ТЕОРИЯ ВЕРОЯТНОСТЕЙ

И ЕЕ ПРИМЕНЕНИЯ

Журнал имени А.Н.Колмогорова

ТОМ 50

(отдельный оттиск)

© 2005 г. *HADJILIADIS O.*, MOUSTAKIDES V.***

OPTIMAL AND ASYMPTOTICALLY OPTIMAL CUSUM RULES FOR CHANGE POINT DETECTION IN THE BROWNIAN MOTION MODEL WITH MULTIPLE ALTERNATIVES[1]

В статье изучается задача последовательного обнаружения изменения равного константе сноса броуновского движения в случае многих альтернатив. Как мера качества предлагается некоторое обобщение критерия Лордена. В случае, когда коэффициенты сноса, возможные после разладки, имеют одинаковый знак, доказано, что метод кумулятивных сумм (CUSUM) является оптимальным при обнаружении наименьшего по абсолютной величине сноса. В случае, когда коэффициенты сноса имеют разные знаки, предъявляется специальное 2-CUSUM правило, которое является асимптотически оптимальным при стремлении частоты ложных тревог к бесконечности.

Ключевые слова и фразы: обнаружение моментов изменения, скорейшее обнаружение, метод кумулятивных сумм (CUSUM), двусторонний метод кумулятивных сумм (2-CUSUM).

1. Introduction and mathematical formulation of the problem.
We begin by considering the observation process $\{\xi_t\}_{t>0}$ with the following dynamics:

$$d\xi_t = \begin{cases} dw_t, & t \leqslant \theta, \\ \mu_i \, dt + dw_t, & t > \theta, \end{cases} \quad i = 1, 2,$$

where θ, the time of change, is assumed deterministic but unknown; μ_i, the possible drifts the process can change to, are assumed known, but the specific drift the process is changing to is assumed to be unknown. Our goal is to detect the change and *not to infer which of the changes occurred.*

The probabilistic setting of the problem can be summarized as follows.
— The space of continuous functions $\Omega = C[0, \infty]$.
— The filtration $\{\mathscr{F}_t\}$ with $\mathscr{F}_t = \sigma\{\xi_s, \ 0 < s \leqslant t\}$ and $\mathscr{F}_\infty = \bigcup_{t>0} \mathscr{F}_t$.
— The families of probability measures:
1) $\{\mathbf{P}_\theta^i\}$, $\theta \in [0, \infty)$, whenever the change is μ_i, $i = 1, 2$;

* Department of Statistics, Columbia University, Rm 1005, MC 4609, 1255 Amsterdam Ave, New York, NY 10027, USA; e-mail: oh66@columbia.edu
** Institut National de Recherche en Informatique et en Automatique (INRIA), Rennes, France and Department of Computer and Communication Engineering, University of Thessaly, Volos, Greece.

2) $\mathbf{P}_\infty$, the Wiener measure.

The objective is to detect the change as soon as possible while at the same time controlling the frequency of false alarms. This is achieved through the means of a stopping rule τ adapted to the filtration $\mathscr{F}_t$. One of the possible performance measures of the detection delay, suggested by Lorden in [6], considers the worst detection delay over all paths before the change and all possible change points θ. It is

$$J(\tau) = \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta[(\tau - \theta)^+ \mid \mathscr{F}_\theta], \qquad (1)$$

giving rise to the following constrained stochastic optimization problem:

$$\inf_\tau J(\tau), \qquad \mathbf{E}_\infty[\tau] \geqslant T. \qquad (2)$$

Other performance measures include the Stationary Average Delay Time (SADT), first advocated by Shiryaev in [13] and the Conditional Average Delay Time (CADT): $\sup_\theta \mathbf{E}_\theta[(\tau - \theta) \mid \tau > \theta]$. The former is used in the comparison between Roberts' EWMA rule (see [11]) with Page's CUSUM rule (see [9]) and the Shiryaev–Roberts rule (see [13] and [12]) appearing in the paper by Srivastava and Wu [16] for the one-sided alternative in the Brownian motion model. The latter is used in [10], where the Shiryaev–Roberts rule is compared with the CUSUM rule for the same problem. In the multiple and two-sided alternative case, Tartakovsky in [17] proves the asymptotic optimality of the N-CUSUM rule as the frequency of false alarms tends to infinity by considering the CADT for all changes as a performance measure in the exponential family model. Lorden in [6] proves the first-order asymptotic optimality of the generalized CUSUM rule for two-sided alternatives in the exponential family model. This result was further improved by Dragalin in [3].

In order to incorporate the different possibilities for the μ_i we extend Lorden's performance measure inspired by the idea of the worst detection delay regardless of the change (along the lines of [4]). It is

$$J_L(\tau) = \max_i \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta^i[(\tau - \theta)^+ \mid \mathscr{F}_\theta], \qquad (3)$$

which results in a corresponding optimization problem of the form:

$$\inf_\tau J_L(\tau), \qquad \mathbf{E}_\infty[\tau] \geqslant T. \qquad (4)$$

It is easily seen, that in seeking solutions to the above problem, as suggested in [7], we can restrict our attention to stopping times that satisfy the false alarm constraint with equality. This is because, if $\mathbf{E}_\infty[\tau] > T$, we can produce a stopping time that achieves the constraint with equality without increasing the detection delay, simply by randomizing between τ and the stopping time that is identically 0. To this effect, we introduce the following definition.

D e f i n i t i o n 1. Define $\mathscr{K}$ to be the set of all stopping rules τ that are adapted to $\mathscr{F}_t$ and that satisfy $\mathbf{E}_\infty[\tau] = T$.

The paper is organized as follows. In Section 2 the one-sided CUSUM stopping rule along with its optimal character is presented. Section 3 is devoted to the presentation of the 2-CUSUM stopping rules and certain families amongst them that display interesting properties. Finally, in Section 4, two asymptotic optimality results are provided as $T \to \infty$.

2. The one-sided CUSUM stopping time.
The CUSUM statistic process and the corresponding one-sided CUSUM stopping time are defined as follows.

D e f i n i t i o n 2. Let $\lambda \in \mathbf{R}$ and $\nu \in \mathbf{R}_+$. Define the following processes:

1) $u_t(\lambda) = \lambda \xi_t - \frac{1}{2}\lambda^2 t$; $m_t(\lambda) = \inf_{0 \leqslant s \leqslant t} u_s(\lambda)$;
2) $y_t(\lambda) = u_t(\lambda) - m_t(\lambda) \geqslant 0$, which is the CUSUM statistic process;
3) $\tau_c(\lambda, \nu) = \inf\{t \geqslant 0; \ y_t(\lambda) \geqslant \nu\}$, which is the CUSUM stopping time.

We are now in a position to examine two very important properties of the one-sided CUSUM stopping time. The first is a characteristic specifically inherent to the CUSUM statistic and is summarized in the following lemma.

Lemma 1. *Fix $\theta \in [0, \infty)$. Let $t \geqslant \theta$ and consider the process*

$$y_{t,\theta} = u_t - u_\theta - \inf_{\theta \leqslant s \leqslant t}(u_s - u_\theta).$$

This is the CUSUM process when starting at time θ. We have that $y_t \geqslant y_{t,\theta}$ with equality if $y_\theta = 0$.

P r o o f. The proof is a matter of noticing that we can write

$$y_t = y_{t,\theta} + \left(\inf_{\theta \leqslant s \leqslant t}(u_s - u_\theta) + y_\theta\right)^+ \geqslant y_{t,\theta} \qquad (5)$$

and that $\inf_{\theta \leqslant s \leqslant t}(u_s - u_\theta) \leqslant 0$.

By its definition it is clear that $y_{t,\theta}$ depends only on information received *after* time θ. Thus, we conclude that all contribution of the observation process $\{\xi_t\}$ before time θ is summarized in y_θ. Relation (5), therefore, suggests that, as a function of the information before θ, the worst detection delay occurs whenever $y_\theta = 0$. In other words,

$$\operatorname{ess\,sup} \mathbf{E}_\theta\left[(\tau_c(\lambda, \nu) - \theta)^+ \mid \mathscr{F}_\theta\right] = \mathbf{E}_\theta\left[(\tau_c(\lambda, \nu) - \theta)^+ | y_\theta = 0\right]$$
$$= \mathbf{E}_0\left[\tau_c(\lambda, \nu)\right]. \qquad (6)$$

Equation (6) states that the CUSUM stopping time is an equalizer rule over θ, in the sense that its performance does not depend on the value of this parameter.

The second property of the one-sided CUSUM comes as a result of noticing that m_t is nonincreasing and that when it changes (decreases) we necessarily have $m_t = u_t$. In other words, when m_t changes, y_t attains its smallest value, that is 0. When this happens we will say that the CUSUM statistic process *restarts*. This important observation combined with standard results appearing in [5] allow for the computation of the CUSUM delay function.

Lemma 2. *Suppose a CUSUM stopping rule is based on the CUSUM statistic with drift parameter $\lambda \in \mathbf{R}$ and has threshold $\nu \in \mathbf{R}_+$. Then the detection delay when the observation process ξ_t has drift $\mu \in \mathbf{R}$ is given by* $\mathbf{E}[\tau_c(\lambda, \nu)] = (2/\lambda^2)\, g(\nu, \rho)$, *where*

$$g(\nu, \rho) = \frac{e^{-\rho\nu} + \rho\nu - 1}{\rho^2} \quad and \quad \rho = 2\frac{\mu}{\lambda} - 1.$$

P r o o f. Consider the function $f(y) = (2/\lambda^2)[g(\nu, \rho) - g(y, \rho)]$. Then f is a twice continuously differentiable function of y satisfying

$$\rho f'(y) + f''(y) = -1, \quad \text{with} \quad f'(0) = f(\nu) = 0.$$

Using standard Itô calculus on the process $f(y_t)$ and the results appearing in [5, pp. 149, 210] it is easy to show that for any stopping time τ with $\mathbf{E}[\tau] < \infty$, we have:

$$\mathbf{E}[f(y_\tau)] - f(y_0) = -\mathbf{E}[\tau].$$

The desired formula follows by noticing that $y_0 = 0$ and for the CUSUM stopping time we have $y_{\tau_c} = \nu$ (for more details see also [8]).

Notice that for $\alpha \neq 0$ we have $\alpha^{-2} g(\nu, \rho) = g(\nu/|\alpha|, \rho|\alpha|)$. This suggests the following alternative expression for the delay function:

$$\mathbf{E}[\tau_c(\lambda, \nu)] = 2g\left(\frac{\nu}{|\lambda|}, \ \text{sign}\,(\lambda)(2\mu - \lambda)\right). \tag{7}$$

In [2] and [14] it is shown that when there is only one possible alternative for the drift μ, the CUSUM stopping rule $\tau_c(\mu, \nu)$, with ν satisfying $(2/\mu^2)\, g(\nu, -1) = T$, solves the optimization problem defined in (2). It is also interesting to note that in [8], after a proper modification of Lorden's criterion that replaces expected delays with Kullback–Leibler divergences, the optimality of the CUSUM can be extended to cover detection of general changes in Itô processes.

When the sign of the alternative drifts is the same, with the help of the following lemma we can show that the one-sided CUSUM stopping rule that detects the smallest in absolute value drift is the optimal solution of the problem in (4).

Lemma 3. *For every path of the Brownian motion w_t, the process $y_t(\lambda)$ is an increasing (decreasing) function of the drift of the observation process ξ_t when $\lambda > 0$ ($\lambda < 0$).*

P r o o f. Consider two possible drift values μ_1, μ_2 with $\mu_1 < \mu_2$. We define the following two observation processes $\xi_t(\mu_i) = \mu_i(t - \theta)^+ + w_t$, $i = 1, 2$, that lead to the corresponding CUSUM processes

$$u_t(\lambda, \mu_i) = \lambda \xi_t(\mu_i) - \frac{1}{2}\lambda^2 t = \lambda\{w_t + \mu_i(t - \theta)^+\} - \frac{1}{2}\lambda^2 t,$$

$$m_t(\lambda, \mu_i) = \inf_{0 \leqslant s \leqslant t} u_s(\lambda, \mu_i),$$

$$y_t(\lambda, \mu_i) = u_t(\lambda, \mu_i) - m_t(\lambda, \mu_i).$$

Consider the difference $y_t(\lambda, \mu_2) - y_t(\lambda, \mu_1) = \delta(t - \theta)^+ - m_t(\lambda, \mu_2) + m_t(\lambda, \mu_1)$, where $\delta = \lambda(\mu_2 - \mu_1)$. Notice now that $\lambda > 0$ implies $\delta > 0$ and we can write

$$u_s(\lambda, \mu_2) = u_s(\lambda, \mu_1) + \delta(s - \theta)^+ \leqslant u_s(\lambda, \mu_1) + \delta(t - \theta)^+.$$

Taking the infimum over $0 \leqslant s \leqslant t$ we get $m_t(\lambda, \mu_2) \leqslant m_t(\lambda, \mu_1) + \delta(t - \theta)^+$ from which, by rearranging terms, we get that $y_t(\lambda, \mu_2) \geqslant y_t(\lambda, \mu_1)$. The case $\lambda < 0$ can be shown similarly. Lemma 3 is proved.

From Lemma 3 it also follows that $\mu_1 \leqslant \mu_2$ implies $\mathbf{E}^1[\tau_c(\lambda, \nu)] \geqslant \mathbf{E}^2[\tau_c(\lambda, \nu)]$ when $\lambda > 0$ and the opposite when $\lambda < 0$. As a direct consequence of this fact comes our first optimality result concerning drifts with the same sign.

Theorem 1. *Let $0 < \mu_1 \leqslant \mu_2$ or $\mu_2 \leqslant \mu_1 < 0$, then the one-sided CUSUM stopping time $\tau_c(\mu_1, \nu_1)$ with ν_1 satisfying $(2/\mu_1^2) g(\nu_1, -1) = T$ solves the optimization problem defined in (4).*

P r o o f. The proof is straightforward. Since ν_1 was selected so that $\tau_c(\mu_1, \nu_1)$ satisfies the false alarm constraint, we have $\tau_c(\mu_1, \nu_1) \in \mathcal{K}$. Then, for all $\tau \in \mathcal{K}$ we have

$$J_L(\tau) = \max_i \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta^i[(\tau - \theta)^+ \mid \mathscr{F}_\theta] \geqslant \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta^1[(\tau - \theta)^+ \mid \mathscr{F}_\theta]$$

$$\geqslant \mathbf{E}_0^1[\tau_c(\mu_1, \nu_1)] = \max_i \mathbf{E}_0^i[\tau_c(\mu_1, \nu_1)] = J_L(\tau_c(\mu_1, \nu_1)) = \frac{2}{\mu_1^2} g(\nu_1, 1).$$

The last inequality comes from the optimality of the one-sided CUSUM stopping rule and the last three equalities are due to Lemma 3, the definition of the performance measure $J_L(\tau)$ in (3), and Lemma 2. Theorem 1 is proved.

It is worth pointing out that if we had n alternative drifts (instead of two) of the form $0 < \mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_n$ or $0 > \mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$ and we used the extended Lorden's criterion in (3), the optimality of $\tau_c(\mu_1, \nu_1)$, presented in Theorem 1, would still be valid. Our result should be compared to [4] (which refers to discrete time and the exponential family), where for the same type of changes only asymptotically optimum schemes are offered.

We also have the following corollary of Lemma 3.

Corollary 1. *Let* $0 < |\mu_1| \leqslant |\mu_2|$ *and define* η_i, $i = 1, 2$, *so that* $(2/\mu_i^2) g(\eta_i, -1) = T > 0$. *Then we have*

$$\frac{1}{\mu_1^2} g(\eta_1, 1) \geqslant \frac{1}{\mu_2^2} g(\eta_2, 1). \tag{8}$$

P r o o f. Since the result is independent of the sign of the two drifts, without loss of generality we may assume $0 < \mu_1 \leqslant \mu_2$. Consider the two CUSUM rules $\tau_c(\mu_i, \eta_i)$, $i = 1, 2$. Because the two thresholds η_i were selected to satisfy the false alarm constraint, using Lemma 1, Lemma 3 and the optimality of the one-sided CUSUM stopping time, the following inequalities hold for all $\tau \in \mathcal{K}$:

$$\begin{aligned}
\frac{2}{\mu_1^2} g(\eta_1, 1) &= \mathbf{E}_0^1[\tau_c(\mu_1, \eta_1)] \geqslant \mathbf{E}_0^2[\tau_c(\mu_1, \eta_1)] \\
&= \sup_\theta \text{ess}\sup \mathbf{E}_\theta^2[(\tau_c(\mu_1, \eta_1) - \theta)^+ \mid \mathscr{F}_\theta] \\
&\geqslant \inf_\tau \sup_\theta \text{ess}\sup \mathbf{E}_\theta^2[(\tau - \theta)^+ \mid \mathscr{F}_\theta] \\
&= \mathbf{E}_0^2[\tau_c(\mu_2, \eta_2)] = \frac{2}{\mu_2^2} g(\eta_2, 1).
\end{aligned}$$

Corollary 1 is proved.

3. Different drift signs and the 2-CUSUM stopping time. Let us now consider the case $\mu_2 < 0 < \mu_1$. The very interesting problem of knowing the amplitude of the drift but not the sign falls into this setting. What has traditionally been done in the literature, dating as far back as Barnard in [1], is to use the minimum of the stopping rules $\tau_c(\mu_1, \nu_1)$ and $\tau_c(\mu_2, \nu_2)$ each tuned to detect the respective changes μ_1 and μ_2. To this effect. we introduce the following 2-CUSUM stopping rule.

D e f i n i t i o n 3. Let $\lambda_2 < 0 < \lambda_1$. The 2-CUSUM stopping time $\tau_{2c}(\lambda_1, \lambda_2, \nu_1, \nu_2)$ is defined as follows: $\tau_{2c}(\lambda_1, \lambda_2, \nu_1, \nu_2) = \tau_c(\lambda_1, \nu_1) \wedge \tau_c(\lambda_2, \nu_2)$.

We will, from now on, denote all 2-CUSUM rules by τ_{2c} unless it is necessary to give emphasis to their four parameters. By the definition of the 2-CUSUM stopping rule it is apparent that it consists of running in parallel the two CUSUM statistic processes $y_t(\lambda_1)$ and $y_t(\lambda_2)$ and stopping whenever one of the two hits its corresponding threshold for the first time. From Lemma 1 we can conclude that

$$\begin{aligned}
\text{ess}\sup \mathbf{E}_\theta^i[(\tau_{2c} - \theta)^+ \mid \mathscr{F}_\theta] &= \mathbf{E}_\theta^i[(\tau_{2c} - \theta)^+ \mid y_\theta(\lambda_1) = y_\theta(\lambda_2) = 0] \\
&= \mathbf{E}_0^i[\tau_{2c}], \tag{9}
\end{aligned}$$

from which we get

$$J_L(\tau_{2c}) = \max_i \sup_\theta \text{ess}\sup \mathbf{E}_\theta^i[(\tau_{2c} - \theta)^+ \mid \mathscr{F}_\theta] = \max_i \mathbf{E}_0^i[\tau_{2c}].$$

As we have seen the 2-CUSUM stopping rule is characterized by the four parameters, $\lambda_1, \lambda_2, \nu_1$, and ν_2. Since our intention is to propose a specific rule as the «preferable» one, we need to come up with a specific selection of these parameters. For this purpose, up to this point, we only have one equation available, namely, the false alarm constraint $\mathbf{E}_\infty[\tau_{2c}] = T$. Hence, we will gradually impose additional constraints on our 2-CUSUM structure in order to arrive to a unique stopping rule. Once our rule is specified we will support its selection by demonstrating that it enjoys a strong asymptotic optimality property.

3.1. A special class of 2-CUSUM rules. First we shed our attention to a specific class of 2-CUSUM stopping rules that allow for the exact computation of their performance.

D e f i n i t i o n 4. Define

$$\mathcal{G} = \{\tau_{2c}(\lambda_1, \lambda_2, \nu_1, \nu_2); \ \nu_1 = |\lambda_1|\nu \text{ and } \nu_2 = |\lambda_2|\nu\}.$$

For $\tau_{2c} \in \mathcal{G}$ we have the following characteristic property.

Lemma 4. *Let $\tau_{2c} \in \mathcal{G}$ then, when τ_{2c} stops, one of its CUSUM statistic processes hits its corresponding threshold while the other necessarily restarts.*

P r o o f. Although the proof given in [15, p. 28] for discrete time and the exponential family, applies here as well (without major changes), we prefer to give an alternative (hopefully easier) proof. Consider the process $\{Y_t\}$ with

$$Y_t = |\lambda_2|\, y_t(\lambda_1) + |\lambda_1|\, y_t(\lambda_2) = -\frac{1}{2}\left(|\lambda_2|\lambda_1^2 + |\lambda_1|\lambda_2^2\right)t - |\lambda_2|\, m_t(\lambda_1) - |\lambda_1|\, m_t(\lambda_2).$$

Since $y_t(\lambda_i) \geqslant 0$ we clearly have $Y_t \geqslant 0$. Let us suppose that $Y_t > 0$. Then we notice that, when both processes $m_t(\lambda_i)$, $i = 1, 2$, stay constant, Y_t *decreases* linearly in time. From this we conclude that Y_t can increase only when at least one of the two processes $m_t(\lambda_i)$ changes (decreases). This implies that the corresponding CUSUM processes $y_t(\lambda_i)$ restarts. We obviously cannot have both CUSUM processes restarting, since that would yield $Y_t = 0$. By its definition, the 2-CUSUM rule stops when one of the two CUSUM processes hits its corresponding threshold. At this instant, we necessarily have $Y_t \geqslant |\lambda_1\lambda_2|\nu$. In fact we are going to argue that equality holds. Indeed we can see that when Y_t hits the level $|\lambda_1\lambda_2|\nu > 0$ *for the first time*, since Y_t attains a new level, it has to be during an increase. But the latter can only happen when one of the two CUSUM processes restarts while the other necessarily hits its threshold. Lemma 4 is proved.

The following lemma uses the above property to derive a formula for the expected delay of the 2-CUSUM rule.

Lemma 5. *Let $\tau_{2c} = \tau_1 \wedge \tau_2$ with $\tau_{2c} \in \mathscr{G}$ and τ_1, τ_2 the corresponding one-sided CUSUM branches. Then the expected delay of the 2-CUSUM stopping time τ_{2c} is related to the corresponding delays of its one-sided CUSUM branches through the formula*

$$(\mathbf{E}[\tau_{2c}])^{-1} = (\mathbf{E}[\tau_1])^{-1} + (\mathbf{E}[\tau_2])^{-1}. \tag{10}$$

P r o o f. The proof basically repeats the one presented in [15, p. 28] for the discrete time case.

3.2. 2-CUSUM equalizer rules. It is well known that min-max problems, such as (4), are solved by *equalizer rules*. In other words, by stopping rules that demonstrate the same performance under the two changes. Thus, we further restrict ourselves among the class of equalizer rules.

D e f i n i t i o n 5. Define

$$\mathscr{D} = \{\tau_{2c} \in \mathscr{G}; \ \mathbf{E}_0^1[\tau_{2c}] = \mathbf{E}_0^2[\tau_{2c}]\}.$$

By the definition of the class of equalizer rules it follows that $\mathscr{D} \subset \mathscr{G}$. Let us now find a simple condition that guarantees this property.

By using (7), (10) we get for $i = 1, 2$

$$\mathbf{E}_0^i[\tau_{2c}] = \left(\frac{1}{2g(\nu, \text{sign}\,(\lambda_1)(2\mu_i - \lambda_1))} + \frac{1}{2g(\nu, \text{sign}\,(\lambda_2)(2\mu_i - \lambda_2))} \right)^{-1}. \tag{11}$$

From (11) we can see that in order to have $\tau_{2c} \in \mathscr{D}$ we need

$$\text{sign}\,(\lambda_1)(2\mu_1 - \lambda_1) = \text{sign}\,(\lambda_2)(2\mu_2 - \lambda_2), \tag{12}$$

$$\text{sign}\,(\lambda_2)(2\mu_1 - \lambda_2) = \text{sign}\,(\lambda_1)(2\mu_2 - \lambda_1). \tag{13}$$

One can now easily verify that both of the above equations (12) and (13) are satisfied whenever

$$\lambda_1 + \lambda_2 = 2(\mu_1 + \mu_2). \tag{14}$$

In other words, if we select λ_1, λ_2 to satisfy (14), then the corresponding 2-CUSUM stopping rule has the same performance under both drifts μ_1, μ_2.

By limiting ourselves to the class $\mathscr{D}$ (i.e., selecting $\nu_1 = \lambda_1 \nu$, $\nu_2 = |\lambda_2|\nu$ and using (14)), apart from the false alarm constraint, we impose two additional constraints on our four parameters. In order for the 2-CUSUM rule to be completely specified we need one final condition. Our intention is to select the parameter λ_1 so that the corresponding detection delay is asymptotically (as $T \to \infty$) minimized.

Theorem 2. *Let $\mu_2 < 0 < \mu_1$ with $|\mu_1| \leqslant |\mu_2|$. Consider all 2-CUSUM stopping times $\tau_{2c} \in \mathscr{K} \cap \mathscr{D}$. Then among all such stopping rules the one with $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$ is asymptotically optimal as $T \to \infty$.*

P r o o f. Since $\mu_1 + \mu_2 \leqslant 0$, for any $\lambda_1 > 0$, from equation (14) we get $|\lambda_1| \leqslant |\lambda_2|$. Let us first consider the false alarm constraint. Using (7), (10) with $\mu = 0$ and $\nu_1 = |\lambda_1|\nu$, $\nu_2 = |\lambda_2|\nu$, we get

$$\mathbf{E}_\infty[\tau_{2c}] = \left(\frac{1}{2g(\nu, -|\lambda_1|)} + \frac{1}{2g(\nu, -|\lambda_2|)}\right)^{-1} = T. \tag{15}$$

By carefully examining the exponential rates of the two terms in (15) we conclude that the leading term is the one containing λ_1. Hence, we get

$$\lambda_1 \nu = \ln T(1 + o(1)). \tag{16}$$

For the common detection delay, using equation (11) and substituting $\lambda_2 = 2(\mu_1 + \mu_2) - \lambda_1$ we have the following estimates:

$$\mathbf{E}_0^i[\tau_{2c}] = \left(\frac{1}{2g(\nu, 2\mu_1 - \lambda_1)} + \frac{1}{2g(\nu, 2\mu_2 - \lambda_1)}\right)^{-1}$$

$$= \begin{cases} \dfrac{2\nu}{2\mu_1 - \lambda_1}(1 + o(1)) & \text{for} \quad 2\mu_1 > \lambda_1 \geqslant 0, \\[2mm] \nu^2(1 + o(1)) & \text{for} \quad 2\mu_1 = \lambda_1, \\[2mm] \dfrac{2e^{\nu|2\mu_1 - \lambda_1|}}{(2\mu_1 - \lambda_1)^2}(1 + o(1)) & \text{for} \quad 2\mu_1 < \lambda_1. \end{cases} \tag{17}$$

The objective is to minimize the detection delay with respect to λ_1 in order to find the best selection for this parameter. From (17) it is clear that it is sufficient to limit ourselves to the case $0 \leqslant \lambda_1 < 2\mu_1$, since for $\lambda_1 \geqslant 2\mu_1$ the detection delay increases significantly faster as ν increases. For $0 \leqslant \lambda_1 < 2\mu_1$, the detection delay, after substituting ν from (16), can be written as

$$\frac{2\ln T}{\lambda_1(2\mu_1 - \lambda_1)}(1 + o(1)),$$

which is clearly minimized, asymptotically, for $\lambda_1 = \mu_1$. Using equation (14), we also get $\lambda_2 = 2\mu_2 + \mu_1$. Theorem 2 is proved.

Let us now summarize our results. We propose the following 2-CUSUM rule for the case $\mu_2 < 0 < \mu_1$: when $|\mu_1| \leqslant |\mu_2|$ select $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$, $\nu_1 = |\mu_1|\nu$, $\nu_2 = |2\mu_2 + \mu_1|\nu$. If $|\mu_1| \geqslant |\mu_2|$, then $\lambda_1 = 2\mu_1 + \mu_2$, $\lambda_2 = \mu_2$, $\nu_1 = |2\mu_1 + \mu_2|\nu$, $\nu_2 = |\mu_2|\nu$. Finally, the parameter ν is selected so as to satisfy the false alarm constraint (15).

4. Asymptotic optimality in opposite sign drifts. For the specific 2-CUSUM rule introduced at the end of the previous section, we are going to demonstrate two asymptotic optimality results. By means of an upper and a lower bound on the performance of the unknown optimal stopping

rule, we will show that in the case of equal in absolute value drifts the *differ-
ence* in performance between the unknown optimum rule and the proposed
2-CUSUM rule tends to a constant as $T \to \infty$. In the case of different in
absolute value drifts we have a stronger asymptotic result. In particular, we
will demonstrate that the *difference* in performance between the unknown
optimal rule and the proposed 2-CUSUM rule tends to 0 as $T \to \infty$. This
should be compared to most existing asymptotic optimality results, where it
is shown that the *ratio* between the performance of the optimum and the pro-
posed scheme tends to unity (first order optimality). Our form of asymptotic
optimality is clearly stronger since it implies first order optimality, while the
opposite is not necessarily true.

Let τ_{2c} denote the specific 2-CUSUM rule proposed in the previous
section with the threshold ν selected so that the false alarm constraint is
satisfied with equality. Since τ_{2c} constitutes a possible choice in the class $\mathscr{K}$,
equation (9) and Lemma 2 imply that for all $\tau \in \mathscr{K}$

$$\mathbf{E}_0^1[\tau_{2c}] = \mathbf{E}_0^2[\tau_{2c}] = J_L(\tau_{2c}) \geqslant \inf_\tau J_L(\tau). \tag{18}$$

To find a lower bound, we observe that for all $\tau \in \mathscr{K}$ we can write

$$\inf_\tau J_L(\tau) = \inf_\tau \max_i \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta^i[(\tau - \theta)^+ \mid \mathscr{F}_\theta]$$

$$\geqslant \max_i \left(\inf_\tau \sup_\theta \operatorname{ess\,sup} \mathbf{E}_\theta^i[(\tau - \theta)^+ \mid \mathscr{F}_\theta] \right) = \max_i \frac{2}{\mu_i^2} g(\eta_i, 1), \tag{19}$$

where for the last equality we used the optimality of the one-sided CUSUM
stopping rule and the expression for its worst detection delay from Lemma 2.
The two thresholds η_i, $i = 1, 2$, are selected to satisfy the false alarm con-
straint $(2/\mu_i^2) g(\eta_i, -1) = T$. The asymptotic results that follow examine
the way the two bounds approach each other. Since the performance of the
optimal stopping rule is between the two bounds, this will also determine
the rate with which the 2-CUSUM approaches the optimal solution.

4.1. The case of equal in absolute value drifts. We first consider
the special case $\mu_1 = -\mu_2 = \mu$. Here our parameter selection takes the
form $\lambda_1 = \mu_1 = \mu$ and $\lambda_2 = 2\mu_2 + \mu_1 = \mu_2 = -\mu$ which coincides with the
2-CUSUM scheme proposed in the literature. Let us now examine the two
bounds. The upper bound, from (11), with this specific parameter selection
becomes

$$J_L(\tau_{2c}) = \mathbf{E}_0^i[\tau_{2c}] = \left(\frac{1}{2g(\nu, \mu)} + \frac{1}{2g(\nu, -3\mu)} \right)^{-1}, \qquad i = 1.2. \tag{20}$$

with the threshold ν computed from the false alarm constraint (15) that
takes the form

$$\mathbf{E}_\infty[\tau_{2c}] = \left(\frac{1}{2g(\nu, -\mu)} + \frac{1}{2g(\nu, -\mu)} \right)^{-1} = g(\nu, -\mu) = T. \tag{21}$$

Similarly, the lower bound becomes $(2/\mu^2)\,g(\eta,1)$ with the threshold η satisfying $(2/\mu^2)\,g(\eta,-1) = T$.

Theorem 3. *The difference in the performance between the proposed 2-CUSUM stopping rule and the optimal stopping rule is asymptotically, as the false alarm constraint $T \to \infty$, bounded by the constant $(2\log 2)/\mu^2$.*

P r o o f. Solving for ν from (21) we obtain $\mu\nu = \ln T + \ln(\mu^2/2) + \ln 2 + o(1)$. On the other hand, we can write (20) as $J_L(\tau_{2c}) = (2/\mu^2)\{\mu\nu + e^{-\mu\nu} - 1\}\{1 + O(\mu\nu e^{-3\mu\nu})\}$. Substituting the estimate for ν we get

$$J_L(\tau_{2c}) = \frac{2}{\mu^2}\left\{\ln T + \ln\frac{\mu^2}{2} - 1 + \ln 2 + o(1)\right\}.$$

Similarly, for the lower bound we have that the threshold η as a function of T becomes $\eta = \ln T + \ln(\mu^2/2) + o(1)$. Therefore, the lower bound is of the form $(2/\mu^2)\{\ln T + \ln(\mu^2/2) - 1 + o(1)\}$. Since the difference between the upper and the lower bound bounds the difference $J_L(\tau_{2c}) - \inf_\tau J_L(\tau)$, we conclude that $0 \leqslant J_L(\tau_{2c}) - \inf_\tau J_L(\tau) \leqslant (2/\mu^2)\{\ln 2 + o(1)\}$, from which the result follows by letting $T \to \infty$. Theorem 3 is proved.
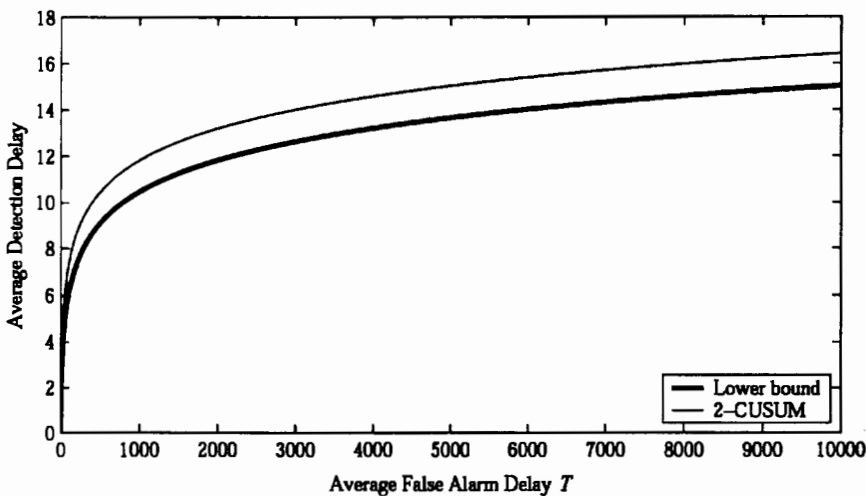

Fig. 1. Typical form of the upper and lower bounds of the performance of the optimum stopping rule for the case $\mu_1 = -\mu_2 = 1$.

Figure 1 depicts the upper and lower bound as a function of the false alarm constraint T for the case $\mu_1 = -\mu_2 = 1$. Since, as we can see, the difference of the two bounds is increasing with T, the constant proposed by Theorem 3 corresponds to a *worst case* performance attained only in the limit as $T \to \infty$.

4.2. The case of different in absolute value drifts.

Theorem 4. *The difference in the performance between the proposed 2-CUSUM stopping rule and the optimal stopping rule tends to 0 as the false alarm constraint $T \to \infty$.*

P r o o f. We will only examine the case $|\mu_1| < |\mu_2|$. From Corollary 1 and equation (8) it follows that the maximum in the lower bound in (19) is achieved for μ_1. Hence, as in Theorem 3, we get $(2/\mu_1^2)\{\ln T + \ln(\mu_1^2)/2 - 1 + o(1)\}$ for the lower bound.

The upper bound is the detection delay of the proposed 2-CUSUM stopping time τ_{2c}. From (11), with $\lambda_1 = \mu_1$, $\lambda_2 = 2\mu_2 + \mu_1$, we have

$$
J_L(\tau_{2c}) = \mathbf{E}_0^i[\tau_{2c}] = \left(\frac{1}{2g(\nu, \mu_1)} + \frac{1}{2g(\nu, 2\mu_2 - \mu_1)} \right)^{-1}
$$
$$
= \frac{2}{\mu_1^2} \{e^{-\mu_1 \nu} + \mu_1 \nu - 1\}\{1 + O(\mu_1 \nu e^{(2\mu_2 - \mu_1)\nu})\}, \quad (22)
$$

where ν is selected to satisfy the false alarm constraint, which by (15) takes the form

$$
\mathbf{E}_\infty[\tau_{2c}] = \left(\frac{1}{2g(\nu, -\mu_1)} + \frac{1}{2g(\nu, 2\mu_2 + \mu_1)} \right)^{-1} = T. \quad (23)
$$

From (23) we get the estimate $\mu_1 \nu = \ln T - \ln(\mu_1^2 \, 2) - o(1)$. This, when substituted in (22), produces

$$
J_L(\tau_{2c}) = \mathbf{E}_0^i[\tau_{2c}] = \frac{2}{\mu_1^2} \left\{ \ln T + \ln \frac{\mu_1^2}{2} - 1 - o\,1 \right\}. \quad (24)
$$

Subtracting now the lower bound expression from the upper bound expression in (24) we obtain

$$
0 \leqslant J_L(\tau_{2c}) - \inf_\tau J_L(\tau) \leqslant o\,1\,.
$$

which tends to 0 as $T \to \infty$. Theorem 4 is proved.

In Figure 2 we present the two bounds for $\mu_1 = 1$ and $\mu_2 = -1.05, -1.15, -1.3$. We recall that the upper bound is the detection delay of the 2-CUSUM rule $\tau_{2c} \in \mathscr{G} \cap \mathscr{K}$ with parameters $\lambda_1 = \mu_1$ and $\lambda_2 = 2\mu_2 + \mu_1$. We can see that the difference between the two curves is tending to zero as the false alarm tends to infinity, thus corroborating Theorem 4. What is more interesting, however, is the fact that the two curves rapidly approach each other, *uniformly* over T, as the ratio $|\mu_2|/|\mu_1|$ of the two drifts increases. As we can see, in the case $\mu_1 = 1$, $\mu_2 = -1.3$ the two bounds become almost indistinguishable. This suggests that the proposed 2-CUSUM rule can be (extremely) close to the unknown optimal rule, not only asymptotically, as proposed by Theorem 4, but also uniformly over all false alarm values.
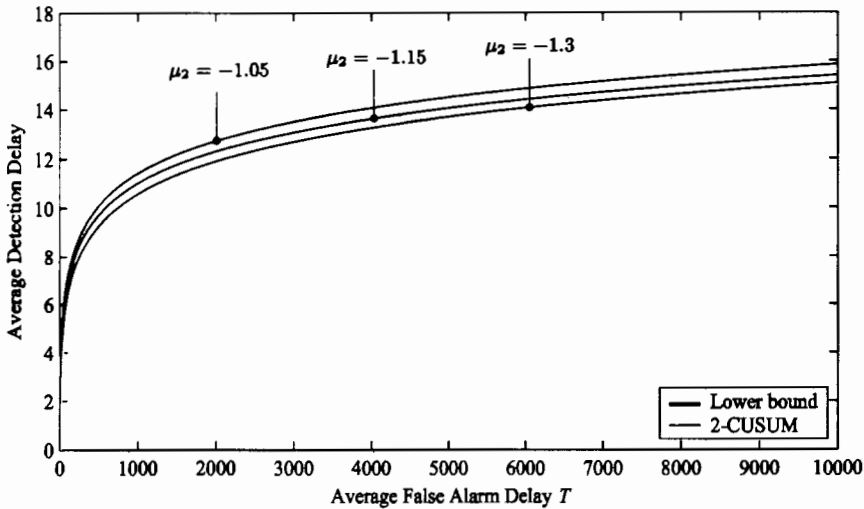
Fig. 2. Typical form of the upper and lower bounds of the performance of the optimal stopping rule for the case $\mu_2 < 0 < \mu_1$, with $\mu_1 = 1$ and $\mu_2 = -1.05, -1.15, -1.3$.

It is also worth noting that the difference in the performance of the optimal rule and any 2-CUSUM rule in $\mathscr{G}$ with parameters $\lambda_1 = \mu_1$ and $\lambda_2 \in (-\mu_1, 2\mu_2 + \mu_1]$ (one such possibility is the selection proposed in the literature $\lambda_1 = \mu_1$, $\lambda_2 = \mu_2$) also tends to 0 as $T \to \infty$. Therefore, asymptotically optimal solutions allow for many different choices. It is, however, our selection that leads to an equalizer rule.

5. Acknowledgement. The authors are grateful to Professor Shiryaev for his thoughtful suggestions.

REFERENCES

1. *Barnard G. A.* Control charts and stochastic processes. — J. Roy. Statist. Soc. Ser. B, 1959, v. 21, p. 239–271.
2. *Beibel M.* A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure. — Ann. Statist., 1996, v. 24, №4, p. 1804–1812.
3. *Драгалин В. П.* Оптимальность обобщенного алгоритма кумулятивных сумм в задаче скорейшего обнаружения разладки. — Труды МИРАН, 1993, т. 202, с. 132–148.
4. *Dragalin V. P.* The design and analysis of 2-CUSUM procedure. — Comm. Statist. Simulation Comput., 1997, v. 26, №1, p. 67–81.
5. *Karatzas I., Shreve S. E.* Brownian Motion and Stochastic Calculus. New York: Springer-Verlag, 1991, 470 p.
6. *Lorden G.* Procedures for reacting to a change in distribution. — Ann. Math. Statist., 1971, v. 42, pp. 1897–1908.
7. *Moustakides G. V.* Optimal stopping rules for detecting changes in distributions. — Ann. Statist., 1986, v. 14, p. 1379–1387.

8. *Moustakides G. V.* Optimality of the CUSUM procedure in continuous time. — Ann. Statist., 2004, v. 32, № 1, p. 302–315.

9. *Page E. S.* Continuous inspection schemes. — Biometrika, 1954, v. 41, p. 100–115.

10. *Pollak M., Siegmund D.* A diffusion process and its application to detecting a change in the drift of Brownian Motion. — Biometrika, 1985, v. 72, № 2, p. 267–280.

11. *Roberts S.* Control chart tests based on geometric moving average. — Technometrics, 1959, v. 1, p. 239–250.

12. *Roberts S.* A comparison of some control chart procedures. — Technometrics, 1966, v. 8, p. 411–430.

13. *Ширяев А. Н.* Об оптимальных методах в задачах скорейшего обнаружения. — Теория вероятн. и ее примен., 1963, т. 8, в. 1, с. 26–51.

14. *Ширяев А. Н.* Минимаксная оптимальность метода кумулятивных сумм (CUSUM) в случае непрерывного времени. — Успехи матем. наук, 1996, т. 51, № 4, с. 173–174.

15. *Siegmund D.* Sequential Analysis. New York: Springer-Verlag, 1985, 272 p.

16. *Srivastava M., Wu Y. H.* Comparison of EWMA, CUSUM and Shiryayev–Roberts procedures for detecting a shift in the mean. — Ann. Statist., 1993, v. 21, № 2, p. 645–670.

17. *Тартаковский А. Г.* Асимптотически минимаксное многоальтернативное последовательное правило обнаружения разладки. — Труды МИАН, 1993, т. 202, с. 287–295.