SEQUENTIAL ANALYSIS
2024, VOL. 43, NO. 3, 359–378
https://doi.org/10.1080/07474946.2024.2373118

Taylor & Francis
Taylor & Francis Group

Bayesian quickest change detection for unnormalized and score-based models

Taposh Banerjee[a] and Vahid Tarokh[b]

[a]Department of Industrial Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania, USA;
[b]Department of Electrical and Computer Engineering, Duke University, Durham, North Carolina, USA

ABSTRACT

Score-based algorithms are proposed for the quickest detection of changes in unnormalized statistical models. These are models where the densities are known within a normalizing constant. These algorithms can also be applied to score-based models where the score, i.e., the gradient of log density, is known to the decision maker. Bayesian performance analysis is provided for these algorithms and compared with their classical counterparts. It is shown that strong performance guarantees can be provided for these score-based algorithms where the Kullback-Leibeler divergence between pre- and post-change densities is replaced by their Fisher divergence.

ARTICLE HISTORY
Received 31 October 2023
Revised 23 February 2024
Accepted 12 June 2024

KEYWORDS
Fisher divergence; Hyvärinen score; nonlinear renewal theory; score-matching; unnormalized models

1. INTRODUCTION

In the literature on quickest change detection (Tartakovsky 2019; Tartakovsky, Nikiforov, and Basseville 2014; Basseville and Nikiforov 1993; Poor and Hadjiliadis 2009; Veeravalli and Banerjee 2014), algorithms are developed to detect an abrupt change in a sequence of random variables. The strongest results are available under the assumption that the distributions of the observations are known before and after the change point. When the distributions are unknown, the algorithms developed are one of four types: (1) generalized likelihood ratio tests, (2) mixture-based tests, (3) robust tests, and (4) nonparametric tests (see Section 2 below for details).

In many modern machine-learning applications, the data are high-dimensional and the distribution of the observations can only be learned in an unnormalized form (Hyvärinen 2005). Due to the high dimension of the data, the normalizing constant cannot be evaluated using numerical integration. Using the modern techniques of score-matching, it is now also possible to learn the score of the density (gradient of log density) from data using deep neural networks (Song and Ermon 2019; Song et al. 2021; Vincent 2011; Hyvärinen 2005). The classical algorithms from the quickest change detection literature cannot be applied to these modern models. In our recent work (Wu et al. 2023b), we have proposed a score-based cumulative sum algorithm for quickest change detection in these models and also provided its performance analysis. In the score-based method, we replace the log score (which is sensitive to normalizing constants) with the Hyvärinen score (which is invariant to

CONTACT Taposh Banerjee taposh.banerjee@pitt.edu Department for Taposh Banerjee is Industrial Engineering, University of Pittsburgh, Pittsburgh, PA, USA.

normalizing constants) (Hyvärinen 2005). We have then shown that this Hyvärinen score-based test can be designed similarly to the classical cumulative sum algorithm (Lorden 1971; Lai 1998). We review our work in Wu et al. (2023b) in Section 3.

In this article, we develop a Bayesian theory for quickest change detection in unnormalized and score-based statistical models. Specifically, we use the Hyvärinen score to propose score-based variants of classical Shiryaev and Shiryaev-Roberts algorithms. We then provide analysis for the average detection delay and the probability of a false alarm for these algorithms. Our analysis reveals that these score-based algorithms can be designed similarly to their classical counterparts. Also, the Kullback-Leibler divergence term appearing in the delay analysis of classical algorithms is replaced by the Fisher divergence between the pre- and post-change distributions.

The article is organized as follows. In Section 2, we discuss our motivation and the required mathematical background. In Section 3, we review the score-based cumulative sum algorithm and its analysis (Wu et al. 2023b). In Section 4, we propose the score-based Shiryaev algorithm. In Section 5, we provide the false-alarm analysis of the proposed algorithm, and in Section 6, we provide its delay analysis. In Section 7, we provide an example from the Gaussian family of distributions for which the performance of the score-based algorithm and the classical algorithm coincide. Finally, the Bayesian performance of the score-based Shiryaev-Roberts algorithm and the score-based CUSUM algorithm are respectively provided in Sections 8 and 9.

2. BACKGROUND AND MOTIVATION

In the problem of quickest change detection (QCD), a decision maker observes a sequence of random variables $\{X_n\}$. At the time ν, called the change point, the distribution of the variables changes. In the problems studied in Page (1954), Lorden (1971), Moustakides (1986), and Lai (1998), the variables are independent and identically distributed (i.i.d.) before a time ν with density f_0 and i.i.d. with density f_1 after ν:

$$X_n \sim \begin{cases} f_0, & \forall n < \nu, \\ f_1, & \forall n \geq \nu. \end{cases} \tag{2.1}$$

An optimal algorithm in minimax settings (Pollak 1985; Moustakides 1986; Lorden 1971; Lai 1998) is given by the cumulative sum (CUSUM) algorithm:

$$\tau_c = \min\{n \geq 1 : W_n \geq B\}, \tag{2.2}$$

where the CUSUM statistic W_n is given by

$$W_n = \left(W_{n-1} + \log \frac{f_1(X_n)}{f_0(X_n)} \right)^+, \quad W_0 = 0, \tag{2.3}$$

where $(x)^+ = \max\{x, 0\}$. Specifically, this algorithm is asymptotically optimal for the minimax problem formulation of Pollak (Pollak 1985; Lai 1998) and exactly optimal for the minimax problem formulation of Lorden (Lorden 1971; Moustakides 1986). The CUSUM algorithm consistently detects the change because before the change,

$$\mathsf{E}_\infty \left[\log \frac{f_1(X_n)}{f_0(X_n)} \right] = -D(f_0 \mid\mid f_1) < 0,$$

and after the change,

$$\mathsf{E}_1\left[\log\frac{f_1(X_n)}{f_0(X_n)}\right] = D(f_1 \;||\; f_0) > 0.$$

Here,

$$D(f_1 \;||\; f_0) := \int f_1(x)\log\frac{f_1(x)}{f_0(x)}dx \;\; > \;\; 0$$

is the Kullback-Leibler divergence between f_1 and f_0. Also, we have used the notation that E_ν is the expectation when the change occurs at time ν. Thus, the algorithm works because the *drift* of the random walk $\sum_{k=1}^{n}\log\frac{f_1(X_k)}{f_0(X_x)}$ is negative before the change and positive after the change. In addition, if the threshold is set to $B = \log\gamma$, then it can be shown that (Lorden 1971; Lai 1998)

$$\mathsf{E}_\infty[\tau_c] \geq \gamma.$$

Thus, a universal bound (valid for any pair of densities f_0 and f_1) on the mean time to false alarm $\mathsf{E}_\infty[\tau_c]$ can be obtained for the CUSUM algorithm. Finally, as $\gamma \to \infty$,

$$\mathsf{E}_1[\tau_c] = \frac{\log\gamma}{D(f_1 \;||\; f_0)}(1 + o(1)).$$

Here, $o(1) \to 0$ as $\gamma \to \infty$. Thus, the delay of the CUSUM algorithm inversely depends on $D(f_1 \;||\; f_0)$, the Kullback-Leibler divergence between f_1 and f_0. The larger the divergence, the smaller the average detection delay.

In the Bayesian version of the problem studied in Shiryaev (1963, 2007) and Tartakovsky and Veeravalli (2005), it is assumed that the change point is a random variable. The optimal solution is the Shiryaev test

$$\tau_s = \min\{n \geq 1 : \mathsf{P}(\nu \leq n|X_1, ..., X_n) \geq A\}. \tag{2.4}$$

The asymptotic optimality of this test is established in Tartakovsky and Veeravalli (2005). This test is exactly optimal when the change point is a geometrically distributed random variable: $\nu \sim \text{Geom}(\rho)$. In this case, the Shiryaev statistic has a simple recursion: if $p_n = \mathsf{P}(\nu \leq n|X_1, ..., X_n)$, then $R_n = \frac{p_n}{1-p_n}$ can be written as

$$R_n = \frac{1}{(1-\rho)^n}\sum_{k=1}^{n}(1-\rho)^{k-1}\rho \;\; \prod_{i=k}^{n}\frac{f_1(X_i)}{f_0(X_i)}, \tag{2.5}$$

and the statistic R_n has the simple recursion:

$$R_n = \frac{R_{n-1} + \rho}{1-\rho} \;\; \frac{f_1(X_n)}{f_0(X_n)}, \quad R_0 = 0. \tag{2.6}$$

By setting the threshold $A = 1 - \alpha$, we get

$$\text{PFA}(\tau_s) = \mathsf{P}(\tau_s < \nu) \leq \alpha.$$

Thus, like the CUSUM algorithm, we can provide a universal guarantee on the false alarm rate (valid for any f_0 and f_1) and also for the Shiryaev algorithm. Using the results from nonlinear renewal theory (Woodroofe 1982), it can be further shown that as $\alpha \to 0$,

$$\mathsf{E}_1[\tau_s] = \frac{|\log \alpha|}{D(f_1 \;||\; f_0) + |\log(1-\rho)|}(1 + o(1)).$$

Thus, for both minimax and Bayesian settings, the performance of the optimal algorithm depends inversely on the Kullback-Leibler divergence between f_1 and f_0.

Classical QCD algorithms have also been extended to non-i.i.d. models, including Markov models. Also, if the change is not persistent, a delay penalty may not be appropriate. We refer the readers to Veeravalli and Banerjee (2014); Lai (1998); Tartakovsky and Veeravalli (2005); Tartakovsky (2017); Tartakovsky, Nikiforov, and Basseville (2014); Sarnowski and Szajowski (2011); Xie et al. (2021);and Polunchenko and Tartakovsky (2012) for details.

In many science and engineering applications of QCD, the densities f_0 and f_1 are not known. In the QCD literature, this issue is addressed through four fundamental methods:

1. Generalized likelihood ratio (GLR) tests: In this class of tests, the pre-change density f_0 is generally assumed known and the post-change density f_1 is assumed to belong to a parametric family of densities, i.e., $f_1 = f_\theta,\ \theta \in \Theta \subset \mathbb{R}^d$. The optimal test is then obtained by replacing the unknown parameter θ with its maximum likelihood estimate (Lorden 1971; Lai 1998; Tartakovsky, Nikiforov, and Basseville 2014; Tartakovsky 2019).
2. Mixture-based tests: In this class of tests, it is again assumed that $f_1 = f_\theta,\ \theta \in \Theta \subset \mathbb{R}^d$, with θ having a prior density $\pi(\theta)$. The optimal test is then obtained by integrating the likelihood ratio over the prior density (Lai 1998; Tartakovsky, Nikiforov, and Basseville 2014; Tartakovsky 2019).
3. Robust tests: In this class of tests, it is assumed that the post-change family of distributions has a member that is least favorable in a well-defined sense, and the optimal test designed using this least favorable member is robust optimal over the entire post-change class (Unnikrishnan, Veeravalli, and Meyn 2011; Hou et al. 2023; Oleyaeimotlagh et al. 2023). The paradigm of robust tests has two major benefits. First, it allows the post-change class to be infinite-dimensional. Second, the optimal test is often computationally efficient to implement. The GLR and mixture tests, in general, cannot be implemented using a recursively computable statistic.
4. Nonparametric tests: Often the assumptions mentioned above are not satisfied and we need to resort to nonparametric tests. These tests are based on either signs or ranks or other universal strategies employed in the theory of nonparametric statistics (Gordon and Pollak 1994; Pawlak and Steland 2013; Liang and Veeravalli 2022; Banerjee, Firouzi, and Hero 2018; Lau, Tay, and Veeravalli 2019; Konev and Vorobeychikov 2017; Brodsky and Darkhovsky 1993; Darkhovsky and Piryatinska 2018a, 2018b; Darkhovsky and Piryatinska, 2014a, 2014b, 2015). In this family of tests, the desire for optimality is replaced with the need to obtain performance guarantees on the detection delay and the rate of false alarms.

In modern machine-learning applications, two new classes of models have emerged:

(1) *Unnormalized statistical models*: In these models, we know the distribution within a normalizing constant. Specifically, we have

$$f_0(x) = \frac{\tilde{f}_0(x)}{Z_0}, \quad \text{and} \quad f_1(x) = \frac{\tilde{f}_1(x)}{Z_1}. \tag{2.7}$$

Here, Z_0 and Z_1 are normalizing constants:

$$Z_0 = \int_x \tilde{f}_0(x)dx, \quad \text{and} \quad Z_1 = \int_x \tilde{f}_1(x)dx. \tag{2.8}$$

The variable x is high-dimensional and Z_0 and Z_1 are hard (or even impossible) to calculate by numerical integration. Thus, the normalizing constants Z_0 and Z_1 are assumed to be unknown. The unnormalized models $\tilde{f}_0(x)$ and $\tilde{f}_1(x)$ are known in precise functional forms. Examples include continuous-valued Markov random fields or undirected graphical models, which are used for image modeling. We refer the reader to Hyvärinen (2005) and Wu et al. (2023b) for detailed discussions on unnormalized models.

(2) *Score-based models*: In many modern machine-learning applications, even $\tilde{f}_0(x)$ and $\tilde{f}_1(x)$ are unknown. But we may learn the scores:

$$\nabla_x \log f_0(x), \quad \text{and} \quad \nabla_x \log f_1(x),$$

from data. Here ∇_x is the gradient operator. This is possible using the idea of *score-matching*. Specifically, these scores can be learned using a deep neural network. We refer the readers to Hyvärinen (2005), Song and Ermon (2019), Vincent (2011), and Wu et al. (2023a, 2023b) for details. We note that a score-based model is also unnormalized where the exact form of the unnormalized function is hard to estimate. </NL>

While score-matching is used for generative modeling in machine learning, we show that it can also be used for quickest change detection. Clearly, the classical CUSUM or Shiryaev algorithms cannot be applied to unnormalized and score-based models. In recent work, Wu et al. (2023b), we have developed a score-based CUSUM algorithm to detect changes in unnormalized and score-based models. In this article, we develop a Bayesian theory for the quickest change detection in these models. In this theory, and also in the analysis provided in Wu et al. (2023b), the Kullback-Leibler divergence is replaced by the Fisher divergence (defined precisely below) between the pre- and post-change densities. We note that there exist score-based approaches to hypothesis testing and change detection (e.g., Kang and Song 2017; Song and Kang 2020). However, the definition of the score used in these articles is different from the notion of the score used in our article.

In the rest of the article, we assume that for any two pdfs f and g appearing in the article, pdfs have full support on the Euclidean space, the pdf f is differentiable, the model score function $\nabla_x \log f(x)$ is differentiable, the expectations $\mathbb{E}_{X\sim f}[||\nabla_x \log f(X)||_2^2]$ and $\mathbb{E}_{X\sim f}[||\nabla_x \log g(X)||_2^2]$ are finite, and $f(x)\nabla_x \log g(x) \to 0$ when $||x|| \to \infty$. We refer to Hyvärinen (2005) for details.

3. SCORE-BASED CUSUM (SCUSUM) ALGORITHM

To address the limitations of the classical CUSUM algorithm for unnormalized or score-based models, we have developed a score-based CUSUM (SCUSUM) algorithm in Wu et al. (2023b). In this subsection, we review this algorithm and its performance analysis.

In the performance of the SCUSUM algorithm (to be provided below), the Kullback-Leibler divergence term is replaced by the Fisher divergence between the densities. We define the Fisher divergence between two densities f and g as

$$\mathbb{D}_{\mathrm{F}}(f \;||\; g) = \mathbb{E}_{X\sim f}\left[\frac{1}{2}||\nabla_x \log f(X) - \nabla_x \log g(X)||_2^2\right], \tag{3.1}$$

where $||\cdot||_2$ denotes the Euclidean norm. It is not a metric since it is not symmetric, but is zero if and only if the densities are identical. It also does not depend on the normalizing constants. The SCUSUM algorithm is based on the score introduced by Hyvärinen in Hyvärinen (2005). The Hyvärinen score for a density f is defined as

$$\mathcal{S}_{\mathrm{H}}(X,f) = \frac{1}{2}||\nabla_x \log f(X)||_2^2 + \Delta_x \log f(X), \tag{3.2}$$

whenever it can be well defined. Here, ∇_x and $\Delta_x = \sum_{i=1}^{d} \frac{\partial^2}{\partial x_i^2}$ are the gradient and the Laplacian operators acting on $X = (x_1, ..., x_d)^\top$. Since the score is a function of the gradient of log density, it is not a function of any normalizing constants. Under some mild regularity conditions on f and g, it can be shown that

$$\mathbb{D}_{\mathrm{F}}(f \;||\; g) = \mathbb{E}_{X\sim f}\left[\frac{1}{2}||\nabla_{\mathbf{x}} \log f(X)||_2^2 + \mathcal{S}_{\mathrm{H}}(X,g)\right].$$

Using the log score notation $\mathcal{S}_{\mathrm{L}}(X,g) = -\log g(X)$, the CUSUM algorithm can be expressed as

$$\begin{aligned} W_n &= \left(W_{n-1} + \mathcal{S}_{\mathrm{L}}(X_n,f_0) - \mathcal{S}_{\mathrm{L}}(X_n,f_1)\right)^+, \quad W_0 = 0 \\ \tau_c &= \min\{n \geq 1 : W_n > A\}. \end{aligned}$$

Motivated by this, in Wu et al. (2023b), we used the Hyvärinen score difference

$$z_\lambda(X) = \lambda(\mathcal{S}_{\mathrm{H}}(X,f_0) - \mathcal{S}_{\mathrm{H}}(X,f_1)) \tag{3.3}$$

and obtained a score-based CUSUM algorithm:

$$\begin{aligned} Y_n &= \left(Y_{n-1} + \lambda(\mathcal{S}_{\mathrm{H}}(X,f_0) - \mathcal{S}_{\mathrm{H}}(X,f_1))\right)^+, \quad Y_0 = 0 \\ \tau_{sc} &= \min\{n \geq 1 : Y_n > A\}. \end{aligned} \tag{3.4}$$

The parameter λ plays an important role in the analysis of the algorithm. We have proved the following main results in Wu et al. (2023b) regarding the score-based CUSUM algorithm:

(1) Consider the instantaneous SCUSUM score function $X \mapsto z_\lambda(X)$ as defined in Equation (3.3). Then,

$$\begin{aligned} \mathbb{E}_\infty[z_\lambda(X)] &= -\;\lambda\; \mathbb{D}_{\mathrm{F}}(f_0 \;||\; f_1) < 0, \\ \mathbb{E}_1[z_\lambda(X)] &= \lambda\; \mathbb{D}_{\mathrm{F}}(f_1 \;||\; f_0) > 0. \end{aligned} \tag{3.5}$$

Thus, the SCUSUM algorithm can consistently detect the change for any choice of $\lambda > 0$.

(2) In the analysis of the CUSUM algorithm, a fundamental role is played by the fact that the process

$$\exp\left(\sum_{i=1}^{n}\left(\mathcal{S}_L(X_i,f_0)-\mathcal{S}_L(X_i,f_1)\right)\right)=\prod_{i=1}^{n}\frac{f_1(X_i)}{f_0(X_i)},\quad \forall n\geq 1,$$

is a $P_\infty-$ martingale ($\mathcal{S}_L$ is the log score $\mathcal{S}_{\mathrm{L}}(X,g)=-\log g(X)$). This means we can use martingale theory to design the CUSUM test (Lai 1998). This argument cannot be utilized for the SCUSUM algorithm because

$$\exp\left(\sum_{i=1}^{n}(\mathcal{S}_H(X_i,f_0)-\mathcal{S}_H(X_i,f_1))\right),\ \forall n\geq 1,$$

is not a $P_\infty-$ martingale, in general. However, we have shown in Wu et al. (2023b) that there always exists a $\lambda > 0$ such that

$$\mathbb{E}_\infty[\exp(z_\lambda(X))]\leq 1, \tag{3.6}$$

and this implies that

$$\exp\left(n\delta+\lambda\sum_{i=1}^{n}\left(\mathcal{S}_H(X_i,f_0)-\mathcal{S}_H(X_i,f_1)\right)\right),\ \forall n\geq 1,$$

is a$P_\infty-$martingale where $\delta=-\log\left(\mathbb{E}_\infty[\exp(z_\lambda(X))]\right)$. This novel martingale characterization allowed us to prove the following statement: Consider the stopping rule τsc defined in Equation (3.4) with λ satisfying (3.6). Then, for any $A > 0$,

$$\mathbb{E}_\infty[\tau_{sc}]\geq e^A. \tag{3.7}$$

Thus, setting $A = \log\gamma$ implies

$$\mathbb{E}_\infty[\tau_{sc}]\geq \gamma.$$

Thus, similar to the CUSUM algorithm, there exists a universal bound (valid for every f_0 and f_1) for the mean time to false alarm for the SCUSUM algorithm.

(3) In addition to the guarantee on the false alarm rate, we have established the following asymptotic delay guarantee for the algorithm. Consider the stopping rule τ_{sc} defined in (3.4) with $A = \log\gamma$. Then

$$\mathbb{E}_1[\tau_{sc}]\sim\frac{\log\gamma}{\lambda\mathbb{D}_{\mathrm{F}}(f_1\ ||\ f_0)}, \tag{3.8}$$

as $\gamma\to\infty$. Thus, the expected detection delay depends inversely on the Fisher divergence between f_1 and f_0. Thus, the role of KL-divergence in classical quickest change detection is replaced by the Fisher divergence in the score-based CUSUM algorithm.

We note that, as discussed in Remark 2 in Wu et al. (2023b), except in some pathological cases, we can find a λ^* that satisfies (3.6) with equality. In this case, we may select λ (using empirical methods such as Langevin algorithm (Andrieu and Thoms 2008) or Stein variational gradient descent (Liu and Wang 2016) close or equal to λ^* for optimal performance.

4. HYVÄRINEN SCORE-BASED SHIRYAEV ALGORITHM

In this section, we use the Hyvärinen score to define a score-based version of the classical Shiryaev algorithm. Let ν be the random variable for the change point with the prior

$$\pi_n = \mathsf{P}(\nu = n).$$

Also, let $\Pi_n = \mathsf{P}(\nu > n)$. The score-based Shiryaev statistic is defined as

$$S_n = \frac{1}{\Pi_n} \sum_{k=1}^{n} \pi_k \ e^{\sum_{i=k}^{n} Z_\lambda(X_i)},$$

where

$$Z_\lambda(X_i) = \lambda\big(\mathcal{S}_{\mathrm{H}}(X_i, f_0) - \mathcal{S}_{\mathrm{H}}(X_i, f_1)\big).$$

Then, S_n can be written recursively as:

$$S_n = \frac{\Pi_{n-1}}{\Pi_n}\left(S_{n-1} + \frac{\pi_n}{\Pi_{n-1}}\right) \ e^{Z_\lambda(X_n)}.$$

If ν is a geometrically distributed random variable:

$$\pi_n = (1-\rho)^{n-1}\rho, \quad n \geq 1.$$

Then we get a simpler recursion for S_n:

$$S_n = \frac{1}{1-\rho}\left(S_{n-1} + \rho\right) \ e^{Z_\lambda(X_n)}. \tag{4.1}$$

In this article, we focus on geometrically distributed change point random variables.

We first show that, for a carefully selected λ, the statistic process $\{S_n\}$ is a nonnegative submartingale.

Lemma 4.1. *Let ν be a geometrically distributed random variable with parameter ρ sufficiently close to zero. Let λ be such that*

$$\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right] = (1-\rho). \tag{4.2}$$

Then, for this choice of λ, the process $\{S_n\}$ is a non-negative submartingale with respect to its natural filtration.

Proof. We note that the moment generating function $\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right]$ is a convex function of λ. Also,

$$\frac{d}{d\lambda}\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right]\Big|_{\lambda=0} = -\mathbb{D}_F(f_0 \ || \ f_1) < 0.$$

So, finding a λ that satisfies (4.2) is possible for all ρ sufficiently close to zero. The fact that the statistic is non-negative follows from its definition. Now,

$$\mathsf{E}_\infty[S_n|\mathcal{F}_{n-1}] = \left(\frac{1}{1-\rho}(S_{n-1}+\rho)\right)\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right] = S_{n-1}+\rho \geq S_{n-1}.$$

Here the terms $\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right]$ and $1-\rho$ cancel each other because of the assumption. Thus, S_n is a nonnegative submartingale. □

5. FALSE ALARM ANALYSIS OF SCORE-BASED SHIRYAEV ALGORITHM

In this section, we provide a false-alarm analysis of the score-based Shiryaev algorithm. We note for this and the subsequent analysis that, since the score-based methods are based on scores and not likelihoods, the standard proofs from the literature are not directly usable, and subtle changes and assumptions are needed to make the classical proofs work (see Tartakovsky, Nikiforov, and Basseville 2014).

Now, let

$$\tau_{ss} = \min\{n \geq 1 : S_n \geq A\}, \tag{5.1}$$

where S_n is the score-based Shiryaev algorithm defined in (4.1). We refer to the stopping rule τ_{ss} as the score-based Shiryaev stopping rule. We use the notation

$$\mathsf{P}^\pi = \sum_{n=1}^{\infty} \pi_n \ \mathsf{P}_n,$$

where P_n is the law under which the change occurs at time n. The probability of a false alarm for a stopping rule or time τ is defined as

$$\mathrm{PFA}(\tau) = \mathsf{P}^\pi(\tau < \nu). \tag{5.2}$$

Note that

$$\mathsf{P}_n(\tau < n) = \mathsf{P}_\infty(\tau < n)$$

because the event $\{\tau < n\}$ belongs to the sigma algebra generated by $X_1, \ldots, X_{n-1}$. As a result, the probability of this event under P_n and P_∞ is the same.

The next theorem provides a universal guarantee for the probability of a false alarm for τ_{ss}.

Theorem 5.1. *Let ν be a geometrically distributed random variable: $\nu \sim \mathrm{Geom}(\rho)$. Then there exists a ρ_0 such that for all $\rho \leq \rho_0$ and for λ selected as in Lemma 4.1, i.e.,*

$$\mathsf{E}_\infty\left[e^{Z_\lambda(X_1)}\right] = (1-\rho).$$

we have that setting $A = \frac{1-\rho}{\alpha}$ gives us

$$\mathsf{P}^\pi(\tau_s < \nu) \leq \alpha.$$

Proof. From Lemma 4.1, we know that a λ satisfying $\mathsf{E}_\infty\left[e^{Z_\lambda(X_1)}\right] = (1-\rho)$ can always be found for a ρ sufficiently close to zero.

First note that

$$\begin{aligned}\mathrm{PFA}(\tau_s) = \mathsf{P}^{\pi}(\tau_s < \nu) = \sum_{n=1}^{\infty} \pi_n \ \mathsf{P}_n(\tau_s < n) \quad &= \sum_{n=1}^{\infty} \pi_n \ \mathsf{P}_{\infty}(\tau_s < n) \\ &= \sum_{n=1}^{\infty} \pi_n \ \mathsf{P}_{\infty}\Big(\max_{1\le k<n} S_k \ge A\Big).\end{aligned} \tag{5.3}$$

Next, by Lemma 4.1, the statistic $\{S_n\}$ is a submartingale. Hence by Doob's submartingale inequality,

$$\begin{aligned}\mathrm{PFA}(\tau_s) = \mathsf{P}^{\pi}(\tau_s < \nu) \quad &= \sum_{n=1}^{\infty} \pi_n \ \mathsf{P}_{\infty}\Big(\max_{1\le k<n} S_k \ge A\Big) \\ &\le \sum_{n=1}^{\infty} \pi_n \ \frac{1}{A}\mathsf{E}_{\infty}[S_{n-1}] \\ &= \frac{1}{A}\sum_{n=1}^{\infty} \pi_n \ \mathsf{E}_{\infty}[S_{n-1}].\end{aligned} \tag{5.4}$$

Now, the expected value $\mathsf{E}_{\infty}[S_{n-1}]$ satisfies the recursion

$$\mathsf{E}_{\infty}[S_n] \quad = \left(\frac{1}{1-\rho}(\mathsf{E}[S_{n-1}] + \rho)\right)\mathsf{E}_{\infty}\big[e^{Z_{\lambda}(X_n)}\big] = \mathsf{E}[S_{n-1}] + \rho,$$

where the second equality follows by canceling $\mathsf{E}_{\infty}\big[e^{Z_{\lambda}(X_n)}\big]$ with $(1-\rho)$. Using the fact that $S_0 = 0$, we have

$$\mathsf{E}_{\infty}[S_n] = n\rho.$$

Substituting this in the expression for the PFA, we have

$$\mathrm{PFA}(\tau_s) \quad \le \frac{1}{A}\sum_{n=1}^{\infty} \pi_n \ \mathsf{E}_{\infty}[S_{n-1}] = \frac{1}{A}\sum_{n=1}^{\infty} \pi_n \ (n-1)\rho = \frac{\rho}{A}\left(\frac{1}{\rho} - 1\right) = \frac{1-\rho}{A}$$

Thus, setting $A = \frac{1-\rho}{\alpha}$ gives us the desired bound α on the PFA. □

6. DELAY ANALYSIS OF THE SCORE-BASED SHIRYAEV ALGORITHM

In this section, we provide the delay analysis of the proposed score-based Shiryaev algorithm. For the delay analysis, we first express the stopping rule in a form that is amenable to analysis using non-linear renewal theory (Woodroofe 1982). To this end, note that the statistic S_n can be written as

$$\begin{aligned}S_n \quad &= \frac{1}{(1-\rho)^n}\sum_{k=1}^{n}(1-\rho)^{k-1}\rho \ e^{\sum_{i=k}^{n} Z_{\lambda}(X_i)} \\ &= \rho\frac{e^{\sum_{i=1}^{n} Z_{\lambda}(X_i)}}{(1-\rho)^n}\sum_{k=1}^{n}(1-\rho)^{k-1} \ e^{-\sum_{i=1}^{k-1} Z_{\lambda}(X_i)} \\ &= \rho\frac{e^{\sum_{i=1}^{n} Z_{\lambda}(X_i)}}{(1-\rho)^n}\left(1 + \sum_{k=1}^{n-1}(1-\rho)^{k} \ e^{-\sum_{i=1}^{k} Z_{\lambda}(X_i)}\right).\end{aligned} \tag{6.1}$$

Here we assume that $\sum_{k=1}^{0} Y_k = 0$ for any sequence $\{Y_k\}$. Thus,

$$\begin{aligned}\log\left(\frac{S_n}{\rho}\right) &= \sum_{i=1}^{n}(Z_\lambda(X_i) + |\log(1-\rho)|) + \log\left(1 + \sum_{k=1}^{n-1}(1-\rho)^k \; e^{-\sum_{i=1}^{k} Z_\lambda(X_i)}\right) \\ &:= Z_n + \ell_n.\end{aligned} \tag{6.2}$$

Here, Z_n is the random walk $\sum_{i=1}^{n}(Z_\lambda(X_i) + |\log(1-\rho)|)$ and ℓ_n is the disturbance term

$$\ell_n = \log\left(1 + \sum_{k=1}^{n-1}(1-\rho)^k \; e^{-\sum_{i=1}^{k} Z_\lambda(X_i)}\right).$$

Thus,

$$\begin{aligned}\tau_{ss} = \min\{n \geq 1 : S_n \geq A\} &= \min\left\{n \geq 1 : \log\left(\frac{S_n}{\rho}\right) \geq \log\left(\frac{A}{\rho}\right)\right\} \\ &= \min\left\{n \geq 1 : Z_n + \ell_n \geq \log\left(\frac{A}{\rho}\right)\right\}.\end{aligned} \tag{6.3}$$

Define

$$b = \log\left(\frac{A}{\rho}\right),$$

then the score-based Shiryaev stopping rule is given by

$$\tau_{ss} = \min\{n \geq 1 : Z_n + \ell_n \geq b\}. \tag{6.4}$$

Thus, the stopping time τ_{ss} can be written as the hitting time for a random walk and a 'slowly changing' term. This brings us to the domain of nonlinear renewal theory (Woodroofe 1982) and allows us to prove the following theorem. Let

$$\mu = \lambda \; \mathbb{D}_{\mathrm{F}}(f_1 \;||\; f_0) + |\log(1-\rho)|. \tag{6.5}$$

Theorem 6.1. *Let ν be a geometrically distributed random variable: $\nu \sim \mathrm{Geom}(\rho)$. Then there exists a ρ_0 such that for all $\rho \leq \rho_0$ and for λ selected such that*

$$\mathsf{E}_\infty\left[e^{Z_\lambda(X_1)}\right] = (1-\rho),$$

we have the following results:

(1) *The stopping rule stops almost surely:*

$$\tau_{ss} < \infty, \;\; \text{almost surely}, \quad \forall b \geq 0.$$

and

$$\frac{\tau_{ss}}{\lfloor b/\mu \rfloor} \to 1, \;\; \text{almost surely}, \qquad \text{as } b \to \infty.$$

(2) *The expected value of the stopping rule is also finite and has the following asymptotics: when* $\sigma^2 = \mathrm{Var}(Z_\lambda(X_1)) < \infty$, *then*

$$\mathsf{E}_1[\tau_{ss}] = \frac{b}{\mu}(1 + o(1)) = \frac{b}{\lambda\ \mathbb{D}_{\mathrm{F}}(f_1||f_0) + |\log(1-\rho)|}(1 + o(1)), \quad \text{as} \quad b \to \infty.$$

Thus, by setting $A = \frac{1-\rho}{\alpha}$, *we get* $\mathrm{PFA}(\tau_{ss}) \leq \alpha$ *and the delay then becomes*

$$\mathsf{E}_1[\tau_{ss}] = \frac{|\log\alpha|}{\lambda\ \mathbb{D}_{\mathrm{F}}(f_1||f_0) + |\log(1-\rho)|}(1 + o(1)), \quad \text{as} \quad \alpha \to 0.$$

(3) *The stopping rule is asymptotically normal: when* $\sigma^2 = \mathrm{Var}(Z_\lambda(X_1)) < \infty$, *and let* $N_b = \lfloor b/\mu \rfloor$, *then*

$$\frac{\tau_s - N_b}{\sqrt{N_b}}$$

is asymptotically normal with mean zero and variance $\frac{\sigma^2}{\mu^2}$, *as* $b \to \infty$.

Proof. First, note that

$$\ell_n \ \uparrow\ \ell := \log\left(1 + \sum_{k=1}^{\infty}(1-\rho)^k\ e^{-\sum_{i=1}^{k} Z_\lambda(X_i)}\right), \quad \text{as} \quad n \to \infty.$$

Here we used $\uparrow$ to denote a monotonic limit, i.e., ℓ_n monotonically increases to ℓ, as $n \to \infty$. The expected value of ℓ is given by (using Jensen's inequality)

$$\begin{aligned}
\mathsf{E}_1[\ell] &= \mathsf{E}_1 \log\left(1 + \sum_{k=1}^{\infty}(1-\rho)^k\ e^{-\sum_{i=1}^{k} Z_\lambda(X_i)}\right) \\
&\leq \log\left(1 + \sum_{k=1}^{\infty}(1-\rho)^k\ \mathsf{E}_1\left[e^{-\sum_{i=1}^{k} Z_\lambda(X_i)}\right]\right) \\
&= \log\left(1 + \sum_{k=1}^{\infty}(1-\rho)^k\ \mathsf{E}_1\left[e^{-Z_\lambda(X_1)}\right]^k\right).
\end{aligned} \tag{6.6}$$

For all ρ small enough, the corresponding λ will be close to zero. This would ensure that

$$\mathsf{E}_1\left[e^{-Z_\lambda(X_1)}\right] \leq 1$$

because

$$\frac{d}{d\lambda}\mathsf{E}_1\left[e^{-Z_\lambda(X_1)}\right] = -\mathsf{E}_1\left[Z_1(X_1)e^{-Z_\lambda(X_1)}\right].$$

Thus, at $\lambda = 0$, the slope of the moment generating function is

$$-\mathsf{E}_1[Z_1(X_1)] = -\mathbb{D}_F(f_1\ ||\ f_0) < 0.$$

Thus,

$$\begin{aligned} \mathsf{E}_1[\ell] &\leq \log\left(1+\sum_{k=1}^{\infty}(1-\rho)^k\ \mathsf{E}_1\left[e^{-Z_\lambda(X_1)}\right]^k\right) \\ &\leq \log\left(1+\sum_{k=1}^{\infty}(1-\rho)^k\right) = \log\left(1+\frac{1-\rho}{\rho}\right) = \log\left(\frac{1}{\rho}\right). \end{aligned} \tag{6.7}$$

The first part of the theorem now follows from Lemma 4.1 in Woodroofe (1982) because

$$\frac{1}{n}\max\{|\ell_1|,|\ell_2|,\ldots,|\ell_n|\} \quad \to \quad 0, \qquad \text{as} \qquad n\to\infty. \tag{6.8}$$

The previous assertion is true simply because $\ell_n \to \ell < \infty$.

The second part of the theorem follows from Theorem 4.4 in Woodroofe (1982) because the process $\{\ell_n\}$ satisfies (6.8) and

$$\sum_{n=1}^{\infty}\mathsf{P}_1(\ell_n \leq -n\epsilon) < \infty, \qquad \text{for some } \epsilon, \quad 0<\epsilon<\mu.$$

The last condition is satisfied because $\ell_n \geq 0$, for all n.

The result on asymptotic normality follows from Lemma 4.2 in Woodroofe (1982) because $\{\ell_n\}$ satisfies (6.8),

$$\frac{\ell_n}{\sqrt{n}} \to 0, \qquad \text{almost surely, as } n\to\infty,$$

and the sequence $\{\ell_n\}$ is uniformly continuous in probability: for every $\epsilon>0$, there is $\delta>0$ for which

$$\mathsf{P}_1\left(\max_{0\leq k\leq \delta n}|\ell_{n+k}-\ell_n|>\epsilon\right) < \epsilon, \quad \forall\ n\geq 1. \tag{6.9}$$

The last two assertions are true again because $\ell_n \to \ell$, a finite limit. □

7. ASYMPTOTIC PERFORMANCE FOR GAUSSIAN RANDOM VARIABLES

We now give an example in which the performance of the Shiryaev algorithm and the score-based Shiryaev algorithm are asymptotically identical. Let

$$f_0 = \mathcal{N}(0,1), \quad f_1 = \mathcal{N}(\mu,1), \quad \mu\neq 0.$$

Then

$$\begin{aligned} \nabla_x \log f_0(x) &= \nabla_x \log \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} = -x \\ \nabla_x \log f_1(x) &= \nabla_x \log \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2}} = -x+\mu. \end{aligned} \tag{7.1}$$

The Fisher divergence is given by

$$\begin{aligned}\mathbb{D}_{\mathrm{F}}(f_1 \;||\; f_0) &= \mathsf{E}_{X\sim f_1}\left[\frac{1}{2}||\nabla_x \log f_1(X) - \nabla_x \log f_0(X)||_2^2\right] \\ &= \mathsf{E}_{X\sim f_1}\left[\frac{1}{2}||-(X-\mu)+X||_2^2\right] = \frac{\mu^2}{2}.\end{aligned} \tag{7.2}$$

The Kullback-Leibler divergence is given by

$$\begin{aligned}\mathbb{D}(f_1 \;||\; f_0) &= \mathsf{E}_{X\sim f_1}\left[\log\frac{f_1(X)}{f_0(X)}\right] \\ &= \mathsf{E}_{X\sim f_1}\left[X\mu - \frac{\mu^2}{2}\right] = \frac{\mu^2}{2}.\end{aligned} \tag{7.3}$$

Thus, the two divergences coincide. But note that the performance of the score-based Shiryaev algorithm is governed by the product $\lambda\mathbb{D}_{\mathrm{F}}(f_1 \;||\; f_0)$, where $\lambda > 0$ is such that

$$\mathsf{E}_{\infty}\left[e^{\lambda(\mathcal{S}_{\mathrm{H}}(X,f_0)-\mathcal{S}_{\mathrm{H}}(X,f_1))}\right] = 1. \tag{7.4}$$

We now show that we can choose $\lambda = 1$ in the above equation. To see this, note that

$$\begin{aligned}\mathcal{S}_{\mathrm{H}}(X,f_0) &= \frac{1}{2}||\nabla_x \log f_0(X)||_2^2 + \Delta_x \log f_0(X) = \frac{X^2}{2} - 1 \\ \mathcal{S}_{\mathrm{H}}(X,f_1) &= \frac{1}{2}||\nabla_x \log f_1(X)||_2^2 + \Delta_x \log f_1(X) = \frac{(X-\mu)^2}{2} - 1 \\ \mathcal{S}_{\mathrm{H}}(X,f_0) - \mathcal{S}_{\mathrm{H}}(X,f_1) &= \frac{X^2}{2} - \frac{(X-\mu)^2}{2} = X\mu - \frac{\mu^2}{2}.\end{aligned} \tag{7.5}$$

Thus,

$$\mathsf{E}_{\infty}\left[e^{\lambda(\mathcal{S}_{\mathrm{H}}(X,f_0)-\mathcal{S}_{\mathrm{H}}(X,f_1))}\right] = \mathsf{E}_{\infty}\left[e^{\lambda\left(X\mu-\frac{\mu^2}{2}\right)}\right] = e^{-\frac{\lambda\mu^2}{2}}\mathsf{E}_{\infty}\left[e^{\lambda\mu X}\right] = e^{-\frac{\lambda\mu^2}{2}}e^{\frac{\lambda^2\mu^2}{2}}. \tag{7.6}$$

For $\lambda > 0$ to satisfy (7.4), we must have

$$-\frac{\lambda\mu^2}{2} + \frac{\lambda^2\mu^2}{2} = 0.$$

This implies that $\lambda = 1$. These calculations show that for the Shiryaev stopping rule τ_s and the score-based Shiryaev stopping rule τ_{ss},

$$\mathsf{E}_1[\tau_s] \sim \mathsf{E}_1[\tau_{ss}] \sim \frac{|\log\alpha|}{\lambda\;\mathbb{D}_{\mathrm{F}}(f_1||f_0) + |\log(1-\rho)|} = \frac{|\log\alpha|}{\frac{\mu^2}{2} + |\log(1-\rho)|}, \quad \text{as} \quad \alpha \to 0. \tag{7.7}$$

We note that the above arguments can be used to show that the SCUSUM algorithm is asymptotically optimal (Wu et al. 2023b). However, a similar statement for the score-based Shiryaev algorithm does not follow from our analysis because the delay analysis is only provided for $\nu = 1$ and is not averaged over all possible values of the change point. For a more general statement about multivariate Gaussian data, we refer the readers to Wu et al. (2023b).

We make a few additional remarks on the applicability of the score-based algorithms:

1. In general, $\lambda\ \mathbb{D}_{\mathrm{F}}(f_1\ ||\ f_0) < \mathbb{D}(f_1\ ||\ f_0)$, and the score-based methods are suboptimal.
2. For high-dimensional data, we can only provide an analytical comparison between the classical and score-based algorithms as the classical algorithms cannot be implemented in practice (due to the lack of knowledge of the exact likelihood).
3. For comparison of score-based methods with other competing methods for QCD, we refer the readers to Wu et al. (2024).

8. BAYESIAN ANALYSIS OF THE SCORE-BASED SHIRYAEV-ROBERTS ALGORITHM

In this section, we consider the score-based Shiryaev-Roberts algorithm and provide its performance analysis. We define the statistic for this algorithm as

$$R_n = \sum_{k=1}^{n} e^{\sum_{i=k}^{n} Z_\lambda(X_i)},$$

where recall that

$$Z_\lambda(X_i) = \lambda\big(\mathcal{S}_{\mathrm{H}}(X_i, f_0) - \mathcal{S}_{\mathrm{H}}(X_i, f_1)\big).$$

The stopping time for this algorithm is defined as

$$\tau_{ssr} = \min\{n \geq 1 : R_n \geq B\}.$$

Similar to the classical likelihood ratio-based Shiryaev-Roberts statistic (Tartakovsky, Nikiforov, and Basseville 2014; Pollak 1985), this statistic R_n also has a recursion:

$$R_n = (1 + R_{n-1})e^{Z_\lambda(X_n)}.$$

The following theorems provide false-alarm and delay guarantees for the score-based Shiryaev-Roberts algorithm τ_{ssr}.

Theorem 8.1. *Let ν be a geometrically distributed random variable: $\nu \sim \mathrm{Geom}(\rho)$. Let the value of λ be selected to satisfy*

$$\mathsf{E}_\infty[e^{Z_\lambda(X_1)}] = 1.$$

we have that setting $B = \frac{1-\rho}{\rho\alpha}$ gives us

$$\mathsf{P}(\tau_{ssr} < \nu) \leq \alpha.$$

Proof. As discussed in Wu et al. (2023b), such a λ satisfying the above equation can always be found for all non-trivial change detection problems. Also, note that, unlike the analysis of score-based Shiryaev algorithm, we do not need to constrain the value of the parameter ρ. Next, note that

$$\mathsf{E}_\infty[R_n|\mathcal{F}_{n-1}] = (R_{n-1} + 1)\mathsf{E}_\infty[e^{Z_\lambda(X_n)}] = R_{n-1} + 1 \geq R_{n-1}.$$

Thus, R_n is also a non-negative submartingale. Furthermore, the expected value of R_n satisfies the recursion

$$\mathsf{E}_\infty[R_n] \quad = (\mathsf{E}_\infty[R_{n-1}]+1)\mathsf{E}_\infty\left[e^{Z_\lambda(X_n)}\right] = \mathsf{E}_\infty[R_{n-1}]+1.$$

Thus, $\mathsf{E}_\infty[R_n] = n$. The probability of a false alarm can again be bounded as follows:

$$\begin{aligned}\mathrm{PFA}(\tau_{ssr}) = \mathsf{P}(\tau_{ssr}<\nu) = \sum_{n=1}^{\infty}\pi_n\ \mathsf{P}_n(\tau_{ssr}<n) &= \sum_{n=1}^{\infty}\pi_n\ \mathsf{P}_\infty(\tau_{ssr}<n)\\ &= \sum_{n=1}^{\infty}\pi_n\ \mathsf{P}_\infty\left(\max_{1\le k<n} R_k \ge B\right)\\ &\le \sum_{n=1}^{\infty}\pi_n\ \frac{1}{B}\mathsf{E}_\infty[R_{n-1}]\\ &= \frac{1}{B}\sum_{n=1}^{\infty}\pi_n\ (n-1) = \frac{1}{B}\left(\frac{1}{\rho}-1\right) = \frac{1-\rho}{\rho B}.\end{aligned}$$

Here again, the inequality follows from Doob's submartingale inequality. Thus, setting $B = \frac{1-\rho}{\rho\alpha}$ gives us the desired bound α on the PFA. □

Theorem 8.2. *Let ν be a geometrically distributed random variable: $\nu \sim \mathrm{Geom}(\rho)$. Let λ selected such that*

$$\mathsf{E}_\infty\left[e^{Z_\lambda(X_1)}\right] = 1.$$

Then we have the following results:

(1) *The stopping rule stops almost surely:*

$$\tau_{ssr} < \infty, \ \text{almost surely}, \quad \forall b \ge 0.$$

and

$$\frac{\tau_{ssr}}{\lfloor \log B/(\lambda\ \mathbb{D}_{\mathrm{F}}(f_1\ ||\ f_0))\rfloor} \to 1, \ \text{almost surely}, \qquad \text{as } B\to\infty.$$

(2) The expected value of the stopping rule has the following asymptotics:

$$\mathsf{E}_1[\tau_{ssr}] \le \frac{\log B}{\lambda\ \mathbb{D}_{\mathrm{F}}(f_1\ ||\ f_0)}(1+o(1)), \quad \text{as} \quad B\to\infty.$$

Thus, by setting $B = \frac{1-\rho}{\rho\alpha}$, we get $\mathrm{PFA}(\tau_{sr}) \le \alpha$ *and the delay then* becomes

$$\mathsf{E}_1[\tau_{ssr}] \le \frac{|\log\alpha|}{\lambda\ \mathbb{D}_{\mathrm{F}}(f_1\ ||\ f_0)}(1+o(1)), \quad \text{as} \quad \alpha\to 0.$$

Proof. The only observation we make here is that

$$\sum_{k=1}^{n} e^{\sum_{i=k}^{n} Z_\lambda(X_i)} \ge e^{\sum_{i=1}^{n} Z_\lambda(X_i)}.$$

Hence, the stopping time τ_{ssr} can be bounded by the hitting time of the exponential random walk $e^{\sum_{i=1}^{n} Z_i(\lambda)}$. The rest of the arguments follow from the delay analysis of the score-based Shiryaev algorithm. □

9. BAYESIAN ANALYSIS OF THE SCORE-BASED CUSUM ALGORITHM

In this section, we consider the score-based cumulative sum statistic

$$e^{W_n} = \max_{1\le k\le n} e^{\sum_{i=k}^{n} Z_\lambda(X_i)}$$

and stopping time

$$\tau_{sc} = \min\{n \ge 1 : W_n \ge B\}.$$

The statistic W_n has a recursion:

$$W_n = (W_{n-1} + Z_\lambda(X_n))^+.$$

This algorithm was analyzed in the minimax settings in Wu et al. (2023b).

The following theorems provide a guarantee for the probability of a false alarm for the score-based CUSUM algorithm.

Theorem 9.1. *Let ν be a geometrically distributed random variable: $\nu \sim \text{Geom}(\rho)$. Let the value of λ be selected to satisfy*

$$\mathsf{E}_\infty[e^{Z_\lambda(X_1)}] = 1.$$

we have that setting $B = \frac{1-\rho}{\rho\alpha}$ gives us

$$\mathsf{P}(\tau_{sc} < \nu) \le \alpha.$$

Proof. For the proof, we simply note that

$$R_n \ge e^{W_n}.$$

□

The delay analysis when the change occurs at time 1 is given in Wu et al. (2023b).

10. CONCLUSION

We proposed the Hyvärinen score-based Shiryaev algorithm. We showed that the statistic is a nonnegative submartingale and used it for analyzing the probability of a false alarm for the algorithm. We then showed that the statistic can also be written as a random walk and a slowly changing term, and used this fact to obtain the average detection delay for the algorithm using nonlinear renewal theory. The analysis shows that, similar to the classical Shiryaev algorithm, the threshold of the score-based algorithm can be chosen to guarantee a universal guarantee on the probability of a false alarm. Moreover, while the delay of the classical Shiryaev algorithm is inversely proportional to the Kullback-Leibler divergence between pre- and post-chance densities, the delay of the score-based algorithm is inversely proportional to their Fisher divergence. We also

analyzed score-based variants of the classical Shiryaev-Roberts algorithm and the CUSUM algorithm.

ACKNOWLEDGMENT

We thank the editor, the associate editor, and the referees for their constructive suggestions and comments. This article has significantly improved due to their feedback.

DISCLOSURE

The authors have no conflicts of interest to report.

FUNDING

This material is based upon work supported by the U.S. National Science Foundation under award numbers 2334898 and 2334897.

REFERENCES

Andrieu, Christophe, and Johannes Thoms. 2008. "A Tutorial on Adaptive MCMC." *Statistics and Computing* 18 (4): 343–373. https://doi.org/10.1007/s11222-008-9110-y.

Banerjee, Taposh, Hamed Firouzi, and Alfred O. Hero. 2018. "Quickest Detection for Changes in Maximal kNN Coherence of Random Matrices." *IEEE Transactions on Signal Processing* 66 (17): 4490–4503. https://doi.org/10.1109/TSP.2018.2855644.

Basseville, Michele, and Igor V. Nikiforov. 1993. *Detection of Abrupt Changes: Theory and Application*. Vol. 104. Englewood Cliffs, NJ: Prentice Hall.

Brodsky, Emily, and Boris S. Darkhovsky. 1993. *Nonparametric Methods in Change Point Problems*. Vol. 243. Norwell, MA: Springer Science & Business Media.

Darkhovsky, Boris, and Alexandra Piryatinska. 2014a. "Quickest Detection of Changes in the Generating Mechanism of a Time Series via the ε-Complexity of Continuous Functions." *Sequential Analysis* 33 (2): 231–250. https://doi.org/10.1080/07474946.2014.896698.

Darkhovsky, Boris, and Alexandra Piryatinska. 2014b. "New Approach to the Segmentation Problem for Time Series of Arbitrary Nature." *Proceedings of the Steklov Institute of Mathematics* 287 (1): 54–67. https://doi.org/10.1134/S0081543814080045.

Darkhovsky, Boris, and Alexandra Piryatinska. 2015. "Novel Methodology of Change-Points Detection for Time Series with Arbitrary Generating Mechanisms." In *Stochastic Models, Statistics and Their Applications: Wrocław, Poland, February 2015*, edited by Ansgar Steland, Ewaryst Rafajłowicz and Krzysztof Szajowski, 241–51. Switzerland: Springer.

Darkhovsky, Boris, and Alexandra Piryatinska. 2018a. "Classification of Multivariate Time Series of Arbitrary Nature Based on the ε-Complexity Theory." In *Statistics and Simulation*, 231–242.

Darkhovsky, Boris, and Alexandra Piryatinska. 2018b. "Model-Free Offline Change-Point Detection in Multidimensional Time Series of Arbitrary Nature via ε-Complexity: Simulations and Applications." *Applied Stochastic Models in Business and Industry* 34 (5): 633–644. https://doi.org/10.1002/asmb.2303.

Gordon, Louis, and Moshe Pollak. 1994. "An Efficient Sequential Nonparametric Scheme for Detecting a Change of Distribution." *The Annals of Statistics* 22 (2): 763–804. https://doi.org/10.1214/aos/1176325495.

Hou, Yingze, Yousef Oleyaeimotlagh, Rahul Mishra, Hoda Bidkhori, and Taposh Banerjee. 2023. "Robust Quickest Change Detection in Non-Stationary Processes." arXiv preprint arXiv: 2310.09673.

Hyvärinen, Aapo. 2005. "Estimation of Non-Normalized Statistical Models by Score Matching." *Journal of Machine Learning Research* 6 (4): 695–709.
Kang, Jiwon, and Junmo Song. 2017. "Score Test for Parameter Change in Poisson Autoregressive Models." *Economics Letters* 160: 33–37. https://doi.org/10.1016/j.econlet.2017.08.021.
Konev, Victor, and Sergey Vorobeychikov. 2017. "Quickest Detection of Parameter Changes in Stochastic Regression: Nonparametric CUSUM." *IEEE Transactions on Information Theory* 63 (9): 1. https://doi.org/10.1109/TIT.2017.2673825.
Lai, Tze Leung. 1998. "Information Bounds and Quick Detection of Parameter Changes in Stochastic Systems." *IEEE Transactions on Information Theory* 44 (7): 2917–2929.
Lau, Tze Siong, Wee Peng Tay, and Venugopal V. Veeravalli. 2019. "A Binning Approach to Quickest Change Detection with Unknown Post-Change Distribution." *IEEE Transactions on Signal Processing* 67 (3): 609–621. https://doi.org/10.1109/TSP.2018.2881666.
Liang, Yuchen, and Venugopal V. Veeravalli. 2022. "Non-Parametric Quickest Mean-Change Detection." *IEEE Transactions on Information Theory* 68 (12): 8040–8052. https://doi.org/10.1109/TIT.2022.3191957.
Liu, Qiang, and Dilin Wang. 2016. "Stein Variational Gradient Descent: A General Purpose Bayesian Inference Algorithm." Advances in Neural Information Processing Systems 29, Barcelona, Spain.
Lorden, G. 1971. "Procedures for Reacting to a Change in Distribution." *The Annals of Mathematical Statistics* 42 (6): 1897–1908. https://doi.org/10.1214/aoms/1177693055.
Moustakides, G. V. 1986. "Optimal Stopping Times for Detecting Changes in Distributions." *The Annals of Statistics* 14 (4): 1379–1387.
Oleyaeimotlagh, Yousef, Taposh Banerjee, Ahmad Taha, and Eugene John. 2023. "Quickest Change Detection in Statistically Periodic Processes with Unknown Post-Change Distribution." *Sequential Analysis* 42 (4): 404–437. https://doi.org/10.1080/07474946.2023.2247035.
Page, E. S. 1954. "Continuous Inspection Schemes." *Biometrika* 41 (1-2): 100–115. https://doi.org/10.1093/biomet/41.1-2.100.
Pawlak, Mirosław, and Ansgar Steland. 2013. "Nonparametric Sequential Signal Change Detection under Dependent Noise." *IEEE Transactions on Information Theory* 59 (6): 3514–3531. https://doi.org/10.1109/TIT.2013.2243200.
Pollak, Moshe. 1985. "Optimal Detection of a Change in Distribution." *The Annals of Statistics* 13 (1): 206–227.
Polunchenko, Aleksey S., and Alexander G. Tartakovsky. 2012. "State-of-the-Art in Sequential Change-Point Detection." *Methodology and Computing in Applied Probability* 14 (3): 649–684. https://doi.org/10.1007/s11009-011-9256-5.
Poor, H. Vincent, and Olympia Hadjiliadis. 2009. *Quickest Detection*. New York, USA: Cambridge University Press.
Sarnowski, Wojciech, and Krzysztof Szajowski. 2011. "Optimal Detection of Transition Probability Change in Random Sequence." *Stochastics* 83 (4-6): 569–581. https://doi.org/10.1080/17442508.2010.540015.
Shiryaev, Albert N. 1963. "On Optimum Methods in Quickest Detection Problems." *Theory of Probability & Its Applications* 8 (1): 22–46. https://doi.org/10.1137/1108002.
Shiryaev, Albert N. 2007. *Optimal Stopping Rules*. Vol. 8. New York, USA: Springer Science & Business Media.
Song, Junmo, and Jiwon Kang. 2020. "Sequential Change Point Detection in ARMA-GARCH Models." *Journal of Statistical Computation and Simulation* 90 (8): 1520–1538. https://doi.org/10.1080/00949655.2020.1734807.
Song, Yang, and Stefano Ermon. 2019. "Generative Modeling by Estimating Gradients of the Data Distribution." Advances in Neural Information Processing Systems 32, Vancouver, Canada.
Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. "Score-Based Generative Modeling through Stochastic Differential Equations." In International Conference on Learning Representations ICLR 2021, Vienna, Austria.
Tartakovsky, Alexander. 2019. *Sequential Change Detection and Hypothesis Testing: General Non-i.i.d. Stochastic Models and Asymptotically Optimal Rules*. Boca Raton, FL: CRC Press.

Tartakovsky, Alexander G. 2017. "On Asymptotic Optimality in Sequential Changepoint Detection: Non-Iid Case." *IEEE Transactions on Information Theory* 63 (6): 3433–3450. https://doi.org/10.1109/TIT.2017.2683496.

Tartakovsky, Alexander G., Igor V. Nikiforov, and Michele. Basseville. 2014. *Sequential Analysis: Hypothesis Testing and Change-Point Detection*. Statistics. Boca Raton, FL: CRC Press.

Tartakovsky, Alexander G., and V. V. Veeravalli. 2005. "General Asymptotic Bayesian Theory of Quickest Change Detection." *Theory of Probability & Its Applications* 49 (3): 458–497. https://doi.org/10.1137/S0040585X97981202.

Unnikrishnan, Jayakrishnan, Venugopal V. Veeravalli, and Sean P. Meyn. 2011. "Minimax Robust Quickest Change Detection." *IEEE Transactions on Information Theory* 57 (3): 1604–1614. https://doi.org/10.1109/TIT.2011.2104993.

Veeravalli, Venugopal V., and Taposh Banerjee. 2014. "Quickest Change Detection." In *Academic Press Library in Signal Processing*, Vol. 3, 209–255. Elsevier. https://www.sciencedirect.com/bookseries/academic-press-library-in-signal-processing.

Vincent, Pascal. 2011. "A Connection between Score Matching and Denoising Autoencoders." *Neural Computation* 23 (7): 1661–1674. https://doi.org/10.1162/NECO_a_00142.

Woodroofe, Michael. 1982. *Nonlinear Renewal Theory in Sequential Analysis*. Philadelphia, PA: SIAM.

Wu, Suya, Enmao Diao, Taposh Banerjee, Jie Ding, and Vahid Tarokh. 2023a. "Robust Quickest Change Detection for Unnormalized Models." Conference on Uncertainty in Artificial Intelligence (UAI), Pittsburgh, PA.

Wu, Suya, Enmao Diao, Taposh Banerjee, Jie Ding, and Vahid Tarokh. 2023b. "Score-Based Change Point Detection for Unnormalized Models." International Conference on Artificial Intelligence and Statistics (AISTATS), Valencia, Spain.

Wu, Suya, Enmao Diao, Taposh Banerjee, Jie Ding, and Vahid Tarokh. 2024. "Quickest Change Detection for Unnormalized Statistical Models." *IEEE Transactions on Information Theory* 70 (2): 1220–1232. https://doi.org/10.1109/TIT.2023.3328274.

Xie, Liyan, Shaofeng Zou, Yao Xie, and Venugopal V. Veeravalli. 2021. "Sequential (Quickest) Change Detection: Classical Results and New Directions." *IEEE Journal on Selected Areas in Information Theory* 2 (2): 494–514. https://doi.org/10.1109/JSAIT.2021.3072962.