

This article was downloaded by: [University of Illinois at Urbana-Champaign]

On: 24 February 2012, At: 16:42

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Sequential Analysis: Design Methods and Applications

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/lsqa20>

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

Taposh Banerjee<sup>a</sup> & Venugopal V. Veeravalli<sup>a</sup>

<sup>a</sup> ECE Department and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Available online: 24 Feb 2012

To cite this article: Taposh Banerjee & Venugopal V. Veeravalli (2012): Data-Efficient Quickest Change Detection with On-Off Observation Control, Sequential Analysis: Design Methods and Applications, 31:1, 40-77

To link to this article: <http://dx.doi.org/10.1080/07474946.2012.651981>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## 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, Illinois, USA

**Abstract:** In this article we extend Shiryaev’s quickest change detection formulation by also accounting for the cost of observations used before the change point. The observation cost is captured through the average number of observations used in the detection process before the change occurs. The objective is to select an on–off observation control policy that decides whether or not to take a given observation, along with the stopping time at which the change is declared, to minimize the average detection delay, subject to constraints on both the probability of false alarm and the observation cost. By considering a Lagrangian relaxation of the constraint problem and using dynamic programming arguments, we obtain an *a posteriori* probability-based two-threshold algorithm that is a generalized version of the classical Shiryaev algorithm. We provide an asymptotic analysis of the two-threshold algorithm and show that the algorithm is asymptotically optimal—that is, the performance of the two-threshold algorithm approaches that of the Shiryaev algorithm—for a fixed observation cost, as the probability of false alarm goes to zero. We also show, using simulations, that the two-threshold algorithm has good observation cost-delay trade-off curves and provides significant reduction in observation cost compared to the naïve approach of fractional sampling, where samples are skipped randomly. Our analysis reveals that, for practical choices of constraints, the two thresholds can be set independent of each other: one based on the constraint of false alarm and another based on the observation cost constraint alone.

**Keywords:** Asymptotic optimality; Nonlinear renewal theory; Observation control; Quickest change detection.

**Subject Classifications:** 62L05; 62L10; 62L15; 62F05; 62F15; 60G40.

Received May 24, 2011, Revised November 16, 2011, and December 12, 2011, Accepted December 17, 2011

Recommended by A. G. Tartakovsky

Address correspondence to Venugopal V. Veeravalli, ECE Department, University of Illinois at Urbana-Champaign, 106 Coordinated Science Laboratory, 1308 West Main Street, Urbana, IL 61801, USA; E-mail: vvv@illinois.edu

## 1. INTRODUCTION

In the Bayesian quickest change detection problem proposed by Shiryaev (1963), there is a sequence of random variables,  $\{X_n\}$ , whose distribution changes at a random time  $\Gamma$ . It is assumed that before  $\Gamma$ ,  $\{X_n\}$  are independent and identically distributed (i.i.d.) with density  $f_0$ , and after  $\Gamma$  they are i.i.d. with density  $f_1$ . The distribution of  $\Gamma$  is assumed to be known and modeled as a geometric random variable with parameter  $\rho$ . The objective is to find a stopping time  $\tau$ , 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 article we extend Shiryaev's formulation by explicitly accounting for the cost of the observations used in the detection process. We capture the observation penalty (cost) through the average number of observations used before the change point  $\Gamma$  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  $\tau$ , so that the average detection delay is minimized subject to constraints on the probability of false alarm and the observation cost. The motivation for this model comes from the consideration of the following engineering applications.

In many monitoring applications—for example, infrastructure monitoring, environmental 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 Mainwaring et al. (2002), 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 (Rice et al., 2010). This change could be a fault in the structures in infrastructure monitoring (Rice et al., 2010), the arrival of the species to the habitat (Mainwaring et al., 2002), etc.

In industrial quality control, statistical control charts are designed that can detect a sustained deviation of the industrial process from normal behavior (Stoumbos et al., 2000). Often there is a cost associated with acquiring the statistics for the control charts and it is of interest to consider designing *economic-statistical* control chart schemes (Assaf, 1988, Assaf et al., 1993, Makis, 2008, Stoumbos and Reynolds, 2005, Stoumbos et al., 2000, Tagaras, 1998, Yakir, 1996). One approach to economic-statistical control chart design has been to use algorithms from the change detection literature, such as Shewhart, exponential weighted moving average (EWMA), and cumulative sum (CUSUM), as control charts, and optimize over the choice of sample size, sampling interval, and control limits (Tagaras, 1998, Stoumbos and Reynolds, 2005). Another approach has been to find optimal sampling rates in the problem of detection of a change in the drift of a sequence of Brownian motions with global false alarm constraint (Assaf et al., 1993, Yakir, 1996). Thus, these approaches are essentially non-Bayesian. In the Bayesian setting,

the problem of dynamic sampling for detecting a change in the drift of a standard Brownian motion is considered for an exponentially distributed change point in Assaf (1988). For practical applications, it is of interest to consider the economic design of Bayesian control charts in discrete time. The design of a Bayesian economic-statistical control chart is considered for a shift in the mean vector of a multivariate Gaussian model in Makis (2008). But, the problem is modeled as an optimal stopping problem where the long-term average cost is minimized and, hence, there is no control on the number of observations used at each time step. The process control problem is fundamentally a quickest change detection problem, and it is therefore appropriate that economic-statistical schemes for process control are developed in this framework.

In most of the above-mentioned or similar applications, changes are rare and quick detection is often required. So, ideally we would like to take as few observations as possible before change to reduce the observation cost, and skip as few as possible after change to minimize delay, while maintaining an acceptable probability of false alarm.

There have been other formulations of the Bayesian quickest change detection problem that are relevant to sensor networks: see Banerjee et al. (2011), Mei (2005), Tartakovsky and Veeravalli (2008), Veeravalli (2001), and Zacharias and Sundaresan (2007). The change detection problem studied here was earlier considered in a similar set up for sensor networks in Premkumar and Kumar (2008). But due 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 article 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. We extend Shiryaev's formulation by also accounting for the cost of observations used before the change point, and obtain an *a posteriori* probability-based two-threshold algorithm that is asymptotically optimal. Specifically, we show that the probability of false alarm and the average detection delay of the two-threshold algorithm approaches that of the Shiryaev algorithm, for a fixed observation cost constraint, as the probability of false alarm goes to zero. Even for moderate values of the false alarm probability, we will show using simulations that the two-threshold algorithm provides good performance. We also provide an asymptotic analysis of the two-threshold algorithm; that is, we obtain expressions for the delay, probability of false alarm, and the average number of observations used before and after change, using which the thresholds can be set to meet the constraints on probability of false alarm and observation cost.

The layout of the article is as follows: In the following section, we set up the data-efficient quickest change detection problem with on-off observation control and introduce the two-threshold algorithm. In Section 3, we provide an asymptotic analysis of the two-threshold algorithm. In Section 4, we provide approximations using which the analytical expressions in Section 3 can be computed and validate the approximations by comparing them with the corresponding values obtained via simulations. In Section 5, we prove the asymptotic optimality of the two-threshold algorithm, provide its false alarm-delay-observation cost trade-off curves, and also

compare its performance with the naïve approach of fractional sampling, where observations are skipped randomly.

## 2. PROBLEM FORMULATION AND THE TWO-THRESHOLD ALGORITHM

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

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

where  $\mathbb{I}$  is the indicator function, and  $\pi_0$  represents the probability of the change having happened before the observations are taken. Typically  $\pi_0$  is set to 0.

In order to minimize the average number of observations used before  $\Gamma$ , at each time instant a decision is made on whether to use the observation in the next time step, based on all available information. Let  $S_k \in \{0, 1\}$ , with  $S_k = 1$  if it had been decided to take the observation at time  $k$ ; that is,  $X_k$  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  $k - 1$ ; that is,

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

with  $\mu$  denoting the control law and  $I$  defined as:

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

Here,  $X_i^{(S_i)}$  represents  $X_i$  if  $S_i = 1$ ; otherwise,  $X_i$  is absent from the information vector  $I_k$ . The choice of  $S_1$  is based on the prior  $\pi_0$ .

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  $\tau$ , we can define the average detection delay (ADD) as

$$\text{ADD} = \mathbf{E}[(\tau - \Gamma)^+].$$

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

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

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

$$\text{ANO} = \mathbf{E} \left[ \sum_{k=1}^{\min\{\tau, \Gamma-1\}} S_k \right].$$

This metric captures the cost of observations used before the change point. The cost of observations used after the change point is captured by the ADD. This way of capturing the observation cost is practically relevant, as in many applications, changes are rare and quick detection is often required.

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

$$\begin{aligned} & \underset{\gamma}{\text{minimize}} && \text{ADD}(\gamma), \\ & \text{subject to} && \text{PFA}(\gamma) \leq \alpha, \text{ and } \text{ANO}(\gamma) \leq \beta, \end{aligned} \quad (2.1)$$

where  $\alpha$  and  $\beta$  are given constraints. Toward solving (2.1), we consider a Lagrangian relaxation of this problem that can be approached using dynamic programming:

$$J^* = \min_{\gamma} \text{ADD}(\gamma) + \lambda_f \text{PFA}(\gamma) + \lambda_e \text{ANO}(\gamma), \quad (2.2)$$

where  $\lambda_f$  and  $\lambda_e$  are Lagrange multipliers,  $\lambda_f \geq 0$  and  $\lambda_e \geq 0$ . It is easy to see that if  $\lambda_f$  and  $\lambda_e$  can be found such that the solution to (2.2) achieves the PFA and ANO constraints with equality, then the solution to (2.2) is also the solution to (2.1).

The problem in (2.2) can be converted to an appropriate Markov control problem using steps similar to those followed in Premkumar and Kumar (2008).

Let  $\Theta_k$  denote the state of the system at time  $k$ . After the stopping time  $\tau$  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

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

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

Furthermore, let  $D_k$  denote the stopping decision variable at time  $k$ ; that is,  $D_k = 0$  if  $k < \tau$  and  $D_k = 1$  otherwise. Then the optimization problem in (2.2) can be written as a minimization of an additive cost over time:

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

with

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

Using standard arguments (Bertsekas, 2007) it can be seen that this optimization problem can be solved using infinite horizon dynamic programming with sufficient statistic (belief state) given by:

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

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

$$\Phi^{(0)}(p_k) = p_k + (1 - p_k)\rho \quad (2.3)$$

and

$$\Phi^{(1)}(X_{k+1}, p_k) = \frac{\Phi^{(0)}(p_k)L(X_{k+1})}{\Phi^{(0)}(p_k)L(X_{k+1}) + [1 - \Phi^{(0)}(p_k)]} \quad (2.4)$$

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 (2.2) can be obtained from the solution to the Bellman equation:

$$J(p_k) = \min_{d_k, s_{k+1}} \lambda_f(1 - p_k)\mathbb{I}_{\{d_k=1\}} + \mathbb{I}_{\{d_k=0\}} [p_k + A_J(p_k)], \quad (2.5)$$

where

$$A_J(p_k) = B_0(p_k)\mathbb{I}_{\{s_{k+1}=0\}} + [\lambda_e(1 - p_k) + B_1(p_k)]\mathbb{I}_{\{s_{k+1}=1\}},$$

with

$$B_0(p_k) = J[\Phi^{(0)}(p_k)]$$

and

$$B_1(p_k) = E\{J[\Phi^{(1)}(X_{k+1}, p_k)]\}.$$

It can be shown by an induction argument (see, e.g., Premkumar and Kumar, 2008) that  $J$ ,  $B_0$ , and  $B_1$  are all nonnegative concave functions on the interval  $[0, 1]$ , and that  $J(1) = B_0(1) = B_1(1) = 0$ . Also, by Jensen's inequality

$$B_1(p) \leq J\{E[\Phi^{(1)}(X, p)]\} = B_0(p), \quad p \in [0, 1].$$

Let

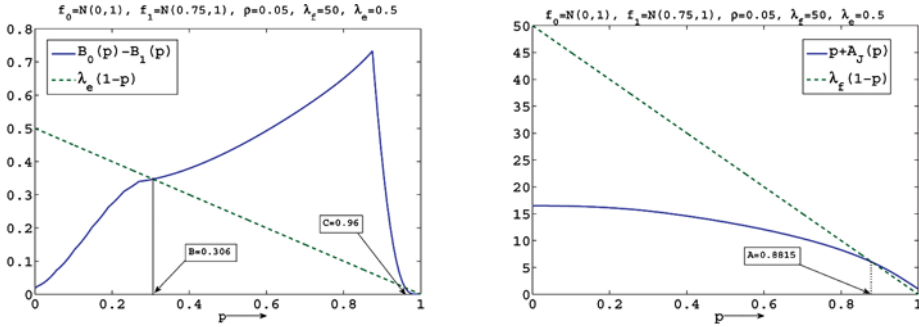
$$d(p_k) = B_0(p_k) - B_1(p_k).$$

Then, from the above properties of  $J$ ,  $B_0$ , and  $B_1$ , it is easy to show that the optimal policy  $\gamma^* = (\tau^*, \mu_0^*, \mu_1^*, \dots, \mu_{\tau-1}^*)$  for the problem given in (2.2) has the following structure:

$$S_{k+1}^* = \mu_k^*(p_k) = \begin{cases} 0 & \text{if } d(p_k) < \lambda_e(1 - p_k) \\ 1 & \text{if } d(p_k) \geq \lambda_e(1 - p_k) \end{cases} \quad (2.6)$$

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

**Remark 2.1.** Since  $d(p_k) \geq 0 \forall p_k$ , the algorithm in (2.6) reduces to the classical Shiryaev algorithm when  $\lambda_e = 0$  (Shiryaev, 1963).



(a)  $d(p) = B_0(p) - B_1(p)$  and  $\lambda_e(1 - p)$  as a function of  $p$  (b)  $p + A_J(p)$  and  $\lambda_f(1 - p)$  as a function of  $p$

**Figure 1.** Example where a two-threshold policy is optimal:  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$ ,  $\rho = 0.05$ ,  $\lambda_f = 50$ , and  $\lambda_e = 0.5$ . Value iteration: number of iterations = 1, 500, number of points = 2, 000 (color figure available online).

The optimal stopping rule  $\tau^*$  is similar to the one of the Shiryaev problem. But, the observation control is not explicit and one has to evaluate the differential cost function  $d(p_k)$  at  $p_k$  at each time step to choose  $S_{k+1}$ .

In Figure 1(a) we plot the differential cost function  $d(p) = B_0(p) - B_1(p)$  and  $\lambda_e(1 - p)$  as a function of  $p$ . We note that, although  $B_0(p)$  and  $B_1(p)$  are concave in  $p$ , their difference  $d(p)$  is not. Thus, the line  $\lambda_e(1 - p)$  can intersect  $d(p)$  at more than two points. However, in Figure 1(a) we see that there are exactly two points of intersection, one at  $B = 0.306$  and another at  $C = 0.96$ . In Figure 1(b) we plot the functions  $p + A_J(p)$  and  $\lambda_f(1 - p)$  as a function of  $p$ . This figure shows that the stopping threshold is  $A = 0.8815 < 0.96 = C$ . Thus, from Figures 1(a) and (b) we see that the optimal policy has two thresholds. For most of the system parameters we have tried, the cost functions behave in this way, and hence for these values the following two-threshold policy is optimal.

**Algorithm 2.1** (Two-Threshold Policy:  $\gamma(A, B)$ ). Start with  $p_0 = 0$  and use the following control, with  $B < A$ , for  $k \geq 0$ :

$$S_{k+1} = \mu_k(p_k) = \begin{cases} 0 & \text{if } p_k < B \\ 1 & \text{if } p_k \geq B \end{cases} \quad (2.7)$$

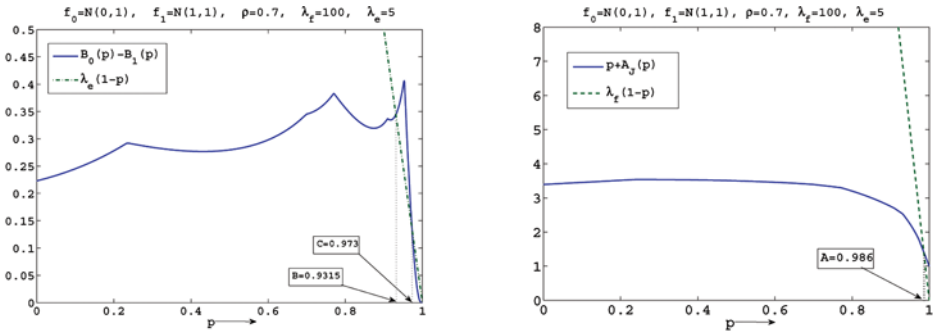
$$\tau = \inf \{k \geq 1 : p_k > A\}.$$

The probability  $p_k$  is updated using (2.3) and (2.4).

**Remark 2.2.** When  $B = 0$ ,  $\gamma(A, B)$  in (2.7) reduces to the classical Shiryaev algorithm (Shiryaev, 1963), and  $p_k$  is updated using only the recursion (2.4).

Extensive numerical studies of the Bellman equation (2.5) also show that there exist choices of  $\rho$ ,  $f_0$ ,  $f_1$ ,  $\lambda_f$ , and  $\lambda_e$  for which (2.7) is not optimal. In Figure 2 we plot one such case. Note from Figure 2(a) that again there are two points of intersection of the plotted curves, one at  $B = 0.9315$  and another at  $C = 0.973$ . But Figure 2(b) shows that  $A = 0.986 > 0.973 = C$ . Thus, in this case the optimal policy





(a)  $d(p) = B_0(p) - B_1(p)$  and  $\lambda_e(1 - p)$  as a function of  $p$  (b)  $p + A_J(p)$  and  $\lambda_f(1 - p)$  as a function of  $p$

**Figure 2.** Example where a two-threshold policy is not optimal:  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(1, 1)$ ,  $\rho = 0.7$ ,  $\lambda_f = 100$ , and  $\lambda_e = 5$ . Value iteration: number of iterations = 1, 500, number of points = 2, 000 (color figure available online).

has three thresholds. But, note that the value of  $\rho = 0.7$  is quite large and hence impractical. Also, simulations with these choices of thresholds show that the ANO is approximately zero. In all of the cases we have found for which the two-threshold policy is not optimal, the value of  $\rho$  is large and ANO is almost zero.

From a practical point of view, even if a two-threshold policy or algorithm (2.7) is not optimal, one would like to use the algorithm for the following reasons. First, as the asymptotic analysis given in Section 3 will reveal, if the PFA constraint is moderate to small and the ANO constraint is not very severe, then the thresholds  $A$  and  $B$  in  $\gamma(A, B)$  can be set independently: the threshold  $A$  can be set only based on the constraint  $\alpha$ , and the threshold  $B$  can be set based on the constraint  $\beta$  alone. Second, apart from being simple, the two-threshold algorithm (2.7) is asymptotically optimal as the PFA  $\rightarrow 0$ . Finally,  $\gamma(A, B)$  has good trade-off curves; that is, the ANO of  $\gamma(A, B)$  can be reduced by up to 70%, by keeping the ADD of the  $\gamma(A, B)$  within 10% of the ADD of the Shiryaev algorithm.

It is interesting to note that a two-threshold algorithm similar to that in (2.7) was shown to be exactly optimal in Girshick and Rubin (1952) for a different but related problem of quality control where inspection costs are considered or when the tests are destructive.

### 3. ASYMPTOTIC ANALYSIS OF $\gamma(A, B)$

In this section we derive asymptotic approximations for ADD, PFA, and ANO for the two-threshold algorithm  $\gamma(A, B)$ . To that end, we first convert the recursion for  $p_k$  (see (2.3) and (2.4)) 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

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

we can write the recursions (2.3) and (2.4) in terms of  $Z_k$ .

For  $k \geq 1$ ,

$$Z_{k+1} = Z_k + \log L(X_{k+1}) + |\log(1 - \rho)| + \log(1 + \rho e^{-Z_k}), \quad \text{if } Z_k \in [b, a] \quad (3.1)$$

and

$$Z_{k+1} = Z_k + |\log(1 - \rho)| + \log(1 + \rho e^{-Z_k}), \quad \text{if } Z_k \notin [b, a] \quad (3.2)$$

with

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

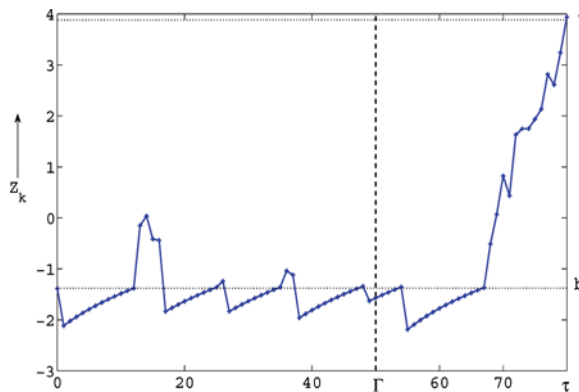
Here we have used the fact that  $S_{k+1} = 1$  if  $p_k \in [B, A]$ , and  $S_{k+1} = 0$  otherwise (see (2.6)). The crossing of thresholds  $A$  and  $B$  by  $p_k$  is equivalent to the crossing of thresholds  $a$  and  $b$  by  $Z_k$ . Thus, the stopping time for  $\gamma(A, B)$  (equivalently  $\gamma(a, b)$  with some abuse of notation) is

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

In this section we study the asymptotic behavior of  $\gamma(a, b)$  in terms of  $Z_k$  under various limits of  $a, b$ , and  $\rho$ . Specifically, we provide two asymptotic expressions for ADD, one for fixed thresholds  $a, b$ , as  $\rho \rightarrow 0$ , and another for fixed  $b$  and  $\rho$ , as  $a \rightarrow \infty$ . We also provide, for a fixed  $b$ , an asymptotic expression for PFA as  $a \rightarrow \infty$ . Finally, we also provide asymptotic estimates of the average number of observations used before (ANO) and after the change point  $\Gamma$ . Note that the limit of  $a \rightarrow \infty$  corresponds to PFA  $\rightarrow 0$ .

Figure 3 shows a typical evolution of  $\gamma(a, b)$ ; that is, of  $Z_k$  using (3.1) and (3.2), starting at time 0. Note that for  $Z_k \in [b, a]$ , recursion (3.1) is employed, while outside that interval, recursion (3.2), which only uses the prior  $\rho$ , is employed. As a result,  $Z_k$  increases monotonically outside  $[b, a]$ .

From Figure 3 again, each time  $Z_k$  crosses  $b$  from below, it can either increase to  $a$  (point  $\tau$ ) or it can go below  $b$  and approach  $b$  monotonically from below, at which



**Figure 3.** 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  $a = 3.89$  and  $b = -1.38$ , corresponding to the  $p_k$  thresholds  $A = 0.98$  and  $B = 0.2$ , respectively. Also  $Z_0 = b$  (color figure available online).

time it faces a similar set of alternatives. Thus, the passage to threshold  $a$  possibly involves multiple cycles of the evolution of  $Z_k$  below  $b$ . We will show in Section 3.3 that after the change point  $\Gamma$ , following a finite number of cycles below  $b$ ,  $Z_k$  grows up to cross  $a$ , and the time spent on the cycles below  $b$  is insignificant compared to  $\tau - \Gamma$ , as  $a \rightarrow \infty$ . In fact, we show that, asymptotically, the time to reach  $a$  is equal to the time taken by the classical Shiryaev algorithm to reach  $a$ . (Note that for the classical Shiryaev algorithm the evolution of  $Z_k$  would be based on (3.1).)

When  $Z_k$  crosses  $a$  from below, it does so with an overshoot. Overshoots play a significant role in the performance of many sequential algorithms (see Siegmund, 1985, Tartakovsky and Veeravalli, 2005) and they are central to the performance of  $\gamma(a, b)$  as well. In Section 3.2, we show that PFA depends on the threshold  $a$  and the overshoot ( $Z_\tau - a$ ) as  $a \rightarrow \infty$ , but is *not* a function of the threshold  $b$ .

The number of observations taken during the detection process is the total time spent by  $Z_k$  between  $b$  and  $a$ . As  $a \rightarrow \infty$ ,  $Z_k$  crosses  $a$  only after change point  $\Gamma$ , with high probability. The total number of observations taken can thus be divided in to two parts: the part taken before  $\Gamma$  (ANO), which is the fraction of time  $Z_k$  is above  $b$  (and hence depends only on  $b$ ), and the part taken after  $\Gamma$ . In Section 3.4 we show that, asymptotically, the average number of observations taken after  $\Gamma$  is approximately equal to the delay itself.

In Section 4 we provide approximations using which the asymptotic expressions obtained in this section can be computed and provide numerical results to demonstrate the accuracy of the approximations. In Section 5 we use the asymptotic expressions for ADD and PFA to show asymptotic optimality of  $\gamma(a, b)$ .

We begin our analysis by first obtaining the asymptotic overshoot distribution for  $(Z_\tau - a)$  using nonlinear renewal theory (Siegmund, 1985, Woodroffe, 1982). As mentioned above, this will be critical to the PFA analysis. For convenience of reference, in Table 1 we provide a glossary of important terms used in this article.

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

### 3.1. Asymptotic Overshoot

In this section we characterize the overshoot distribution of  $Z_k$  as it crosses  $a$  as  $a \rightarrow \infty$ . In analyzing the trajectory of  $Z_k$ , it useful to allow for an arbitrary starting point  $Z_0$  (shifting the time axis). We first combine the recursions in (3.1) and (3.2) to get:

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

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

$$\begin{aligned} Z_n &= \sum_{k=1}^n Y_k + \log(e^{Z_0} + \rho) + \sum_{k=1}^{n-1} \log(1 + e^{-Z_k} \rho) - \sum_{k=1}^n \mathbb{I}_{\{Z_k < b\}} \log L(X_k) \\ &= \sum_{k=1}^n Y_k + \eta_n. \end{aligned} \tag{3.3}$$

**Table 1.** Glossary

Symbol	Definition/Interpretation
ADD	Average detection delay
PFA	Probability of false alarm
ANO	Average number of observations used before change
ANO <sub>1</sub>	Average number of observations used after change
$\{X_k\}$	Observation sequence
$p_k$	<i>a posteriori</i> probability of change
$Z_k$	$\log \frac{p_k}{1-p_k} = \sum_{i=1}^k Y_i + \eta_k$ ,
$\tau$	First time for $p_k$ to cross $A$ or first time for $Z_k$ to cross $a = \log \frac{A}{1-A}$
$\{\eta_k\}$	Slowly changing sequence
$R(x), \bar{r}$	Asymptotic distribution and mean of overshoot when $\sum_{i=1}^k Y_i$ crosses a large threshold
$t(x, y)$	Time for $Z_k$ to reach $y$ starting at $x$ using (3.2)
$v(x, y)$	Time for $Z_k$ to reach $y$ starting at $x$ using (3.1) also, time for Shiryaev algorithm to reach $y$ starting at $x$
$v_b, v_0$	$v(b, a)$ and $v(-\infty, a)$
$\lambda$	Starting at $b$ , first time $Z_k$ is outside $[b, a)$
$\Lambda$	Starting at $b$ , first time $Z_k$ crosses $a$ or crosses $b$ from below
ADD <sup>s</sup>	Starting at $b$ , time for $Z_k$ to reach $a$ under $P_1$ , when $Z_k$ is reset to $b$ each time it crosses $b$ from below
$\hat{\lambda}(x)$	Starting at $x \geq b$ , first time $Z_k$ is outside $[b, a)$
$\Lambda(x)$	Starting at $x \geq b$ , first time $Z_k$ crosses $a$ or crosses $b$ from below
$\hat{\lambda}$	Starting at $b$ , first time $Z_k < b$ with $a = \infty$
$\hat{\lambda}(x)$	Starting at $x \geq b$ , first time $Z_k < b$ with $a = \infty$
$T_b$	Time spent by $Z_k$ below $b$ , after $\Gamma$ , when $\tau \geq \Gamma$
$\tilde{\Lambda}^x$	Starting at $x \geq b$ , first time $Z_k > a$ , or crosses $b$ from below, or is stopped by occurrence of change
$\delta^x$	The fraction of time $Z_k$ is above $b$ , when stopped by $\tilde{\Lambda}^x$
$\tilde{v}_b (\hat{v}_b)$	Starting at $b$ , time for $Z_k$ to reach $a$ , when $Z_k$ is reflected at $b$ (reset to $b$ when it crosses $b$ from below)

Here  $\eta_n$  is used to represent all terms other than the first in the equation above:

$$\eta_n = \log(e^{Z_0} + \rho) + \sum_{k=1}^{n-1} \log(1 + e^{-Z_k} \rho) - \sum_{k=1}^n \mathbb{I}_{\{Z_k < b\}} \log L(X_k). \quad (3.4)$$

As defined in Siegmund (1985),  $\eta_n$  is a *slowly changing* sequence if

$$n^{-1} \max\{|\eta_1|, \dots, |\eta_n|\} \xrightarrow{i.p.} 0, \quad (3.5)$$

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

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n\delta} |\eta_{n+k} - \eta_n| > \epsilon \right\} < \epsilon. \quad (3.6)$$

If indeed  $\{\eta_n\}$  is a slowly changing sequence, then the distribution of  $Z_\tau - a$ , as  $a \rightarrow \infty$ , 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 3.1.** *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  $\rho$  and  $b$ , under  $P_1$ , we have the following:*

1.  $\{\eta_n\}$  is a slowly changing sequence.
2.  $R(x)$  is the distribution of  $Z_\tau - a$  as  $a \rightarrow \infty$ ; that is,

$$\lim_{a \rightarrow \infty} \mathbf{P}[Z_\tau - a \leq x \mid \tau \geq \Gamma] = R(x). \tag{3.7}$$

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

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

It was shown in Tartakovsky and Veeravalli (2005) 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 (3.4) vanishes, and  $\eta_n$  in (3.4) behaves like the  $\eta_n$  for  $b = -\infty$ . We prove the theorem using these observations. The detailed proof is given in the Appendix to this section.  $\square$

### 3.2. PFA Analysis

We first obtain an expression for PFA as a function of the overshoot when  $Z_k$  crosses  $a$ .

**Lemma 3.1.** *For fixed  $\rho$  and  $b$ ,*

$$\text{PFA} = \mathbf{E}[1 - p_\tau] = e^{-a} \mathbf{E}[e^{-(Z_\tau - a)} \mid \tau \geq \Gamma] [1 + o(1)] \text{ as } a \rightarrow \infty.$$

*Proof.* See the Appendix for the proof.  $\square$

From Lemma 3.1, it is evident that PFA depends on the overshoot when  $Z_k$  crosses  $a$  as  $a \rightarrow \infty$ . Since the overshoot has an asymptotic distribution (Theorem 3.1) that depends only on densities  $f_0, f_1$ , and prior  $\rho$  and is independent of  $b$ , it is natural to expect that as  $a \rightarrow \infty$ , 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 in the following theorem.

**Theorem 3.2.** *For a fixed  $b$  and  $\rho$ ,*

$$\text{PFA}[\gamma(a, b)] = \left( e^{-a} \int_0^\infty e^{-x} dR(x) \right) [1 + o(1)] \text{ as } a \rightarrow \infty. \tag{3.8}$$

*Proof.* The theorem follows from Theorem 3.1 and Lemma 3.1.  $\square$

### 3.3. Delay Analysis

The PFA for  $\gamma(a, b)$  have the following bound:

$$\text{PFA} = \mathbb{E}[1 - p_\tau] \leq 1 - A = \frac{1}{1 + e^a} \leq e^{-a}. \quad (3.9)$$

Using this upper bound we can show that the ADD of  $\gamma(a, b)$  is given by:

$$\begin{aligned} \text{ADD} &= \mathbb{E}[(\tau - \Gamma)^+] \\ &= \mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma][1 + o(1)] \quad \text{as } a \rightarrow \infty. \end{aligned} \quad (3.10)$$

In the following we provide two different expressions for  $\mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma]$ . The first one is obtained by keeping  $b$  fixed and taking  $\rho \rightarrow 0$ . This expression will be used to get accurate delay estimates for  $\gamma(a, b)$  in Section 4.

Next, we will provide another asymptotic expression for  $\mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma]$  for a fixed  $b, \rho$  and as  $a \rightarrow \infty$ . We show that in this limit,  $\mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma]$  converges to the Shiryaev delay. This fact will be used to prove the asymptotic optimality of  $\gamma(a, b)$  in Section 5.

It was discussed in reference to Figure 3 that each time  $Z_k$  crosses  $b$  from below, it faces two alternatives, to cross  $a$  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  $a$  is through multiple such cycles. Motivated by this we define the following stopping times  $\lambda$  and  $\Lambda$ :

$$\lambda \triangleq \inf\{k \geq 1 : Z_k \notin [b, a), Z_0 = b\}, \quad (3.11)$$

and

$$\Lambda \triangleq \inf\{k \geq 1 : Z_k > a \text{ or } \exists k \text{ s.t. } Z_{k-1} < b \text{ and } Z_k \geq b, Z_0 = b\}. \quad (3.12)$$

Let  $t(x, y)$  be the constant time taken by  $Z_k$  to move from  $Z_0 = x$  to  $y$  using the recursion (3.2); that is,

$$t(x, y) \triangleq \inf\{k \geq 0 : Z_k > y, Z_0 = x, x, y \notin [b, a)\}. \quad (3.13)$$

Then, we can write  $\Lambda$  as a function of  $\lambda$  using (3.13):

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

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  $a$  is the sum of a finite but random number of random variables with distribution of  $\Lambda$ , say  $\Lambda_1, \Lambda_2, \dots, \Lambda_N$ . For  $i = 1, \dots, N - 1$ ,  $Z_{\Lambda_i} < b$ , and  $Z_{\Lambda_N} > a$ . Thus, the time to reach  $a$  in this case is  $\mathbb{E}_1\left[\sum_{k=1}^N \Lambda_k\right]$ . Let

$$\text{ADD}^s \triangleq \mathbb{E}_1\left[\sum_{k=1}^N \Lambda_k\right].$$

The behavior of the delay path depends on  $Z_\Gamma$ , the value of  $Z_k$  at the change point  $\Gamma$ , and how  $Z_k$  evolves after that point. We use  $\{Z_k \nearrow b\}$  to indicate that  $Z_k$  approaches  $b$  from below for some  $k > \Gamma$ ; that is,  $\exists k > \Gamma, s.t., Z_{k-1} < b, Z_k \geq b$ , and use  $\{Z_k \nearrow a\}$  to represent the event that  $Z_k$  crossed  $a$  without ever coming back to  $b$ ; that is,  $Z_k \geq b, \forall k > \Gamma$ . We define the following three disjoint events:

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

Thus, under event  $\mathcal{A}$ , the process  $Z_k$  starts below  $b$  at  $\Gamma$  and reaches  $a$  after multiple upcrossings of the threshold  $b$ . Under event  $\mathcal{B}$ , the process  $Z_k$  starts above  $b$  at  $\Gamma$  and crosses  $b$  before  $a$ . It then has multiple up-crossings of  $b$ , similar to the case of event  $\mathcal{A}$ . Under event  $\mathcal{C}$ , the process  $Z_k$  starts above  $b$  at  $\Gamma$  and reaches  $a$  without ever coming below  $b$ .

Also define

$$\lambda(x) = \inf\{k \geq 1 : Z_k \notin [b, a], Z_0 = x, b \leq x < a\}, \tag{3.14}$$

and let  $\Lambda(x)$  be defined with  $Z_0 = x$  similar to (3.12). Thus,  $\lambda$  and  $\lambda(b)$  have the same distribution. Similarly,  $\Lambda$  and  $\Lambda(b)$  are identically distributed.

The following theorem gives an asymptotic expression for the conditional delay.

**Theorem 3.3.** *For a fixed value of the thresholds  $a, b$ , the conditional delay is given by*

$$\begin{aligned} E[\tau - \Gamma | \tau \geq \Gamma] &= [ADD^s P(\mathcal{A} \cup \mathcal{B} | \tau \geq \Gamma) + E[\Lambda(Z_\Gamma) | \mathcal{C}, \tau \geq \Gamma] P(\mathcal{C} | \tau \geq \Gamma) \\ &\quad + E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma] P(\mathcal{A} | \tau \geq \Gamma) + E[\Lambda(Z_\Gamma) | \mathcal{B}, \tau \geq \Gamma] \\ &\quad \times P(\mathcal{B} | \tau \geq \Gamma)][1 + o(1)] \text{ as } \rho \rightarrow 0. \end{aligned} \tag{3.15}$$

*Proof.* The proof is provided in the Appendix. □

In Section 4 we will provide approximations for various terms in (3.15) to get an accurate estimate of ADD. In Lemma 3.2 we provide expressions for  $ADD^s$ .

Let  $\Psi$  represent the Shiryaev recursion; that is, updating  $Z_k$  using only (3.1). Define

$$v(x, y) = \inf\{k \geq 1 : \Psi(Z_{k-1}) > y, Z_0 = x\}. \tag{3.16}$$

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

$$v_b = v(b, a), \tag{3.17}$$

and

$$v_0 = v(-\infty, a). \tag{3.18}$$

Note that  $v_0$  is the stopping time for the classical Shiryaev algorithm (Shiryaev, 1963) and  $v_b$  is its modified form, which starts at  $b$ . We have the following asymptotic expression.

**Lemma 3.2.** For a fixed  $b$  and  $\rho$ ,  $\text{ADD}^s$ , the average time for  $Z_k$  to cross  $a$  starting at  $b$ , under  $P_1$ , with  $Z_k$  reset to  $b$  each time it crosses  $b$  from below, is given by

$$\text{ADD}^s = \frac{E_1[\lambda] + E_1[t(Z_\lambda, b) | \{Z_\lambda < b\}]P_1(Z_\lambda < b)}{P_1(Z_\lambda > a)} \quad (3.19)$$

and is asymptotically equal to the time taken by the Shiryaev algorithm to move from  $b$  to  $a$ ; that is,

$$\text{ADD}^s = E_1[v_b][1 + o(1)] \quad \text{as } a \rightarrow \infty. \quad (3.20)$$

*Proof.* We have

$$\begin{aligned} \text{ADD}^s &= E_1 \left[ \sum_{k=1}^N \Lambda_k \right] \\ &\stackrel{(i)}{=} E_1[N]E_1[\Lambda] \\ &\stackrel{(ii)}{=} \frac{E_1[\Lambda]}{P_1(Z_\lambda > a)} \\ &= \frac{E_1[\lambda] + E_1[t(Z_\lambda, b) | \{Z_\lambda < b\}]P_1(Z_\lambda < b)}{P_1(Z_\lambda > a)}. \end{aligned}$$

In the above equation, equality (i) follows from Wald's lemma (Siegmund, 1985), and equality (ii) follows because  $N \sim \text{Geom}[P(Z_\lambda > a)]$ . To obtain (3.20), the main idea of the proof is to find stopping times that upper and lower bound the Shiryaev time on average and have delay equal to  $\frac{E_1[\lambda]}{P_1(Z_\lambda > a)}$  as  $a \rightarrow \infty$ . The details are provided in the Appendix.  $\square$

Note that Theorem 3.3 takes  $\rho \rightarrow 0$ . We now provide another expression for  $E[\tau - \Gamma | \tau \geq \Gamma]$ , for a fixed  $b$  and  $\rho$  as  $a \rightarrow \infty$ , which will be used to prove the asymptotic optimality of  $\gamma(a, b)$  in Section 5.

**Theorem 3.4.** For a fixed  $b$  and  $\rho$ , we have as  $a \rightarrow \infty$

$$E[\tau - \Gamma | \tau \geq \Gamma] \leq \text{ADD}^s [1 + o(1)], \quad (3.21)$$

and, hence, we have

$$E[\tau - \Gamma | \tau \geq \Gamma] = \left[ \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} \right] [1 + o(1)] \quad \text{as } a \rightarrow \infty, \quad (3.22)$$

where  $D(f_1, f_0)$  is the K-L divergence between  $f_0$  and  $f_1$ .

*Proof.* To get (3.21), we show that  $\text{ADD}^s$  is the dominant term in an upper bound to  $E[\tau - \Gamma | \tau \geq \Gamma]$  as  $a \rightarrow \infty$ . The steps followed are very similar to those used to obtain (3.15). The proof is given in the Appendix.

To obtain (3.22), from Lemma 3.2 and (3.21) we have

$$E[\tau - \Gamma | \tau \geq \Gamma] \leq E_1[v_b][1 + o(1)] \quad \text{as } a \rightarrow \infty.$$



To evaluate  $E_1[v_b]$ , following steps similar to those in Section 3.1, it is easy to show that evolution of  $Z_k$  from  $b$  to  $a$ , with  $Z_0 = b$ , is according to the random walk  $\sum_k \log L(X_k) + |\log(1 - \rho)|$  and a slowly changing term. Thus, according to Lemma 9.1.3, p. 191 of Siegmund (1985),

$$E_1[v_b] = \left[ \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} \right] [1 + o(1)] \quad \text{as } a \rightarrow \infty,$$

and

$$E[\tau - \Gamma | \tau \geq \Gamma] \leq \left[ \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} \right] [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

To complete the proof of Theorem 3.4, we now show that  $E[\tau - \Gamma | \tau \geq \Gamma]$  is asymptotically lower bounded by  $E_1[v_b]$ . From Theorem 1 in Tartakovsky and Veeravalli (2005),

$$E[v_0 - \Gamma | v_0 \geq \Gamma] \geq \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

Also, from Theorem 3.2,

$$P[\tau < \Gamma] = P[v_0 < \Gamma] [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

Thus, we have

$$E[\tau - \Gamma | \tau \geq \Gamma] \geq E[v_0 - \Gamma | v_0 \geq \Gamma] [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

This is true because the Shiryaev algorithm is optimal for problem (2.1) with  $\beta = \infty$ . This completes the proof.  $\square$

### 3.4. Computation of ANO

In this section we provide an asymptotic expression for ANO for a fixed  $b$ , as  $a \rightarrow \infty$ , and as  $\rho \rightarrow 0$ , with  $a \rightarrow \infty$  before  $\rho \rightarrow 0$ . In Section 4.2 we show that this expression provides a good estimate of ANO and can be used to choose the threshold  $b$  for the test.

First note that

$$\begin{aligned} \text{ANO} &= E \left[ \sum_{k=1}^{\min\{\tau, \Gamma-1\}} S_k \right] \\ &= E \left[ \sum_{k=1}^{\Gamma-1} S_k \mid \tau \geq \Gamma \right] P(\tau \geq \Gamma) + E \left[ \sum_{k=1}^{\tau} S_k \mid \tau < \Gamma \right] P(\tau < \Gamma) \\ &= E \left[ \sum_{k=1}^{\Gamma-1} S_k \mid \tau \geq \Gamma \right] [1 + o(1)] \quad \text{as } a \rightarrow \infty. \end{aligned}$$

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

Following (3.11), we define

$$\hat{\lambda} = \inf\{k \geq 1 : Z_k < b, Z_0 = b, a = \infty\}. \quad (3.23)$$

The following theorem provides an asymptotic expression for ANO.

**Theorem 3.5.** *For fixed  $b$ , we have as  $a \rightarrow \infty$ , and as  $\rho \rightarrow 0$ , with  $a \rightarrow \infty$  before  $\rho \rightarrow 0$ ,*

$$\text{ANO} = \frac{\mathbb{E}_\infty[\hat{\lambda}]}{\mathbb{P}_\infty[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)]} \frac{1}{1 + e^b} [1 + o(1)],$$

where  $\hat{\lambda}$  is as defined in (3.23).

*Proof.* Let  $t(b)$  be the first time  $Z_k$  crossed  $b$  from below; that is,  $t(b) = t(z_0, b)$ . Using the fact that observations are used only after  $t(b)$ , we can write the following:

$$\begin{aligned} \text{ANO} &= \mathbb{E} \left[ \sum_{k=1}^{\Gamma-1} S_k \mid \tau \geq \Gamma \right] \\ &= \mathbb{E} \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), \tau \geq \Gamma \right] \mathbb{P}[\Gamma > t(b) \mid \tau \geq \Gamma]. \end{aligned} \quad (3.24)$$

We now compute each of the two terms in (3.24). For the first term in (3.24), we have the following lemma.

**Lemma 3.3.** *For a fixed  $b$ , as  $a \rightarrow \infty$ ,  $\rho \rightarrow 0$ , with  $a \rightarrow \infty$  before  $\rho \rightarrow 0$ ,*

$$\mathbb{E} \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), \tau \geq \Gamma \right] = \frac{\mathbb{E}_\infty[\hat{\lambda}]}{\mathbb{P}_\infty[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)]} [1 + o(1)].$$

*Proof.* Note that

$$\lim_{a \rightarrow \infty} \mathbb{E} \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), \tau \geq \Gamma \right] = \mathbb{E} \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), a = \infty \right].$$

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.  $\square$

For the second term in (3.24), we show that  $\mathbb{P}(\Gamma > t(b) \mid \tau \geq \Gamma)$  is equal to  $\frac{1}{1+e^b}$  in the limit and is independent of  $z_0$ .

**Lemma 3.4.** *As  $a \rightarrow \infty$ ,  $\rho \rightarrow 0$ , with  $a \rightarrow \infty$  before  $\rho \rightarrow 0$ ,*

$$\mathbb{P}(\Gamma > t(b) \mid \tau \geq \Gamma) = \frac{1}{1 + e^b} + o(1).$$

*Proof.* The proof is provided in the Appendix.  $\square$

Lemmas 3.3 and 3.4 taken together complete the proof of Theorem 3.5.  $\square$

Define

$$\text{ANO}_1 = \mathbb{E} \left[ \sum_{k=\Gamma}^{\tau} S_k \mid \tau \geq \Gamma \right].$$

Thus,  $\text{ANO}_1$  is the average number of observations used after the change point  $\Gamma$ . In some applications it might be of interest to have an estimate of  $\text{ANO}_1$  as well. The following theorem shows that  $\text{ANO}_1$  is approximately equal to the delay itself.

**Theorem 3.6.** *For fixed  $b$  and  $\rho$ , we have*

$$\text{ANO}_1 = \mathbb{E}_1[v_b][1 + o(1)], \quad \text{as } a \rightarrow \infty.$$

*Proof.* The number of observations used after  $\Gamma$  can be written as the difference between the time for  $Z_k$  to reach  $a$  and the time spend by it below  $b$ . For this we define the variable

$$T_b \triangleq \mathbb{E} \left[ \sum_{k=\Gamma}^{\tau} 1_{\{Z_k < b\}} \mid \tau \geq \Gamma \right].$$

Thus,

$$\text{ANO}_1 = \mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma] - T_b + 1.$$

We know from Theorem 3.4 that  $\mathbb{E}[\tau - \Gamma \mid \tau \geq \Gamma] \approx \mathbb{E}_1[v_b]$ . As  $a \rightarrow \infty$ ,  $T_b$  converges, and therefore  $\text{ANO}_1 \approx \mathbb{E}_1[v_b]$  for large  $a$  as well. The detailed proof is given in the Appendix.  $\square$

## 4. APPROXIMATIONS AND NUMERICAL RESULTS

In Sections 3.2–3.4, we have obtained asymptotic expressions for ADD, PFA, and ANO as a function of the system parameters: the thresholds  $a$ ,  $b$ ; the densities  $f_0$  and  $f_1$ ; and the prior  $\rho$ . We now provide approximations for some of the analytical expressions obtained in these sections and also provide numerical results to validate the analysis. The observations are assumed to be 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  $\mathbb{E}[1 - p_\tau]$ . This guarantees a faster convergence for small values of PFA.

### 4.1. Numerical Results for PFA

By Theorem 3.2, we have the following approximation for PFA:

$$\text{PFA} \approx e^{-a} \int_0^\infty e^{-x} dR(x).$$

**Table 2.** PFA: for  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(\theta, 1)$ 

$\theta$	$\rho$	$a$	$b$	PFA simulations	PFA analysis
0.4	0.01	3.0	0	$3.78 \times 10^{-2}$	$3.94 \times 10^{-2}$
0.4	0.01	6.0	2.0	$1.955 \times 10^{-3}$	$1.96 \times 10^{-3}$
0.75	0.01	9.0	-2.0	$7.968 \times 10^{-5}$	$7.964 \times 10^{-5}$
2.0	0.01	5.0	-4.0	$2.15 \times 10^{-3}$	$2.155 \times 10^{-3}$
0.75	0.005	7.6	3.0	$3.231 \times 10^{-4}$	$3.235 \times 10^{-4}$
0.75	0.1	4.0	-3.0	$1.143 \times 10^{-2}$	$1.157 \times 10^{-2}$

We note that  $\int_0^\infty e^{-x} dR(x)$  and  $\bar{r}$  can be computed numerically, at least for Gaussian observations (Siegmund, 1985). In this section we provide numerical results to show the accuracy of the above expression for PFA.

In Table 2 we compare the analytical approximation with the PFA obtained using simulations of  $\gamma(a, b)$  for various choices of  $\rho$ , thresholds  $a, b$ , and post-change mean  $\theta$ . From the table we see that the analytical approximation is quite good.

In Table 3 we show that PFA is not a function of  $b$  for large values of  $a$ . We fix  $a = 4.6$  and increase  $b$  from  $-2.2$  to  $0.85$ . We notice that PFA is unchanged in simulations when  $b$  is changed in this way. This is also captured by the analysis and it is quite accurate.

#### 4.2. Approximations and Numerical Results for ANO and ANO<sub>1</sub>

We recall the expressions for ANO from Theorem 3.5 and for ANO<sub>1</sub> from Theorem 3.6:

$$\text{ANO} \approx \frac{E_\infty[\hat{\lambda}]}{P_\infty[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)]} \frac{1}{1 + e^b}$$

$$\text{ANO}_1 = E_1[v_b].$$

We first simplify the expression for ANO. Note that

$$P_\infty[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)] = 1 - P_\infty[\Gamma > \hat{\lambda} + t(Z_{\hat{\lambda}}, b)]$$

$$= 1 - E_\infty[(1 - \rho)^{\hat{\lambda} + t(Z_{\hat{\lambda}}, b)}].$$

**Table 3.** PFA for  $\rho = 0.01$ ,  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$ 

$a$	$b$	Simulations	Analysis
4.6	-2.2	$6.44 \times 10^{-3}$	$6.48 \times 10^{-3}$
4.6	-1.5	$6.44 \times 10^{-3}$	$6.48 \times 10^{-3}$
4.6	-0.85	$6.44 \times 10^{-3}$	$6.48 \times 10^{-3}$
4.6	0	$6.44 \times 10^{-3}$	$6.48 \times 10^{-3}$
4.6	0.85	$6.44 \times 10^{-3}$	$6.48 \times 10^{-3}$

Thus, using binomial approximation we get

$$P_\infty[\Gamma \leq \hat{\lambda} + t(Z_{\hat{\lambda}}, b)] \approx \rho \left\{ E_\infty[\hat{\lambda}] + E_\infty[t(Z_{\hat{\lambda}}, b)] \right\}.$$

Thus, we have

$$\text{ANO} \approx \frac{\rho^{-1} E_\infty[\hat{\lambda}]}{E_\infty[\hat{\lambda}] + E_\infty[t(Z_{\hat{\lambda}}, b)]} \frac{1}{1 + e^b}. \tag{4.1}$$

We now provide approximation to compute  $E_\infty[\hat{\lambda}]$  and  $E_\infty[t(Z_{\hat{\lambda}}, b)]$  in (4.1). Invoking Wald's lemma (Siegmund, 1985), we write  $E_\infty[\hat{\lambda}]$  as

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

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

$$E_\infty[\hat{\lambda}] \approx \frac{\bar{r} + \log(1 + \rho e^{-b})}{D(f_1, f_0) - |\log(1 - \rho)|}. \tag{4.2}$$

Here,  $\log(1 + \rho e^{-b})$  is an approximation to  $E_\infty[\eta_{\hat{\lambda}}]$  by ignoring all the random terms after  $b$  is factored out of it. This extra  $b$  will cancel with the  $b$  in  $E_\infty[Z_{\hat{\lambda}}] = b + E_\infty[Z_{\hat{\lambda}} - b]$ . We approximate  $E_\infty[b - Z_{\hat{\lambda}}]$  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 (3.3)).

For the term  $E_\infty[t(Z_{\hat{\lambda}}, b)]$ , we have the following lemma.

**Lemma 4.1.** *For fixed values of  $x$  and  $y$ , we have*

$$t(x, y) = \left[ \frac{\log(1 + e^y) - \log(1 + e^x)}{|\log(1 - \rho)|} \right] (1 + o(1)) \text{ as } \rho \rightarrow 0. \tag{4.3}$$

*Proof.* The proof is provided in the Appendix. □

We use (4.3) to get the following approximation:

$$E_\infty[t(Z_{\hat{\lambda}}, b)] \approx \int_0^\infty \frac{\log(1 + e^b) - \log(1 + e^{b-x})}{|\log(1 - \rho)|} dR(x). \tag{4.4}$$

Thus, we approximate the distribution of  $(b - Z_{\hat{\lambda}})$  by  $R(x)$ .

Based on the second-order approximation for  $E_1[v_0]$  developed in Tartakovsky and Veeravalli (2005), we have obtained the following approximation for  $E_1[v_b]$ :

$$E_1[v_b] = \frac{a - E[\eta(b)] + \bar{r}}{D(f_1, f_0) + |\log(1 - \rho)|} + o(1) \quad \text{as } a \rightarrow \infty, \quad (4.5)$$

where  $\eta(b)$  is the a.s. limit of the slowly changing sequence  $\eta_n$  with  $Z_0 = b$  under  $f_1$  (see (3.4)) and

$$\bar{r} = \int_0^\infty x dR(x), \quad (4.6)$$

with  $R(x)$  as in Theorem 3.1.

In Table 4 we demonstrate the accuracy of approximations for ANO and ANO<sub>1</sub> for various values of  $\rho$ , thresholds  $a, b$ , and post-change mean  $\theta$ . The table shows that the approximations are quite accurate for the parameters chosen.

### 4.3. Approximations and Numerical Results for ADD

Theorem 3.4 gave a first-order approximation for  $E[\tau - \Gamma | \tau \geq \Gamma]$ :

$$E[\tau - \Gamma | \tau \geq \Gamma] \approx \left[ \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} \right].$$

Note that, from Tartakovsky and Veeravalli (2005), this is also the first-order approximation for the ADD of the Shiryaev algorithm and gives a good estimate of the delay when PFA is small. For the Shiryaev delay, a second-order approximation was developed in Tartakovsky and Veeravalli (2005) (also see (4.5)):

$$E_1[v_0] = \left\{ \frac{a - E[\eta(-\infty)] + \bar{r}}{D(f_1, f_0) + |\log(1 - \rho)|} \right\} + o(1) \quad \text{as } a \rightarrow \infty.$$

So, instead of using  $\frac{a}{D(f_1, f_0) + |\log(1 - \rho)|}$ , we propose to use the following:

$$E[\tau - \Gamma | \tau \geq \Gamma] \approx \left\{ \frac{a - E[\eta(-\infty)] + \bar{r}}{D(f_1, f_0) + |\log(1 - \rho)|} \right\}. \quad (4.7)$$

For the Shiryaev algorithm, (4.7) provides a very good estimate of the delay even for moderate values of PFA. In case of  $\gamma(a, b)$ , the accuracy of (4.7) depends

**Table 4.**  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(\theta, 1)$

$\theta$	$\rho$	$a$	$b$	ANO		ANO <sub>1</sub>	
				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 5.**  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(\theta, 1)$

$\theta$	$\rho$	ADD				PFA		
		$a$	$b$	Simulations	Analysis	Simulations	Analysis	ANO%
				$E[\tau - \Gamma   \tau \geq \Gamma]$	(4.7)			
0.4	0.01	8.5	-2.2	104.9	111.7	$1.608 \times 10^{-4}$	$1.608 \times 10^{-4}$	66
0.75	0.01	6.467	-2.2	32.3	29.5	$1.002 \times 10^{-3}$	$1.004 \times 10^{-3}$	35
2.0	0.01	7.5	-4.0	6.1	6.23	$1.77 \times 10^{-4}$	$1.768 \times 10^{-4}$	43
0.75	0.005	8.7	-3.0	42.6	40.4	$1.076 \times 10^{-4}$	$1.076 \times 10^{-4}$	77
0.75	0.1	8.5	0.0	23.9	22.18	$1.286 \times 10^{-4}$	$1.285 \times 10^{-4}$	26

on the choice of  $b$  and hence on the constraint  $\beta$ , as having  $b > -\infty$  increases the delay. Before we demonstrate this through numerical and simulation results we introduce the following concept:

$$\text{ANO\%} = \text{ANO expressed as a percentage of } E[\Gamma]. \tag{4.8}$$

For example, if  $\rho = 0.05$ , and for some choice of system parameters  $\text{ANO} = 15$ , then  $\text{ANO\%} = 15 * 0.05 = 75\%$ . Thus, the concept of ANO% captures the reduction in the average number of observations used before change by employing  $\gamma(a, b)$ .

In Table 5 we provide various numerical examples where (4.7) is a good approximation for  $E[\tau - \Gamma | \tau \geq \Gamma]$ . Since (4.7) is a good approximation for the Shiryaev delay as well, it follows that, for these parameter values, the delay of  $\gamma(a, b)$  is approximately equal to the Shiryaev delay. It might be intuitive that if we are aiming for large ANO% values of, say, 90%, then the delay will be close to the Shiryaev delay. But values in Table 5 shows that it is possible to achieve considerably smaller values of ANO% without significantly affecting the delay.

However, if the ANO% value is small, then this means that the value of  $b$  is large and, further, that the delay is large. In this case, it might happen that (4.7) is a good approximation only for values of PFA which are very small. This is demonstrated in Table 6. It is clear from the table that, for the parameter values considered, estimating the delay with less than 10% error is only possible at PFA values of the order of  $\text{PFA} \approx 10^{-22}$ .

This motivates the need for a more accurate estimate of the delay. This is provided below.

From Theorem 3.3, recall that we had the following three events:

$$\mathcal{A} = \{Z_\Gamma < b\},$$

**Table 6.**  $\rho = 0.05$ ,  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$

$a$	$b$	Simulations $E[\tau - \Gamma   \tau \geq \Gamma]$	Analysis (4.7)	ANO%	PFA
5.0	1.0	30	13	7.5	$4.3 \times 10^{-3}$
9.0	1.0	42	25	7.5	$7.9 \times 10^{-5}$
13.0	1.0	54	37	7.5	$1.4 \times 10^{-6}$
18.0	1.0	69	52	7.5	$9.7 \times 10^{-9}$
50.0	1.0	165	149	7.5	$1.23 \times 10^{-22}$

Downloaded by [University of Illinois at Urbana-Champaign] at 16:42 24 February 2012

$$\mathcal{B} = \{Z_\Gamma \geq b; Z_k \nearrow b\},$$

$$\mathcal{C} = \{Z_\Gamma \geq b; Z_k \nearrow a\}.$$

As a first step toward the approximations, we ignore the event  $\mathcal{B}$ :  $\mathbf{P}(\mathcal{B}) \approx 0$ . That is, we assume that if  $Z_\Gamma > b$ , then  $Z_k$  climbs to  $a$ . Define

$$P_b = \mathbf{P}(Z_\Gamma \geq b | \tau \geq \Gamma).$$

Then (3.15),

$$\mathbf{E}[\tau - \Gamma | \tau \geq \Gamma] \approx P_b \mathbf{E}[\lambda(Z_\Gamma) | \mathcal{C}, \tau \geq \Gamma] + (1 - P_b) \{ \mathbf{E}[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma] + \text{ADD}^s \}. \quad (4.9)$$

From Lemma 3.2, it is easy to show the following:

$$\begin{aligned} \text{ADD}^s &= \mathbf{E}_1[\lambda | \{Z_\lambda > a\}] + (\mathbf{E}_1[\lambda | \{Z_\lambda < b\}] + \mathbf{E}_1[t(Z_\lambda, b) | \{Z_\lambda < b\}]) \\ &\quad \times \frac{\mathbf{P}_1(Z_\lambda < b)}{1 - \mathbf{P}_1(Z_\lambda < b)}. \end{aligned}$$

We now use the following approximations:

$$\mathbf{E}_1[\lambda | \{Z_\lambda > a\}] \approx \mathbf{E}[\lambda(Z_\Gamma) | \mathcal{C}, \tau \geq \Gamma] \approx \frac{a - \mathbf{E}[\eta(-\infty)] + \bar{r}}{D(f_1, f_0) + |\log(1 - \rho)|},$$

$$\mathbf{E}_1[\lambda | \{Z_\lambda < b\}] \approx \frac{\bar{r} + \log(1 + \rho e^{-b})}{D(f_1, f_0) - |\log(1 - \rho)|},$$

$$\mathbf{E}_1[t(Z_\lambda, b) | \{Z_\lambda < b\}] \approx t(b - \bar{r}, b) \approx \frac{\log(1 + e^b) - \log(1 + e^{b-\bar{r}})}{|\log(1 - \rho)|}.$$

To compute (4.9), we also need approximations for  $\mathbf{P}_1(Z_\lambda < b)$ ,  $P_b$ , and  $\mathbf{E}[t(Z_\Gamma, b) | \mathcal{A}]$ . Those are provided below. Setting  $a = \infty$  we have, by Wald's likelihood identity, Siegmund (1985, Proposition 2.24, p. 13),

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

Under  $\mathbf{P}_\infty$ ,  $\lambda$  a.s. ends in  $b$ , and with high probability it takes very small values. Hence, this expression can be computed using Monte Carlo simulations. Further,

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

We already have the approximations for  $\mathbf{E}_\infty[\hat{\lambda}]$  and  $\mathbf{E}_\infty[t(Z_\lambda, b)]$  from Section 4.2. The approximation for  $\mathbf{E}[t(Z_\Gamma, b) | \mathcal{A}]$  can be obtained as follows (all expectations conditioned on  $\{\tau \geq \Gamma\}$ ):

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



This can be computed using

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

and

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

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

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

and

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

Thus, we have obtained approximations for all the terms for the new approximation for  $E[\tau - \Gamma | \tau \geq \Gamma]$  in (4.9).

In Table 7, we now reproduce Table 6 with a new column containing delay estimates computed using the new ADD (for  $E[\tau - \Gamma | \tau \geq \Gamma]$ ) approximation (4.9). The values shows that all estimates are nearly within 10% of the actual value.

In Table 8, we show the accuracy of the new ADD approximation (4.9), for various values of the system parameters, by comparing it with simulations and also with (4.7). We also set PFA around  $1 \times 10^{-3}$ . The table clearly demonstrates that the new ADD approximation predicts the ADD with less than 10% error.

**Table 7.**  $\rho = 0.05$ ,  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$

$a$	$b$	Simulations	Analysis	New analysis		PFA
		$E[\tau - \Gamma   \tau \geq \Gamma]$	(4.7)	ADD from (4.9)	ANO%	
5.0	1.0	30	13	34	7.5	$4.3 \times 10^{-3}$
9.0	1.0	42	25	46	7.5	$7.9 \times 10^{-5}$
13.0	1.0	54	37	58	7.5	$1.4 \times 10^{-6}$
18.0	1.0	69	52	73	7.5	$9.7 \times 10^{-9}$
50.0	1.0	165	149	169	7.5	$1.23 \times 10^{-22}$

**Table 8.**  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$ , PFA  $\approx 10^{-3}$ , ANO=10% of Shiryaev's ANO

$\rho$	$a$	$b$	ADD			ANO%
			Simulations	Analysis new (4.9)	Analysis (4.7)	
0.01	6.4	2.7	250	260	14.42	0.33
0.005	6.45	0.6	181	190	22.09	1.5
0.001	6.47	-2.7	75	80	33.68	7.6
0.0005	6.47	-3.49	74	79	36.49	8.4
0.0001	6.47	-5.2	76	80	42.56	9.6

## 5. ASYMPTOTIC OPTIMALITY AND PERFORMANCE OF $\gamma(a, b)$

### 5.1. Asymptotic Optimality of $\gamma(a, b)$

In Theorem 3.4 we saw that for a fixed  $b$  and  $\rho$ ,

$$\mathbb{E}[\tau - \Gamma | \tau \geq \Gamma] = \left[ \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|} \right] [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

We recall that from Tartakovsky and Veeravalli (2005), this is also the asymptotic delay of the Shiryaev algorithm.

Moreover, from Theorem 3.2, the PFA for  $\gamma(a, b)$  is

$$\text{PFA} = \left( e^{-a} \int_0^\infty e^{-x} dR(x) \right) [1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

Again from Tartakovsky and Veeravalli (2005), this is the PFA for the Shiryaev algorithm. We thus have the following asymptotic optimality result for  $\gamma(a, b)$ .

**Theorem 5.1.** *With  $\gamma = \{\tau, S_1, \dots, S_\tau\}$  define*

$$\Delta(\alpha, \beta) = \{\gamma : \text{PFA}(\gamma) \leq \alpha; \text{ANO}(\gamma) \leq \beta\},$$

*then for a fixed  $\beta$  and  $\rho$ ,*

$$\text{ADD}\{\gamma[a(\alpha, \beta), b(\alpha, \beta)]\} = \left[ \inf_{\gamma \in \Delta(\alpha, \beta)} \text{ADD}(\gamma) \right] [1 + o(1)] \quad \text{as } \alpha \rightarrow 0. \quad (5.1)$$

*Here, for each  $\alpha, \beta$ ,  $b(\alpha, \beta)$  is the smallest  $b$  such that  $\text{ANO}\{\gamma[a(\alpha, \beta), b(\alpha, \beta)]\} \leq \beta$  as  $a \rightarrow \infty$ .*

*Proof.* Fix  $b$  such that  $\text{ANO}[\gamma(a, b)] \leq \beta$  as  $a \rightarrow \infty$ . It may happen that the constraint  $\beta$  is not met with equality. Then we choose the smallest  $b$  that satisfies the constraint  $\beta$  as  $a \rightarrow \infty$ . This choice of threshold  $b$  is unique for a given  $\beta$  because ANO is not a function of threshold  $a$  as  $a \rightarrow \infty$ .

As  $a \rightarrow \infty$ , the PFA and ADD both approach the Shiryaev PFA (3.8) and Shiryaev delay (3.22), respectively. Thus, as  $a \rightarrow \infty$ ,  $\gamma(a, b)$  is optimal over the class of all control policies  $\Delta(\alpha, \beta)$  that satisfy the constraints  $\alpha$  and  $\beta$ .  $\square$

## 5.2. Trade-off Curves: Performance of $\gamma(a, b)$ for a Fixed and Moderate $\alpha$

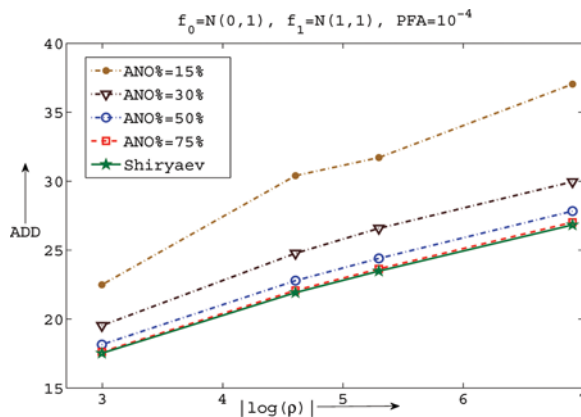
Theorem 5.1 shows that for small values of PFA,  $\gamma(a, b)$  is approximately optimal; that is, it is not possible to outperform  $\gamma(a, b)$  by a significant margin. But for moderate values of PFA, it is not clear whether there exist algorithms that can significantly outperform  $\gamma(a, b)$ . Our aim is to partially address this issue in this section.

In Figure 4 we plot the ANO-ADD trade-off for the two-threshold algorithm. Specifically, we compare the two-threshold algorithm with the classical Shiryaev algorithm and study how much ANO can be reduced without significantly losing in terms of ADD. For Figure 4 we pick four values of  $\rho$ : 0.05, 0.01, 0.005, 0.001. For a fixed  $\rho$ , we fix  $b = -\infty$  and select threshold  $a$  such that the  $\text{PFA}[\gamma(a, b)] = 10^{-4}$ . We then increase the threshold  $b$  to have ANO% values of 75, 50, 30, and 15. We note that it was possible to reduce the ANO to 15% of  $E[\Gamma]$  by increasing the threshold  $b$  this way, without affecting the probability of false alarm. Figure 4 shows that we can reduce ANO by up to 25% while obtaining approximately the same ADD performance as that of the Shiryaev algorithm. Moreover, if we allow for a 10% increase in ADD compared to that of the Shiryaev algorithm, then we can reduce ANO by up to 70% (see plot for ANO% = 30%).

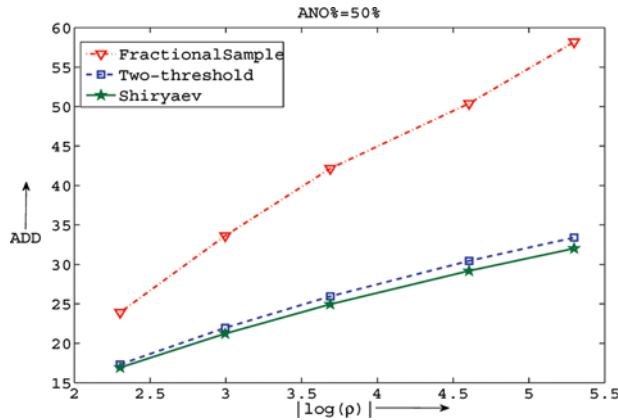
Such a behavior was also observed in Table 5, where we saw that the delay for  $\gamma(a, b)$  is approximately equal to the Shiryaev delay for moderate to large ANO% values. Thus, for moderate PFA values, when the ANO% is moderate to large,  $\gamma(a, b)$  is approximately optimal.

## 5.3. Comparison with Fractional Sampling

In this section we compare the performance of  $\gamma(a, b)$  with the naïve approach of fractional sampling, in which an ANO% of  $\epsilon\%$  is achieved by employing the Shiryaev algorithm and using a sample with probability  $\frac{\epsilon}{100}$ . Also, in fractional



**Figure 4.** Trade-off curves comparing performance of two-threshold algorithm with the Shiryaev algorithm for ANO% of 75, 50, 30, and 15.  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(1, 1)$ , and  $\text{PFA} = 10^{-4}$  (color figure available online).



**Figure 5.** Trade-off curves comparing performance of the two-threshold algorithm with the fractional sampling scheme for ANO% of 50.  $f_0 \sim \mathcal{N}(0, 1)$ ,  $f_1 \sim \mathcal{N}(0.75, 1)$ , and  $\text{PFA} = 10^{-3}$  (color figure available online).

sampling, when a sample is skipped, the posterior probability  $p_k$  is updated using (2.3). Figure 5 compares the two schemes for ANO% of 50. We also plot the performance of the Shiryaev algorithm for the same values of PFA and  $\rho$ . The figure shows that  $\gamma(a, b)$  helps in reducing the observation cost by a significant margin compared to the fractional sampling scheme.

From our approximations, we know that for large  $a$

$$\text{ADD}[\gamma(a, b)] \approx \frac{a}{D(f_1, f_0) + |\log(1 - \rho)|}.$$

When the K-L distance  $D(f_1, f_0)$  dominates the sum  $D(f_1, f_0) + |\log(1 - \rho)|$ , then we would expect that any scheme that ignores the past observations for observation control will perform poorly compared to the one that relies on the state of the system to decide whether or not to take a sample in the next time slot. This is verified by the figure: as  $\rho \rightarrow 0$ , we see a significant difference in performances of  $\gamma(a, b)$  and the fractional sampling scheme. The figure also shows that as  $\rho$  becomes large, and begins to dominate the sum  $D(f_1, f_0) + |\log(1 - \rho)|$ , the ADD performance of the fractional sampling scheme approach that of the two-threshold algorithm  $\gamma(a, b)$ .

## 6. 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 occurs. We obtained a two-threshold Bayesian algorithm that is asymptotically optimal, has good trade-off curves, and is easy to design. We derived analytical approximations for the ADD, PFA, and ANO performance of the two-threshold algorithm using which we can design the algorithm by choosing the thresholds. In particular, we showed that, when the constraint on the PFA is moderate to small and that on the ANO is not very small, the two thresholds can be set independent

of each other. We also provided extensive numerical and simulation results that validate our analysis. Our results indicate that our two-threshold algorithm can significantly save on the number of observations taken before the change while maintaining the delay relatively unchanged. A comparison with the naive approach of fractional sampling shows that the two-threshold algorithm is indeed very efficient in using observations to detect the change. Our two-threshold algorithm has many engineering applications in settings 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.

An important problem for future research is to see whether two-threshold policies are optimal in non-Bayesian (e.g., minimax) settings, where we do not have a prior on  $\Gamma$ . In particular, it is of interest to understand how to update the algorithm metric in a non-Bayesian setting when we skip an observation. From an application point of view, one can design a two-threshold algorithm based on the Shiryaev-Roberts or CUSUM approaches (Tartakovsky and Moustakides, 2010) 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 algorithms, as is done for the standard quickest change detection problem (Lai, 1998, Tartakovsky and Veeravalli, 2005). Also, such lower bounds can possibly help in obtaining insights for cases where the observations are not i.i.d. (Lai, 1998, Tartakovsky and Veeravalli, 2005). Other interesting problems in this area include the design of data-efficient optimal algorithms for robust change detection and nonparametric change detection.

**APPENDIX TO SECTION 3.1**

*Proof of Theorem 3.1.* We first show that  $\eta_n$  with  $b = -\infty$ , and  $Z_0$  a random variable, is a slowly changing sequence. Let  $Z_0$  takes value  $z_0$ , then

$$\eta_n = \log \left[ e^{z_0} + \sum_{k=0}^{n-1} \rho(1 - \rho)^k \prod_{i=1}^k \frac{f_0(X_i)}{f_1(X_i)} \right] \xrightarrow[n \rightarrow \infty]{P_1 - a.s.} \log \left[ e^{z_0} + \sum_{k=0}^{\infty} \rho(1 - \rho)^k \prod_{i=1}^k \frac{f_0(X_i)}{f_1(X_i)} \right].$$

Define

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

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  $Z_0 = z_0$ ,

$$\begin{aligned} E[\eta(z_0)] &\leq \log \left[ e^{z_0} + \sum_{k=0}^{\infty} \rho(1 - \rho)^k E_1 \left( \prod_{i=1}^k \frac{f_0(X_i)}{f_1(X_i)} \right) \right] \\ &= \log \left[ e^{z_0} + \sum_{k=0}^{\infty} \rho(1 - \rho)^k \right] = \log(e^{z_0} + 1). \end{aligned}$$

Thus

$$\eta_n \xrightarrow[b=-\infty]{P_1\text{-a.s.}} \eta(Z_0) = \log(e^{Z_0} + \rho) + \sum_{k=1}^{\infty} \log(1 + e^{-Z_k} \rho). \quad (\text{A.1})$$

This implies that  $\sum_{k=1}^{\infty} \log(1 + e^{-Z_k} \rho)$  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  $\Gamma$ . We first show that starting with the above  $Z_0$ , the sequence  $\eta_n$  generated in (3.4) is slowly changing.

To verify the first condition (3.5), from (3.4) note that

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

Since  $Z_k \rightarrow \infty$  a.s.,  $\log(1 + e^{-Z_k} \rho) \rightarrow 0$ , also,  $\mathbb{I}_{\{Z_k < b\}} \rightarrow 0$  a.s. Thus, both the sequences  $\{\log(1 + e^{-Z_k} \rho)\}$  and  $\{(|\log L(X_k)|) \mathbb{I}_{\{Z_k < b\}}\}$  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 (3.6), we first obtain a bound on  $|\eta_{n+k} - \eta_n|$ .

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

Thus,

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

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

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

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

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

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

This is because  $d_n^1$  behaves like a partial sum of a series of type in (A.1). Since the series in (A.1) 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; that is,  $\mathbb{P}(d_n^1 > \epsilon | E) \rightarrow 0$ . Select  $n = n_1^*$  such that  $\forall n > n_1^*$ ,  $\mathbb{P}(d_n^1 > \epsilon | E) < \epsilon$ .

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,

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

Since  $L_Z$  is almost surely finite,  $\mathbb{P}_{E'}(L_Z \geq n) \rightarrow 0$  as  $n \rightarrow \infty$ . Thus, we can select  $n = n_2^*$  such that  $\forall n > n_2^*$ ,  $\mathbb{P}_{E'}(L_Z \geq n) < \epsilon/2$ . 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 (A.1), with  $Z_0$  replaced by  $Z_{L_Z}$ . Since the series in (A.1) 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; that is,  $\mathbb{P}_{E'}(d_n^1 > \epsilon | L_Z < n) \rightarrow 0$ . Select,  $n = n_3^*$  such that  $\forall n > n_3^*$ ,  $\mathbb{P}(d_n^1 > \epsilon | L_Z < n) < \epsilon/2$ . Then  $n^* = \max\{n_1^*, n_2^*, n_3^*\}$ , is the desired  $n^*$  and pick any  $\delta > 0$ . Then for  $n > n^*$ ,

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

Since the sequence  $\eta_n$  is slowly changing, according to Siegmund (1985), 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:

$$\lim_{a \rightarrow \infty} \mathbb{P}_\ell [Z_\tau - a \leq x | \tau \geq l] = R(x),$$

where  $\mathbb{P}_\ell$  is the probability measure with change happening at  $l$ . Now,

$$\mathbb{P}[Z_\tau - a \leq x | \tau \geq \Gamma] = \sum_{l=1}^{\infty} \mathbb{P}_l [Z_\tau - a \leq x | \tau \geq l] \mathbb{P}(\Gamma = l | \tau \geq \Gamma),$$

and

$$\lim_{a \rightarrow \infty} \mathbb{P}_l [Z_\tau - a \leq x | \tau \geq l] \mathbb{P}(\Gamma = l | \tau \geq \Gamma) = R(x)\mathbb{P}(\Gamma = l) \leq 1.$$

Hence, we have the desired result by dominated convergence theorem.  $\square$

### APPENDIX TO SECTION 3.2

*Proof of Lemma 3.1.* 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:

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

Also,

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

Thus,

$$\text{PFA} = \text{E}[e^{-Z_\tau}][1 + o(1)] = e^{-a} \text{E}[e^{-(Z_\tau - a)}][1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

Now note that

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

Since  $\text{P}(\tau < \Gamma) = \text{E}[1 - p_\tau] \leq 1 - A \leq e^{-a}$ , we can write

$$\text{PFA} = e^{-a} \text{E}[e^{-(Z_\tau - a)} \mid \tau \geq \Gamma][1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

This proves the lemma. □

### APPENDIX TO SECTION 3.3

*Proof of Theorem 3.3.* Each time  $Z_k$  crosses  $b$  from below, it satisfies

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

Define  $b_1 \triangleq b + \log \frac{1}{1 - \rho} + \log(1 + e^{-b}\rho)$ . Then  $b_1 \rightarrow b$  as  $\rho \rightarrow 0$ . Also, each time  $Z_k$  crosses  $b$  from below, the average time for  $Z_k$  to reach  $a$  can be decreased by setting  $Z_k = b_1$  and increased by setting  $Z_k = b$ . Let  $N(N_1)$  be one plus the number of times  $Z_k$  goes below  $b$  before it crosses  $a$ , when it is reset to  $b$  ( $b_1$ ), each time it crosses  $b$  from below.



Now recall the three disjoint events:

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

We can write

$$\begin{aligned}E[\tau - \Gamma | \tau \geq \Gamma] &= E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] + E[\tau - \Gamma; \mathcal{B} | \tau \geq \Gamma] \\ &\quad + E[\tau - \Gamma; \mathcal{C} | \tau \geq \Gamma].\end{aligned}\tag{A.2}$$

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  $a$  after multiple up-crossings of the threshold  $b$ . Then,

$$\begin{aligned}E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] &\leq E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma]P(\mathcal{A} | \tau \geq \Gamma) \\ &\quad + E_1 \left[ \sum_{k=1}^N \Lambda_k(b) \right] P(\mathcal{A} | \tau \geq \Gamma).\end{aligned}\tag{A.3}$$

This upper bound was obtained by resetting  $Z_k$  to  $b$  each time it crosses  $b$  from below. Similarly, we can get a lower bound by setting  $Z_k = b_1$  each time  $Z_k$  crosses  $b$  from below. Thus,

$$\begin{aligned}E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] &\geq E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma]P(\mathcal{A} | \tau \geq \Gamma) \\ &\quad + E_1 \left[ \sum_{k=1}^{N_1} \Lambda_k(b_1) \right] P(\mathcal{A} | \tau \geq \Gamma).\end{aligned}$$

Now by Wald's lemma (Siegmund, 1985),

$$\begin{aligned}E_1 \left[ \sum_{k=1}^{N_1} \Lambda_k(b_1) \right] &= E_1[N_1]E_1[\Lambda(b_1)] \\ \xrightarrow{\rho \rightarrow 0} E_1[N]E_1[\Lambda(b)] &= E_1 \left[ \sum_{k=1}^N \Lambda_k(b) \right] = \text{ADD}^\rho.\end{aligned}$$

Thus,

$$\begin{aligned}E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] &= [E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma]P(\mathcal{A} | \tau \geq \Gamma) \\ &\quad + \text{ADD}^\rho P(\mathcal{A} | \tau \geq \Gamma)][1 + o(1)] \quad \text{as } \rho \rightarrow 0.\end{aligned}$$

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

$$\begin{aligned}E[\tau - \Gamma; \mathcal{B} | \tau \geq \Gamma] &= [E[\Lambda(Z_\Gamma) | \mathcal{B}, \tau \geq \Gamma]P(\mathcal{B} | \tau \geq \Gamma) \\ &\quad + \text{ADD}^\rho P(\mathcal{B} | \tau \geq \Gamma)][1 + o(1)] \quad \text{as } \rho \rightarrow 0.\end{aligned}$$

Similarly, considering event  $\mathcal{C}$ , we get

$$E[\tau - \Gamma; \mathcal{C} | \tau \geq \Gamma] = [E[\Lambda(Z_\Gamma) | \mathcal{C}, \tau \geq \Gamma]P(\mathcal{C} | \tau \geq \Gamma)][1 + o(1)] \quad \text{as } \rho \rightarrow 0.$$

Substituting in (A.2) we get the desired result (3.15).  $\square$

*Proof of Lemma 3.2.* Based on  $\Psi$ , 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:

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

These two stopping times stochastically upper and lower bound the Shiryaev stopping time  $v_b$  defined in (3.17); that is,

$$E_1[\tilde{v}_b] \leq E_1[v_b] \leq E_1[\hat{v}_b]. \quad (\text{A.4})$$

Recall from (3.16) that

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

Using Wald's lemma (Siegmund, 1985), we can get the following expressions:

$$E_1[\tilde{v}_b] = \frac{E_1[\lambda]}{P_1(Z_\lambda > a)}, \quad E_1[\hat{v}_b] = \frac{E_1[\lambda] + E_1[v(Z_\lambda, b); \{Z_\lambda < b\}]}{P_1(Z_\lambda > a)}. \quad (\text{A.5})$$

Multiplying and dividing  $\text{ADD}^s$  by  $E_1[\lambda]$  we get

$$\begin{aligned} \text{ADD}^s &= \frac{E_1[\lambda] + E_1[t(Z_\lambda, b); \{Z_\lambda < b\}]}{E_1[\lambda]} \frac{E_1[\lambda]}{P_1(Z_\lambda > a)} \\ &= E_1[\tilde{v}_b] \frac{E_1[\lambda] + E_1[t(Z_\lambda, b); \{Z_\lambda < b\}]}{E_1[\lambda]} \\ &= E_1[\tilde{v}_b][1 + o(1)] \quad \text{as } a \rightarrow \infty. \end{aligned}$$

The last equality follows because  $E_1[\lambda] \rightarrow \infty$  as  $a \rightarrow \infty$ , while  $E_1[t(Z_\lambda, b); \{Z_\lambda < b\}]$  is not a function of  $a$ . Similarly, multiplying and dividing  $\text{ADD}^s$  by  $E_1[\lambda] + E_1[v(Z_\lambda, b); \{Z_\lambda < b\}]$  we get

$$\text{ADD}^s = E_1[\hat{v}_b][1 + o(1)] \quad \text{as } a \rightarrow \infty.$$

Using these two expressions for  $\text{ADD}^s$  and the relationship that  $E_1[\tilde{v}_b] \leq E_1[v_b] \leq E_1[\hat{v}_b]$ , we have

$$\text{ADD}^s = E_1[v_b][1 + o(1)] \quad \text{as } a \rightarrow \infty. \quad \square$$

*Proof of Theorem 3.4.* Consider the upper bound (A.3):

$$E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] \leq E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma]P(\mathcal{A} | \tau \geq \Gamma) + \text{ADD}^s P(\mathcal{A} | \tau \geq \Gamma).$$

Similarly, the upper bounds corresponding to the other two events  $\mathcal{B}$  and  $\mathcal{C}$  are

$$E[\tau - \Gamma; \mathcal{B} | \tau \geq \Gamma] \leq E[\Lambda(Z_\Gamma) | \mathcal{B}, \tau \geq \Gamma]P(\mathcal{B} | \tau \geq \Gamma) + \text{ADD}^s P(\mathcal{B} | \tau \geq \Gamma)$$

and

$$\begin{aligned} E[\tau - \Gamma; \mathcal{C} | \tau \geq \Gamma] &= E[\Lambda(Z_\Gamma) | \mathcal{C}, \tau \geq \Gamma]P(\mathcal{C} | \tau \geq \Gamma) \\ &\leq E_1[\Lambda(b) | Z_{\Lambda(b)} > a]P(\mathcal{C} | \tau \geq \Gamma) \\ &\leq \text{ADD}^s P(\mathcal{C} | \tau \geq \Gamma). \end{aligned}$$

Substituting in (A.2) we get

$$\begin{aligned} E[\tau - \Gamma | \tau \geq \Gamma] &= E[\tau - \Gamma; \mathcal{A} | \tau \geq \Gamma] + E[\tau - \Gamma; \mathcal{B} | \tau \geq \Gamma] + E[\tau - \Gamma; \mathcal{C} | \tau \geq \Gamma]. \\ &\leq \text{ADD}^s + E[t(Z_\Gamma, b) | \mathcal{A}, \tau \geq \Gamma] + E[\Lambda(Z_\Gamma) | \mathcal{B}, \tau \geq \Gamma]. \end{aligned} \quad (\text{A.6})$$

In Eq. (A.6), we observe that except for  $\text{ADD}^s$ , other terms are not a function of threshold  $a$ . Thus, we have

$$E[\tau - \Gamma | \tau \geq \Gamma] \leq \text{ADD}^s[1 + o(1)] \quad \text{as } a \rightarrow \infty. \quad \square$$

## APPENDIX TO SECTION 3.4

*Proof of Lemma 3.3.* Each time  $Z_k$  crosses  $b$  from below, it satisfies:

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

Define  $b_1 \triangleq b + \log \frac{1}{1 - \rho} + \log(1 + e^{-b\rho})$ . Then  $b_1 \rightarrow b$  as  $\rho \rightarrow 0$ . Also, each time  $Z_k$  crosses  $b$  from below, the average number of observations used before  $\Gamma$  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 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:

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

Thus,  $\tilde{\Lambda}^x$  is the time for  $Z_k$  to start at  $Z_0 = x$  with  $a = \infty$  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  $\tilde{\Lambda}^x$ ; that is, a fraction of  $\tilde{\Lambda}^x$  when  $Z_k \geq b$ . If  $\{\tilde{\Lambda}_k^b\}$  and  $\{\tilde{\Lambda}_k^{b_1}\}$  are 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

$$\begin{aligned} E_\infty[L^b]E_\infty[\tilde{\Lambda}^b \delta^b] &= E_\infty \left[ \sum_{k=1}^{L^b} \tilde{\Lambda}_k^b \delta_k^b \right] \leq E \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), a = \infty \right] \\ &\leq E_\infty \left[ \sum_{k=1}^{L^{b_1}} \tilde{\Lambda}_k^{b_1} \delta_k^{b_1} \right] = E_\infty[L^{b_1}]E_\infty[\tilde{\Lambda}^{b_1} \delta^{b_1}]. \end{aligned}$$

Here the equalities follow from Wald's lemma (Siegmund, 1985).

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

$$\frac{P_\infty[\Gamma \leq \tilde{\Lambda}^{b_1}]}{P_\infty[\Gamma \leq \tilde{\Lambda}^b]} \rightarrow 1 \quad \text{as } \rho \rightarrow 0.$$

Further, for  $x = b_1$  or  $x = b$ , define  $\hat{\lambda}(x)$  based on (3.23) as

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

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

$$\begin{aligned} E_\infty[\tilde{\Lambda}^x \delta^x] &= E_\infty[\tilde{\Lambda}^x \delta^x \mid \Gamma \leq \tilde{\Lambda}^x \delta^x] P_\infty[\Gamma \leq \tilde{\Lambda}^x \delta^x] + E_\infty[\tilde{\Lambda}^x \delta^x \mid \Gamma > \tilde{\Lambda}^x \delta^x] P_\infty[\Gamma > \tilde{\Lambda}^x \delta^x] \\ &\rightarrow E_\infty[\hat{\lambda}(x)] \quad \text{as } \rho \rightarrow 0. \end{aligned}$$

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

$$E_\infty[\hat{\lambda}(b_1)] \rightarrow E_\infty[\hat{\lambda}(b)] = E_\infty[\hat{\lambda}].$$

Thus,

$$E \left[ \sum_{k=t(b)}^{\Gamma-1} S_k \mid \Gamma > t(b), a = \infty \right] = \frac{E_\infty[\hat{\lambda}]}{P_\infty[\Gamma \leq \tilde{\Lambda}^b]} [1 + o(1)] \quad \text{as } \rho \rightarrow 0. \quad \square$$

*Proof of Lemma 3.4.* Since  $P\{\tau \geq \Gamma\} \rightarrow 1$  as  $a \rightarrow \infty$ ,

$$\begin{aligned} P[\Gamma > t(b) \mid \tau \geq \Gamma] &= P[\Gamma > t(b)] + o(1) \quad \text{as } a \rightarrow \infty \\ &= \frac{1}{1 + z_0} (1 - \rho)^{t(b)} + o(1) \quad \text{as } a \rightarrow \infty. \end{aligned}$$

From (4.3) in Lemma 4.1, with  $y = b$  and  $x = z_0$ , we have

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

From this, it is easy to show that

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

By substituting this in the expression for  $P(\Gamma > t(b) | \tau \geq \Gamma)$  we get the desired result.  $\square$

*Proof of Theorem 3.6.* Using Theorem 3.4 we write  $\text{ANO}_1$  as

$$\begin{aligned} \text{ANO}_1 &= E[\tau - \Gamma | \tau \geq \Gamma] \left( 1 - \frac{T_b - 1}{E[\tau - \Gamma | \tau \geq \Gamma]} \right) \\ &= E_1[v_b] \left( 1 - \frac{T_b - 1}{E[\tau - \Gamma | \tau \geq \Gamma]} \right) [1 + o(1)] \text{ as } a \rightarrow \infty. \end{aligned}$$

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

Recall that  $\mathcal{A}$  and  $\mathcal{B}$  are the events under which excursions below  $b$  are possible. The passage to  $a$  is through multiple cycles below  $b$ , and the time spent 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

$$T_b - 1 \leq T_b \leq P_1(\mathcal{A})t(-\infty, b)E[N_{\mathcal{A}}] + P_1(\mathcal{B})t(-\infty, b)E[N_{\mathcal{B}}].$$

The averages  $E[N_{\mathcal{A}}]$  and  $E[N_{\mathcal{B}}]$  can be written as a series of probabilities, where each term corresponds to the event that  $Z_k$  goes below  $b$ , and not above  $a$ , 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,  $E[N_{\mathcal{A}}] \leq E[M]$  and  $E[N_{\mathcal{B}}] \leq E[M]$ . This gives a bound on  $T_b - 1$

$$T_b - 1 \leq t(-\infty, b)E[M].$$

By using (A.4) we get as  $a \rightarrow \infty$ ,

$$\frac{T_b - 1}{E[\tau - \Gamma | \tau \geq \Gamma]} \leq \frac{t(-\infty, b)E[M]}{E_1[v_b]} [1 + o(1)] \leq \frac{t(-\infty, b)E[M]}{E_1[\tilde{v}_b]} [1 + o(1)].$$

From (A.5) we know that  $E_1[\tilde{v}_b] = E_1[\lambda]E[M]$ . Thus, the upper bound on  $\frac{T_b - 1}{E[\tau - \Gamma | \tau \geq \Gamma]}$  goes to 0 as  $a \rightarrow \infty$ . This proves the theorem.  $\square$

## APPENDIX TO SECTION 4.2

*Proof of Lemma 4.1.* First note that by definition (3.13),  $Z_{t(x,y)} > y \geq Z_{t(x,y)-1}$ . Also, from (3.2)

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

Thus,

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

and equivalently,

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

Further, the recursion (3.2) 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  $t(x, y)$ :

$$\frac{\log(1 + e^y - \rho) - \log(1 + e^x)}{|\log(1 - \rho)|} \leq t(x, y) \leq \frac{\log\left(1 + e^y \frac{(1+e^{-y}\rho)}{(1-\rho)} - \rho\right) - \log(1 + e^x)}{|\log(1 - \rho)|}. \quad (\text{A.7})$$

By keeping  $x, y$  fixed and taking  $\rho \rightarrow 0$  we get (4.3).  $\square$

## ACKNOWLEDGMENTS

We thank the Associate Editor and the referees for their constructive suggestions and comments. We would also like to thank Dr. K. Premkumar for useful discussions. This research was 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, and by the U.S. Defense Threat Reduction Agency through subcontract 147755 at the University of Illinois from prime award HDTRA1-10-1-0086.

## REFERENCES

- Assaf, D. (1988). A Dynamic Sampling Approach for Detecting a Change in Distribution, *Annals of Statistics* 16: 236–253.
- Assaf, D., Pollak, M., Ritov, Y., and Yakir, B. (1993). Detecting a Change of a Normal Mean by Dynamic Sampling with a Probability Bound on a False Alarm, *Annals of Statistics* 21: 1155–1165.

- Banerjee, T., Sharma, V., Kavitha, V., and Jayaprakasam, A. (2011). Generalized Analysis of a Distributed Energy Efficient Algorithm for Change Detection, *IEEE Transactions on Wireless Communication* 10: 91–101.
- Bertsekas, D. P. (2007). *Dynamic Programming and Optimal Control*, Vols. 1 and 2, third edition, Nashua: Athena Scientific.
- Girshick, M. A. and Rubin, H. (1952). A Bayes Approach to a Quality Control Model, *Annals of Mathematical Statistics* 23: 114–125.
- Lai, T. L. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917–2929.
- Mainwaring, A., Culler, D., Polastre, J., Szewczyk, R. and Anderson, J. (2002). Wireless Sensor Networks for Habitat Monitoring, in *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA '02)*, September 28, pp. 88–97, Atlanta, GA.
- Makis, V. (2008). Multivariate Bayesian Control Chart, *Operations Research* 56: 487–496.
- Mei, Y. (2005). Information Bounds and Quickest Change Detection in Decentralized Decision Systems, *IEEE Transactions on Information Theory* 51: 2669–2681.
- Premkumar, K. and Kumar, A. (2008). Optimal Sleep/Wake Scheduling for Quickest Intrusion Detection Using Sensor Networks, in *Proceedings of 27th IEEE Conference on Computer Communications (INFOCOM 2008)*, April 15–17, Phoenix, AZ.
- Rice, J. A., Mechitov, K., Sim, S., Nagayama, T., Jang, S., Kim, R., Spencer, B. F., Agha G., and Fujino, Y. (2010). Flexible Smart Sensor Framework for Autonomous Structural Health Monitoring, *Smart Structures and Systems* 6: 423–438.
- Shiryayev, A. N. (1963). On Optimal Methods in Quickest Detection Problems, *Theory of Probability and Its Applications* 8: 22–46.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer.
- Stoumbos, Z. G. and Reynolds, M. R., Jr. (2005). Economic Statistical Design of Adaptive Control Schemes for Monitoring the Mean and Variance: An Application to Analyzers, *Nonlinear Analysis: Real World Applications* 6: 817–844.
- Stoumbos, Z. G., Reynolds, M. R., Jr., Ryan T. P., and Woodall, W. H. (2000). The State of Statistical Process Control as We Proceed into the 21st Century, *Journal of American Statistical Association* 95: 992–998.
- Tagaras, G. (1998). A Survey of Recent Developments in the Design of Adaptive Control Charts, *Journal of Quality Technology* 30: 212–231.
- Tartakovsky, A. G. and Moustakides, G. V. (2010). State-of-the-Art in Bayesian Changepoint Detection, *Sequential Analysis* 29: 125–145.
- Tartakovsky, A. G. and Veeravalli, V. V. (2005). General Asymptotic Bayesian Theory of Quickest Change Detection, *Theory of Probability and Its Applications* 49: 458–497.
- Tartakovsky, A. G. and Veeravalli, V. V. (2008). Asymptotically Optimal Quickest Change Detection in Distributed Sensor Systems, *Sequential Analysis*, 27: 441–475.
- Veeravalli, V. V. (2001). Decentralized Quickest Change Detection, *IEEE Transactions on Information Theory* 47: 1657–1665.
- Woodroffe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Philadelphia: SIAM.
- Yakir, B. (1996). Dynamic Sampling Policy for Detecting a Change in Distribution, with a Probability Bound on False Alarm, *Annals of Statistics* 24: 2199–2214.
- Zacharias, L. and Sundaresan, R. (2007). Decentralized Sequential Change Detection using Physical Layer Fusion, in *Proceedings of 2007 IEEE International Symposium on Information Theory (ISIT)*, June 24–29, pp. 4999–5008, Nice, France.