosses b from below, the average number of observations used before Γ can be increased by setting $Z_k = b_1$ and decreased by setting $Z_k = b$. This is because of the geometric nature of change. Let $Z_k = x$ when it crosses b from below, and suppose we reset Z_k to b_1. Then, the number of observations used before change, on an average, would be the number of observations used before Z_k reaches x from b_1, plus the number of observations used there onwards as if the process started at x. Similar reasoning can be given to explain why the average number of observations used decreases, if we reset Z_k to b, each time it crosses b from below.

Define the following stopping time:

$$\tilde{\Lambda}^x = \inf\{k \geq 1 : Z_{k-1} < b \text{ and } Z_k \geq b \text{ or } k \geq \Gamma, Z_0 = x \geq b, c = \infty\}.$$

Thus, $\tilde{\Lambda}^x$ is the time for Z_k, to start at $Z_0 = x$ with $c = \infty$, and stop the first time, either Z_k approaches b from below, or when change happens. Also, let $\delta^x \in (0,1)$ be such that $\tilde{\Lambda}^x\delta^x$ is the number of observations used before Z_k was stopped by $\tilde{\Lambda}^x$, i.e., fraction of $\tilde{\Lambda}^x$ when $Z_k \geq b$. If $\{\tilde{\Lambda}^b_k\}$ and $\{\tilde{\Lambda}^{b_1}_k\}$ be sequences with distribution of $\tilde{\Lambda}^b$ and $\tilde{\Lambda}^{b_1}$ respectively and if L^x is the number of times Z_k crosses b from below and is set to x at each such instant, then,

$$\begin{aligned}
\mathrm{E}_\infty[L^b] \;\; \mathrm{E}_\infty[\tilde{\Lambda}^b\delta^b] = \mathrm{E}_\infty\left[\sum_{k=1}^{L^b}\tilde{\Lambda}^b_k\delta^b_k\right] \;&\leq\; \mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), c = \infty\right] \\
&\leq\; \mathrm{E}_\infty\left[\sum_{k=1}^{L^{b_1}}\tilde{\Lambda}^{b_1}_k\delta^{b_1}_k\right] = \mathrm{E}_\infty[L^{b_1}] \;\; \mathrm{E}_\infty[\tilde{\Lambda}^{b_1}\delta^{b_1}].
\end{aligned}$$

Here the equalities follows from Wald's lemma [14].

In the above, L^x is $\mathrm{Geom}(\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^x])$, and hence $\mathrm{E}_\infty[L^{b_1}] = \frac{1}{\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^{b_1}]}$. Also note that

$$\frac{\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^{b_1}]}{\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^{b}]} \to 1 \text{ as } \rho \to 0.$$

Further, for $x = b_1$ or $x = b$, define $\hat{\lambda}(x)$ based on (33) as

$$\hat{\lambda}(x) = \inf\{k \geq 1 : Z_k < b, Z_0 = x \geq b, c = \infty\}.$$

It is clear that $\hat{\lambda}(b) = \hat{\lambda}$. Thus we have, for both $x = b_1$ and $x = b$,

$$\begin{aligned} \mathrm{E}_\infty[\tilde{\Lambda}^x \delta^x] &= \mathrm{E}_\infty[\tilde{\Lambda}^x \delta^x | \Gamma \leq \tilde{\Lambda}^x \delta^x] \mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^x \delta^x] + \mathrm{E}_\infty[\tilde{\Lambda}^x \delta^x | \Gamma > \tilde{\Lambda}^x \delta^x] \mathrm{P}_0[\Gamma > \tilde{\Lambda}^x \delta^x] \\ &\to \mathrm{E}_\infty[\hat{\lambda}(x)] \text{ as } \rho \to 0. \end{aligned}$$

Here, the result follows because as $\rho \to 0$, $\tilde{\Lambda}^x \delta^x$ converges a.s. to a finite limit and $\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^x \delta^x] \to 0$. Also for the same reason, $\mathrm{P}_0[\Gamma > \tilde{\Lambda}^x \delta^x] \to 1$ as $\rho \to 0$. Moreover, since $b_1 \to b$ as $\rho \to 0$, we have as $\rho \to 0$

$$\mathrm{E}_\infty[\hat{\lambda}(b_1)] \to \mathrm{E}_\infty[\hat{\lambda}(b)] = \mathrm{E}_\infty[\hat{\lambda}].$$

Thus,

$$\mathrm{E}\left[\sum_{k=t(b)}^{\Gamma-1} S_k \middle| \Gamma > t(b), c = \infty \right] = \frac{\mathrm{E}_\infty[\hat{\lambda}]}{\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^b]} (1 + o(1)) \quad \text{as } \rho \to 0.$$

Using Binomial expansion we can obtain an approximation for $\mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^b]$:

$$\begin{aligned} \mathrm{P}_0[\Gamma \leq \tilde{\Lambda}^b] &= \mathrm{P}_0[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)] = 1 - \mathrm{P}_0[\Gamma > \hat{\lambda} + t(Z_{\hat{\lambda}}, b)] \\ &= 1 - \mathrm{E}_\infty[(1-\rho)^{\hat{\lambda} + t(Z_{\hat{\lambda}}, b)}] \\ &\overset{(i)}{=} \rho \left(\mathrm{E}_\infty[\hat{\lambda}] + \mathrm{E}_\infty[t(Z_{\hat{\lambda}}, b)] \right) (1 + o(1)) \quad \text{as } \rho \to 0. \end{aligned} \tag{54}$$

To see why (i) is true we note that,

$$t(Z_{\hat{\lambda}}, b) \leq t(-\infty, b) \approx \frac{\log\left(1 + e^b\right)}{|\log\left(1 - \rho\right)|}.$$

Using L'Hopital's rule it is easy to show that as $\rho \to 0$, followed by $b \to -\infty$,

$$\rho^n \left(\frac{\log\left(1 + e^b\right)}{\log(1 - \rho)} \right)^n \to 0.$$

Using this in the Binomial expansion of $\mathrm{E}_\infty[(1-\rho)^{\hat{\lambda} + t(Z_{\hat{\lambda}}, b)}]$ we get equality (i) in (54). ■

References

[1] A. N. Shiryaev, “On optimal methods in quickest detection problems,” Theory Probab. Appl., Vol. 8, pp. 22-46, 1963.

[2] Z. G. Stoumbos, M. R. Reynolds Jr., “Economic statistical design of adaptive control schemes for monitoring the mean and variance: An application to analyzers,” Nonlinear Analysis: Real World Applications, Volume 6, Issue 5, Pages 817-844, December 2005.

[3] Z. Yin and V. Makis, “Economic and economic-statistical design of a multivariate Bayesian control chart for condition-based maintenance,” IMA J Management Math 22: pp. 47-63, 2011.

[4] G. Tagaras, “A survey of recent developments in the design of adaptive control charts,” Journal of Quality Technology. Vol. 30, no. 3, pp. 212-231. July 1998.

[5] V. Makis, “Multivariate Bayesian Control Chart,” Oper. Res. 56, 2, pp. 487-496, March 2008.

[6] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson, “Wireless sensor networks for habitat monitoring,” In Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications (WSNA ’02), ACM, New York, NY, USA, 88-97, 2002.

[7] J. A. Rice, K. Mechitov, S. Sim, T. Nagayama, S. Jang, R. Kim, B. F. Spencer, G. Agha and Y. Fujino, “Flexible Smart Sensor Framework for Autonomous Structural Health Monitoring,” Smart Structures and Systems, vol. 6, no. 5, pp. 423-438, 2010.

[8] V.V. Veeravalli, “Decentralized quickest change detection,” IEEE Transactions on Information Theory, vol. 47, pp. 1657-1665, 2001.

[9] Y. Mei, “Information bounds and quickest change detection in decentralized decision systems,” IEEE Transactions on Information Theory, vol. 51, pp. 2669–2681, 2005.

[10] A.G. Tartakovsky and V.V. Veeravalli, “Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems,” Sequential Analysis, 27(4): 441-475, October 2008.

[11] L. Zacharias and R. Sundaresan,“Decentralized sequential change detection using physical layer fusion,” Proc. ISIT, France, Jun. 2007.

[12] T. Banerjee, V. Sharma, V. Kavitha and A. Jayaprakasam,“Generalized Analysis of a Distributed Energy Efficient Algorithm for Change Detection,” Wireless Communication, IEEE Transactions on, vol. 10, pp. 91–101, 2011.

[13] K. Premkumar and A. Kumar, “Optimal Sleep/Wake Scheduling for Quickest Intrusion Detection Using Sensor Networks,” IEEE INFOCOM, Arizona, USA, 2008.

[14] D Siegmund, “Sequential analysis: tests and confidence intervals”, Springer Series in Statistics, 1985.

[15] M. Woodroofe, “Nonlinear Renewal Theory in Sequential Analysis,” Philadelphia: SIAM, 1982.

[16] K. W. Ross, “Randomized and Past-Dependent Policies for Markov Decision Processes with Multiple Constraints,” Operations Research Vol. 37, No. 3, pp. 474-477, May - Jun., 1989.

[17] A.G. Tartakovsky and V.V. Veeravalli, “General Asymptotic Bayesian Theory of Quickest Change Detection,” SIAM Theory of Probability and its Applications, Vol 49, No. 3, pp. 458-497, 2005.

[18] D. P. Bertsekas, ”Dynamic Programming and Optimal Control,” Vol. I and II (3rd ed.). Athena Scientific. 2007.

[19] A.G. Tartakovsky and G.V. Moustakides, “State-of-the-Art in Bayesian Changepoint Detection,” Sequential Analysis, vol. 29, no. 2, pp. 125-145, 2010. Special issue ”Celebration of the 75th Anniversary of Albert Shiryaev”, Guest Editor A.G. Tartakovsky.

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