Sequential Analysis, 29: 483–486, 2010

ISSN: 0747-4946 print/1532-4176 online
DOI: 10.1080/07474946.2010.520633

Taylor & Francis
Taylor & Francis Group

A Note on "The Optimal Stopping Time for Detecting Changes in Discrete Time Markov Processes" by Han and Tsung

George V. Moustakides[1] and Alexander G. Tartakovsky[2]
[1]Department of Electrical and Computer Engineering, University of Patras, Patras, Greece
[2]Department of Mathematics, University of Southern California, Los Angeles, California, USA

Abstract: We analyze the article by Han and Tsung (2009) "The Optimal Stopping Time for Detecting Changes in Discrete Time Markov Processes," and demonstrate that it is seriously flawed.

Keywords: Change-point detection; CUSUM procedure.

Subject Classifications: 62L10; 62N10.

1. INTRODUCTION

Han and Tsung (2009) consider Markov observations $X_1, X_2, \dots$ with two possible stationary transition probabilities $Pr(X_n = y \mid X_{n-1} = x) = p_i(x, y)$, $i = 0, 1$. Up to some point in time t the samples follow $p_0(x, y)$ and after t they switch to $p_1(x, y)$. The goal is to detect the change as soon as possible.

As a possible detection procedure, the authors propose the CUSUM stopping time T_p, which is defined with the help of the CUSUM statistic

$$V_n = \max\{V_{n-1}, 1\} l(X_{n-1} \mid X_n); \quad l(X_{n-1} \mid X_n) = \frac{p_1(X_{n-1}, X_n)}{p_0(X_{n-1}, X_n)}; \quad V_0 = 1,$$

Received January 15, 2010, Revised June 3, 2010, Accepted July 15, 2010
Recommended by Nitis Mukhopadhyay
Address correspondence to G. V. Moustakides, Department of Electrical and Computer Engineering, University of Patras, Patras, Rio, 26500 Greece; E-mail: moustaki@upatras.gr or A. G. Tartakovsky, Department of Mathematics, University of Southern California, KAP-108, Los Angeles, CA 90089-2532, USA; E-mail: tartakov@math.usc.edu

as $T_p = \inf\{n \geq 1 : V_n \geq \mu\}$. We emphasize that the threshold μ, as in the case of the classical CUSUM, is considered *constant* and selected to satisfy the false alarm constraint $E_0[T_p] = L$, where $L > 1$ is some pre-specified constant.

As a performance measure, the authors use the Lorden (1971) criterion

$$D(T) = \sup_{m\geq 1} \operatorname{ess\,sup} E^{(m)}[(T - m + 1) \mid X_1, \ldots, X_{m-1}].$$

In fact the definition of the Lorden measure given by the authors in Lemma 2.1 is incomplete, since it is missing the first supremum over $m \geq 1$. The goal is to minimize $D(T)$ over all stopping times T that satisfy the false alarm constraint $E_0[T] \geq L$.

Following Moustakides (1986), the authors show in Lemma 2.1 that a lower bound can be found, namely

$$D(T) \geq \frac{E_0\Big[\sum_{n=0}^{T-1} \max\{V_n, 1\}\Big]}{E_0\Big[\sum_{n=0}^{T-1} (1 - V_n)^+\Big]} = \overline{D}(T), \tag{1}$$

and then instead of optimizing $D(T)$ over the class of stopping times $E_0[T] \geq L$, they optimize the lower bound $\overline{D}(T)$ by minimizing the numerator and maximizing the denominator of $\overline{D}(T)$ with the help of Theorem 2.1. In order for this double optimization to lead to the minimization of $\overline{D}(T)$, it is clear that *the same* (optimal) stopping time should be a solution to both optimization problems (for the numerator and the denominator).

2. DISCUSSION OF THE PROOF OF THEOREM 2.1

Theorem 2.1 in Han and Tsung (2009) claims optimality of CUSUM for the general constrained optimization problem

$$\sup_T E_0\left[\sum_{n=0}^{T-1} \varphi(V_n)\right] = E_0\left[\sum_{n=0}^{T_p-1} \varphi(V_n)\right] \tag{2}$$

for all T that satisfy $E_0[T] = L$, where $\varphi(x)$ is a continuous nonincreasing function. This theorem is a key result for demonstrating the optimality of CUSUM in the sense of Lorden.

Consider the optimization problem in (2), which can be solved using the Optimal Stopping Theory in Shiryaev (1978). Note that in order to apply Shiryaev's results the decision statistic has to be a *Markov process*, and for the CUSUM procedure (with a constant threshold) to be optimal the CUSUM statistic $\{V_n\}_{n\geq 1}$ should be a homogeneous Markov process, as in the i.i.d. case. However, for the Markov model considered the CUSUM statistic is NOT Markov. In order to obtain a Markov process we need to combine V_n with the observation X_n. In other words, the two-dimensional process $\{(V_n, X_n)\}_{n\geq 1}$ has the desired Markov property.

To optimize (2) we reduce it into an unconstrained problem and consider

$$U(v, x) = \sup_T E_0\left[\sum_{n=0}^{T-1} \{\varphi(V_n) - \lambda\}\right]$$

where $V_0 = v$, $X_0 = x$. The optimum cost function $U(v, x)$ depends, of course, on the initial state of the Markov process $\{(V_n, X_n)\}$. By Theorem 2.23 of Shiryaev (1978), $U(v, x)$ satisfies the following equation

$$U(v, x) = \max\{0, \varphi(v) - \lambda + E_0[U(V_1, X_1) \mid V_0 = v, X_0 = x]\}.$$

It is clear that $E_0[U(V_1, X_1) \mid V_0 = v, X_0 = x] = \overline{U}(v, x)$ is a function of both quantities v, x. Consequently, finding the optimum stopping boundary requires solving the equation $\varphi(v) - \lambda + \overline{U}(v, x) = 0$, which has a solution of the form $v = \mu(x)$, but by no means a constant boundary $v = \mu$. In other words, the optimal threshold is *a function of* x, and the CUSUM test needs to be modified as $T_p = \inf\{n : V_n \geq \mu(X_n)\}$. Let us call this CUSUM test the $(V, X)-$CUSUM, to emphasize the difference with the traditional CUSUM.

One might suggest that we can simply adopt this modification for the CUSUM procedure, i.e., introduce a variable threshold, and this will be sufficient to guarantee the desired optimality. However, we note that the threshold $\mu(x)$ depends in general on the form of the function $\varphi(x)$. Since Theorem 2.1 is used to optimize, separately, the numerator and the denominator in $\overline{D}(T)$ defined in (1), there is absolutely no guarantee that the optimum threshold will turn out to be the same in both optimization problems.

Let us disregard even this latter complication and assume that there exists a unique variable threshold $\mu(x)$ for CUSUM that optimizes the lower bound $\overline{D}(T)$. Does this necessarily imply optimization of the original Lorden measure $D(T)$ as in the case of i.i.d. observations?

3. DISCUSSION OF THE OPTIMIZATION OF LORDEN'S MEASURE

Note from (1) that the Lorden measure $D(T)$ is lower-bounded by $\overline{D}(T)$. By our previous assumption, there exists a variable threshold CUSUM version T_p that optimizes $\overline{D}(T)$. In order for this to also imply optimization of the original Lorden measure $D(T)$, the equality $D(T_p) = \overline{D}(T_p)$ should hold. In other words, for $(V, X)-$CUSUM, Lorden's measure and the lower bound must coincide. Indeed, if this were the case, then from (1), by minimizing over T both sides, we could argue that $\inf_T D(T) \geq \inf_T \overline{D}(T) = \overline{D}(T_p) = D(T_p)$. This would also imply $\inf_T D(T) \geq D(T_p)$, i.e., optimality of the $(V, X)-$CUSUM procedure in Lorden's sense. It is therefore imperative to show that for $(V, X)-$CUSUM $D(T_p)$ and $\overline{D}(T_p)$ are equal. Unfortunately this is not the case, as we briefly explain next.

Note that due to the Markov structure of the sequence $\{(V_n, X_n)\}_{n\geq 1}$

$$E^{(m)}\big[(T_p - m + 1)^+ \mid X_0, \ldots, X_{m-1}\big] = \psi(V_{m-1}, X_{m-1})$$

for some suitable function $\psi(v, x)$; where, we recall, $E^{(m)}[\cdot]$ denotes expectation with respect to the probability measure induced by the change occurring at time m. Since *for a given* X_{m-1} the statistic V_n, $n \geq m$, is path-wise increasing in V_{m-1}, we conclude that the detection delay $\psi(V_{m-1}, X_{m-1})$ is decreasing in V_{m-1}, implying that $\psi(V_{m-1}, X_{m-1}) \leq \psi(1, X_{m-1})$. In other words, for a given X_{m-1} the conditional expectation is maximized when we restart the CUSUM process. Unlike the i.i.d. case

$\psi(1, X_{m-1})$ does not correspond to the worst performance of CUSUM, since we need to further maximize over X_{m-1}. Indeed, note that

$$\operatorname{ess\,sup} E^{(m)}\big[(T_p - m + 1)^+ \mid X_0, \dots, X_{m-1}\big] = \operatorname{ess\,sup} \psi(V_{m-1}, X_{m-1}) = \sup_x \psi(1, x).$$

This has a grave consequence on the relationship between $D(T_p)$ and $\overline{D}(T_p)$. The lower bound $\overline{D}(T_p)$ is obtained by summing the following quantities (see Lemma 3 in Moustakides, 1986)

$$\begin{aligned} E_0\big[E^{(m)}[(T_p - m + 1)^+ \mid X_0, \dots, X_{m-1}](1 - V_{m-1})^+\big] &= E_0[\psi(V_{m-1}, X_{m-1})(1 - V_{m-1})^+] \\ &= E_0[\psi(1, X_{m-1})(1 - V_{m-1})^+] \\ &\le \Big\{\sup_x \psi(1, x)\Big\} E_0[(1 - V_{m-1})^+]. \end{aligned}$$

The second equality is true because every time $V_n \le 1$ the CUSUM process restarts from 1. Note now that the last inequality *is strict* whenever $\psi(1, x)$ is not a constant. While equality holds for i.i.d. observations (since $\psi(V, x) = \psi(V)$), the same property is NOT valid for Markov observations. Thus the previous inequality is strict yielding the strict inequality $D(T_p) > \overline{D}(T_p)$ between Lorden's measure and its lower bound. Consequently, even if the $(V, X)-$CUSUM test optimizes $\overline{D}(T)$, this does not necessarily imply that it optimizes the original Lorden measure $D(T)$.

4. CONCLUSION

Although we are sure that the conventional CUSUM is not optimal in the Markov case, we would like to stress that in this note we did not rigorously prove that Han and Tsung's claim regarding optimality of CUSUM is wrong. Rather we showed that crucial parts of their proof are problematic. We believe that finding the optimal detection procedure in the Markov case requires fresh ideas that go far beyond immediate extensions of existing methodologies. Furthermore, if a CUSUM-like procedure turns out to be optimal, we believe that not only the stopping threshold, but also the restarting barrier (which is 1 in the classical case) will be functions of the current observation X_n.

REFERENCES

Han, D. and Tsung, F. (2009). The Optimal Stopping Time for Detecting Changes in Discrete Time Markov Processes, *Sequential Analysis* 28: 115–135.

Lorden, G. (1971). Procedures for Reacting to a Change in Distribution, *Annals of Mathematical Statistics* 42: 1897–1908.

Moustakides, G. V. (1986). Optimal Stopping Times for Detecting Changes in Distributions, *Annals of Statistics* 14: 1379–1387.

Shiryaev, A. N. (1978). *Optimal Stopping Rules*, New York: Springer-Verlag.