OPTIMALITY OF THE 2-CUSUM DRIFT EQUALIZER RULES FOR DETECTING TWO-SIDED ALTERNATIVES IN THE BROWNIAN MOTION MODEL

OLYMPIA HADJILIADIS,* *Columbia University*

Abstract

This work employs the Brownian motion model in which observations are taken sequentially. The objective is to detect a two-sided change in the constant drift by means of a stopping rule. As a performance measure, an extended Lorden criterion is used. The goal is to minimize the worst-case detection delay subject to a constraint in the frequency of false alarms. In a companion paper, attention is drawn to a first category of 2-CUSUM rules for which the harmonic mean rule holds. It is further seen that a special class of 2-CUSUM stopping rules within this category, called drift equalizer rules, perform strictly better than non-equalizer rules, according to this specific performance measure.

Keywords: Sequential change detection; detection delay; CUSUM; two-sided CUSUM

2000 Mathematics Subject Classification: Primary 62L15
Secondary 60G40; 91A60

1. Introduction

The need for statistical surveillance, the problem of detecting abrupt changes in a stochastic process through sequential observations, has been noted in many different areas. Applications include quality control, onset detection in seismic signal processing [3], target detection in multiple-resolution radar [1], [11], [5], statistical pattern recognition [9], fault detection in navigational systems, centralized detection, epidemiology [24], medicine [8], [18], etc.

We seek a stopping rule T that detects the change point τ while simultaneously controlling the frequency of false alarms. In other words, at each decision time t we want to discriminate between the two states of the process: the state $\{t < \tau\}$ and the state $\{t \geq \tau\}$. More specifically, the stopping rule T balances the trade-off between controlling the mean time between false alarms and minimizing the detection delay of the change.

In the case of one-sided alternatives where the change is a known constant, the traditional Page cumulative sum (CUSUM) [15] was proven to be optimal for any fixed value of the frequency of false alarms [22], [4]. In fact, for the extended Lorden criterion, proposed in [10], it turns out that the one-sided CUSUM rule maintains its optimal character for one-sided alternatives even when the change is known only to be in a given interval of all positive or negative values. It is also worth mentioning that the CUSUM stopping rule has been proven to be optimal for an alternative to Lorden's criterion that uses the Kullback–Leibler divergence whenever the change is a function of time [27] or a measurable function of the observations [14].

The problem of two-sided alternatives is considerably more difficult than that of one-sided alternatives. The first author to suggest the use of cumulative sum charts for two-sided

alternatives was Barnard [2]. Later, the problem of multiple alternatives in the discrete-time exponential family model was examined by Lorden [12]. He proposed the generalized CUSUM rule and proved that it is first-order asymptotically optimal as the frequency of false alarms tends to infinity. Dragalin [6] improved on this result and was able to prove that the generalized CUSUM stopping rule is second-order asymptotically optimal for a specific choice of threshold as the frequency of false alarms tends to infinity. The problem of multiple alternatives was subsequently addressed by Tartakovsky in [26]. His objective was to find a rule that would not only detect the change, but also infer what change it was. For all $i = 1, \ldots, N$, indexing the N possible changes, he considered $\sup_\tau \mathrm{E}_\tau^i[T - \tau \mid T > \tau]$ as a performance measure, and found that the N-CUSUM stopping rule is asymptotically optimal as the frequency of false alarms tends to infinity. The 2-CUSUM stopping rule was proposed as an alternative to the generalized CUSUM rule in [7]. Although the author there only considered one-sided alternatives in the discrete-time exponential family model, he used a min–max-type criterion for the performance measure subject to the usual false alarm constraint. In [10], it was shown that specific 2-CUSUM rules chosen from within the class of equalizer rules have asymptotically the best performance for two-sided alternatives in the specific extended min–max Lorden performance setting that considers the worst detection delay regardless of the change.

For other performance measures, see [21] or [17]. To broaden the scope of stopping rules considered for the traditional change point detection problem, it is worth mentioning the Shiryaev–Roberts rule [21], [20] and the exponentially weighted moving average rule [19]. For a comparison of their performances in the detection of constant drift in the Brownian motion model, we refer the reader to [16] and [25].

2. Problem formulation

We sequentially observe a process $\{\xi_t\}$ with the following dynamics:

$$
\mathrm{d}\xi_t = \begin{cases} \mathrm{d}w_t, & t \leq \tau, \\ \mu_1\mathrm{d}t + \mathrm{d}w_t \quad \text{or} \quad -\mu_2\mathrm{d}t + \mathrm{d}w_t, & t \geq \tau. \end{cases}
$$

Here τ, the time of change, is assumed to be deterministic but unknown, and μ_i, $i = 1, 2$, the possible drifts the process can change to, are assumed to be known, while the specific drift the process is changing to is assumed to be unknown. Both μ_1 and μ_2 are assumed to be positive. Without loss of generality, we can assume that $\mu_2 \geq \mu_1$. Our goal is to detect the change and not to infer which of the two changes occurred.

The probabilistic setting of the problem can be summarized as follows. We consider

- the space of continuous functions $\Omega = C[0, \infty]$;

- the filtration $\{\mathcal{F}_t\}$, with $\mathcal{F}_t = \sigma\{\xi_s, \ 0 < s \leq t\}$ and $\mathcal{F}_\infty = \bigcup_{t>0} \mathcal{F}_t$;

- the following families of probability measures:

  (i) $\{\mathcal{P}_\tau^i\}$, $\tau \in [0, \infty)$, whenever the change is to μ_i, $i = 1, 2$, and

  (ii) $\mathcal{P}_\infty$, the Wiener measure.

In order to incorporate the different possibilities for the μ_i, Hadjiliadis and Moustakides [10] extended Lorden's performance measure by including an extra maximization of the measure over the different detection delays for each of the probability measures generated after the

change. This extension was inspired by consideration of the worst detection delay regardless of the change, and is

$$J_L(T) = \max_i \sup_\tau \operatorname{ess\,sup} \operatorname{E}^i_\tau[(T - \tau)^+ \mid \mathcal{F}_\tau].$$

This gives rise to the following min–max constraint optimization problem:

$$\inf_T J_L(T), \qquad \operatorname{E}_\infty[T] \geq \gamma. \tag{1}$$

Notice that P_∞ is the Wiener measure that corresponds to there being no change ($\tau = \infty$). Therefore, $\operatorname{E}_\infty[T]$ is the mean time to the first false alarm.

As discussed first in [13] and later in [17], in seeking solutions to the above problem we can restrict our attention to stopping times that satisfy the false alarm constraint with equality. This follows since, if $\operatorname{E}_\infty[T] > \gamma$, we can produce a stopping time that achieves the constraint with equality, without increasing the detection delay, simply by randomizing between T and the stopping time that is identically 0. In order to find a solution to (1), we therefore look for stopping rules that are $\mathcal{F}_t$-adapted for all t and that satisfy the false alarm constraint with equality.

In the next section, we turn our attention to the 2-CUSUM stopping rules, since they display asymptotically optimal performance as the frequency of false alarms tends to infinity [10], [26]. From among them we select a special class of 2-CUSUM rules that obey a particular rule, called the harmonic mean rule. We then further restrict ourselves to those harmonic 2-CUSUM rules that are equalizer rules, that is, rules that have the same performance under either the positive or the negative change. In the penultimate section, we prove that (harmonic) 2-CUSUM equalizer rules perform strictly better than all other (harmonic) 2-CUSUM rules. We thus conclude, in the last section, that (harmonic) 2-CUSUM equalizer rules are preferred, and we mention their strong asymptotic properties.

3. The 2-CUSUM rules and the harmonic mean rule

We begin by defining the CUSUM statistics.

Definition 1. The normalized CUSUM statistics with drift parameters $\lambda_1 > 0$ and $\lambda_2 > 0$, tuned to detect the positive and negative changes in the drift of the Brownian motion, are defined as follows:

(i) $y_t^+(\lambda_1)/\lambda_1 = \xi_t - \frac{1}{2}\lambda_1 t - \inf_{s \leq t}(\xi_s - \frac{1}{2}\lambda_1 s)$,

(ii) $y_t^-(\lambda_2)/\lambda_2 = -\xi_t - \frac{1}{2}\lambda_2 t - \inf_{s \leq t}(-\xi_s - \frac{1}{2}\lambda_2 s)$.

We now proceed to define the 2-CUSUM stopping rules.

Definition 2. The 2-CUSUM stopping rule with drift parameters $\lambda_1 > 0$ and $\lambda_2 > 0$ and threshold parameters $\nu_1 > 0$ and $\nu_2 > 0$ is defined as

$$T(\lambda_1, \lambda_2, \nu_1, \nu_2) = T^1 \wedge T^2,$$

where

(i) $T^1 = \inf\{t > 0 \colon y_t^+(\lambda_1)/\lambda_1 > \nu_1\}$,

(ii) $T^2 = \inf\{t > 0 \colon y_t^-(\lambda_2)/\lambda_2 > \nu_2\}$.

We now consider a smaller, very interesting class of 2-CUSUM rule: those that obey the harmonic mean rule. In particular, we will consider all 2-CUSUM rules whose two CUSUM stopping time branches, T^1 and T^2, have the same threshold. The harmonic mean rule enables us to explicitly compute the expected value of the 2-CUSUM stopping rule in terms of the expected values of its corresponding one-sided CUSUM stopping times. To this end, we introduce the following class of stopping rule.

Definition 3. Define $\mathcal{G} = \{T(\lambda_1, \lambda_2, \nu_1, \nu_2) \colon \nu_1 = \nu_2\}$.

Henceforth, we only consider 2-CUSUM rules in $\mathcal{G}$, and denote them by $T(\lambda_1, \lambda_2, \nu)$. Using the proof in [10] or [23, p. 28] it is possible to show that, under any of the measures P_0^1, P_0^2, or P_∞, we have

$$\frac{1}{\mathrm{E}[T(\lambda_1, \lambda_2, \nu)]} = \frac{1}{\mathrm{E}[T^1]} + \frac{1}{\mathrm{E}[T^2]}. \tag{2}$$

At this point, it is worth noting that, for any CUSUM stopping rule T, the worst detection delay occurs when $y_\tau^+ = 0$ and $y_\tau^- = 0$. This is a simple consequence of the nonnegativity of the CUSUM statistic processes. Hence,

$$J_L(T) = \max_i \sup_\tau \operatorname*{ess\,sup} \mathrm{E}_\tau^i[(T - \tau)^+ \mid \mathcal{F}_\tau]$$

$$= \max\{\mathrm{E}_0^1[T], \mathrm{E}_0^2[T]\}.$$

As shown in [10], by applying Itô's rule and using existing results in stochastic analysis we obtain

$$\frac{1}{2}\mathrm{E}_\infty[T^1] = \frac{h(\lambda_1 \nu)}{\lambda_1^2}, \tag{3}$$

$$\frac{1}{2}\mathrm{E}_\infty[T^2] = \frac{h(\lambda_2 \nu)}{\lambda_2^2}, \tag{4}$$

$$\frac{1}{2}\mathrm{E}_0^1[T^1] = \frac{h((\lambda_1 - 2\mu_1)\nu)}{(\lambda_1 - 2\mu_1)^2}, \tag{5}$$

$$\frac{1}{2}\mathrm{E}_0^1[T^2] = \frac{h((\lambda_2 + 2\mu_1)\nu)}{(\lambda_2 + 2\mu_1)^2}, \tag{6}$$

$$\frac{1}{2}\mathrm{E}_0^2[T^1] = \frac{h((\lambda_1 + 2\mu_2)\nu)}{(\lambda_1 + 2\mu_2)^2}, \tag{7}$$

$$\frac{1}{2}\mathrm{E}_0^2[T^2] = \frac{h((\lambda_2 - 2\mu_2)\nu)}{(\lambda_2 - 2\mu_2)^2}, \tag{8}$$

where $h(x) = \mathrm{e}^x - x - 1$.

4. Equalizer rules are best

We now proceed to inspect the dynamics of the CUSUM statistic processes when the change is respectively μ_1 and $-\mu_2$. See Tables 1 and 2.

We notice that if

$$\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1 \tag{9}$$

holds, then y_t^+/λ_1, when the change is μ_1, has the same law as y_t^-/λ_2, when the change is $-\mu_2$, and that y_t^+/λ_1, when the change is $-\mu_2$, has the same law as y_t^-/λ_2, when the change is

TABLE 1: The dynamics of the two CUSUM statistics when the change is μ_1.

Statistic	Dynamics
y_t^+/λ_1	$w_t + (\mu_1 - \frac{1}{2}\lambda_1)t - \inf_{s\leq t}(w_s + (\mu_1 - \frac{1}{2}\lambda_1)s)$
y_t^-/λ_2	$-w_t - (\mu_1 + \frac{1}{2}\lambda_2)t - \inf_{s\leq t}(-w_s - (\mu_1 + \frac{1}{2}\lambda_2)s)$

TABLE 2: The dynamics of the two CUSUM statistics when the change is $-\mu_2$.

Statistic	Dynamics
y_t^+/λ_1	$w_t - (\mu_2 + \frac{1}{2}\lambda_1)t - \inf_{s\leq t}(w_s - (\mu_2 + \frac{1}{2}\lambda_1)s)$
y_t^-/λ_2	$-w_t + (\mu_2 - \frac{1}{2}\lambda_2)t - \inf_{s\leq t}(-w_s + (\mu_2 - \frac{1}{2}\lambda_2)s)$

μ_1. In particular, this means that $T(\lambda_1, \lambda_2, \nu) = T^1 \wedge T^2$ has the same distribution under the measures P_0^1 and P_0^2. Therefore, when (9) holds, $\mathrm{E}_0^1[T(\lambda_1, \lambda_2, \nu)] = \mathrm{E}_0^2[T(\lambda_1, \lambda_2, \nu)]$. This allows us to distinguish from among all (harmonic) 2-CUSUM rules the equalizer rules whose performances are the same under the measures P_0^1 and P_0^2.

Definition 4. We define the class of all equalizer rules as follows:

$$\mathcal{E} = \{T(\lambda_1, \lambda_2, \nu)\colon \lambda_2 - \lambda_1 = 2(\mu_2 - \mu_1),\ \nu > 0\}.$$

In the sequel, we will use S to denote any stopping rule that belongs to the class $\mathcal{E}$ and T to denote any stopping rule that does not. Notice that if $\mu_2 = \mu_1$ (the symmetric case), any choice of $\lambda \in \mathbb{R}_+$ will result in an equalizer rule for $\lambda_2 = \lambda_1 = \lambda$.

Our focus is thus on the case $\mu_2 > \mu_1$. The objective is, for any arbitrary rule T, to find an equalizer rule S that achieves the same frequency of false alarms while lowering the detection delay. In other words, for any arbitrary rule T we want always to be able to find a rule $S \in \mathcal{E}$ that has better performance.

To this end, let us define two classes of non-equalizer rule.

Definition 5. We define the following two classes of non-equalizer rule:

(i) $\mathcal{D}_g = \{T(\lambda_1, \lambda_2, \nu)\colon \lambda_2 - \lambda_1 > 2\mu_2 - 2\mu_1\}$,

(ii) $\mathcal{D}_s = \{T(\lambda_1, \lambda_2, \nu)\colon \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1\}$.

Note that $\mathcal{E}^c = \mathcal{D}_g \cup \mathcal{D}_s$.

Theorem 1. *For all 2-CUSUM rules $T \in \mathcal{D}_g \cup \mathcal{D}_s$, there exists an $S \in \mathcal{E}$ such that*

$$\mathrm{E}_\infty[T] = \mathrm{E}_\infty[S] \tag{10}$$

and

$$\max\{\mathrm{E}_0^1[T], \mathrm{E}_0^2[T]\} > \mathrm{E}_0^1[S] = \mathrm{E}_0^2[S].$$

Proof. We can distinguish the following three cases.

(i) $T(\lambda_1', \lambda_2', \nu) \in \mathcal{D}_g$. There exist $\lambda_2, \lambda_1 > 0$, such that $\lambda_2 < \lambda_2'$, $\lambda_1 > \lambda_1'$, and $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$, for which (10) holds.

(ii) $T(\lambda_1', \lambda_2', \nu) \in \mathcal{D}_s$ and $\lambda_2' > \lambda_1'$. (The justification for the additional assumption, $\lambda_2' > \lambda_1'$, is given in Appendix C and should be read after the sequel). There exist $\lambda_2, \lambda_1 > 0$, such that $\lambda_2 > \lambda_2'$, $\lambda_1' > \lambda_1$, and $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$, for which the same frequency of false alarms can be achieved by an equalizer rule for the same threshold ν. More specifically, there exists an $S(\lambda_1, \lambda_2, \nu)$ for which (10) holds.

(iii) There exists no $\lambda_1 > 0$ such that, with $\lambda_2 > \lambda_2'$ and $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$, the same frequency of false alarms can be achieved by an equalizer rule with the same threshold.

To prove the result in the first case, it suffices to prove the following two inequalities:

$$\frac{1}{\mathrm{E}_0^2[S^1]} - \frac{1}{\mathrm{E}_0^2[T^1]} > \frac{1}{\mathrm{E}_\infty[S^1]} - \frac{1}{\mathrm{E}_\infty[T^1]}, \tag{11}$$

$$\frac{1}{\mathrm{E}_\infty[T^2]} - \frac{1}{\mathrm{E}_\infty[S^2]} > \frac{1}{\mathrm{E}_0^2[T^2]} - \frac{1}{\mathrm{E}_0^2[S^2]}. \tag{12}$$

This is because the right-hand side of (11) is equal to the left-hand side of (12), as can be seen by using (2) and (10). It then follows that the left-hand side of (11) is greater than the right-hand side of (12), and from (2) we obtain $\mathrm{E}_0^2[T] > \mathrm{E}_0^2[S]$. By using (7), (8), (3), and (4,) we can rewrite (11) in the following way:

$$\left[\frac{h((\lambda_1 + 2\mu_2)\nu)}{(\lambda_1 + 2\mu_2)^2} \right]^{-1} - \left[\frac{h((\lambda_1' + 2\mu_2)\nu)}{(\lambda_1' + 2\mu_2)^2} \right]^{-1} > \left[\frac{h(\lambda_1\nu)}{\lambda_1^2} \right]^{-1} - \left[\frac{h(\lambda_1\nu)}{\lambda_1'^2} \right]^{-1}.$$

The result follows by multiplying both sides of the equation by ν^2 and using the convexity of the function $g(x) = x^2/h(x)$ (see Appendix A). We can prove (12) similarly.

In cases (ii) and (iii) the result follows from the inequalities

$$\frac{1}{\mathrm{E}_0^1[S^2]} - \frac{1}{\mathrm{E}_0^1[T^2]} > \frac{1}{\mathrm{E}_\infty[S^2]} - \frac{1}{\mathrm{E}_\infty[T^2]}, \tag{13}$$

$$\frac{1}{\mathrm{E}_\infty[T^1]} - \frac{1}{\mathrm{E}_\infty[S^1]} > \frac{1}{\mathrm{E}_0^1[T^1]} - \frac{1}{\mathrm{E}_0^1[S^1]}. \tag{14}$$

Notice that in cases (ii) and (iii) we have $\mathrm{E}_0^1[T] > \mathrm{E}_0^1[S]$. In case (ii), after using (5), (6), (3), and (4), the two inequalities are a direct consequence of the convexity of the function $g(x) = x^2/h(x)$. In case (iii) the situation is slightly more involved, since, in order to achieve the same frequency of false alarms for an equalizer rule, we need to lower the threshold to $\nu' < \nu$. In other words, we can still find an $S(\lambda_1, \lambda_2, \nu') \in \mathcal{E}$ for which (10) holds, by taking $\lambda_1 = \lambda_1'$ and $\lambda_2 > \lambda_2'$ such that $\lambda_2 - \lambda_1 = 2\mu_2 - 2\mu_1$; consequently, we can find a threshold $\nu' < \nu$. We can now rewrite (13) and (14), with the above choice of parameters, as follows, using (5), (6), (3), and (4):

$$\left[\frac{h((\lambda_2 + 2\mu_1)\nu')}{(\lambda_2 + 2\mu_1)^2} \right]^{-1} - \left[\frac{h((\lambda_2' + 2\mu_1)\nu)}{(\lambda_2' + 2\mu_1)^2} \right]^{-1} > \left[\frac{h(\lambda_2\nu')}{\lambda_2^2} \right]^{-1} - \left[\frac{h(\lambda_2'\nu)}{\lambda_2'^2} \right]^{-1}, \tag{15}$$

$$\left[\frac{h(\lambda_1\nu)}{\lambda_1^2} \right]^{-1} - \left[\frac{h(\lambda_1\nu')}{\lambda_1^2} \right]^{-1} > \left[\frac{h((\lambda_1 - 2\mu_1)\nu)}{(\lambda_1 - 2\mu_1)^2} \right]^{-1} - \left[\frac{h((\lambda_1 - 2\mu_1)\nu')}{(\lambda_1 - 2\mu_1)^2} \right]^{-1}. \tag{16}$$

For a proof of (16), we refer the reader to Appendix D. Notice that the parameters are chosen in such a way that (10) holds; therefore, the right-hand side of (15) is equal to the left-hand side of (16) and they are both negative. Thus, we have

$$\frac{h(\lambda_2' v)}{\lambda_2'^2} < \frac{h(\lambda_2 v')}{\lambda_2^2}.$$

If

$$\frac{h((\lambda_2' + 2\mu_1)v)}{(\lambda_2' + 2\mu_1)^2} \geq \frac{h((\lambda_2 + 2\mu_1)v')}{(\lambda_2 + 2\mu_1)^2}$$

then (15) holds trivially. We will now proceed to examine the opposite case. We make two selections, $\lambda_2'' \in [\lambda_2', \lambda_2]$ and $\lambda_2^x \in [\lambda_2', \lambda_2]$, such that

$$\frac{h(\lambda_2 v')}{\lambda_2^2} = \frac{h(\lambda_2'' v)}{\lambda_2''^2}, \tag{17}$$

$$\frac{h((\lambda_2 + 2\mu_1)v')}{(\lambda_2 + 2\mu_1)^2} = \frac{h((\lambda_2^x + 2\mu_1)v)}{(\lambda_2^x + 2\mu_1)^2}. \tag{18}$$

From Appendix B it follows that $\lambda_2^x < \lambda_2''$ and, since the function $h(x)/x^2$ is strictly increasing for all $x \in \mathcal{R}_+$, we have

$$\left[\frac{h((\lambda_2^x + 2\mu_1)v)}{(\lambda_2^x + 2\mu_1)^2}\right]^{-1} - \left[\frac{h((\lambda_2' + 2\mu_1)v)}{(\lambda_2' + 2\mu_1)^2}\right]^{-1} > \left[\frac{h((\lambda_2'' + 2\mu_1)v)}{(\lambda_2'' + 2\mu_1)^2}\right]^{-1} - \left[\frac{h((\lambda_2' + 2\mu_1)v)}{(\lambda_2' + 2\mu_1)^2}\right]^{-1}. \tag{19}$$

Inequality (15) now readily follows from (19), (17), (18), the convexity of the function $g(x) = x^2/h(x)$ (see Appendix A), and the fact that $\lambda_2'' > \lambda_2'$. This completes the proof.

5. Conclusions

The result presented in this paper allows us to select the drifts λ_1 and λ_2 in such a way that we can construct 2-CUSUM rules that perform strictly better for all frequencies of false alarm, especially in the case that the absolute values of the possible two-sided drifts assumed after the change are not equal. It is interesting to mention that, in this case, the equalizer rule with the best asymptotic performance (as the frequency of false alarms tends to infinity) is the one with the choice of parameters $\lambda_1 = \mu_1$, implying that $\lambda_2 = 2\mu_2 - \mu_1$ (see [10]). It is worth mentioning that the difference between the detection delay of the optimal unknown scheme and the 2-CUSUM stopping rule with this choice of parameters (v selected so that the false alarm constraint is satisfied with equality) in fact tends to 0, even though both of the detection delay quantities are unbounded as the frequency of false alarms tends to infinity. This difference tends to 0 faster as the difference between μ_2 and μ_1 increases. Moreover, in the symmetric case, the choice of 2-CUSUM equalizer rule with drift parameter equal to the absolute value of the change and threshold parameter chosen so as to satisfy the false alarm constraint with equality also displays very good properties. In particular, the difference in the detection delay of the optimum scheme and the specific 2-CUSUM rule tends to the constant $2 \log 2/\mu^2$ as the frequency of false alarms tends to infinity. Here, μ is the absolute value of the two-sided possible changes. Notice that, in this case also, both detection delays become unbounded as the frequency of false alarms tends to infinity. For more details, we refer the reader to [10].

Appendix A.

Lemma 1. *The function $g(x) = [h(x)/x^2]^{-1}$, where $h(x) = e^x - x - 1$, is strictly convex.*

Proof. It suffices to show that

$$g''(x) = \frac{[2(e^x - x - 1) - x^2 e^x](e^x - x - 1) - 2(e^x - 1)[2x(e^x - x - 1)]}{(e^x - x - 1)^3} > 0.$$

To do so, we will show that the function $f(x) = (e^x - x - 1)^3 g''(x)$ is positive for all $x \neq 0$, by showing that $f'(x)$ has the same sign as x (note that $f(0) = 0$). We have

$$f'(x) = xe^x[2xe^x - 6e^x + x^2 + 4x + 6].$$

Let $k(x) = 2xe^x - 6e^x + x^2 + 4x + 6$ and note that $k(0) = 0$. Furthermore, $k'(x) = 2xe^x - 4e^x + 2x + 4$ and $k'(0) = 0$. Hence, $k''(x) = 2e^x[e^{-x} + x - 1] > 0$ for all $x \neq 0$, with $k''(0) = 0$. Therefore, $k(x) > 0$ for all $x \neq 0$, with $k(0) = 0$, and $f'(x)$ has the same sign as x, with $f(0) = 0$. From this, it follows that $f(x) \geq 0$ with equality only at $x = 0$. This completes the proof.

Appendix B.

Lemma 2. *Suppose that we choose x_1, x_2, v', and v to all be positive, with $v > v'$, $x_1 < x_2$, and*

$$\frac{h(x_1 v)}{x_1^2} = \frac{h(x_2 v')}{x_2^2}. \tag{20}$$

Then, for all $a \in \mathcal{R}_+$, we have

$$\frac{h((x_1 + a)v)}{(x_1 + a)^2} > \frac{h((x_2 + a)v')}{(x_2 + a)^2},$$

where $h(x) = e^x - x - 1$.

Proof. Notice that, since $x_1 < x_2$, for (20) to hold we must have $x_1 v < x_2 v'$, implying that

$$\frac{1}{x_1 v} + \frac{1}{(x_1 v)^2} > \frac{1}{x_2 v'} + \frac{1}{(x_2 v')^2}.$$

Using this and the fact that $v > v'$, we find that

$$\left[\frac{x_1 v + 1}{(x_1 v)^2}\right] v^2 - \left[\frac{x_2 v' + 1}{(x_2 v')^2}\right] v'^2 > 0. \tag{21}$$

From (20) and (21), it follows that

$$\frac{e^{x_1 v}}{x_1^2} > \frac{e^{x_2 v'}}{x_2^2} \Leftrightarrow \frac{x_2^2}{x_1^2} > \frac{e^{x_2 v'}}{e^{x_1 v}}. \tag{22}$$

We now have

$$\begin{aligned}
\frac{e^{x_2 v'}}{e^{x_1 v}} &> \frac{e^{x_2 v'} - x_1 v e^{-av}}{e^{x_1 v} - x_1 v e^{-av}} \\
&> \frac{e^{x_2 v'} - x_2 v' e^{-av'}}{e^{x_1 v} - x_1 v e^{-av}} \\
&> \frac{e^{x_2 v'} - x_2 v' e^{-av'} - av' e^{-av'} - e^{-av'}}{e^{x_1 v} - x_1 v e^{-av} - a v e^{-av} - e^{-av}},
\end{aligned} \tag{23}$$

where the first inequality follows from the fact that $x_1 v e^{-av} > 0$, the second inequality from the facts that $x_1 v < x_2 v'$ and $e^{-av'} > e^{-av}$, and the last inequality by noting that the function $(x + 1)e^{-x}$ is decreasing for all $x > 0$. Using (22) and (23), and the facts that $a > 0$ and $e^{av'}/e^{av} < 1$, we find that

$$\frac{(x_2 + a)^2}{(x_1 + a)^2} > \frac{x_2^2}{x_1^2} > \frac{e^{x_2 v'} - (x_2 + a)v' e^{-av'} - e^{-av'}}{e^{x_1 v} - (x_1 + a)v e^{-av} - e^{-av}}$$

$$> \frac{e^{av'}}{e^{av}} \frac{e^{x_2 v'} - (x_2 + a)v' e^{-av'} - e^{-av'}}{e^{x_1 v} - (x_1 + a)v e^{-av} - e^{-av}}. \tag{24}$$

The result follows from the final inequality and a rearrangement of terms.

Appendix C.

To complete the proof of Theorem 1, it remains to justify the assertion that, whenever $E_0^1[T] > E_0^1[S]$ (cases (ii) and (iii)), it is sufficient to consider 2-CUSUM rules T for which the second drift parameter is greater than the first. To this end, let us define the following two classes of stopping rule.

Definition 6. Define

  (i) $\mathcal{C}_1 = \{T(\lambda_1, \lambda_2, v) : \lambda_2 > \lambda_1 > 0, \ \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1, \ v > 0\}$,

  (ii) $\mathcal{C}_2 = \{T(\lambda_1, \lambda_2, v) : 0 < \lambda_2 < \lambda_1, \ \lambda_2 - \lambda_1 < 2\mu_2 - 2\mu_1, \ v > 0\}$.

Note that $\mathcal{C}_1 \cup \mathcal{C}_2 = \mathcal{D}_s$. The following lemma is sufficient to justify our adherence to rules that belong to the class $\mathcal{C}_1$ whenever $E_0^1[T] > E_0^1[S]$.

Lemma 3. *For all $T_{c_2} \in \mathcal{C}_2$, there exists a $T_{c_1} \in \mathcal{C}_1$ such that $E_\infty[T_{c_1}] = E_\infty[T_{c_2}]$ and $E_0^1[T_{c_2}] > E_0^1[T_{c_1}]$.*

Proof. Let $\lambda_1' > \lambda_2' > 0$. Then $T(\lambda_1', \lambda_2', v) \in \mathcal{C}_2$. From (3), (4), and (2), we find that

$$\frac{1}{E_\infty[T(\lambda_1', \lambda_2', v)]} = \left[\frac{h((\lambda_1')v)}{(\lambda_1')^2}\right]^{-1} + \left[\frac{h((\lambda_2')v)}{(\lambda_2')^2}\right]^{-1}. \tag{25}$$

Now let $\lambda_1'' = \lambda_2'$ and $\lambda_2'' = \lambda_1'$. Then $T(\lambda_1'', \lambda_2'', v) \in \mathcal{C}_1$ and

$$\frac{1}{E_\infty[T(\lambda_1'', \lambda_2'', v)]} = \left[\frac{h((\lambda_1'')v)}{(\lambda_1'')^2}\right]^{-1} + \left[\frac{h((\lambda_2'')v)}{(\lambda_2'')^2}\right]^{-1}$$

$$= \left[\frac{h((\lambda_1')v)}{(\lambda_1')^2}\right]^{-1} + \left[\frac{h((\lambda_2')v)}{(\lambda_2')^2}\right]^{-1}. \tag{26}$$

Therefore, the above rules have the same frequency of false alarms.

The desired result comes as a direct consequence of the following two inequalities:

$$\left[\frac{h((\lambda_1' - 2\mu_1)v)}{(\lambda_1' - 2\mu_1)^2}\right]^{-1} - \left[\frac{h((\lambda_1'' - 2\mu_1)v)}{(\lambda_1'' - 2\mu_1)^2}\right]^{-1} < \left[\frac{h((\lambda_1')v)}{(\lambda_1')^2}\right]^{-1} - \left[\frac{h((\lambda_1'')v)}{(\lambda_1'')^2}\right]^{-1}, \tag{27}$$

$$\left[\frac{h((\lambda_2'')v)}{(\lambda_2'')^2}\right]^{-1} - \left[\frac{h((\lambda_2')v)}{(\lambda_2')^2}\right]^{-1} < \left[\frac{h((\lambda_2'' + 2\mu_1)v)}{(\lambda_2'' + 2\mu_1)^2}\right]^{-1} - \left[\frac{h((\lambda_2' + 2\mu_1)v)}{(\lambda_2' + 2\mu_1)^2}\right]^{-1}. \tag{28}$$

Notice that, from (26) and (25), it follows that the right-hand side of (27) is equal to the left-hand side of (28). Therefore, the left-hand side of (27) is greater than the right-hand side of (28). From this result, a rearrangement of terms, and (6), (5), (3), (4), and (2), we obtain

$$\frac{1}{E_0^1[T(\lambda_1'', \lambda_2'', \nu)]} > \frac{1}{E_0^1[T(\lambda_1', \lambda_2', \nu)]},$$

which is the desired inequality.

Inequalities (27) and (28) follow by multiplying both sides of each of them by ν^2 and using the convexity of the function $x^2/h(x)$ (see Appendix A) along with the facts that $\lambda_1' > \lambda_1''$ and $\lambda_2'' > \lambda_2'$, respectively. This completes the proof of the lemma.

Appendix D.

Definition 7. Let $\mu \in [0, \infty)$ and $\nu > \nu' > 0$. Define the following functions, where $h(x) = e^x - x - 1$:

(i) $C_\nu(\mu) = \dfrac{h((\lambda_1 - 2\mu)\nu)}{(\lambda_1 - 2\mu)^2},$

(ii) $C_{\nu'}(\mu) = \dfrac{h((\lambda_1 - 2\mu)\nu')}{(\lambda_1 - 2\mu)^2},$

(iii) $f(\mu) = \dfrac{1}{C_\nu(\mu)} - \dfrac{1}{C_{\nu'}(\mu)}.$

After introducing the above definition we can rewrite (16) as $f(0) > f(\mu_1)$. To prove this, it suffices to show that $f(\mu)$ is strictly decreasing.

Lemma 4. *The function $f(\mu)$ is strictly decreasing.*

Proof. By differentiating $f(\mu)$ with respect to μ, we obtain

$$f'(\mu) = -\frac{C_\nu'(\mu)}{[C_\nu(\mu)]^2} + \frac{C_{\nu'}'(\mu)}{[C_{\nu'}(\mu)]^2}.$$

Hence, it suffices to show that

$$-C_\nu'(\mu)[C_{\nu'}(\mu)]^2 < -C_{\nu'}'(\mu)[C_\nu(\mu)]^2.$$

Using $\nu > \nu'$ and doing a term-by-term comparison gives the result.

Acknowledgements

References

[1] BAKUT, P. *et al.* (1963). *Statistical Radar Theory*, Vol. 1, 1st edn. Soviet Radio, Moscow (in Russian).
[2] BARNARD, G. (1959). Control charts and stochastic processes. *J. R. Statist. Soc. B* **11,** 239–271.
[3] BASSEVILLE, M. AND NIKIFOROV, I. (1993). *Detection of Abrupt Changes: Theory and Applications.* Prentice Hall, Englewood Cliffs, NJ.
[4] BEIBEL, M. (1996). A note on Ritov's Bayes approach to the minimax property of the cusum procedure. *Ann. Statist.* **24,** 1804–1812.

[5] BUSSGANG, J. (1970). Sequential methods in radar detection. *Proc. IEEE* **58,** 731–743.

[6] DRAGALIN, V. (1994). Optimality of a generalized CUSUM procedure in quickest detection problem. In *Statistics and Control of Random Processes* (Proc. Steklov Inst. Math. **202**), American Mathematical Society, Providence, RI, pp. 107–120.

[7] DRAGALIN, V. (1997). The design and analysis of 2-CUSUM procedure. *Commun. Statist. Simul. Comput.* **26,** 67–81.

[8] FRISEN, M. (1992). Evaluations of methods for statistical surveillance. *Statist. Med.* **11,** 1489–1502.

[9] FU, K. (1968). *Sequential Methods in Pattern Recognition and Learning*, 1st edn. Academic Press, New York.

[10] HADJILIADIS, O. AND MOUSTAKIDES, G. (2005). Optimal and asymptotically optimal CUSUM rules for change point detection in the Brownian motion model with multiple alternatives. *Teor. Veroyat. Primen.* **50,** 131–144. English translation: to appear in *Theory Prob. Appl.* **50**.

[11] LONG, M. (1992). *Airborne Early Warning System Concepts*, 1st edn. Artech, Boston, MA.

[12] LORDEN, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42,** 1897–1908.

[13] MOUSTAKIDES, G. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14,** 1379–1387.

[14] MOUSTAKIDES, G. (2004). Optimality of the CUSUM procedure in continuous time. *Ann. Statist.* **32,** 302–315.

[15] PAGE, E. (1954). Continuous inspection schemes. *Biometrika* **41,** 100–115.

[16] POLLAK, M. AND SIEGMUND, D. (1985). A diffusion process and its applications to detecting a change in the drift of Brownian motion. *Biometrika* **72,** 267–280.

[17] POOR, V. (1998). Quickest detection with exponential penalty for delay. *Ann. Statist.* **26,** 2179–2205.

[18] RADAELLI, G. (1992). Using the cuscore technique in the surveillance of rare health events. *J. Appl. Statist.* **19,** 75–81.

[19] ROBERTS, S. (1959). Control chart tests based on geometric moving averages. *Technometrics* **1,** 239–250.

[20] ROBERTS, S. (1966). A comparison of some control chart procedures. *Technometrics* **8,** 411–430.

[21] SHIRYAEV, A. N. (1963). On optimum methods in quickest detection problems. *Theory Prob. Appl.* **13,** 22–46.

[22] SHIRYAEV, A. N. (1996). Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russian Math. Surveys* **51,** 750–751.

[23] SIEGMUND, D. (1985). *Sequential Analysis*, 1st edn. Springer, New York.

[24] SONESSON, C. (2003). Evaluations of some exponentially weighted moving average methods. *J. Appl. Statist.* **30,** 1115–1133.

[25] SRIVASTAVA, M. AND WU, Y. (1993). Comparison of EWMA, CUSUM and Shiryayev-Roberts procedures for detecting a shift in the mean. *Ann. Statist.* **21,** 645–670.

[26] TARTAKOVSKY, A. (1994). Asymptotically minimax multi-alternative sequential rule for disorder detection. In *Statistics and Control of Random Processes* (Proc. Steklov Inst. Math. **202**), American Mathematical Society, Providence, RI, pp. 229–236.

[27] TARTAKOVSKY, A. (1995). Asymptotic properties of CUSUM and Shiryaev's procedures for detecting a change in non-homogeneous Gaussian processes. *Math. Meth. Statist.* **4,** 389–404.