SEQUENTIAL ANALYSIS
2017, VOL. 36, NO. 3, 355–369
https://doi.org/10.1080/07474946.2017.1360088

Taylor & Francis
Taylor & Francis Group

Minimax optimality of Shiryaev-Roberts procedure for quickest drift change detection of a Brownian motion

Taposh Banerjee[a] and George V. Moustakides[b]

[a]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, USA; [b]Department of Electrical and Computer Engineering, University of Patras, Rion, Greece; Department of Computer Science, Rutgers University, New Brunswick, New Jersey, USA

ABSTRACT
The problem of detecting a change in the drift of a Brownian motion is considered. The change point is assumed to have a modified exponential prior distribution with unknown parameters. A worst-case analysis with respect to these parameters is adopted leading to a min–max problem formulation. Analytical and numerical justifications are provided toward establishing that the Shiryaev-Roberts procedure with a specially designed starting point is exactly optimal for the proposed mathematical setup.

ARTICLE HISTORY
Received 21 October 2016
Revised 18 December 2016
Accepted 2 March 2017

KEYWORDS
Brownian motion; drift change detection; minimax optimality; quickest change detection; Shiryaev-Roberts procedure

MATHEMATICS SUBJECT CLASSIFICATIONS
62L05; 62L10; 62L15; 62F05; 62F15; 60G40

1. Introduction

Consider a continuous-time stochastic process $\{\xi_t\}$ of the form

$$d\xi_t = \mu dt \mathbb{1}_{\{t\geq\tau\}} + dw_t, \quad t \geq 0,\ \ \xi_0 = 0,\ \ \mu \neq 0, \tag{1.1}$$

where $\{w_t\}$ is a Wiener process and τ is a real-valued random variable independent of $\{w_t\}$. The variable τ is to be interpreted as the change-point at which there is a change in the drift: for $t \leq \tau$, ξ_t is a standard Brownian motion, and for $t > \tau$, it is a Brownian motion with a known drift $\mu \in \mathbb{R}$. We consider the problem of detecting this change using a stopping time T adapted to the filtration generated by the process $\{\xi_t\}_{t\geq 0}$, with minimum possible delay $T-\tau$, subject to a constraint on false alarms $\{T \leq \tau\}$. To simplify our presentation, from now on, without loss of generality we assume that $\mu = \sqrt{2}$. Indeed, any other value of $\mu \neq 0$ can be reduced to $\sqrt{2}$ just by a simple change in time scale of the process $\{\xi_t\}$ and a change in sign if $\mu < 0$.

When the random variable τ with $\tau \in \mathbb{R}$ has a zero-modified exponential prior, the Bayesian version of the quickest change detection problem was studied by Shiryaev (1963) and the corresponding optimum test is known as the Shiryaev test. In this work, we are interested in the case where the parameters of the zero-modified exponential prior are *unknown* and

CONTACT Taposh Banerjee tbanerjee@seas.harvard.edu School of Engineering and Applied Sciences, Harvard University, 33 Oxford St, MD 321, Cambridge, MA 02138.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsqa.
Recommended by S. Lee

we follow a worst-case analysis to cope with this lack of information. Our claim is that the Shiryaev-Roberts procedure with a specially designed deterministic starting point, known as the SR-r procedure (see Moustakides et al., 2011), is exactly optimal for the proposed formulation. In fact, we provide analytical and numerical evidence to support this claim.

We now state the problem formulation and the main results of this article rigorously. The observation process $\{\xi_t\}$ is as in (1.1). The zero-modified exponential prior for the change-point τ is such that

$$\mathsf{P}(\tau \leq 0) = \pi, \quad \mathsf{P}(\tau \in dt) = (1-\pi)\lambda e^{-\lambda t} dt, \quad t \geq 0, \tag{1.2}$$

for some $\lambda \geq 0$ and $\pi \in [0,1]$. The assumption of a zero-modified exponential prior is fundamental to our work, and will play a crucial role in what follows. However, for analytical convenience, it is necessary to change the corresponding parametrization. In particular, we define $r = \frac{\pi}{\lambda}$ suggesting

$$\mathsf{P}(\tau \leq 0) = r\lambda, \;\; \mathsf{P}(\tau \in dt) = (1-r\lambda)\lambda e^{-\lambda t} dt, \;\; t \geq 0, \tag{1.3}$$

with $\lambda \geq 0$ and $\frac{1}{\lambda} \geq r \geq 0$.

Regarding probability measures, we use P_t to denote the measure incurred, when the change-time takes upon the deterministic value $\tau = t$ and reserve E_t for the corresponding expectation. With this definition we have that P_∞ corresponds to the probability measure when all observations are under the nominal regime while P_0 when the observations are under the alternative. Combining the previous measures with the prior on τ produces $\mathsf{P}_{r,\lambda}$ and $\mathsf{E}_{r,\lambda}$; that is, the probability measure and expectation when the change-time τ is random.

If $\{\mathscr{F}_t\}_{t\geq 0}$ denotes the filtration generated by the observations; that is, $\mathscr{F}_t = \sigma(\xi_s : 0 \leq s \leq t)$, with $\mathscr{F}_0$ the trivial sigma-algebra then, for detection we seek an $\{\mathscr{F}_t\}$-adapted stopping time T that will detect the change in the drift as quickly as possible, subject to a constraint on the false alarm rate. When the pair (r, λ) is known, Shiryaev (1963) proposed the following formulation:

$$\inf_T \mathsf{E}_{r,\lambda}[T - \tau^+ | T > \tau], \quad \text{subject to: } \mathsf{P}_{r,\lambda}(T \leq \tau) \leq \alpha, \tag{1.4}$$

where $x^+ = \max\{x, 0\}$ and $\alpha \in [0,1]$ a known false alarm probability level.

In the current work, unlike (1.4), we consider (r, λ) to be *unknown*. In order to deal with this lack of information, we adopt a worst-case analysis with respect to the parameter pair. We therefore propose the following min–max constrained optimization alternative

$$\inf_T \sup_{r,\lambda} \mathsf{E}_{r,\lambda}[T - \tau^+ | T > \tau], \quad \text{subject to: } \mathsf{E}_\infty[T] \geq \gamma, \tag{1.5}$$

where γ is a constant that constrains the average period of false alarms. The switching from the false alarm probability appearing in (1.4) to the average false alarm period adopted in (1.5) is common for min-max approaches (e.g., see Moustakides, 2014). This change is necessary since the false alarm probability in (1.4) depends on the unknown parameter pair and would therefore require an additional worst-case analysis for the constraint. Unfortunately, the worst-case false alarm probability cannot be efficiently controlled (actually very often it takes the value 1), thus making the constrain meaningless. This is the reason why it is replaced by the average false alarm period that is independent from the unknown parameters.

To complete our introduction, we need some additional definitions that are necessary for our analysis. Consider the process

$$du_t = -dt + \sqrt{2}d\xi_t, \; u_0 = 0, \tag{1.6}$$

then from Peskir and Shiryaev (2006, chapter VI, section 22), and Girsanov's theorem (see Rogers and Williams, 2000), we have

$$\frac{d\mathsf{P}_0}{d\mathsf{P}_\infty}(\mathscr{F}_t) = e^{u_t}, \; t \geq 0,$$

and more generally for $s \geq 0$

$$\frac{d\mathsf{P}_s}{d\mathsf{P}_\infty}(\mathscr{F}_t) = \begin{cases} e^{u_t - u_s} & s \leq t \\ 1 & s > t. \end{cases} \tag{1.7}$$

It is clear that e^{u_t} is the Radon-Nikodym derivative between the two probability measures $\mathsf{P}_0, \mathsf{P}_\infty$ limited to $\mathscr{F}_t$ and $e^{u_t - u_s}$ is the Radon-Nikodym derivative between $\mathsf{P}_s, \mathsf{P}_\infty$ on the same sigma-algebra when $t \geq s$.

Consider now the following statistic, which will play a key role in our analysis:

$$R_t = e^{u_t}\left\{r_* + \int_0^t e^{-u_s} ds\right\}, \tag{1.8}$$

where $r_* \geq 0$ is a specially designed initial point (since $R_0 = r_*$) that will be specified exactly in the sequel. Define now the following function:

$$g(R) = e^{(r_*+\gamma)^{-1}} E_1\left((r_* + \gamma)^{-1}\right) - e^{R^{-1}} E_1\left(R^{-1}\right), \tag{1.9}$$

where $E_1(x) = \int_x^\infty \frac{e^{-z}}{z} dz$ is the exponential integral (see Abramowitz and Stegun, 1965, chapter 5), r_* is the parameter we introduced in the definition of R_t in (1.8), and γ the constraint on the average false alarm period in (1.5). The next lemma contains a number of interesting equalities that will be used throughout our analysis. The most important one consists of providing an alternative form for our performance measure.

Lemma 1.1. *If R_t is as in (1.8) and T an $\{\mathscr{F}_t\}$-adapted stopping time, then we have the following equalities that are valid:*

$$\mathsf{E}_\infty[R_T] = r_* + \mathsf{E}_\infty[T] \tag{1.10}$$

$$\mathsf{E}_t[(T-t)^+ | \mathscr{F}_t] = \mathsf{E}_t[g(R_t) - g(R_T) | \mathscr{F}_t] \mathbb{1}_{\{T>t\}}. \tag{1.11}$$

Furthermore,

$$\mathscr{D}(T, r, \lambda) = \mathsf{E}_{r,\lambda}[T - \tau^+ | T > \tau] = \frac{r\mathsf{E}_0[g(r_*) - g(R_T)] + (1-\lambda r)\int_0^\infty \mathsf{E}_t\left[\left(g(R_t) - g(R_T)\right)\mathbb{1}_{\{T>t\}}\right] e^{-\lambda t} dt}{r + (1-\lambda r)\mathsf{E}_\infty\left[\int_0^T e^{-\lambda t} dt\right]}. \tag{1.12}$$

When $r = r_*$ and $\lambda = 0$, then we can also write

$$\mathscr{D}(T, r_*, 0) = \frac{\mathsf{E}_\infty\left[\int_0^T R_t dt\right]}{r_* + \mathsf{E}_\infty[T]}. \tag{1.13}$$

Proof. The proof of this lemma is presented in Appendix A. □

1.1. Saddle-point problem

With the help of Lemma 1.1, the min–max problem depicted in (1.5) can be equivalently expressed as

$$\inf_T \sup_{r,\lambda} \mathscr{D}(T, r, \lambda), \quad \text{subject to: } \mathsf{E}_\infty[T] \geq \gamma. \tag{1.14}$$

As is the case in most min–max problems, it is possible to obtain their solution by solving a simpler *saddle-point* alternative. Boyd and Vandenberghe (2004, Section 5.1), report that when a saddle-point solution exists it is also the solution of the min-max problem. The opposite is not necessarily true. In particular, we are interested in a triplet $T_*, r_*, \lambda_* = 0$ such that for any $\lambda \geq 0$ and $\frac{1}{\lambda} \geq r \geq 0$ we have validity of the following double inequality:

$$\mathscr{D}(T, r_*, 0) \geq \mathscr{D}(T_*, r_*, 0) \geq \mathscr{D}(T_*, r, \lambda), \quad \text{subject to: } \mathsf{E}_\infty[T] \geq \gamma. \tag{1.15}$$

We should point out that with $\lambda_* = 0$ the exponential prior becomes a degenerate uniform.

As we mentioned, it is a well-established fact that the solution to the saddle-point problem in (1.15) is also the solution to the min–max problem in (1.14). We therefore focus on (1.15).

2. Main results

Our first goal is to specify completely the triplet T_*, r_*, λ_*. So far we have that $\lambda_* = 0$. Let us now define T_* in terms of r_*. For this to be possible, we focus on the first inequality of the saddle-point problem in (1.15), which requires $\mathscr{D}(T, r_*, 0) \geq \mathscr{D}(T_*, r_*, 0)$ for all T satisfying the constraint $\mathsf{E}_\infty[T] \geq \gamma$. In fact, we realize that T_* must solve the following constrained minimization problem:

$$\inf_T \mathscr{D}(T, r_*, 0) = \mathscr{D}(T_*, r_*, 0), \quad \text{subject to: } \mathsf{E}_\infty[T] \geq \gamma. \tag{2.1}$$

Minimizing $\mathscr{D}(T, r_*, 0)$ over T is straightforward and the optimum stopping time is given in the next lemma.

Lemma 2.1. *The stopping time that solves the constrained minimization problem depicted in* (2.1) *is given by*

$$T_* = \inf\{t > 0 : R_t \geq \gamma + r_*\}. \tag{2.2}$$

Proof. To prove this lemma, we use the expression for $\mathscr{D}(T, r_*, 0)$ provided in (1.13). We are interested in showing that among all T that satisfy the false alarm constraint $\mathsf{E}_\infty[T] \geq \gamma$, the stopping time that solves the minimization

$$\inf_T \frac{\mathsf{E}_\infty\left[\int_0^T R_t dt\right]}{r_* + \mathsf{E}_\infty[T]}$$

is T_* defined in (2.2). This is a known result in discrete time (see Polunchenko and Tartakovsky, 2010). The continuous-time version follows a similar line of proof and uses classical optimal stopping arguments. The analysis presents no special difficulties; for this reason, we do not provide any further details. We only point out that T_* satisfies the constraint with equality. Indeed, from (1.10) and since $R_{T_*} = r_* + \gamma$, we have $r_* + \gamma = \mathsf{E}_\infty[R_{T_*}] = r_* + \mathsf{E}_\infty[T_*]$, from which we conclude that $\mathsf{E}_\infty[T_*] = \gamma$. □

The candidate stopping time T_* is specified in terms of r_*, which is still unknown. To define r_*, we make use of the second inequality in the saddle-point problem (1.15), namely, that $\mathscr{D}(T_*, r_*, 0) \geq \mathscr{D}(T_*, r, \lambda)$ for all $\lambda \geq 0$ and $\frac{1}{\lambda} \geq r \geq 0$. Since the second inequality in (1.15) must be true for all $\lambda \geq 0$ it must certainly be valid for $\lambda = 0$. This implies that r_* must be such that for any $r \geq 0$ we have $\mathscr{D}(T_*, r_*, 0) \geq \mathscr{D}(T_*, r, 0)$. In other words, r_* must maximize $\mathscr{D}(T_*, r, 0)$ over r. In (1.12), substituting $T = T_*, \lambda = 0$, recalling that $R_{T_*} = r_* + \gamma$ and $g(R_{T_*}) = g(r_* + \gamma) = 0$, we can write

$$\mathscr{D}(T_*, r, 0) = \frac{rg(r_*) + \int_0^\infty \mathsf{E}_t\left[g(R_t)\mathbb{1}_{\{T_*>t\}}\right]dt}{r + \mathsf{E}_\infty[T_*]}$$

$$= \frac{rg(r_*) + \int_0^\infty \mathsf{E}_\infty\left[g(R_t)\mathbb{1}_{\{T_*>t\}}\right]dt}{r + \mathsf{E}_\infty[T_*]} = \frac{rg(r_*) + \mathsf{E}_\infty\left[\int_0^{T_*} g(R_t)dt\right]}{r + \gamma},$$

where the second equality is due to the fact that $g(R_t)\mathbb{1}_{\{T_*>t\}}$ is $\mathscr{F}_t$-measurable and on $\mathscr{F}_t$ we know that P_t coincides with P_∞. To maximize $\mathscr{D}(T_*, r, 0)$ over r, we observe in the last ratio, that both the numerator and the denominator are linear functions of r; therefore, the ratio is maximized either for $r = 0$ or $r = \infty$. In order for the maximum to be attained by any other value between these two extremes, we need

$$g(r_*) = \frac{\mathsf{E}_\infty\left[\int_0^{T_*} g(R_t)dt\right]}{\gamma}, \quad \text{or, equivalently,} \quad \mathsf{E}_\infty\left[\int_0^{T_*} \big(g(R_t) - g(r_*)\big)dt\right] = 0, \tag{2.3}$$

where for the last equation we used the fact that $\mathsf{E}_\infty[T_*] = \gamma$. Condition (2.3) is the equation through which we can compute r_*. Interestingly, the same condition also assures that $\mathscr{D}(T_*, r, 0) = g(r_*)$, that is, that $\mathscr{D}(T_*, r, 0)$ is constant independent of r, namely, an *equalizer* over r.

Summarizing: For the solution of the min–max problem in (1.5) we propose the candidate stopping time T_* defined in (2.2), where the parameter r_* is obtained by solving (2.3). Regarding (2.3), in the next section we offer a more analytic expression.

2.1. Optimality of the proposed test

The optimality of our candidate stopping time is assured if we can show that the two inequalities in the saddle-point problem (1.15) are true. We note that T_* was constructed so that the first inequality is valid for all T satisfying the false alarm constraint. Regarding the second inequality, by selecting r_* through (2.3) we guarantee $g(r_*) = \mathscr{D}(T_*, r, 0)$ for all $r \geq 0$. However, for optimality we need to demonstrate the stronger version

$$g(r_*) \geq \mathscr{D}(T_*, r, \lambda). \tag{2.4}$$

The next lemma presents a condition that can replace (2.4) and is easier to verify.

Lemma 2.2. *The inequality in (2.4) is equivalent to*

$$\mathsf{E}_\infty\left[\int_0^{T_*} e^{-\lambda t}\big(g(R_t)-g(r_*)\big)dt\right]\leq 0,\ \forall\lambda\geq 0. \tag{2.5}$$

Proof. The proof is simple. Replacing T with T_* in the definition of $\mathscr{D}(T,r,\lambda)$ in (1.12) and using the boundary condition $g(R_{T_*})=g(r_*+\gamma)=0$, we conclude that (2.4) is true iff

$$g(r_*)\geq\frac{rg(r_*)+(1-\lambda r)\mathsf{E}_\infty\left[\int_0^{T_*}e^{-\lambda t}g(R_t)dt\right]}{r+(1-\lambda r)\mathsf{E}_\infty\left[\int_0^{T_*}e^{-\lambda t}dt\right]},$$

is valid for all $\lambda\geq 0$ and $\frac{1}{\lambda}\geq r\geq 0$. The above inequality is clearly equivalent to (2.5) for $\frac{1}{\lambda}>r\geq 0$, whereas it is trivially valid when $r=\frac{1}{\lambda}$. □

The next Lemma provides a differential equation and suitable conditions for the computation of the left-hand-side expectation in (2.5).

Lemma 2.3. *Fix $\lambda\geq 0$, if $f_\lambda(R)$ is a twice differentiable function of R that is the solution of the ODE (ordinary differential equation).*

$$-\lambda f_\lambda(R)+f_\lambda'(R)+R^2f_\lambda''(R)=-\big(g(R)-g(r_*)\big)=e^{R^{-1}}E_1\left(R^{-1}\right)-e^{r_*^{-1}}E_1\left(r_*^{-1}\right), \tag{2.6}$$

with $f_\lambda(R)$ bounded when $R\in[0,r_*+\gamma]$ and $f_\lambda(r_*+\gamma,\lambda)=0$, then

$$f_\lambda(r_*)=\mathsf{E}_\infty\left[\int_0^{T_*}e^{-\lambda t}\big(g(R_t)-g(r_*)\big)dt\right]. \tag{2.7}$$

Proof. The proof is detailed in Appendix A. □

An analytic form for $f_0(R)$ (i.e., $f_\lambda(R)$ when $\lambda=0$) and how this function can be used in order to obtain an integral instead of a differential equation for $f_\lambda(R)$ when $\lambda>0$ is presented in the next lemma.

Lemma 2.4. *If $f_\lambda(R)$ is as in Lemma 2.3, then for $\lambda=0$ the corresponding function $f_0(R)$ is equal to*

$$f_0(R)=\{1-e^{r_*^{-1}}E_1(r_*^{-1})\}(R-r^*-\gamma)+\int_{(r^*+\gamma)^{-1}}^{R^{-1}}E_1(x)d\big(E_i(x)\big), \tag{2.8}$$

while $f_\lambda(R)$ when $\lambda>0$ satisfies the following integral equation:

$$\begin{aligned}f_\lambda(R)=f_0(R)-\lambda\Bigg\{&\Big(E_i\big((r^*+\gamma)^{-1}\big)-e^{(r^*+\gamma)^{-1}}(r^*+\gamma)-E_i(R^{-1})+e^{R^{-1}}R\Big)\\&\times\int_{(r^*+\gamma)^{-1}}^{\infty}f_\lambda(z^{-1})d(e^{-z})-\int_{(r^*+\gamma)^{-1}}^{R^{-1}}f_\lambda(z^{-1})d\big(e^{-z}E_i(z)\big)\\&+\Big(E_i(R^{-1})-e^{R^{-1}}R\Big)\int_{(r^*+\gamma)^{-1}}^{R^{-1}}f_\lambda(z^{-1})d(e^{-z})\Bigg\},\end{aligned} \tag{2.9}$$

where $E_i(x)=\int_{-\infty}^{x}\frac{e^z}{z}dz$ is the second version of the exponential integral.

Proof. The details of the proof are presented in Appendix A. □

The function $f_0(R)$ enjoys an additional notable property. Comparing (2.7) with (2.3), we observe that we can recover the expectation in (2.3) by computing $f_0(r_*)$. This suggests that the corresponding equation can be written as $f_0(r_*) = 0$. Using (2.8) and substituting $R = r_*$, we obtain the final form of the equation that identifies r_* and replaces (2.3):

$$f_0(r_*) = -\gamma\{1 - e^{r_*^{-1}}E_1(r_*^{-1})\} + \int_{(r_*+\gamma)^{-1}}^{r_*^{-1}} E_1(x)d\big(E_i(x)\big) = 0. \tag{2.10}$$

To complete the proof of optimality for T_*, we need to establish the validity of (2.5), which, because of (2.7), is equivalent to showing that

$$f_\lambda(r_*) \leq 0, \tag{2.11}$$

where $f_\lambda(R)$ satisfies the integral equation in (2.9). Unfortunately, this last step *was not possible to demonstrate analytically*. Therefore, we state the following claim:

Conjecture. *The inequality $f_\lambda(r_*) \leq 0$ is true for all $\lambda \geq 0$.*

The validity of this claim establishes exact optimality of the candidate stopping time T_* defined in (2.2) in the sense that it is min–max optimum according to the problem proposed in (1.14). Of course, our conjecture constitutes a crucial part of the optimality proof for T_*. Even though we cannot support our claim analytically, we intend to provide *numerical evidence* for its validity by directly computing $f_\lambda(r_*)$ and examining its sign. To achieve this goal, we develop a simple computational method by borrowing ideas from Moustakides et al. (2011). In fact, as we will see next, the expressions for $f_0(R)$ and $f_\lambda(R)$ proposed in (2.8) and (2.9), respectively, are properly set for the numerical computation of the two functions.

2.2. Numerical method

To numerically evaluate $f_0(R)$ and $f_\lambda(R)$, we need to compute integrals of the form $\int_\alpha^\beta a(x)d\big(b(x)\big)$ where $a(x), b(x)$ are functions of x and $d\big(b(x)\big) = b'(x)dx$ denotes the differential of $b(x)$. If we sample the interval $[\alpha, \beta]$ (not necessarily canonically) at the points $\alpha = x_0 < x_1 < \cdots < x_N = \beta$, then using the simple trapezoidal rule we can approximate the corresponding integral by the following sum:

$$\begin{aligned}\int_\alpha^\beta a(x)d\big(b(x)\big) &\approx \sum_{n=1}^{N} \frac{a(x_n) + a(x_{n-1})}{2}\big(b(x_n) - b(x_{n-1})\big)\\ &= \frac{b(x_1) - b(x_0)}{2}a(x_0) + \sum_{n=1}^{N-1} \frac{b(x_{n+1}) - b(x_{n-1})}{2}a(x_n)\\ &\quad + \frac{b(x_N) - b(x_{N-1})}{2}a(x_N).\end{aligned} \tag{2.12}$$

The last sum in (2.12) can be clearly written as the inner product $\mathbf{b}^t\mathbf{a}$ of the two vectors

$$\mathbf{a} = [a(x_0), a(x_1), \ldots, a(x_N)]^t$$

$$\mathbf{b} = \frac{1}{2}[b(x_1) - b(x_0), b(x_2) - b(x_0), \ldots, b(x_N) - b(x_{N-2}), b(x_N) - b(x_{N-1})]^t.$$

This straightforward idea can be applied in (2.10) for the computation of the corresponding integral and the evaluation of the function $f_0(r_*)$ for any given r_*. Furthermore, with the help of an elementary bisection method, we can then easily approximate the root of the equation $f_0(r_*) = 0$ and obtain the initializing point $R_0 = r_*$ of our test statistic R_t.

Once r_* is specified, we can attempt to solve the integral equation (2.9) in order to compute the function $f_\lambda(R)$. In fact, as we can see from the equations, it is more convenient to compute $f_\lambda(R^{-1})$ since it is the actual function used in the corresponding integrals. We first sample the interval $(0, r_* + \gamma]$ at a sufficient number of points. A small but crucial technicality is that we must avoid the value $R = 0$ because it is the source of numerical instability. We can instead select a point that is sufficiently close to 0 but avoids the product of a very large with a very small number (which is the source of the observed instability). Among our sampled values we must include r_* since we are interested in (the sign of) $f_\lambda(r_*)$.

Call $\mathbf{f}_\lambda$ the vector version of the samples of $f_\lambda(R)$ and $\mathbf{f}_0$ the corresponding vector for the samples of $f_0(R)$. In the sampled version of the integral equation (2.9), if we approximate the three integrals using the idea proposed in (2.12), we end up with the following system of linear equations:

$$\mathbf{f}_\lambda = \mathbf{f}_0 - \lambda \mathbf{P} \mathbf{f}_\lambda.$$

Matrix $\mathbf{P}$ summarizes the contribution of the three integrals that use the function $f_\lambda(R)$. The reason we need a matrix (and not a vector) is because we evaluate (2.9) for the complete collection of samples of R. Each sample requires its own vector $\mathbf{b}$, which contributes a row to the matrix $\mathbf{P}$. It is clear that the product $\mathbf{P}\mathbf{f}_\lambda$ evaluates the sum of the three integrals for all sampled values of R at the same time. Solving for $\mathbf{f}_\lambda$ yields

$$\mathbf{f}_\lambda = (\mathbf{I} + \lambda \mathbf{P})^{-1} \mathbf{f}_0. \tag{2.13}$$

From the solution vector $\mathbf{f}_\lambda$, we only need to retain the term corresponding to $f_\lambda(r_*)$. We note that $\mathbf{f}_0, \mathbf{P}$ must be computed only once, since they do not depend on λ. By changing the value of the scalar λ, we can then find $f_\lambda(r_*)$ for different values of this parameter and examine its sign to verify the validity of (2.11).

3. Examples

Let us apply the numerical method we introduced above to the case where the average false alarm period takes the values $\gamma = 5$ and 20. The next two figures depict our numerical results.

In Figure 1(a) we plot $f_0(r_*)$ as a function of r_* for $\gamma = 5$. For the computation of the integral in (2.10) we used 501 samples in the interval $[r_*, r_* + \gamma]$. The bisection method estimated the root of $f_0(r_*) = 0$ to be $r_* = 1.0707$. This is the value we adopted for this parameter. For the computation of $f_\lambda(R)$, we sampled the interval $[2 \times 10^{-3}, r_* + \gamma]$ at 2,001

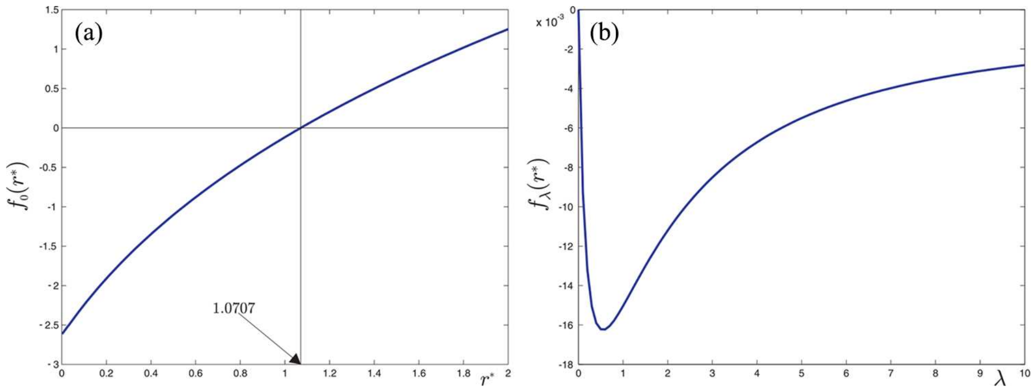

Figure 1. (a) Plot of $f_0(r_*)$ as a function of r_* for $\mu = \sqrt{2}$ and $\gamma = 5$. The point at which the function becomes 0 is $r_* = 1.0707$. (b) Plot of $f_\lambda(r_*)$ as a function of $\lambda \geq 0$.

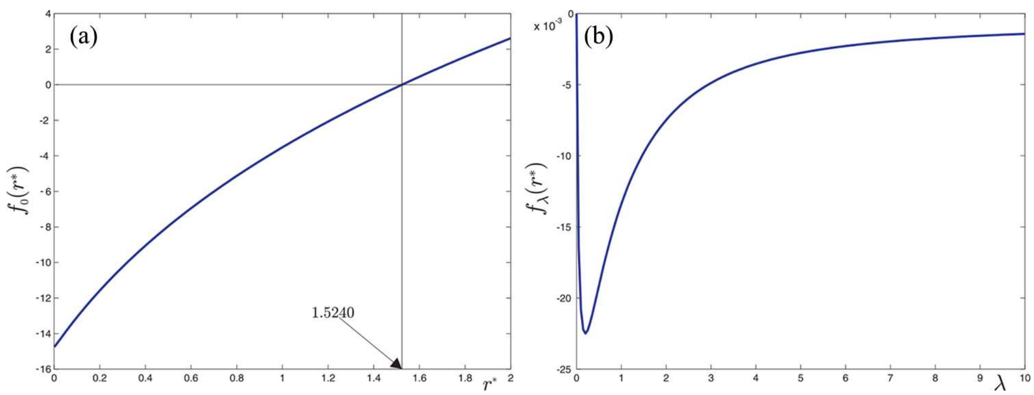

Figure 2. (a) Plot of $f_0(r_*)$ as a function of r_* for $\mu = \sqrt{2}$ and $\gamma = 20$. The point at which the function becomes 0 is $r_* = 1.5240$. (b) Plot of $f_\lambda(r_*)$ as a function of $\lambda \geq 0$.

points retaining the 501 we used for the determination of r_*. We then solved the linear system in (2.13) for 100 values of λ selected canonically from the interval (0, 10]. The resulting $f_\lambda(r_*)$ appears in Figure 1(b). We can see that this function is clearly negative thus supporting our conjecture.

In Figures 2(a) and 2(b), we present our numerical results for the false alarm value $\gamma = 20$. Here the bisection method yielded $r_* = 1.5240$. For the computation of $f_0(r_*)$ and $f_\lambda(R)$, we used 1,001 and 4,001 samples, respectively, where for the latter case, as before, we sampled the interval $[2 \times 10^{-3}, r_* + \gamma]$. Finally, we canonically selected 200 samples for λ from the interval (0, 10]. The resulting function $f_\lambda(r_*)$ is depicted in Figure 2(b) and, as we can see, it is again negative, thus supporting, once more, our claim.

We should mention that we have performed numerous similar computations for various γ that ranged from small to large values. In all cases, $f_\lambda(r_*)$ turned out to be a negative function of λ. Of course, it is understood that these observations cannot serve, by any means, as a formal proof of optimality for T_*. However, finding the proper formulas for the numerical computation demanded a serious mathematical analysis, and the final outcome, undeniably, supports our conjecture and the optimality of our detector.

4. Discussion

We must point out that the proposed stopping time T_* in (2.2) is known as the Shiryaev-Roberts-r (SR-r) test and has already been considered in the analysis of Pollak's performance measure (Pollak, 1985):

$$J_{\mathrm{P}}(T) = \sup_{t\geq 0} \mathsf{E}_t[T - t | T > t]. \quad (4.1)$$

In the same article, the following constrained min-max optimization problem was suggested:

$$\inf_T J_{\mathrm{P}}(T) = \inf_T \sup_{t\geq 0} \mathsf{E}_t[T - t | T > t], \text{ subject to: } \mathsf{E}_\infty[T] \geq \gamma > 0, \quad (4.2)$$

for the determination of an optimum detection strategy. Pollak (1985) was able to prove that the discrete time analog of the stopping time T_* in (2.2) is *asymptotically optimum* in a very strong sense, provided that the deterministic starting point $R_0 = r_*$ is replaced by a *random variable* that follows the quasi stationary distribution. More precisely, he demonstrated that

$$J_{\mathrm{P}}(T_*) - \inf_T J_{\mathrm{P}}(T) = o(1), \text{ as } \gamma \to \infty.$$

This type of asymptotic solution is called third order and has the important characteristic that, although the quantities $J_{\mathrm{P}}(T_*)$ and $\inf_T J_{\mathrm{P}}(T)$ tend to infinity as $\gamma \to \infty$, their distance tends to zero. We recall for completeness that first-order asymptotic optimality is when $J_{\mathrm{P}}(T_*)/\inf_T J_{\mathrm{P}}(T) \to 1$ and second when $J_{\mathrm{P}}(T_*) - \inf_T J_{\mathrm{P}}(T)$ is bounded uniformly in γ.

A third order asymptotic optimality property was proven by Tartakovsky et al. (2012) for the SR-r test, namely, the analog of T_* in discrete time, with a deterministic and specially designed initialization r_*.

In continuous time, there exist similar optimality claims. Specifically, Polunchenko (2017) shows that T_* can solve (4.2) in the third-order sense, when r_* is random and follows the quasi-stationary distribution. This is the continuous-time analog of Pollak's (1985) result. To obtain the equivalent of Tartakovsky et al.'s (2012) conclusions, one must demonstrate that the T_* in (2.2) can enjoy third-order asymptotic optimality with proper deterministic initialization. Regarding the initializing value r_* of R_t in continuous time, we can be very precise. Since we are under an asymptotic regime with $\gamma \to \infty$, if we refer to (2.10), divide by γ, and let $\gamma \to \infty$, we arrive at the equation

$$1 - e^{r_*^{-1}} E_1(r_*^{-1}) = 0,$$

from which we compute $r_* = 2.299812$. Consequently, the claim is that T_* when initialized with $r_* = 2.299812$ becomes a third-order asymptotic solution of the min–max problem defined in (4.2). Unfortunately, the proof of this statement is still an open problem.

We note that the Pollak metric in (4.1) does not rely on any prior distribution (for τ). It turns out that we can recover this criterion by considering a generic performance measure of the form $\mathsf{E}[T - \tau^+ | T > \tau]$ where the prior for τ is *unknown*. If we follow a worst-case approach over all possible priors, then, as is reported in Moustakides (2008), we recover the Pollak criterion. For this general case as we mentioned above, when R_t is initialized with $r_* = 2.299812$, the conjecture is that T_* is third-order asymptotically optimum.

In our current work, we limit ourselves to prior belonging to the two-parameter zero-modified exponential family. We assume lack of exact knowledge of these parameters and we follow a worst-case analysis *with respect to the two unknowns*. Since we adopt a significantly smaller class of distributions for the change-time (compared to Pollak's metric), our optimality claim can become stronger: We conjecture *exact* optimality for T_* in the sense that it is the exact solution of the min–max constrained optimization problem proposed in (1.5). For T_* to enjoy this optimality property, the initialization parameter r_* must depend on γ through (2.10). Even though we do not provide a complete analytical proof, we do make a thorough mathematical analysis and supply strong numerical evidence supporting the validity of our conjecture.

Appendix A: Proofs of lemmas

In all proofs that follow, we denote the threshold $r_* + \gamma$ with A in order to simplify our mathematical analysis and the corresponding manipulations.

Proof of Lemma 1.1. To show (1.10) we recall that $\{e^{u_t}\}$ is an $\{\mathscr{F}_t\}$-martingale with respect to P_∞, consequently, $\mathsf{E}_\infty[e^{u_t}|\mathscr{F}_s] = e^{u_s}$ when $t \geq s$. This can be extended to stopping times using optional sampling in the sense that $\mathsf{E}_\infty[e^{u_T}|\mathscr{F}_s] = e^{u_s}$ on the event $\{T \geq s\}$. With the help of this observation, we can write

$$\begin{aligned}\mathsf{E}_\infty[R_T] &= \mathsf{E}_\infty\left[e^{u_T}\left\{r_* + \int_0^T e^{-u_s}ds\right\}\right]\\ &= r_*\mathsf{E}_\infty[e^{u_T}] + \int_0^\infty \mathsf{E}_\infty\big[\mathsf{E}_\infty[e^{u_T-u_s}|\mathscr{F}_s]\mathbb{1}_{\{T>t\}}\big]ds\\ &= r_* + \int_0^\infty \mathsf{E}_\infty[\mathbb{1}_{\{T>t\}}]ds = r_* + \mathsf{E}_\infty[T],\end{aligned}$$

which proves the desired expression.

For (1.11), we use Itô calculus and observe that under the P_0 measure we have the following SDE for R_t:

$$dR_t = (2R_t + 1)\,dt + \sqrt{2}R_t dw_t,\ R_0 = r_*, \tag{A.1}$$

while under P_∞ the sde becomes

$$dR_t = dt + \sqrt{2}R_t dw_t,\ R_0 = r_*. \tag{A.2}$$

Consider now $dg(R_t)$ under P_0, that we have

$$dg(R_t) = \{(2R_t + 1)g'(R_t) + R_t^2 g''(R_t)\}dt + \sqrt{2}R_t g'(R_t)dw_t.$$

Integrating and taking expectation with respect to P_t, since we consider R_T for $\{T > t\}$, we are under the post change regime, namely, P_0. This yields

$$\begin{aligned}\mathsf{E}_t[g(R_T) - g(R_t)|\mathscr{F}_t]\mathbb{1}_{\{T>t\}} &= \mathsf{E}_t\left[\int_t^T \{(2R_t + 1)g'(R_t) + R_t^2 g''(R_t)\}dt|\mathscr{F}_t\right]\mathbb{1}_{\{T>t\}}\\ &= -\mathsf{E}_t[T - t|\mathscr{F}_t]\mathbb{1}_{\{T>t\}} = -\mathsf{E}_t[(T-t)^+|\mathscr{F}_t],\end{aligned}$$

where we used the fact that $\{T > t\}$ is $\mathscr{F}_t$-measurable. We note that the second equality is true because, as we can verify, $g(R)$ defined in (1.9) is the solution of the ode $(2R + 1)g'(R) + R^2 g''(R) = -1$.

To prove (1.12), after we note that $\{T > \tau\} = \{T > \tau^+\}$ because $T > 0$, we observe that

$$\mathsf{E}_{r,\lambda}[T - \tau^+ | T > \tau] = \frac{\mathsf{E}_{r,\lambda}[(T - \tau^+)^+]}{\mathsf{P}_{r,\lambda}(T > \tau)}.$$

We consider the numerator and denominator separately. We start with the denominator for which we can write

$$\begin{aligned}\mathsf{P}_{r,\lambda}(T > \tau) &= \mathsf{P}_0(T > 0)\mathsf{P}(\tau \leq 0) + \int_0^\infty \mathsf{E}_{r,\lambda}[\mathbb{1}_{\{T>t\}}\mathbb{1}_{\{\tau\in dt\}}] \\ &= \pi + \int_0^\infty \mathsf{E}_t[\mathbb{1}_{\{T>t\}}]\mathsf{P}(\tau \in dt) = \pi + \int_0^\infty \mathsf{E}_t[\mathbb{1}_{\{T>t\}}](1-\pi)\lambda e^{-\lambda t} dt.\end{aligned}$$

We note that when the time of change is at $\tau = t$, since $\{T > t\}$ is $\mathscr{F}_t$-measurable, it is a pre-change event. But on $\mathscr{F}_t$ the probability measure P_t coincides with the nominal P_∞; therefore, the previous formula can be modified as follows:

$$\begin{aligned}\mathsf{P}_{r,\lambda}(T > \tau) &= \pi + (1-\pi)\int_0^\infty \mathsf{E}_\infty[\mathbb{1}_{\{T>t\}}]\lambda e^{-\lambda t} dt \\ &= \pi + (1-\pi)\mathsf{E}_\infty\left[\int_0^T \lambda e^{-\lambda t} dt\right] \\ &= \lambda\left\{r + (1-\lambda r)\mathsf{E}_\infty\left[\int_0^T e^{-\lambda t} dt\right]\right\}. \qquad \text{(A.3)}\end{aligned}$$

Following a similar line of reasoning for the numerator, we obtain

$$\mathsf{E}_{r,\lambda}[(T - \tau^+)^+] = \pi \mathsf{E}_0[T] + (1-\pi)\int_0^\infty \mathsf{E}_t[(T-t)^+]\lambda e^{-\lambda t} dt.$$

Replacing $\mathsf{E}_t[(T-t)^+]$ from (1.11) yields

$$\begin{aligned}&\mathsf{E}_{r,\lambda}[(T - \tau^+)^+] \\ &= \lambda\left\{r\mathsf{E}_0[g(R_0) - g(R_T)] + (1-\lambda r)\int_0^\infty \mathsf{E}_t\left[\big(g(R_t) - g(R_T)\big)\mathbb{1}_{\{T>t\}}\right] e^{-\lambda t} dt\right\}. \qquad \text{(A.4)}\end{aligned}$$

Taking the ratio of the numerator expression (A.4) and the expression for the denominator in (A.3) and also recalling that $R_0 = r_*$ yields the desired equality.

To prove the last equality of this lemma, we consider the denominator of $\mathscr{D}(T, r_*, 0)$, normalize it by λ and then take the limit as $\lambda \to 0$. As we can then see from (A.3), the denominator becomes $r_* + \mathsf{E}_\infty[T]$. For the numerator, we propose the following alternative way to express $\mathsf{E}_t[(T-t)^+]$ that avoids the use of the function $g(R)$,

$$\begin{aligned}\mathsf{E}_t[(T-t)^+] &= \mathsf{E}_t\left[\int_t^\infty \mathbb{1}_{\{T>t\}}\mathbb{1}_{\{T>s\}} ds\right] \\ &= \int_t^\infty \mathsf{E}_t[\mathbb{1}_{\{T>t\}}\mathbb{1}_{\{T>s\}}] ds \\ &= \int_t^\infty \mathsf{E}_\infty\left[\mathsf{E}_t[\mathbb{1}_{\{T>s\}}|\mathscr{F}_t]\mathbb{1}_{\{T>t\}}\right] ds\end{aligned}$$

$$
\begin{aligned}
&= \int_t^\infty \mathsf{E}_\infty\left[\mathsf{E}_\infty[e^{u_s-u_t}\mathbb{1}_{\{T>s\}}|\mathscr{F}_t]\mathbb{1}_{\{T>t\}}\right]ds\\
&= \int_t^\infty \mathsf{E}_\infty[e^{u_s-u_t}\mathbb{1}_{\{T>t\}}\mathbb{1}_{\{T>s\}}]ds\\
&= \mathsf{E}_\infty\left[\mathbb{1}_{\{T>t\}}\int_t^T e^{u_s-u_t}ds\right],
\end{aligned}
$$

where in the fourth equality we used (1.7) and the fact that $\{T > s\}$ is $\mathscr{F}_s$-measurable. Normalizing the numerator by λ and then letting $\lambda \to 0$ and using the previous expression, we obtain

$$
\begin{aligned}
r_*\mathsf{E}_0[T] &+ \int_0^\infty \mathsf{E}_t[(T-t)^+]e^{-\lambda t}dt\\
&= r_*\mathsf{E}_\infty\left[\int_0^T e^{u_s}ds\right] + \mathsf{E}_\infty\left[\int_0^T\left(\int_t^T e^{u_s}ds\right)e^{-u_t}dt\right]\\
&= r_*\mathsf{E}_\infty\left[\int_0^T e^{u_s}ds\right] + \mathsf{E}_\infty\left[\int_0^T\left(\int_0^s e^{-u_t}dt\right)e^{u_s}ds\right]\\
&= \mathsf{E}_\infty\left[\int_0^T e^{u_s}\left\{r_* + \int_0^s e^{-u_t}dt\right\}ds\right] = \mathsf{E}_\infty\left[\int_0^T R_s ds\right].
\end{aligned}
\tag{A.5}
$$

Dividing the expression for the normalized numerator in (A.5) with the expression for the normalized denominator $r_* + \mathsf{E}_\infty[T]$ yields the desired result. This concludes the proof of the lemma. □

Proof of Lemma 2.3. To prove this lemma, we use methodology similar to the one applied in Lemma 1.1. Consider $f_\lambda(R)$ to be twice differentiable, then under P_∞ we have

$$
d\big(e^{-\lambda t}f_\lambda(R_t)\big) = e^{-\lambda t}\{-\lambda f_\lambda(R_t) + g'(R_t) + R_t^2 g''(R_t)\}dt + \sqrt{2}e^{-\lambda t}R_t g'(R_t)dw_t,
$$

from which we conclude that

$$
\mathsf{E}_\infty[e^{-\lambda T_*}f_\lambda(R_{t_*}) - f_\lambda(R_0)] = \mathsf{E}_\infty\left[\int_0^{T_*} e^{-\lambda t}\{-\lambda f_\lambda(R_t) + g'(R_t) + R_t^2 g''(R_t)\}dt\right].
$$

We select $f_\lambda(R)$ to satisfy the ode

$$
-\lambda f_\lambda(R) + f_\lambda'(R) + R^2 f_\lambda''(R) = -\big(g(R) - g(r_*)\big)
$$

and to be bounded in $[0, A]$ with the boundary condition $f(A) = 0$. If we substitute in the previous equality, after recalling that $R_0 = r_*$ and $R_{T_*} = A$, we prove the desired result. □

Proof of Lemma 2.4. We have that the function $f_0(R)$ satisfies the ode

$$
f_0'(R) + R^2 f_0''(R) = e^{R^{-1}}E_1(R^{-1}) - e^{r_*^{-1}}E_1(r_*^{-1}), \tag{A.6}
$$

and it is bounded for $R \in [0, A]$ with $f_0(A) = 0$. With direct substitution, we can verify that the desired solution has the following form:

$$
\begin{aligned}
f_0(R) &= \{1 - e^{r_*^{-1}}E_1(r_*^{-1})\}(R - A) + \int_{A^{-1}}^{R^{-1}} \frac{e^x}{x}E_1(x)dx\\
&= \{1 - e^{r_*^{-1}}E_1(r_*^{-1})\}(R - A) + \int_{A^{-1}}^{R^{-1}} E_1(x)d\big(E_i(x)\big).
\end{aligned}
$$

We can apply similar ideas in the differential equation (2.6) that defines $f_\lambda(R)$. Multiplying both sides with $e^{-R^{-1}}\frac{1}{R^2}$ yields

$$\left(e^{-R^{-1}}f_\lambda'(R)\right)' = -\lambda e^{-R^{-1}}\frac{1}{R^2}f_\lambda(R) + \left(e^{-R^{-1}}f_0'(R)\right)',$$

where for the last term we used the ode in (A.6) that defines $f_0(R)$. Integrating both sides, we obtain

$$\begin{aligned}
f_\lambda(R) &= f_0(R) - \lambda\int_R^A e^{x^{-1}}\left(\int_0^x \frac{e^{-z^{-1}}}{z^2}f_\lambda(z)dz\right)dx\\
&= f_0(R) - \lambda\int_{A^{-1}}^{R^{-1}}\frac{e^x}{x^2}\left(\int_x^\infty e^{-z}f_\lambda(z^{-1})dz\right)dx\\
&= f_0(R) - \lambda\left\{\int_{A^{-1}}^{R^{-1}} e^{-z}f_\lambda(z^{-1})\left(\int_{A^{-1}}^{z}\frac{e^x}{x^2}dx\right)dz + \left(\int_{A^{-1}}^{R^{-1}}\frac{e^x}{x^2}dx\right)\int_{R^{-1}}^\infty e^{-z}f_\lambda(z^{-1})dz\right\},
\end{aligned}$$

where the second equality is the result of applying the change of variables $x \to x^{-1}$ and $z \to z^{-1}$ and the third is obtained by changing the order of integration in the double integral combined with careful housekeeping of the integration regions. The next step is to observe that the indefinite integral of $\frac{e^x}{x^2}$ is equal to $E_i(x) - \frac{e^x}{x}$. This applied in the previous expression yields

$$\begin{aligned}
f_\lambda(R) &= f_0(R)\\
&\quad - \lambda\left\{\left(E_i(A^{-1}) - e^{A^{-1}}A\right)\int_{A^{-1}}^{R^{-1}} f_\lambda(z^{-1})d(e^{-z}) + \int_{A^{-1}}^{R^{-1}} f_\lambda(z^{-1})\left(e^{-z}E_i(z) - z^{-1}\right)dz\right.\\
&\quad \left. + \left(E_i(A^{-1}) - e^{A^{-1}}A - E_i(R^{-1}) + e^{R^{-1}}R\right)\int_{R^{-1}}^\infty f_\lambda(z^{-1})d(e^{-z})\right\}.
\end{aligned}$$

Combining terms and observing that the indefinite integral of $e^{-z}E_i(z) - z^{-1}$ is $-e^{-z}E_i(z)$ yields

$$\begin{aligned}
f_\lambda(R) = f_0(R) - \lambda\Bigg\{&\left(E_i(A^{-1}) - e^{A^{-1}}A - E_i(R^{-1}) + e^{R^{-1}}R\right)\int_{A^{-1}}^\infty f_\lambda(z^{-1})d(e^{-z})\\
&- \int_{A^{-1}}^{R^{-1}} f_\lambda(z^{-1})d\left(e^{-z}E_i(z)\right) + \left(E_i(R^{-1}) - e^{R^{-1}}R\right)\int_{A^{-1}}^{R^{-1}} f_\lambda(z^{-1})d(e^{-z})\Bigg\},
\end{aligned}$$

which is the final expression. This concludes the proof of Lemma 2.4. □

Acknowledgment

We thank the Editor, the Associate Editor, and the referees for their constructive suggestions and comments.

Funding

This work was supported by the U.S. National Science Foundation under Grant CIF 1513373, through Rutgers University.

References

Abramowitz, M. and Stegun, I. A. (1965). *Handbook of Mathematical Functions*, New York: Dover.
Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*, Cambridge: Cambridge University Press.
Moustakides, G. V. (2008). Sequential Change Detection Revisited, *Annals of Statitics* 36: 787–807.
Moustakides, G. V. (2014). Multiple Optimality Properties of the Shewhart Test, *Sequential Analysis* 33: 318–344.
Moustakides, G. V., Polunchenko, A. S., and Tartakovsky, A. G. (2011). A Numerical Approach to Performance Analysis of Quickest Change-Point Detection Procedures, *Statistica Sinica* 21: 571–596.
Peskir, G. and Shiryaev, A. (2006). *Optimal Stopping and Free-Boundary Problems*, Basel: Birkhäuser.
Pollak, M. (1985). Optimal Detection of a Change in Distribution, *Annals of Statistics* 13: 206–227.
Polunchenko, S. (2017). Asymptotic Near-Minimaxity of the Shiryaev-Roberts-Pollak Change-Point Detection Procedure in Continuous Time. *Theory of Probability and Its Applications* 62: 4.
Polunchenko, S. and Tartakovsky, A. G. (2010). On Optimality of the Shiryaev-Roberts Procedure for Detecting a Change in Distribution, *Annals of Statistics* 38: 3445–3457.
Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes, and Martingales: Volume 2. Itô Calculus*, Cambridge: Cambridge University Press.
Shiryaev, A. N. (1963). On Optimal Methods in Quickest Detection Problems, *Theory of Probability and its Applications* 8: 22–46.
Tartakovsky, A. G., Pollak, M., and Polunchenko A. S. (2012). Third Order Asymptotic Optimality of the Generalized Shiryaev-Roberts Changepoint Detection Procedure, *Theory of Probability and Its Applications* 56: 457–484.