INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Optimality of the CUSUM Procedure

George V. Moustakides

N° 4541

September 2002

THÈME 1

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4541--FR+ENG

Optimality of the CUSUM Procedure

George V. Moustakides

Thème 1 — Réseaux et systèmes
Projet IRIS

Rapport de recherche n° 4541 — September 2002 — 14 pages

Abstract: 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.

Key-words: CUSUM, change-point, disorder problem, Kullback-Leibler divergence, sequential detection.

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu, 35042 Rennes Cedex (France)
Téléphone : +33 2 99 84 71 00 — Télécopie : +33 2 99 84 71 71

Optimalité de la Procédure CUSUM

Résumé : On considère l'optimalité de CUSUM avec un critère du type Lorden. On démontre l'optimalité du test CUSUM pour les processus d'Itô dans un cadre similaire de Lorden, mais avec un critère qui remplace les délais moyens avec la divergence du Kullback-Leibler correspondante.

Mots-clés : CUSUM, détection des ruptures, divergence du Kullback-Leibler, détection séquentielle.

Contents

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. 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), that yielded also 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 optimum 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), Irle (1984) and Galtchouk (2001) 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 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 provides also 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

$$d\xi_t = \begin{cases} dw_t, & 0 \le t \le \tau \\ \alpha_t dt + dw_t, & \tau < t, \end{cases} \tag{1}$$

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 régime, 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 (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 $\mathbb{P}_\infty$ 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 (1) is well defined. Following Øksendal (2000), page 44, we require the process α to be $\mathcal{F}$ adapted and to satisfy

$$\mathbb{P}_0\left[\int_0^t |\alpha_s| ds < \infty\right] = 1, \quad \forall\, t \in [0,\infty). \tag{2}$$

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

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

which, because of (1), satisfies

$$du_t = \begin{cases} \alpha_t dw_t - \frac{1}{2}\alpha_t^2 dt, & 0 \le t \le \tau \\ \alpha_t dw_t + \frac{1}{2}\alpha_t^2 dt, & \tau < t, \;\; 0 \le t < \infty \end{cases} \tag{3}$$

In order that this process be well defined under both hypotheses, again from Øksendal (2000), page 44, we need to assume that for every $0 \le t < \infty$ we have

$$\mathbb{P}_0\left[\int_0^t \alpha_s^2 ds < \infty\right] = \mathbb{P}_\infty\left[\int_0^t \alpha_s^2 ds < \infty\right] = 1. \tag{4}$$

Since $(\int_0^t |\alpha_s| ds)^2 \le t \int_0^t \alpha_s^2 ds$, it is clear that (4) also implies (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 necessary condition that can guarantee this fact is, for example, the Novikov condition:

$$\mathbb{E}_\infty\left[\exp\left(\int_0^t \frac{1}{2}\alpha_s^2 ds\right)\right] < \infty,\ \ \forall\, t \in [0,\infty), \tag{5}$$

or alternatively

$$\mathbb{E}_\infty\left[\exp\left(\int_{t_{n-1}}^{t_n} \frac{1}{2}\alpha_s^2 ds\right)\right] < \infty, \tag{6}$$

where $\{t_n\}_{n=0}^\infty$ is a strictly increasing sequence of positive real numbers that tends to infinity (for details, see Karatzas and Shreve (1988), page 198).

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

$$\frac{d\mathbb{P}_0}{d\mathbb{P}_\infty}(\mathcal{F}_t) = e^{u_t},\ \ 0 \le t < \infty, \tag{7}$$

or more generally,

$$\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_t) = e^{u_t - u_\tau},\ \ \text{for } 0 \le \tau \le t < \infty. \tag{8}$$

Following Liptser and Shiryayev (1978), page 224, we impose a final condition on α

$$\mathbb{P}_0\left[\int_0^\infty \alpha_t^2 dt = \infty\right] = \mathbb{P}_\infty\left[\int_0^\infty \alpha_t^2 dt = \infty\right] = 1, \tag{9}$$

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 (4) and (9); moreover, e^{u_t} is assumed to be a martingale with respect to $\mathbb{P}_\infty$, with (5) or (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 (4), (9) be valid, suppose that $\{e^{u_t},\ 0 \le t < \infty\}$ is a martingale; then we have*

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

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

Proof of Lemma 1: From (9) and (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

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

and leads directly to

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

■

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,

$$J_L(T) = \sup_{\tau\in[0,\infty)} \operatorname{essup} \mathbb{E}_\tau \left[(T-\tau)^+ | \mathcal{F}_\tau\right], \tag{10}$$

and suggested the following min-max problem as a criterion for defining an optimal detection scheme: to minimize $J_L(T)$ of (10) 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 224, we propose the following alternative performance measure

$$J(T) = \sup_{\tau\in[0,\infty)} \operatorname{essup} \mathbb{E}_\tau \left[\int_\tau^T \frac{1}{2}\alpha_t^2 dt \,\middle|\, \mathcal{F}_\tau\right], \tag{11}$$

where integrals are considered zero whenever the upper limit is smaller than the lower. This gives rise to the min-max optimization problem

$$\inf_{T\in\mathcal{J}_\gamma} J(T) = \inf_{T\in\mathcal{J}_\gamma} \sup_{\tau\in[0,\infty)} \operatorname{essup} \mathbb{E}_\tau \left[\int_\tau^T \frac{1}{2}\alpha_t^2 dt \,\middle|\, \mathcal{F}_\tau\right] \tag{12}$$

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

$$\mathbb{E}_\infty\left[\int_0^T \frac{1}{2}\alpha_t^2 dt\right] \geq \gamma. \tag{13}$$

Clearly, when α is constant the above criterion and optimization problem of (11)-(13) 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 (8) and by taking (3) into account, we conclude that the K-L divergence can be written as

$$\begin{aligned}
\mathbb{E}_\tau\left[\log\left(\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_t)\right) \,\middle|\, \mathcal{F}_\tau\right] &= \mathbb{E}_\tau\left[\int_\tau^t \alpha_s dw_s + \int_\tau^t \frac{1}{2}\alpha_s^2 ds \,\middle|\, \mathcal{F}_\tau\right] \\
&= \mathbb{E}_\tau\left[\int_\tau^t \frac{1}{2}\alpha_s^2 ds \,\middle|\, \mathcal{F}_\tau\right], \text{ for } 0 \leq \tau \leq t < \infty,
\end{aligned} \tag{14}$$

with equality in (14) whenever the displayed quantity is finite.

Remark: In view of (14), one might wonder why not define the performance measure using directly the K-L divergence, that is,

$$J(T) = \sup_{\tau\in[0,\infty)} \operatorname{essup} \mathbb{E}_\tau \left[\log\left(\frac{d\mathbb{P}_\tau}{d\mathbb{P}_\infty}(\mathcal{F}_T) \right) \mathbb{1}_{\{T>\tau\}} \,\middle|\, \mathcal{F}_\tau \right], \tag{15}$$

instead of the seemingly arbitrary definition of (11). 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 (14). Secondly, 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 (15) requires to limit even further the class of stopping times to bounded ones. In order to bypass these two problems, we introduced arbitrarily the measure (11), making only a loose connection to the K-L divergence. Let us therefore, with a slight abuse of definition, call the quantities in (11) and (13) 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,

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

then the CUSUM process is defined as

$$y_t = u_t - m_t, \quad 0 \le t < \infty. \tag{16}$$

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

$$\mathcal{S}_\nu = \inf\{t \ge 0 : \ y_t \ge \nu\}, \tag{17}$$

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, that are summarized in the following lemma. They are consequences of very standard results in stochastic analysis (see Karatzas and Shreve (1998), 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*

$$\int_0^\infty f(y_t)dm_t = 0. \tag{18}$$

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

$$df(y_t) = f'(y_t)(du_t - dm_t) + \frac{1}{2}\alpha_t^2 f''(y_t)dt. \tag{19}$$

With the next theorem we compute the K-L detection and false alarm divergence for the CUSUM stopping time of (17).

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

$$\mathbb{P}_\tau[\mathcal{S}_\nu = \infty | \mathcal{F}_\tau] = 0, \;\; \mathbb{P}_\tau\text{-a.s.} \tag{20}$$

$$\mathbb{P}_\infty[\mathcal{S}_\nu = \infty | \mathcal{F}_\tau] = 0, \;\; \mathbb{P}_\infty\text{-a.s.} \tag{21}$$

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

$$\mathbb{E}_\tau\left[\int_\tau^{\mathcal{S}_\nu} \frac{1}{2}\alpha_t^2 dt \,\middle|\, \mathcal{F}_\tau\right] = [g(\nu) - g(y_\tau)]\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}} \tag{22}$$

$$\mathbb{E}_\infty\left[\int_\tau^{\mathcal{S}_\nu} \frac{1}{2}\alpha_t^2 dt \,\middle|\, \mathcal{F}_\tau\right] = [h(\nu) - h(y_\tau)]\mathbb{1}_{\{\mathcal{S}_\nu > \tau\}}. \tag{23}$$

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

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

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

Proof of Theorem 1: Let T_n denote the stopping time

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

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 obtain

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

On $\{\mathcal{S}_\nu^n \geq t\}$ we have $y_t \leq \nu$, consequently

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

suggesting that the expectation of the stochastic integral is zero. On the other hand we have $g'(0) = 0$ and thus $\int_0^\infty g'(y_t) dm_t = 0$ from (18). We end up with

$$\mathbb{E}_\tau[g(y_{\mathcal{S}_\nu^n}) - g(y_\tau)|\mathcal{F}_\tau] = \mathbb{E}_\tau\left[\int_\tau^{\mathcal{S}_\nu^n} \frac{1}{2}\alpha_t^2 dt \,\bigg|\, \mathcal{F}_\tau\right]. \tag{24}$$

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

$$g(\nu) = g(\nu) - g(0) \geq \mathbb{E}_\tau[g(y_{\mathcal{S}_\nu^n}) - g(y_\tau)|\mathcal{F}_\tau] = \mathbb{E}_\tau\left[\int_\tau^{\mathcal{S}_\nu^n} \frac{1}{2}\alpha_t^2 dt \,\bigg|\, \mathcal{F}_\tau\right].$$

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

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

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 (20).

If we now return to (24); let $n \to \infty$, use monotone convergence on the right-hand side and bounded convergence on the left, and (20), we can prove (22). Following similar steps we can show (21) and (23). ■

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*

$$\mathbb{E}_\tau\left[\int_\tau^{T_\nu} \frac{1}{2}\alpha_t^2 dt \,\bigg|\, \mathcal{F}_\tau\right] = \mathbb{E}_\tau[g(y_{T_\nu}) - g(y_\tau)|\mathcal{F}_\tau]\,\mathbb{1}_{\{T_\nu>\tau\}} \tag{25}$$

$$\mathbb{E}_\infty\left[\int_\tau^{T_\nu} \frac{1}{2}\alpha_t^2 dt \,\bigg|\, \mathcal{F}_\tau\right] = \mathbb{E}_\infty[h(y_{T_\nu}) - h(y_\tau)|\mathcal{F}_\tau]\,\mathbb{1}_{\{T_\nu>\tau\}} \tag{26}$$

Proof of Corollary 1: The proof follows by another application of Itô's rule. Expectation of the stochastic integral is zero, 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 zero, since $g'(0) = h'(0) = 0$. ■

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*

$$\mathbb{E}_\tau[f(y_{T_\nu})|\mathcal{F}_\tau] = \mathbb{E}_\infty[e^{u_{T_\nu}-u_\tau} f(y_{T_\nu})|\mathcal{F}_\tau], \ \ \mathbb{P}_\infty\text{-a.s.} \tag{27}$$

Proof of Corollary 2: It should be noted that (27) is not obvious because Girsanov's theorem is valid only for bounded stopping times. Let $M > 0$ then

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

Notice now that for $T_\nu \geq \tau \geq 0$ we have $u_{T_\nu} - u_\tau \leq u_{T_\nu} - m_{T_\nu} = y_{T_\nu} \leq \nu$ therefore we obtain the following bounds for the last two terms

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

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

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

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

Let $\nu_\star$ be the threshold for which the corresponding CUSUM stopping time satisfies the false alarm constraint (13) with equality, that is

$$\mathbb{E}_\infty\left[\int_0^{\mathcal{S}_{\nu_\star}} \frac{1}{2}\alpha_t^2 dt\right] = h(\nu_\star) = e^{\nu_\star} - \nu_\star - 1 = \gamma. \tag{28}$$

For every γ there is a unique $\nu_\star$ satisfying (28). The worst K-L detection divergence of $S_{\nu_\star}$ can be obtained from Theorem 1 using the increase of $g(\cdot)$, specifically

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

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 (12), (13).

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 (13) 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 of Theorem 2: Since $T \geq T_\nu$ we have

$$J(T) \geq J(T_\nu) \geq \mathbb{E}_\tau\left[\int_\tau^{T_\nu} \frac{1}{2}\alpha_t^2 dt \,\middle|\, \mathcal{F}_\tau\right], \tag{29}$$

$$J(T) \geq J(T_\nu) \geq \mathbb{E}_0\left[\int_0^{T_\nu} \frac{1}{2}\alpha_t^2 dt\right], \tag{30}$$

for any $0 \leq \tau < \infty$, thanks to (12). Applying Corollary 1 on the right hand side and Corollary 2 on both sides of (29), we obtain

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

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

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

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

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

which leads to

$$J(T)\mathbb{E}_\infty[e^{y_{T_\nu}} - e^{u_{T_\nu}}] \geq \mathbb{E}_\infty[(e^{y_{T_\nu}} - e^{u_{T_\nu}})g(y_{T_\nu})]. \tag{31}$$

Focusing now on (30), recalling that $\mathcal{F}_0$ is the trivial σ-algebra, using Corollaries 1 & 2, and that $y_0 = 0$, we end up with

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

By adding this relation, term by term, to (31) we obtain

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

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 and S_ν the CUSUM stopping time with threshold ν, let $T_\nu = T \wedge S_\nu$ and define the function $\psi_T(\nu) = \mathbb{E}_\infty\left[\int_0^{T_\nu} \frac{1}{2}\alpha_t^2 dt\right]$, then $\psi_T(\nu)$ is continuous and increasing in ν with $\psi_T(0) = 0$ and $\psi_T(\infty) = \mathbb{E}_\infty[\int_0^T 0.5\alpha_t^2 dt]$.*

Proof of Lemma 3: 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

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

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 $\{T > S_\nu\}$ 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 (13) with equality. Indeed, if a stopping time T has $\mathbb{E}_\infty[\int_0^T 0.5\alpha_t^2 dt] > \gamma$ 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 (13) 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 (13) with equality, has a K-L detection divergence $J(T)$ that is no less than $g(\nu_\star)$.*

Proof of Theorem 3: 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

$$\frac{\mathbb{E}_\infty\left[e^{y_{T_{\nu_\varepsilon}}} g(y_{T_{\nu_\varepsilon}})\right]}{\mathbb{E}_\infty[e^{y_{T_{\nu_\varepsilon}}}]} \geq g(\nu_\star) - \varepsilon. \tag{32}$$

To prove (32), 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

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

From Corollary 1, we have $\mathbb{E}_\infty\left[\int_0^{T_{\nu_\varepsilon}} 0.5\alpha_t^2 dt\right] = \mathbb{E}[h(y_{T_{\nu_\varepsilon}})]$, which suggests that

$$\mathbb{E}[h(y_{T_{\nu_\varepsilon}})] \geq \gamma - \varepsilon. \tag{33}$$

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

$$U(T_{\nu_\varepsilon}) = \mathbb{E}_\infty\left[e^{y_{T_{\nu_\varepsilon}}}[g(y_{T_{\nu_\varepsilon}}) - g(\nu_\star)] - h(y_{T_{\nu_\varepsilon}}) + h(\nu_\star)\right]. \tag{34}$$

If we define the function $p(y) = e^y[g(y) - g(\nu_\star)] - h(y) + h(\nu_\star)$, it is a simple exercise to verify that its derivative has the same sign as $y - \nu_\star$. This suggests that $p(y)$ exhibits 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 (34) along with (33) and recalling from (28) that $h(\nu_\star) = \gamma$, yields

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

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 (5). 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 optimum stopping time. There are of course situations where (5) does not hold, as for example, in transient changes where the process returns to its nominal statistics after finite time. For such cases CUSUM is not necessarily optimum 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

$$d\Xi_t = \begin{cases} A_t dt + \Sigma_t dW_t, & 0 \le t \le \tau \\ B_t dt + \Sigma_t dW_t, & \tau < t, \end{cases}$$

where W is a vector Brownian motion; A, B are adapted vector process and Σ is an adapted matrix process, then our previous analysis goes through without significant modifications. The log-likelihood ratio in this case satisfies

$$du_t = \mathcal{A}_t^T dW_t \pm \frac{1}{2} \mathcal{A}_t^T \mathcal{A}_t dt$$

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

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

Here the role of α_t^2 plays the quantity $\mathcal{A}_t^T \mathcal{A}_t$ 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, then the modified criterion is equivalent to the original Lorden criterion. In other words CUSUM is optimum, 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.

Let us now present two examples that fall into our class of processes (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 non-homogeneous 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 (4), (5) and (9) are satisfied and therefore CUSUM is optimum in the proposed generalized sense.

A more interesting situation occurs when $\alpha_t = -\alpha\xi_t$, where α 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 (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 (1988), page 200, applies showing validity of (6).

For (9), to show first $\mathbb{P}_\infty[\int_0^\infty \xi_t^2 dt = \infty] = 1$ we observe, using Schwartz 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 $\sigma_t^2 = c_2 t^2$, where c_1, c_2 constants. If $M > 0$ we can then write

$$\begin{aligned}\mathbb{P}_\infty\left[\int_0^\infty \xi_s^2 ds \leq M\right] &\leq \mathbb{P}_\infty\left[\int_0^t \xi_s^2 ds \leq M\right] \leq \mathbb{P}_\infty\left[|z_t| \leq \sqrt{M}\right] \\ &= \Phi((\sqrt{M}-\mu_t)/\sigma_t) - \Phi(-(\sqrt{M}+\mu_t)/\sigma_t),\end{aligned}$$

where $\Phi(z)$ is the standard Gaussian cumulative distribution. The last term tends to zero as t tends to infinity. For a different proof see Problem 6.30 of Karatzas and Shreve (1988), page 217.

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

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

For the process z_t we can then show that its expected value is of the form $\mu_t = c_1 t + o(t)$ and its variance $\sigma_t^2 = c_2 t + o(t)$ with c_1, 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

$$\begin{aligned}\mathbb{P}_0\left[\int_0^\infty \xi_s^2 ds \leq M\right] &\leq \mathbb{P}_0[z_t \leq M] = \mathbb{P}_0\left[\left(\frac{\mu_t - z_t}{\sigma_t}\right)^2 \geq \left(\frac{\mu_t - M}{\sigma_t}\right)^2\right] \\ &\leq \left(\frac{\sigma_t}{\mu_t - M}\right)^2.\end{aligned}$$

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

7 Acknowledgement.

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

References

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

[2] GALTCHOUK, L. (2001). Optimality of the Wald SPRT for processes with continuous time parameter, *Advances in Model-Oriented Design and Analysis. Contributions in Statistics*, 97–107, Springer.

[3] KARATZAS, I., SHREVE S. E. (1988). *Brownian Motion and Stochastic Calculus*, Springer, New York.

[4] IRLE, A. (1984). Extended optimality of SPRT, *Ann. Stat.* **12**, 380–386.

[5] LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes II*, Springer.

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

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

[8] ØKSENDAL, B. K. (2000). *Stochastic Differential Equations: An Introduction with Applications*, 5-th edition, Universitext.

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

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

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

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

[13] SHIRYAYEV, A. N. (1996). Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russian Mathem. Surv.* **51**, 750-751.

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

[15] WALD, A. (1947). *Sequential Analysis*, John Wiley, New York.

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

[17] YASHIN, A. (1983). On a problem of sequential hypothesis testing, *Th. Proba. Applic.* **28**, 157–165.

Unité de recherche INRIA Rennes
IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifi que
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399