The Annals of Statistics
2004, Vol. 32, No. 1, 302–315

OPTIMALITY OF THE CUSUM PROCEDURE IN CONTINUOUS TIME

BY GEORGE V. MOUSTAKIDES

Institut National de Recherche en Informatique et en Automatique and University of Thessaly

The optimality of CUSUM under a Lorden-type criterion setting is considered. We demonstrate the optimality of the CUSUM test for Itô processes, in a sense similar to Lorden's, but with a criterion that replaces expected delays by the corresponding Kullback–Leibler divergence.

1. Introduction. The cumulative sum (CUSUM) test was proposed by Page (1954) as a means to detect sequentially changes in distributions of discrete-time random processes. Lorden (1971) introduced a min–max criterion for the change detection problem, and established the asymptotic optimality of the CUSUM test under his proposed performance measure. Moustakides (1986) proved optimality, under Lorden's criterion, for the i.i.d. case and for known distributions before and after the change. Ritov (1990) demonstrated a Bayesian optimality property of CUSUM, based on which he also provided an alternative proof for optimality in Lorden's sense. Moustakides (1998) extended the optimality of CUSUM to a special class of dependent processes. Finally, optimum CUSUM procedures were proposed by Poor (1998) for exponentially penalized detection delays.

In continuous time, the optimality of CUSUM has been established for Brownian motion with constant drift by Beibel (1996), in the Bayesian setting of Ritov (1990), which also yielded optimality in Lorden's sense, and by Shiryayev (1996). These results should be compared to the significantly richer and more general ones available for the other popular sequential test, the sequential probability ratio test (SPRT). In continuous time, the SPRT was shown to be optimal in Wald's sense [Wald (1947)] for Brownian motion with constant drift by Shiryayev [(1978), page 180]. However, when one replaces in Wald's criterion the expected delay by the Kullback–Leibler (K–L) divergence, then Liptser and Shiryayev [(1978), page 224] demonstrated the optimality of the SPRT for Itô processes. This result was subsequently extended by Yashin (1983) and Irle (1984) to more general continuous-time processes.

It is the goal of this work to demonstrate a similar extension for the optimality of CUSUM. In particular, we shall show that the CUSUM is optimum in detecting changes in the statistics of Itô processes, in a Lorden-like sense, when the expected delay is replaced in the criterion by the corresponding K–L divergence. It should

Received July 2002; revised March 2003.

AMS 2000 subject classifications. Primary 62L10; secondary 62L15, 62C10.

Key words and phrases. CUSUM, change point, disorder problem, sequential detection, Kullback–Leibler divergence.

be noted that, for the special case of Brownian motion with constant drift, the original Lorden criterion and the modified one proposed here coincide; thus, our result also provides a different proof for the Lorden min–max problem considered in Beibel (1996) and Shiryayev (1996).

2. Assumptions and background results. Let ξ be a continuous-time process, and define the filtration $\mathcal{F}$ given by $\mathcal{F}_t = \sigma(\xi_s; 0 \le s \le t)$. We are interested in the case where ξ is an Itô process satisfying

$$d\xi_t = \alpha_t \mathbb{1}_{\{t>\tau\}}\, dt + dw_t, \tag{2.1}$$

where α is a process adapted to the filtration $\mathcal{F}$, w is a standard Brownian motion with respect to the same filtration and $\tau \in [0, \infty]$ denotes the time of change of regime, which is considered deterministic but otherwise unknown. Moreover, we assume that $\mathcal{F}_0$ is the trivial σ-algebra.

Given that ξ is observed sequentially, and assuming exact knowledge of the model (2.1) before and after the change, our goal is to detect the change time τ as soon as possible using a sequential scheme.

Let us introduce several definitions, assumptions and key results that are necessary for our analysis. Let $\mathbb{P}_\tau$ denote the probability measure when the change is at time τ and $\mathbb{E}_\tau[\cdot]$ the corresponding expectation. With this notation, $\mathbb{P}_0$ is the measure corresponding to the case of all observations being under the alternative model, whereas $\mathbb{P}_\infty$ corresponds to all observations being under the nominal one. In other words, $\mathbb{P}_\infty$ is the Wiener measure on the canonical space of continuous functions, and $\mathbb{P}_0$ is the measure induced on this space by the process $w_t + \int_0^t a_s\, ds$.

We now need a first condition to ensure that ξ introduced in (2.1) is well defined. Following Øksendal [(1998), page 44] we require the process α to be $\mathcal{F}$ adapted and to satisfy

$$\mathbb{P}_0\left[\int_0^t |\alpha_s|\, ds < \infty\right] = 1 \qquad \forall\, t \in [0, \infty). \tag{2.2}$$

The next step is to impose conditions that will guarantee the existence of the Radon–Nikodym derivative $d\mathbb{P}_\tau / d\mathbb{P}_\infty$ and validity of Girsanov's theorem. For this purpose, consider the process

$$u_t = \int_0^t \alpha_s\, d\xi_s - \tfrac{1}{2}\int_0^t \alpha_s^2\, ds,$$

which, because of (2.1), satisfies

$$du_t = \begin{cases} \alpha_t\, dw_t - \frac{1}{2}\alpha_t^2\, dt, & 0 \le t \le \tau, \\ \alpha_t\, dw_t + \frac{1}{2}\alpha_t^2\, dt, & \tau < t < \infty. \end{cases} \tag{2.3}$$

In order that this process be well defined under both hypotheses, again from Øksendal [(1998), page 44] we need to assume that for every $0 \le t < \infty$ we have

$$\mathbb{P}_0\left[\int_0^t \alpha_s^2\, ds < \infty\right] = \mathbb{P}_\infty\left[\int_0^t \alpha_s^2\, ds < \infty\right] = 1. \tag{2.4}$$

Since $(\int_0^t |\alpha_s|\,ds)^2 \le t\int_0^t \alpha_s^2\,ds$, it is clear that (2.4) also implies (2.2).

To ensure now that u_t can play the role of log-likelihood between $\mathbb{P}_0$ and $\mathbb{P}_\infty$, we need to assume that e^{u_t} is a martingale with respect to $\mathbb{P}_\infty$. A sufficient condition that can guarantee this fact is, for example, the Novikov condition,

$$\mathbb{E}_\infty\left[\exp\left(\int_0^t \tfrac{1}{2}\alpha_s^2\,ds\right)\right] < \infty \qquad \forall\, t \in [0,\infty), \tag{2.5}$$

or, alternatively,

$$\mathbb{E}_\infty\left[\exp\left(\int_{t_{n-1}}^{t_n} \tfrac{1}{2}\alpha_s^2\,ds\right)\right] < \infty, \tag{2.6}$$

where $\{t_n\}_{n=0}^\infty$ is a strictly increasing sequence of positive real numbers that tends to ∞ [for details, see Karatzas and Shreve (1991), page 198].

If e^{u_t} is a martingale with respect to $\mathbb{P}_\infty$, then Girsanov's theorem applies and we can write

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_\infty}(\mathcal{F}_t) = e^{u_t}, \qquad 0 \le t < \infty, \tag{2.7}$$

or, more generally,

$$\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_t) = e^{u_t - u_\tau} \qquad \text{for } 0 \le \tau \le t < \infty. \tag{2.8}$$

Following Liptser and Shiryayev [(1978), page 225] we impose a final condition on α:

$$\mathbb{P}_0\left[\int_0^\infty \alpha_t^2\,dt = \infty\right] = \mathbb{P}_\infty\left[\int_0^\infty \alpha_t^2\,dt = \infty\right] = 1, \tag{2.9}$$

which, as we will see in the next section, ensures a.s. finiteness of the optimal scheme.

To summarize: the process α is required to satisfy (2.4) and (2.9); moreover, e^{u_t} is assumed to be a martingale with respect to $\mathbb{P}_\infty$, with (2.5) or (2.6) being sufficient conditions that guarantee this property. Let us now present a lemma that will be needed later in our analysis.

LEMMA 1. *Let* (2.4) *and* (2.9) *be valid, and suppose that* $\{e^{u_t}, 0 \le t < \infty\}$ *is a martingale under* $\mathbb{P}_\infty$; *then we have*

$$\mathbb{P}_\tau\left[\int_\tau^\infty \alpha_t^2\,dt = \infty \,\middle|\, \mathcal{F}_\tau\right] = 1, \qquad \mathbb{P}_\tau\text{-a.s.}$$

for each $0 \le \tau < \infty$.

PROOF. From (2.9) and (2.4), it is seen that $\mathbb{P}_0[\int_\tau^\infty \alpha_t^2\,dt = \infty] = 1$ holds for every $\tau \in [0,\infty)$, hence also $\mathbb{P}_\tau[\int_\tau^\infty \alpha_t^2\,dt = \infty] = 1$ since $\mathbb{P}_\tau \ll \mathbb{P}_0$. This suggests that

$$1 = \mathbb{P}_\tau\left[\int_\tau^\infty \alpha_t^2\,dt = \infty\right] = \mathbb{E}_\tau\left[\mathbb{P}_\tau\left[\int_\tau^\infty \alpha_t^2\,dt = \infty \,\middle|\, \mathcal{F}_\tau\right]\right],$$

and leads directly to

$$\mathbb{P}_\tau\Big[\int_\tau^\infty \alpha_t^2\,dt = \infty \Big| \mathcal{F}_\tau\Big] = 1, \qquad \mathbb{P}_\tau\text{-a.s.},$$

which is the desired relation. □

3. Lorden's criterion and proposed modification. Detection of the change time τ is performed with the help of a stopping time T. Lorden (1971) introduced the following *maximal possible conditional delay in issuing the alarm*, as a measure of performance for T,

$$J_{\mathrm{L}}(T) = \sup_{\tau\in[0,\infty)} \operatorname{ess\,sup} \mathbb{E}_\tau[(T-\tau)^+ | \mathcal{F}_\tau], \tag{3.1}$$

and suggested the following min–max problem as a criterion for defining an optimal detection scheme: to minimize $J_{\mathrm{L}}(T)$ of (3.1) over all stopping times T of $\mathcal{F}$ that satisfy the false-alarm constraint

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

Here $\gamma > 0$ is a given constant. In other words, we are interested in the stopping time that has the smallest worst conditional mean detection delay, under the constraint that false alarms should occur with a mean period no smaller than γ.

Proceeding along the same lines as in Liptser and Shiryayev [(1978), page 225] we propose the following alternative performance measure:

$$J(T) = \sup_{\tau\in[0,\infty)} \operatorname{ess\,sup} \mathbb{E}_\tau\Big[\mathbb{1}_{\{T>\tau\}} \int_\tau^T \tfrac{1}{2}\alpha_t^2\,dt \Big| \mathcal{F}_\tau\Big]. \tag{3.2}$$

This gives rise to the min–max optimization problem

$$\inf_{T\in\mathcal{J}_\gamma} J(T) = \inf_{T\in\mathcal{J}_\gamma} \sup_{\tau\in[0,\infty)} \operatorname{ess\,sup} \mathbb{E}_\tau\Big[\mathbb{1}_{\{T>\tau\}} \int_\tau^T \tfrac{1}{2}\alpha_t^2\,dt \Big| \mathcal{F}_\tau\Big], \tag{3.3}$$

where $\mathcal{J}_\gamma$ is the class of $\mathcal{F}$-stopping times T that satisfy the false-alarm constraint

$$\mathbb{E}_\infty\Big[\int_0^T \tfrac{1}{2}\alpha_t^2\,dt\Big] \geq \gamma. \tag{3.4}$$

Clearly, when α is constant the above criterion and optimization problem of (3.2)–(3.4) are equivalent to the original ones defined by Lorden.

We should mention that the proposed modification is motivated by the K–L divergence. Indeed, from (2.8) and by taking (2.3) into account, we conclude that, for $\infty > t \geq \tau \geq 0$, the K–L divergence can be written as

$$\begin{aligned}\mathbb{E}_\tau\Big[\log\Big(\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_t)\Big)\Big|\mathcal{F}_\tau\Big] &= \mathbb{E}_\tau\Big[\int_\tau^t \alpha_s\,dw_s + \int_\tau^t \tfrac{1}{2}\alpha_s^2\,ds \Big|\mathcal{F}_\tau\Big] \\ &= \mathbb{E}_\tau\Big[\int_\tau^t \tfrac{1}{2}\alpha_s^2\,ds \Big|\mathcal{F}_\tau\Big],\end{aligned} \tag{3.5}$$

with equality in (3.5) whenever the displayed quantity is finite.

REMARK. In view of (3.5), one might wonder why not define the performance measure using directly the K–L divergence, that is,

$$J(T) = \sup_{\tau \in [0,\infty)} \operatorname{ess\,sup} \mathbb{E}_\tau \left[\log \left(\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_T) \right) \mathbb{1}_{\{T>\tau\}} \Big| \mathcal{F}_\tau \right], \tag{3.6}$$

instead of the seemingly arbitrary definition of (3.2). Unfortunately, this approach presents certain technical difficulties. First, we need to limit ourselves to stopping times that satisfy $\mathbb{E}_i[\int_0^T \alpha_s^2\, ds] < \infty$, $i = 0, \infty$, in order to assure validity of (3.5). Second, there is a more serious problem coming from Girsanov's theorem: with the usual conditions, the equality $d\mathbb{P}_0/d\mathbb{P}_\infty(\mathcal{F}_t) = e^{u_t}$ is assured *only* for finite t. Consequently, defining our measure as in (3.6) requires limiting even further the class of stopping times to bounded ones. To bypass these two problems, we introduced arbitrarily the measure (3.2), making only a loose connection to the K–L divergence. Let us therefore, with a slight abuse of definition, call the quantities in (3.2) and (3.4) the *K–L detection divergence* and the *K–L false-alarm divergence*, respectively, keeping in mind that there exists a rich class of stopping times for which each of these quantities indeed coincides with the corresponding K–L divergence.

4. The CUSUM process. Let us now introduce the CUSUM process. If m_t denotes the running minimum of u_t, that is,

$$m_t = \inf_{0 \le s \le t} u_s, \qquad 0 \le t < \infty,$$

then the CUSUM process is defined as

$$y_t = u_t - m_t, \qquad 0 \le t < \infty. \tag{4.1}$$

For $\nu \in (0, \infty)$ a given threshold, the CUSUM stopping time with threshold ν is defined as

$$\mathcal{S}_\nu = \inf\{t \ge 0 : y_t \ge \nu\} \tag{4.2}$$

if the indicated set is not empty; otherwise, $\mathcal{S}_\nu = \infty$.

At this point, it is appropriate to introduce certain key properties for the two processes y, m, which are summarized in the following lemma. They are consequences of very standard results in stochastic analysis [see Karatzas and Shreve (1991), pages 149 and 210].

LEMMA 2. *Let m, y be defined as above.*

(i) *The process y is always nonnegative. The process m is nonincreasing and flat off the set $\{y_t = 0\}$; equivalently, if $f(y)$ is a continuous function with $f(0) = 0$, then*

$$\int_0^\infty f(y_t)\, dm_t = 0. \tag{4.3}$$

(ii) *If a function $f(y)$ is twice continuously differentiable, then*

$$df(y_t) = f'(y_t)(du_t - dm_t) + \tfrac{1}{2}\alpha_t^2 f''(y_t)\,dt. \tag{4.4}$$

With the next theorem, we compute the K–L detection and false-alarm divergence for the CUSUM stopping time of (4.2).

THEOREM 1. *The CUSUM stopping time $\mathcal{S}_\nu$ is a.s. finite in the sense that*

$$\mathbb{P}_\tau[\mathcal{S}_\nu = \infty | \mathcal{F}_\tau] = 0, \qquad \mathbb{P}_\tau\text{-a.s.}, \tag{4.5}$$

$$\mathbb{P}_\infty[\mathcal{S}_\nu = \infty | \mathcal{F}_\tau] = 0, \qquad \mathbb{P}_\infty\text{-a.s.} \tag{4.6}$$

For any $0 \le \tau < \infty$, the conditional K–L divergence is given by

$$\mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}} \int_\tau^{\mathcal{S}_\nu} \tfrac{1}{2}\alpha_t^2\,dt \,\middle|\, \mathcal{F}_\tau\right] = [g(\nu) - g(y_\tau)]\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}}, \tag{4.7}$$

$$\mathbb{E}_\infty\left[\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}} \int_\tau^{\mathcal{S}_\nu} \tfrac{1}{2}\alpha_t^2\,dt \,\middle|\, \mathcal{F}_\tau\right] = [h(\nu) - h(y_\tau)]\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}}. \tag{4.8}$$

Here the functions $g(y)$, $h(y)$ are defined as

$$g(y) = y + e^{-y} - 1, \qquad h(y) = e^y - y - 1.$$

They are both strictly increasing and strictly convex on $[0, \infty)$, with $g(0) = h(0) = 0$ and $g(\infty) = h(\infty) = \infty$.

PROOF. Let T_n denote the stopping time

$$T_n = \inf\left\{t \ge \tau : \int_\tau^t \tfrac{1}{2}\alpha_s^2\,ds \ge n\right\}.$$

Because of Lemma 1, T_n is $\mathbb{P}_\tau$-a.s. finite. If $\mathcal{S}_\nu^n$ denotes $\mathcal{S}_\nu^n = \mathcal{S}_\nu \wedge T_n$, then $\mathcal{S}_\nu^n$ is also $\mathbb{P}_\tau$-a.s. finite. Applying Itô's rule to $g(y_t)$ and using the observation $g'(y) + g''(y) = 1$, we can write

$$\begin{aligned}&\mathbb{E}_\tau\left[g\left(y_{\mathcal{S}_\nu^n}\right) - g(y_\tau)\middle|\mathcal{F}_\tau\right]\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}}\\ &\quad= \mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}} \int_\tau^{\mathcal{S}_\nu^n} \tfrac{1}{2}\alpha_t^2\,dt + g'(y_t)\alpha_t\,dw_t - g'(y_t)\,dm_t \,\middle|\, \mathcal{F}_\tau\right].\end{aligned}$$

Furthermore, on $\{\mathcal{S}_\nu^n \ge t\}$ we have $y_t \le \nu$. Consequently,

$$\mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}} \int_\tau^{\mathcal{S}_\nu^n} \tfrac{1}{2}\alpha_t^2 g'(y_t)^2\,dt \,\middle|\, \mathcal{F}_\tau\right] \le \left(g'(\nu)\right)^2 n < \infty,$$

suggesting that the expectation of the stochastic integral is 0. On the other hand, we have $g'(0) = 0$ and thus $\int_0^\infty g'(y_t)\,dm_t = 0$ from (4.3). Thus, we end up with

$$\mathbb{E}_\tau\left[g\left(y_{\mathcal{S}_\nu^n}\right) - g(y_\tau)\middle|\mathcal{F}_\tau\right]\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}} = \mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}} \int_\tau^{\mathcal{S}_\nu^n} \tfrac{1}{2}\alpha_t^2\,dt \,\middle|\, \mathcal{F}_\tau\right]. \tag{4.9}$$

Now $y_{\mathcal{S}_\nu^n} \leq \nu$ and $g(\cdot)$ is increasing. Therefore,

$$g(\nu) = g(\nu) - g(0) \geq \mathbb{E}_\tau\big[g\big(y_{\mathcal{S}_\nu^n}\big) - g(y_\tau)|\mathcal{F}_\tau\big]\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}}$$

$$= \mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu^n > \tau\}} \int_\tau^{\mathcal{S}_\nu^n} \tfrac{1}{2}\alpha_t^2\, dt \Big| \mathcal{F}_\tau\right].$$

Because of Lemma 1, as n tends to ∞ T_n tends to ∞ as well and $\mathcal{S}_\nu^n$ tends to $\mathcal{S}_\nu$. This yields

$$g(\nu) \geq \mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}} \int_\tau^{\mathcal{S}_\nu} \tfrac{1}{2}\alpha_t^2\, dt \Big| \mathcal{F}_\tau\right] \geq \mathbb{E}_\tau\left[\mathbb{1}_{\{\mathcal{S}_\nu = \infty\}} \int_\tau^{\infty} \tfrac{1}{2}\alpha_t^2\, dt \Big| \mathcal{F}_\tau\right].$$

Using again Lemma 1, we conclude that $\mathbb{P}_\tau[\mathcal{S}_\nu = \infty|\mathcal{F}_\tau] = 0$, $\mathbb{P}_\tau$-a.s., which is (4.5).

If we now return to (4.9), let $n \to \infty$, use monotone convergence on the right-hand side and bounded convergence on the left and use (4.5), we can prove (4.7). Following similar steps we can show (4.6) and (4.8). $\square$

We have the following two corollaries of Theorem 1.

COROLLARY 1. *Let T be a stopping time and $\mathcal{S}_\nu$ the CUSUM stopping time with threshold ν. If $T_\nu = T \wedge \mathcal{S}_\nu$, then*

$$\mathbb{E}_\tau\left[\mathbb{1}_{\{T_\nu > \tau\}} \int_\tau^{T_\nu} \tfrac{1}{2}\alpha_t^2\, dt \Big| \mathcal{F}_\tau\right] = \mathbb{E}_\tau\big[g\big(y_{T_\nu}\big) - g(y_\tau)|\mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}}, \tag{4.10}$$

$$\mathbb{E}_\infty\left[\mathbb{1}_{\{T_\nu > \tau\}} \int_\tau^{T_\nu} \tfrac{1}{2}\alpha_t^2\, dt \Big| \mathcal{F}_\tau\right] = \mathbb{E}_\infty\big[h\big(y_{T_\nu}\big) - h(y_\tau)|\mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}}. \tag{4.11}$$

PROOF. The proof follows by another application of Itô's rule. Expectation of the stochastic integral is 0, because for $0 \leq t \leq T_\nu \leq \mathcal{S}_\nu$ we have $0 \leq y_t \leq \nu$; therefore, $g'(y_t)$ and $h'(y_t)$ are again bounded, and from Theorem 1 we have $\mathbb{E}_i[\int_0^{T_\nu} \alpha_t^2\, dt|\mathcal{F}_\tau] < \infty$. Finally, the Stieltjes integral involving dm_t is again 0, since $g'(0) = h'(0) = 0$. $\square$

COROLLARY 2. *Let T be a stopping time and $T_\nu = T \wedge \mathcal{S}_\nu$. If the function $f(y)$ is continuous and bounded for $0 \leq y \leq \nu$, then*

$$\mathbb{E}_\tau\big[f\big(y_{T_\nu}\big)|\mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}} = \mathbb{E}_\infty\big[e^{u_{T_\nu} - u_\tau} f\big(y_{T_\nu}\big)|\mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}}, \qquad \mathbb{P}_\infty\text{-a.s.} \tag{4.12}$$

PROOF. It should be noted that (4.12) is not obvious because Girsanov's theorem is valid only for bounded stopping times. Let $M > 0$. Then on $\{T_\nu > \tau\}$

we can write

$$\begin{aligned}\mathbb{E}_\tau\big[f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big] &= \mathbb{E}_\tau\big[\mathbb{1}_{\{T_\nu\le M\}}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big] + \mathbb{E}_\tau\big[\mathbb{1}_{\{T_\nu>M\}}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big]\\ &= \mathbb{E}_\infty\big[\mathbb{1}_{\{T_\nu\le M\}}e^{u_{T_\nu}-u_\tau}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big] + \mathbb{E}_\tau\big[\mathbb{1}_{\{T_\nu>M\}}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big]\\ &= \mathbb{E}_\infty\big[e^{u_{T_\nu}-u_\tau}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big] - \mathbb{E}_\infty\big[\mathbb{1}_{\{T_\nu>M\}}e^{u_{T_\nu}-u_\tau}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big]\\ &\quad + \mathbb{E}_\tau\big[\mathbb{1}_{\{T_\nu>M\}}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big].\end{aligned}$$

Notice now that on $\{T_\nu > \tau\}$ we have $u_{T_\nu} - u_\tau \le u_{T_\nu} - m_{T_\nu} = y_{T_\nu} \le \nu$. Therefore, we obtain the following bounds for the last two terms:

$$\begin{aligned}\big|\mathbb{E}_\tau\big[\mathbb{1}_{\{T_\nu>M\}}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big]\big| &\le \max_{0\le y\le\nu}|f(y)|\mathbb{P}_\tau[T_\nu > M|\mathcal{F}_\tau]\\ &\le \max_{0\le y\le\nu}|f(y)|\mathbb{P}_\tau[\mathcal{S}_\nu > M|\mathcal{F}_\tau],\\ \big|\mathbb{E}_\infty\big[\mathbb{1}_{\{T_\nu>M\}}e^{u_{T_\nu}-u_\tau}f\big(y_{T_\nu}\big)\big|\mathcal{F}_\tau\big]\big| &\le e^\nu\max_{0\le y\le\nu}|f(y)|\mathbb{P}_\infty[T_\nu > M|\mathcal{F}_\tau]\\ &\le e^\nu\max_{0\le y\le\nu}|f(y)|\mathbb{P}_\infty[\mathcal{S}_\nu > M|\mathcal{F}_\tau].\end{aligned}$$

Both bounds, because of Theorem 1, tend to 0 as $M \to \infty$. This concludes the proof. □

Using Theorem 1, the K–L false-alarm divergence of $\mathcal{S}_\nu$ satisfies

$$\mathbb{E}_\infty\bigg[\int_0^{\mathcal{S}_\nu}\tfrac{1}{2}\alpha_t^2\,dt\bigg] = h(\nu) - h(0) = h(\nu).$$

Let $\nu_\star$ be the threshold for which the corresponding CUSUM stopping time satisfies the false-alarm constraint (3.4) with equality, that is,

$$\mathbb{E}_\infty\bigg[\int_0^{\mathcal{S}_{\nu_\star}}\tfrac{1}{2}\alpha_t^2\,dt\bigg] = h(\nu_\star) = e^{\nu_\star} - \nu_\star - 1 = \gamma. \tag{4.13}$$

For every γ there is a unique $\nu_\star$ satisfying (4.13). The worst K–L detection divergence of $S_{\nu_\star}$ can be obtained from Theorem 1 using the increase of $g(\cdot)$. Specifically,

$$J(\mathcal{S}_{\nu_\star}) = \sup_{\tau\in[0,\infty)}\operatorname{ess\,sup}\{g(\nu_\star) - g(y_\tau)\} = g(\nu_\star) - g(0) = g(\nu_\star) = \nu_\star + e^{-\nu_\star} - 1.$$

It is the goal of the next section to show that the CUSUM stopping time with threshold $\nu_\star$ is, in fact, the one that solves the min–max optimization problem defined by (3.3) and (3.4).

5. Optimality of the CUSUM stopping time. To prove the optimality of $\mathcal{S}_{\nu_\star}$, it is sufficient to show that for any stopping time T satisfying the false-alarm constraint (3.4) we have $J(T) \geq g(\nu_\star)$. We will show this fact following similar steps as in Moustakides (1986). We first obtain a convenient lower bound for $J(T)$.

THEOREM 2. *Let T be a stopping time, let $\mathcal{S}_\nu$ be the CUSUM stopping time with threshold ν and define $T_\nu = T \wedge \mathcal{S}_\nu$. Then*

$$J(T) \geq \frac{\mathbb{E}_\infty[e^{y_{T_\nu}} g(y_{T_\nu})]}{\mathbb{E}_\infty[e^{y_{T_\nu}}]}.$$

PROOF. Since $T \geq T_\nu$ we have

$$J(T) \geq J(T_\nu) \geq \mathbb{E}_\tau\left[\mathbb{1}_{\{T_\nu > \tau\}} \int_\tau^{T_\nu} \tfrac{1}{2}\alpha_t^2\, dt \,\middle|\, \mathcal{F}_\tau\right], \tag{5.1}$$

$$J(T) \geq J(T_\nu) \geq \mathbb{E}_0\left[\int_0^{T_\nu} \tfrac{1}{2}\alpha_t^2\, dt\right], \tag{5.2}$$

for any $0 \leq \tau < \infty$, owing to (3.2). Applying Corollary 1 on the right-hand side and Corollary 2 on both sides of (5.1), we obtain

$$J(T)\mathbb{E}_\infty\big[e^{u_{T_\nu} - u_\tau} | \mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}} \geq \mathbb{E}_\infty\big[e^{u_{T_\nu} - u_\tau}\big[g\big(y_{T_\nu}\big) - g(y_\tau)\big]\big|\mathcal{F}_\tau\big]\mathbb{1}_{\{T_\nu > \tau\}}.$$

Integrating both sides with $-dm_\tau$ and recalling that m_t is nonincreasing, then taking expectation with respect to $\mathbb{P}_\infty$, yields

$$J(T)\mathbb{E}_\infty\left[\int_0^{T_\nu} e^{u_{T_\nu} - u_\tau}(-dm_\tau)\right] \geq \mathbb{E}_\infty\left[\int_0^{T_\nu} e^{u_{T_\nu} - u_\tau}\big[g\big(y_{T_\nu}\big) - g(y_\tau)\big](-dm_\tau)\right].$$

Using from Lemma 2 the fact that the process m is flat off the set $\{\tau \geq 0 : y_\tau = 0\} = \{\tau \geq 0 : u_\tau = m_\tau\}$ and also that $g(0) = 0$, we can write the previous relation as

$$J(T)\mathbb{E}_\infty\left[\int_0^{T_\nu} e^{u_{T_\nu} - m_\tau}(-dm_\tau)\right] \geq \mathbb{E}_\infty\left[\int_0^{T_\nu} e^{u_{T_\nu} - m_\tau} g\big(y_{T_\nu}\big)(-dm_\tau)\right],$$

which leads to

$$J(T)\mathbb{E}_\infty\big[e^{y_{T_\nu}} - e^{u_{T_\nu}}\big] \geq \mathbb{E}_\infty\big[\big(e^{y_{T_\nu}} - e^{u_{T_\nu}}\big)g\big(y_{T_\nu}\big)\big]. \tag{5.3}$$

Focusing now on (5.2), recalling that $\mathcal{F}_0$ is the trivial σ-algebra, using Corollaries 1 and 2, and that $y_0 = 0$, we end up with

$$J(T)\mathbb{E}_\infty\big[e^{u_{T_\nu}}\big] \geq \mathbb{E}_\infty\big[e^{u_{T_\nu}} g\big(y_{T_\nu}\big)\big].$$

By adding this relation, term by term, to (5.3), we obtain

$$J(T)\mathbb{E}_\infty\big[e^{y_{T_\nu}}\big] \geq \mathbb{E}_\infty\big[e^{y_{T_\nu}} g\big(y_{T_\nu}\big)\big].$$

Finally, since $e^{\nu} \geq e^{y_{T_\nu}} \geq 1$, we conclude that

$$J(T) \geq \frac{\mathbb{E}_\infty[e^{y_{T_\nu}} g(y_{T_\nu})]}{\mathbb{E}_\infty[e^{y_{T_\nu}}]},$$

which proves the theorem. □

At this point, we need the following technical lemma.

LEMMA 3. *Let T be a stopping time, let S_ν be the CUSUM stopping time with threshold ν, let $T_\nu = T \wedge S_\nu$ and define the function $\psi_T(\nu) = \mathbb{E}_\infty[\int_0^{T_\nu} \frac{1}{2}\alpha_t^2\, dt]$. Then $\psi_T(\nu)$ is continuous and increasing in ν with $\psi_T(0) = 0$ and $\psi_T(\infty) = \mathbb{E}_\infty[\int_0^{T} \frac{1}{2}\alpha_t^2\, dt]$.*

PROOF. Since for $\nu < \mu$ we have $S_\nu \leq S_\mu$, we conclude that $\psi_T(\nu)$ is increasing in ν. By observing that $S_0 = 0$ and $S_\infty = \infty$, we can verify the correctness of the two values $\psi_T(0)$ and $\psi_T(\infty)$. To show continuity, let $\nu < \mu$ and consider the difference

$$\begin{aligned}
\psi_T(\mu) - \psi_T(\nu) &= \mathbb{E}_\infty\left[\int_{T_\nu}^{T_\mu} \tfrac{1}{2}\alpha_t^2\, dt\right] \\
&= \mathbb{E}_\infty\left[\mathbb{1}_{\{T > S_\nu\}} \int_{T_\nu}^{T_\mu} \tfrac{1}{2}\alpha_t^2\, dt\right] + \mathbb{E}_\infty\left[\mathbb{1}_{\{T \leq S_\nu\}} \int_{T_\nu}^{T_\mu} \tfrac{1}{2}\alpha_t^2\, dt\right] \\
&= \mathbb{E}_\infty\left[\mathbb{1}_{\{T > S_\nu\}} \int_{T_\nu}^{T_\mu} \tfrac{1}{2}\alpha_t^2\, dt\right] \leq \mathbb{E}_\infty\left[\int_{S_\nu}^{S_\mu} \tfrac{1}{2}\alpha_t^2\, dt\right] \\
&= h(\mu) - h(\nu),
\end{aligned}$$

where we have used the property that for $\nu < \mu$ we have $\mathcal{S}_\nu \leq \mathcal{S}_\mu$. Therefore, on the set $\{T \leq S_\nu\}$ we have that $T = T_\nu = T_\mu$, whereas on $\{T > S_\nu\}$ we have that $S_\nu = T_\nu \leq T_\mu \leq S_\mu$. Continuity of $\psi_T(\nu)$ is a consequence of the continuity of $h(\nu)$. □

We are now in a position to show the optimality of CUSUM. We first observe that we can limit ourselves to stopping times that satisfy the false-alarm constraint (3.4) with equality. Indeed, if a stopping time T has $\mathbb{E}_\infty[\int_0^T \frac{1}{2}\alpha_t^2\, dt] > \gamma$, then from Lemma 3 we conclude that we can select a threshold ν such that the stopping time $T_\nu = T \wedge \mathcal{S}_\nu$ satisfies (3.4) with equality. Since $T \geq T_\nu$, this yields $J(T) \geq J(T_\nu)$, which suggests that T_ν is preferable to T.

THEOREM 3. *Any stopping time T that satisfies the false-alarm constraint* (3.4) *with equality has a K–L detection divergence $J(T)$ that is no less than $g(\nu_\star)$.*

PROOF. Based on Theorem 2, it is sufficient to show that for every $\varepsilon > 0$ we can find a threshold ν_ε such that $T_{\nu_\varepsilon} = T \wedge \mathcal{S}_{\nu_\varepsilon}$ satisfies

$$\frac{\mathbb{E}_\infty[e^{y_{T_{\nu_\varepsilon}}} g(y_{T_{\nu_\varepsilon}})]}{\mathbb{E}_\infty[e^{y_{T_{\nu_\varepsilon}}}]} \geq g(\nu_\star) - \varepsilon. \tag{5.4}$$

To prove (5.4), let T be a stopping time satisfying the false-alarm constraint with equality and consider any $\varepsilon > 0$. Then, because of Lemma 3, we can select a sufficiently large threshold ν_ε such that

$$\gamma \geq \mathbb{E}_\infty\left[\int_0^{T_{\nu_\varepsilon}} \tfrac{1}{2}\alpha_t^2\, dt\right] \geq \gamma - \varepsilon.$$

From Corollary 1, we have $\mathbb{E}_\infty[\int_0^{T_{\nu_\varepsilon}} \frac{1}{2}\alpha_t^2\, dt] = \mathbb{E}_\infty[h(y_{T_{\nu_\varepsilon}})]$, which suggests that

$$\mathbb{E}_\infty\big[h\big(y_{T_{\nu_\varepsilon}}\big)\big] \geq \gamma - \varepsilon. \tag{5.5}$$

Let $U(T_{\nu_\varepsilon})$ be the following expression:

$$U(T_{\nu_\varepsilon}) = \mathbb{E}_\infty\big[e^{y_{T_{\nu_\varepsilon}}}\big[g\big(y_{T_{\nu_\varepsilon}}\big) - g(\nu_\star)\big] - h\big(y_{T_{\nu_\varepsilon}}\big) + h(\nu_\star)\big]. \tag{5.6}$$

If we define the function $p(y) = e^y[g(y) - g(\nu_\star)] - h(y) + h(\nu_\star)$, we can then verify that its derivative satisfies $p'(y) = e^y[g(y) - g(\nu_\star)]$. Due to the strict increase in $g(y)$, this suggests that $p'(y)$ has the same sign as $y - \nu_\star$, and therefore $p(y)$ has a minimum at $y = \nu_\star$. Since $p(\nu_\star) = 0$ we conclude that $p(y) \geq 0$. Consequently, we also have $U(T_{\nu_\varepsilon}) \geq 0$. Using this fact in (5.6) along with (5.5) and recalling from (4.13) that $h(\nu_\star) = \gamma$ yields

$$\begin{aligned}\mathbb{E}_\infty\big[e^{y_{T_{\nu_\varepsilon}}} g\big(y_{T_{\nu_\varepsilon}}\big)\big] &\geq g(\nu_\star)\mathbb{E}_\infty\big[e^{y_{T_{\nu_\varepsilon}}}\big] + \mathbb{E}_\infty\big[h\big(y_{T_{\nu_\varepsilon}}\big)\big] - h(\nu_\star)\\ &\geq g(\nu_\star)\mathbb{E}_\infty\big[e^{y_{T_{\nu_\varepsilon}}}\big] - \varepsilon \geq [g(\nu_\star) - \varepsilon]\mathbb{E}_\infty\big[e^{y_{T_{\nu_\varepsilon}}}\big],\end{aligned}$$

where the last inequality holds because $e^y \geq 1$. This proves the theorem and establishes optimality of the CUSUM stopping time. □

6. Discussion and examples. A key property for the validity of our result is (2.9). In fact, this condition imposes a form of persistency in the difference between the statistics of the two hypotheses, thus ensuring the a.s. finiteness of the optimal stopping time. There are, of course, situations where (2.9) does not hold, as, for example, in transient changes where the process returns to its nominal statistics after a finite time. For such cases, CUSUM is not necessarily optimal since the previous analysis is no longer valid (it is Theorem 1 that fails).

Extension of our result to multidimensional processes is straightforward. In particular, if the observation is a vector Itô process Ξ of the form

$$d\Xi_t = \begin{cases} A_t\, dt + \boldsymbol{\Sigma}_t\, dW_t, & 0 \leq t \leq \tau, \\ B_t\, dt + \boldsymbol{\Sigma}_t\, dW_t, & \tau < t < \infty, \end{cases}$$

where W is a vector Brownian motion, A and B are adapted vector processes and $\mathbf{\Sigma}$ is an adapted matrix process, then our previous analysis goes through without significant modifications. The log-likelihood ratio in this case satisfies

$$du_t = \mathcal{A}_t^T \, d\Xi_t - \tfrac{1}{2}\mathcal{A}_t^T \mathcal{A}_t \, dt = \mathcal{A}_t^T \, dW_t \pm \tfrac{1}{2}\mathcal{A}_t^T \mathcal{A}_t \, dt,$$

where the superscript T denotes transpose, the $-$ sign corresponds to the case before the change, the $+$ after, and the process $\mathcal{A}$ is defined as

$$\mathcal{A}_t = \mathbf{\Sigma}_t^{-1}(B_t - A_t).$$

Here the quantity $\mathcal{A}_t^T \mathcal{A}_t$ plays the role of α_t^2, and as was mentioned above all results go through without major difficulty.

What is interesting to note in this more general setting is the fact that when $\mathcal{A}_t^T \mathcal{A}_t$ is equal to a constant, the modified criterion is equivalent to the original Lorden criterion. In other words, CUSUM is optimal in the original Lorden sense not only for detecting changes in the constant drift of a Brownian motion but also changes in which the process $\mathcal{A}_t^T \mathcal{A}_t$ is constant.

One might also consider the possibility of extending our result to the discrete-time case. In particular, we are interested in observations ξ_n that satisfy the model

$$\xi_n = \alpha_{n-1} \mathbb{1}_{\{n>\tau\}} + w_n,$$

with α adapted to the filtration $\mathcal{F}$, where $\mathcal{F}_n = \sigma\{\xi_k; 0 \le k \le n\}$ and w is an i.i.d. standard, zero-mean, unit-variance, Gaussian process. Unfortunately, this generalization cannot be obtained by following a similar approach as in the continuous-time case. The main bottleneck comes from Theorem 1; specifically, it is not clear whether the two conditional divergences of the CUSUM test depend *only* on y_τ and, furthermore, whether expressions similar to Corollary 1 hold for any other stopping time.

Let us now present two examples that fall into our class of processes (2.1). Consider the case where $\alpha_t = \alpha(t)$, with $\alpha(t)$ a deterministic function of time. This is the problem considered in Tartakovski (1995), where one is interested in detecting changes in nonhomogeneous Gaussian processes. If for every finite $t \ge 0$ we have $\int_0^t \alpha(s)^2 \, ds < \infty$ and $\lim_{t\to\infty} \int_0^t \alpha(s)^2 \, ds = \infty$, then (2.4), (2.5) and (2.9) are satisfied and therefore CUSUM is optimal in the proposed generalized sense.

A more interesting situation occurs when $\alpha_t = -\alpha\xi_t$, where α is a positive constant. This corresponds to a standard Brownian motion without drift under nominal conditions and to an Ornstein–Uhlenbeck process under change. Notice that, under $\mathbb{P}_\infty$, $\xi_t = \xi_0 + w_t$ is Gaussian with mean ξ_0 and variance equal to t, whereas under $\mathbb{P}_0$, $\xi_t = \xi_0 e^{-\alpha t} + \int_0^t e^{-\alpha(t-s)} \, dw_s$ is Gaussian with mean $\xi_0 e^{-\alpha t}$ and variance $(1 - e^{-2\alpha t})/2\alpha$.

For (2.4) to be true, it is sufficient to have $\mathbb{E}_i[\int_0^t \xi_s^2 \, ds] < \infty$, $i = 0, \infty$, which can be directly verified.

To show that e^{u_t} is a martingale, Corollary 5.16 from Karatzas and Shreve (1991), page 200, applies showing validity of (2.6).

For (2.9), to show first $\mathbb{P}_\infty[\int_0^\infty \xi_t^2\,dt=\infty]=1$, we observe, using the Schwarz inequality, that $\int_0^t \xi_s^2\,ds \geq (\int_0^t \xi_s\,ds)^2/t$. If we call $z_t=\int_0^t \xi_s\,ds/\sqrt{t}$, then z_t is Gaussian with mean $\mu_t=c_1\sqrt{t}$ and variance $\sigma_t^2=c_2t^2$, where c_1 and c_2 are constants. If $M>0$, we can then write

$$\mathbb{P}_\infty\Big[\int_0^\infty \xi_s^2\,ds\leq M\Big]\leq \mathbb{P}_\infty\Big[\int_0^t \xi_s^2\,ds\leq M\Big]\leq \mathbb{P}_\infty\big[|z_t|\leq\sqrt{M}\,\big]$$
$$\leq \Phi\Big(\frac{\sqrt{M}-\mu_t}{\sigma_t}\Big)-\Phi\Big(-\frac{\sqrt{M}+\mu_t}{\sigma_t}\Big),$$

where $\Phi(z)$ is the standard Gaussian cumulative distribution. The last term tends to 0 as t tends to ∞. For a different proof, see Problem 6.30 of Karatzas and Shreve [(1991), page 217].

To prove $\mathbb{P}_0[\int_0^\infty \xi_t^2\,dt=\infty]=1$, we use Itô's rule and conclude

$$z_t=\int_0^t \xi_s^2\,ds=\Big(t-\xi_t^2+\xi_0^2+2\int_0^t \xi_s\,dw_s\Big)\Big/2\alpha.$$

For the process z_t, we can then show that its expected value is of the form $\mu_t=c_1t+o(t)$ and its variance $\sigma_t^2=c_2t+o(t)$, with c_1 and c_2 positive constants. We can now use Chebyshev's inequality and for any $M>0$ and sufficiently large t (such that $\mu_t>M$) we can write

$$\mathbb{P}_0\Big[\int_0^\infty \xi_s^2\,ds\leq M\Big]\leq \mathbb{P}_0[z_t\leq M]=\mathbb{P}_0\Big[\Big(\frac{\mu_t-z_t}{\sigma_t}\Big)^2\geq\Big(\frac{\mu_t-M}{\sigma_t}\Big)^2\Big]$$
$$\leq\Big(\frac{\sigma_t}{\mu_t-M}\Big)^2.$$

The right-hand side term in the last inequality can be seen to tend to 0 as t tends to ∞. Therefore, our optimality result applies to this case as well.

Acknowledgment. The author is grateful to Professor I. Karatzas for the meticulous reading of the manuscript and for his valuable remarks.

REFERENCES

BEIBEL, M. (1996). A note on Ritov's Bayes approach to the minmax property of the CUSUM procedure. *Ann. Statist.* **24** 1804–1812.

IRLE, A. (1984). Extended optimality of sequential probability ratio tests. *Ann. Statist.* **12** 380–386.

KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer, New York.

LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes*. **2**. *Applications*. Springer, New York.

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

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

MOUSTAKIDES, G. V. (1998). Quickest detection of abrupt changes for a class of random processes. *IEEE Trans. Inform. Theory* **44** 1965–1968.

ØKSENDAL, B. K. (1998). *Stochastic Differential Equations*: *An Introduction with Applications*, 5th ed. Springer, Berlin.

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

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

RITOV, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18** 1464–1469.

SHIRYAYEV, A. N. (1978). *Optimal Stopping Rules*. Springer, New York.

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

TARTAKOVSKI, A. (1995). Asymptotic properties of CUSUM and Shiryayev's procedures for detecting a change in a nonhomogeneous Gaussian process. *Math. Methods Statist.* **4** 389–404.

WALD, A. (1947). *Sequential Analysis*. Wiley, New York.

WALD, A. and WOLFOWITZ, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19** 326–339.

YASHIN, A. (1983). On a problem of sequential hypothesis testing. *Theory Probab. Appl.* **28** 157–165.

INRIA–IRISA
CAMPUS DE BEAULIEU
35042 RENNES CEDEX
FRANCE
E-MAIL: George.Moustakides@inria.fr

DEPARTMENT OF COMPUTER AND
COMMUNICATION ENGINEERING
UNIVERSITY OF THESSALY
382 21 VOLOS
GREECE
E-MAIL: moustaki@inf-uth.gr