

SUBOPTIMAL PROPERTIES OF PAGE'S CUSUM AND SHIRYAYEV-ROBERTS PROCEDURES IN CHANGE-POINT PROBLEMS WITH DEPENDENT OBSERVATIONS

Yajun Mei

Georgia Institute of Technology

Abstract: We construct a simple counterexample to the conjectures of Pollak (1985) and Yakir, Krieger and Pollak (1999), which state that Page's CUSUM procedure and the Shirayev-Roberts procedure are asymptotically minimax optimal for dependent observations. Moreover, our example shows that the close relationship between open-ended tests and change-point detection procedures no longer holds for dependent observations. As a consequence, the standard approach which constructs change-point detection procedures based on asymptotically optimal open-ended tests does not in general provide asymptotically optimal procedures for dependent observations.

Key words and phrases: Asymptotic optimality, CUSUM, open-ended tests, quality control, Shirayev-Roberts, statistical process control.

1. Introduction

Sequential change-point detection problems have many important applications, including industrial quality control, reliability, fault detection, signal detection, surveillance and security systems. Extensive research has been done in this field during the last few decades. For recent reviews, we refer readers to Basseville and Nikiforov (1993), Lai (1995, 2001), and the references therein.

In change-point problems, one observes a sequence of observations X_1, X_2, \dots from some process. Initially, the process is "in control," i.e., the X 's have some distribution f . At some unknown time ν , the process may go "out of control" and the distribution of the observations $\{X_n\}$ changes abruptly to another distribution g . It is desired to raise an alarm as soon as the change occurs so that appropriate action can be taken.

In the simplest situation where the observations $\{X_n\}$ are independent and both the pre-change distribution f and the post-change distribution g are completely specified, the problem is well understood and has been solved under a variety of criteria. Two efficient detection schemes are Page's cumulative sum (CUSUM) procedure and the Shirayev-Roberts procedure, see, for example,

Page (1954), Shiriyayev (1963), Roberts (1966), Lorden (1971), Moustakides (1986), Pollak (1985) and Ritov (1990).

In practice, the assumption of independent observations is too restrictive, and change-point problems involving dependent observations have been important topics in the literature, see, for example, Lai (1995, 2001). It is natural to extend Page's CUSUM procedure or the Shiriyayev-Roberts procedure, both of which are based on likelihood ratios, to dependent observations by simply replacing the probability densities with the corresponding conditional densities. But it is unclear whether Page's CUSUM and the Shiriyayev-Roberts procedures are still efficient in the presence of dependent observations.

There are two standard mathematical formulations for studying the optimality properties of Page's CUSUM and the Shiriyayev-Roberts procedures. The first one is a Bayesian formulation, due to Shiriyayev (1963), in which the change-point ν is assumed to have a known prior distribution. There are a few papers in the literature that used this approach to study dependent observations, see, for example, Yakir (1994), Beibel (1997) and Lai (1998), and all show that Page's CUSUM and the Shiriyayev-Roberts procedures are asymptotically optimal under Bayesian formulations in their respective special models.

The second option is a minimax formulation, proposed by Lorden (1971), in which the change-point ν is assumed to be unknown (possibly ∞) but non-random. Lai (1998) showed that Page's CUSUM procedure is still asymptotically minimax optimal for dependent observations under some conditions which are difficult to verify in general. Pollak (1987) showed that the Shiriyayev-Roberts procedure is asymptotically minimax optimal in change-point problems for post-change distributions that are a certain type of mixture. Fuh (2003, 2004) proved asymptotic minimax optimality of Page's CUSUM and the Shiriyayev-Roberts procedures for hidden Markov models. Pollak conjectured that Page's CUSUM and the Shiriyayev-Roberts procedures are asymptotically minimax optimal for dependent observations in a wide context, see page 226 of Pollak (1985) and page 1905-1906 of Yakir, Krieger and Pollak (1999).

In this paper, we use the second of these formulations and disprove Pollak's conjecture by constructing a counterexample where the pre-change distribution is a so-called "mixture distribution" and the post-change distribution is fully specified. Section 2 states our counterexample, and Section 3 shows that Page's CUSUM and the Shiriyayev-Roberts procedures are not asymptotically minimax optimal in this setting. Numerical simulation results are given in Section 4.

Our example also shows that the close relationship between "open-ended tests," developed by Robbins (1970) and Robbins and Siegmund (1970, 1973), and change-point detection procedures no longer holds for dependent observations. Since Lorden (1971) established such a relationship in the setting of

independent observations and completely specified pre-change and post-change distributions, it has been a standard approach to study change-point problems via open-ended tests, for example, see Pollak and Siegmund (1975), Basseville and Nikiforov (1993), and Lai (1995, 1998). Actually, it is this relationship that leads to Pollak's conjecture, see Section 5 of Yakir, Krieger and Pollak (1999). In Section 5 of this paper, we demonstrate that this relationship fails for dependent observations, and thus the standard approach of studying change-point problems via open-ended tests could be misleading when observations are dependent.

2. The Counterexample

Consider three given probability densities f_1, f_2 and g such that

$$\mathbf{E}_g\left(\log\frac{g(X)}{f_j(X)}\right)^2 < \infty \quad \text{and} \quad \mathbf{E}_{f_j}\left(\log\frac{f_1(X)}{f_2(X)}\right)^2 < \infty, \quad \text{for } j = 1, 2, \quad (2.1)$$

and $I_1 > I_2 > 0$, where $I_j = I(g, f_j) = \mathbf{E}_g(\log(g(X)/f_j(X)))$, $j = 1, 2$, are the Kullback-Leibler information numbers. Denote respectively by $\mathbf{P}_{f_1}, \mathbf{P}_{f_2}$ and \mathbf{P}_g the probability measures when X_1, X_2, \dots are independent and identically distributed (i.i.d.) with densities f_1, f_2 and g . Choose a constant $\pi_0 \in (0, 1)$, say $\pi_0 = 1/3$, and define $\mathbf{P}_f = \pi_0\mathbf{P}_{f_1} + (1 - \pi_0)\mathbf{P}_{f_2}$. That is, under \mathbf{P}_f , X_1, \dots, X_n have a "mixture" joint density

$$f(x_1, \dots, x_n) = \pi_0 \prod_{i=1}^n f_1(x_i) + (1 - \pi_0) \prod_{i=1}^n f_2(x_i). \quad (2.2)$$

To simplify notations, we also denote by $f(\cdot|X_1, \dots, X_{n-1})$ and $g(\cdot|X_1, \dots, X_{n-1})$ the conditional density functions of X_n given X_1, \dots, X_{n-1} under \mathbf{P}_f and \mathbf{P}_g , respectively. Note that $g(\cdot|X_1, \dots, X_{n-1}) = g(\cdot)$ because observations are independent under \mathbf{P}_g .

In change-point problems, we are interested in detecting a change in distribution from \mathbf{P}_f to \mathbf{P}_g . For $1 \leq \nu < \infty$, let $\mathbf{P}^{(\nu)}$ and $\mathbf{E}^{(\nu)}$ denote the probability measure and expectation, respectively, when the change in distribution from \mathbf{P}_f to \mathbf{P}_g occurs at the ν th observation, so that $X_1, \dots, X_{\nu-1}$ have a mixture joint density f and $X_\nu, X_{\nu+1}, \dots$ are i.i.d. with density g . Denote by \mathbf{P}_f and \mathbf{E}_f the probability measure and expectation when there is no change, i.e. $\nu = \infty$, in which case X_1, X_2, \dots are distributed with mixture joint density f .

A procedure for detecting that a change has occurred is defined as a stopping time N with respect to $\{X_n\}_{n \geq 1}$. The interpretation of N is that, when $N = n$, we stop taking observations at time n and declare that a change occurred somewhere in the first n observations. We want to find a stopping time N which will stop as soon as possible after a change occurs, but will continue taking observations

as long as possible if no change occurs. Thus, the performance of a stopping time N is evaluated by two criteria: the long and short Average Run Lengths (ARL). The long ARL is defined by $\mathbf{E}_f(N)$. Imagining repeated applications of such procedures, practitioners refer to the frequency of false alarms as $1/\mathbf{E}_f(N)$ and the ARL to false alarm as $\mathbf{E}_f(N)$. The short ARL can be defined by the following worst case detection delay, proposed by Lorden (1971):

$$\bar{\mathbf{E}}_g(N) = \sup_{1 \leq \nu < \infty} \left(\text{ess sup } \mathbf{E}^{(\nu)}[(N - \nu + 1)^+ | X_1, \dots, X_{\nu-1}] \right).$$

In our theorems we can also use the average detection delay, proposed by Shirayev (1963) and Pollak (1985):

$$\sup_{1 \leq \nu < \infty} \mathbf{E}^{(\nu)}(N - \nu | N \geq \nu),$$

which is asymptotically equivalent to $\bar{\mathbf{E}}_g(N)$.

Similar to the standard minimax formulation for independent observations, our problem can be stated as follows: find a stopping time N that minimizes the detection delay $\bar{\mathbf{E}}_g(N)$ subject to

$$\mathbf{E}_f(N) \geq \gamma, \quad (2.3)$$

where γ is a given, fixed lower bound.

For this problem, Page's CUSUM procedure has stopping time

$$\begin{aligned} T_{CM}(A) &= \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \prod_{i=k}^n \frac{g(X_i | X_1, \dots, X_{i-1})}{f(X_i | X_1, \dots, X_{i-1})} \geq A \right\} \\ &= \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \frac{\prod_{i=k}^n g(X_i)}{\pi_{k-1} \prod_{i=k}^n f_1(X_i) + (1 - \pi_{k-1}) \prod_{i=k}^n f_2(X_i)} \geq A \right\}, \end{aligned}$$

where

$$\pi_k = \frac{\pi_0 \prod_{i=1}^k f_1(X_i)}{\pi_0 \prod_{i=1}^k f_1(X_i) + (1 - \pi_0) \prod_{i=1}^k f_2(X_i)}.$$

Similarly, the Shirayev-Roberts procedure has stopping time

$$\begin{aligned} T_{SR}(A) &= \inf \left\{ n \geq 1 : \sum_{k=1}^n \prod_{i=k}^n \frac{g(X_i | X_1, \dots, X_{i-1})}{f(X_i | X_1, \dots, X_{i-1})} \geq A \right\} \\ &= \inf \left\{ n \geq 1 : \sum_{k=1}^n \frac{\prod_{i=k}^n g(X_i)}{\pi_{k-1} \prod_{i=k}^n f_1(X_i) + (1 - \pi_{k-1}) \prod_{i=k}^n f_2(X_i)} \geq A \right\}. \end{aligned}$$

3. Suboptimal Properties in Change-Point Problems

We first establish sharp asymptotic lower bounds for the detection delays of any procedures satisfying (2.3). Later these bounds will be used to prove the asymptotic suboptimal properties of $T_{CM}(A)$ and $T_{SR}(A)$, Page's CUSUM and the Shiriyayev-Roberts procedures.

Theorem 1. *Let $n(\gamma)$ be the infimum of $\bar{\mathbf{E}}_g(N)$ as N ranges over the class of stopping times satisfying (2.3). Then, as $\gamma \rightarrow \infty$,*

$$n(\gamma) = (1 + o(1)) \frac{\log \gamma}{I_1}. \quad (3.1)$$

Proof. It is easy to see from the definition of the mixture distribution f that for any stopping time N ,

$$\mathbf{E}_f(N) = \pi_0 \mathbf{E}_{f_1}(N) + (1 - \pi_0) \mathbf{E}_{f_2}(N). \quad (3.2)$$

Thus, for any stopping time N satisfying (2.3), $\mathbf{E}_{f_1}(N) \geq \gamma$ or $\mathbf{E}_{f_2}(N) \geq \gamma$. Note that since the observations are independent under \mathbf{P}_{f_1} or \mathbf{P}_{f_2} , by the classical lower bound on the detection delay for independent observations (see Theorem 3 of Lorden (1971)),

$$\bar{\mathbf{E}}_g(N) \geq (1 + o(1)) \frac{\log \gamma}{I_1} \quad \text{or} \quad \bar{\mathbf{E}}_g(N) \geq (1 + o(1)) \frac{\log \gamma}{I_2}.$$

Since $I_1 > I_2$, we have

$$\bar{\mathbf{E}}_g(N) \geq (1 + o(1)) \frac{\log \gamma}{I_1}, \quad \text{as } \gamma \rightarrow \infty.$$

To prove the reverse inequality, consider Page's CUSUM procedure for detecting a change in distribution of independent observations from f_1 to g , which has a stopping time

$$T_1 = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \prod_{i=k}^n \frac{g(X_i)}{f_1(X_i)} \geq \frac{\gamma}{\pi_0} \right\}. \quad (3.3)$$

By the classical results for independent observations (see Theorem 1 of Lorden (1971)),

$$\bar{\mathbf{E}}_g(T_1) = (1 + o(1)) \frac{\log(\frac{\gamma}{\pi_0})}{I_1} = (1 + o(1)) \frac{\log \gamma}{I_1} \quad \text{as } \gamma \rightarrow \infty,$$

while for all γ , $\mathbf{E}_{f_1}(T_1) \geq \gamma/\pi_0$, showing that T_1 satisfies (2.3) by virtue of (3.2). This property of T_1 shows the lower bound $n(\gamma)$ is asymptotically no larger than the right-hand side of (3.1), completing the proof.

Next, we consider the asymptotic behavior of $T_{CM}(A)$ and $T_{SR}(A)$ for large values of A , regardless of the constraint (2.3). For these procedures, it is well-known that their detection delays and the logarithms of their ARLs to false alarm are of order $\log A$, in the setting of independent observations. The following theorem, whose proof is given in the appendix, shows that similar conclusions hold for our example.

Theorem 2. *Let T be either $T_{CM}(A)$ or $T_{SR}(A)$. Then, as $A \rightarrow \infty$,*

$$\log \mathbf{E}_f(T) \leq (1 + o(1)) \log A, \quad (3.4)$$

$$\overline{\mathbf{E}}_g(T) \geq (1 + o(1)) \frac{\log A}{I_2}. \quad (3.5)$$

We are now in a position to establish the asymptotic suboptimal properties of $T_{CM}(A)$ and $T_{SR}(A)$ in the change-point problems with the constraint on the ARL to false alarm.

Theorem 3. *Under the constraint (2.3), $T_{CM}(A)$ and $T_{SR}(A)$ do not asymptotically minimize the detection delay.*

Proof. Let T be either $T_{CM}(A)$ or $T_{SR}(A)$. In order to satisfy (2.3), the threshold $A = A_\gamma$ must be chosen so that $\log A \geq (1 + o(1)) \log \gamma$, by (3.4). Combining this with (3.5) yields $\overline{\mathbf{E}}_g(T) \geq (1 + o(1))(\log \gamma)/I_2$. By Theorem 1, we have $\overline{\mathbf{E}}_g(T)/n(\gamma) \geq (1 + o(1))(I_1/I_2)$. Since $I_1 > I_2$, the detection delay of T is asymptotically larger than the sharp lower bound $n(\gamma)$ for sufficient large γ , so the theorem is proved.

Remarks. The first-order asymptotic expansions in Theorems 1–3 can be improved to second order (up to $O(1)$) by nonlinear renewal theory, see Theorem 4.5 in Woodroffe (1982), or Theorem 9.28 in Siegmund (1985). For our purpose, the first-order expansions are sufficient since Page's CUSUM and the Shirayayev-Roberts procedures are not even first-order asymptotically optimal.

As one referee pointed out, if one considers a Bayesian version of our example by assuming that the change-point ν is geometrically distributed, then $T_{CM}(A)$ and $T_{SR}(A)$ are asymptotically optimal in the Bayesian formulation, see, for example, Beibel (1997) and Lai (1998).

A subtle point here lies in the criteria that one uses to assess change-point detection procedures. In the Bayesian formulation, the false alarm criterion is $\mathbf{P}_f(N < \nu)$, which is asymptotically equivalent to $\mathbf{P}_{f_2}(N < \nu)$ for our example in Section 2. Thus, the Bayesian formulation of our example is asymptotically equivalent to the problem of detecting a change from f_2 to g . Meanwhile, in the standard minimax formulation, the false alarm criterion is $\mathbf{E}_f(N)$, which is asymptotically equivalent to $\mathbf{E}_{f_1}(N)$ for our example. Thus, the standard

minimax formulation of our example is asymptotically equivalent to the problem of detecting a change from f_1 to g .

By Theorems 1–3, $T_{CM}(A)$ and $T_{SR}(A)$ perform poorly when detecting a change from f_1 to g , although they can effectively detect a change from f_2 to g . Therefore, they are asymptotically optimal under the Bayesian formulation, but are asymptotically suboptimal under the standard minimax formulation. At a small additional cost of detection delay, either of the following procedures is able to detect both changes well:

$$\begin{aligned} M(a) &= \inf \left\{ n \geq 1 : \min_{j=1,2} \min_{1 \leq k \leq n} \frac{1}{I_j} \sum_{i=k}^n \log \frac{g(X_i)}{f_j(X_i)} \geq a \right\}; \\ M^*(a) &= \inf \left\{ n \geq 1 : \min_{1 \leq k \leq n} \min_{j=1,2} \frac{1}{I_j} \sum_{i=k}^n \log \frac{g(X_i)}{f_j(X_i)} \geq a \right\}. \end{aligned} \quad (3.6)$$

The properties of these last procedures were studied in Mei (2006). In particular, it can be shown that they are asymptotically first-order optimal in both Bayesian and minimax formulations for our Section 2 example.

4. A Numerical Example

The purpose of this section is to illustrate Theorem 3 about the suboptimality of Page's CUSUM and the Shiriyayev-Roberts procedures within the traditional minimax framework.

Table 4.1 compares the results of a 2,500-repetition Monte Carlo experiment in MATLAB for our Section 2 change-point problem with $f_1 = N(1, 1)$, $f_2 = N(-0.5, 1)$, $g = N(0, 1)$ and $\pi_0 = 1/3$. Note that the expected values of sample means are 0 under both the pre-change distribution f and the post-change distribution g .

Four different procedures are considered in Table 4.1: Page's CUSUM procedure $T_{CM}(A)$; Shiriyayev-Roberts procedure $T_{SR}(A)$; $T_1(\gamma)$, defined by (3.3); and $M(a)$, defined by (3.6).

For each of these four procedures τ , the threshold value was first determined from the criterion $\mathbf{E}_f(\tau) \approx 1,000$. Rather than simulating $\mathbf{E}_f(\tau)$ directly (which is complicated because it requires generating dependent random numbers), an efficient, easy-to-implement algorithm is to simulate $\mathbf{E}_{f_1}(\tau)$ and $\mathbf{E}_{f_2}(\tau)$, and then to calculate $\mathbf{E}_f(\tau)$ by (3.2).

Next, using the obtained threshold value, we ran 2,500-repetitions to simulate $\mathbf{E}_g(\tau)$, the expected sample size when the change happens at time $\nu = 1$. By definition, this gives a lower bound of the detection delay $\overline{\mathbf{E}}_g(\tau)$ in $\mathbf{E}_g(\tau)$ for all procedures including $T_{CM}(A)$ and $T_{SR}(A)$. Moreover, the renewal property of

the CUSUM statistics implies that the detection delay $\overline{\mathbf{E}}_g(\tau)$ for the procedures $T_1(\gamma)$ and $M(a)$ is just $\mathbf{E}_g(\tau)$.

Table 4.1. Comparisons of four stopping times.

	$\mathbf{E}_{f_1}(\tau)$	$\mathbf{E}_{f_2}(\tau)$	$\mathbf{E}_f(\tau)$	$\overline{\mathbf{E}}_g(\tau)$
$T_{CM}(A)$ ($A = 89.5$)	557 ± 11	1225 ± 25	1002	$\geq 33.1 \pm 0.3$
$T_{SR}(A)$ ($A = 675$)	1218 ± 25	895 ± 17	1003	$\geq 32.5 \pm 0.3$
$T_1(\gamma)$ ($\gamma = 156$)	2997 ± 62	7 ± 0.1	1004	12.4 ± 0.1
$M(a)$ ($a = 12.24$)	2928 ± 61	46 ± 1	1007	17.2 ± 0.1

Table 4.1 indicates that $T_1(\gamma)$ performs better than both $T_{CM}(A)$ and $T_{SR}(A)$ in the sense that $T_1(\gamma)$ has a much smaller detection delay. This conclusion still holds if $T_1(\gamma)$ is replaced by $M(a)$. Table 4.1 also shows that under the standard minimax formulation, the ARL to false alarm under f_2 does not play any serious role. To overcome this, a new approach, proposed by Mei (2006), is to specify the required detection delay $\overline{\mathbf{E}}_g(\tau)$ while trying to maximize the ARLs to false alarm under both f_1 and f_2 . See also Section 3 of Mei (2003).

5. Open-Ended Hypothesis Testing Problems

In the literature, a standard approach to studying the problem of detecting a change in distribution from f to g is to relate it to the following open-ended hypothesis testing problems which were developed by Robbins (1970) and Robbins and Siegmund (1970, 1973). Suppose that X_1, X_2, \dots are sampled from a true distribution p , and we are interested in testing the null hypothesis $H_0 : p = f$ against the alternative hypothesis $H_1 : p = g$. Assume that if H_0 is true, sampling costs nothing and our preferred action is just to observe X_1, X_2, \dots without stopping. On the other hand, if H_1 is true, each observation costs a fixed amount and we want to stop sampling as soon as possible and reject the null hypothesis H_0 .

An open-ended test is a statistical procedure for an open-ended hypothesis testing problem. Since there is only one terminal decision, an open-ended test is defined by a stopping time τ . The null hypothesis H_0 is rejected if and only if $\tau < \infty$. A good open-ended test τ should keep the error probabilities $\mathbf{P}_f(\tau < \infty)$ small while keeping $\mathbf{E}_g(\tau)$ small. Thus, a standard formulation for open-ended hypothesis testing problems is to seek an open-ended test τ that minimizes $\mathbf{E}_g(\tau)$ subject to the constraint

$$\mathbf{P}_f(\tau < \infty) \leq \alpha, \quad (5.1)$$

where $\alpha \in (0, 1)$ is a given constant.

Given an open-ended test τ , a standard method for finding change-point procedures is to define a new stopping time

$$N = \min_{k \geq 1} (\tau_k + k - 1), \quad (5.2)$$

where τ_k is the stopping time obtained by applying τ to X_k, X_{k+1}, \dots . Lorden (1971) showed that if the observations are independent and τ is asymptotically optimal in the open-ended hypothesis testing problem, then the stopping time N in (5.2) is asymptotically optimal in the corresponding change-point problem. This important result illustrates the close relationship between open-ended test and change-point procedures for independent observations. We now show that this close relationship no longer holds for dependent observations.

Consider our example in Section 2, where the distributions f and g are defined in (2.1) and (2.2). In the corresponding open-ended hypothesis testing problem, as an open-ended test the one-sided sequential probability ratio test (SPRT) is defined by

$$\begin{aligned} \tau_A &= \inf \left\{ n \geq 1 : \prod_{i=1}^n \frac{g(X_i | X_1, \dots, X_{i-1})}{f(X_i | X_1, \dots, X_{i-1})} \geq A \right\} \\ &= \inf \left\{ n \geq 1 : \frac{\prod_{i=1}^n g(X_i)}{\pi_0 \prod_{i=1}^n f_1(X_i) + (1 - \pi_0) \prod_{i=1}^n f_2(X_i)} \geq A \right\}. \end{aligned}$$

Page's CUSUM procedure $T_{CM}(A)$ can still be constructed from τ_A by using (5.2). However, Theorem 3 shows that $T_{CM}(A)$ is not asymptotically optimal in the change-point problem even though the next theorem shows that τ_A is asymptotically optimal (up to $O(1)$) in the open-ended hypothesis testing problem. In other words, asymptotically optimal open-ended tests do not in general lead to asymptotically optimal change-point detection procedures for dependent observations.

Theorem 4. *For any $0 < \alpha < 1$ let $A = 1/\alpha$. Then the one-sided SPRT, τ_A , satisfies (5.1) and, as $\alpha \rightarrow 0$,*

$$\mathbf{E}_g(\tau_A) \leq \frac{|\log \alpha|}{I_2} + O(1). \quad (5.3)$$

Moreover, if $\{\tau(\alpha)\}$ is a family of stopping times such that (5.1) holds, then

$$\mathbf{E}_g(\tau(\alpha)) \geq \frac{|\log \alpha|}{I_2} + O(1), \quad \text{as } \alpha \rightarrow 0. \quad (5.4)$$

Thus, τ_A , the one-sided SPRT, minimizes the expected sample size $\mathbf{E}_g(\tau_A)$ up to $O(1)$ among all open-ended tests satisfying (5.1).

Proof. It follows at once from Wald's likelihood ratio identity that τ_A satisfies (5.1). Before we prove (5.3), let us first prove (5.4). For any stopping time τ ,

$$\mathbf{P}_f(\tau < \infty) = \pi_0 \mathbf{P}_{f_1}(\tau < \infty) + (1 - \pi_0) \mathbf{P}_{f_2}(\tau < \infty) \quad (5.5)$$

by the definition of f . Thus, for any stopping time τ satisfying (5.1), we have

$$\mathbf{P}_{f_2}(\tau < \infty) \leq \frac{\alpha}{1 - \pi_0}.$$

Relation (5.4) follows at once from the well-known fact (Proposition 2.38 of Siegmund (1985)) that

$$\mathbf{E}_g(\tau) \geq \frac{|\log \mathbf{P}_{f_2}(\tau < \infty)|}{I_2}.$$

Now let us prove (5.3). To find an upper bound for $\mathbf{E}_g(\tau_A)$, define $\tau_A^* = \inf \{n \geq 1 : S_n \geq \log A \text{ and } U_n \leq 0\}$, where

$$S_n = \sum_{i=1}^n \log \frac{g(X_i)}{f_2(X_i)} \quad \text{and} \quad U_n = \sum_{i=1}^n \log \frac{f_1(X_i)}{f_2(X_i)}. \quad (5.6)$$

It is obvious that $\tau_A \leq \tau_A^*$, and so it suffices to show that $\mathbf{E}_g(\tau_A^*)$ satisfies (5.3).

Following the same lines as the proof of equations (2.21)–(2.25) of Kiefer and Sacks (1963) (a modification is also used to prove Lemma 1 in the Appendix), we have

$$\mathbf{E}_g(\tau_A^*) \leq \mathbf{E}_g(t_A) + \mathbf{E}_g(t_+) \mathbf{E}_g(\eta), \quad (5.7)$$

where the stopping times t_A , t_+ and η are defined by $t_A = \inf \{n \geq 1 : S_n \geq \log A\}$, $t_+ = \inf \{n \geq 1 : S_n \geq 0\}$, and $\eta = \text{last time } U_n \geq 0$. Since the summands in U_n have mean $I_2 - I_1 < 0$ and finite variance under \mathbf{P}_g , it is well known that $\mathbf{E}_g(\eta) < \infty$, see, for example, Theorem D in Kiefer and Sacks (1963). Moreover, note that since S_n is a random walk with mean $\mathbf{E}_g(S_1) = I_2 > 0$ and finite variance under \mathbf{P}_g , by standard renewal theory,

$$\mathbf{E}_g(t_A) = \frac{\log A}{I_2} + O(1) \quad \text{and} \quad \mathbf{E}_g(t_+) < \infty.$$

Combining these with (5.7) shows that $\mathbf{E}_g(\tau_A^*)$ satisfies (5.3), which completes the proof.

One may wonder why an asymptotically optimal open-ended test τ leads to asymptotically optimal change-point procedure N defined in (5.2) for independent observations, but does not do so for our example with dependent observations. The reasons behind this difference are given by the following heuristic

arguments. For independent observations, the open-ended test τ and the change-point detection procedure N constructed from τ by (5.2) are closely related because it is often true that

$$\mathbf{E}_f(N) \sim \frac{1}{\mathbf{P}_f(\tau < \infty)} \quad (5.8)$$

as $\mathbf{P}_f(\tau < \infty) \rightarrow 0$. Here and everywhere below, $x \sim y$ means that x/y converges to a finite positive value as y goes to ∞ or 0 . However, (5.8) fails in general for dependent observations, particularly in our example where f is defined by (2.2). To see this, first note that (5.8) holds for f_1 and f_2 because observations are independent under \mathbf{P}_{f_1} and \mathbf{P}_{f_2} . Then by (3.2),

$$\mathbf{E}_f(N) = \pi_0 \mathbf{E}_{f_1}(N) + (1 - \pi_0) \mathbf{E}_{f_2}(N) \sim \frac{\pi_0}{\mathbf{P}_{f_1}(\tau < \infty)} + \frac{1 - \pi_0}{\mathbf{P}_{f_2}(\tau < \infty)}.$$

Assume $\mathbf{P}_{f_1}(\tau < \infty)$ is much smaller than $\mathbf{P}_{f_2}(\tau < \infty)$, i.e.,

$$\mathbf{P}_{f_1}(\tau < \infty) \ll \mathbf{P}_{f_2}(\tau < \infty). \quad (5.9)$$

Then $\mathbf{E}_f(N) \sim \pi_0 / \mathbf{P}_{f_1}(\tau < \infty)$. On the other hand, by (5.5) and (5.9),

$$\mathbf{P}_f(\tau < \infty) = \pi_0 \mathbf{P}_{f_1}(\tau < \infty) + (1 - \pi_0) \mathbf{P}_{f_2}(\tau < \infty) \sim (1 - \pi_0) \mathbf{P}_{f_2}(\tau < \infty).$$

Hence, $\mathbf{E}_f(N) \mathbf{P}_f(\tau < \infty) \sim \pi_0 (1 - \pi_0) \mathbf{P}_{f_2}(\tau < \infty) / \mathbf{P}_{f_1}(\tau < \infty)$, which goes to ∞ by virtue of (5.9). Therefore, (5.8) fails for our example when observations are dependent under the pre-change distribution.

Acknowledgement

The author would like to thank his advisor, Professor Gary Lorden, for his support, guidance, and encouragement, and Professor Moshe Pollak for stimulating discussions. The author would also like to thank an associate editor and an anonymous referee for suggestions on how to improve the presentation of the paper. This work was supported in part by the National Institutes of Health under Grant R01 AI055343.

Appendix. Proof of Theorem 2

Let T be either $T_{CM}(A)$ or $T_{SR}(A)$. Note that

$$\begin{aligned} \prod_{i=k}^n \frac{g(X_i | X_1, \dots, X_{i-1})}{f(X_i | X_1, \dots, X_{i-1})} &= \left(\prod_{i=k}^n \frac{g(X_i)}{f_1(X_i)} \right) \left(\frac{\pi_0 + (1 - \pi_0) \prod_{i=1}^n \frac{f_2(X_i)}{f_1(X_i)}}{\pi_0 + (1 - \pi_0) \prod_{i=1}^{k-1} \frac{f_2(X_i)}{f_1(X_i)}} \right)^{-1} \\ &\geq \left(\prod_{i=k}^n \frac{g(X_i)}{f_1(X_i)} \right) \left(1 + \frac{1 - \pi_0}{\pi_0} \prod_{i=1}^n \frac{f_2(X_i)}{f_1(X_i)} \right)^{-1}. \end{aligned}$$

Thus, if we define a new stopping time

$$M_1 = \inf \left\{ n \geq 1 : \prod_{i=1}^n \frac{f_2(X_i)}{f_1(X_i)} \leq \log A, \text{ and } \max_{1 \leq k \leq n} \prod_{i=k}^n \frac{g(X_i)}{f_1(X_i)} \geq K_A \right\}, \tag{A.1}$$

where $K_A = A(1 + (1 - \pi_0)\pi_0^{-1} \log A)$, then $T \leq M_1$. Applying Lemma 1 (below), we have $\mathbf{E}_{f_1}(T) \leq O(A \log A)$. Similarly, $\mathbf{E}_{f_2}(T) \leq O(A \log A)$. Combining these with (3.2), we have $\mathbf{E}_f(T) \leq O(A \log A)$, which proves (3.4).

To prove (3.5), note that $T_{CM}(A) \geq T_{SR}(A)$ and $\bar{\mathbf{E}}_g(T) \geq \mathbf{E}_g(T)$, so it suffices to show that $\mathbf{E}_g(T_{SR}(A))$ satisfies (3.5). Rewrite $T_{SR}(A)$ as

$$T_{SR}(A) = \inf \left\{ n \geq 1 : S_n + \log \frac{\pi_0 W_n^{(1)} + (1 - \pi_0) W_n^{(2)}}{\pi_0 \exp(U_n) + (1 - \pi_0)} \geq \log A \right\},$$

where S_n and U_n are defined in (5.6) and, for $j = 1, 2$,

$$W_n^{(j)} = 1 + \sum_{k=1}^{n-1} \prod_{i=1}^k \frac{f_j(X_i)}{g(X_i)}.$$

Here $\sum_{k=1}^0$ is denoted 0. To find a lower bound for $\mathbf{E}_g(T_{SR}(A))$, define a new stopping time $T_A^* = \inf \left\{ n \geq 1 : S_n + \eta_n \geq \log A \right\}$, where

$$\eta_n = \log \frac{\pi_0 W_n^{(1)} + (1 - \pi_0) W_n^{(2)}}{1 - \pi_0}.$$

Clearly, $T_{SR}(A) \geq T_A^*$ since $\exp(U_n) \geq 0$. So it suffices to show that

$$\mathbf{E}_g(T_A^*) \geq (1 + o(1)) \frac{\log A}{I_2} \quad \text{as } A \rightarrow \infty. \tag{A.2}$$

We prove this inequality using the argument in the proof of Lemma 9.13 of Siegmund (1985). The crucial observation is that the increasing sequences $W_n^{(j)}, n = 1, 2, \dots$ converge to a finite random variable $W^{(j)}$ under \mathbf{P}_g for $j = 1, 2$, see the proof of Theorem 3 in Pollak (1987). Thus the increasing $\eta_n, n \geq 1$, converge to a finite random variable under \mathbf{P}_g , and they are slowly changing in the sense that $n^{-1} \max[\eta_1, \dots, \eta_n] \rightarrow 0$ in probability. The Strong Law of Large Numbers implies that $n^{-1} \max_{1 \leq k \leq n} S_k$ converges to I_2 with probability one, so by the slowly changing property of the $\eta_n, n^{-1} \max_{1 \leq k \leq n} (S_k + \eta_k) \rightarrow I_2$ in probability. For every $0 < \epsilon < 1$, let $n_1 = (1 - \epsilon)(\log A)/I_2$. Since $\log A = n_1 I_2 / (1 - \epsilon)$,

$$\begin{aligned} \mathbf{P}_g \left(T_A^* \leq (1 - \epsilon) \frac{\log A}{I_2} \right) &= \mathbf{P}_g \left(\max_{1 \leq k \leq n_1} (S_k + \eta_k) \geq \log A \right) \\ &= \mathbf{P}_g \left(\max_{1 \leq k \leq n_1} (S_k + \eta_k) \geq \frac{n_1 I_2}{1 - \epsilon} \right) \rightarrow 0 \end{aligned}$$

as $A \rightarrow \infty$. Hence,

$$\mathbf{P}_g\left(T_A^* \geq (1 - \epsilon) \frac{\log A}{I_2}\right) \rightarrow 1$$

and, therefore,

$$\mathbf{E}_g(T_A^*) \geq (1 - \epsilon + o(1)) \frac{\log A}{I_2}$$

as $A \rightarrow \infty$. Since ϵ is arbitrary, it follows that T_A^* satisfies (A.2), and the proof of (3.5) is complete.

We need the following lemma to complete the proof of Theorem 2.

Lemma 1. *As $A \rightarrow \infty$, $\mathbf{E}_{f_1}(M_1) = O(A \log A)$, where M_1 is defined in (A.1).*

Proof. Let $V_n = \sum_{i=1}^n \log(g(X_i)/f_1(X_i))$ for $n = 1, 2, \dots$, and $V_0 = 0$. If $W_n = \max_{0 \leq k \leq n-1} (V_n - V_k)$, then M_1 can be written as

$$M_1 = \inf\{n \geq 1 : U_n \geq -\log \log A \quad \text{and} \quad W_n \geq \log(K_A)\},$$

where U_n is defined in (5.6). Using an idea of Kiefer and Sacks (1963), let v_1 be the first n such that $W_n \geq \log(K_A)$, v_2 the second n such that $W_n \geq \log(K_A)$, etc. Let ϕ_t be the indicator function of the set where $U_{v_t} < -\log \log A$, $t = 1, 2, \dots$. Then as shown on page 719 of Kiefer and Sacks (1963),

$$M_1 = v_1 + \sum_{j=1}^{\infty} (v_{j+1} - v_j) \prod_{t=1}^j \phi_t.$$

Let $v_{j+1}^* - v_j$ be the first m such that $\max_{1 \leq k \leq m} (V_{m+v_j} - V_{k+v_j}) \geq \log(K_A)$. Evidently, $v_{j+1}^* - v_j \geq v_{j+1} - v_j$. Since $v_{j+1}^* - v_j$ depends on X 's whose indices are greater than v_j , it follows that $v_{j+1}^* - v_j$ is independent of ϕ_1, \dots, ϕ_j . Consequently

$$\mathbf{E}_{f_1}(M_1) \leq \mathbf{E}_{f_1}(v_1) + \sum_{j=1}^{\infty} \left[\mathbf{E}_{f_1}(v_{j+1}^* - v_j) \mathbf{E}_{f_1} \left(\prod_{t=1}^j \phi_t \right) \right].$$

Now $v_{j+1}^* - v_j$ has the same distribution as v_1 (note that in Kiefer and Sacks (1963) and the proof of Theorem 1 of the present paper, $v_{j+1}^* - v_j$ is defined to have same distribution as the first n for which $V_n \geq 0$), so

$$\mathbf{E}_{f_1}(M_1) \leq \mathbf{E}_{f_1}(v_1) \left[1 + \sum_{j=1}^{\infty} \mathbf{E}_{f_1} \left(\prod_{t=1}^j \phi_t \right) \right].$$

Observe that

$$v_1 = \inf \{n \geq 1 : W_n \geq \log(K_A)\} = \inf \left\{ n \geq 1 : \max_{1 \leq k \leq n} \prod_{i=k}^n \frac{g(X_i)}{f_1(X_i)} \geq K_A \right\},$$

which is just Page's CUSUM procedure for detecting a change in distribution from f_1 to g . Now apply the standard bounds on the ARL to false alarm, $\mathbf{E}_{f_1}(v_1) = O(K_A)$, see, for example, the proof of Theorem 1(i) on page 753 of Pollak (1987).

To estimate $\mathbf{E}_{f_1}(\prod_{t=1}^j \phi_t)$, we let σ be the *last* time $S_n < -\log \log A$. Since $v_j \geq j$, we have

$$\sum_{j=1}^{\infty} \mathbf{E}_{f_1} \left(\prod_{t=1}^j \phi_t \right) \leq \sum_{j=1}^{\infty} \mathbf{P}_{f_1}(\sigma \geq v_j) \leq \sum_{j=1}^{\infty} \mathbf{P}_{f_1}(\sigma \geq j) = \mathbf{E}_{f_1}(\sigma).$$

Since the summands in U_n have positive mean and finite variance under \mathbf{P}_{f_1} , by our assumption in (2.1), it is well known that $\mathbf{E}_{f_1}(\sigma) < \infty$, see, for example, Theorem D in Kiefer and Sacks (1963). Therefore,

$$\mathbf{E}_{f_1}(M_1) \leq \mathbf{E}_{f_1}(v_1) \left[1 + \mathbf{E}_{f_1}(\sigma) \right] = O(K_A)[1 + O(1)] = O(A \log A),$$

and the lemma is proved.

References

- Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, Englewood Cliffs.
- Beibel, M. (1997). Sequential change-point detection in continuous time when the post-change drift is unknown. *Bernoulli* **3**, 457-478.
- Fuh, C.-D. (2003). SPRT and CUSUM in hidden Markov models. *Ann. Statist.* **31**, 942-977.
- Fuh, C.-D. (2004). Asymptotic operating characteristics of an optimal change point detection in hidden Markov models. *Ann. Statist.* **32**, 2305-2339.
- Kiefer, J. and Sacks, J. (1963). Asymptotically optimum sequential inference and design. *Ann. Math. Statist.* **34**, 705-750.
- Lai, T. L. (1995). Sequential change-point detection in quality control and dynamical systems (with discussion). *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **57**, 613-658.
- Lai, T. L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory* **44**, 2917-2929.
- Lai, T. L. (2001). Sequential analysis: some classical problems and new challenges. *Statist. Sinica* **11**, 303-408.
- Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* **42**, 1897-1908.
- Mei, Y. (2003). Asymptotically optimal methods for sequential change-point detection. Ph.D. dissertation, Calif. Inst. Technol., Pasadena, CA. Available at <http://etd.caltech.edu/etd/available/etd-05292003-133431/>.

- Mei, Y. (2006). Sequential change-point detection when unknown parameters are present in the pre-change distribution. *Ann. Statist.* **34**, 92-112.
- Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14**, 1379-1387.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100-115.
- Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13**, 206-227.
- Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Statist.* **15**, 749-779.
- Pollak, M. and Siegmund, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3**, 1267-1282.
- Ritov, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18**, 1464-1469.
- Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics* **8**, 411-430.
- Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41**, 1397-1409.
- Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the Wiener process and partial sums. *Ann. Math. Statist.* **41**, 1410-1429.
- Robbins, H. and Siegmund, D. (1973). Statistical tests of power one and the integral representation of solutions of certain partial differential equations. *Bull. Inst. Math. Acad. Sinica* **1**, 93-120.
- Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.* **8**, 22-46.
- Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals*. Springer-Verlag, New York.
- Woodroffe, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*, Society for Industrial and Applied Mathematics, Philadelphia.
- Yakir, B. (1994). Optimal detection of a change in distribution when the observations form a Markov chain with a finite state space. In *Change-Point Problems* (Edited by E. Carlstein, H. Muller and D. Siegmund), 346-358. Inst. Math. Statist., Hayward, CA.
- Yakir, B., Krieger, A. M. and Pollak, M. (1999). Detecting a change in regression: first order optimality. *Ann. Statist.* **27**, 1896-1913.

School of Industrial and Systems Engineering, Georgia Institute of Technology, 765 Ferst Drive NW, Atlanta, GA 30332-0205, U.S.A.

E-mail: ymei@isye.gatech.edu

(Received January 2004; accepted July 2005)