

An Optimal Retrospective Change Point Detection Policy

Running headline: Optimal Change Point Detection Policy

ALBERT VEXLER

Department of Biostatistics, The State University of New York at Buffalo

CHENGQING WU

Department of Epidemiology and Public Health, Yale University

ABSTRACT. Since the middle of the twentieth century, the problem of making inferences about the point in a surveyed series of observations at which the underlying distribution changes has been extensively addressed in the economics, biostatistics and statistics literature. CUSUM type statistics have commonly been thought to play a central role in non-sequential change point detections. Alternatively, we present and examine an approach based on the Shiriyayev-Roberts scheme. We show that retrospective change point detection policies based on Shiriyayev-Roberts statistics are non-asymptotically optimal in the context of the most powerful testing.

Key words: bayes factors, change point, CUSUM, likelihood, mixture type testing, most powerful testing, optimality, Shiriyayev-Roberts

1 Introduction

Often, formalized econometric or epidemiological studies correspond to the following problem of testing hypotheses (e.g., Csörgő & Horváth, 1997). There are n independent data points X_1, X_2, \dots, X_n . The assumption that observations X_i are identically distributed (say, $X_i \sim f_0(x; \theta_0)$, $1 \leq i \leq n$) presents the baseline hypothesis H_0 , while the alternative hypothesis H_1 corresponds to $X_k \sim f_0(x; \theta_0)$, $X_j \sim f_1(x; \theta_1)$, $k = 1, \dots, \nu - 1$, $j = \nu, \dots, n$, where θ_0 and θ_1 are parameters. Thus, we suppose that if a change in the distribution did occur at time ν , then it is unique and the observations after the change all have the same distribution, which differs from the distribution of the observations before the change. Let P_ν and E_ν be the probability measure and expectation conditional on ν such that, under P_ν , $X_i \sim f_0(x; \theta_0)I\{i < \nu\} + f_1(x; \theta_1)I\{i \geq \nu\}$, $1 \leq i \leq n$, where $I\{\cdot\}$ is the indicator function and $\nu = 1, \dots, n + 1$.

(The case with $\nu = n + 1$ corresponds to H_0 .)

When the parameters ν , θ_0 , and θ_1 are known, the likelihood methodology suggests the test for H_0 versus H_1 based on the ratio

$$\Lambda_n(\nu, \theta_0, \theta_1) = \prod_{i=\nu}^n \frac{f_1(X_i; \theta_1)}{f_0(X_i; \theta_0)}, \quad (1)$$

i.e., for a threshold C , the event $\{\Lambda_n(\nu, \theta_0, \theta_1) > C\}$ declares rejection of H_0 . By virtue of the elementary inequality

$$(A - B)(I\{A > B\} - \delta) \geq 0, \quad \text{for all } A, B, \text{ and } \delta \in [0, 1] \quad (2)$$

with $A = \Lambda_n(\nu, \theta_0, \theta_1)$ and $B = C$, we have $(\Lambda_n(\nu, \theta_0, \theta_1) - C)I\{\Lambda_n(\nu, \theta_0, \theta_1) > C\} \geq (\Lambda_n(\nu, \theta_0, \theta_1) - C)\delta$. Therefore

$$E_{n+1}((\Lambda_n(\nu, \theta_0, \theta_1) - C)I\{\Lambda_n(\nu, \theta_0, \theta_1) > C\}) \geq E_{n+1}((\Lambda_n(\nu, \theta_0, \theta_1) - C)\delta),$$

where

$$E_{n+1}(\Lambda_n(\nu, \theta_0, \theta_1)\delta) = \int \prod_{i=\nu}^n \frac{f_1(x_i; \theta_1)}{f_0(x_i; \theta_0)} \delta \prod_{i=1}^n f_0(x_i; \theta_0) \prod_{i=1}^n dx_i = E_\nu(\delta), \quad (3)$$

for any statistic δ based on $\{X_1, \dots, X_n\}$. Hence, we obtain

$$P_\nu\{\Lambda_n(\nu, \theta_0, \theta_1) > C\} - CP_{n+1}\{\Lambda_n(\nu, \theta_0, \theta_1) > C\} \geq P_\nu\{\delta = 1\} - CP_{n+1}\{\delta = 1\},$$

for any $\delta = 0, 1$. In this article, we will assume that $\delta = 0, 1$ is any decision rule based on $\{X_1, \dots, X_n\}$ (for determinacy, say, the event $\{\delta = 1\}$ rejects H_0). Thus, supposing that the event $\{\delta = 1\}$ rejects H_0 with type I error $P_{n+1}\{\delta = 1\} = \alpha$, we fix the threshold $C : P_{n+1}\{\Lambda_n(\nu, \theta_0, \theta_1) > C\} = \alpha$ and conclude that the likelihood ratio statistic (1) provides the most powerful test.

In the case, where the baseline and post-change parameters θ_0 and θ_1 are known but ν is unknown, the maximum likelihood estimation of ν yields the CUSUM statistic

$$\Lambda_n(\theta_0, \theta_1) = \max_{1 \leq k \leq n} \Lambda_n(k, \theta_0, \theta_1), \quad \Lambda_n(\theta_0, \theta_1) = \frac{f_1(X_n; \theta_1)}{f_0(X_n; \theta_0)} \max(\Lambda_{n-1}(\theta_0, \theta_1), 1). \quad (4)$$

Decision rules based on CUSUM type statistics are well-accepted change point detection schemes (e.g., Page, 1954, 1955; Ploberger & Kramer, 1992; Sen & Srivastava, 1975; Pettitt, 1980; Gombay & Horvath, 1994 as well as Gurevich & Vexler, 2005). It is widely

known in change point literature (e.g., Lai, 1995) that likelihood ratio type tests have high power, therefore, evaluation of their significance level is a major issue. Note that $\Lambda_n(\nu, \theta_0, \theta_1)$ is an H_0 -martingale, whereas its estimator $\Lambda_n(\theta_0, \theta_1)$ is an H_0 -submartingale (i.e. by the definitions (1) and (4), for $\nu < n$: $E_{n+1}(\Lambda_n(\nu, \theta_0, \theta_1) | X_1, \dots, X_{n-1}) = \Lambda_{n-1}(\nu, \theta_0, \theta_1)$, but $E_{n+1}(\Lambda_n(\theta_0, \theta_1) | X_1, \dots, X_{n-1}) = \max(\Lambda_{n-1}(\theta_0, \theta_1), 1) \geq \Lambda_{n-1}(\theta_0, \theta_1)$), and moreover, $\Lambda_n(\theta_0, \theta_1) \geq \Lambda_n(\nu, \theta_0, \theta_1)$. Thus, the type I error of the test based on the CUSUM statistic can be expected to be raised too high, and hence the idea of a non-asymptotic (as n is fixed) optimality of the retrospective CUSUM test is problematic. (We note, however, that, Moustakides (1986) has shown a non-asymptotical optimality of the sequential change point detection procedure based on the CUSUM statistic.) James *et al.* (1987) have conjectured that, in the context of the retrospective change point problem, there is no uniformly most powerful test.

Because of the martingale property of the optimal test statistic $\Lambda_n(\nu, \theta_0, \theta_1)$, it is natural to require that any variation of the detection scheme due to the estimation of ν should preserve this martingale structure (e.g., Broström, 1997).

In the field of sequential change point detections, professor Pollak and his co-authors have proposed and examined optimal procedures in which a quasi-Bayesian approach to dealing with Shiriyayev-Roberts statistics has been applied (e.g., Pollak, 1985). In the context of retrospective detections of changes in complicated models, Vexler (2006) has proposed the Shiriyayev-Roberts approach to fix the significance levels of decision rules. The Shiriyayev-Roberts (SR) statistic has the form of

$$R_n(\theta_0, \theta_1) = \sum_{k=1}^n \Lambda_n(k, \theta_0, \theta_1), \quad R_n(\theta_0, \theta_1) = \frac{f_1(X_n; \theta_1)}{f_0(X_n; \theta_0)} (R_{n-1}(\theta_0, \theta_1) + 1). \quad (5)$$

Under H_0 , definition (5) provides $E_{n+1}(R_n | X_1, \dots, X_{n-1}) - n = R_{n-1} - (n-1)$. Therefore, the $R_n - n$ statistic is an H_0 -martingale. Thus, we can intuitively assume that the retrospective test based on the SR statistic has a significance level that is comparable with the type I error of the likelihood ratio test. Hence, an optimality of the SR test can be shown. Here the expression "optimality" corresponds to the most powerful testing in a variety of contexts. For example, in the next section we demonstrate that the test based on (5) is optimal with respect to the average power (via $\nu = 1, \dots, n$) of change point procedures. In this paper, we use the term "optimality" in a manner similar to Lorden's (1967) usage of the term for

sequential testing.

The main goal of the paper is to present non-asymptotic optimal tests based on the SR statistic for the change point detection. We consider complicated cases of the change point detection where the baseline parameter and post-change parameter are unknown. In the cases where X s are stated in terms of a regression model or X s are dependent, the proposed approach can be easily extended with support for the optimality of the change point test. A Monte Carlo study confirms that the proposed change point detection policies, in many cases, have power properties that are superior to those of the commonly used CUSUM procedures.

2 Pre-and-post-change parameters are known

We begin with the case where parameters θ_0 and θ_1 are known. (Throughout this paper, ν is unknown.) Although this case is simple, the situations where baseline and post-change parameters are known, can be associated with real-data problems (e.g., Gurevich & Vexler, 2005). Here, the analysis is relatively clear, and it has the basic ingredients for more general cases.

The test is given by the following statement: reject H_0 if and only if

$$R_n(\theta_0, \theta_1) > C, \quad (6)$$

where $R_n(\theta_0, \theta_1)$ is the SR statistic (5) and the threshold $C > 0$.

Proposition 1 *For the test (6) the following holds:*

(i) (6) is the average most powerful change point detection, i.e.

$$\begin{aligned} & \frac{1}{n} \sum_{k=1}^n P_k \{R_n(\theta_0, \theta_1) > C\} - \frac{C}{n} P_{n+1} \{R_n(\theta_0, \theta_1) > C\} \\ & \geq \frac{1}{n} \sum_{k=1}^n P_k \{\delta \text{ rejects } H_0\} - \frac{C}{n} P_{n+1} \{\delta \text{ rejects } H_0\}, \end{aligned}$$

where $\delta = 0, 1$ is any decision rule based on $\{X_1, \dots, X_n\}$;

(ii) Under H_0 , $R_n(\theta_0, \theta_1) \xrightarrow[n \rightarrow \infty]{dist} \gamma$, where $\gamma = \sum_{k=1}^{\infty} \prod_{i=1}^k \frac{f_1(X_i; \theta_1)}{f_0(X_i; \theta_0)}$ is a.s. a finite-valued random variable.

Proof. The expectation under P_{n+1} applied to the inequality (2) with $A = R_n(\theta_0, \theta_1)$ and

$B = C$ leads to the inequality

$$\begin{aligned} & \sum_{k=1}^n E_{n+1} (\Lambda_n(k, \theta_0, \theta_1) I \{R_n(\theta_0, \theta_1) > C\}) - CP_{n+1} \{R_n(\theta_0, \theta_1) > C\} \geq \\ & \sum_{k=1}^n E_{n+1} (\Lambda_n(k, \theta_0, \theta_1) I \{\delta \text{ rejects } H_0\}) - CP_{n+1} \{\delta \text{ rejects } H_0\}. \end{aligned}$$

Now, by virtue of equation (3), proposition (i) follows.

Lemma 3 of Pollak (1985, p. 210) directly proves statement (ii).

In accordance with Proposition 1, if some test $\delta(X_1, \dots, X_n)$ has type I error $P_{n+1}\{\delta(X_1, \dots, X_n) \text{ rejects } H_0\} = \alpha$, then for the fixed threshold $C_\alpha : P_{n+1}\{R_n(\theta_0, \theta_1) > C_\alpha\} = \alpha$, we have

$$\frac{1}{n} \sum_{k=1}^n P_k \{R_n(\theta_0, \theta_1) > C_\alpha\} \geq \frac{1}{n} \sum_{k=1}^n P_k \{\delta \text{ rejects } H_0\}.$$

Therefore, test (6) has an average power which is greater than that of $\delta(X_1, \dots, X_n)$.

Since the stated problem is parametric, the significance level of (6) can be evaluated by Monte Carlo simulations. Alternatively, Proposition 1 can be applied to evaluate asymptotically (as $n \rightarrow \infty$) the type I error of the test (6).

Remark. The proof scheme of statement (i) of Proposition 1 provides the next issue. A generalized form of the test (6) can be presented as follows. For $q_k \geq 0$, we reject H_0 if

$$R_n^q(\theta_0, \theta_1) = \sum_{k=1}^n q_k \Lambda_n(k, \theta_0, \theta_1) > C. \quad (7)$$

In other words, we can take into account a sequence of weights, $\{q_k\}$, that can be considered as a prior distribution of the change point, i.e., $q_k = P_{H_1}(\nu = k)$. Following the proof of Proposition 1, for all decision rules $\delta(X_1, \dots, X_n) = 0, 1$ with $P_{n+1}\{\delta(X_1, \dots, X_n) \text{ rejects } H_0\} = \alpha$, we conclude that

$$\begin{aligned} \sum_{k=1}^n q_k P_k \{R_n^q(\theta_0, \theta_1) > C_\alpha^q\} & \geq \sum_{k=1}^n q_k P_k \{\delta \text{ rejects } H_0\}, \\ C_\alpha^q : P_{n+1} \{R_n^q(\theta_0, \theta_1) > C_\alpha^q\} & = \alpha. \end{aligned} \quad (8)$$

(The test (7) is the average most powerful change point detection with respect to the weights q_k , $k = 1, \dots, n$.) Hence, when the q_k , $k = 1, \dots, n$ are not fixed (frequently, in Bayesian cases, q_k , $k = 1, \dots, n$ can be pre-specified), these weights can be chosen with respect to a special area of change point occurrence, where the maximum of the test's power is desired corresponding to the interests that belong to a tester.

3 Post-change parameter is unknown

Let θ_0 be known, and suppose θ_1 is unknown. This case has been dealt with extensively in the literature, and it corresponds to a control problem where the baseline in-control distribution is known and the post-change out-of-control distribution is not. One application of this problem is found in epidemiological studies where it is often assumed that observations have zero expectation prior to a certain unknown change point. (In several situations, when the baseline parameter of an observed sample is unknown, observations can be transformed in order to set a baseline parameter of the transformed sample to a known value, e.g. Brown *et al.* (1975) as well as Yakir (1998).)

The unknown post-change parameter is usually estimated. To preserve the H_0 -martingale structure of the test statistics, the post-change parameter's estimation has to be adapted with a loss of efficiency (e.g., Lai, 2001 (p. 398), Lorden & Pollak, 2004 and Gurevich & Vexler, 2005). (In a general context, Vexler (2008) proposed an approach to obtain H_0 -martingale type test-statistics with improved power properties.) Alternatively, the literature on sequential change point detection (e.g., Krieger *et al.*, 2003) introduces mixture type tests.

To present the SR statistic when θ_1 is not known, we introduce a mixing measure $H(\theta)$ and define

$$R_n(\theta_0) = \sum_{k=1}^n \int \Lambda_n(k, \theta_0, \theta) dH(\theta). \quad (9)$$

(Here we can pretend that $\theta_1 \sim H$, but this assumption is not vital for the execution of the change point detection.) To formulate the next proposition, we point out that the notation

$$P_k \left[L(X_1, \dots, X_n) < u \mid X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right]$$

means a probability distribution function (depending on k and θ) of a statistic L based on $X_i \sim f_0(x; \theta_0)I\{i < k\} + f_1(x; \theta)I\{i \geq k\}$, $1 \leq i \leq n$.

Proposition 2 *For the test based on (9) the following hold:*

(i) *The statistic $R_n(\theta_0) - n$ is an H_0 -martingale with respect to the σ -algebra generated by X_1, \dots, X_n ;*

(ii) *(9) denotes the average integrated most powerful change point detection with respect*

to $H(\theta)$, i.e.

$$\begin{aligned}
& \frac{1}{n} \sum_{k=1}^n \int P_k \left[R_n(\theta_0) > C \mid X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right] dH(\theta) \\
& - \frac{C}{n} P_{n+1} \{R_n(\theta_0) > C\} \\
& \geq \frac{1}{n} \sum_{k=1}^n \int P_k \left[\delta \text{ rejects } H_0 \mid X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right] dH(\theta) \\
& - \frac{C}{n} P_{n+1} \{\delta \text{ rejects } H_0\},
\end{aligned}$$

where $\delta = 0, 1$ is any decision rule based on $\{X_1, \dots, X_n\}$.

Proof. Since

$$\begin{aligned}
E_{n+1} \left(R_n(\theta_0) \mid X_1, \dots, X_{n-1} \right) - n &= \sum_{k=1}^{n-1} \int \prod_{i=k}^{n-1} \frac{f_1(X_i; \theta)}{f_0(X_i; \theta_0)} E_{n+1} \left(\frac{f_1(X_n; \theta)}{f_0(X_n; \theta_0)} \right) dH(\theta) \\
&+ \int E_{n+1} \left(\frac{f_1(X_n; \theta)}{f_0(X_n; \theta_0)} \right) dH(\theta) - n \\
&= \sum_{k=1}^{n-1} \int \prod_{i=k}^{n-1} \frac{f_1(X_i; \theta)}{f_0(X_i; \theta_0)} \left(\int \frac{f_1(x; \theta)}{f_0(x; \theta_0)} f_0(x; \theta_0) dx \right) dH(\theta) \\
&+ \int \left(\int \frac{f_1(x; \theta)}{f_0(x; \theta_0)} f_0(x; \theta_0) dx \right) dH(\theta) - n = R_{n-1}(\theta_0) - (n-1),
\end{aligned}$$

the definition (9) preserves the H_0 -martingale property of the test statistic. It is clear that, in a similar manner to the proof scheme of Proposition 1.(i), one can show that, for any test $\delta(X_1, \dots, X_n)$ with significance level $P_{n+1} \{\delta \text{ rejects } H_0\} = \alpha$,

$$\begin{aligned}
& \frac{1}{n} \sum_{k=1}^n \int P_k \left[R_n(\theta_0) > C_\alpha(H) \mid X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right] dH(\theta) \\
& \geq \frac{1}{n} \sum_{k=1}^n \int P_k \left[\delta \text{ rejects } H_0 \mid X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right] dH(\theta),
\end{aligned}$$

where the threshold $C_\alpha(H)$ preserves the significance level α of the test based on (9), i.e. $C_\alpha(H)$ is such that $P_{n+1} \{R_n(\theta_0) > C_\alpha(H)\} = \alpha$.

3.1 Choosing a prior $H(\theta)$

In the context of sequential change point detection, Krieger *et al.* (2003) have proposed several forms of a prior $H(\theta)$. The method of Krieger *et al.* (2003) can be adapted for the problem stated in this paper. For example,

$$H(\theta) = \Phi\left(\frac{\mu}{\sigma}\right)^{-1} \left(\Phi\left(\frac{\theta - \mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right) \right)^+ \quad (10)$$

if we suspect that the observations after the possible change have a distribution that differs greatly from the distribution of the observations before the change (Φ is the standard normal distribution function); a somewhat broader prior is

$$H(\theta) = \frac{1}{2} \left(\Phi\left(\frac{\theta - \mu}{\sigma}\right) + \Phi\left(\frac{\theta + \mu}{\sigma}\right) \right). \quad (11)$$

Note that the parameters μ and $\sigma > 0$ of the distributions $H(\theta)$ can be chosen arbitrarily, e.g., Krieger *et al.* (2003) have recommended $\mu = 0$, and $\sigma = 1$ so that (10) and (11) are simplified. Marden (2000) has reviewed the Bayesian approach applied to hypothesis testing. The function H can be defined in accordance with rules mentioned in the literature cited in Section 3 of Marden's paper.

Consider, e.g., for a fixed σ ,

$$H \in \Theta = \left\{ \frac{1}{2} \left(\Phi\left(\frac{\theta - \mu}{\sigma}\right) + \Phi\left(\frac{\theta + \mu}{\sigma}\right) \right), \mu \in (\mu_{lower}, \mu_{upper}) \right\} \quad (12)$$

(here Θ is a set of distribution functions). In this case, the definition (9) provides

$$R_n^a(\theta_0) = \frac{1}{2} \sum_{k=1}^n \int \Lambda_n(k, \theta_0, \theta) \frac{\partial \left(\Phi\left(\frac{\theta - a}{\sigma}\right) + \Phi\left(\frac{\theta + a}{\sigma}\right) \right)}{\partial \theta} d\theta. \quad (13)$$

Let f_0 and f_1 be normal densities with standard deviation 1 and means $\theta_0 = 0$ and θ_1 , respectively. Then the test statistic $R_n^a(\theta_0)$ has the form of

$$R_n^a(0) = \sum_{k=1}^n \frac{1}{2(\sigma^2(n - k + 1) + 1)^{1/2}} \exp\left\{-\frac{a^2}{2\sigma^2}\right\} \\ \times \left[\exp\left\{\frac{\sigma^2 \left(\sum_{i=k}^n X_i + a/\sigma^2\right)^2}{2(\sigma^2(n - k + 1) + 1)}\right\} + \exp\left\{\frac{\sigma^2 \left(\sum_{i=k}^n X_i - a/\sigma^2\right)^2}{2(\sigma^2(n - k + 1) + 1)}\right\} \right]$$

(this form is the average integrated most powerful statistic with respect to the distribution function (11), where $\mu = a$). Hence, we represent the average integrated power of $R_n^a(0)$ in

the form of

$$\begin{aligned}
U(a) &= \sum_{v=1}^n \frac{1}{2} \sum_{j=1}^2 p_{vj}, \quad p_{vj} = P \left\{ \sum_{k=1}^n \frac{1}{2(\sigma^2(n-k+1)+1)^{1/2}} \exp \left\{ -\frac{a^2}{2\sigma^2} \right\} \right. \\
&\times \left[\exp \left\{ \frac{\sigma^2 (\sum_{i=k}^n \varepsilon_i + \xi_j(n-k+1)I\{k \geq v\} + \xi_j(n-v+1)I\{k < v\} + a/\sigma^2)^2}{2(\sigma^2(n-k+1)+1)} \right\} \right. \\
&+ \left. \exp \left\{ \frac{\sigma^2 (\sum_{i=k}^n \varepsilon_i + \xi_j(n-k+1)I\{k \geq v\} + \xi_j(n-v+1)I\{k < v\} - a/\sigma^2)^2}{2(\sigma^2(n-k+1)+1)} \right\} \right] \\
&> C_\alpha(a) \left. \right\}, \quad C_\alpha(a) : P_{n+1} \{R_n^a(0) > C_\alpha(a)\} = \alpha, \quad \varepsilon_i \sim iid N(0, 1), \\
&\xi_j \sim N(aI\{j=1\} - aI\{j=2\}, \sigma^2), \quad E\varepsilon_i \xi_j = 0.
\end{aligned} \tag{14}$$

To depict the function $U(a)$ from (14), we define $\alpha = 0.05$ and $n = 20$ for different values of σ . For each a and C , the probability $P_{n+1} \{R_n^a(0) > C\}$ was Monte Carlo-estimated via 50000 repetitions of sample $X_1, \dots, X_n \sim N(0, 1)$. In this way, the approximate values of the roots $C_{0.05}(a)$ were obtained. Similarly, for each a , $v = 1, \dots, n$ and $j = 1, 2$, the probability p_{vj} by (14) was estimated by Monte Carlo simulation based on 50000 repetitions of the sample $\{\varepsilon_i, \xi_j, i = 1, \dots, n, j = 1, 2\}$.

Figure 1

Thus, in these cases, if we pretend that $\theta_1 \sim \Theta$ (were Θ is given by (12)) then

$$H(\theta) = \frac{1}{2} \left(\Phi \left(\frac{\theta - \mu}{\sigma} \right) + \Phi \left(\frac{\theta + \mu}{\sigma} \right) \right), \quad \mu = \arg \max_{a=\mu_{lower}, \mu_{upper}} U(a)$$

can be suggested to maximize the average integrated power $U(a)$, when μ_{lower}, μ_{upper} are fixed corresponding to practical meanings of the testing-statement of problem.

4 Pre-and-post-change parameters are unknown

Assume that θ_0 and θ_1 are unknown. Let $\hat{\theta}_0$ denote an estimator of θ_0 based on X_1, \dots, X_n , e.g. $\hat{\theta}_0 = \arg \max_a \prod_{i=1}^n f_0(X_i; a)$. Obviously, given an estimation method of θ_0 , an asymptotic (as $n \rightarrow \infty$) closeness of

$$R_n(\hat{\theta}_0) = \sum_{k=1}^n \int \Lambda_n(k, \hat{\theta}_0, \theta) dH(\theta)$$

to the optimal test statistic (9) can be obtained and evaluated. However, in this paper, since we would like to present non-asymptotically optimal procedures, we define

$$R_n = \sum_{k=1}^n \int \prod_{i=1}^{k-1} \frac{f_0(X_i; \bar{\theta}_0)}{f_0(X_i; \hat{\theta}_0)} \Lambda_n(k, \hat{\theta}_0, \bar{\theta}_1) dH(\bar{\theta}_0, \bar{\theta}_1), \quad (15)$$

where the likelihood ratios $\Lambda_j(\cdot)$ are denoted by (1) and we pretend that $(\theta_0, \theta_1) \sim H$. Denoting the estimate of the type I error $P_{n+1} \{\delta \text{ rejects } H_0\}$ in the form of

$$\begin{aligned} & \hat{P} \left\{ \delta \text{ rejects } H_0 \mid X_1, \dots, X_n \sim f_0(x; \hat{\theta}_0) \right\} \\ & := \int \prod_{i=1}^n f_0(x_i; \hat{\theta}_0(x_1, \dots, x_n)) I \{ \delta(x_1, \dots, x_n) \text{ rejects } H_0 \} \prod_{i=1}^n dx_i = \alpha \end{aligned} \quad (16)$$

(here $\hat{P} = P_{n+1}$, if $\hat{\theta}_0 \equiv \theta_0$), we formulate the next result.

Proposition 3 *Let the estimation method of the unknown baseline parameter be fixed and let $\hat{\theta}_0(X_1, \dots, X_n)$ be the estimate of θ_0 . Then, for all decision rules $\{\delta(X_1, \dots, X_n) \in [0, 1]\}$ with*

$$\hat{P} \left\{ \delta \text{ rejects } H_0 \mid X_1, \dots, X_n \sim f_0(x; \hat{\theta}_0) \right\} = \alpha \quad (\alpha \text{ is pre-specified}),$$

the change point detection based on the statistic (15) is average integrated most powerful with respect to the function H .

Proof. Since $(R_n - C)(I\{R_n > C\} - \delta) \geq 0$, we have

$$\begin{aligned} & \sum_{k=1}^n \int \prod_{i=1}^{k-1} \frac{f_0(X_i; \bar{\theta}_0)}{f_0(X_i; \theta_0)} \prod_{i=k}^n \frac{f_1(X_i; \bar{\theta}_1)}{f_0(X_i; \theta_0)} I\{R_n > C\} dH(\bar{\theta}_0, \bar{\theta}_1) - C \prod_{i=1}^n \frac{f_0(X_i; \hat{\theta}_0)}{f_0(X_i; \theta_0)} I\{R_n > C\} \\ & \geq \sum_{k=1}^n \int \prod_{i=1}^{k-1} \frac{f_0(X_i; \bar{\theta}_0)}{f_0(X_i; \theta_0)} \prod_{i=k}^n \frac{f_1(X_i; \bar{\theta}_1)}{f_0(X_i; \theta_0)} \delta dH(\bar{\theta}_0, \bar{\theta}_1) - C \prod_{i=1}^n \frac{f_0(X_i; \hat{\theta}_0)}{f_0(X_i; \theta_0)} \delta. \end{aligned} \quad (17)$$

Similar to the proof scheme of the previous propositions, it is clear that deriving E_{n+1} of (17) completes the proof of Proposition 3.

Thus, (15) is the optimal test-statistic in any set of test-statistics in which the estimation method $\hat{\theta}_0$ is applied.

Remark 1. The notation $\hat{P} \left\{ \delta \text{ rejects } H_0 \mid X_1, \dots, X_n \sim f_0(x; \hat{\theta}_0) \right\}$ is not the classical type I error $P_{n+1} \{\delta \text{ rejects } H_0\}$ (\hat{P} is the estimate of P_{n+1} given the method $\hat{\theta}_0$). When $\hat{\theta}_0 =$

$\arg \max_a \prod_{i=1}^n f_0(X_i; a)$ (the maximum likelihood estimator), however, any test, which satisfies the conditions of Proposition 3, has the type I error that is in classical control

$$\begin{aligned} & \sup_{\theta_0} P_{n+1} \{\delta \text{ rejects } H_0\} \leq \alpha, \text{ since} \\ \alpha &= \int \prod_{i=1}^n f_0(x_i; \hat{\theta}_0(x_1, \dots, x_n)) I \{\delta(x_1, \dots, x_n) \text{ rejects } H_0\} \prod_{i=1}^n dx_i \\ &\geq \int \prod_{i=1}^n f_0(x_i; \theta_0) I \{\delta(x_1, \dots, x_n) \text{ rejects } H_0\} \prod_{i=1}^n dx_i, \text{ for all } \theta_0. \end{aligned}$$

Note that, if θ_0 is unknown, comparing decision rules from the set

$\{\delta : \sup_{\theta_0} P_{n+1}(\delta \text{ rejects } H_0) \leq \alpha\}$ (where α is fixed) is a very complex problem. Suppose we have two tests, (A) and (B), for hypothesis H'_0 versus H'_1 , given by

$$(A) : \text{reject } H'_0 \text{ if statistic } L > C^L; \text{ and } (B) : \text{reject } H'_0 \text{ if statistic } D > C^D,$$

where L, D are based on observations $\{Z_i, i = 1, \dots, m \geq 1\}$, and C^L, C^D are both thresholds. We would like to compare (A) with (B), when the parameters under H'_0 (say, η_0) and parameters under H'_1 are unknown. The suggestion of fixing the type I errors of (A) and (B) as α and then contrasting the powers of these tests is problematic. First, in general, we cannot easily choose C_α^L, C_α^D such that $\sup_{\eta_0} P_{H'_0}\{L > C_\alpha^L\}$ and $\sup_{\eta_0} P_{H'_0}\{D > C_\alpha^D\} \leq \alpha$. Since Monte Carlo evaluations of $\sup_{\eta_0} P_{H'_0}$ are usually complex and biased, analytical presentations of $P_{H'_0}\{L > C\}$ and $P_{H'_0}\{D > C\}$ are often required to derive C_α^L and C_α^D . Second, assuming that C_α^L and C_α^D are known or evaluated, then comparing $P_{H'_1}\{L > C_\alpha^L\}$ with $P_{H'_1}\{D > C_\alpha^D\}$ is an arduous task ($P_{H'_1}\{\cdot\}$ depends on unknown parameters). Alternatively, we suggest fixing H'_0 -parameters' estimates (say, $\hat{\eta}_0(Z_1, \dots, Z_m)$) as well as

$$\begin{aligned} \hat{P}_{H'_0}\{H'_0 \text{ is rejected} \mid \hat{\eta}_0\} &:= \int \psi(z_1; \hat{\eta}_0(z_1, \dots, z_m)) \cdots \psi(z_m; \hat{\eta}_0(z_1, \dots, z_m)) \\ &\quad \times I\{H'_0 \text{ is rejected}\} dz_1 \cdots dz_m = \alpha \end{aligned}$$

(where $\psi(z; \eta_0)$ is a density function of Z under H'_0), and then evaluating integrated powers of (A) and (B). (When the $\hat{\eta}_0$ is equal to true parameters under H'_0 , the denoted $\hat{P}_{H'_0}$ is the significance level.) For example, Proposition 3 provides a method for comparing the test statistic (15) with the CUSUM statistic $\max_{1 \leq k \leq n} \max_a \Lambda_n(k, \hat{\theta}_0, a)$ under the condition, $\hat{\theta}_0 = \arg \max_a \prod_{i=1}^n f_0(X_i; a)$ (see Section 5). Obviously, thresholds $C_\alpha^L, C_\alpha^D : \hat{P}_{H'_0}\{L > C_\alpha^L \mid \hat{\eta}_0\} = \hat{P}_{H'_0}\{D > C_\alpha^D \mid \hat{\eta}_0\} = \alpha$ can be easily obtained by Monte Carlo methods. (We consider calculation α -relevant test-thresholds in Section 5.)

Remark 2. For the test based on (15), we have

$$\begin{aligned}\alpha &= \hat{P} \left\{ R_n > C \mid X_1, \dots, X_n \sim f_0(x; \hat{\theta}_0) \right\} = \int \prod_{i=1}^n f_0 \left(x_i; \hat{\theta}_0(x_1, \dots, x_n) \right) \\ &\times I \left\{ \prod_{i=1}^n f_0(x_i; \hat{\theta}_0) \leq \frac{1}{C} \sum_{k=1}^n \int \prod_{i=1}^{k-1} f_0(X_i; \bar{\theta}_0) \prod_{i=k}^n f_1(x_i; \bar{\theta}_1) dH(\bar{\theta}_0, \bar{\theta}_1) \right\} \prod_{i=1}^n dx_i \\ &\leq \frac{n}{C}.\end{aligned}$$

That is, we have the upper bound (that is independent of different conditions on the distribution of X_1, \dots, X_n) for α . If $\hat{\theta}_0$ is the maximum likelihood estimator, then

$$\sup_{\theta_0} P_{n+1} \{ R_n > C \} \leq \hat{P} \left\{ R_n > C \mid X_1, \dots, X_n \sim f_0(x; \hat{\theta}_0) \right\} \leq \frac{n}{C}.$$

Thus, selecting $C = n/\alpha$ determines a test with the level of significance that does not exceed α and ensures a p-value of the test. In accordance with the inequality $\alpha \leq n/C$, theoretically, values of α can be chosen as small as desired. For example, the following test statistics have the same non-asymptotic distribution-free upper bounds for the significance levels:

$$\begin{aligned}\text{SR:} & \quad \sum_{k=1}^n \prod_{i=1}^{k-1} \frac{f_0(X_i; \hat{\theta}_0^{(1,i-1)})}{f_0(X_i; \hat{\theta}_0)} \prod_{i=k}^n \frac{f_1(X_i; \hat{\theta}_1^{(k,i-1)})}{f_0(X_i; \hat{\theta}_0)}; \\ \text{CUSUM:} & \quad \max_{1 \leq k \leq n} \prod_{i=1}^{k-1} \frac{f_0(X_i; \hat{\theta}_0^{(1,i-1)})}{f_0(X_i; \hat{\theta}_0)} \prod_{i=k}^n \frac{f_1(X_i; \hat{\theta}_1^{(k,i-1)})}{f_0(X_i; \hat{\theta}_0)}; \\ \text{CUSUM:} & \quad \max_{1 \leq k \leq n} \int \prod_{i=1}^{k-1} \frac{f_0(X_i; \bar{\theta}_0)}{f_0(X_i; \hat{\theta}_0)} \Lambda_n(k, \hat{\theta}_0, \bar{\theta}_1) dH(\bar{\theta}_0, \bar{\theta}_1),\end{aligned}$$

where $\hat{\theta}^{(r,m)}$ is any estimator (e.g., the MLE) of θ based on $\{X_r, \dots, X_m\}$.

5 Monte Carlo simulations

Often in the change point literature, the CUSUM tests have been referred to as the best change point detection policy, particularly in the cases where observations have normal distributions.

Thus, we expect good properties of the CUSUM test, when X_i has the density function

$$(2\pi)^{-0.5} e^{-x^2 I\{i < \nu\}/2 - (x - \theta_1)^2 I\{i \geq \nu\}/2}, \quad i = 1, \dots, n \quad (\theta_0 = 0). \quad (18)$$

In this section, we present simulations to compare the power of the proposed tests with the power of the respective CUSUM detection schemes. To this end, when we estimate the probability of an event A , based on X_1, \dots, X_n , we repeat sampling from (18) 50000 times. The type I error of the considered tests is fixed at $\alpha = 0.05$.

The case when θ_0 and θ_1 are known. In this case, the proposed test has the form (6), and the CUSUM statistic is $\max_{1 \leq k \leq n} \exp\{\theta_1 \sum_{i=k}^n X_i - (n-k+1)\theta_1^2/2\}$. The results of Table 1 are predicted by Proposition 1, i.e. for different θ_1 and n :

$$\frac{1}{n} \sum_{k=1}^n P_k \{R_n(\theta_0, \theta_1) > C_{0.05}\} > \frac{1}{n} \sum_{k=1}^n P_k \left\{ \max_{1 \leq k \leq n} e^{\theta_1 \sum_{i=k}^n X_i - (n-k+1)\frac{\theta_1^2}{2}} > D_{0.05} \right\}.$$

Table 1

In this simulation study, we observed that the SR test was not only the more powerful on average, but it was also a more powerful detection scheme than CUSUM policy was. As an example we present Figure 2.

Figure 2

However, we doubt that

$$P_\nu \{R_n(0, \theta_1) > C_{0.05}\} \geq P_\nu \left\{ \max_{1 \leq k \leq n} e^{\theta_1 \sum_{i=k}^n X_i - (n-k+1)\frac{\theta_1^2}{2}} > D_{0.05} \right\}$$

for all θ_1 , ν , and n . To demonstrate the asymptotic result presented in Proposition 1, we define $\gamma = \sum_{k=1}^{50000} \exp\{\theta_1 \sum_{i=1}^k X_i - k\theta_1^2/2\}$ and depict the empirical tail probabilities $P_{n+1}\{R_n > u\}$ and $P_\infty\{\gamma > u\}$, where $\theta_1 = 0.5$ and $n = 20, 50, 75, 100$. (100000 Monte Carlo repetitions of R_n and γ were conducted.)

Figure 3

Figure 3 shows that when $n = 20$, the tail of the γ -distribution function cannot be approximated by the distribution function of the SR statistic. However, when $n > 50$, the tail probability of SR is close to that of γ .

The case when θ_1 is unknown. Here we analyze the test statistic (9) with $H(\theta) = \Phi\left(\frac{\theta-a}{\sigma}\right)$, i.e.

$$R_n(0) = \sum_{k=1}^n \frac{1}{(\sigma^2(n-k+1)+1)^{1/2}} \exp \left\{ -\frac{a^2}{2\sigma^2} + \frac{\sigma^2 (\sum_{i=k}^n X_i + a/\sigma^2)^2}{2(\sigma^2(n-k+1)+1)} \right\}.$$

(In this case, H is a commonly used conjugate prior in the context of Bayes Factors methods.)

The CUSUM statistic has the form of $\max_{1 \leq k \leq n} \exp\{0.5(\sum_{i=k}^n X_i)^2 / (n-k+1)\}$. Table 2

illustrates item (ii) of Proposition 2, where the equation

$$\begin{aligned} & \int P_k \left[\delta \text{ rejects } H_0 \middle| X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta), & i \geq k \end{cases}, i = 1, \dots, n \right] dH(\theta) \\ &= P_k \left[\delta \text{ rejects } H_0 \middle| X_i \sim \begin{cases} f_0(x; \theta_0), & i < k \\ f_1(x; \theta_1), & i \geq k \end{cases}, i = 1, \dots, n, \theta_1 \sim N(a, 1) \right] \end{aligned}$$

is applied with $\sigma^2 = 1$.

Table 2

However, in accordance with Table 3, the SR test is on average more powerful than the CUSUM test, even when we smooth the unknown θ_1 around $\theta_0 = 0$, which is known (i.e. $a = 0$).

Table 3

Figure 4 shows that when $a \neq \theta_0 = 0$, the SR test is more powerful than the CUSUM test.

Figure 4

The case, where θ_0 and θ_1 are unknown. Let the function $dH(\bar{\theta}_0, \bar{\theta}_1)$ in (15) have the simple form of $dH(\bar{\theta}_0, \bar{\theta}_1) = \varphi\left(\frac{\bar{\theta}_0 - a_0}{\sigma_0}\right) \varphi\left(\frac{\bar{\theta}_1 - a_1}{\sigma_1}\right) d\bar{\theta}_0 d\bar{\theta}_1$, where $\varphi = d\Phi$. Under H_0 , the maximum likelihood estimator of θ_0 is $\hat{\theta}_0 = \sum_{i=1}^n X_i/n$. Hence

$$\begin{aligned} R_n &= \sum_{k=1}^n \frac{\exp\left\{-0.5 \left(\sum_{i=1}^n X_i\right)^2/n - 0.5 \left(a_0^2/\sigma_0^2 + a_1^2/\sigma_1^2\right)\right\}}{\left(\sigma_0^2(k-1) + 1\right)^{1/2} \left(\sigma_1^2(n-k+1) + 1\right)^{1/2}} \\ &\quad \times \exp\left\{\frac{\sigma_0^2 \left(\sum_{i=1}^{k-1} X_i + a_0/\sigma_0\right)^2}{2\left(\sigma_0^2(k-1) + 1\right)} + \frac{\sigma_1^2 \left(\sum_{i=k}^n X_i + a_1/\sigma_1\right)^2}{2\left(\sigma_1^2(n-k+1) + 1\right)}\right\}, \text{ and} \\ \text{CUSUM} &= \max_{k=1, \dots, n} \exp\left\{\frac{\left(\sum_{i=k}^n X_i\right)^2}{2(n-k+1)} + \frac{(n-k+1)}{2} \left(\frac{\sum_{i=1}^n X_i}{n}\right)^2\right. \\ &\quad \left. - \frac{\sum_{i=1}^n X_i \sum_{i=k}^n X_i}{n}\right\}. \end{aligned}$$

Let $\sigma_0 = \sigma_1 = 1$ and assume that $\theta_0 = 0$. Following Proposition 3 for the SR and CUSUM tests, we fix $\hat{\theta}_0(X_1, \dots, X_n) = \sum_{i=1}^n X_i/n$ and obtain test-thresholds $C_{0.05}$ and $D_{0.05}$ by

$$\begin{aligned} \alpha = 0.05 &= \int \prod_{i=1}^n f_0 \left(x_i; \hat{\theta}_0(x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i}{n} \right) I \{ \delta \text{ rejects } H_0 \} \prod_{i=1}^n dx_i \\ &= \frac{1}{(2\pi)^{n/2}} \int e^{-\frac{(\sum_{i=1}^n x_i)^2}{2n}} I \{ \delta \text{ rejects } H_0 \} e^{-\frac{\sum_{i=1}^n x_i^2}{2}} \prod_{i=1}^n dx_i \\ &= E_{n+1} \left(e^{-\frac{(\sum_{i=1}^n X_i)^2}{2n}} I \{ \delta(X_1, \dots, X_n) \text{ rejects } H_0 \} \right). \end{aligned}$$

(We numerically evaluated the corresponding classical type I error of the SR and CUSUM procedures: $P_{n+1}\{\text{SR rejects } H_0\} \simeq P_{n+1}\{\text{CUSUM rejects } H_0\} \simeq 0.02$.)

Similar to Table 2 in the previous paragraph, Table 4 relates to Proposition 3.

Table 4

In the considered cases, Table 5 concludes that regardless of a prior distribution of the unknown parameters, the SR test is on average more powerful than the CUSUM test. For example, when we set $n = 100$, $\theta_1 = 0.7$, and $dH(\bar{\theta}_0, \bar{\theta}_1) = \varphi(\bar{\theta}_0) \varphi(\bar{\theta}_1 - 1) d\bar{\theta}_0 d\bar{\theta}_1$, the average power of the SR test is 0.5048 whereas the average power of the CUSUM test is 0.2389.

Table 5

On the level of the P_k -regime, the situations of $\{a_0 = 0, a_1 = 0\}$, $\{a_0 = 0, a_1 = 0.5\}$, $\{a_0 = 0.5, a_1 = 0\}$, and $\{a_0 = 0, a_1 = 1\}$ are displayed in Figure 5.

Figure 5

Note that, certainly, the case $\{a_0 = 0, a_1 = 0\}$ (both the pre-and-post change parameters are expected to be around 0) is strange and the case $\{a_0 = 0.5, a_1 = 0\}$ corresponds to the misunderstanding of a physical process presented by an analyzed data set. (In this situation, (10) or (11) type functions for H can be recommended because such functions regard no information about $\theta_0 < \theta_1$.) Even in these cases, however, in most situations, the SR procedure is still more powerful than the CUSUM procedure.

Thus, the results of the Monte Carlo simulations experimentally confirm the existence of a practical meaning of the proposed tests. Although we did not utilize any technique to choose priors for the unknown parameters, the Monte Carlo study showed that the proposed change point detections are somewhat better than the well-known CUSUM-procedures.

6 Remarks

1. Assume that θ_0 and θ_1 are multidimensional parameters. In this case, the proposed change point detection method does not need to be changed. When the dimension of the vector of unknown parameters is close to the number of observations, we suspect that the proposed testing is very reasonable. Note that the proposed method can be easily adapted to regression models. In several situations, biostatisticians deal with "wide and short" data, where the number of explanatory variables is close to the number of observations. In these situations, applications of tests based on the estimation of unknown parameters are problematic.

2. The proposed method can be modified if X_i , $i \geq 1$ are non-independent random variables. For example, in this case, we denote

$$R_n(\theta_0, \theta_1) = \sum_{k=1}^n \prod_{i=k}^n \frac{f_1(X_i|X_1, \dots, X_{i-1}; \theta_1)}{f_0(X_i|X_1, \dots, X_{i-1}; \theta_0)}.$$

Obviously, the modified tests have the optimal properties that are mentioned in the propositions of this paper.

3. Consider the problem of change-points with epidemic alternatives (e.g., Vexler, 2006, 2008), i.e., X_i has density $f_0 I\{i < \nu_1\} + f_1 I\{\nu_1 \leq i < \nu_2\} + f_0 I\{i \geq \nu_2\}$, where $i = 1, \dots, n$ and $\nu_1 < \nu_2$ are unknown. In order to test for $\nu_1 > n$ versus $\nu_1 \leq n$, the proposed method can be easily extended, i.e.

$$R_n(\theta_0, \theta_1) = \sum_{k=1}^n \sum_{r=1}^k \prod_{i=r}^k \frac{f_1(X_i; \theta_1)}{f_0(X_i; \theta_0)}.$$

4. In accordance with Remark of Section 2, Propositions 2 and 3, weights $\{q_k, k = 1, \dots, n\}$ and/or prior H can be chosen with respect to a special area of parameters ν, θ_0, θ_1 , under the alternative hypothesis, where the maximum of the test's power is desired. Thus, the proposed change point tests provide maximum integrated power in the area that can correspond to the tester's interests.

Acknowledgements

This research was supported by the Intramural Research Program of the National Institute

of Child Health and Human Development, National Institutes of Health. The opinions expressed are those of the authors and not necessarily of the National Institutes of Health. The authors are indebted to Professors Kai Fun Yu and Aiyi Liu for many helpful discussions and comments. We are grateful to the editor, associate editor, and two referees for their helpful comments that clearly improved this paper.

References

- Broström, G. (1997). A martingale approach to the changepoint problem *J. Amer. Statist. Assoc.* **92**, 1177-1183.
- Brown, R. L., Durbin, J. & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time (with discussion). *J. Roy. Statist. Soc. Ser. B* **37**, 149-192.
- Csörgő, M. & Horváth, L. (1997). *Limit theorems in change-point analysis*. Wiley, New York.
- Gombay, E. & Horvath, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Process. Appl.* **50**, 161-171.
- Gurevich, G. & Vexler, A. (2005). Change point problems in the model of logistic regression. *J. Statist. Plann. Inference* **131**, 313-331.
- James, B., James K. L. & Siegmund D. (1987). Tests for a change-point. *Biometrika* **74**, 71-83.
- Krieger, A. M., Pollak, M. & Yakir, B. (2003). Surveillance of a simple linear regression. *J. Amer. Statist. Assoc.* **98**, 456-469.
- Lai, T.L. (1995). Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B.* **57** 4, 613-658.
- Lai, T.L. (2001). Sequential analysis: some classical problems and new challenges. *Statist. Sinica* **11**, 303-408.

- Lorden, G. (1967). Integrated risk of asymptotically bayes sequential tests. *Ann. Math. Statist.* **38**, 1399-1422.
- Lorden, G. & Pollak, M. (2004). Non-anticipating estimation applied to sequential analysis and changepoint detection. *Ann. Statist.* **33**, 1422-1454.
- Marden, J. I. (2000). Hypothesis testing: from p values to Bayes factors. *J. Amer. Statist. Assoc.* **95**, 1316-1320.
- 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**, 100-114.
- Page, E. S. (1955). A test for a change in a parameter occurring at an unknown Point. *Biometrika* **42**, 523-527.
- Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point with zero-one observations. *Biometrika* **67**, 79-84.
- Ploberger, W. & Kramer, W. (1992). The Cusum Test with Ols Residuals. *Econometrica* **60**, 271-285.
- Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.* **13**, 206-227.
- Sen, A. & Srivastava, M. S. (1975). On tests for detecting change in mean. *Ann. Statist.* **1**, 98-108.
- Vexler, A. (2006). Guaranteed testing for epidemic changes of a linear regression model. *J. Statist. Plann. Inference* **136**, 3101-3120.
- Vexler, A. (2008). Martingale type statistics applied to change points detection. *Comm. Statist. Theory Methods* **37**, 1207-1224.
- Yakir, B. (1998). On the average run length to false alarm in surveillance problems which possess an invariance structure. *Ann. Statist.* **26**, 1198-1214.

Albert Vexler, Department of Biostatistics, The State University of New York, Buffalo, NY,
14214, USA.

E-mail: avexler@buffalo.edu

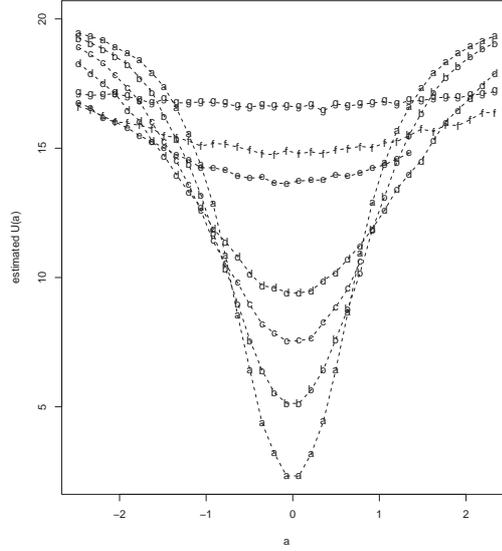


Figure 1: The Monte Carlo estimator of the function $U(a)$ from (14). The curves -a-, -b-, -c-, -d-, -e-, -f- and -g- correspond to $\sigma = 0.25, 0.5, 0.75, 1, 2, 2.5$ and 4 , respectively.

Table 1: The Monte Carlo estimators of the average powers $\sum_{k=1}^n P_k\{\delta \text{ rejects } H_0\}/n$, where the decision rules δ are the SR and CUSUM tests.

		$\theta_1 = 0.1$	$\theta_1 = 0.2$	$\theta_1 = 0.3$	$\theta_1 = 0.5$	$\theta_1 = 0.7$	$\theta_1 = 1$
$n = 10$	SR	0.0738	0.1076	0.1516	0.2708	0.4132	0.6217
	CUSUM	0.0732	0.1067	0.1457	0.2651	0.4087	0.6123
$n = 20$	SR	0.0892	0.1551	0.2543	0.4292	0.5999	0.7898
	CUSUM	0.0842	0.1356	0.2074	0.3953	0.5858	0.7758
$n = 100$	SR	0.1549	0.3503	0.5639	0.8230	0.9165	0.9591
	CUSUM	0.1488	0.3343	0.5452	0.8042	0.9025	0.9544

Table 2: The Monte Carlo estimators of the average powers $\sum_{k=1}^n P_k\{\delta \text{ rejects } H_0\}/n$, where the decision rules δ are the SR and CUSUM tests and the unknown post-change parameter θ_1 is generated from normal distributions.

	θ_1	$\sim N(0, 1)$	$\sim N(0.5, 1)$	$\sim N(1, 1)$
$n = 20$	SR	0.4883	0.5371	0.6352
	CUSUM	0.4625	0.5027	0.6081
$n = 100$	SR	0.6997	0.7491	0.8332
	CUSUM	0.6879	0.7178	0.7923

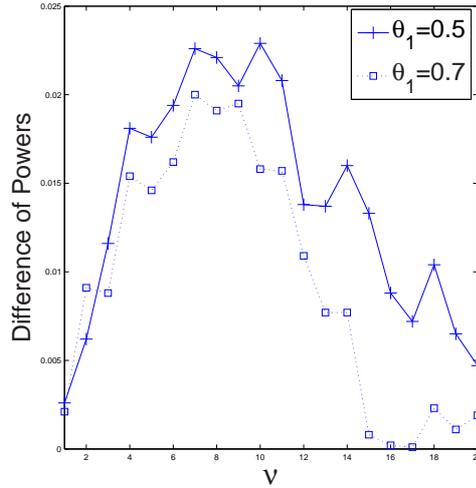


Figure 2: The Monte Carlo estimators of $P_\nu\{\text{SR rejects } H_0\} - P_\nu\{\text{CUSUM rejects } H_0\}$, where $P_{n+1}\{\text{SR rejects } H_0\} = P_{n+1}\{\text{CUSUM rejects } H_0\} = 0.05$ and $n = 20$.

Table 3: The Monte Carlo estimators of the average powers of the SR and CUSUM tests, for different values of the post-change parameter θ_1 .

$\sigma^2 = 1$		$n = 20$			$n = 100$		
		$a = 0.0$	$a = 0.5$	$a = 1.0$	$a = 0.0$	$a = 0.5$	$a = 1.0$
$\theta_1 = 0.1$	CUSUM		0.0598			0.0813	
	SR	0.0610	0.0699	0.0774	0.1009	0.1059	0.1233
$\theta_1 = 0.5$	CUSUM		0.2751			0.7127	
	SR	0.3071	0.3539	0.3741	0.7649	0.7798	0.8091
$\theta_1 = 0.7$	CUSUM		0.4707			0.8526	
	SR	0.4835	0.5509	0.5699	0.8696	0.8874	0.8959

Table 4: The Monte Carlo estimators of the average powers $\sum_{k=1}^n P_k\{\delta \text{ rejects } H_0\}/n$, where the zdecision rules δ are the SR and CUSUM tests and the unknown pre-post-change parameters θ_0, θ_1 are generated from normal distributions.

$\theta_j \sim N(a_j, 1), j = 0, 1$		$a_0 = 0, a_1 = 0$	$a_0 = 0, a_1 = .5$	$a_0 = .5, a_1 = .0$	$a_0 = 0, a_1 = 1$
$n = 20$	RS	0.2824	0.3191	0.3129	0.3697
	CUSUM	0.2173	0.2366	0.2359	0.2902
$n = 100$	RS	0.5534	0.5872	0.5854	0.6489
	CUSUM	0.4524	0.4731	0.4724	0.5292

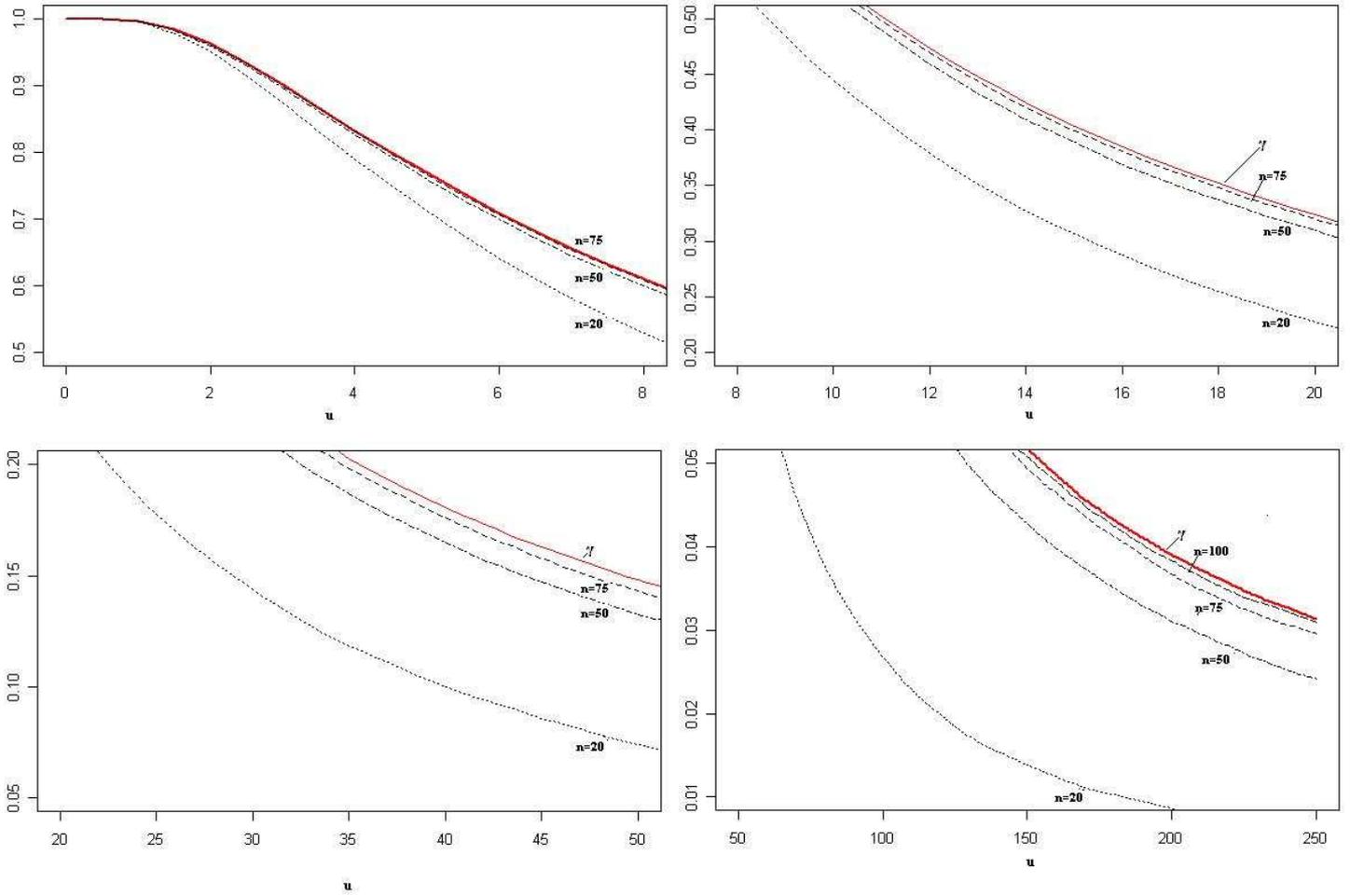


Figure 3: The Monte Carlo probabilities $P_{n+1}\{R_n > u\}$ and $P_{\infty}\{\gamma > u\}$ are depicted by the line types that are specified as "dotted", "dotdash", "dashed" and "longdash" corresponding to $n = 20, 50, 75$ and $n = 100$, respectively. Here γ is denoted in Proposition 1 and $\theta_1 = 0.5$.

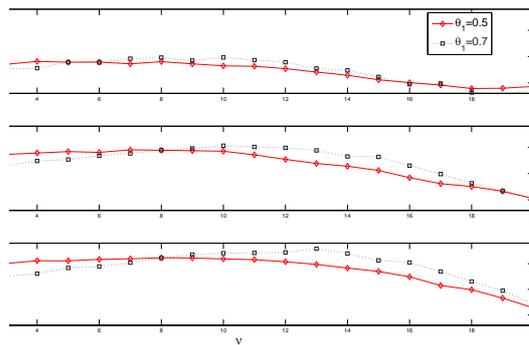


Figure 4: The Monte Carlo estimators of $P_{\nu}\{SR \text{ rejects } H_0\} - P_{\nu}\{CUSUM \text{ rejects } H_0\}$, where $P_{n+1}\{SR \text{ rejects } H_0\} = P_{n+1}\{CUSUM \text{ rejects } H_0\} = 0.05$ and $n = 20$.

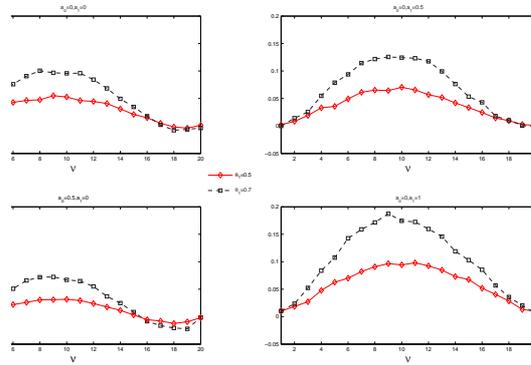


Figure 5: The Monte Carlo estimators of $P_\nu\{\text{SR rejects } H_0\} - P_\nu\{\text{CUSUM rejects } H_0\}$, where $\hat{P}_{n+1}\{\text{SR rejects } H_0\} = \hat{P}_{n+1}\{\text{CUSUM rejects } H_0\} = 0.05$ and $n = 20$.

Table 5: The Monte Carlo estimators of the average powers of the SR and CUSUM tests, for different values of the post-change parameter θ_1 (the pre-change parameter θ_0).

		$\sigma_0 = \sigma_1 = 1$	$a_0 = 0, a_1 = 0$	$a_0 = 0, a_1 = .5$	$a_0 = .5, a_1 = .0$	$a_0 = 0, a_1 = 1$
$n = 20$	$\theta_1 = 0.1$	RS	0.0220	0.0200	0.0186	0.0254
		CUSUM	0.0153			
	$\theta_1 = 0.5$	RS	0.0538	0.0632	0.0401	0.0864
		CUSUM	0.0331			
	$\theta_1 = 0.7$	RS	0.0905	0.1104	0.0766	0.1460
		CUSUM	0.0558			
$n = 100$	$\theta_1 = 0.1$	RS	0.0169	0.0189	0.020	0.0239
		CUSUM	0.0148			
	$\theta_1 = 0.5$	RS	0.1999	0.2289	0.2264	0.2451
		CUSUM	0.0958			
	$\theta_1 = 0.7$	RS	0.4245	0.4689	0.4519	0.5048
		CUSUM	0.2389			