Asymptotic Optimality of Change-Point Detection Schemes in General Continuous-Time Models

M. Baron
Department of Mathematical Sciences, University of Texas at Dallas,
Richardson, Texas, USA
IBM Research Division, Yorktown Heights, New York, USA

A. G. Tartakovsky
Department of Mathematics, University of Southern California,
Los Angeles, California, USA

Abstract: In the early 1960s, Shiryaev obtained the structure of Bayesian stopping rules for detecting abrupt changes in independent and identically distributed sequences as well as in a constant drift of the Brownian motion. Since then, methodology of optimal change-point detection concentrates on the search of stopping rules that achieve the best balance of the mean detection delay and the rate of false alarms or minimize the mean delay under a fixed false alarm probability. In this respect, analysis of the performance of the Shiryaev procedure has been an open problem. Recently, Tartakovsky and Veeravalli (2005) investigated asymptotic performance of the Shiryaev Bayesian change detection procedure, the Page procedure, and the Shiryaev-Roberts procedure when the false alarm probability goes to zero for general discrete-time models. In this paper, we investigate the asymptotic performance of Shiryaev and Shiryaev-Roberts procedures for general continuous-time stochastic models for small false alarm probability and small cost of detection delay. We show that the Shiryaev procedure has asymptotic optimality properties under mild conditions, while the Shiryaev-Roberts procedure may or may not be asymptotically optimal depending on the type of the prior distribution. The presented asymptotic Bayesian detection theory substantially generalizes previous work in the field of change-point detection for continuous-time processes.

Key Words: Asymptotic theory; Change-point detection; Continuous-time stochastic models; Optimal stopping; Sequential detection.

Subject Classifications: 62L15; 60G40; 62F12; 62F15.

1. INTRODUCTION

The problem of detecting abrupt changes in stochastic systems arises across various branches of science and engineering, including such important applications as biomedical signal and image processing, quality control engineering, financial markets, link failure detection in communication, intrusion detection in computer networks and security systems, chemical or biological warfare agent detection systems (as a protection tool against terrorist attacks), detection of the onset of an epidemic, failure detection in manufacturing systems

Address correspondence to A. G. Tartakovsky, Department of Mathematics, University of Southern California, KAP-108, Los Angeles, CA 90089-2532, USA. E-mail: tartakov@math.usc.edu

and large machines, and target detection in surveillance systems (see, e.g., Baron, 2002; Basseville and Nikiforov, 1993; Blažek et al., 2001; Kent, 2000; MacNeill and Mao, 1993; Tartakovsky, 1991; Tartakovsky and Veeravalli, 2002, 2004; Tartakovsky et al., 2006; Wang et al., 2002; Willsky, 1976). In all these applications sensors monitoring the environment take observations that undergo a change in distribution in response to a change in the environment. The change occurs at an unknown instant, and the practitioners' goal is to detect it as quickly as possible while avoiding frequent false alarms.

Therefore, the desired quickest (sequential) change detection procedures usually optimize the tradeoff between a measure of detection delay (speed of detection) and a measure of the frequency of false detections (false alarm rate). There are two standard formulations for this optimization problem—minimax (see Lorden, 1971; Pollak, 1985) and Bayesian (see Shiryaev, 1961, 1963, 1978). In the Bayesian formulation a change point is usually assumed to have a geometric prior distribution for discrete-time models or an exponential prior distribution for continuous-time models, and the goal is to minimize the expected delay subject to an upper bound on false alarm probability or the average risk represented usually by a weighted sum of the average detection delay and the false alarm probability.

Furthermore, in continuous time the Bayesian solution is only available for the very limited set of models—for detecting a change in the constant drift of a Brownian motion (Shiryaev, 1963, 1978) and in the constant intensity of a Poisson process (Peskir and Shiryaev, 2002). Also, its practical implementation is computationally rather difficult as it requires computation of the payoff function as a solution of a nonhomogeneous integral equation (Shiryaev, 1978; Baron, 2001). Here we show that for a wide class of prior distributions and for a wide class of statistical models that are not restricted to a conventional i.i.d. (independent and identically distributed) assumption, the stopping rule (2.13) is asymptotically equivalent to the Bayes rule. That is, applying (2.13) in change-point detection, a practitioner is guaranteed to have a Bayes risk that is equal, up to the higher-order and asymptotically negligible terms, to the overall minimum Bayes risk, as the probability of a false alarm or the cost of detection delay converge to 0.

The stopping rule (2.13) has the same general form as Shiryaev's Bayes rule under the exponential prior distribution (for aforementioned Brownian motion and Poisson models). Thus, we will refer to it as the Shiryaev procedure. The study of asymptotic optimality properties under the fixed and small probability of a false detection is important not only for theory but also for a variety of applied problems. Analogous to the given probability of type I error, this property is desired in all practical applications where frequent false alarms cannot be tolerated.

The optimality results are obtained through the study of asymptotic behavior of the Shiryaev change detection procedure, which is Bayesian, and of the Shiryaev-Roberts procedure, which is not Bayesian.

A large number of publications relate to the *design* of reasonable detection procedures for detecting changes in stochastic systems that are driven by more or less general models, such as hidden Markov models and autoregressive models. However, until recently the detection *theory* has been limited to i.i.d. models (in pre-change and post-change modes with different common distributions). There is only a handful of works where the asymptotic study goes beyond the i.i.d. case (see, e.g., Bansal and Papantoni-Kazakos, 1986; Beibel, 2000; Davis et al., 1995; Fuh, 2003, 2004; Lai, 1998; Moustakides, 2004; Tartakovsky, 1995, 1998c; Yakir, 1994).

For i.i.d. data models, asymptotic performance of various change-point detection procedures in a minimax context with the constraint on the mean time to false alarm is well understood (see, e.g., Basseville and Nikiforov, 1993; Beibel, 1996; Moustakides, 1986; Pollak, 1987; Ritov, 1990; Shiryaev, 1996; Siegmund, 1985; Srivastava and Wu, 1993; Tartakovsky, 1991, 1994; Yakir, 1997). Asymptotically pointwise optimal change-point detection rules in the sense of Bickel and Yahav (1967) (see also Ghosh, Mukhopadhyay, and Sen, 1997, sect. 5.4), as the rate of false alarms tends to zero, have been derived by Baron (2002). Asymptotic optimality of the cumulative sum (CUSUM) procedure with an increasing threshold when the constraint is imposed on the false alarm probability has been proven by Borovkov (1998).

Generalizations for continuous-time Brownian motion models with time-dependent coefficients for a minimax problem were given by Tartakovsky (1995) and in a Bayesian problem by Beibel (2000). However, it was Lai (1998) who first developed general asymptotic detection theory for discrete-time models. In particular, he showed that the CUSUM test is asymptotically optimal under certain quite general conditions, not only in the minimax but also in the Bayesian framework. In the minimax setting, similar asymptotic results for the CUSUM and the Shiryaev-Roberts detection procedures were reported in Tartakovsky (1998). Recently, Fuh (2003, 2004) proved asymptotic optimality of the CUSUM and Shiryaev-Roberts-Pollak procedures for hidden Markov models.

At the same time, such asymptotic results have not been obtained for the Shiryaev procedure even for the i.i.d. data models. Recently, Tartakovsky and Veeravalli (2005) investigated this problem for non-i.i.d. discrete-time models that generalize previous results far beyond the restrictive i.i.d. assumption. In particular, asymptotic optimality of the Shiryaev procedure in terms of the average detection delay versus the average false alarm probability has been proven in that paper. The goal of the present paper is to address this important open problem for general, continuous-time stochastic models.

The paper is organized as follows. In Section 2, we formulate the problem and describe the Shiryaev detection procedure that will be analyzed in subsequent sections. Main results are presented in Section 3. These results show that, under mild conditions, the Shiryaev detection procedure is asymptotically optimal when the false alarm probability and/or the cost of the detection delay tend to vanish. We show that it is asymptotically optimal not only with respect to the average detection delay, but also uniformly asymptotically optimal in the sense of minimizing the conditional expected delay for every change point. Moreover, we show that under certain conditions the Shiryaev procedure minimizes higher moments of the detection delay. In Section 4, we study the behavior of the Shiryaev detection procedure for the processes with i.i.d. increments. In Section 5, we analyze the asymptotic performance of another well-known change detection procedure, the Shiryaev-Roberts-Pollak procedure, in the Bayesian framework. Finally, in Section 6, we consider examples that illustrate general results. These examples are motivated by some of the application areas listed in the beginning of this section, e.g. by target detection and intrusion detection problems. Proofs of main results are given in Appendix.

The asymptotic theory that we developed relies upon the strong law of large numbers for the log-likelihood ratio process and the rates of convergence in the strong law similar to those used by Hsu and Robbins (1947) and later by Baum and Katz (1965), Strassen (1967), and Lai (1976) that lead to the so-called complete and $r-$quick convergence. Similar approach has been previously used by Lai (1981), Tartakovsky (1998a), and Dragalin et al. (1999) to establish asymptotic optimality of sequential hypothesis tests.

2. THE SHIRYAEV DETECTION PROCEDURE AND A CLASS OF STOPPING RULES

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbf{P})$, $t \geqslant 0$, be a stochastic basis with standard assumptions about monotonicity and right-continuity of the σ-algebras $\mathcal{F}_t$. The sub-σ-algebra $\mathcal{F}_t = \mathcal{F}_t^X = \sigma(X^t)$ of $\mathcal{F}$ is assumed to be generated by the process $X^t = \{X_v, 0 \leqslant v \leqslant t\}$ observed up to time t, which is defined on the space $(\Omega, \mathcal{F})$. Consider the following change-point detection problem. Let $\mathbf{F}_0$ and $\mathbf{F}_1$ be completely known probability measures, which are mutually locally absolutely continuous. In a normal mode, the observed data X_t follow the measure $\mathbf{F}_0$. At an unknown point in time λ, $\lambda \geqslant 0$, something happens and X_t follow the measure $\mathbf{F}_1$. The goal is to detect the change as soon as possible after it occurs, subject to constraints on the false alarm rate.

We will be interested in a Bayesian setting where the change point λ is assumed to be random with prior probability distribution $\pi_t = \mathbf{P}(\lambda \leqslant t)$, $t \geqslant 0$.

In mathematical terms, a sequential detection procedure is identified with a stopping time τ for the observed process $\{X_t\}_{t \geqslant 0}$, i.e. τ is an extended continuous-valued random variable, such that the event

$\{\tau \leqslant t\}$ belongs to the sigma-algebra $\mathcal{F}_t^X$. A false alarm is raised whenever $\tau < \lambda$. A good detection procedure should guarantee a "stochastically small" detection lag $(\tau - \lambda)$ provided that there is no false alarm (i.e. $\tau \geqslant \lambda$), while the rate of false alarms should be kept at a given, usually low level.

Let $\mathbf{P}_u$ and $\mathbf{E}_u$ denote the probability measure and the corresponding expectation when the change occurs at time $\lambda = u$. In what follows, $\mathbf{P}^\pi$ stands for the "average" probability measure which, for every measurable set $\mathcal{A}$, is defined as $\mathbf{P}^\pi(\mathcal{A}) = \int_0^\infty \mathbf{P}_u(\mathcal{A})\, d\pi_u$ and $\mathbf{E}^\pi$ denotes the expectation with respect to $\mathbf{P}^\pi$.

In the Bayesian setting, a reasonable measure of the detection lag is the average detection delay (ADD)

$$\mathrm{ADD}(\tau) = \mathbf{E}^\pi(\tau - \lambda | \tau \geqslant \lambda), \tag{2.1}$$

while the false alarm rate is usually measured by the (average) probability of false alarm (PFA), which is defined as follows

$$\mathrm{P_{FA}}(\tau) = \mathbf{P}^\pi(\tau < \lambda) = \int_0^\infty \mathbf{P}_u(\tau < u)\, d\pi_u. \tag{2.2}$$

Alternatively, the PFA can be defined as $\mathbf{P}_\infty(\tau < \infty)$. However, in the rest of the paper we will consider the "classical" average PFA given in (2.2). See Remark 2 in Section 7 for a brief discussion of possible modifications when the PFA is formulated in terms of the probability $\mathbf{P}_\infty(\tau < \infty)$.

Define the class of change-point detection procedures $\mathbf{\Delta}(\alpha) = \{\tau : \mathrm{P_{FA}}(\tau) \leqslant \alpha\}$ for which the false alarm probability does not exceed the predefined number α. In a variational formulation of the optimization problem, an optimal Bayesian detection procedure is a procedure for which ADD is minimized, while the probability $\mathrm{P_{FA}}(\tau)$ should be maintained at a given level α, i.e. it is described by the stopping time

$$\nu = \arg \inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathrm{ADD}(\tau). \tag{2.3}$$

Since

$$\mathrm{ADD}(\tau) = \mathbf{E}^\pi(\tau - \lambda)^+ / \mathbf{P}^\pi(\tau \geqslant \lambda),$$

ν minimizes also $\mathbf{E}^\pi(\tau - \lambda)^+$ whenever $\mathbf{P}^\pi(\nu < \lambda) = \alpha$. Hereafter $x^+ = \max(0, x)$.

In addition to the above variational Bayes formulation, we will also be interested in a "purely" Bayes problem with the loss function, which for the fixed $\lambda = u$ and $\tau = t$, is given by

$$L(t, u) = c \cdot (t - u) \mathbb{1}_{\{t \geqslant u\}} + \mathbb{1}_{\{t < u\}}, \tag{2.4}$$

where $\mathbb{1}_{\{\mathcal{A}\}}$ is an indicator of a set $\mathcal{A}$ and c is a cost associated with the detection delay. In this case, the problem is to find a detection procedure ν which minimizes the average loss (risk),

$$\nu = \arg \inf_\tau \left\{ \mathbf{P}^\pi(\tau < \lambda) + c \mathbf{E}^\pi(\tau - \lambda)^+ \right\}, \tag{2.5}$$

where infimum is taken over all Markov times. A more general loss $L_m(t, u) = c(t - u)^m \mathbb{1}_{\{t \geqslant u\}} + \mathbb{1}_{\{t < u\}}$ for $m > 1$ will also be considered.

In continuous time, the structure of the optimal procedure in Bayesian problems (2.3) and (2.5) is known only for the two particular examples—for detecting a change in the drift of a Wiener process and in the intensity of a Poisson process. To be specific, let $p_t = \mathbf{P}(\lambda \leqslant t | \mathcal{F}_t^X)$ be the posterior probability that the change occurred before time t and let $A \in (0, 1)$. Define the detection procedure that raises an alarm at the first time such that the posterior probability p_t exceeds a threshold A, i.e.

$$\nu(A) = \inf \{t \geqslant 0 : p_t \geqslant A\}. \tag{2.6}$$

Hereafter without special emphasis we use a convention that $\inf\{\varnothing\} = \infty$, i.e. $\nu(A) = \infty$ if no such t exists. Shiryaev(1961, 1963, 1978) and Peskir and Shiryaev (2002) have proved that the procedure (2.6) is optimal in the aforementioned two detection problems if the prior distribution of the change point is exponential,

$$\pi_t = (1 - \pi_0)\left(1 - e^{-\rho t}\right) \mathbb{1}_{\{t>0\}} + \pi_0 \mathbb{1}_{\{t=0\}}, \tag{2.7}$$

where $\rho > 0$ and $0 \leqslant \pi_0 < 1$.

This result can be undoubtedly extended to any problem with i.i.d. increments (Levy processes). However, it is very difficult, if not impossible, to find the optimal threshold in the optimization problem (2.5). In the variational Bayes problem (2.3), the threshold $A = A_\alpha$ should be chosen so that $\mathrm{P}_{\mathrm{FA}}(\nu(A)) = \alpha$. It turns out that except for the cases where the trajectories of the process X_t are continuous it is difficult to find a threshold that exactly matches the given PFA. Also, there are no results related to the ADD evaluation of this optimal procedure except for the case of detecting the change in the constant drift of the Wiener process (see Shiryaev, 1978).

The exact match of the false alarm probability is related to the estimation of the overshoot in the stopping rule (2.6) and is usually problematic in discontinuous cases. However, ignoring the overshoot, a simple upper bound can be easily derived. Indeed, since $\mathbf{P}^\pi\left\{\nu(A) < \lambda\right\} = \mathbf{E}^\pi\left\{1 - p_{\nu(A)}\right\}$ and $1 - p_{\nu(A)} \leqslant 1 - A$ on $\{\nu(A) < \infty\}$, we obtain that

$$\mathrm{P}_{\mathrm{FA}}(\nu(A)) \leqslant 1 - A. \tag{2.8}$$

Therefore, setting $A = A_\alpha = 1 - \alpha$ guarantees the inequality $\mathrm{P}_{\mathrm{FA}}(\nu(A_\alpha)) \leqslant \alpha$.

In the rest of the paper, we consider the following two classes of prior distributions of the point of change:

A1. Prior distributions with exponential right tail:

$$\lim_{t\to\infty} \frac{\log(1-\pi_t)}{t} = -\rho, \quad \rho > 0$$

(e.g., exponential, gamma, logistic distributions, i.e., models with bounded hazard rate);

A2. Asymptotically "flat" (or heavy-tailed) prior distributions:

$$\lim_{t\to\infty} \frac{\log(1-\pi_t)}{t} = 0; \quad \limsup_{t\to\infty} \frac{1-\pi_{kt}}{1-\pi_t} < 1 \text{ for some } k > 1 \tag{2.9}$$

(e.g., Weibull distribution with the shape parameter $\gamma < 1$, Pareto, lognormal distributions, i.e., models with vanishing hazard rate).

In the case A1, we will write $\pi \in \mathcal{E}(\rho)$, and in the case A2, $\pi \in F$. The case A2 can be further generalized by replacing 0 on the right with a negative number $-\rho_\alpha$ that vanishes when $\alpha \to 0$.

Practitioners further narrow the set of considered prior distributions based on the characteristics of their specific applied problem. If occurrence of a change point becomes approximately memoryless in the remote future (although its hazard rate may still vary), the prior will be selected from the class $\mathcal{E}(\rho)$. If the process has memory, and longer periods of no change make a practitioner decide that a change point during the next minute is *a priori* less likely, then $\pi \in F$.

The set of models excluded from our consideration consists of priors with unboundedly increasing hazard rates, for example, the Gompertz distribution. In such models, time intervals containing a change point with the same probability p (given that it has not occurred earlier) decrease to 0. Hence the change is likely to be detected at early stages, and the asymptotic study is impractical. Also, the second part of (2.9) eliminates some very non-typical extremely sparse distributions where increasingly large density is supported on more and more dispersed intervals as $t \to \infty$.

Note that we do not restrict ourselves to the case of absolutely continuous prior distributions. The prior distribution π_t may have positive masses at some values of λ including the case of $\pi_0 = \mathbf{P}(\lambda = 0) > 0$.

For $\pi \in F$, the study of a mean delay in Sections 3.2 and 3.4 will require a finite expectation of the prior distribution,

$$\mathbf{E}\lambda = \int_0^\infty s d\pi_s < \infty, \tag{2.10}$$

which automatically holds for all $\pi \in \mathcal{E}(\rho)$.

Remark 1. The second condition in (2.9) can be further relaxed. In fact, this condition is used in Theorems 2, 3, and 5, but it is not needed at all in Theorems 1, 4, and 7.

By

$$Z_t^u := \log \frac{d\mathbf{P}_u^t}{d\mathbf{P}_\infty^t}(X^t), \quad t \geqslant u, \tag{2.11}$$

we will denote the log-likelihood ratio (LLR) for the hypotheses that the change occurred at the point $\lambda = u$ and at $\lambda = \infty$ (no change at all). Hereafter $\mathbf{P}^t$ stands for a restriction of the probability measure $\mathbf{P}$ to the sigma-algebra $\mathcal{F}_t^X$. For $t \geqslant 0$, define the statistic

$$\Lambda_t = \frac{1}{1-\pi_t}\int_0^t e^{Z_t^s}\, d\pi_s. \tag{2.12}$$

It is easily verified that $\Lambda_t = p_t/(1-p_t)$ for $t \geqslant 0$. Therefore, the optimal stopping rule (2.6) can be written in the following form that is more convenient for asymptotic study,

$$\nu_B = \inf\left\{t \geqslant 0 : \Lambda_t \geqslant B\right\}, \quad \text{where } B = A/(1-A). \tag{2.13}$$

This rule has been proposed and proven to be the Bayes stopping rule by Shiryaev under the geometric prior distribution for i.i.d. discrete time models or under the exponential prior distribution for the Brownian motion (Shiryaev, 1963, 1978). In the sequel, we study its performance for the general situation, always referring to it as the Shiryaev rule.

Using inequality (2.8), we obtain that

$$B_\alpha = (1-\alpha)/\alpha \quad \text{implies} \quad \nu_{B_\alpha} \in \mathbf{\Delta}(\alpha). \tag{2.14}$$

In addition to the Bayesian ADD defined in (2.1), we will also be interested in the behavior of the conditional ADD (CADD) for the fixed change point $\lambda = u$, which is defined by $\mathrm{CADD}_u(\tau) = \mathbf{E}_u(\tau - u | \tau \geqslant u)$, $u \geqslant 0$, as well as higher moments of the detection delay $\mathbf{E}_u\left\{(\tau-u)^m | \tau \geqslant u\right\}$ and $\mathbf{E}^\pi\left\{(\tau-\lambda)^m | \tau \geqslant \lambda\right\}$ for $m > 1$.

In the next section, we study the operating characteristics of the Shiryaev procedure (2.13) for $\alpha \to 0$ and $c \to 0$ in a general non-i.i.d. case. It turns out that under mild conditions this detection procedure with the thresholds $B_\alpha = (1-\alpha)/\alpha$ and $B_c = O(1/c)$ asymptotically minimizes the $\mathrm{ADD}(\tau)$ in the class $\mathbf{\Delta}(\alpha)$ and the average risk $\mathcal{R}_c(\tau) = \mathrm{P}_{\mathrm{FA}}(\tau) + c(1 - \mathrm{P}_{\mathrm{FA}}(\tau))\mathrm{ADD}(\tau)$ in the class of all stopping times, respectively. It is interesting that it also minimizes $\mathrm{CADD}_u(\tau)$ for all $u \geqslant 0$ as well as higher moments of the detection delay. In Section 3.4, these results will be specified for the processes with i.i.d. increments.

3. ASYMPTOTIC BAYESIAN DETECTION THEORY

3.1. Heuristics

It turns out that the asymptotic performance of the Shiryaev detection procedure is different for the prior distributions that have exponential tail, for example for the exponential prior (2.7), and for asymptotically flat prior distributions A2 for which the right tail decays slower than an exponent. We start with a heuristic argument that explains the reason for that.

For simplicity, assume that the data X_t have i.i.d. increments. Let $\mathbf{E}_0 Z_1^0 = I$ denote the Kullback-Leibler information number. By (2.12), the statistic $\log \Lambda_t$ can be represented in the form

$$\log \Lambda_t = |\log(1-\pi_t)| + Z_t^0 + \log \int_{(0,t]} \frac{d\mathbf{P}_s^t}{d\mathbf{P}_0^t}(X^t)\, d\pi_s.$$

Substituting $t = \nu_B$, taking expectation, and ignoring the overshoot, we obtain

$$\log B \approx \mathbf{E}_0 |\log(1-\pi_{\nu_B})| + I \cdot \mathbf{E}_0 \nu_B + \mathbf{E}_0 \log \int_{(0,\nu_B]} \frac{d\mathbf{P}_s^{\nu_B}}{d\mathbf{P}_0^{\nu_B}}(X^{\nu_B})\, d\pi_s.$$

For large B, the last term on the right-hand side can be approximately estimated as

$$\mathbf{E}_0 \log \int_{(0,\nu_B]} \frac{d\mathbf{P}_s^{\nu_B}}{d\mathbf{P}_0^{\nu_B}}(X^{\nu_B})\, d\pi_s \approx \mathbf{E}_0 \log \int_{(0,\infty)} \frac{d\mathbf{P}_s^{\nu_B}}{d\mathbf{P}_0^{\nu_B}}(X^{\nu_B})\, d\pi_s \leqslant \log \int_0^\infty d\pi_s = 0.$$

In the case of the exponential prior distribution (2.7), $\mathbf{E}_0 |\log(1-\pi_{\nu_B})| = \rho \mathbf{E}_0 \nu_B$, and we obtain that for large B

$$\mathbf{E}_0 \nu_B \approx \frac{\log B}{I+\rho}.$$

On the other hand, for asymptotically flat prior distributions $\mathbf{E}_0 |\log(1-\pi_{\nu_B})| = o(\mathbf{E}_0 \nu_B)$, as $B \to \infty$, which yields

$$\mathbf{E}_0 \nu_B \approx \frac{\log B}{I}.$$

Therefore, we expect that the average detection delay

$$\mathrm{ADD}(\nu_B) \sim \frac{\log B}{I_\rho} \quad \text{as } B \to \infty,$$

where $I_\rho = I + \rho$ for the exponential prior distribution (more generally for $\pi \in \mathcal{E}(\rho)$) and $I_\rho = I$ for asymptotically flat prior distributions.

A similar heuristic argument can be used to conjecture that under certain conditions for $m > 0$

$$\mathbf{E}^\pi \left\{ (\nu_B - \lambda)^m | \nu_B \geqslant \lambda \right\} \sim \left(\frac{\log B}{I_\rho} \right)^m \quad \text{as } B \to \infty.$$

"Physical" meaning of this difference is the following. The exponential prior distribution imposes quite rigid constraint on the expected time to the change, which is expressed by the factor $1/\rho$. For small difference between pre-change and post-change distributions, i.e. for small I that characterizes a magnitude of the change, the prior information on the point of change is more important, which is expressed in the fact that the average detection delay is on the order of $\log B/\rho$, since one expects that the change will occur every $1/\rho$ "seconds" (on average). By contrast, for asymptotically flat prior distributions prior information is limited and the situation is *a priori* uncertain. Therefore, observations have a complete control. In particular, when $I \to 0$, the average detection delay goes to infinity. Mathematical details are given below.

3.2. Asymptotic Performance in the Class $\mathbf{\Delta}(\alpha)$

We begin with deriving asymptotic lower bounds for moments of the detection delay (in particular, for ADD and CADD). To this end, we will require the following condition.

Let q be a positive finite number and suppose that

$$\mathbf{P}_u\left\{\frac{1}{M}\sup_{0\leqslant t\leqslant M} Z^u_{u+t}\geqslant(1+\varepsilon)q\right\}\xrightarrow[M\to\infty]{} 0 \quad \text{for all } \varepsilon>0 \text{ and } u\geqslant 0. \tag{3.1}$$

Remark 2. The condition (3.1) holds whenever

$$\frac{1}{t}Z^u_{u+t}\xrightarrow[t\to\infty]{\mathbf{P}_u\text{-a.s.}} q \quad \text{for every } u<\infty. \tag{3.2}$$

This last condition always holds for the processes with i.i.d. increments with $q=I=\mathbf{E}_0 Z^0_1$ being the Kullback-Leibler information number.

In Theorem 1 below, we establish asymptotic lower bounds for moments of the detection delay of any procedure from the class $\mathbf{\Delta}(\alpha)$ under the condition postulated above. It shows that the number q plays the fundamental role in the asymptotic detection theory.

Remark 3. The log-likelihood ratio Z^u_t can be interpreted as a measure of "distance" between pre- and post-change distributions in the time interval $[u,t]$. Therefore, $(t-u)^{-1}Z^u_t$ can be regarded as an "instantaneous (local) distance" at time t, i.e., the amount of change. Substantial changes result in large values of q in (3.1) and (3.2), the quantity that always appears in the denominator of the first-order asymptotics of ADD and CADD in the following theorems. Hence the delay term is inversely proportional to the amount of change. According to this, larger changes are detected more promptly, which is intuitively obvious. Thus, q measures both the magnitude of change and our ability of its fast detection. On the other hand, the value of ρ characterizes the amount of our prior knowledge about the change occurrence. It also impacts the average detection delay, as our study shows (see below).

In the rest of the paper, for the sake of brevity we use the generic notation $q_\rho=q+\rho$ for the prior distributions with exponential right tails and asymptotically flat prior distributions with the understanding that $\rho=0$ in the latter case, i.e. $q_\rho=q_0=q$ for $\pi\in F$.

Theorem 1. *Let for some positive finite number q condition* (3.1) *hold. Then as $\alpha\to 0$, for all $m>0$*

$$\inf_{\tau\in\mathbf{\Delta}(\alpha)}\mathbf{E}_u\left\{(\tau-u)^m|\tau\geqslant u\right\}\geqslant\left(\frac{|\log\alpha|}{q_\rho}\right)^m(1+o(1)) \quad \textit{for every } u\geqslant 0; \tag{3.3}$$

$$\inf_{\tau\in\mathbf{\Delta}(\alpha)}\mathbf{E}^\pi\left\{(\tau-\lambda)^m|\tau\geqslant \lambda\right\}\geqslant\left(\frac{|\log\alpha|}{q_\rho}\right)^m(1+o(1)). \tag{3.4}$$

A proof of this theorem as well as proofs of further results presented in this section are given in Appendix.

Next, we show that under certain conditions the Bayes Shiryaev rule ν_{B_α} attains the asymptotic lower bounds (3.3) and (3.4). Thus, in particular, it is an asymptotically optimal change detection procedure in the sense of minimizing the CADD and ADD in the class $\mathbf{\Delta}(\alpha)$.

For every $u\geqslant 0$ and $\varepsilon>0$, define a random variable

$$T^{(u)}_\varepsilon=\sup\left\{t\geqslant 0: \left|t^{-1}Z^u_{u+t}-q\right|>\varepsilon\right\},$$

where $\sup\{\varnothing\}=0$. Clearly, in terms of $T^{(u)}_\varepsilon$ the almost sure convergence (3.2) may be written as $\mathbf{P}_u\{T^{(u)}_\varepsilon<\infty\}=1$ for all $\varepsilon>0$.

While the a.s. convergence condition (3.2) is sufficient for obtaining lower bounds for moments of the detection delay, it should be strengthened if we wish to establish asymptotic optimality properties of the detection procedure ν_B. Indeed, in general this condition does not even guarantee finiteness of $\mathrm{CADD}_u(\nu_B)$ and $\mathrm{ADD}(\nu_B)$. In order to study asymptotic operating characteristics we will impose the following constraints on the rate of convergence in the strong law for Z^u_{u+t}/t:

$$\mathbf{E}_u T_\varepsilon^{(u)} < \infty \quad \text{for all } \varepsilon > 0 \text{ and } u \geqslant 0, \tag{3.5}$$

and

$$\mathbf{E}^\pi T_\varepsilon^{(\lambda)} = \int_0^\infty \mathbf{E}_u T_\varepsilon^{(u)} \, d\pi_u < \infty \quad \text{for all } \varepsilon > 0. \tag{3.6}$$

Note that (3.5) is closely related to the condition

$$\int_0^\infty \mathbf{P}_u \left\{ \left| Z^u_{u+t} - qt \right| > \varepsilon t \right\} \, dt < \infty \quad \text{for all } \varepsilon > 0 \text{ and } u \geqslant 0,$$

which is nothing but the complete convergence of $t^{-1} Z^u_{u+t}$ to q (see Hsu and Robbins, 1947, for discrete-time sequences). In what follows we will use the notation

$$t^{-1} Z^u_{u+t} \xrightarrow[t \to \infty]{\mathbf{P}_u-\text{completely}} q$$

for this mode of convergence.

The following theorem establishes the asymptotic optimality result for general statistical models when the conditions (3.5)–(3.6) hold.

Theorem 2. *(i) Let condition* (3.5) *hold for some positive q. Then,*

$$\mathrm{CADD}_u(\nu_B) \sim \frac{\log B}{q_\rho} \quad \textit{as } B \to \infty \textit{ for all } u \geqslant 0. \tag{3.7}$$

If $B = B_\alpha = (1-\alpha)/\alpha$, then $\nu_{B_\alpha} \in \mathbf{\Delta}(\alpha)$ and

$$\inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathrm{CADD}_u(\tau) \sim \mathrm{CADD}_u(\nu_{B_\alpha}) \sim \frac{|\log \alpha|}{q_\rho} \quad \textit{as } \alpha \to 0 \textit{ for all } u \geqslant 0. \tag{3.8}$$

(ii) Let conditions (2.10) *and* (3.6) *hold for some positive q. Then,*

$$\mathrm{ADD}(\nu_B) \sim \frac{\log B}{q_\rho} \quad \textit{as } B \to \infty. \tag{3.9}$$

If $B = B_\alpha = (1-\alpha)/\alpha$, then $\nu_{B_\alpha} \in \mathbf{\Delta}(\alpha)$ and

$$\inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathrm{ADD}(\tau) \sim \mathrm{ADD}(\nu_{B_\alpha}) \sim \frac{|\log \alpha|}{q_\rho} \quad \textit{as } \alpha \to 0. \tag{3.10}$$

Strengthening the complete convergence conditions (3.5) and (3.6), Theorem 2 can be generalized for higher moments of the detection delay. Specifically, assume that for some $r > 1$

$$\mathbf{E}_u [T_\varepsilon^{(u)}]^r < \infty \quad \text{for all } \varepsilon > 0 \text{ and } u \geqslant 0, \tag{3.11}$$

and

$$\mathbf{E}^\pi [T_\varepsilon^{(\lambda)}]^r = \int_0^\infty \mathbf{E}_u [T_\varepsilon^{(u)}]^r \, d\pi_u < \infty \quad \text{for all } \varepsilon > 0. \tag{3.12}$$

If (3.11) holds, it is said that $t^{-1} Z^u_{u+t}$ converges $r-$quickly to the constant q under $\mathbf{P}_u$ (see Lai ,1976; Strassen, 1967). Also, we will now assume finite prior moments,

$$\mathbf{E}\lambda^r = \int_0^\infty s^r \, d\pi_s < \infty. \tag{3.13}$$

Theorem 3. *Suppose that* (3.11)–(3.13) *are satisfied for some* $r \geqslant 1$ *and* $q > 0$. *Then for all* $m \leqslant r$

$$\mathbf{E}_u[(\nu_B - u)^m | \nu_B \geqslant u] \sim \left(\frac{\log B}{q_\rho}\right)^m \quad \textit{as } B \to \infty \textit{ for all } u \geqslant 0; \tag{3.14}$$

$$\mathbf{E}^\pi[(\nu_B - \lambda)^m | \nu_B \geqslant \lambda] \sim \left(\frac{\log B}{q_\rho}\right)^m \quad \textit{as } B \to \infty. \tag{3.15}$$

If $B = B_\alpha = (1-\alpha)/\alpha$, *then* $\nu_{B_\alpha} \in \boldsymbol{\Delta}(\alpha)$ *and for all* $m \leqslant r$ *as* $\alpha \to 0$

$$\inf_{\tau \in \boldsymbol{\Delta}(\alpha)} \mathbf{E}_u[(\tau - u)^m | \tau \geqslant u] \sim \mathbf{E}_u[(\nu_{B_\alpha} - u)^m | \nu_{B_\alpha} \geqslant u] \sim \left(\frac{|\log \alpha|}{q_\rho}\right)^m; \tag{3.16}$$

$$\inf_{\tau \in \boldsymbol{\Delta}(\alpha)} \mathbf{E}^\pi[(\tau - \lambda)^m | \tau \geqslant \lambda] \sim \mathbf{E}^\pi[(\nu_{B_\alpha} - \lambda)^m | \nu_{B_\alpha} \geqslant \lambda] \sim \left(\frac{|\log \alpha|}{q_\rho}\right)^m. \tag{3.17}$$

It is worth noting that in Theorems 2 and 3 the complete and $r-$quick convergence conditions can be relaxed. See Remark 1 in Section 7.

3.3. Weak Optimality

As we established in Section 3.2, the almost sure convergence condition (3.2) is sufficient for obtaining the lower bound for the moments of the detection delay (see Theorem 1 and Remark 2). However, as we also discussed above, this condition is not sufficient for asymptotic optimality with respect to the moments of the detection delay. In particular, it does not even guarantee finiteness of ADD. It is interesting to ask whether some asymptotic optimality result still can be obtained under this condition. The answer is affirmative. In fact, the following theorem shows that the procedure ν_{B_α} is asymptotically optimal in a certain "weak" probabilistic sense.

Theorem 4. *Assume that there exists a finite positive number* q *such that*

$$\frac{1}{t} Z_{u+t}^u \xrightarrow[t \to \infty]{\mathbf{P}_u - a.s.} q \quad \textit{for all } u \geqslant 0. \tag{3.18}$$

(i) Then

$$\frac{(\nu_B - u)^+}{\log B} \xrightarrow[B \to \infty]{\mathbf{P}_u - a.s.} \frac{1}{q_\rho} \quad \textit{for all } u \geqslant 0; \tag{3.19}$$

$$\frac{(\nu_B - \lambda)^+}{\log B} \xrightarrow[B \to \infty]{\mathbf{P}^\pi - a.s.} \frac{1}{q_\rho}, \tag{3.20}$$

where $q_\rho = q + \rho$ *if the prior distribution belongs to the class* $\mathcal{E}(\rho)$ *and* $q_\rho = q$ *if the prior distribution belongs to the class* F.

(ii) Let $B = B_\alpha = (1-\alpha)/\alpha$. *Then for every* $0 < \varepsilon < 1$

$$\inf_{\tau \in \boldsymbol{\Delta}(\alpha)} \mathbf{P}_u \left\{ (\tau - u)^+ > \varepsilon (\nu_{B_\alpha} - u)^+ \right\} \xrightarrow[\alpha \to 0]{} 1 \quad \textit{for all } u \geqslant 0; \tag{3.21}$$

$$\inf_{\tau \in \boldsymbol{\Delta}(\alpha)} \mathbf{P}^\pi \left\{ (\tau - \lambda)^+ > \varepsilon (\nu_{B_\alpha} - \lambda)^+ \right\} \xrightarrow[\alpha \to 0]{} 1. \tag{3.22}$$

3.4. Asymptotic Optimality in the Purely Bayesian Setting

We now consider purely Bayesian setting with the loss function

$$L_m(t,u) = \mathbb{1}_{\{t<u\}} + c(t-u)^m \mathbb{1}_{\{t \geqslant u\}},$$

where $m > 0$, which reduces to (2.4) in the case of $m = 1$. The average risk associated with the detection procedure τ is given by

$$\mathcal{R}_{\rho,c,m}(\tau) = \mathbf{P}^\pi(\tau < \lambda) + c\mathbf{E}^\pi[(\tau-\lambda)^+]^m = \mathrm{P}_{\mathrm{FA}}(\tau) + c[1-\mathrm{P}_{\mathrm{FA}}(\tau)]\mathbf{E}^\pi[(\tau-\lambda)^m | \tau \geqslant \lambda)],$$

where $\rho > 0$ for the class of prior distributions with exponential right tail and $\rho = 0$ for asymptotically flat prior distributions.

Let ν_B be as in (2.13). The first question is how to choose the threshold $B(c,m)$ in the detection procedure ν_B to optimize the performance for small values of the cost c. To answer this question, we observe that ignoring the overshoot $\mathrm{P}_{\mathrm{FA}}(\nu_B) \approx 1/B$ and that by (3.15) $\mathbf{E}^\pi[(\nu_B-\lambda)^m | \nu_B \geqslant \lambda] \approx (q_\rho^{-1} \log B)^m$. Therefore, for large B the average risk is approximately equal to

$$\mathcal{R}_{\rho,c,m}(\nu_B) \approx 1/B + c(q_\rho^{-1} \log B)^m = g_{c,m}(B).$$

The "optimal" threshold value $B = B(c,m)$ that minimizes $g_{c,m}(B)$, $B > 0$, is a solution of the equation

$$m(B/q_\rho)\left(\frac{\log B}{q_\rho}\right)^{m-1} = 1/c. \tag{3.23}$$

In particular, for $m = 1$ we obtain $B(c,1) = q_\rho/c$.

It is intuitively appealing that the procedure $\nu_{B(c,m)}$ with the threshold $B(c,m)$ that satisfies the equation (3.23) is asymptotically optimal as $c \to 0$. In the next theorem, whose proof is given in Appendix, we establish sufficient conditions under which this procedure is indeed asymptotically optimal at least in the first-order sense.

Theorem 5. *Suppose that conditions* (3.12) *and* (3.13) *hold for some positive q and some $r \geqslant 1$. Then, for all $m \leqslant r$,*

$$\mathcal{R}_{\rho,c,m}(\nu_B) \sim c\left(\frac{\log B}{q_\rho}\right)^m \quad \text{as } B \to \infty. \tag{3.24}$$

Let $B = B(c,m)$ be the solution of the equation (3.23)*. Then*

$$\inf_{\tau \in \mathcal{T}} \mathcal{R}_{\rho,c,m}(\tau) \sim \mathcal{R}_{\rho,c,m}(\nu_{B(c,m)}) \sim c\left(\frac{-\log c}{q_\rho}\right)^m \quad \text{as } c \to 0, \tag{3.25}$$

where $\mathcal{T}$ is a set of all $\{\mathcal{F}_t^X\}$-stopping times τ.

4. ASYMPTOTIC OPTIMALITY IN THE I.I.D. CASE

Consider now the change-point detection problem for the processes with i.i.d. increments that, conditioned on $\lambda = u$, follow a distribution F_0 before u and a distribution F_1 after u. Let $I = I(\mathbf{P}_0, \mathbf{P}_\infty) = \mathbf{E}_0 Z_1^0$ be the Kullback-Leibler information between the measures $\mathbf{P}_0$ and $\mathbf{P}_\infty$ for $X_t, 0 \leqslant t \leqslant 1$. In this case, the Kullback-Leibler number I plays the role of the number q that appeared in Theorems 1-5.

The last entry times $T_\varepsilon^{(u)}$ now have the form

$$T_\varepsilon^{(u)} = \sup\left\{t \geqslant 0 : \left|t^{-1} Z_{u+t}^u - I\right| > \varepsilon\right\}. \tag{4.1}$$

Obviously, $\mathbf{E}_u T_\varepsilon^{(u)} = \mathbf{E}_0 T_\varepsilon^{(0)}$ and the r-quick convergence condition

$$\mathbf{E}_0 [T_\varepsilon^{(0)}]^r < \infty \quad \text{for all } \varepsilon > 0 \tag{4.2}$$

implies the "average" r-quick convergence condition (3.12) for any proper prior distribution.

In the case of the processes with i.i.d. increments, the condition $\mathbf{E}_0 \left|Z_1^0\right|^{r+1} < \infty$ implies the $r-$quick convergence (4.2). This can be shown by analogy to the Baum-Katz rate of convergence in the strong law (Baum and Katz, 1965).

Therefore, applying Theorem 3, we obtain that the Shiryaev procedure with the threshold $B_\alpha = (1-\alpha)/\alpha$ minimizes moments of detection delay up to the $r-$th order when $\mathbf{E}_0 |Z_1^0|^{r+1} < \infty$. Also, Theorem 5 can be used to show that the Shiryaev procedure with the threshold $B(c,m)$ is asymptotically optimal as $c \to 0$. More specifically, the following result holds.

Theorem 6. *Let $\{X_t, t \geqslant 0\}$ be the process with i.i.d. increments that has one distribution before change and another distribution after change (conditioned on $\lambda = u$). Suppose that $\mathbf{E}_0 \left|Z_1^0\right|^{r+1} < \infty$ and $\int_0^\infty s^r \, d\pi_s < \infty$ for some $r \geqslant 1$.*

(i) For all $m \leqslant r$ as $B \to \infty$

$$\mathbf{E}_u[(\nu_B - u)^m | \nu_B \geqslant u] \sim \left(\frac{\log B}{I_\rho}\right)^m \quad \textit{for all } u \geqslant 0; \tag{4.3}$$

$$\mathbf{E}^\pi[(\nu_B - \lambda)^m | \nu_B \geqslant \lambda] \sim \left(\frac{\log B}{I_\rho}\right)^m. \tag{4.4}$$

(ii) If $B = B_\alpha = (1-\alpha)/\alpha$, then $\nu_{B_\alpha} \in \mathbf{\Delta}(\alpha)$ and for all $m \leqslant r$ as $\alpha \to 0$

$$\inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathbf{E}_u[(\tau - u)^m | \tau \geqslant u] \sim \mathbf{E}_u[(\nu_{B_\alpha} - u)^m | \nu_{B_\alpha} \geqslant u] \sim \left(\frac{|\log \alpha|}{I_\rho}\right)^m; \tag{4.5}$$

$$\inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathbf{E}^\pi[(\tau - \lambda)^m | \tau \geqslant \lambda] \sim \mathbf{E}^\pi[(\nu_{B_\alpha} - \lambda)^m | \nu_{B_\alpha} \geqslant \lambda] \sim \left(\frac{|\log \alpha|}{I_\rho}\right)^m. \tag{4.6}$$

(iii) If $B = B(c,m)$, where $B(c,m)$ is the solution of (3.23), then for all $m \leqslant r$

$$\inf_{\tau \in \mathcal{T}} \mathcal{R}_{\rho,c,m}(\tau) \sim \mathcal{R}_{\rho,c,m}(\nu_{B(c,m)}) \sim c\left(\frac{-\log c}{I_\rho}\right)^m \quad \textit{as } c \to 0, \tag{4.7}$$

where $I_\rho = I + \rho$ for $\pi \in \mathcal{E}(\rho)$ and $I_\rho = I_0 = I$ for $\pi \in F$.

We conjecture that finiteness of higher order moments of the LLR in this theorem can be relaxed to finiteness of the first absolute moment $\mathbf{E}_0|Z_1^0| < \infty$, i.e., the Shiryaev procedure asymptotically minimizes all positive moments of the detection delay under the assumption $\mathbf{E}_0|Z_1^0| < \infty$. See Theorem 5 in Tartakovsky and Veeravalli (2005) for a related result for discrete-time i.i.d. sequences.

5. ASYMPTOTIC PERFORMANCE OF THE SHIRYAEV-ROBERTS-POLLAK PROCEDURE

It is known that in the discrete-time i.i.d. case where the change point is considered as a non-random, unknown parameter, Page's CUSUM test (Page, 1954) and the randomized version of the Shiryaev-Roberts detection procedure proposed by Pollak (1985) are optimal with respect to the minimax expected detection delay, subject to a constraint on the mean time to false alarm (see Basseville and Nikiforov, 1993; Lorden, 1971; Moustakides, 1986; Pollak, 1985; Tartakovsky, 1991, 1998c). The latter procedure will be referred to as the Shiryaev-Roberts-Pollak (SRP) detection procedure. Recently, Tartakovsky and Veeravalli (2005) proved that these procedures are not optimal (even asymptotically) with respect to the Bayesian criterion when the prior distribution of the change point belongs to the class $\mathcal{E}(\rho)$, but remain asymptotically optimal for asymptotically flat prior distributions A2. See also Lai (1998) for related results regarding the CUSUM test.

For continuous-time models, beyond the problem of detecting a change in the drift of Brownian motion little is known about the properties of CUSUM and SRP procedures. Shiryaev (1996) and Beibel (1996) proved that the CUSUM test is strictly optimal in the minimax sense in the class of procedures for which the average run length $\mathbf{E}_\infty \tau \geqslant T, T > 0$ for detecting a change in the drift of the Brownian motion. Pollak and Siegmund (1985) studied the asymptotic performance of these procedures for homogeneous Brownian motion models in terms of the average run lengths $\mathbf{E}_0 \tau$ and $\mathbf{E}_\infty \tau$ and showed that they have almost the same performance: the SRP procedure only slightly outperforms the CUSUM test when the post-change drift is exactly specified. Tartakovsky (1995) studied the nonhomogeneous case when both the drift and diffusion are varying functions in time. See also Beibel (2000) for a discussion in a specific (non-classical) Bayesian setting that is different from our classical approach.

In this section, we outline asymptotic properties of the SRP change-point detection procedure in the Bayesian problem for general statistical models. The conclusion is that it loses the asymptotic optimality property under the Bayesian criterion for prior distributions with exponential right tails $\pi \in \mathcal{E}(\rho)$, but remains asymptotically optimal for asymptotically flat prior distributions $\pi \in F$.

The non-randomized SRP procedure has the form

$$\hat{\nu}_B = \inf \{t \geqslant 0 : R_t \geqslant B\},$$

where the statistic R_t is defined as follows

$$R_t = \int_0^t e^{Z_t^s} ds, \quad R_0 = 0.$$

Note that the statistic $R_t = \lim_{\rho \to 0} \Lambda_t / \rho$ when the prior distribution of λ is exponential (2.7) with $\pi_0 = 0$.

In order to estimate the PFA of the SRP procedure, we note that the statistic $R_t - t$ is a zero-mean $(\mathbf{P}_\infty, \mathcal{F}_t^X)$-martingale. Therefore, R_t is a submartingale with mean $\mathbf{E}_\infty R_t = t$. Using Doob's submartingale inequality, we obtain

$$\mathbf{P}_\infty \{\hat{\nu}_B < t\} = \mathbf{P}_\infty \left\{ \max_{0 \leqslant s < t} R_s \geqslant B \right\} \leqslant \mathbf{E}_\infty (R_t)/B = t/B, \tag{5.1}$$

which yields

$$\widehat{\mathrm{P}}_{\mathrm{FA}}(B) = \mathbf{P}^\pi (\hat{\tau}_B < \lambda) = \int_0^\infty \mathbf{P}_\infty \{\hat{\tau}_B < t\} \, d\pi_t \leqslant \bar{\lambda}/B,$$

where $\bar{\lambda} = \int_0^\infty s d\pi_s$ is the mean value of λ. Thus, choosing $B = B_\alpha = \bar{\lambda}/\alpha$ guarantees $\hat{\nu}_{B_\alpha} \in \mathbf{\Delta}(\alpha)$. In particular, $\bar{\lambda} = 1/\rho$ for the exponential prior distribution (2.7).

The following theorem establishes the asymptotic operating characteristics of the SRP procedure with respect to the average detection delay and average risk function.

Theorem 7. *Let $B = B_\alpha = \bar{\lambda}/\alpha$. Then under the conditions of Theorem 2(i)*

$$\mathrm{CADD}_u(\hat{\tau}_{B_\alpha}) \sim |\log \alpha|/q \quad \textit{for all } u \geqslant 0 \textit{ as } \alpha \to 0 \tag{5.2}$$

and under the conditions of Theorem 2(ii)

$$\mathrm{ADD}(\hat{\tau}_{B_\alpha}) \sim |\log \alpha|/q \quad \textit{as } \alpha \to 0. \tag{5.3}$$

If $B = B_c = q/c$, then under the conditions of Theorem 2(ii)

$$\mathcal{R}_{\rho,c,1}(\hat{\tau}_{B_c}) \sim c|\log c|/q \quad \textit{as } c \to 0. \tag{5.4}$$

The proof of this theorem runs along the lines of the proofs of Theorems 2 and 5. A sketch of the proof is presented in Appendix. In a similar manner we can establish counterparts of Theorems 3 and 5 for higher moments of detection delay. More specifically, Theorem 7 can be formulated in terms of higher moments of detection delay of the SRP procedure under conditions postulated in Theorems 3 and 5.

Comparing Theorems 2, 5, and 7 shows that the SRP procedure is not asymptotically optimal for prior distributions $\pi \in \mathcal{E}(\rho)$, but it is optimal for prior distributions $\pi \in F$, which is not surprising since the SRP statistic R_t uses flat prior.

The asymptotic relative efficiency (ARE) of the Shiryaev procedure with respect to the SPR procedure

$$\mathrm{ARE} = \lim_{\alpha \to 0} \frac{\mathrm{ADD}(\nu_{B_\alpha})}{\mathrm{ADD}(\hat{\tau}_{B_\alpha})} = \frac{q}{q+\rho}$$

depends on q, the amount of change, and ρ, the amount of our prior knowledge about the change. If the situation is *a priori* uncertain, ρ should be chosen small, in which case $\mathrm{ARE} \approx 1$, i.e. the SPR procedure is asymptotically optimal.

6. APPLICATIONS AND EXAMPLES

This section illustrates the considered change-point detection methods in several applied problems. It also shows how conditions (3.11), (3.12) (in particular, (3.5), (3.6)) of Theorems 2 and 3 can be verified in practice.

Example 1. Detection of a deterministic signal with an unknown appearance time in correlated Gaussian noise. This example is motivated by the problem of detecting a target that appears at an unknown random instant λ. Radar, infrared, and acoustic surveillance systems deal with the detection of moving targets that appear and disappear randomly at unknown points in time. The most challenging problem for these systems is the rapid detection of dim targets against heavily cluttered backgrounds. Typically sensors observe the signal S_t from the target in the presence of correlated clutter V_t and white Gaussian noise $\dot{W}_t$ (sensor noise). See, e.g., Tartakovsky (1991); Tartakovsky and Blažek (2000); Tartakovsky and Veeravalli (2004).

Let the observation process obey the following Itô stochastic differential equation

$$dX_t = [S_t \mathbb{1}_{\{\lambda \leqslant t\}} + V_t]dt + \sqrt{N}\, dW_t, \quad t \geqslant 0,$$

where S_t is a deterministic function (signal), $\{V_t\}_{t \geqslant 0}$ is an L_2-continuous Gaussian process, $\{W_t\}_{t\geqslant 0}$ is a standard Wiener process, N is its intensity, and $\mathbb{1}_{\{\mathcal{A}\}}$ is the indicator of the set $\mathcal{A}$.

Let $\widehat{V}_t = \mathbf{E}_\infty \left[V_t \mid \mathcal{F}_t^X\right]$ be an optimal (in the mean-square-error sense) filtering estimate of the process V_t observed in white Gaussian noise. Since V_t is a Gaussian process, $\widehat{V}_t$ is a linear functional

$$\widehat{V}_t = \int_0^t C(t,v) dX_v,$$

where $C(t, v)$ is a characteristic of an optimal filter that satisfies the well-known Wiener-Hopf equation (see, e.g., Liptser and Shiryaev, 1977). For $u < t$ define

$$\widetilde{S}_{u,t} = S_t - \int_u^t S_v C(t, v)\, dv, \quad \widetilde{X}_{u,t} = X_t - \int_u^t C(t, v)\, dX_v.$$

With the use of Theorems 7.12 and 7.15 of Liptser and Shiryaev (1977) it can be shown that the process $\widetilde{X}_{u,t}$ may be represented via an innovation (standard Wiener) process $\widetilde{W}_t$ as

$$\widetilde{X}_{u,t} = \int_u^t \widetilde{S}_{u,v}\, dv + \sqrt{N}\, \widetilde{W}_t \tag{6.1}$$

and that the LLR is of the form

$$Z_t^u = \frac{1}{N} \int_u^t \widetilde{S}_{u,v}\, d\widetilde{X}_{u,v} - \frac{1}{2N} \int_u^t \widetilde{S}_{u,v}^2\, dv \tag{6.2}$$

(cf. Tartakovsky, 1991, 1998a, 1998b).

Using (6.1) and (6.2), we obtain that, under the hypothesis $H_u : \lambda = u$,

$$Z_t^u = \frac{1}{2N} \int_u^t \widetilde{S}_{u,v}^2\, dv + N^{-1/2} \int_u^t \widetilde{S}_{u,v}\, d\widetilde{W}_v. \tag{6.3}$$

Assume that for every $u \geqslant 0$

$$\frac{1}{2N} \int_u^t \widetilde{S}_{u,v}^2\, dv = \frac{\mu}{2N} \cdot t\,(1 + o(1)) \quad \text{as } t \to \infty, \tag{6.4}$$

where μ is positive and finite, and $o(1)$ converges to 0 as $t \to \infty$ uniformly in u, $u > 0$.

The value of $\frac{1}{2N} \int_u^t \widetilde{S}_{u,v}^2\, dv$ represents the cumulative signal-to-noise ratio (SNR) in the time interval $[u, t]$ at the output of the whitening filter (that whitens clutter V_t). Therefore, the value of $q = \mu/(2N)$ can be interpreted as instantaneous SNR (per time unit) in the stationary regime and μ as the instantaneous power of the signal. In target detection applications, condition (6.4) typically holds. Specific examples are considered below.

Using (6.3) it is readily shown that

$$t^{-1} Z_{u+t}^u \xrightarrow[t\to\infty]{\mathbf{P}_u - r - \text{quickly}} q = \mu/(2N) \quad \text{for every } u \geqslant 0 \text{ and all } r > 0.$$

Indeed, write

$$Q_{u,u+t} = \frac{1}{2Nt} \int_u^{u+t} \widetilde{S}_{u,v}^2\, dv \quad \text{and} \quad \Delta_{u,u+t} = Q_{u,u+t} - q.$$

It is easy to check that

$$\begin{aligned}\mathbf{P}_u \left\{ \left| Z_{u+t}^u - qt \right| > \varepsilon t \right\} &= \Phi \left\{ -\frac{(\varepsilon + \Delta_{u,u+t})\sqrt{t}}{\sqrt{2Q_{u,u+t}}} \right\} + \Phi \left\{ -\frac{(\varepsilon - \Delta_{u,u+t})\sqrt{t}}{\sqrt{2Q_{u,u+t}}} \right\} \\ &\leqslant 2\Phi \left\{ -\frac{(\varepsilon - |\Delta_{u,u+t}|)\sqrt{t}}{\sqrt{2(q + \Delta_{u,u+t})}} \right\},\end{aligned}$$

where $\Phi(x)$ is the standard normal distribution function. By condition (6.4), there exists T_0, independent of u because of the uniform convergence, such that $|\Delta_{u,u+t}| < \varepsilon/2$ for all $t > T_0$. Then, for all $r > 0$ and $\varepsilon > 0$,

$$\int_0^\infty t^{r-1} \mathbf{P}_u \left\{ \left| Z_{u+t}^u - qt \right| > \varepsilon t \right\} dt \leqslant \int_0^{T_0} t^{r-1} dt + \int_{T_0}^\infty t^{r-1} \Phi \left\{ -\frac{\varepsilon \sqrt{t}}{2\sqrt{2q + \varepsilon}} \right\} dt < \infty, \tag{6.5}$$

and hence, condition (3.11) holds for all positive r with $q = \mu/2N$.

Also, since the middle part of (6.5) does not depend on u, we have

$$\int_0^\infty d\pi_u \left(\int_0^\infty t^{r-1} \mathbf{P}_u \left\{ \left| Z_{u+t}^u - qt \right| > \varepsilon t \right\} dt \right) < \infty,$$

which implies condition (3.12).

By Theorem 2, the asymptotic equalities (3.7)–(3.10) hold with $q = \mu/2N$, and the detection procedure ν_{B_α} is asymptotically optimal within the class $\mathbf{\Delta}(\alpha)$ for small α with respect to ADD and CADD. Furthermore, according to Theorem 3 all the moments of the detection delay are minimized as $\alpha \to 0$. In particular, for all $m > 0$

$$\inf_{\tau \in \mathbf{\Delta}(\alpha)} \mathbf{E}_u[(\tau - u)^m | \tau \geqslant u] \sim \mathbf{E}_u[(\nu_{B_\alpha} - u)^m | \nu_{B_\alpha} \geqslant u] \sim \left(\frac{|\log \alpha|}{q_\rho} \right)^m, \quad u \geqslant 0.$$

Condition (6.4), which is sufficient for optimality of the Shiryaev procedure, is satisfied in many applications. To illustrate this fact, we now consider several special cases.

Deterministic signals in Markov Gaussian noise. Let V_t be a zero-mean Markov Gaussian process that is described by the Itô stochastic equation

$$dV_t = -\beta V_t \, dt + \sqrt{\sigma} \, dw_t, \quad V_0 \sim \mathcal{N}(0, \delta^2), \tag{6.6}$$

where $\{w_t, t \geqslant 0\}$ is a standard Wiener process. Assume that $N = 0$, i.e. there is no white noise in observations. Then, the LLR is given by (6.2) where N is replaced with σ and where

$$\widetilde{X}_{u,t} = \widetilde{X}_t = \beta X_t + \dot{X}_t, \quad \widetilde{S}_{u,t} = \widetilde{S}_t = \beta S_t + \dot{S}_t.$$

Hereafter $\dot{Y}_t = dY_t/dt$ denotes a time derivative. It follows that

$$\mu = \lim_{t \to \infty} \frac{1}{t} \int_0^t \widetilde{S}_v^2 \, dv \quad \text{and} \quad q = \mu/2\sigma.$$

Let S_t be a constant, $S_t = S_0$ for $t \geqslant 0$. Then $\widetilde{S}_t = \beta S_0$ and $q = \beta^2 S_0^2/2\sigma$.

Assume now that $S_t = A \sin \omega t$ is a harmonic signal with the amplitude $A > 0$ and frequency ω. Then $\widetilde{S}_t = A(\beta \sin \omega t + \omega \cos \omega t)$ and

$$\int_0^t \widetilde{S}_v^2 \, dv = A^2 \left[\frac{1}{2}(\beta^2 + \omega^2)t - \frac{1}{4}(\beta^2/\omega + \omega) \sin 2\omega t - \beta \sin \omega t \right].$$

Thus, condition (6.4) holds with $\mu = A^2(\beta^2 + \omega^2)/2$ and the value of q is equal to $A^2(\beta^2 + \omega^2)/4\sigma$. A similar result can be obtained for the sequence of harmonic pulses – the signal that is typically used in radar applications.

Finally, suppose that, conditioned on $\lambda = u$, $S_t = A(t - u)$ for $u \leqslant t \leqslant u + t_0$ and $S_t = A$ for $t \geqslant u + t_0$ (gradual linear increase). Then for $t > t_0$

$$\int_0^t \widetilde{S}_v^2 \, dv = A^2[t_0 + \beta^2 t_0^3/3 + \beta t_0 + \beta^2(t - t_0)],$$

and condition (6.4) holds with $\mu = A^2\beta^2$; the value of q is equal to $A^2\beta^2/2\sigma$.

Deterministic signals in Markov Gaussian clutter and white noise. Let V_t be as in (6.6) but $N > 0$, i.e. there is white Gaussian noise with intensity N in the observations. As we mentioned above, this model arises in many applications of detection theory where the signal is observed in the presence of clutter V_t, which is described by the Markov Gaussian process, and sensor noise $\dot{W}_t$, which is described by white Gaussian noise.

In the Markov case, it is easier to exploit the Kalman-Bucy approach rather than the Wiener-Hopf approach described above. Indeed, by Theorem 10.1 in Liptser and Shiryaev (1977) for $t \geqslant 0$ the optimal estimate $\widehat{V}_t$ satisfies the system of Kalman-Bucy equations

$$d\widehat{V}_t = -\left(\beta + K_t/N\right)\widehat{V}_t dt + N^{-1}K_t dX_t, \quad \widehat{V}_0 = 0; \tag{6.7}$$

$$\dot{K}_t = -\left(2\beta K_t + K_t^2/N\right) + \sigma, \quad K_0 = \delta^2, \tag{6.8}$$

where $K_t = \mathbf{E}_\infty[V_t - \widehat{V}_t]^2$ is the mean-square filtering error. In what follows we set $\delta^2 = \sigma/2\beta$, in which case the Markov process V_t is stationary, $\mathbf{E}V_tV_{t+\tau} = \delta^2 \exp\{-\beta|\tau|\}$.

Solving equation (6.7), we obtain that for $t > u$

$$\widetilde{S}_{u,t} = S_t - \frac{1}{N}\exp\left\{-\int_u^t (\beta + K_v/N)\, dv\right\}\int_u^t S_v K_v \exp\left\{\int_u^v (\beta + K_\tau/N)\, d\tau\right\} dv. \tag{6.9}$$

Now, solving the Ricatti equation (6.8), we get

$$\frac{1}{t}\int_0^t K_v\, dv = N\beta\left(\sqrt{1 + 2\delta^2/N\beta} - 1\right) + o(1) \quad \text{as } t \to \infty. \tag{6.10}$$

Let $S_t = S_0$ for all $t \geqslant 0$. Then, applying relations (6.9) and (6.10) together yields

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \widetilde{S}_{0,v}^2\, dv = \mu = \frac{S_0^2}{1 + 2\delta^2/N\beta}\,.$$

Therefore, condition (6.4) holds and

$$q = \frac{S_0^2}{2N(1 + 2\delta^2/N\beta)}$$

in this case.

In a particular case where $V_t = 0$ (no clutter), the value of q is equal to $S_0^2/2N$. This problem is a classical problem of detecting the abrupt change of the drift of the Wiener process that has been extensively studied by Shiryaev (1978) in a non-asymptotic setting.

Now, let $S_t = A\sin\omega t$ be the harmonic signal and let $V_t = 0$. Then $\widetilde{S}_{u,t} = A\sin\omega t$, condition (6.4) holds with $\mu = A^2/2$, and $q = A^2/4N$.

Finally, let $V_t = 0$ and $S_t = Af(t-u)$, $t \geqslant u$ (conditioned on $\lambda = u$), where $f(t)$ is a monotone nondecreasing function such that $f(0) = 0$ and $f(\infty) = 1$. This problem corresponds to the detection of a gradual change in the drift of the Brownian motion. In this case, $\widetilde{S}_{u,t} = Af(t-u)$ and condition (6.4) holds with

$$\mu = A^2 \lim_{t\to\infty} t^{-1}\int_0^t f^2(v)\, dv.$$

Example 2. Change-point detection in additive Itô processes. Rapid detection of additive changes in discrete-time linear state-space models (hidden Markov models) is a conventional problem on fault detection in various dynamical systems (see, e.g., Basseville and Nikiforov, 1993; Willsky, 1976). In this example we consider a linear continuous-time model driven by the Brownian motion where the additive (unobservable) component is described by a general Itô process.

Let the observed process X_t have the Itô stochastic differential

$$dX_t = \begin{cases} S_0(t)dt + \sqrt{N}\,dW(t) & \text{for } \lambda > t, \\ S_1(t)dt + \sqrt{N}\,dW(t) & \text{for } \lambda \leqslant t, \end{cases} \tag{6.11}$$

where $S_0(t)$ and $S_1(t)$ are Itô stochastic processes. In what follows, we will always assume that

$$\int_0^t \mathbf{E}S_i^2(v)\,dv < \infty,\ i = 0, 1 \quad \text{for every } t < \infty.$$

Define the functionals $\widehat{S}_1(u,t) = \mathbf{E}_u[S_1(t) \mid \mathcal{F}_t^X]$ and $\widehat{S}_0(t) = \mathbf{E}_\infty[S_0(t) \mid \mathcal{F}_t^X]$.

By Theorem 7.12 of Liptser and Shiryaev (1977), there exist standard Wiener (innovation) processes $\{\widetilde{W}_i(t)\}$, $i = 0, 1$ such that under the hypothesis "$H_u : \lambda = u$" the process X_t allows for the following "minimal" representation in the form of diffusion-type processes

$$dX_t = \begin{cases} \widehat{S}_0(t)dt + \sqrt{N}\,d\widetilde{W}_0(t) & \text{for } u > t, \\ \widehat{S}_1(u,t)dt + \sqrt{N}\,d\widetilde{W}_1(t) & \text{for } u \leqslant t. \end{cases} \tag{6.12}$$

The processes $\widetilde{W}_i(t)$ are statistically equivalent to the original Wiener process $W(t)$.

Using this representation along with the results of absolute continuity of probability measures of diffusion-type processes with respect to the Wiener measure $\mathbf{P}_W$ (see, e.g., Theorems 7.6 and 7.7 of Liptser and Shiryaev, 1977), we obtain

$$\log \frac{\mathbf{P}_u^t}{\mathbf{P}_W^t}(X^t) = \frac{1}{N}\int_0^u \widehat{S}_0(v)\,dX_v - \frac{1}{2N}\int_0^u \widehat{S}_0^2(v)\,dv + \frac{1}{N}\int_u^t \widehat{S}_1(u,v)\,dX_v - \frac{1}{2N}\int_u^t \widehat{S}_1^2(u,v)\,dv;$$
$$\log \frac{\mathbf{P}_\infty^t}{\mathbf{P}_W^t}(X^t) = \frac{1}{N}\int_0^t \widehat{S}_0(v)\,dX_v - \frac{1}{2N}\int_0^t \widehat{S}_0^2(v)\,dv.$$

Applying these last formulas yields

$$Z_t^u = \frac{1}{N}\int_u^t \left[\widehat{S}_1(u,v) - \widehat{S}_0(v)\right]\,dX_v - \frac{1}{2N}\int_u^t \left[\widehat{S}_1^2(u,v) - \widehat{S}_0^2(v)\right]\,dv. \tag{6.13}$$

Now, using (6.12) and (6.13), we get that under H_u

$$Z_t^u = \frac{1}{\sqrt{N}}\int_u^t \left[\widehat{S}_1(u,v) - \widehat{S}_0(v)\right] d\widetilde{W}_1(v) + \frac{1}{2N}\int_u^t \left[\widehat{S}_1(u,v) - \widehat{S}_0(v)\right]^2 dv. \tag{6.14}$$

Assume that

$$\frac{1}{t}\int_u^{u+t} \widehat{S}_i(u,v)\,d\widetilde{W}_1(v) \xrightarrow[t\to\infty]{\mathbf{P}_u-\text{completely}} 0, \quad i = 0, 1$$

and

$$\frac{1}{t}\int_u^{u+t} \left[\widehat{S}_1(u,v) - \widehat{S}_0(v)\right]^2 dv \xrightarrow[t\to\infty]{\mathbf{P}_u\text{-completely}} \mu,$$

where μ is positive and finite. Then, obviously,

$$\frac{1}{t}Z_{u+t}^u \xrightarrow[t\to\infty]{\mathbf{P}_u-\text{completely}} q = \mu/2N \quad \text{for every } u \geqslant 0$$

and the Shiryaev detection procedure is asymptotically optimal.

In most cases

$$\mu = \lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{E}_0 \left[\widehat{S}_1(0,v) - \widehat{S}_0(v)\right]^2 dv.$$

This fact is illustrated in the next example for Gaussian Markov processes.

Example 3. Change-point detection in Gaussian hidden Markov models. Consider a continuous-time Gaussian hidden Markov model which is a particular case of the Itô model (6.11) with $S_i(t) = \mu_i(t) + A_i(t)\theta_i(t)$, $i = 0, 1$, where $A_i(t)$ and $\mu_i(t)$ are deterministic continuous functions and $\{\theta_i(t)\}$, $i = 0, 1$, are mutually independent Markov Gaussian processes. For $\lambda = u$, $u \geqslant 0$, the processes $\theta_0(t)$ and $\theta_1(t)$ obey the following Itô stochastic differential equations

$$\begin{aligned} d\theta_0(t) &= a_0(t)\theta_0(t)dt + \sqrt{\sigma_0(t)}dw_0(t), \quad t \geqslant 0, \quad \theta_0(0) \sim \mathcal{N}(0, \delta_0^2); \\ d\theta_1(t) &= a_1(t)\theta_1(t)dt + \sqrt{\sigma_1(t)}dw_1(t), \quad t \geqslant u, \quad \theta_1(u) \sim \mathcal{N}(0, \delta_1^2), \end{aligned}$$

where $\{w_0(t)\}_{t\geqslant 0}$ and $\{w_1(t)\}_{t\geqslant 0}$ are independent standard Wiener processes both independent of the Wiener process $W(t)$; $a_i(t)$ and $\sigma_i(t)$, $i = 0, 1$ are deterministic continuous functions. Note that $\theta_1(t) = \theta_1(u,t)$ depends on u.

Assume that the functions $A_i(t)$, $\mu_i(t)$, $a_i(t)$ and $\sigma_i(t)$ are such that for every finite $T > 0$

$$\int_0^T A_i^2(t)dt < \infty, \quad \int_0^T \mu_i^2(t)dt < \infty, \quad \int_0^T |a_i(t)|dt < \infty, \quad \int_0^T \sigma_i(t)dt < \infty.$$

Write $\hat\theta_1(u,t) = \mathbf{E}_u[\theta_1(t) \mid \mathcal{F}_t^X]$, $\hat\theta_0(t) = \mathbf{E}_\infty[\theta_0(t) \mid \mathcal{F}_t^X]$, $\widehat{S}_0(t) = \mu_0(t) + A_0(t)\hat\theta_0(t)$ and $\widehat{S}_1(u,t) = \mu_1(t) + A_1(t)\hat\theta_1(u,t)$.

The formulas (6.13) and (6.14) hold with $\widehat{S}_i(t)$ defined above. The functional $\hat\theta_1(u,t)$ is nothing but the optimal mean-square-error estimate (filtering estimate) of the Markov process $\theta_1(t) = \theta_1(u,t)$ based on the data $X^t = \{X_v, 0 \leqslant v \leqslant t\}$. For $t \geqslant u$, it satisfies the system of Kalman-Bucy equations (see, e.g., Liptser and Shiryaev, 1977; Tartakovsky, 1991, 1998a)

$$\begin{aligned} d\hat\theta_1(u,t) &= \left(a_1(t) - \frac{K_1(u,t)A_1^2(t)}{N}\right)\hat\theta_1(u,t)dt \\ &\quad + \frac{K_1(u,t)A_1(t)}{N}(dX_t - \mu_1(t)dt), \quad \hat\theta_1(u) = 0, \end{aligned} \tag{6.15}$$

$$\frac{dK_1(u,t)}{dt} = 2a_1(t)K_1(u,t) - A_1^2(t)K_1^2(u,t)/N + \sigma_1(t), \quad K_1(u,u) = \delta_1^2, \tag{6.16}$$

where $K_1(u,t) = \mathbf{E}_0[\theta_1(u,t) - \hat\theta_1(u,t)]^2$ is the mean-square filtering error. For $t \geqslant 0$, the same system describes the behavior of the filtering estimate $\hat\theta_0(t)$ of the process $\theta_0(t)$ if we set $u = 0$ and replace $a_1(t)$ with $a_0(t)$, $A_1(t)$ with $A_0(t)$, $\sigma_1(t)$ with $\sigma_0(t)$, and $K_1(u,t)$ with $K_0(t) = \mathbf{E}_\infty[\theta_0(t) - \hat\theta_0(t)]^2$, $t \geqslant 0$, $K_0(0,0) = \delta_0^2$. In the rest of this subsection, we will omit the index u when $u = 0$. For example, we will write $\widehat{S}_1(t)$ in place of $\widehat{S}_1(0,t)$.

Suppose that

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t \mathbf{E}_0 \left[\widehat{S}_1(v) - \widehat{S}_0(v)\right]^2 dv = \mu, \tag{6.17}$$

where μ is positive and finite. Then, using (6.14) and an argument similar to that used in Example 1 of Tartakovsky (1998a), it can be proven that the value of $t^{-1}\int_u^{u+t} \widehat{S}_i(v)d\widetilde{W}_1(v)$ converges completely to zero and the normalized LLR converges completely to the constant $q = \mu/2N$,

$$\frac{1}{t} Z_{u+t}^u \xrightarrow[t\to\infty]{\mathbf{P}_u\text{-completely}} q \quad \text{for every } u \geqslant 0.$$

Therefore, Theorem 2 applies with $q = \mu/2N$, where μ is defined in (6.17). In particular, since trajectories of the statistic Λ_t are continuous, there is no overshoot and $B = B_\alpha = (1-\alpha)/\alpha$ implies $\mathrm{P}_{\mathrm{FA}} = \alpha$. With this threshold the detection procedure ν_{B_α} is asymptotically optimal in the class $\boldsymbol{\Delta}(\alpha)$,

$$\inf_{\tau \in \boldsymbol{\Delta}(\alpha)} \mathrm{CADD}_u(\tau) \sim \mathrm{CADD}_u(\nu_{B_\alpha}) \sim \frac{|\log \alpha|}{q_\rho} \quad \text{as } \alpha \to 0 \text{ for all } u \geqslant 0,$$

where $q_\rho = q + \rho$ for the exponential prior and $q_\rho = q$ for asymptotically flat priors.

In order to find the value of q explicitly, let us confine ourselves to the particular case of stationary hidden Markov models where $\mu_i(t) = \mu_i$, $a_i(t) = -\beta_i$, $\sigma_i(t) = \sigma_i$, $A_i(t) = 1$, and $\delta_i = \sigma_i/2\beta_i$ ($\beta_i > 0$, $\sigma_i > 0$). Then, the Markov processes $S_i(t)$ are stationary with the parameters

$$\mathbf{E}S_i(t) = \mu_i, \quad \mathbf{E}[S_i(t) - \mu_i][S_i(t+v) - \mu_i] = \delta_i^2 e^{-\beta_i |v|}.$$

We first observe that

$$\mathbf{E}_0 \left[\widehat{S}_1(t) - \widehat{S}_0(t)\right]^2 = (\mu_1 - \mu_0)^2 - 2(\mu_1 - \mu_0)\mathbf{E}_0\hat{\theta}_0(t) + \mathbf{E}_0 \left[\hat{\theta}_1(t) - \hat{\theta}_0(t)\right]^2; \tag{6.18}$$

$$\mathbf{E}_0 \left[\hat{\theta}_1(t) - \hat{\theta}_0(t)\right]^2 = D(t) - K_1(t), \tag{6.19}$$

where $D(t) = \mathbf{E}_0[\theta_1(t) - \hat{\theta}_0(t)]^2$ is the mean-square error of estimating $\theta_1(t)$ by the filter $\hat{\theta}_0(t)$ and $K_1(t) = K_1(0,t)$. It is not difficult to show that

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \mathbf{E}_0 \hat{\theta}_0(v)\, dv = 0. \tag{6.20}$$

(If $\mu_1 = \mu_0$, then $\mathbf{E}_0\hat{\theta}_0(t) = 0$ for all $t \geqslant 0$.)

Similar to (6.10)

$$\frac{1}{t} \int_0^t K_1(v)\, dv = N\beta_1(\zeta_1 - 1) + o(1) \quad \text{as } t \to \infty, \tag{6.21}$$

where $\zeta_1 = \sqrt{1 + 2\delta_1^2/N\beta_1}$. Finally, straightforward but cumbersome computations show that

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t D(v)\, dv = N\beta_0(\zeta_0 - 1) - \frac{\sigma_1(\beta_1^2 - \beta_0^2)}{2\beta_0\beta_1\zeta_0(\beta_1 + \beta_0\zeta_0)} + \frac{\sigma_1 - \sigma_0}{2\beta_0\zeta_0}, \tag{6.22}$$

where $\zeta_0 = \sqrt{1 + 2\delta_0^2/N\beta_0}$.

Combining (6.18)–(6.22) yields

$$q = \frac{(\mu_1 - \mu_0)^2}{2N} + \frac{\beta_0(\zeta_0 - 1) - \beta_1(\zeta_1 - 1)}{2} - \frac{\sigma_1(\beta_1^2 - \beta_0^2)}{4N\beta_0\beta_1\zeta_0(\beta_1 + \beta_0\zeta_0)} + \frac{\sigma_1 - \sigma_0}{4N\beta_0\zeta_0}.$$

The computations are simplified in the case where $S_0(t) = 0$. Then

$$q = \frac{\mu_1^2}{2N} + \frac{\delta_1^2}{2N} \frac{Q}{\left(1 + \sqrt{1 + Q}\right)^2}, \quad \text{where } Q = 2\delta_1^2/(\beta_1 N).$$

In the latter, stationary case, it can be shown that not only ADD but also all the moments of the detection delay are asymptotically minimized.

It is worth remarking that similar results hold in case of correlated processes $\theta_0(t)$ and $\theta_1(t)$. In particular, if $w_1(t) = w_0(t) = w(t)$ the same asymptotic result holds true.

Example 4. Change-point detection in the intensity of a nonhomogeneous Poisson process. In previous Gaussian models the trajectories of the decision statistic were continuous. Consider now the example of the process with jumps. Suppose, under hypothesis "$H_u : \lambda = u$", the observed process $\{X_t, t \geqslant 0\}$ represents a non-stationary Poisson random process with intensity $\lambda_0(t)$ before the change point u and with intensity $\lambda_1(t)$ after the change. Then, for $t \geqslant u$,

$$Z_t^u = \int_u^t \log\left(\frac{\lambda_1(v)}{\lambda_0(v)}\right) dX_v - \int_u^t (\lambda_1(v) - \lambda_0(v))\, dv;$$

$$\mu_{u,t} = \mathbf{E}_u Z_t^u = \int_u^t \left\{\lambda_0(v) + \lambda_1(v)\left[\log\left(\frac{\lambda_1(v)}{\lambda_0(v)}\right) - 1\right]\right\} dv;$$

$$\widetilde{Z}_t^u = Z_t^u - \mu_{u,t} = \int_u^t \log\left(\frac{\lambda_1(v)}{\lambda_0(v)}\right) d\left[X_v - \int_u^v \lambda_1(s)\, ds\right];$$

$$D_{u,t} = \mathbf{E}_u(\widetilde{Z}_t^u)^2 = \int_u^t \lambda_1(v)\left(\log\frac{\lambda_1(v)}{\lambda_0(v)}\right)^2 dv.$$

One of the important applications for this example is found in network security (Kent, 2000). Network intrusions, such as denial of service attacks, occur at unknown points in time and should be detected as rapidly as possible. To detect denial of service attacks, one can monitor the number of packets (TCP, ICMP, UDP), Blažek et al. (2001); Tartakovsky et al. (2006); Tartakovsky and Veeravalli (2004); Wang et al. (2002). The attack leads to an abrupt change of the average number of packets of a particular type. Models for network traffic are usually nonstationary and even bursty. The nonhomogeneous Poisson model may be particularly useful.

Suppose that $\lambda_i(t) = \gamma_i f(t)$, where γ_i are positive numbers and $f(t)$ is such function that

$$\int_0^t f(v)\, dv = C\, t(1 + o(1)) \quad \text{as } t \to \infty, \quad C > 0.$$

Then, as $t \to \infty$,

$$\begin{gathered}\int_0^t \lambda_i(v)\, dv \sim \gamma_i\, t, \quad \mu_{0,t} \sim C\left(\gamma_1 \log\frac{\gamma_1}{\gamma_0} + \gamma_0 - \gamma_1\right) t,\\ D_{0,t} \sim C\gamma_1\left(\log\frac{\gamma_1}{\gamma_0}\right)^2 t.\end{gathered} \tag{6.23}$$

We now show that

$$t^{-1}\widetilde{Z}_{u+t}^u \xrightarrow[t\to\infty]{\mathbf{P}_u-\text{completely}} 0 \quad \text{for all } u \geqslant 0. \tag{6.24}$$

To this end, we have to show that

$$\int_0^\infty \mathbf{P}_u\left\{|\widetilde{Z}_{u+t}^u| > \varepsilon t\right\} dt < \infty \quad \text{for all } \varepsilon > 0. \tag{6.25}$$

Let $m > 2$. By the Chebyshev inequality,

$$\mathbf{P}_u\left\{|\widetilde{Z}_{u+t}^u| > \varepsilon t\right\} \leqslant \frac{\mathbf{E}_u|\widetilde{Z}_{u+t}^u|^m}{\varepsilon^m t^m}.$$

Further, let $\Delta\widetilde{Z}_{u+t}^u = \widetilde{Z}_{u+t}^u - \lim_{v\uparrow t}\widetilde{Z}_{u+v}^u$ denote a jump of the processes $\{\widetilde{Z}_{u+v}^u\}$ at time $u + t$.

Under $\mathbf{P}_u$, the process $\widetilde{Z}^u_{u+t}$, $t \geqslant 0$, is a square integrable martingale with independent increments (with respect to $\{\mathcal{F}^X_t\}$). By the moment inequalities for martingales there exist universal constants C_m and c_m such that

$$\mathbf{E}_u|\widetilde{Z}^u_{u+t}|^m \leqslant C_m D^{m/2}_{u,u+t} + c_m \mathbf{E}_u|\Delta\widetilde{Z}^u_{u+t}|^m \quad \text{for any } m > 2 \tag{6.26}$$

(see, e.g., Lemma 5 in Tartakovsky, 1998b). Using (6.23) and (6.26) along with the fact that

$$\mathbf{E}_u|\Delta\widetilde{Z}^u_{u+t}|^m = \left|\log\frac{\lambda_1(u+t)}{\lambda_0(u+t)}\right|^m \mathbf{E}_u|\Delta X_{u+t}|^m \leqslant |\log(\gamma_1/\gamma_0)|^m,$$

we obtain that, as $t \to \infty$,

$$\mathbf{E}_u|\widetilde{Z}^u_{u+t}|^m \leqslant \widetilde{C}_m|\log(\gamma_1/\gamma_0)|^m\, t^{m/2}(1+o(1)) + c_m|\log\gamma_1/\gamma_0)|^m,$$

where $\widetilde{C}_m = CC_m$.

It follows that for large t and every $m > 2$

$$\mathbf{P}_u\left\{|\widetilde{Z}^u_{u+t}| > \varepsilon t\right\} \leqslant \frac{c_m|\log(\gamma_1/\gamma_0)|^m}{\varepsilon^m t^m} + \frac{\widetilde{C}_m|\log(\gamma_1/\gamma_0)|^m}{\varepsilon^m t^{m/2}}(1+o(1)), \tag{6.27}$$

which implies (6.25) and hence (6.24).

Thus, for every $u \geqslant 0$

$$\frac{1}{t}Z^u_{u+t} \xrightarrow[t\to\infty]{\mathbf{P}_u-\text{completely}} q = C\left(\gamma_1\log\frac{\gamma_1}{\gamma_0} + \gamma_0 - \gamma_1\right).$$

Moreover, since (6.27) holds for all $m > 2$,

$$\int_0^\infty t^{r-1}\mathbf{P}_u\left\{|\widetilde{Z}^u_{u+t}| > \varepsilon t\right\}dt < \infty \quad \text{for all } \varepsilon > 0 \text{ and } r \geqslant 1.$$

This implies that $t^{-1}Z^u_{u+t}$ converges $r-$quickly to q under $\mathbf{P}_u$ for all $r > 0$. Hence Theorem 3 applies to show that the Shiryaev procedure asymptotically minimizes all the positive moments of the detection delay.

Also, since the upper bound in (6.27) is independent of u,

$$\int_0^\infty d\pi_u\left(\int_0^\infty t^{r-1}\mathbf{P}_u\left\{\left|Z^u_{u+t} - qt\right| > \varepsilon t\right\}dt\right) < \infty.$$

If, for example, $f(t) = A^2\sin^2 t$, then the results are valid with $C = A^2/2$.

The above consideration can also be applied to the detection of the gradual change in the intensity when $\lambda_0(t) = \gamma_0$ and $\lambda_1(t) = \gamma_1 f(t-u)$, $t \geqslant u$ (conditioned on $\lambda = u$), where $f(t)$ is a monotone nondecreasing function such that $f(0) = \gamma_0/\gamma_1$ and $f(\infty) = 1$. The details are omitted.

7. REMARKS

1. The $r-$quick convergence conditions (3.11) and (3.12) are sufficient but not necessary for Theorems 2 and 3 to be true. In particular, proofs of these theorems show that the last entry time $T^{(u)}_\varepsilon$ in the corresponding conditions can be replaced by the one-sided (left-tail) version:

$$\tilde{T}^{(u)}_\varepsilon = \sup\left\{t : t^{-1}Z^u_{u+t} - q < -\varepsilon\right\},$$

i.e. it suffices to require $\mathbf{E}_u[\tilde{T}_\varepsilon^{(u)}]^r < \infty$ and $\mathbf{E}^\pi[\tilde{T}_\varepsilon^{(\lambda)}]^r < \infty$ for all $\varepsilon > 0$. We also conjecture that these last conditions can be further relaxed into

$$\lim_{t\to\infty} \mathbf{P}_u \left\{ t^{-1} Z_{u+t}^u - q \leqslant -\varepsilon \right\} = 0 \quad \forall \varepsilon > 0 \text{ and } u \geqslant 0.$$

However, we find it convenient and natural to formulate conditions in terms of rates of convergence in the strong law of large numbers.

2. Similar asymptotic optimality results can be obtained for a different false alarm constraint, $\mathbf{P}_\infty(\tau < \infty) \leqslant \alpha$. While for the CUSUM and SRP procedures the latter probability is equal to 1, it can be fixed at a given level $\alpha < 1$ for the Bayesian problem with proper prior distributions. Indeed, replacing the statistic Λ_t in the definition of the stopping time ν_B by the statistic

$$\tilde{\Lambda}_t = \int_0^\infty \frac{d\mathbf{P}_s}{d\mathbf{P}_\infty}(X^t) d\pi_s = \int_0^t \frac{d\mathbf{P}_s}{d\mathbf{P}_\infty}(X^t) d\pi_s + 1 - \pi_t,$$

denoting the corresponding stopping time by $\tilde{\nu}_B$, and noting that $\tilde{\Lambda}_t = \frac{d\mathbf{P}^\pi}{d\mathbf{P}_\infty}(X^t)$, we obtain

$$\mathbf{P}_\infty(\tilde{\nu}_B < \infty) = \mathbf{E}^\pi(1/\tilde{\Lambda}_{\tilde{\nu}_B} \mathbb{1}_{\{\tilde{\nu}_B < \infty\}}) \leqslant 1/B.$$

Therefore, setting $B = B_\alpha = 1/\alpha$ guarantees $\mathbf{P}_\infty(\tilde{\nu}_B < \infty) \leqslant \alpha$.

Further details will be given elsewhere. Here we only note that Beibel (2000) considered a purely Bayesian problem with the risk function

$$\mathcal{R}(c, \tau) = \mathbf{P}_\infty(\tau < \infty) + cq\mathbf{E}^\pi(\tau - \lambda)^+,$$

for detecting changes in a drift of the Brownian motion (as $c \to 0$) for the composite post-change hypothesis. For the case of the simple hypothesis, the results of Beibel (2000) can be generalized for arbitrary continuous-time models whenever the complete convergence condition (3.6) holds using the method of the present paper.

APPENDIX: PROOFS

Derivation of the lower bounds in Theorem 1 is based of the Chebyshev inequality that involves certain probabilities. In the following lemma we establish that these probabilities approach 0 when $\alpha \to 0$ under extremely weak conditions.

Define

$$\begin{aligned} L_\alpha(q_\rho) &= \frac{|\log \alpha|}{q_\rho}, \quad \gamma_{\varepsilon,\alpha}(\tau) = \mathbf{P}^\pi \left\{ \lambda \leqslant \tau \leqslant \lambda + (1-\varepsilon) L_\alpha(q_\rho) \right\}, \\ \gamma_{\varepsilon,\alpha}^{(u)}(\tau) &= \mathbf{P}_u \left\{ u \leqslant \tau \leqslant u + (1-\varepsilon) L_\alpha(q_\rho) \right\}, \end{aligned} \tag{7.1}$$

where $0 < \varepsilon < 1$.

Lemma 1. *Assume that for some $q > 0$ condition* (3.1) *holds. Then, for every $0 < \varepsilon < 1$ and $u \geqslant 0$,*

$$\lim_{\alpha\to 0} \sup_{\tau \in \boldsymbol{\Delta}(\alpha)} \gamma_{\varepsilon,\alpha}^{(u)}(\tau) = 0 \tag{7.2}$$

and for all $0 < \varepsilon < 1$

$$\lim_{\alpha\to 0} \sup_{\tau \in \boldsymbol{\Delta}(\alpha)} \gamma_{\varepsilon,\alpha}(\tau) = 0. \tag{7.3}$$

Proof. Despite the fact that the proof of this lemma is almost identical to the proof of Lemma 1 in Tartakovsky and Veeravalli (2005) (for the discrete-time case), we present the proof for the sake of completeness.

(i) Let $v_\alpha = v_\alpha(u) = u + (1-\varepsilon)L_\alpha(q_\rho)$. By changing measures from $\mathbf{P}_\infty$ to $\mathbf{P}_u$, we obtain that for any $C > 0$ and $\varepsilon \in (0,1)$

$$\begin{aligned}
&\mathbf{P}_\infty\{u \leqslant \tau \leqslant v_\alpha\} = \mathbf{E}_\infty\left\{\mathbb{1}_{\{u\leqslant\tau\leqslant v_\alpha\}}\right\} = \mathbf{E}_u\left\{\mathbb{1}_{\{u\leqslant\tau\leqslant v_\alpha\}}e^{-Z^u_\tau}\right\} \\
&\geqslant \mathbf{E}_u\left\{\mathbb{1}_{\{u\leqslant\tau\leqslant v_\alpha, Z^u_\tau<C\}}e^{-Z^u_\tau}\right\} \geqslant e^{-C}\mathbf{P}_u\left\{u \leqslant \tau \leqslant v_\alpha, \sup_{0\leqslant t\leqslant v_\alpha-u} Z^u_{u+t} < C\right\} \\
&\geqslant e^{-C}\left[\mathbf{P}_u\{u\leqslant\tau\leqslant v_\alpha\} - \mathbf{P}_u\left\{\sup_{0\leqslant t\leqslant v_\alpha-u} Z^u_{u+t} \geqslant C\right\}\right].
\end{aligned} \tag{7.4}$$

Taking $C = (1-\varepsilon^2)qL_\alpha(q_\rho) = q(1+\varepsilon)(v_\alpha - u)$, we get from (7.4) that

$$\gamma^{(u)}_{\varepsilon,\alpha}(\tau) = \mathbf{P}_u\{u\leqslant\tau\leqslant v_\alpha\} \leqslant p_u(\alpha,\varepsilon) + \beta_u(\alpha,\varepsilon), \tag{7.5}$$

where

$$p_u(\alpha,\varepsilon) = e^C\mathbf{P}_\infty\{u\leqslant\tau\leqslant v_\alpha\}$$

and

$$\beta_u(\alpha,\varepsilon) = \mathbf{P}_u\left\{\sup_{0\leqslant t\leqslant v_\alpha-u} Z^u_{u+t} \geqslant q(1+\varepsilon)(v_\alpha-u)\right\}.$$

By condition (3.1), for every $0<\varepsilon<1$,

$$\beta_u(\alpha,\varepsilon) = \mathbf{P}_u\left\{\frac{1}{v_\alpha-u}\sup_{0\leqslant t\leqslant v_\alpha-u} Z^u_{u+t} \geqslant (1+\varepsilon)q\right\} \xrightarrow[\alpha\to 0]{} 0 \tag{7.6}$$

because $v_\alpha - u = (1-\varepsilon)|\log\alpha|/q_\rho \to \infty$ as $\alpha\to 0$.

Next, we observe that

$$\begin{aligned}
\alpha &\geqslant \mathrm{P}_{\mathrm{FA}}(\tau) = \mathbf{P}^\pi(\tau<\lambda) \geqslant \mathbf{P}(\{\tau\leqslant v\}\cap\{\lambda>v\}) \\
&= \mathbf{P}(\tau\leqslant v|\lambda>v)\mathbf{P}(\lambda>v) = \mathbf{P}_\infty(\tau\leqslant v)(1-\pi_v)
\end{aligned}$$

for any $\tau\in\boldsymbol{\Delta}(\alpha)$ and $v\geqslant 0$, from where

$$\mathbf{P}_\infty(\tau\leqslant v_\alpha) \leqslant \frac{\alpha}{1-\pi_{v_\alpha}}. \tag{7.7}$$

It follows that

$$p_u(\alpha,\varepsilon) \leqslant e^C\mathbf{P}_\infty\{\tau\leqslant v_\alpha\} \leqslant e^{q(1+\varepsilon)(v_\alpha-u)}\frac{\alpha}{1-\pi_{v_\alpha}} = \tilde{p}_u(\alpha,\varepsilon). \tag{7.8}$$

To prove (7.2), it remains to show that

$$\tilde{p}_u(\alpha,\varepsilon)\to 0 \quad \text{as } \alpha\to 0 \text{ for all } u\geqslant 0 \text{ and } 0<\varepsilon<1. \tag{7.9}$$

From definitions of v_α and L_α, we have

$$\alpha = e^{-q_\rho L_\alpha(q_\rho)} = \exp\left\{-q_\rho\left(\frac{v_\alpha-u}{1-\varepsilon}\right)\right\},$$

and from (7.8),

$$\tilde{p}_u(\alpha,\varepsilon)=\frac{e^{q(1+\varepsilon)(v_\alpha-u)}}{1-\pi_{v_\alpha}}\exp\left\{-q_\rho\left(\frac{v_\alpha-u}{1-\varepsilon}\right)\right\}=\frac{\exp\left\{-\frac{q\varepsilon^2+\rho}{1-\varepsilon}(v_\alpha-u)\right\}}{1-\pi_{v_\alpha}}. \tag{7.10}$$

It follows that

$$\lim_{\alpha\to 0}\frac{\log\tilde{p}_u(\alpha,\varepsilon)}{v_\alpha-u}=-\frac{q\varepsilon^2+\rho}{1-\varepsilon}-\lim_{\alpha\to 0}\frac{\log(1-\pi_{v_\alpha})}{v_\alpha-u},$$

where the latter limit is equal to ρ for $\pi\in\mathcal{E}(\rho)$ and 0 for $\pi\in F$. Therefore, for $\pi\in F\cup\mathcal{E}(\rho)$

$$\lim_{\alpha\to 0}\frac{\log\tilde{p}_u(\alpha,\varepsilon)}{v_\alpha-u}=-\frac{\varepsilon(q\varepsilon+\rho)}{1-\varepsilon},$$

completing the proof of (7.9).

Therefore, we obtain that for every $\tau\in\mathbf{\Delta}(\alpha)$ and $\varepsilon>0$,

$$\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\leqslant\tilde{p}_u(\alpha,\varepsilon)+\beta_u(\alpha,\varepsilon), \tag{7.11}$$

where both $\tilde{p}_u(\alpha,\varepsilon)$ and $\beta_u(\alpha,\varepsilon)$ do not depend on a particular stopping time τ. Since by (7.6) and (7.9), β_u and $\tilde{p}_u$ converge to zero as $\alpha\to 0$, we obtain

$$\sup_{\tau\in\mathbf{\Delta}(\alpha)}\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\to 0\quad\text{as }\alpha\to 0,$$

i.e. (7.2) follows for arbitrary $0<\varepsilon<1$.

(ii) Obviously, for every $0<\varepsilon<1$,

$$\gamma_{\varepsilon,\alpha}(\tau)=\int_0^\infty\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\,d\pi_u\leqslant\int_0^{\varepsilon L_\alpha(q_\rho)}\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\,d\pi_u+1-\pi_{\varepsilon L_\alpha(q_\rho)}, \tag{7.12}$$

where $1-\pi_{\varepsilon L_\alpha(q_\rho)}\to 0$ as $\alpha\to 0$.

Using (7.11), we have

$$\int_0^{\varepsilon L_\alpha(q_\rho)}\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\,d\pi_u\leqslant\int_0^{\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)d\pi_u+\int_0^\infty\beta_u(\alpha,\varepsilon)d\pi_u. \tag{7.13}$$

The second term on the right-hand side,

$$\int_0^\infty\beta_u(\alpha,\varepsilon)d\pi_u\to 0\quad\text{as }\alpha\to 0$$

by the condition (3.1) and the Lebesgue dominated convergence theorem.

By (7.10),

$$\begin{aligned}\int_0^{\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)d\pi_u&\leqslant\sup_{u\leqslant\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)=\frac{\exp\{(q\varepsilon^2+\rho)L_\alpha(q_\rho)\}}{\inf_{u\leqslant\varepsilon L_\alpha(q_\rho)}(1-\pi_{v_\alpha(u)})}\\&=\frac{\exp\{(q\varepsilon^2+\rho)L_\alpha(q_\rho)\}}{1-\pi_{v_\alpha(\varepsilon L_\alpha(q_\rho))}}=\frac{\exp\{(q\varepsilon^2+\rho)L_\alpha(q_\rho)\}}{1-\pi_{L_\alpha(q_\rho)}},\end{aligned}$$

where the latter equality follows from the fact that $v_\alpha(u)=L_\alpha(q_\rho)$ for $u=\varepsilon L_\alpha(q_\rho)$. Thus,

$$\frac{\log\int_0^{\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)d\pi_u}{L_\alpha(q_\rho)}\leqslant-(q\varepsilon^2+\rho)-\frac{\log\left(1-\pi_{L_\alpha(q_\rho)}\right)}{L_\alpha(q_\rho)}.$$

Since

$$\lim_{\alpha\to 0}\frac{\log\left(1-\pi_{L_\alpha(q_\rho)}\right)}{L_\alpha(q_\rho)}=-\rho,$$

we obtain that for any $\pi\in F\cup\mathcal{E}(\rho)$

$$\lim_{\alpha\to 0}\frac{\log\int_0^{\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)d\pi_u}{L_\alpha(q_\rho)}\leqslant -q\varepsilon^2,$$

which shows that for any $\pi\in F\cup\mathcal{E}(\rho)$,

$$\int_0^{\varepsilon L_\alpha(q_\rho)}\tilde{p}_u(\alpha,\varepsilon)d\pi_u\to 0\quad\text{as } \alpha\to 0.$$

We showed that all terms in the right-hand sides of inequalities (7.12) and (7.13) converge to zero as $\alpha\to 0$. Since none of these terms depends on τ, (7.3) follows, and the proof is complete. □

Proof of Theorem 1. (i) By the Chebyshev inequality, for any $m>0$,

$$\mathbf{E}_u[(\tau-u)^+]^m\geqslant(1-\varepsilon)^m L_\alpha^m(q_\rho)\mathbf{P}_u\left\{\tau>u+(1-\varepsilon)L_\alpha(q_\rho)\right\}.$$

Obviously,

$$\mathbf{P}_u\left\{\tau>u+(1-\varepsilon)L_\alpha(q_\rho)\right\}=\mathbf{P}_u\left\{\tau\geqslant u\right\}-\gamma_{\varepsilon,\alpha}(\tau).$$

It follows that

$$\begin{aligned}\mathbf{E}_u\left\{(\tau-u)^m|\tau\geqslant u\right\}&=\frac{\mathbf{E}_u[(\tau-u)^+]^m}{\mathbf{P}_u\left\{\tau\geqslant u\right\}}\\&\geqslant(1-\varepsilon)^m L_\alpha^m(q_\rho)\left[1-\frac{\gamma^{(u)}_{\varepsilon,\alpha,q_\rho}(\tau)}{\mathbf{P}_u\left\{\tau\geqslant u\right\}}\right].\end{aligned}\tag{7.14}$$

By (7.7), for any $\tau\in\mathbf{\Delta}(\alpha)$ and $\alpha<1/(1-\pi_u)$,

$$\mathbf{P}_u\left\{\tau\geqslant u\right\}=1-\mathbf{P}_\infty\left\{\tau<u\right\}\geqslant 1-\alpha(1-\pi_u)^{-1},\tag{7.15}$$

and therefore, for every $\tau\in\mathbf{\Delta}(\alpha)$, $0<\varepsilon<1$ and $\alpha<1/(1-\pi_u)$,

$$\mathbf{E}_u\left\{(\tau-u)^m|\tau\geqslant u\right\}\geqslant(1-\varepsilon)^m L_\alpha^m(q_\rho)\left[1-\frac{\gamma^{(u)}_{\varepsilon,\alpha}(\tau)}{1-\alpha(1-\pi_u)^{-1}}\right].\tag{7.16}$$

By (7.2), $\gamma^{(u)}_{\varepsilon,\alpha}(\tau)\to 0$ as $\alpha\to 0$ uniformly over all $\tau\in\mathbf{\Delta}(\alpha)$, which implies

$$\inf_{\tau\in\mathbf{\Delta}(\alpha)}\mathbf{E}_u\left\{(\tau-u)^m|\tau\geqslant u\right\}\geqslant(1-\varepsilon)^m L_\alpha^m(q_\rho)(1+o(1))\quad\text{as } \alpha\to 0.$$

Since this inequality holds for arbitrary $0<\varepsilon<1$, the lower bound (3.3) follows.

(ii) Similar to (7.14),

$$\mathbf{E}^\pi\left\{(\tau-\lambda)^m|\tau\geqslant\lambda\right\}\geqslant(1-\varepsilon)^m L_\alpha^m(q_\rho)\left[1-\frac{\gamma_{\varepsilon,\alpha}(\tau)}{\mathbf{P}\left\{\tau\geqslant\lambda\right\}}\right].\tag{7.17}$$

Since for any $\tau\in\mathbf{\Delta}(\alpha)$, $\mathbf{P}^\pi\left\{\tau\geqslant\lambda\right\}\geqslant 1-\alpha$, it follows that for every $\tau\in\mathbf{\Delta}(\alpha)$ and $0<\varepsilon<1$

$$\mathbf{E}^\pi\left\{(\tau-\lambda)^m|\tau\geqslant\lambda\right\}\geqslant[(1-\varepsilon)L_\alpha(q_\rho)]^m\left[1-\frac{\gamma_{\varepsilon,\alpha}(\tau)}{1-\alpha}\right].$$

Since ε is arbitrary and, by (7.3), $\sup_{\tau\in\mathbf{\Delta}(\alpha)}\gamma_{\varepsilon,\alpha}(\tau)\to 0$ as $\alpha\to 0$, asymptotic lower bound (3.4) follows for arbitrary $m>0$.

□

Proof of Theorem 2. (i) We begin with obtaining the asymptotic lower bound for $\mathrm{CADD}_u(\nu_B)$ for large values of B. To this end, define

$$\gamma^{(u)}_{\varepsilon,B}(\nu_B) = \mathbf{P}_u\left\{u \leqslant \nu_B \leqslant u + (1-\varepsilon)q_\rho^{-1}\log B\right\}$$

(cf. (7.1)). Replacing α by $1/B$ in Lemma 1 shows that

$$\gamma^{(u)}_{\varepsilon,B}(\nu_B) \to 0 \quad \text{as } B \to \infty \text{ for all } u \geqslant 0. \tag{7.18}$$

Similar to (7.16),

$$\mathrm{CADD}_u(\nu_B) \geqslant (1-\varepsilon)\frac{\log B}{q_\rho}\left(1 - \frac{\gamma^{(u)}_{\varepsilon,B}(\nu_B)}{\mathbf{P}_u\left\{\nu_B \geqslant u\right\}}\right),$$

where, for $B > \pi_u/(1-\pi_u)$,

$$\mathbf{P}_u\left\{\nu_B \geqslant u\right\} = 1 - \mathbf{P}_\infty\left\{\nu_B < u\right\} \geqslant 1 - \frac{1}{(1+B)(1-\pi_u)}. \tag{7.19}$$

Since ε is arbitrary, it follows that

$$\mathrm{CADD}_u(\nu_B) \geqslant \frac{\log B}{q_\rho}(1+o(1)) \quad \text{as } B \to \infty. \tag{7.20}$$

To complete the proof of (3.7) it suffices to show that the right side of (7.20) is also the upper bound (asymptotically).

For any $u \in (0,t)$,

$$\Lambda_t = \frac{1}{1-\pi_t}\int_0^t e^{Z_t^s}\,d\pi_s \geqslant \frac{1}{1-\pi_t}\int_u^t e^{Z_t^s}\,d\pi_s = \frac{e^{Z_t^u}}{1-\pi_t}\int_u^t e^{-Z_u^s}\,d\pi_s. \tag{7.21}$$

Choose arbitrary $\varepsilon \in (0,q)$ and $\delta > 0$. Then, substituting $t = \nu_B - \delta$ in (7.21), we obtain

$$\log B > \log \Lambda_{\nu_B-\delta} \geqslant Z^u_{\nu_B-\delta} - \log(1-\pi_{\nu_B-\delta}) + Y_u(\pi). \tag{7.22}$$

where $Y_u(\pi) = \log\int_u^{\nu_B-\delta} e^{-Z_u^s}\,d\pi_s$.

Next, (2.9) guarantees existence of $k > 1$, $N > 0$, and $\gamma > 0$ such that

$$1 - \pi_{kt} \leqslant (1-\gamma)(1-\pi_t) \quad \text{for all } t \geqslant N. \tag{7.23}$$

If $\pi \in \mathcal{E}(\rho)$, choose arbitrary $k > 1$, $a \in (0,\rho)$, and $b \in (\rho,\infty)$ and define

$$T^* = \sup\left\{t \geqslant N : \ -t^{-1}\log(1-\pi_t) \notin [a,b]\right\} \tag{7.24}$$

and

$$U = U_{u,\varepsilon} = u + T^{(u)}_\varepsilon + T^*. \tag{7.25}$$

For $\pi \in F$, set $T^* = N$.

Further, let $A_{u,\varepsilon}$ be the following event:

$$A_{u,\varepsilon} = \left\{\nu_B - \delta \geqslant kU_{u,\varepsilon}\right\}.$$

On the set $A_{u,\varepsilon}$, three terms in the right-hand side of (7.22) can be bounded as follows.

$$Z^u_{\nu_B-\delta} \geqslant (q-\varepsilon)(\nu_B - \delta - u) \text{ because } \nu_B - \delta - u > T^{(u)}_\varepsilon;$$

$$-\log(1-\pi_{\nu_B-\delta}) \geqslant a(\nu_B-\delta) \text{ for } \pi\in\mathcal{E}(\rho) \text{ because } \nu_B-\delta > T^*;$$

$$\begin{aligned}
Y_u(\pi) &= \int_u^{\nu_B-\delta} e^{-Z_s^u}\,d\pi_s \\
&\geqslant \log\int_U^{kU} e^{-Z_s^u}\,d\pi_s && (7.26)\\
&\geqslant \log\int_U^{kU} e^{-(q+\varepsilon)(s-u)}\,d\pi_s && (7.27)\\
&\geqslant \log\left\{(\pi_{kU}-\pi_U)e^{-(q+\varepsilon)(kU-u)}\right\} && (7.28)\\
&= \log\{(1-\pi_U)-(1-\pi_{kU})\}-(q+\varepsilon)(kU-u)\\
&\geqslant \log\{(1-\pi_U)-(1-\gamma)(1-\pi_U)\}-(q+\varepsilon)(kU-u) && (7.29)\\
&= \log\gamma-(q+\varepsilon)(kU-u)+\log(1-\pi_U)\\
&\geqslant \log\gamma-(q+\varepsilon)(kU-u)+\begin{cases} -bU & \text{if} \quad \pi\in\mathcal{E}(\rho)\\ \log(1-\pi_{\nu_B-\delta}) & \text{if} \quad \pi\in F.\end{cases} && (7.30)
\end{aligned}$$

In this derivation, (7.26) holds because $u \leqslant U$ by (7.25) and $\nu_B-\delta \geqslant kU$ on $A_{u,\varepsilon}$; (7.27) holds because $U \geqslant u+T_\varepsilon^{(u)}$ and $Z_s^u \leqslant (q+\varepsilon)(s-u)$ for all $s \geqslant u+T_\varepsilon^{(u)}$; (7.28) is obtained replacing the integrand in (7.27) by its lowest value on $[U;kU]$; (7.29) follows from (7.23), which is applicable since $kU > U > T^* \geqslant N$; and (7.30) follows from (7.24) for $\pi\in\mathcal{E}(\rho)$ because $U > T^*$ and from the inequality $U < kU \leqslant \nu_B-\delta$ on $A_{u,\varepsilon}$ for $\pi\in F$.

Using these inequalities in (7.22), we obtain for $\pi\in F$,

$$\log B > (q-\varepsilon)(\nu_B-\delta-u)+\log\gamma-(q+\varepsilon)(kU-u) \quad \text{on } A_{u,\varepsilon},$$

so that

$$\nu_B-u \leqslant \left(\frac{\log B-\log\gamma+(q+\varepsilon)(kU-u)}{q-\varepsilon}+\delta\right)\mathbb{1}_{\{A_{u,\varepsilon}\}}+(kU+\delta-u)\mathbb{1}_{\{\bar{A}_{u,\varepsilon}\}}. \tag{7.31}$$

Since T^* is not random, by condition (3.5), $\mathbf{E}_u U<\infty$. Hence,

$$\mathbf{E}_u(\nu_B-u)^+ \leqslant \frac{\log B}{q-\varepsilon}+O(1) \quad \text{as } B\to\infty. \tag{7.32}$$

Similarly, for $\pi\in\mathcal{E}(\rho)$,

$$\log B > (q-\varepsilon)(\nu_B-\delta-u)+a(\nu_B-\delta)+\log\gamma-(q+\varepsilon)(kU-u)-bU \quad \text{on } A_{u,\varepsilon},$$

so that

$$\nu_B-u \leqslant \left(\frac{\log B-\log\gamma+\{k(q+\varepsilon)+b\}U-(q+a+\varepsilon)u}{q+a-\varepsilon}+\delta\right)\mathbb{1}_{\{A_{u,\varepsilon}\}}+(kU+\delta-u)\mathbb{1}_{\{\bar{A}_{u,\varepsilon}\}}, \tag{7.33}$$

which implies

$$\mathbf{E}_u(\nu_B-u)^+ \leqslant \frac{\log B}{q+a-\varepsilon}+O(1) \quad \text{as } B\to\infty. \tag{7.34}$$

Since (7.32) and (7.34) hold for arbitrary $\varepsilon>0$ and $a<\rho$, part (i) of the theorem follows.

(ii) To prove (3.9) and (3.10), we take expectations in (7.33) and (7.31) with respect to the measure $\mathbf{P}^\pi$ and obtain, from the arbitrary choice of ε and a, that

$$\mathbf{E}^\pi(\nu_B - \lambda)^+ \leqslant \frac{\log B}{q_\rho} + O(1) \quad \text{as } B \to \infty.$$

The lower bound can be obtained similar to (7.20). In fact, see the proof of (7.36) below.

Combining this with Theorem 1 completes the proof.

□

Proof of Theorem 3. We give only a proof of (3.15) and (3.17). The proof of (3.14) and (3.16) is similar and omitted.

Replacing α with $1/B$ in (7.1), i.e. setting

$$L_B(q_\rho) = \frac{\log B}{q_\rho} \quad \text{and} \quad \gamma_{\varepsilon,B}(\nu_B) = \mathbf{P}^\pi\left\{\lambda \leqslant \nu_B \leqslant \lambda + (1-\varepsilon)L_B(q_\rho)\right\},$$

and using the change-of-measure argument identical to that exploited in the proof of Lemma 1 shows that

$$\gamma_{\varepsilon,B}(\nu_B) \to 0 \quad \text{as } B \to \infty. \tag{7.35}$$

In analogy with (7.17),

$$\mathbf{E}^\pi[(\nu_B - \lambda)^m | \nu_B \geqslant \lambda] \geqslant [(1-\varepsilon)L_B(q_\rho)]^m \left[1 - \frac{\gamma_{\varepsilon,B}(\nu_B)}{\mathbf{P}^\pi\left\{\nu_B \geqslant \lambda\right\}}\right],$$

where $\mathbf{P}^\pi\left\{\nu_B \geqslant \lambda\right\} \geqslant B/(1+B) \to 1$. Using this last inequality along with (7.35) shows that, for an arbitrary $0 < \varepsilon < 1$ and all $m > 0$,

$$\mathbf{E}^\pi\left\{(\nu_B - \lambda)^m | \nu_B \geqslant \lambda\right\} \geqslant \left[(1-\varepsilon)\frac{\log B}{q_\rho}\right]^m (1 + o(1)) \quad \text{as } B \to \infty,$$

which proves the asymptotic lower bound

$$\mathbf{E}^\pi\left\{(\nu_B - \lambda)^m | \nu_B \geqslant \lambda\right\} \geqslant \left(\frac{\log B}{q_\rho}\right)^m (1 + o(1)) \quad \text{as } B \to \infty. \tag{7.36}$$

For the upper bound, we notice that $U = U_{u,\varepsilon} = u + T_\varepsilon^{(u)} + const$ in the proof of Theorem 2 (see (7.25)). Therefore, assumptions (3.11) and (3.12) remain valid with $T_\varepsilon^{(u)}$ replaced by U. Hence, one can take m-th moments in (7.31) and (7.33), and the upper bound for $\mathbf{E}^\pi\{(\nu_B - \lambda)^m | \nu_B \geqslant \lambda\}$ follows. Combining it with (7.36) yields (3.15).

Finally, setting $B = (1-\alpha)/\alpha$ in (3.15) and using Theorem 1 completes the proof of (3.17).

□

Proof of Theorem 4. (i) Obviously, for every $\delta > 0$

$$\frac{\log \Lambda_{\nu_B - \delta}}{\nu_B} \leqslant \frac{\log B}{\nu_B} \leqslant \frac{\log \Lambda_{\nu_B}}{\nu_B}. \tag{7.37}$$

Consider $u = 0$. The statistic $\log \Lambda_t$ can be written as

$$\log \Lambda_t = Z_t^0 + \log \int_0^t \exp\{-Z_s^0\} d\pi_s - \log(1 - \pi_t).$$

Since $t^{-1}\log(1-\pi_t)\to\rho$ for all $\pi\in\mathcal{E}(\rho)\cup F$, by the almost sure convergence condition (3.18),

$$\frac{1}{t}\log\Lambda_t \xrightarrow[t\to\infty]{\mathbf{P}_0-\text{a.s.}} q_\rho .$$

Therefore, both the left and right sides in the inequalities (7.37) converge a.s.-$\mathbf{P}_0$ to q_ρ, which implies (3.19) for $u=0$. For $u>0$ the proof is essentially the same. Convergence (3.20) follows from (3.19).

(ii) By Lemma 1, for every $0<\varepsilon<1$

$$\begin{aligned}
&\lim_{\alpha\to 0}\inf_{\tau\in\boldsymbol{\Delta}(\alpha)}\mathbf{P}_u\left\{\tau-u\geqslant\varepsilon q_\rho^{-1}|\log\alpha|\right\}=1,\\
&\lim_{\alpha\to 0}\inf_{\tau\in\boldsymbol{\Delta}(\alpha)}\mathbf{P}^\pi\left\{\tau-\lambda\geqslant\varepsilon q_\rho^{-1}|\log\alpha|\right\}=1.
\end{aligned}\tag{7.38}$$

From (3.19), (3.20), and (7.38), the asymptotic relations (3.21) and (3.22) follow immediately.

□

Proof of Theorem 5. Obviously, for any stopping time τ

$$\mathcal{R}_{\rho,c,m}(\tau)=\mathrm{P}_{\mathrm{FA}}(\tau)+c[1-\mathrm{P}_{\mathrm{FA}}(\tau)]\mathbf{E}^\pi[(\tau-\lambda)^m|\tau\geqslant\lambda].$$

Since $\mathrm{P}_{\mathrm{FA}}(\nu_B)\leqslant 1/B$ and, by Theorem 3,

$$\mathbf{E}^\pi[(\nu_B-\lambda)^m|\nu_B\geqslant\lambda]=\left(\frac{\log B}{q_\rho}\right)^m(1+o(1))\quad\text{as } B\to\infty,$$

the asymptotic relation (3.24) follows.

In the case of $m=1$, the threshold $B(c,1)=q_\rho/c$ and we immediately obtain

$$\mathcal{R}_{\rho,c,1}(\nu_{B(c,1)})\sim c\frac{|\log c|}{q_\rho}\quad\text{as } c\to 0.$$

Next, it is easily seen that, for any $m>1$, the threshold $B(c,m)$ goes to infinity as $c\to 0$ in such a way that $\log B(c,m)\sim|\log c|$. Moreover, since $cB(c,m)(\log B(c,m))^{m-1}=q_\rho^m/m$, it follows that

$$cB(c,m)(\log B(c,m))^m\sim(q_\rho^m/m)|\log c|\quad\text{as } c\to 0.$$

Therefore,

$$\mathcal{R}_{\rho,c,m}(\nu_{B(c,m)})\sim c\left(\frac{|\log c|}{q_\rho}\right)^m(1+1/[cB(c,m)(\log B(c,m))^m])\sim c\left(\frac{|\log c|}{q_\rho}\right)^m\quad\text{as } c\to 0.$$

It remains to prove that

$$\inf_{\tau\in\mathcal{T}}\mathcal{R}_{\rho,c,m}(\tau)\geqslant c\left(\frac{|\log c|}{q_\rho}\right)^m(1+o(1)),$$

where $\mathcal{T}$ is a set of all stopping rules. Moreover, since

$$\lim_{c\to o}\frac{g_{c,m}(B(c,m))}{c(|\log c|/q_\rho)^m}=1,$$

it suffices to prove that

$$\liminf_{c\to 0}\frac{\inf_{\tau\in\mathcal{T}}\mathcal{R}_{\rho,c,m}(\tau)}{g_{c,m}(B(c,m))}\geqslant 1.\tag{7.39}$$

The proof is conducted by contradiction. Suppose that (7.39) is not true, that is, there exist stopping rules $\tau = \tau_c$ such that

$$\limsup_{c\to 0} \frac{\mathcal{R}_{\rho,c,m}(\tau_c)}{g_{c,m}(B(c,m))} < 1. \tag{7.40}$$

Let $\alpha_c = \mathrm{P}_{\mathrm{FA}}(\tau_c)$. Since

$$\alpha_c \leqslant \mathcal{R}_{\rho,c,m}(\tau_c) < g_{c,m}(B(c,m))(1+o(1)) \to 0 \quad \text{as } c \to 0$$

and, by Theorem 1,

$$\mathbf{E}^{\pi}[(\tau_c-\lambda)^m | \tau_c \geqslant \lambda] \geqslant \left(\frac{\log(1/\alpha_c)}{q_\rho}\right)^m (1+o(1)),$$

it follows that

$$\mathcal{R}_{\rho,c,m}(\tau_c) \geqslant \alpha_c + \left(\frac{\log(1/\alpha_c)}{q_\rho}\right)^m (1+o(1)) \quad \text{as } c \to 0.$$

Thus,

$$\frac{\mathcal{R}_{\rho,c,m}(\tau_c)}{g_{c,m}(B(c,m))} \geqslant \frac{g_{c,m}(1/\alpha_c)}{\min_{b>0} g_{c,m}(b)}(1+o(1)) \geqslant 1+o(1),$$

which contradicts (7.40). Hence, (7.39) is proven. This completes the proof of the theorem.

□

Proof of Theorem 7. Write $b = \log B$ and define

$$\hat{\gamma}^{(u)}_{\varepsilon,B} = \mathbf{P}_u\left\{u \leqslant \hat{\nu}_B \leqslant u + (1-\varepsilon)q^{-1}b\right\}.$$

Similar to (7.5),

$$\hat{\gamma}^{(u)}_{\varepsilon,B} \leqslant p_u(\varepsilon,B) + \beta_u(\varepsilon,B),$$

where

$$p_u(\varepsilon,B) = B^{1-\varepsilon^2}\mathbf{P}_\infty\left\{u \leqslant \hat{\nu}_B \leqslant u + (1-\varepsilon)q^{-1}b\right\},$$

$$\beta_u(\varepsilon,B) = \mathbf{P}_u\left\{\frac{1}{(1-\varepsilon)q^{-1}b}\sup_{0\leqslant t<(1-\varepsilon)q^{-1}b} Z^u_{u+t} \geqslant (1+\varepsilon)q\right\}.$$

Since by conditions of the theorem Z^u_{u+t}/t converges $\mathbf{P}_u$-completely to q, the probability $\beta_u(\varepsilon,B)$ goes to 0 as $B \to \infty$.

Next, by (5.1),

$$p_u(\varepsilon,B) \leqslant \frac{u + (1-\varepsilon)q^{-1}\log B}{B^{\varepsilon^2}} \xrightarrow[B\to\infty]{} 0$$

for all $u \geqslant 0$ and $0 < \varepsilon < 1$. Therefore,

$$\hat{\gamma}^{(u)}_{\varepsilon,B} \xrightarrow[B\to\infty]{} 0 \quad \text{for all } u \geqslant 0 \text{ and } 0 < \varepsilon < 1.$$

Applying Markov's inequality (see proof of Theorem 1), yields the asymptotic lower bound

$$\mathrm{CADD}_u(\hat{\nu}_B) \geqslant q^{-1}\log B(1+o(1)) \quad \text{as } B \to \infty. \tag{7.41}$$

Using estimate (7.13), we also obtain that

$$\mathbf{P}^{\pi}\left\{\lambda \leqslant \hat{\nu}_B \leqslant \lambda + (1-\varepsilon)q^{-1}b\right\} \xrightarrow[B\to\infty]{} 0,$$

which along with the Markov inequality shows that (7.41) holds for $\mathrm{ADD}(\hat{\nu}_B)$.

To obtain the upper bound, write

$$R_t = \int_0^t e^{Z_t^s} ds \geqslant \int_u^t e^{Z_t^s} ds = e^{Z_t^u} \int_u^t e^{-Z_s^u} ds \geqslant e^{Z_t^u + Y_t^u},$$

where $Y_t^u = \log \int_u^t e^{-Z_s^u} ds$.

Obviously,

$$\log B > \log R_{\hat{\nu}_B - \delta} \geqslant Z^u_{\hat{\nu}_B - \delta} + Y^u_{\hat{\nu}_B - \delta}$$

and $Z^u_{\hat{\nu}_B - \delta} > (q - \varepsilon)(\hat{\nu}_B - \delta - u)$ on the set $A_{u,\varepsilon} = \{\hat{\nu}_B - \delta - u > T_\varepsilon^{(u)} + 1\}$. Hence,

$$\begin{aligned} \hat{\nu}_B - \delta - u &= (\hat{\nu}_B - \delta - u)\mathbb{1}_{\{\bar{A}_{u,\varepsilon}\}} + (\hat{\nu}_B - \delta - u)\mathbb{1}_{\{A_{u,\varepsilon}\}} \\ &\leqslant T_\varepsilon^{(u)} + 1 + \frac{\log B - Y^u_{\hat{\nu}_B - \delta}}{q - \varepsilon}\mathbb{1}_{\{A_{u,\varepsilon}\}}. \end{aligned}$$

It remains to bound $Y^u_{\hat{\nu}_B - \delta}\mathbb{1}_{\{A_{u,\varepsilon}\}}$ from below. To this end, we observe that if $A_{u,\varepsilon}$ occurs, then

$$Y^u_{\hat{\nu}_B - \delta} = \log \int_u^{\hat{\nu}_B - \delta} e^{-Z_s^u} ds \geqslant \log \int_{T_\varepsilon^u + u}^{T_\varepsilon^u + u + 1} e^{-Z_s^u} ds.$$

Since $Z_s^u < (q + \varepsilon)(s - u)$ for $s > u + T_\varepsilon^{(u)}$, we obtain

$$\begin{aligned} Y^u_{\hat{\nu}_B - \delta}\mathbb{1}_{\{A_{u,\varepsilon}\}} &\geqslant \log \int_{T_\varepsilon^{(u)} + u}^{T_\varepsilon^{(u)} + u + 1} e^{-(q+\varepsilon)(s-u)} ds \\ &\geqslant \log \min \left\{ e^{-(q+\varepsilon)(s-u)}, T_\varepsilon^{(u)} + u \leqslant s \leqslant T_\varepsilon^{(u)} + u + 1 \right\} \\ &\geqslant \log e^{-(q+\varepsilon)(T_\varepsilon^{(u)} + 1)} = -(q + \varepsilon)(T_\varepsilon^{(u)} + 1). \end{aligned}$$

Hence,

$$\hat{\nu}_B - \delta - u \leqslant T_\varepsilon^{(u)} + 1 + \frac{\log B + (q + \varepsilon)(T_\varepsilon^{(u)} + 1)}{q - \varepsilon} = \frac{\log B + 2q(T_\varepsilon^{(u)} + 1)}{q - \varepsilon}, \tag{7.42}$$

so that for an arbitrary $\varepsilon > 0$

$$\mathbf{E}_u(\hat{\nu}_B - u)^+ \leqslant \frac{\log B}{q - \varepsilon} + \frac{2q(1 + \mathbf{E}_u T_\varepsilon^{(u)})}{q - \varepsilon} + \delta = \frac{\log B}{q - \varepsilon} + O(1).$$

Therefore,

$$\mathbf{E}_u(\hat{\nu}_B - u)^+ \leqslant \frac{\log B}{q} + O(1) \quad \text{as } B \to \infty.$$

Combined with (7.41), it yields (5.2). Asymptotics (5.3) is obtained similarly, by taking $\mathbf{P}^\pi$-expectations in (7.42). Then, (5.4) follows similarly to Theorem 5.

□

ACKNOWLEDGEMENTS

This research was supported in part by the U.S. Office of Naval Research Grants N00014-03-1-0027 and N00014-03-1-0823 at the University of Southern California. We are grateful to reviewers whose comments led to an improved style and smoother exposition of results.

REFERENCES

Bansal, R.K. and Papantoni-Kazakos, P. (1986). An Algorithm for Detecting a Change in a Stochastic Process, *IEEE Transactions on Information Theory* 32: 227-235.

Baron, M. (2001). Bayes Stopping Rules in a Change-Point Model with a Random Hazard Rate, *Sequential Analysis* 20: 147-163.

Baron, M. (2002). Bayes and Asymptotically Pointwise Optimal Stopping Rules for the Detection of Influenza Epidemics, *Case Studies of Bayesian Statistics*, vol. 6, C. Gatsonis, R. E. Kass, A. Carriquiry, A. Gelman, D. Higdon, D. K. Pauler and I. Verdinelli, eds., pp. 153-163, New York: Springer-Verlag.

Basseville, M. and Nikiforov, I. V. (1993). *Detection of Abrupt Changes: Theory and Application*, Englewood Cliffs, N.J.: PTR Prentice-Hall, Inc.

Baum, L. and Katz, M. (1965). Convergence Rates in the Law of Large Numbers, *Transactions of American Mathematical Society* 120: 108-123.

Beibel, M. (1996). A Note on Ritov's Bayes Approach to the Minimax Property of the Cusum Procedure, *Ann. Statist.* 24: 1804–1812.

Beibel, M. (2000). Sequential Detection of Signals with Known Shape and Unknown Magnitude, *Statistica Sinica* 10: 715-729.

Bickel, P. J. and Yahav, J. A. (1967). Asymptotically Pointwise Optimal Procedures in Sequential Analysis, *Proc. 5th Berkeley Symp. Math. Statist. Prob.*, vol. I: Statistics, pp. 401-413, Berkeley, CA: University of California Press.

Blažek, R., Kim, H., Rozovskii, B., and Tartakovsky, A. (2001). A Novel Approach to Detection of "Denial-of-Service" Attacks via Adaptive Sequential and Batch-Sequential Change-Point Detection Methods, *Proc. IEEE Systems, Man, and Cybernetics Information Assurance Workshop*, West Point, NY.

Borovkov, A. (1998). Asymptotically Optimal Solutions in the Change-Point Problem, *Theory of Probability and Applications* 43(4): 539-561.

Davis, R.A., Huang, D., and Yao, Y.-C. (1995). Testing for a Change in the Parameter Values and Order of an Autoregressive Model, *Annals of Statistics* 23: 282-304.

Dragalin, V.P., Tartakovsky, A.G., and Veeravalli, V.V. (1999). Multihypothesis Sequential Probability Ratio Tests, Part I: Asymptotic Optimality, *IEEE Transactions on Information Theory* 45: 2448-2461.

Fuh, C. (2003). SPRT and CUSUM in Hidden Markov Models, *Annals of Statistics* 31: 942-977.

Fuh, C. (2004). Asymptotic Operating Characteristics of an Optimal Change Point Detection in Hidden Markov Models, *Annals of Statistics* 32: 2305-2339.

Ghosh, M., Mukhopadhyay, N., and Sen, P.K. (1997). *Sequential Estimation*, New York: Wiley.

Hsu, P.L. and Robbins, H. (1947). Complete Convergence and the Law of Large Numbers, *Proceedings of National Academy of Science U.S.A.* 33: 25-31.

Kent, S. (2000). On the Trial of Intrusions into Information Systems, *IEEE Spectrum* 37(12): 52-56.

Lai, T. (1976). On r-quick Convergence and a Conjecture of Strassen, *Annals of Probability* 4: 612-627.

Lai, T. (1981). Asymptotic Optimality of Invariant Sequential Probability Ratio Tests, *Annals of Statistics* 9: 318-333.

Lai, T. (1998). Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems, *IEEE Transactions on Information Theory* 44: 2917-2929.

Liptser, R. and Shiryaev, A. (1977). *Statistics of Random Processes*, vol. 1, New York: Springer-Verlag.

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897-1908.

MacNeill, I. B. and Mao, Y. (1993). Change-Point Analysis for Mortality and Morbidity Rate, *Journal of Applied Statistical Science* 1: 359-377.

Moustakides, G. V. (2004). Optimality of the Cusum Procedure in Continuous Time, *Annals of Statistics* 32: 302-315.

Moustakides, G. V. (1986). Optimal Stopping Times for Detecting a Change in Distribution, *Annals of Statistics* 14: 1379-1388.

Page, E. S. (1954). Continuous Inspection Schemes, *Biomterika* 41: 100-115.

Peskir, G. and Shiryaev, A. N. (2002). Solving the Poisson Disorder Problem, *Advances in Finance and Stochastics*, pp. 295-312, Berlin: Springer.

Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206-227.

Pollak, M. (1987). Average Run Lengths of an Optimal Method of Detecting a Change in Distribution, *Annals of Statistics* 15: 749-779.

Pollak, M. and Siegmund, D. (1985). A Diffusion Process and its Applications to Detecting a Change in the Drift of Brownian Motion, *Biometrika* 72: 267-280.

Ritov, Y. (1990). Decision Theoretic Optimality of the Cusum Procedure, *Ann. Statist.* 18: 1466-1469.

Shiryaev, A. N. (1961). The Detection of Spontaneous Effects, *Soviet Mathematical Doklady* 2: 740-743.

Shiryaev, A. N. (1963). On Optimum Methods in Quickest Detection Problems, *Theory Probability and Applications* 8: 22-46.

Shiryaev, A. N. (1978). *Optimal Stopping Rules*, New York: Springer-Verlag.

Shiryaev, A. N. (1996). Minimax Optimality of the Method of Cumulative Sum (Cusum) in the Case of Continuous Time, *Russian Mathematical Surveys* 51(4): 750-751.

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*, New York: Springer-Verlag.

Srivastava, M. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryaev-Roberts Procedures for Detecting a Shift in the Mean, *Annals of Statistics* 21: 645-670.

Strassen, V. (1967). Almost Sure Behavior of Sums of Independent Random Variables and Martingales, *Proceedings of Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, pp. 315-343, Berkeley, CA: University of California Press.

Tartakovsky, A. G. (1991). *Sequential Methods in the Theory of Information Systems (in Russian)*, Moscow: Radio i Svyaz'.

Tartakovsky, A. G. (1994). Asymptotically Minimax Multialternative Sequential Rule for Disorder Detection, *Statistics and Control of Random Processes: Proceedings of the Steklov Institute of Mathematics*, vol. 202, pp. 229-236, Providence, RI: American Mathematical Society.

Tartakovsky, A. G. (1995). Asymptotic Properties of Cusum and Shiryaev's Procedures for Detecting a Change in a Nonhomogeneous Gaussian Process, *Mathematical Methods of Statistics* 4: 389-404.

Tartakovsky, A. G. (1998a). Asymptotic Optimality of Certain Multihypothesis Sequential Tests: Non-i.i.d. Case, *Statistical Inference for Stochastic Processes* 1(3): 265-295.

Tartakovsky, A. G. (1998b). Asymptotically Optimal Sequential Tests for Nonhomogeneous Processes, *Sequential Analysis* 17(1): 33-62.

Tartakovsky, A. G. (1998c). *Extended Asymptotic Optimality of Certain Change-Point Detection Procedures*, Preprint, Center for Applied Mathematical Sciences, University of Southern California.

Tartakovsky, A. G. and Blažek, R. (2000). Effective Adaptive Spatial-Temporal Technique for Clutter Rejection in IRST, *SPIE Proceedings: Signal and Data Processing of Small Targets*, vol. 4048, pp. 85-95.

Tartakovsky, A. G., Rozovskii, B.L., Blažek, R., and Kim, H. (2006). Detection of Intrusions in Information Systems by Sequential Change-Point Methods, *Statistical Methodology* (to appear).

Tartakovsky, A. G. and Veeravalli, V. (2002). An Efficient Sequential Procedure for Detecting Changes in Multichannel and Distributed Systems, *Proceeding of the 5th International Conference on Information Fusion*, vol. 1, pp. 41-48, Annapolis, MD: International Society of Information Fusion.

Tartakovsky, A. G. and Veeravalli, V. (2004). Change-Point Detection in Multichannel and Distributed Systems with Applications, *Applications of Sequential Methodologies*, N. Mukhopadhyay, S. Datta and S. Chattopadhyay, eds., pp. 339-370, New York. Marcel Dekker, Inc.

Tartakovsky, A. G. and Veeravalli, V. (2005). General Asymptotic Bayesian Theory of Quickest Change Detection, *Theory Probab. Appl.* 49(4): 458-497.

Wang, H., Zhang, D., and Shin, K.G. (2002). Detecting SYN Flooding Attacks, *Proceedings of INFOCOM 2002, 21st Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 3, pp. 1530-1539, New York: IEEE.

Willsky, A. (1976). A Survey of Design Methods for Failure Detection in Dynamical Systems, *Automatica* 12: 601-611.

Yakir, B. (1994). Optimal Detection of a Change in Distribution when the Observations Form a Markov Chain with a Finite State Space, *Change-Point Problems*, E. Carlstein, H. Muller, and D. Siegmund, eds., pp. 346-358, Hayward, CA: Institute of Mathematical Statistics.

Yakir, B. (1997). A Note on Optimal Detection of a Change in Distribution, *Annals of Statistics* 25: 2117-2126.