MARTINGALE TYPE STATISTICS APPLIED TO CHANGE POINTS DETECTION

Albert Vexler
The State University of New York
Department of Biostatistics
Farber Hall, Rm 246
Buffalo, NY 14214
avexler@buffalo.edu

Key Words: change point; CUSUM statistics; epidemic alternative; martingale; maximum likelihood; Shiryayev-Roberts statistics.

AMS 2000 *subject classifications.* Primary: 62F03; Secondary: 62H15, 62F05.

ABSTRACT

The Shiryayev-Roberts approach has been adapted to detection of various types of changes in distributions of non-i.i.d. observations. By utilizing martingale properties of Shiryayev-Roberts statistics, this technique provides distribution-free non-asymptotic upper bounds for the significance levels of asymptotic power one tests for change points with epidemic alternatives. Since optimal Shiryayev-Roberts sequential procedures are well-investigated, the proposed methodology yields a simple approach for obtaining analytical results related to retrospective testing. In the case when distributions of data are known up to parameters, the paper presents an adaptive estimation that is more efficient than a well-accepted non-anticipating estimation described in the change point literature. The proposed adaptive procedure can also be used in the context of sequential change point detection.

1 INTRODUCTION

The observed sample is $\{Y_i\}_{i=1}^n$, where Y_i has conditional density $f_{i\,\theta_i}(u|Y_1,\ldots,Y_{i-1})$,

$$f_{i\,\theta_i}(u|Y_1,\ldots,Y_{i-1}) = \begin{cases} f_{0\,\theta_0}(u|Y_1,\ldots,Y_{i-1}), & i<\nu; \\ f_{1\,\theta_1}(u|Y_1,\ldots,Y_{i-1}), & i\in[\nu,\gamma); \\ f_{0\,\theta_0}(u|Y_1,\ldots,Y_{i-1}), & i\geq\gamma, \end{cases} \tag{1.1}$$
$$(\nu<\gamma,\ i=1,\ldots,n),$$

with parameters θ_0 and θ_1. (We will assume here that Y_i is independent of future realizations of the random variables $\{Y_k\}_{k>i}$.) In accordance with the definition (1.1), we suppose that if $\nu\leq n$ then an epidemic state of data runs from time ν through $\gamma-1$ after which (if $\gamma<n$) the normal state is restored.

Specific cases of (1.1) have been widely addressed in the literature, both applied and theoretical. Yao (1993), Ramanayake and Gupta (2003) as well as Vexler (2006) have considered the model (1.1) that corresponds to independently distributed $Y_1,\ldots,Y_n$; the definition (1.1), where $Y_1,\ldots,Y_n$ are non-independent was introduced in a sequential context by Lai (1995), Yakir and Pollak (1998) as well as Tartakovsky and Veeravalli (2005); etc. For other applications and investigations of such models in technology, econometrics and biostatistics see, for example, Brown *et al.* (1975), Broemeling and Tsurumi (1987), Hansen (2000), Braun *et al.* (2000) as well as Koul *et al.* (2003).

Obviously, use of the model (1.1) generally requires derivation of statistics to test if density functions of Ys are subject to an epidemic state. Thus, obtaining a test for the hypothesis,

$$\mathbf{H_0}\ :\ \nu>n, \quad \text{versus} \quad \mathbf{H_1}\ :\ 1\leq\nu\leq n; \quad \nu,\gamma:\ 0<\nu<\gamma \text{ are unknown}, \tag{1.2}$$

must be addressed. Note that, If $\gamma>n$ then (1.2) is a standard change point problem (e.g. Lai, 1995; Lai, 2001; Julious, 2001; Gurevich and Vexler, 2005 as well as Vexler and Gurevich, 2006). The paper essentially considers a retrospective statement of problem (1.1), (1.2), i.e. the number of observations n is fixed.

Since, in general, tests based on the maximum likelihood ratio have high power (e.g. Lai, 1995), we propose and examine a class of generalized maximum likelihood asymptotic power one tests for various types of problem (1.1), (1.2). It is widely known in change point literature that evaluation of the significance level of generalized maximum likelihood tests is a major and complicated issue. Commonly, in the context of problem (1.1), (1.2) with i.i.d. observations prior to and after the change, the asymptotic ($n \to \infty$) significance level of the test can be derived upon special conditions on distribution functions of $Y_1, \ldots, Y_n$ and the epidemic state's parameter θ_1 (e.g. Yao, 1993; Ramanayake and Gupta, 2003; Vexler, 2006). Alternatively, for a specific n and case of (1.1) (1.2) (e.g. Y_i, $i = 1, \ldots, n$ are independent and θ_0 is known), the significant level can be estimated by Monte Carlo simulations (e.g. Chan and Lai, 2005). However even in these situations, obtaining a guaranteed non-asymptotic upper bound for the significance level has practical meaning (e.g. Krieger, *et al.*, 2003; Chan and Lai, 2005; Gurevich and Vexler, 2005; Vexler and Gurevich, 2006). It is clear that if the significance level of the test is asymptotically in proportional to its upper bound then that reasonably ensures a *p-value* of the test.

In the context of specific linear regression forms of (1.1), Vexler (2006) has adapted the methods, which were developed to solve change point problems in sequential analysis (Yakir *et al.*, 1999; Krieger, *et al.*, 2003; Lorden and Pollak, 2005), and proposed tests for problem (1.2) based on the Shiryayev-Roberts approach. The main objective of the paper is an extension of this methodology to the case when no model assumptions on (1.1) are assumed and Ys are dependent observations. The proposed tests have a guaranteed non-asymptotic upper bound for significance level, that ensures a *p-value*. The proof of this result utilizes a martingale structure of the test-statistics. Since the upper bound does not depend on the densities of $Y_1, \ldots, Y_n$, nothing can be assumed about these densities. In particular cases of the stated problem, asymptotic results demonstrated closeness of the upper bound to the significance levels of the tests that are obtained in this paper.

Note that, the problem of preserving a martingale structure of test-statistics especially arises in a variety of change point detection issues where density functions of observations have to be estimated. The non-anticipating estimation is functional in schemes of a change point detection (e.g. Robbins and Siegmund, 1973; Dragalin, 1997; Krieger, *et al.* 2003; Lorden and Pollak, 2005; Chan and Lai, 2005; Gurevich and Vexler, 2005; Vexler and Gurevich, 2006 as well as Vexler, 2006). However, this manner of estimation sometimes leads to the information related to the observed sample being disregarded (Lai, 2001). To relax this losing of information, an alternative approach is proposed. This suggested adaptive technique can also be used in sequential change point detection procedures.

The paper is organized as follows. Section 2 introduces the simple case of problem (1.1), (1.2), where parameter θ_0 of the stable state and parameter θ_1 of the epidemic state of model (1.1) are known. The analysis of this section is relatively clear, has basic ingredients for more general cases, and is the obvious case for the paper's issue. Section 3 assumes that the parameters of the epidemic state are unknown. The adaptive procedures are represented in this section. When all parameters of model (1.1) are unknown, Section 4 presents the test-statistic for (1.2). We point out remarks in Section 5. Section 6 demonstrates the results of Monte Carlo simulations.

2 SIMPLE CASE OF THE PROBLEM

Let P_0 and D_0 denote the probability measure and density of P_0 under H_0 as well as P_{km} and D_{km} denote the probability measure and respective density under H_1 with $\nu = k, \gamma = m$. Likewise, let E_0 and E_{km} denote expectation under P_0 and P_{km}, respectively. Assume the parameters θ_0 and θ_1 of model (1.1) are known. Hence the likelihood ratio is,

$$\Lambda_{km} \quad = \quad \frac{D_{km}(Y_1, \dots, Y_n)}{D_0(Y_1, \dots, Y_n)} = \prod_{i=k}^{m} \frac{f_{1\,\theta_1}(Y_i|Y_1, \dots, Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1, \dots, Y_{i-1})}. \tag{2.1}$$

The maximum likelihood methodology applied to the problem (1.2) (where ν and γ are unknown) provides the modified CUSUM test-statistic $\max_{1 \le m \le n} \max_{1 \le k \le m} \Lambda_{km}$ (e.g. Yao, 1993). Alternatively, we propose the test, which is based on the Shiryaev-Roberts technique, in the form of

$$\frac{1}{n} \max_{1 \le m \le n} R_m \quad > \quad C, \tag{2.2}$$

where $R_m = \sum_{k=1}^{m} \Lambda_{km}$, Λ_{km} is defined by (2.1) and $C > 0$ is the threshold value. The same idea appears in Vexler (2006) in the context of testing for epidemic changes of a specific linear regression model. Note that, commonly, sequential procedures based on the Shiryaev-Roberts approach are optimal methods of detecting a change in distribution of observations (e.g. Pollak, 1985, 1987).

Example: Consider, for instance, model (1.1) in the autoregressive form

$$Y_i = \theta_i' Y_{i-1} + \varepsilon_i, \;\; Y_0 = 0, \quad i = 1, \ldots, n, \tag{2.3}$$

with parameters $\theta_i' = \theta_0 I\{i < \nu\} + \theta_1 I\{\nu \le i < \gamma\} + \theta_0 I\{i \ge \gamma\}$ (θ_0 is parameter of the stable state; θ_1 is parameter of the epidemic state of this model; $I\{\cdot\}$ is the indicator function) and independent random disturbance terms ε_i have a density $f(u)$. In this case, we have $\Lambda_{km} = \prod_{i=k}^{m} f\left(Y_i - \theta_1 Y_{i-1}\right) / f\left(Y_i - \theta_0 Y_{i-1}\right)$ and the test-statistic (2.2).

Significance level of test (2.2). Commonly, change point detection procedures based on the Shiryaev-Roberts method have some guaranteed H_0-characteristics. Here we consider a guaranteed non-asymptotic upper bound for the significance level of the proposed test. Since

$$\begin{aligned} E_0 \Lambda_{km} \Big| \{Y_1, \ldots, Y_{m-1}\} &= \Lambda_{km-1} E_0 \frac{f_{1\,\theta_1}\left(Y_m | Y_1, \ldots, Y_{m-1}\right)}{f_{0\,\theta_0}\left(Y_m | Y_1, \ldots, Y_{m-1}\right)} \Big| \{Y_1, \ldots, Y_{m-1}\} \\ &= \Lambda_{km-1} \int \frac{\frac{dP_{km}(u|Y_1,\ldots,Y_{m-1})}{du}}{\frac{dP_0(u|Y_1,\ldots,Y_{m-1})}{du}} \frac{dP_0\left(u|Y_1, \ldots, Y_{m-1}\right)}{du} du = \Lambda_{km-1} \\ \text{and} \qquad E_0 \Lambda_{km} &= \int \cdots \int \frac{D_{km}\left(u_1, \ldots, u_m\right)}{D_0\left(u_1, \ldots, u_m\right)} D_0\left(u_1, \ldots, u_m\right) du_1 \ldots du_m = 1, \end{aligned}$$

the sequence $\{R_m - m\}$ is a martingale under P_0 with zero expectation. Therefore, we have

Proposition 2.1 *The significance level α of the test satisfies:*

$$\alpha \equiv P_0 \left\{ \max_{1 \le m \le n} R_m > nC \right\} \le 1/C.$$

(The proof scheme of Proposition 2.1 of Vexler (2006) is appropriate to details.) Thus, fixing $C = 1/\alpha$ determines the test with the level of significance that does not exceed α.

One can show that when: $Y_1, \ldots, Y_n$ are independent random variables from an exponential family; $f_{0\,\theta_0}(u)$ is a density of some probability measure on a sample space; $f_{1\,\theta_1}(u) = \exp(\theta_1 u - \psi(\theta_1)) f_{0\,\theta_0}(u)$, $\theta_1 \in \Omega$, (Ω is an interval on which $\psi(.)$ is finite, $\theta_1 \psi'(\theta_1) - \psi(\theta_1) < \infty$), we obtain the asymptotic result

Proposition 2.2 *Let $n^* = n^*(C)$ be a sequence that satisfies $\ln(C)/n^* \to 0$ as $C \to \infty$ and $P_{1\infty}$ is nonlattice. Then*

$$\lim_{C\to\infty} C P_0 \left\{ \max_{1\le m\le n^*} R_m > n^* C \right\} = \int_0^\infty e^{-x} dH(x) \le 1,$$

where $H(x)$ is the asymptotic distribution, under the regime $P_{1\infty}$, of the overshoot $\sum_{i=1}^{M(A)} (\theta_1 Y_i - \psi(\theta_1)) - \ln(A)$, $M(A) = inf\left\{ l : \exp\left(\sum_{i=1}^{l} (\theta_1 Y_i - \psi(\theta_1)) \right) \ge A \right\}$ as $A \to \infty$ (see Vexler, 2006).

In the considered case, Proposition 2.2 illustrates the closeness of the true significance level to the upper bound $1/C$.

Power of the test. Without loss of generality and for the sake of clarity of exposition, we assume that random variables $Z_i \equiv \frac{f_{1\,\theta_1}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}$, $i \in [\nu, \min(\gamma-1, n)]$ are independent under H_1 and $a_i \equiv E_{\nu\gamma} \ln(Z_i)$, $i \in [\nu, \min(\gamma-1,n)]$. Since

$$\max_{1\le m\le n} R_m \ge \sum_{k=\nu}^{\min(\gamma-1,n)} \exp\left(\sum_{i=k}^{\min(\gamma-1,n)} \ln(Z_i) \right),$$

where Zs have the nonnegative H_1-expectation:

$$\begin{aligned} a_i &= E_{\nu\gamma} E_{\nu\gamma} \left(\ln(Z_i) \middle| Y_1, \ldots, Y_{i-1} \right) = -E_{\nu\gamma} E_{\nu\gamma} \left(\ln(1/Z_i) \middle| Y_1, \ldots, Y_{i-1} \right) \\ &\ge -E_{\nu\gamma} \ln \left(E_{\nu\gamma} (1/Z_i \middle| Y_1, \ldots, Y_{i-1}) \right) \\ &= -E_{\nu\gamma} \ln \left(\int \frac{f_{0\,\theta_0}(u|Y_1,\ldots,Y_{i-1})}{f_{1\,\theta_1}(u|Y_1,\ldots,Y_{i-1})} f_{1\,\theta_1}(u|Y_1,\ldots,Y_{i-1})\, du \right) = 0, \end{aligned}$$

by applying almost directly the proof scheme of Vexler's (2006) Proposition 2.3, we conclude that the probability of Type II error of test (2.2) is bounded by an exponentially vanishing (as $\min(\gamma-1,n) - \nu \to \infty$ and C is fixed) term, i.e.

Proposition 2.3 *Assume that: for some $T > 0$, a sequence $\{s_i \in (0, \infty),\ i = \nu, \ldots, \min(\gamma - 1, n)\}$ and some $\delta_{\nu\gamma} > 0$:*

$$e^{ta_i} E_{\nu\gamma} |Z_i|^{-t} \le e^{\frac{1}{2} s_i t^2} < \infty,\ \ 0 \le t \le T,\ \ i \in [\nu, \min(\gamma - 1, n)];$$
$$nC < e^{-\delta_{\nu\gamma}} \sum_{k=\nu}^{\min(\gamma-1,n)} e^{\sum_{i=k}^{\min(\gamma-1,n)} a_i}.$$

Then the probability of Type II error satisfies the inequality

$$P_{\nu\gamma}\left\{ \max_{1 \le m \le n} R_m \le nC \right\} \ \le \ \begin{cases} e^{-\frac{\delta^2_{\nu\gamma}}{2S_{\nu\gamma}}}, & \text{if } \delta_{\nu\gamma} \le S_{\nu\gamma} T, \\ e^{-\frac{\delta_{\nu\gamma} T}{2}}, & \text{if } \delta_{\nu\gamma} > S_{\nu\gamma} T, \end{cases}$$

where $S_{\nu\gamma} = \sum_{i=\nu}^{\min(\gamma-1,n)} s_i$.

When $a_i = a_j = a > 0$, Vexler (2006) considered examples that illustrate applications of this proposition. A case, where $a_i \neq a_j,\ i \neq j$ is related to Remark 2 of Section 5.

3 ADAPTIVE PROCEDURES

Section 3 considers a more complicated case, where the parameter of the stable state θ_0 is known and the parameter of the epidemic state θ_1 is unknown. This case has been widely discussed in the literature (e.g. Yakir *et al.*, 1999; Krieger, *et al.*, 2003; Lorden and Pollak, 2005; as well as Gurevich and Vexler, 2005) and corresponds to a control problem where the baseline in-control distribution is known and the epidemic state out-of-control distribution is not. Frequently, in practical applications, θ_0 has zero value.

3.1 Procedure Based on Non-Anticipating Estimation

Denote the estimator of the likelihood ratio (2.1) in the form of

$$\Lambda^{(1)}_{km} \ = \ \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,i-1)}}(Y_i|Y_1, \ldots, Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1, \ldots, Y_{i-1})}, \tag{3.1}$$

where $\widehat{\theta}^{(k,i-1)}$ is an (any reasonable) estimator of θ based upon observations $\{Y_1, \ldots, Y_{i-1}\}$ under the putative $\nu = k$ (say, $\widehat{\theta}^{(k,k-1)}$ is fixed as, for example, $\widehat{\theta}^{(k,k-1)} = \theta_0$). Since $\widehat{\theta}_1^{(k,i-1)}$

does not depend on Y_i, the structure of the non-anticipating estimators $\left\{\widehat{\theta}_1^{(k,i-1)}, \, i \geq k\right\}$ preserves the H_0-martingale property of $R_m^{(1)} - m$, where

$$R_m^{(1)} \;=\; \sum_{k=1}^{m} \Lambda_{km}^{(1)}, \;\; m = 1, \ldots, n \tag{3.2}$$

(i.e. $E_0\left(R_m^{(1)} - m \Big| \{Y_1, \ldots, Y_{m-1}\}\right) = R_{m-1}^{(1)} - (m-1)$ and $E_0\left(R_m^{(1)} - m\right) = 0$).

In the context of example (2.3), we have $\Lambda_{km}^{(1)} = \prod_{i=k}^{m} f\left(Y_i - \widehat{\theta}_1^{(k,i-1)} Y_{i-1}\right) / f\left(Y_i - \theta_0 Y_{i-1}\right)$ (e.g. $\widehat{\theta}_1^{(k,i-1)}$ is the maximum likelihood estimator $\widehat{\theta}_1^{(k,i-1)} = \arg\max_\theta \prod_{j=k}^{i-1} f\left(Y_j - \theta Y_{j-1}\right)$, $\widehat{\theta}_1^{(k,k-1)} = \theta_0$), and hence

$$\begin{aligned}
& E_0(R_m^{(1)} - m)\Big|\{Y_1, \ldots, Y_{m-1}\} = \sum_{k=1}^{m-1} E_0 \Lambda_{km}^{(1)} \Big| \{Y_1, \ldots, Y_{m-1}\} - (m-1) \\
& = \sum_{k=1}^{m-1} \Lambda_{km-1}^{(1)} E_0 \frac{f\left(Y_m - \widehat{\theta}_1^{(k,m-1)} Y_{m-1}\right)}{f\left(Y_m - \theta_0 Y_{m-1}\right)} \Big| \{Y_1, \ldots, Y_{m-1}\} - (m-1) \\
& = R_{m-1}^{(1)} - (m-1).
\end{aligned}$$

Consequently, the proposed test is: reject H_0 iff

$$\frac{1}{n} \max_{1 \leq m \leq n} R_m^{(1)} \;>\; C, \quad \text{where } R_m^{(1)} \text{ by (3.2)}. \tag{3.3}$$

Consider the significance level of the test. Since the sequence $\{R_m^{(1)} - m\}$ is a martingale under P_0, it is clear that

Proposition 3.1 *The significance level $\alpha^{(1)}$ of the test satisfies:*

$$\alpha^{(1)} \equiv P_0 \left\{ \max_{1 \leq m \leq n} R_m^{(1)} > nC \right\} \leq 1/C.$$

In order to make an analogy to Proposition 2.2, assume that observations $Y \sim Gamma(\theta, 1)$ are independent and note that the $Gamma(\theta, 1)$ family can be transformed into an exponential family with canonical form: if $Y \sim Gamma(\theta, 1)$ then a reparameterization $\rho = \theta - \theta_0$ and an appropriate affine transformation Y^* of $\log Y$ yields a family of probability measures of Y^* with densities

$$f_\rho(u) = \exp(\rho u - \psi(\rho)) f_0(u), \;\; \rho \in (-\theta_0, \infty),$$

where $\psi'(0) = 0$. Following Lorden and Pollak (2005), define the estimator $\widehat{\rho}^{(k,i-1)}$ (for the parameter of i-th observation under the putative $\nu = k$) $\widehat{\rho}^{(k,i-1)} = (\sum_{j=k}^{i-1} Y_j + s)/(i-k+t) - \theta_0$ for given $s, t \geq 0$, where $\widehat{\rho}^{(k,k-1)} = 0$ if $s \wedge t = 0$. Let

$$Z_i^{\rho} = \rho Y_i^* - \psi(\rho),\ \Lambda_{km}^{(1)} = \exp\left(\sum_{i=k}^{m} Z_i^{\widehat{\rho}^{(k,i-1)}}\right),\ R_m^{(1)} = \sum_{k=1}^{m} \Lambda_{km}^{(1)},$$
$$\tau^{(1)}(A) = \inf\left\{l > 1 : \quad R_l^{(1)} \geq A\right\},\ M^{(1)}(A) = \inf\left\{l > 1 : \quad \Lambda_{1l}^{(1)} \geq A\right\}.$$

Proposition 3.2 *Let $n^* = n^*(C)$ be a sequence that satisfies $\ln(C)/n^* \to 0$ as $C \to \infty$. Then*

$$\lim_{C\to\infty} \frac{P_0\left\{\max_{1\leq m\leq n^*} R_m^{(1)} > n^* C\right\}}{P_0\left\{\max_{1\leq m\leq n^*} \sum_{i=1}^{m} Z_i^{\widehat{\rho}^{(1,i-1)}} > \ln(C)\right\}} = 1, \text{ and}$$
$$\lim_{C\to\infty} C P_0\left\{\max_{1\leq m\leq n^*} R_m^{(1)} > n^* C\right\} = \int_0^{\infty} e^{-x} dH^{(1)}(x) \leq 1,$$

where $H^{(1)}(x)$ the asymptotic distribution, under the measure $P_{1\infty}$, of the overshoot $\sum_{i=1}^{M^{(1)}(A)} Z_i^{\widehat{\rho}^{(1,i-1)}} - \ln(A)$, as $A \to \infty$.

Proof. The proof is based on an analogy of Yakir's (1995) Lemma 1. This result is obtained in Lorden and Pollak (2005) (see Appendix 1: Sketch of Proof of Theorem 4; Lemma Y1). According to the lemma, we have

$$\frac{P_0\{\tau^{(1)}(A) \leq m\}}{P_0\{M^{(1)}(A/m) \leq m\}} = 1 + o_A(1),\ P_0\left\{M^{(1)}\left(\frac{A}{m}\right) \leq m\right\} = \frac{m}{A}\int_0^{\infty} e^{-x} dH^{(1)}(x)(1 + o_A(1)),$$

where $A = Cn^*$, $m = n^*$ and $o_A(1) \to 0$ (as $A \to \infty$) are defined. From which the proof of Proposition 3.2 follows.

3.2 Weighted Shiryayev-Roberts Procedure

In accordance with Lai (2001, p. 398), the adaptive estimators $\widehat{\theta}_1^{(k,i-1)}$, which replace the unknown θ_1, are easy to use, however an accuracy of this substitution suffers from the loss of information due to ignoring $Y_i, \ldots, Y_m$ that are also available under the putative $\gamma = m$. This loss of information is further intensified when θ_1 is the vector of unknown parameters.

To this end, we consider an alternative of test (3.3). Suppose that a test statistic is based on components in the form of $\tilde{\Lambda}_{km} = \prod_{i=k}^{m} f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)/f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)$, $k \le m = 1,\ldots,n$, where $\widehat{\theta}_1^{(k,m)}$ is the maximum likelihood estimator of θ_1 based upon $Y_k,\ldots,Y_m$ observations. By the definition of the maximum likelihood estimation,

$$
\begin{aligned}
E_0 \prod_{i=k}^{m} \tilde{\Lambda}_{km}\Big| \{Y_1,\ldots,Y_{m-1}\} &\ge E_0 \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}\Big| \{Y_1,\ldots,Y_{m-1}\} \\
= \prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)} & E_0 \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)}{f_{0\,\theta_0}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)}\Big| \{Y_1,\ldots,Y_{m-1}\} = \tilde{\Lambda}_{k\,m-1}
\end{aligned}
$$

whereas $E_0\Lambda_{km}\Big| \{Y_1,\ldots,Y_{m-1}\} = \Lambda_{k\,m-1}$ (i.e. $\tilde{\Lambda}_{km}$ is the bias estimator of Λ_{km}). Therefore, that motives to adjust the test-statistic in order to reduce the bias.

Since the stable state parameter is known, we propose the test: reject H_0 iff

$$
\frac{1}{n}\max_{1\le m\le n} R_m^{(1a)} \;>\; C, \qquad R_m^{(1a)} = \sum_{k=1}^{m}\prod_{i=k}^{m} \omega_{ki}^{(1)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}, \tag{3.4}
$$

$\widehat{\theta}_1^{(k,m)}$ is defined in (3.1), the weights $\omega_{km}^{(1)}$ are such that

$$
\begin{aligned}
\omega_{kk}^{(1)} &= \left(\frac{f_{1\,\widehat{\theta}_1^{(k,k)}}\left(Y_k|Y_1,\ldots,Y_{k-1}\right)}{f_{0\,\theta_0}\left(Y_k|Y_1,\ldots,Y_{k-1}\right)}\right)^{-1}, \\
\prod_{i=k}^{m}\omega_{ki}^{(1)} &= \left(\prod_{i=k}^{m-1}\omega_{ki}^{(1)}\right)\prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)} \\
&\quad\times\left(E_0\prod_{i=k}^{m}\frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}\Big|\{Y_1,\ldots,Y_{m-1}\}\right)^{-1}, \ \prod_{k}^{k-1} = 1.
\end{aligned}
$$

Obviously, definition (3.4) preserves the H_0-martingale property of $R_m^{(1a)} - m$, however calculation of the weights $\omega^{(1)}$ is the complex problem. To relax this calculation, we denote the estimators $\widehat{\theta}_1^{(k,m)} = \arg\max_\theta \prod_{j=k}^{m} f_{1\,\theta}\left(Y_j|Y_1,\ldots,Y_{j-1}\right)$, therefore

$$
\begin{aligned}
\prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)} &= \prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)}{f_{0\,\theta_0}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)} \\
\le \left[\prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}{f_{0\,\theta_0}\left(Y_i|Y_1,\ldots,Y_{i-1}\right)}\right] & \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)}{f_{0\,\theta_0}\left(Y_m|Y_1,\ldots,Y_{m-1}\right)}
\end{aligned}
$$

and hence, if

$$\omega_{ki}^{(2)} = \begin{cases} \left(E_0 \frac{f_{1\,\widehat{\theta}_1^{(k,i)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \Big| \{Y_1,\ldots,Y_{i-1}\}\right)^{-1}, & k < i, \\ \left(\frac{f_{1\,\widehat{\theta}_1^{(k,k)}}(Y_k|Y_1,\ldots,Y_{k-1})}{f_{0\,\theta_0}(Y_k|Y_1,\ldots,Y_{k-1})}\right)^{-1}, & i = k, \end{cases}$$

then

$$R_m^{(1a1)} - m \equiv \sum_{k=1}^{m} \prod_{i=k}^{m} \omega_{ki}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} - m \tag{3.5}$$

is a H_0-supermartingale, because

$$\begin{aligned}
& E_0\left(R_m^{(1a1)} - m\right) \Big| \{Y_1,\ldots,Y_{m-1}\} \\
&= \sum_{k=1}^{m-1} E_0 \prod_{i=k}^{m} \omega_{ki}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \Big| \{Y_1,\ldots,Y_{m-1}\} - (m-1) \\
&= \sum_{k=1}^{m-1} E_0 \prod_{i=k}^{m} \omega_{ki}^{(2)} \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \Big| \{Y_1,\ldots,Y_{m-1}\} - (m-1) \\
&\leq \sum_{k=1}^{m-1} E_0 \prod_{i=k}^{m} \omega_{ki}^{(2)} \prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \\
&\times \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_m|Y_1,\ldots,Y_{m-1})}{f_{0\,\theta_0}(Y_m|Y_1,\ldots,Y_{m-1})} \Big| \{Y_1,\ldots,Y_{m-1}\} - (m-1)
\end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{m-1} E_0 \prod_{i=k}^{m-1} \omega_{ki}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \omega_{km}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_m|Y_1,\ldots,Y_{m-1})}{f_{0\,\theta_0}(Y_m|Y_1,\ldots,Y_{m-1})} \Big| \{Y_l\}_{l=1}^{m-1} \\
&-(m-1) = \sum_{k=1}^{m-1} \prod_{i=k}^{m-1} \omega_{ki}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \\
&\times E_0 \omega_{km}^{(2)} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_m|Y_1,\ldots,Y_{m-1})}{f_{0\,\theta_0}(Y_m|Y_1,\ldots,Y_{m-1})} \Big| \{Y_1,\ldots,Y_{m-1}\} - (m-1) = R_{m-1}^{(1a1)} - (m-1).
\end{aligned}$$

Thus, the alternative test is: reject H_0 iff

$$\frac{1}{n} \max_{1 \leq m \leq n} R_m^{(1a1)} > C, \tag{3.6}$$

By virtue of the H_0-martingale property of the tests statistics, we obtain

Proposition 3.3 *The significance levels $\alpha^{(1a)}$ and $\alpha^{(1a1)}$ of the tests (3.4) and (3.6) satisfies:*

$$\alpha^{(1a)} \equiv P_0\left\{\max_{1\le m\le n} R_m^{(1a)} > nC\right\} \;\le\; 1/C,$$
$$\alpha^{(1a1)} \equiv P_0\left\{\max_{1\le m\le n} R_m^{(1a1)} > nC\right\} \;\le\; 1/C.$$

Examples: 1. Assume that when the process being monitored is in control, it yields independent normally distributed observations, and when the process is out of control only the mean-parameter changes. Let the initial observations $Y_1, \ldots, Y_{\nu-1} \sim \mathbf{N}(0,1)$. Assume $\widehat{\theta}_1^{(k,m)} := \overline{Y}_{k\,m} = \sum_k^m Y_j/(m-k+1)$ is the maximum likelihood estimator. Following the definition of test (3.4), we need to obtain the conditional expectation

$$E_0 \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}\Big|\{Y_l\}_{l=1}^{m-1} = E_0 e^{\left(-\frac{1}{2}\sum_{i=k}^m (Y_i-\overline{Y}_{k\,m})^2+\frac{1}{2}\sum_{i=k}^m Y_i^2\right)}\Big|\{Y_l\}_{l=1}^{m-1}$$
$$= \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{\infty} \exp\left(\frac{m-k+1}{2}\left(\frac{u}{m-k+1}+\overline{Y}_{k\,m-1}\frac{m-k}{m-k+1}\right)^2-\frac{u^2}{2}\right)du$$
$$= \left(\frac{m-k+1}{m-k}\right)^{1/2}\exp\left(\overline{Y}^2_{k\,m-1}\frac{m-k}{2}\right).$$

Therefore, we have

$$\begin{aligned}\omega_{km}^{(1)} &= \exp\left(-\frac{1}{2}\sum_{i=k}^{m-1}(Y_i-\overline{Y}_{k\,m-1})^2+\frac{1}{2}\sum_{i=k}^{m-1}Y_i^2-\overline{Y}^2_{k\,m-1}\frac{m-k}{2}\right)\left(\frac{m-k}{m-k+1}\right)^{1/2}\\ &= \left(\frac{m-k}{m-k+1}\right)^{1/2},\ m>k.\end{aligned}$$

That is, we obtain

$$\begin{aligned}R_m^{(1a)} &= \sum_{k=1}^{m}\prod_{i=k}^{m}\omega_{ki}^{(1)} e^{\left(-\frac{1}{2}(Y_i-\overline{Y}_{k\,m})^2+\frac{1}{2}Y_i^2\right)} = \sum_{k=1}^{m}\left(\prod_{i=k}^{m}\omega_{ki}^{(1)}\right)e^{\left(\frac{m-k+1}{2}\overline{Y}^2_{k\,m}\right)}, \qquad (3.7)\\ \omega_{ki}^{(1)} &= \left(\frac{i-k}{i-k+1}\right)^{1/2} I\{i>k\}+e^{-Y_k^2/2}I\{i=k\},\end{aligned}$$

where the weight $\omega^{(1)}$ is the compensator of applying of the maximum likelihood estimation.

2. Let Ys be independent observations; the pre-change distribution be standard exponential, i. e. $\theta_0 = 1$; the distribution of the epidemic state be exponential with the unknown parameter θ_1, i.e. $f_{1\,\theta_1}(u) = \theta_1 \exp(-\theta_1 u)$; the estimator $\widehat{\theta}_1^{(k,m)} := 1/\overline{Y}_{k\,m} = (m-k+1)/\sum_k^m Y_j$ be the maximum likelihood estimator. Since

$$E_0 \prod_{i=k}^m \frac{f_{1\,\widehat{\theta}_1^{(k,m)}}(Y_i|Y_1,\dots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\dots,Y_{i-1})}\Big|\{Y_l\}_{l=1}^{m-1} = E_0 \frac{1}{\left(\overline{Y}_{k\,m}\right)^{m-k+1}}\, e^{\left(-\frac{\sum_{i=k}^m Y_i}{\overline{Y}_{k\,m}} + \sum_{i=k}^m Y_i\right)}\Big|\{Y_l\}_{l=1}^{m-1}$$
$$= (m-k+1)^{m-k+1} e^{(m-k)\overline{Y}_{k\,m-1}-(m-k+1)} \int_0^\infty \frac{1}{(u+(m-k)\overline{Y}_{k\,m-1})^{m-k+1}} du$$
$$= \left(\frac{m-k+1}{m-k}\right)^{m-k+1} \frac{\exp\left((m-k)\overline{Y}_{k\,m-1} - (m-k+1)\right)}{\left(\overline{Y}_{k\,m-1}\right)^{m-k}},$$

we derive the statistics of test (3.4) as

$$\begin{aligned}
\omega_{ki}^{(1)} &= \left(\frac{i-k}{i-k+1}\right)^{i-k+1} eI\{i>k\} + Y_k e^{1-Y_k} I\{i=k\},\\
R_m^{(1a)} &= \sum_{k=1}^m \left(\prod_{i=k}^m \omega_{ki}^{(1)}\right) \frac{1}{\left(\overline{Y}_{k\,m}\right)^{m-k+1}} e^{(m-k+1)\left(\overline{Y}_{k\,m}-1\right)}.
\end{aligned}$$

3. Write the model (1.1) in the form of the simple segmented linear regression

$$Y_i = \theta_1 x_i I\{\nu \le i < \gamma\} + \varepsilon_i,\ \nu<\gamma,\ i=1,\dots,n, \tag{3.8}$$

where θ_1 is unknown regression parameter, x_i are fixed predictors, ε_i are independent random disturbance terms with the standard normal density. Define the maximum likelihood estimator of the unknown parameter

$$\widehat{\theta}_1^{(k,m)} = \frac{\sum_{j=k}^m Y_j x_j}{\sum_{j=k}^m x_j^2}. \tag{3.9}$$

According to the methodology proposed in this section, we obtain (similarly to the first example) the conditional expectation of the estimator of the likelihood ratio

$$E_0 \prod_{i=k}^m \frac{\exp\left(-\frac{1}{2}\left(Y_i - \widehat{\theta}_1^{(k,m)} x_i\right)^2\right)}{\exp\left(-\frac{1}{2}Y_i^2\right)}\Big|\{Y_l\}_{l=1}^{m-1} = \left(\frac{\sum_{j=k}^m x_j^2}{\sum_{j=k}^{m-1} x_j^2}\right)^{1/2} \exp\left(\frac{\left(\sum_{j=k}^{m-1} Y_j x_j\right)^2}{2\sum_{j=k}^{m-1} x_j^2}\right).$$

For this reason, we write

$$\omega_{ki}^{(1)} = \left(\frac{\sum_{j=k}^{i-1} x_j^2}{\sum_{j=k}^{i} x_j^2}\right)^{1/2} I\{i>k\} + e^{-Y_k^2/2} I\{i=k\}, \tag{3.10}$$
$$R_m^{(1a)} = \sum_{k=1}^m \left(\prod_{i=k}^m \omega_{ki}^{(1)}\right) \exp\left(\frac{\left(\sum_{j=k}^m Y_j x_j\right)^2}{2\sum_{j=k}^m x_j^2}\right).$$

4. One can show regarding the model (2.3) (where $\theta_0 = 0$, f corresponds to $\mathbf{N}(0,1)$ and the estimator of unknown parameter θ_1 is

$$\widehat{\theta}_1^{(k,m)} \quad = \quad \frac{\sum_{j=k}^m Y_j Y_{j-1}}{\sum_{j=k}^m Y_{j-1}^2} \quad), \tag{3.11}$$

we have

$$\begin{aligned} \omega_{ki}^{(1)} &= \left(\frac{\sum_{j=k}^{i-1} Y_{j-1}^2}{\sum_{j=k}^{i} Y_{j-1}^2} \right)^{1/2} I\{i > k\} + e^{-Y_k^2/2} I\{i = k\}, \quad \left(\omega_{12}^{(1)} = \exp(-Y_2^2/2)\right), \quad (3.12) \\ R_m^{(1a)} &= \sum_{k=1}^m \left(\prod_{i=k}^m \omega_{ki}^{(1)} \right) \exp\left(\frac{\left(\sum_{j=k}^m Y_j Y_{j-1}\right)^2}{2\sum_{j=k}^m Y_{j-1}^2} \right). \end{aligned}$$

4 GENERAL CASE OF THE PROBLEM

This section considers a test for hypothesis (1.2), where all parameters of the model (1.1) are unknown. Denote the estimator of the likelihood ratio (2.1)

$$\Lambda_{km}^{(2)} \quad = \quad \prod_{i=1}^{k-1} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}\left(Y_i | Y_1, \ldots, Y_{i-1}\right)}{f_{0\,\widetilde{\theta}_0^{(1,m)}}\left(Y_i | Y_1, \ldots, Y_{i-1}\right)} \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,i-1)}}\left(Y_i | Y_1, \ldots, Y_{i-1}\right)}{f_{0\,\widetilde{\theta}_0^{(1,m)}}\left(Y_i | Y_1, \ldots, Y_{i-1}\right)}, \tag{4.1}$$

where $\widehat{\theta}$ is defined by (3.1) ($\widehat{\theta}^{(k,k-1)} = s$ for given s), $\widetilde{\theta}^{(1,m)}$ is the maximum likelihood estimator of θ based upon $Y_1, \ldots, Y_m$ observations, i.e.

$$\widetilde{\theta}_0^{(1,m)} = \arg\max_{\theta} \prod_{j=1}^m f_{0\,\theta}\left(Y_j | Y_1, \ldots, Y_{j-1}\right). \tag{4.2}$$

The same idea appears in Gurevich and Vexler (2005) in a different context: by virtue of (4.2), we obtain the inequality

$$\prod_{i=1}^m f_{0\,\widetilde{\theta}_0^{(1,m)}}\left(Y_i | Y_1, \ldots, Y_{i-1}\right) \quad \geq \quad \prod_{i=1}^m f_{0\,\theta_0}\left(Y_i | Y_1, \ldots, Y_{i-1}\right) \tag{4.3}$$

that is applied to consider the significance level of the test.

We propose the following test-procedure: reject H_0 iff for a specified threshold $C > 0$

$$\frac{1}{n} \max_{1 \leq m \leq n} R_m^{(2)} \quad > \quad C, \text{ where } R_m^{(2)} = \sum_{k=1}^m \Lambda_{km}^{(2)},\ \Lambda_{km}^{(2)} \text{ by } (4.1). \tag{4.4}$$

Proposition 4.1 *The significance level $\alpha^{(2)}$ of the test satisfies:*

$$\alpha^{(2)} \equiv \sup_{\theta_0} P_0 \left\{ \max_{1\le m\le n} R_m^{(2)} > nC \right\} \le 1/C.$$

Proof. Following (4.3), we obtain that for all $1 \le m \le n$:

$$R_m^{(2)} \le Q_m := \sum_{k=1}^{m} \prod_{i=1}^{k-1} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \prod_{i=k}^{m} \frac{f_{1\,\widehat{\theta}_1^{(k,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}. \tag{4.5}$$

Hence,

$$\alpha^{(2)} \le \sup_{\theta_0} P_0 \left\{ \max_{1\le m\le n} Q_m \ge nC \right\} \tag{4.6}$$

Define a stopping rule $N_n^{(2)}(A) = \min\left\{\tau^{(2)}(A), n\right\}$, where for $A > 0$,

$$\tau^{(2)}(A) = \inf\{l > 1 : \quad Q_l \ge A\}. \tag{4.7}$$

Now

$$\alpha^{(2)} \le \sup_{\theta_0} P_0 \left\{ \max_{1\le m\le n} Q_m \ge nC \right\} = \sup_{\theta_0} P_0 \left\{ Q_{N_n^{(2)}(nC)} > nC \right\} \le \sup_{\theta_0} \frac{E_0 Q_{N_n^{(2)}(nC)}}{nC}. \tag{4.8}$$

Consider the conditional expectation

$$\begin{aligned} E_0(Q_m \big| Y_1,\ldots,Y_{m-1}) &= \sum_{k=1}^{m-1} \prod_{i=1}^{k-1} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} \\ &\times \prod_{i=k}^{m-1} \frac{f_{1\,\widehat{\theta}_1^{(k,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} E_0\left(\frac{f_{1\,\widehat{\theta}_1^{(k,m-1)}}(Y_m|Y_1,\ldots,Y_{m-1})}{f_{0\,\theta_0}(Y_m|Y_1,\ldots,Y_{m-1})} \Big| Y_1,\ldots,Y_{m-1} \right) \\ &+ \prod_{i=1}^{m-1} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})} E_0\left(\frac{f_{1\,\widehat{\theta}_1^{(m,m-1)}}(Y_m|Y_1,\ldots,Y_{m-1})}{f_{0\,\theta_0}(Y_m|Y_1,\ldots,Y_{m-1})} \Big| Y_1,\ldots,Y_{m-1} \right) \\ &= Q_{m-1} + \prod_{i=1}^{m-1} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}. \end{aligned} \tag{4.9}$$

Since (4.9), the sequence

$$Q_m - m \prod_{i=1}^{m} \frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}$$

is a H_0-martingale with zero expectation and hence

$$E_0\left(Q_{N_n^{(2)}(nC)} - N_n^{(2)}(nC)\prod_{i=1}^{N_n^{(2)}(nC)}\frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}\right)=0.$$

By the inequality (4.8) and the definition of $N_n^{(2)}(A)\le n$, we have

$$\alpha^{(2)} \le \frac{1}{C}\sup_{\theta_0} E_0\left(\prod_{i=1}^{N_n^{(2)}(nC)}\frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})}\right). \tag{4.10}$$

It is clear that the sequence $\left\{\prod_{i=1}^{m}\frac{f_{0\,\widehat{\theta}_0^{(1,i-1)}}(Y_i|Y_1,\ldots,Y_{i-1})}{f_{0\,\theta_0}(Y_i|Y_1,\ldots,Y_{i-1})},\ m\ge 1\right\}$ is a H_0-martingale with unit expectation. This completes the proof of Proposition 4.1.

5 REMARKS

1. Note that the proposed tests have the same upper bound for the significance level if we consider model (1.1), where θ_0 and θ_1 are multidimensional parameters.

2. The $AR(1)$ model is considered through the paper. The methodology of this paper can be easily applied to segmented linear/logistic regression models that are commonly applied to epidemiological studies, where it is often assumed that an explanatory variable has no effect on a response prior to a certain unknown change point (e.g. Küchenhoff and Carroll, 1997; Leuraud and Benichou, 2001). (Note that, subindex i in (1.1) can correspond to "age", "weight", "date" etc. variables.) Section 4 provides the analysis of the case where all parameters of a model are unknown. In the context of a linear regression, Vexler (2006) has proposed more complex tests based on using of the special structure of the segmented regression and an application of invariant transformations of observations (under these transformations, the influence of unknown parameters under H_0 is removed). Tests similar to the tests considered in Section 3.2 for regression models are topics that have not been addressed in the literature.

3. We can efficiently evaluate the points, where the changes have occurred, only if we reject hypothesis H_0 (e.g. Dragalin, 1996). Methods for estimation of ν and γ can be found, for example, in Borovkov (1998), Braun *et al.* (2000), Hansen (2000), Koul *et al.* (2003) etc.

4. Assume that observations from (1.1) have been surveyed sequentially ($n = \infty$). Let $N(A)$ be a stopping rule of a sequential scheme for detecting a change point. One typically controls the level of false alarms by considering only stopping rules N which satisfy $E_0 N(A) \geq B$ for some specified level B (e.g. Pollak, 1985). In the case of i.i.d. (prior to and after the change) observations, Lorden and Pollak (2005) investigate the procedure

$$N(A) \quad = \quad \inf\left\{l > 1 : \quad R_l^{(1)} \geq A\right\},$$

where $R_l^{(1)}$ by (3.2). By using the martingale property of $R_l^{(1)} - l$ (i.e. $E_0(R_{N(A)}^{(1)} - N(A)) = 0$) and the definition of $N(A)$ (i.e. $R_{N(A)}^{(1)} \geq A$), these authors show that $E_0 N(A) \geq A$. Following Section 3.2, $E_0 N^{(a1)}(A) \geq A$ and $E_0 N^{(a2)}(A) \geq A$, where the procedures are

$$N^{(a1)}(A) = \inf\left\{l > 1 : \quad R_l^{(1a)} \geq A\right\}, \quad N^{(a2)}(A) = \inf\left\{l > 1 : \quad R_l^{(1a1)} \geq A\right\}.$$

However, here, by applying the weights (which, at least in the considered examples basically, are independent of observations) instead of the non-anticipating estimation based on a part of available observations, we preserve the martingale structure.

In the case of model (3.8) with $x_i = \sin(i), \theta_1 = 0.5, 2$, Monte Carlo simulations were executed in order to evaluate $E_{1\,\infty} N^{(a1)}$ versus $E_{1\,\infty} N$. The conclusion was that the proposed procedure $N^{(a1)}$ raises the alarm more quickly than N, after the change of the model.

5. The proposed martingale methodology is quite general, and that can be applied to complex models. But, of course, this generality may also be looked upon as a disadvantage, because in simple cases of the stated problem, asymptotic results for significant levels of tests can be shown. However, as this case stands, the proposed upper bound is still applicable, since the threshold C can been corrected by the asymptotic approximation. This means that C has been set so that the significance level of the test is no greater than the specified level. Such method has been widely discussed in the change point literature (e.g. Lorden and Pollak, 2005). The proposed methodology is easy to use and, obviously, there are situations, where more accurate calculations of the significant level is impossible or very difficult, or an investigator need the guaranteed conclusion etc.

6 Monte Carlo Simulation Study

Here, we numerically illustrate comparison between the adaptive procedure based on the application of the non-anticipating estimation and the adaptive procedure founded upon the weighted Shiryayev-Roberts statistics. We consider the model mentioned in the example 4 in Section 3.2, i.e. $Y_i = \theta_i' Y_{i-1} + \varepsilon_i,\ Y_0 = 0,\quad i = 1, \ldots, n = 75$, where $\theta_i' = 0.5I\{\nu \le i < \gamma\}$ ($\theta_0 = 0$ and $\theta_1 = 0.5$). For each step of the simulations, 15000 replications of this model were run. The first step is evaluation of the significance levels for tests based on the test-statistics (3.12) and

$$R_m^{(1)} = \sum_{k=1}^{m} \prod_{i=k}^{m} \lambda_i \left(\widehat{\theta}_1^{(k,i-1)}\right), \quad \lambda_i(\theta_1) = e^{-(Y_i - \theta_1 Y_{i-1})^2/2 + (Y_i)^2/2}, \tag{6.1}$$

$$\widehat{\theta}_1^{(a,b)} = \frac{\sum_{j=a}^{b} Y_j Y_{j-1}}{\sum_{j=a}^{b} Y_{j-1}^2} (= 0,\ \text{if } b < a).$$

Figure 1 depicts the histograms of the logarithm of test-statistics $\max_{1 \le m \le n} R_m^{(1a)}/n$ and $\max_{1 \le m \le n} R_m^{(1)}/n$ (under the model, where $\gamma > \nu > n = 75$).

- Figure 1 here -

Thus, for example, if $C = 20$ then the Monte Carlo significance levels are approximately 0.030 and 0.018 for tests (3.12) and (6.1), respectively. It seems that the Monte Carlo significance level for the first test is approximately the Monte Carlo significance level for the second test multiplied by 1.7, for fixed threshold C. For the second step of this simulation study, we denote $\nu = 20$ and $\gamma = 50$. Figure 2 plots realizations of these observations.

- Figure 2 here -

The graphical display of the model realizations illustrates complexity of decision making on existence of epidemic component in the model. Now, we represent the histograms of the logarithm of test-statistics $\max_{1 \le m \le n} R_m^{(1a)}/n$ and $\max_{1 \le m \le n} R_m^{(1)}/n$ (under the model, where $\nu = 20$ and $\gamma = 50$) in Figure 3.

- Figure 3 here -

Thus, for example, if $C = 20$ then the simulated powers are about 0.3909 and 0.0431 of tests (3.12) and (6.1), respectively. It is clear that, in the circumstances, at balance of the simulated significance levels, the test based on the weighted Shiryayev-Roberts statistics is more powerful. The next figure can give an experimental reason for this fact.

- Figure 4 here -

Acknowledgments. I am grateful to Professor Moshe Pollak for valuable discussions and comments as well as for his support to these investigations.

BIBLIOGRAPHY

Borovkov, A. A. (1998). Asymptotically optimal solutions in the change-point problem. *Theory Probab. Appl.*, **43**, 539-561.

Braun, J.V., Braun, R.K. and Müller, H.-G. (2000). Multiple changepoin fitting via quasi-likelihood, with application to DNA sequence segmentation. *Biometrika*, **87**, 301-314.

Broemeling, L.D. and Tsurumi, H. (1987). *Econometrics and Structural Change.* New York: Marcel Dekker.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. R. Statist. Soc. B*, **37**, 149-192.

Chan, H. P. and Lai, T. L. (2005). Importance sampling for generalized likelihood ratio procedures in sequential analysis. *Sequential Analysis*, **24**, 259 - 278.

Dragalin, V.P. (1996). Retrospective change point problem. *Economic Quality Control*, **11**, 3-22.

Dragalin, V.P. (1997). The sequential change point problem. *Economic Quality Control*, **12**, 95-122.

Gurevich, G. and Vexler, A. (2005). Change point problems in the model of logistic regression. *Journal of Statistical Planning and Inference*, **131**, 313-331.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, **68, 3**, 575-603.

Julious, S. A. (2001). Inference and estimation in a change point regression problem. *J. R. Statist. Soc. D*, **50**, 51-61.

Koul, H.L., Qian, L. and Surgailis, D. (2003). Asymptotics of M-estimators in two-phase linear regression models. *Stochastic Processes and their Applications*, **103**, 123-154.

Krieger, A. M., Pollak, M. and Yakir, B. (2003). Surveillance of a simple linear regression. *Journal of the American Statistical Association*, **98**, 456-469.

Küchenhoff, H. and Carroll, R. J. (1997). Segmented regression with errors in predictors: semi-parametric and parametric methods. *Statistics in Medicine*, **16**, 169-188.

Lai, T.L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. R. Statist. Soc. B*, 57 **4**, 613-658.

Lai, T.L. (2001). Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, **11**, 303-408.

Leuraud, K. and Benichou, J.A. (2001). Comparison of several methods to test for the existence of a monotonic dose-response relationship in clinical and epidemiological studies. *Statistics in Medicine*, **20**, 3335-3351.

Lorden, G. and Pollak, M. (2005). Non-Anticipating Estimation Applied to Sequential Analysis and Changepoint Detection. *Ann. Statist.*, **33**, 1422–1454.

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.

Ramanayake, A. and Gupta, A. K. (2003). Tests for an epidemic changes in a sequence of exponentially distributed random variables. *Biometrical Journal*, **45**, 946-958.

Robbins, H. and Siegmund, D. (1973). *A class of stopping rules for testing parametric hypotheses.* Proceedings of the Sixth Berkeley Symposium, Mathematics and Statistical Probability 4, University of California Press.

Tartakovsky, A. G. and Veeravalli, V. V. (2005). General asymptotic bayesian theory of quickest change detection. *Theory Probab. Appl.*, **49**, 458-497.

Vexler, A. (2006). Guaranteed testing for epidemic changes of a linear regression model. *Journal of Statistical Planning and Inference*, **136**, 3101 – 3120.

Vexler, A. and Gurevich, G. (2006). Guaranteed local maximum likelihood detection of a change point in nonparametric logistic regression. *Communications in Statistics—Theory and Methods*, **35**, 711–726.

Yakir, B. (1995). A note on the run length to false alarm of a change-point detection policy. *Ann. Statist.*, **23**, 272-281.

Yakir, B., Krieger, A. M. and Pollak, M. (1999). Detecting a change in regression: first-order optimality. *Ann. Statist.*, **27**, 1896-1913.

Yakir, B. and Pollak, M. (1998). A new representation for a renewal-theoretic constant appearing in asymptotic approximations of large deviations. *Ann. Appl. Probab.*, **8**, 749-774.

Yao, Q. (1993). Tests for change-points with epidemic alternatives. *Biometrika*, **80**, 179-191.

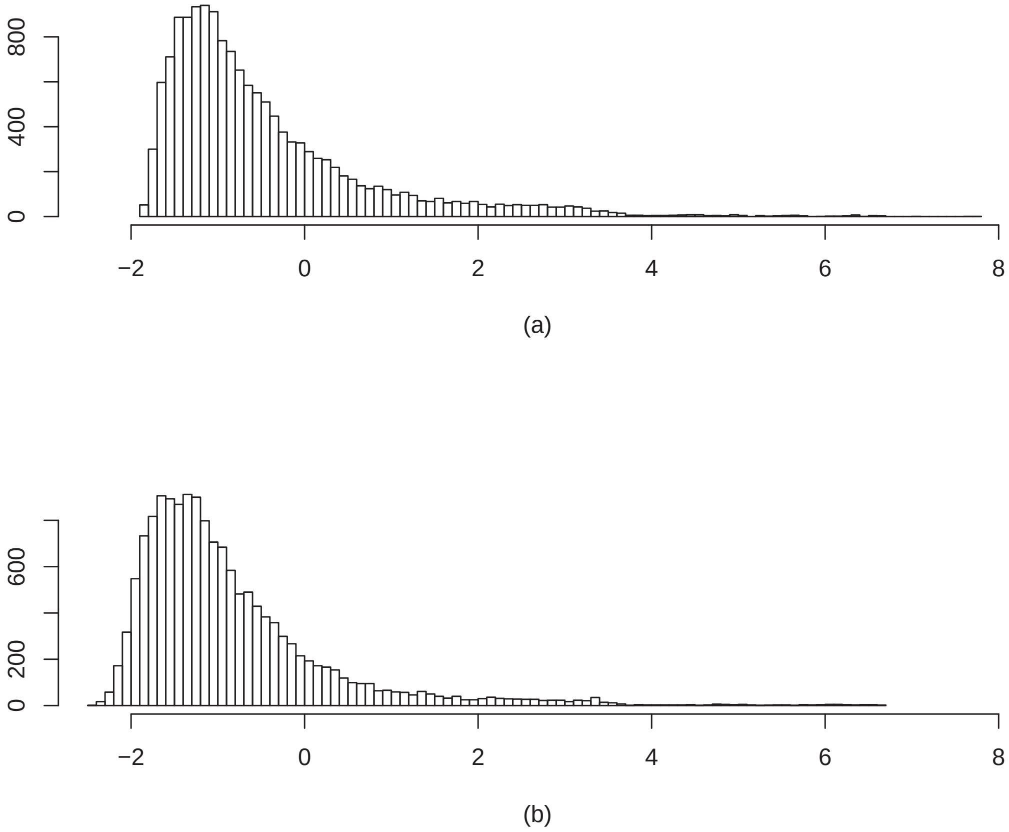

Figure 1: Comparison between the histogram of $\ln\left(\frac{1}{n}\max_{1\leq m\leq n} R_m^{(1a)}\right)$ and $\ln\left(\frac{1}{n}\max_{1\leq m\leq n} R_m^{(1)}\right)$ (the graphs (a) and (b), respectively); number of realizations is 15000.

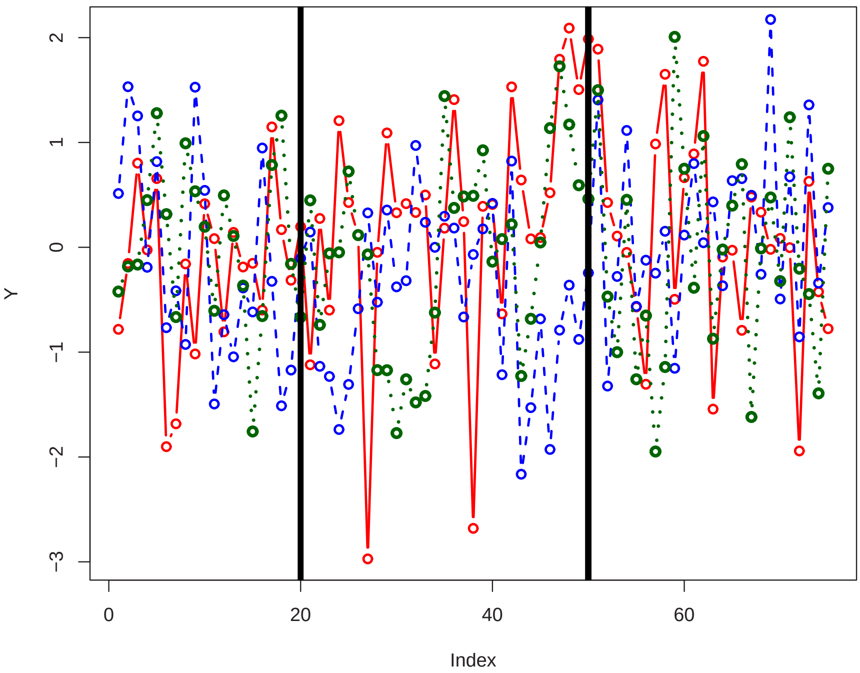

Figure 2: Three realizations ((——), (- - -) and ($\cdots$)) of Y_i, $i = 1, \ldots, 75$, where $\nu = 20$ and $\gamma = 50$.

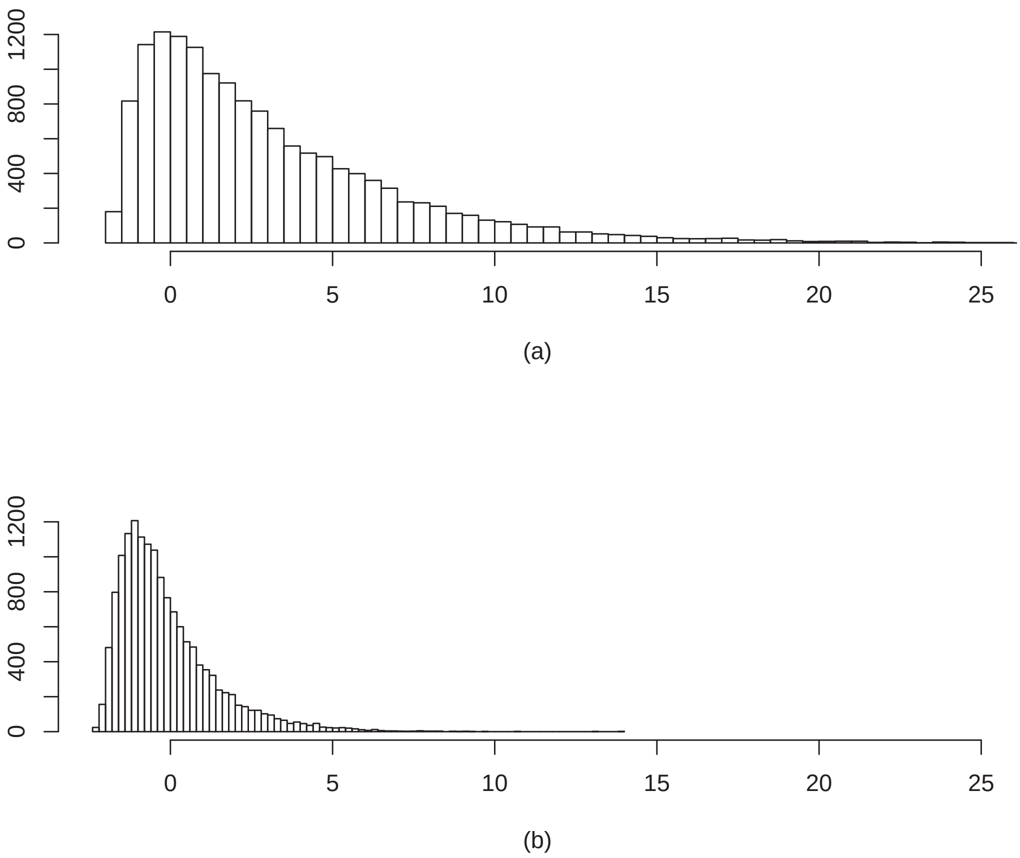

Figure 3: Comparison between the histogram of $\ln\left(\frac{1}{n}\max_{1\le m\le n} R_m^{(1a)}\right)$ and $\ln\left(\frac{1}{n}\max_{1\le m\le n} R_m^{(1)}\right)$ (the graphs (a) and (b), respectively); number of realizations is 15000.

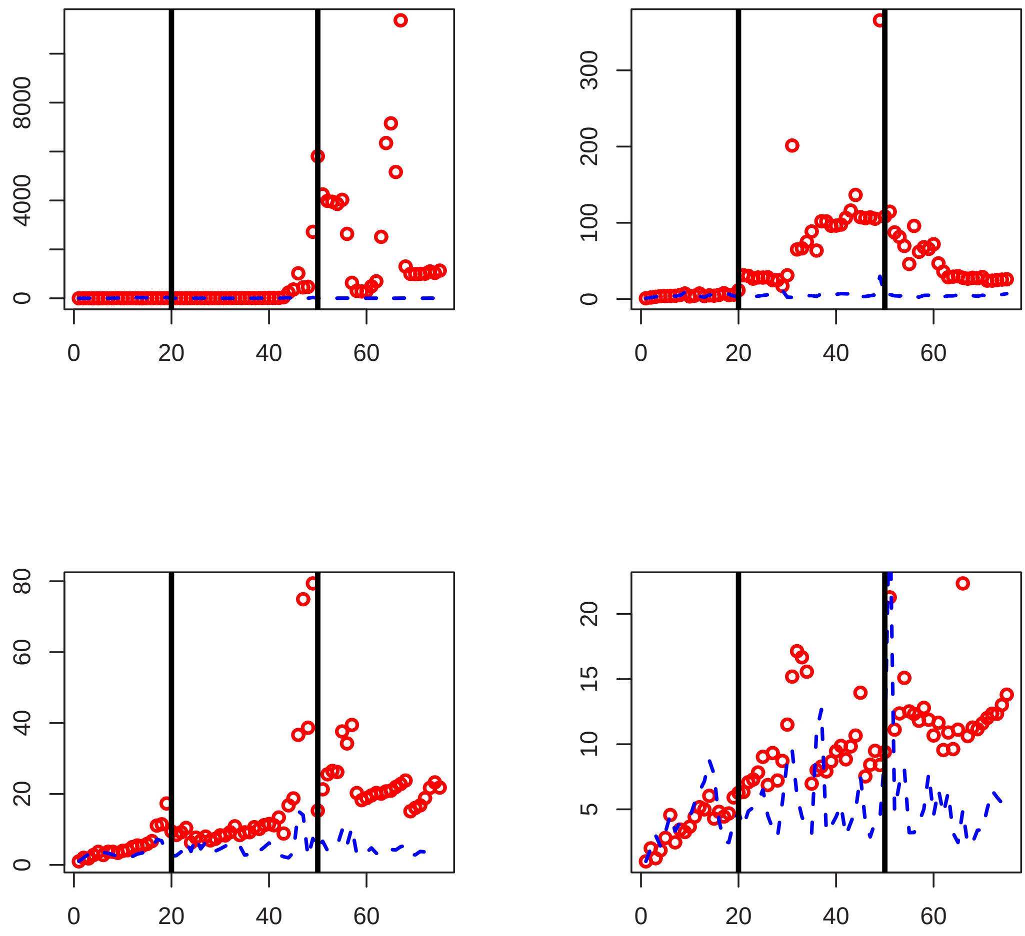

Figure 4: The some realizations of $R_i^{(1a)}$: ($\circ$) and $R_i^{(1)}$: (-) (are based on same observations) plotted against $i = 1, \ldots, 75$ (the axis of abscissae).