An extension of a change point problem

Running title of the paper: Extended Changepoint Problem

ALBERT VEXLER*†, CHENGQING WU‡, AIYI LIU‡, BRIAN W. WHITCOMB‡ and ENRIQUE F. SCHISTERMAN‡

†*Department of Biostatistics, The New York State University at Buffalo, Buffalo, NY 14214, USA*

‡*Division of Epidemiology, Statistics and Prevention Research,NICHD, NIH, USA*

Abstract

We consider a specific classification problem in the context of change point detection. We present generalized classical maximum likelihood tests for homogeneity of the observed sample in a simple form which avoids the complex direct estimation of unknown parameters. The paper proposes a martingale approach to transformation of test statistics. For sequential and retrospective testing problems, we propose adapted Shiryayev-Roberts statistics in order to obtain simple tests with asymptotic power one. An important application of the developed methods is to the analysis of exposures's measurements subject to limits of detection in occupational medicine.

Key Words: Doob decomposition, change point, classification, CUSUM statistics, likelihood ratio, limit of detection, martingale, martingale transforms, Shiryayev-Roberts statistics.

AMS subject classification: Primary: 62C25, 62L10, 62N03, 62F03; Secondary: 60G46.

1 Introduction

Suppose that a series of observations are sequentially or retrospectively surveyed:

$$\begin{aligned} z_i &= x_i I\{\nu_i \geq d\} + y_i I\{\nu_i < d\}, \\ \gamma_i &= \begin{cases} \nu_i, & \text{if } \nu_i \text{ is observed (or known);} \\ z_i, & \text{if } \nu_i \text{ is not observed,} \end{cases} \qquad i \geq 1, \qquad (1.1) \end{aligned}$$

*Correspondence to: Albert Vexler
The State University of New York
Department of Biostatistics
Farber Hall, Rm 246
Buffalo, NY 14214
E-mail: avexler@buffalo.edu

where x_i, y_i, ν_i are some real random variables, d is a fixed threshold value, $I\{\cdot\}$ is the indicator function. Without loss of generality and for the sake of clarity, we assume that $\{x_i,\ i \geq 1\}$ are independent identically distributed ($i.i.d.$) random variables with density function f_x, and are independent of the random variables $\{y_i,\ i \geq 1\}$, also i.i.d. with density function f_y; $\{\nu_i,\ i \geq 1\}$ are independent random variables. We are primarily concerned with testing for homogeneity of the observed samples in (1.1), i.e.

$$\begin{aligned}
\mathbf{H_0} \quad &: \quad z_1, \ldots, z_n \text{ are each distributed according to the density } f_x, \\
&\quad\ \ \textbf{versus} \\
\mathbf{H_1} \quad &: \quad \text{for all } i = 1, \ldots, n,\ \ z_i \text{ is distributed according to the density} \\
&\quad\ \ f_{z\,i}(u; d) \equiv \frac{\partial P\{x_i < u, \nu_i \geq d\}}{\partial u} + \frac{\partial P\{y_i < u, \nu_i < d\}}{\partial u} \\
&\quad\ \ \text{for some unknown } d,
\end{aligned} \tag{1.2}$$

where n is fixed for retrospective testing but random for sequential testing.

The testing for such hypotheses is an extension of the well known change point problems. Clearly, if $\nu_i = -i$ then it becomes the standard change point problem (e.g. Page, 1954; Lai, 1995); if for some a, $\nu_i = (i - a)^2$, then it reduces to the testing for epidemic changes (e.g. Yao, 1993; Vexler, 2006); if $\{\nu_i,\ i \geq 1\}$ are i.i.d. random variables independent of $\{x_i,\ i \geq 1\}$ and $\{y_i,\ i \geq 1\}$, then the problem becomes testing for an identification of a mixture distribution (e.g. Garel, 2005). All these situations have been intensively investigated separately.

In addition to change point related problems, (1.1) with (1.2) covers other important applications as well. Consider the case when $\nu_i = x_i$, for example, which has not been well addressed in the literature in the context of hypothesis testing. Suppose $\nu_i = x_i$ are not observed for all $i \geq 1$, and therefore $\gamma_i = z_i$ in (1.1). Then $f_{z\,i}$ from (1.2) has the form

$$f_{z\,i}(u; d) = f_z(u; d) \quad \equiv \quad f_x(u) I\{u \geq d\} + f_y(u) F_x(d),\ i \geq 1, \tag{1.3}$$

where F_x is the distribution function of x_1. An important application of this reduced model is to the analysis of exposures' measurements subject to some limits of detection (LOD), a situation frequently encountered in many medical areas such as occupational medicine (e.g. Cooper et al., 2002; Helsel, 2005; Vexler et al, 2006). In this case the observed data contain measurements of the exposure $\{x_i\}$ and instrument noises $\{y_i\}$. In order to evaluate the operating characteristics based on such data, various authors have provided parametric (in essence) approaches to managing left-censored data of this type (e.g. Lubinet et al, 2004; Hornung and Reed, 1990; Finkelstein et al, 2001; Schisterman et al, 2006; Vexler et al, 2006).

For measurements of many exposures, the detection threshold is experimentally determined as a function of the variance of a series of blanks or spiked samples of some known concentration. Conventionally, the value corresponding to three standard deviation from the experiment is defined as the LOD and utilized as the detection threshold (e.g. Keith et al, 1983; Helsel, 2005). In these cases, observed data will be a mixture of observations truly below the detection threshold, falsely above the detection threshold and truly above the detection threshold.

The change point models were introduced in the context of quality control for the purpose of determining a time, T, to distinguish between two states-control and lack thereof (e.g. Lai, 1995). Although the LOD problem is in a sense also one of quality control, there are important departures from the classical change point model; the detection threshold problem involves some amount of censored observations and the change point itself is a value taken by a random variable, x_i, rather than some value of T. An empirical approach, applied in practice with the objective to obtain the limit of instrumentation, is close to change point sequential procedures in a general sense (e.g. Helsel, 2005).

Following these motives, in the present paper we focus on the sequential case of (1.1), (1.2), where $\{\nu_i,\, i \geq 1\}$ are not observed. In many problems related to detection of non-homogeneity of the observed sample, a powerful or sufficient test statistic (based, for example, on likelihood ratios) is transformed (sometimes compromising the power of the test, e.g. Lai, 2001: p. 398) to obtain target properties of the detection procedure (e.g. Robbins and Siegmund, 1973; Dragalin, 1997; Lorden and Pollak, 2005; Gurevich and Vexler, 2005; Vexler, 2006). The paper proposes a martingale methodology to provide a technique for such transformations. In Section 2 we introduce the adaptive procedures, and propose a martingale methodology in Section 3 for the targeted adaptive test-statistics. Section 4 represents the results of several Monte Carlo simulations.

2 Adaptive Procedures

Let P_d, E_d denote respectively the probability and expectation for a given d. The case $d = -\infty$ corresponds to the case where observations are from the same density function f_x. The classical construction of test statistics for (1.2) includes a consideration of the likelihood ratio

$$\Lambda_n(d) \equiv \frac{f(z_1, \ldots, z_n| \text{ under } H_1)}{f_x(z_1, \ldots, z_n)} = \prod_{i=1}^{n} \frac{f_{z\,i}(z_i; d)}{f_x(z_i)}. \tag{2.1}$$

Since the parameter d is unknown, we apply the maximum likelihood method to estimate it, however in a specific context. To this end, we arrange the sequence $\{\gamma_i,\ i = 1, \ldots, n\}$ in decreasing order: $\infty = \gamma_{(0:n)} > \gamma_{(1:n)} \geq \gamma_{(2:n)} \geq \ldots \geq \gamma_{(n:n)} > \gamma_{(n+1:n)} = -\infty$. Thus d is estimated by $\gamma_{(k-1:n)}$, where $k = \arg\max_l \prod_i f_{z\,i}(z_i; \gamma_{(l-1:n)})$. Since $1 = \sum_{k=1}^{n+1} I\{\gamma_{(k-1:n)} \geq d > \gamma_{(k:n)}\}$ and therefore $\Lambda_n(d) = \sum_{k=1}^{n+1} \Lambda_n(d) I\{\gamma_{(k-1:n)} \geq d > \gamma_{(k:n)}\}$, we define the maximum likelihood estimator of $\Lambda_n(d)$ in the form $\Lambda_n = \max_k \Lambda_n(\gamma_{(k-1:n)})$.

A common approach in the change point literature is to use the Shiryaev-Roberts (SR) statistic in replacement of the maximum likelihood ratios, leading to the SR change point detection strategy aimed at obtaining guaranteed characteristics (e.g. Pollak, 1985, 1987; Lorden and Pollak, 2005; Vexler, 2006). Note that in a sequential context of a change point detection, SR procedures are optimal methods for statistical monitoring (e.g. Pollak, 1985). Following this remark we propose the test statistics for (1.2) based on some variation of $R_n = \sum_k \Lambda_n(\gamma_{(k-1:n)})$. Formally, we denote, for all $m \geq 1$,

$$\begin{aligned} R_m \equiv & \sum_{k=2}^{m} \Lambda_m(\gamma_{(k-1:m)}) + \prod_{i=1}^{m} \frac{f_y(z_i)}{f_x(z_i)} = \sum_{k=2}^{m} \prod_{i=1}^{m} \frac{f_{z\,i}(z_i; \gamma_{(k-1:m)})}{f_x(z_i)} \\ & + \prod_{i=1}^{m} \frac{f_y(z_i)}{f_x(z_i)} = \sum_{k=2}^{m} \prod_{i=1}^{m} \frac{f_{z\,i}(z_i; \gamma_{k-1})}{f_x(z_i)} + \prod_{i=1}^{m} \frac{f_y(z_i)}{f_x(z_i)} \\ = & \sum_{k=2}^{m} \Lambda_m(\gamma_{k-1}) + \prod_{i=1}^{m} \frac{f_y(z_i)}{f_x(z_i)}, \end{aligned} \tag{2.2}$$

where without ties: $\gamma_{(1:m)} > \gamma_{(2:m)} > \ldots > \gamma_{(m:m)}$ and $\sum_2^1 = 0$. Clearly, if $\{\nu_i = -i,\ i \geq 1\}$ are known then R_m is the classical SR statistic and

Λ_n is the standard CUSUM statistic. Note that (2.2) is a very simple representation of the main component of the test statistic.

Note that, without restrictions on the possible values of the unknown d, the application of the test statistic in the form $\max_{\hat{d}} \Lambda_n(\hat{d})$ is a very complex problem, which depends heavily on the type of density functions of the stated problem. Moreover, using $\max_{-\infty \le \hat{d} \le \infty} \Lambda_n(\hat{d})$ needs compensators for the increasing statistic.

Sequential approach. Suppose one is able to sequentially observe the series of z_1, $z_2, \ldots$ by (1.1). A detection scheme consists of a stopping time N for the process $\{z_1, z_2, \ldots\}$ at which one stops sampling and declares rejection of H_0. In general the stopping time N is determined by

$$N(C) = \inf\left\{n \ge 1 : \; R_n^{(S)} \ge C\right\}, \tag{2.3}$$

where $R_n^{(S)}$ is a reasonable transformation of R_n by (2.2). To control the level of the average run length to false alarms (i.e. $E_{-\infty}N$), we require the stopping rule N satisfy $E_{-\infty}N(C) \ge B$ for some specified level B (e.g. Pollak, 1985,1987; Yakir, 1995).

Retrospective approach. Let the sample size n in (1.1) and (1.2) be fixed. We transform statistic R_n from (2.2) to an appropriate test statistic $R_n^{(R)}$. With R_n^R we reject H_0 iff

$$R_n^{(R)} \quad > \quad C, \tag{2.4}$$

for some threshold $C > 0$. Because $R_n^{(R)}$ is based on generalized maximum likelihood statistics, the behavior of $R_n^{(R)}$ under regime P_d (where $d \ne -\infty$) is quite predictable. Moreover, we define (2.4) as an asymptotic power one test (if $d \ne -\infty$, $C < \infty$ are fixed and $n \to \infty$). It is widely known in change point literature (e.g. Lai, 1995; Gordon and Pollak, 1995; Vexler, 2006) that such tests have high power. However, $R_n^{(R)}$ has been shown to behave very erratically under the null hypothesis H_0. Hence, evaluation of the significance level of test (2.4) is a major issue. In order to control the significance level of the test, we adapt a form of $R_n^{(R)}$, as described below.

Adaptation. A classical approach to the adaption of test statistics is to apply a transformation of the statistic, after which the H_0-martingale property is achieved (e.g. Brostrom, 1997; Krieger, Pollak and Yakir, 2003; Gurevich and Vexler, 2005). Basically, this method can be used in cases where the density of the observed samples, possibly being transformed, under H_0 is completely known (e.g. Yakir, 1998; Krieger, Pollak and Yakir, 2003; Vexler, 2006). Hence, practically, the expectation $E_{-\infty}$ of the test statistic can be evaluated. Consider R_n in (2.2) for example. If $\{\nu_i = -i,\ i \geq 1\}$ are known then $R_n - n$ is the H_0-martingale with zero expectation. Therefore, for a stopping (Markov) time N, by applying the optimal sampling theorem we can show $E_{-\infty}N = E_{-\infty}R_N$. This property is widely used to obtain certain desirable characteristics of change point procedures (e.g. Robbins and Siegmund, 1973; Dragalin, 1997; Lorden and Pollak, 2005; Vexler, 2006). However, if $\nu_i = x_i$ are not observed for all $i \geq 1$, the statistic R_n does not posses the H_0-martingale property. And hence R_n has to be adapted.

3 Decomposition of Test Statistics and Martingale Approximation

3.1 The Technique

Define $\Im_n \equiv \sigma\{z_1, \ldots, z_n\}$ as the sigma algebra based upon $\{z_1, \ldots, z_n\}$ ($\Im_0 = \{\emptyset\}$). Since by (2.2)

$$\begin{aligned} E_{-\infty}R_n\Big|\Im_{n-1} &= \sum_{k=2}^{n-1}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{k-1})}{f_x(z_i)}\left(E_{-\infty}\frac{f_{z\,n}(z_n;\gamma_{k-1})}{f_x(z_n)}\Big|\Im_{n-1}\right) \\ &\quad+\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)}\left(E_{-\infty}\frac{f_{z\,n}(z_n;\gamma_{n-1})}{f_x(z_n)}\Big|\Im_{n-1}\right) \\ &\quad+\prod_{i=1}^{n-1}\frac{f_y(z_i)}{f_x(z_i)}\left(E_{-\infty}\frac{f_y(z_n)}{f_x(z_n)}\Big|\Im_{n-1}\right) \\ &= R_{n-1}+\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)}, \end{aligned} \tag{3.1}$$

one can show that R_n is a positive H_0-submartingale with respect to $\Im_n$. However, in accordance, for example, with Krieger, Pollak and Yakir (2003, Section 2), it is natural to require that $R_n^{(S)} - n$ is a H_0-martingale with respect to $\Im_n$, where $R_n^{(S)}$, being a non-negative H_0-submartingale applied to (2.3), is a reasonable modification of R_n defined to be

$$R_n^{(S)} = W_n R_n, \; W_n = \left(R_{n-1} + \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)} \right)^{-1} (W_{n-1}R_{n-1}+1) \in \Im_{n-1}, \; W_0 = 0, R_0 = 0, \tag{3.2}$$

and hence by virtue of (3.1)

$$E_{-\infty}\left(R_n^{(S)} - n\right) \Big| \Im_{n-1} = R_{n-1}^{(S)} - (n-1).$$

It is clear that if $\gamma_i = -i$ (i.e. $\{\nu_i, i \geq 1\}$ are known) then $W_n = 1$, for all $n \geq 1$.

Brostrom (1997) proposes, for a specific change point problem, to extract a H_0-martingale component of the primary test statistics and to apply it to a test. Here we apply a similar strategy to obtain the adapted test statistics. Suppose under H_0 we can split the test statistic into two parts, a martingale component and a predictable component, where under H_0 the second component, for instance, slowly increases when n increases. While under H_1 the H_0-martingale component explodes (e.g. the H_0-martingale component is based on a variety of f_{H_1}(a sample)$/f_{H_0}$(a sample), where $E_{H_0} f_{H_1}$(an observation)/ f_{H_0}(an observation) $= 1$ and $E_{H_1} \ln(f_{H_1}$ (an observation) / f_{H_0}(an observation)) ≥ 0), but the predictable component still slowly changes (respectively with the first component). Therefore we apply minimum modification to the transformation of the H_0-martingale component, whereas the second component can be more flexibly changed for the target adaption of the test statistics. To this end, we present the Doob decomposition result.

Lemma 3.1. *If S_n is a submartingale with respect to $\Im_n$, then S_n can be uniquely written as $S_n = \langle S\rangle_n^{(M)} + \langle S\rangle_n^{(P)}$ with the following properties: $\langle S\rangle_n^{(M)}$ is a martingale component of S_n (i.e. $(\langle S\rangle_n^{(M)}, \Im_n)$ is a martingale); $\langle S\rangle_n^{(P)}$ is $\Im_{n-1}$-measurable, with that property being called predictable, $\langle S\rangle_n^{(P)} \geq \langle S\rangle_{n-1}^{(P)}$ (a.s.), $\langle S\rangle_1^{(P)} = 1$ and*

$$\langle S\rangle_n^{(P)} - \langle S\rangle_{n-1}^{(P)} = E\left(S_n - S_{n-1}\right) \Big| \Im_{n-1}.$$

In addition, let us represent the widely known definition of martingale transforms. If $\langle G_X \rangle_n^{(M)}$ is a martingale with respect to $\Im_n$ and $\langle G_Y \rangle_n^{(M)}$ is the difference $\langle G_X \rangle_n^{(M)} - \langle G_X \rangle_{n-1}^{(M)}$, then the martingale transform $\langle G'_X \rangle_n^{(M)}$ of $\langle G_X \rangle_n^{(M)}$ is given by the formula

$$\langle G'_X \rangle_n^{(M)} = \langle G'_X \rangle_{n-1}^{(M)} + a_n \langle G_Y \rangle_n^{(M)}, \tag{3.3}$$

where $a_n \in \Im_{n-1}$. Such transformations have a long history and interesting interpretations in terms of gambling. An interpretation is that if we have a fair game, we can choose the size and side of our bet at each stage based on the prior history and the game will continue to be fair. An association of a stochastic game with a change point detection is presented, for example, by Ritov (1990). Therefore, if the martingale component of R_n is important, it is natural to obtain $R_n^{(S)}$ such that $\langle R^{(S)} \rangle_n^{(M)}$ ($\langle R^{(S)} \rangle_n^{(M)} \in \Im_n$, but $\langle R^{(S)} \rangle_n^{(M)} \notin \{\Im_{n-1} \cap \{\langle R \rangle_k^{(M)}\}_{k=1}^{n-1}\}$) is the martingale transform of $\langle R \rangle_n^{(M)}$ ("transition: martingale-martingale"). Note that obtaining $R_n^{(S)}$ that satisfies this condition is not trivial, because by directly using the formula (3.3), we have

$$a_n = \frac{\langle R^{(S)} \rangle_n^{(M)} - \langle R^{(S)} \rangle_{n-1}^{(M)}}{\langle R \rangle_n^{(M)} - \langle R \rangle_{n-1}^{(M)}} \in \Im_n \text{ (we need } \Im_{n-1}).$$

Let the difference $\langle S \rangle_n^{(P)} - \langle S \rangle_{n-1}^{(P)} = E\left(S_n - S_{n-1}\right) \Big| \Im_{n-1}$ be called δ_n-predictor of the sequence S_n. Suppose our objective is a transition from R_n to a nonnegative H_0-submartingale $R_n^{(S)}$ with some specified $\eta_{n-1}^{(S)} \geq 0$, which is δ_n-predictor of $R_n^{(S)}$ (i.e. $E_{-\infty} R_n^{(S)} \Big| \Im_{n-1} = R_{n-1}^{(S)} + \eta_{n-1}^{(S)}$). We have

Proposition 3.1. *Let $\eta_{n-1}^{(S)} \geq 0$, $\eta_{n-1}^{(S)} \in \Im_{n-1}$ be specified and W_n be defined as*

$$W_n = \left(R_{n-1} + \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i; \gamma_{n-1})}{f_x(z_i)} \right)^{-1} \left(W_{n-1} R_{n-1} + \eta_{n-1}^{(S)} \right). \tag{3.4}$$

Then, under H_0,

1. $R_n^{(S)} = W_n R_n$ *is the nonnegative submartingale with respect to* $\Im_n$*;*
2. $\langle R^{(S)} \rangle_n^{(M)}$ *is the martingale transform of* $\langle R \rangle_n^{(M)}$*.*

Proof. In Appendix.

Consider an inverse version of Proposition 3.1.

Proposition 3.2. *Assume, under H_0, $a_n \in \Im_{n-1}$, $n \geq 1$ exist, such that*

$$\langle R' \rangle_n^{(M)} \;=\; \langle R' \rangle_{n-1}^{(M)} + a_n \left(\langle R \rangle_n^{(M)} - \langle R \rangle_{n-1}^{(M)} \right),$$

where R'_n is a submartingale with respect to $\Im_n$ and a specified $\eta_{n-1}^{(S)} \geq 0$, which is δ_n-predictor of R'_n. Then $R'_n = W_n R_n + \varsigma_n$, where W_n is defined by (3.4), $\{\varsigma_n, \Im_n\}$ is a martingale, and ς_n is a martingale transform of $\langle R \rangle_n^{(M)}$.

Proof. In Appendix.

Thus, for a specified $\eta_{n-1}^{(S)} \geq 0$ we can also obtain a submartingale $R'_n = W_n R_n + \varsigma_n$. However, $W_n R_n \,(= R_n^{(S)})$ already includes a component that is a martingale transform of $\langle R \rangle_n^{(M)}$. By considering the basis components, we can limit R'_n up to R_n^S.

3.2 Sequential test

Therefore, the choice of $R_n^{(S)} = W_n R_n$ in (2.3) is proposed by these probabilistic reasons. For the statistical application we define W_n by (3.2) (i.e. $\eta_{n-1}^{(S)} \equiv 1$ in the respective definition of W_n in Proposition 3.1) and consider a lower bound for the average run length to false alarm of procedure (2.3).

Proposition 3.3. *Assume $C > 0$, then $E_{-\infty} N(C) \geq C$.*

Proof. By definition (2.3), we obtain $W_{N(C)} R_{N(C)} \geq C$. From the H_0-martingale structure of $W_n R_n - n$ and the optimal sampling theorem, the proof of Proposition 3.3 follows, since

$$0 = E_{-\infty} \left(W_{N(C)} R_{N(C)} - N(C) \right) \geq C - E_{-\infty} N(C).$$

When we monitor the significant level of the sequential procedure (2.3), let $R_n^{(S)} = W_n R_n$, where W_n is defined by (3.4) with $\eta_{n-1}^{(S)} \equiv 0\,(n > 1)$ and $W_0 = R_0 = 1$. Since, in this case, $R_n^{(S)}$ is a H_0-martingale ($E_{-\infty} R_1^{(S)} = 1$), we obtain

Proposition 3.4.

$$P_{-\infty}\left\{R_n^{(S)} \geq C \text{ for some } n \geq 1\right\} \leq 1/C.$$

Remark 1. When $\nu_i = -i$ in (1.2) and $\eta_{n-1}^{(S)} \equiv 1\,(n > 1)$ in (3.4), the results in Propositions 3.1 and 3.2 have been extended to general submartingale cases by Vexler, Liu and Pollak (2006). In particular, the authors showed that the classical SR statistic $R_n^{(S)} = \sum_{k=1}^{n}\prod_{i=k}^{n}\frac{f_y(z_i)}{f_x(z_i)}$ and the simple CUSUM statistic $\Lambda_n = \max_{1\leq k\leq n}\prod_{i=k}^{n}\frac{f_y(z_i)}{f_x(z_i)}$ have a common H_0-martingale basis, and hence the procedures based on SR statistics and the schemes founded on CUSUM statistics have almost equivalent optimal statistical properties. Moreover, the classical SR statistic $R_n^{(S)}$ is the adapted CUSUM statistic (i.e. $R_n^{(S)}$ is obtained by (3.2) with $R_n = \Lambda_n$).

Remark 2. Assume that a density function f_y depends on the unknown parameter θ_y and $f_{z\,i}(u; d) = f_{z\,i}(u; d; \theta_y)$. In this case, the SR rule is in principle easy to modify (e.g. Krieger, Pollak and Yakir, 2003; Lorden and Pollak, 2005): just define a mixing measure $d\Theta_y(\theta)$ and

$$\begin{aligned}
R_n &= \int\left(\sum_{k=2}^{n}\prod_{i=1}^{n}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta)}{f_x(z_i)} + \prod_{i=1}^{n}\frac{f_y(z_i;\theta)}{f_x(z_i)}\right)d\Theta_y(\theta),\\
W_n &= \left(R_{n-1} + \int\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1};\theta)}{f_x(z_i)}d\Theta_y(\theta)\right)^{-1}\left(W_{n-1}R_{n-1}+1\right),\\
R_n^{(S)} &= W_nR_n, \qquad W_0 = R_0 = 0,
\end{aligned}$$

where Θ_y is a prior probability measure and we pretend that $\theta_y \sim \Theta_y$. At this rate, for example, Proposition 3.3 is valid. Now, let $f_x(u) = f_x(u; \theta_x)$ be known up to parameter $\theta_x \sim \Theta_x$ and $f_{z\,i}(u; d) = f_{z\,i}(u; d; \theta_y, \theta_x)$. Define

$$\begin{aligned}
R_n &= \int\left(\sum_{k=2}^{n}\prod_{i=1}^{n}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x^{1,n})} + \prod_{i=1}^{n}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x^{1,n})}\right)d\Theta_y(\theta_1)d\Theta_x(\theta_2),\\
W_n &= \left\{\int\left(\sum_{k=2}^{n-1}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x^{1,n})} + \prod_{i=1}^{n-1}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x^{1,n})}\right)d\Theta_y(\theta_1)d\Theta_x(\theta_2)\right.\\
&\quad\left. + \int\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x^{1,n})}d\Theta_y(\theta_1)d\Theta_x(\theta_2)\right\}^{-1}\frac{f_x(z_n;\theta_x^{1,n})}{\max_\theta f_x(z_n;\theta)}
\end{aligned}$$

$$\times\left(W_{n-1}R_{n-1}+1\right),$$
$$R_n^{(S)} = W_nR_n, \qquad W_0=R_0=0,$$

where $\theta_x^{1,n}$ is an estimator of θ_x based on $z_1,\dots,z_n$ (e.g. $\theta_x^{1,n}\in\Im_n$ is the maximum likelihood estimator). Since

$$\begin{aligned}
W_nR_n &= \left\{\int\left(\sum_{k=2}^{n-1}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}+\prod_{i=1}^{n-1}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x)}\right)d\Theta_y(\theta_1)\right.\\
&\quad\left. d\Theta_x(\theta_2)+\int\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}d\Theta_y(\theta_1)d\Theta_x(\theta_2)\right\}^{-1}\\
&\quad\times\frac{f_x(z_n;\theta_x)}{\max_\theta f_x(z_n;\theta)}\left(W_{n-1}R_{n-1}+1\right)\\
&\quad\times\int\left(\sum_{k=2}^{n}\prod_{i=1}^{n}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}+\prod_{i=1}^{n}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x)}\right)d\Theta_y(\theta_1)d\Theta_x(\theta_2)\\
&\le\left\{\int\left(\sum_{k=2}^{n-1}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}+\prod_{i=1}^{n-1}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x)}\right)d\Theta_y(\theta_1)d\Theta_x(\theta_2)\right.\\
&\quad\left.+\int\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}d\Theta_y(\theta_1)d\Theta_x(\theta_2)\right\}^{-1}\left(W_{n-1}R_{n-1}+1\right)\\
&\quad\times\int\left(\sum_{k=2}^{n}\prod_{i=1}^{n}\frac{f_{z\,i}(z_i;\gamma_{k-1};\theta_1,\theta_2)}{f_x(z_i;\theta_x)}+\prod_{i=1}^{n}\frac{f_y(z_i;\theta_1)}{f_x(z_i;\theta_x)}\right)d\Theta_y(\theta_1)d\Theta_x(\theta_2),
\end{aligned}$$

$E_{-\infty}R_n^{(S)}\Big|\Im_{n-1}\le R_{n-1}^{(S)}+1$, and hence $\left\{R_n^{(S)}-n,\Im_n\right\}$ is a H_0-super - martingale. Therefore, we have $E_{-\infty}N(C)\ge C$.

Remark 3. Because the H_0-martingale component of the primary test statistic is an important term of the transformation from R_n to $R_n^{(S)}$, by directly applying Lemma 3.1, we can write

$$R_n^{(S)} = \langle R\rangle_n^{(M)}+n=R_n-\sum_k^n\prod_{i=1}^{k-1}\frac{f_{z\,i}(z_i;\gamma_{k-1})}{f_x(z_i)}+n. \tag{3.5}$$

However, in the general context of the stated problem (1.2), this definition does not provide nonnegative H_0-submartingale structure of $R_n^{(S)}$ and use of the optimal sampling theorem is problematic. We will consider the procedure based on (3.5) in Section 4.

3.3 Retrospective Test

Vexler (2006) proposes a modification of SR statistics and applies to retrospective change point detection. In the context of a classical change point problem (the definition of this problem is presented by Yao, 1993) tests based on the modified SR statistics have asymptotic ($(n, C) \to \infty$) significance levels, which are equal to significance levels of the respective CUSUM tests. (For the references related to CUSUM procedures, see Page, 1954; Sen and Srivastava, 1975; Pettitt, 1980; Gombay and Horvath, 1994; Dragalin, 1996; Gurevich and Vexler, 2005.) Here we apply the same approach to the problem of (2.4) with $R_n^{(R)} = \max_{1 \le m \le n} R_m$.

Proposition 3.5. *The significance level* α *of the test satisfies:*

$$\begin{aligned} \alpha \quad &\equiv \quad P_{-\infty}\left\{\max_{1\le m\le n} R_m > C\right\} \le \min\left(\frac{\sum_{m=1}^{n} e_{m-1}}{C},\right. \\ &\left. P_{-\infty}\left\{R_1 > \frac{C}{2}\right\} + \frac{2\sum_{m=1}^{n} E_{-\infty} P_{\gamma'_{m-1}}\left\{R_n > \frac{C}{2}\right\}}{C}\right), \end{aligned}$$

where $e_0 = 1$, $e_m = E_{-\infty} f_{z\,m}(z_m; \gamma_m)/f_x(z_m)$ *and* $\gamma'_0 = -\infty$, $\gamma'_m = \gamma_m$.

Proof. In Appendix.

We therefore have the non-asymptotic upper bound for the significance level of the test (2.4): e.g. $\alpha \le \sum_{m=1}^{n} e_{m-1}/C$. Thus, selecting $C = \sum_{m=1}^{n} e_{m-1}/\alpha$ determines a test with a level of significance that does not exceed α (that ensures a p-value). Note that use and interpretation of the upper bounds similar to the one obtained above is a widely known topic in statistics (e.g. Robbins, 1970; Krieger, Pollak and Yakir, 2003; Gurevich and Vexler, 2005). In simple cases of (1.1), (1.2), one can show $C\alpha/n \to \alpha_L$, where the constant $\alpha_L \in (0, 1)$ (Vexler, 2006).

Remark 4. Since for all $i \ge 1$ and finite d

$$E_d \ln \frac{f_{z\,i}(z_i; d)}{f_x(z_i)} > -\ln E_d \frac{f_x(z_i)}{f_{z\,i}(z_i; d)} = 0, \tag{3.6}$$

for small enough $\eta > 0$, there exists a constant $h > 0$, such that for every $t \in [d-h, d+h]$, $E_d \ln \frac{f_{z\,i}(z_i;\, t)}{f_x(z_i)} > \eta$. Let N be large enough, such that for every $t \in [d-h, d+h]$ and $n \ge N$,

$$\frac{1}{n}\sum_{i=1}^{n} \ln \frac{f_{z\,i}(z_i; t)}{f_x(z_i)} > \eta/2 \ a.s.\ .$$

Then for $n \geq N$,

$$R_n \quad \geq \sum_{\gamma_{k-1} \in [d-h, d+h]} \prod_{i=1}^{n} \frac{f_{z\,i}(z_i; \gamma_{k-1})}{f_x(z_i)} \geq \sum_{\gamma_{k-1} \in [d-h, d+h]} \exp\{n\eta/2\} \quad a.s.,$$

assuming the probability of each event under the summation sign is positive. Therefore, test (2.4) is asymptotically power one, if, e.g., $\ln(C) = o(n)$.

Remark 5. Similar to Remark 2, the retrospective test for (1.2), where f_x and $f_{z\,i}$ are known up to some parameters, can be considered.

4 Implementation and Simulation

In order to demonstrate the efficiency of the procedure (2.3) with $R_n^{(S)}$ defined by (3.2) and the closeness of the lower bound obtained by Proposition 3.3, in this section we present results of several Monte Carlo experiments related to the sequential aspect of the stated problem (1.1), (1.2), where $\nu_i = x_i$, $x_i \sim N(2, 3)$ and $y_i \sim N(0, 1)$.

From the method described in Sections 2 and 3, we choose two procedures $N^{(1)}(C)$ and $N^{(2)}(C)$, which utilize (2.3) with $R_n^{(S)}$ by (3.2) and $R_n^{(S)}$ by (3.5), respectively. We ran 15000 repetitions of the model (where $d = -\infty, 0.5, 0.7, 0.9, 1.1$) for each procedure and at each point $C = 70$, 100, 130, 160, ... Figure 1 depicts the Monte Carlo averages of $N^{(1)}(C)$ and $N^{(2)}(C)$ in these cases, where every simulated observation is from $N(2, 3)$ (i.e. under H_0: $d = -\infty$). It appears that, as usual for change point detections, $E_{-\infty}N^{(1)}(C)/C$ and $E_{-\infty}N^{(2)}(C)/C$ have some limit value, as $C \to \infty$. In the considered case the curve of $E_{-\infty}N^{(1)}(C)$ is close to the lower bound C.

Now consider the situation where the sample is nonhomogeneous. For $d = 0.5, 0.7, 0.9, 1.1$, Figure 2 plots the Monte Carlo averages of $N^{(1)}(C)$ and $N^{(2)}(C)$. The figure shows that both $E_dN^{(1)}(C)$ and $E_dN^{(2)}(C)$ are similar to linear functions of $\ln(C)$. This is also a common property of change point detection policies. However, with

$E_{-\infty}N^{(1)}(C_1) \simeq E_{-\infty}N^{(2)}(C_2)$, $N^{(1)}$ tends to be a quicker detection scheme.

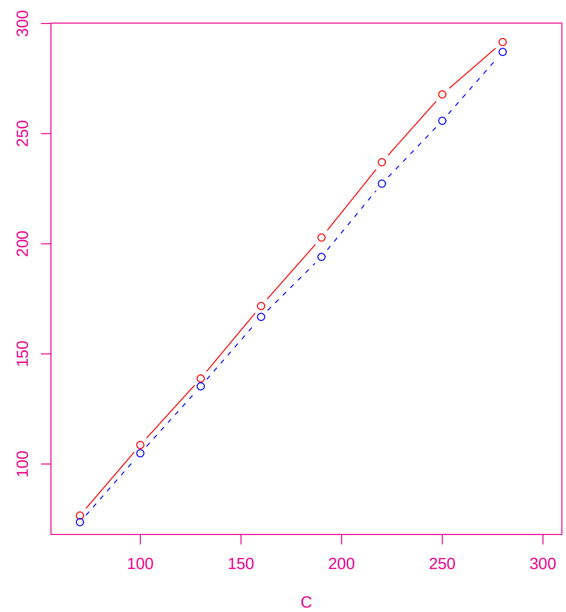

Figure 1: *Comparison between the Monte Carlo estimator of $E_{-\infty}N^{(1)}(C)$ (—○—) and the Monte Carlo estimator of $E_{-\infty}N^{(2)}(C)$ (- -○- -).*

Appendix

Proof of Proposition 3.1. By applying the definition of W_n and (3.1), we obtain

$$E_{-\infty}R_n^{(S)}\Big|\Im_{n-1} = W_{n-1}R_{n-1} + \eta_{n-1}^{(S)} \geq R_{n-1}^{(S)},$$

hence $R_n^{(S)}$ is the nonnegative submartingale with respect to $\Im_n$. According to Lemma 3.1, we have

$$\langle R^{(S)}\rangle_n^{(M)} \quad = \quad R_n^{(S)} - \langle R^{(S)}\rangle_n^{(P)} = R_n^{(S)} - \sum_k^n \eta_{k-1}^{(S)}.$$

Therefore,

$$\langle R^{(S)}\rangle_n^{(M)} - \langle R^{(S)}\rangle_{n-1}^{(M)} = R_n^{(S)} - R_{n-1}^{(S)} - \eta_{n-1}^{(S)} = W_nR_n - W_{n-1}R_{n-1} - \eta_{n-1}^{(S)},$$

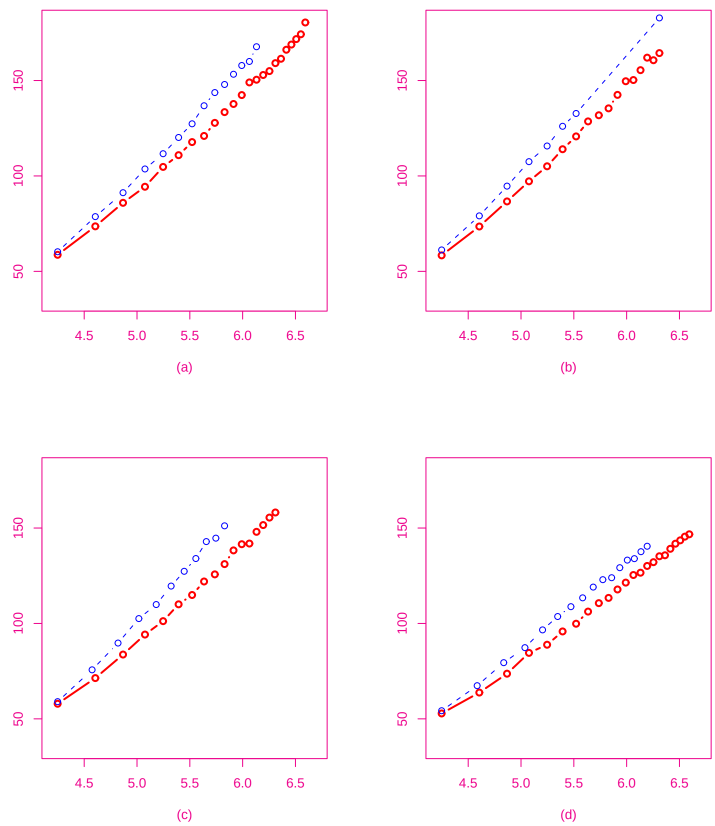

Figure 2: The Monte Carlo estimator of $E_d N^{(1)}(C)$ (—○—) and the Monte Carlo estimator of $E_d N^{(2)}(C)$ (- -○- -) are plotted against $\ln(C)$ (the axis of abscissae), where $d = 0.5, 0.7, 0.9, 1.1$ regard graphs (a)-(d), respectively.

where $W_{n-1}R_{n-1} = W_n R_{n-1} + W_n \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)} - \eta_{n-1}^{(S)}$ by the definition of W_n. Now we have

$$\langle R^{(S)} \rangle_n^{(M)} - \langle R^{(S)} \rangle_{n-1}^{(M)} = W_n \left(R_n - R_{n-1} - \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)} \right).$$

On the other hand, by Lemma 3.1 and (3.1) one can show that

$$\begin{aligned}
\langle R\rangle_n^{(M)} - \langle R\rangle_{n-1}^{(M)} &= R_n - \sum_k^n \prod_{i=1}^{k-1} \frac{f_{z\,i}(z_i;\gamma_{k-1})}{f_x(z_i)} \\
&\quad - R_{n-1} + \sum_k^{n-1} \prod_{i=1}^{k-1} \frac{f_{z\,i}(z_i;\gamma_{k-1})}{f_x(z_i)} \\
&= R_n - R_{n-1} - \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)}.
\end{aligned}$$

Since $W_n \in \Im_{n-1}$, by definition (3.3) the proof of Proposition 3.1 is now complete. ■

Proof of Proposition 3.2. Clearly, we have $R'_n = W_n R_n + \varsigma_n$, where $\varsigma_n = R'_n - W_n R_n \in \Im_n$. Consider the conditional expectation

$$\begin{aligned}
E_{-\infty} R'_n \Big| \Im_{n-1} &= E_{-\infty} W_n R_n \Big| \Im_{n-1} + E_{-\infty} \varsigma_n \Big| \Im_{n-1} \\
&= W_{n-1} R_{n-1} + \eta_{n-1}^{(S)} + E_{-\infty} \varsigma_n \Big| \Im_{n-1} \\
&= R'_{n-1} - \varsigma_{n-1} + \eta_{n-1}^{(S)} + E_{-\infty} \varsigma_n \Big| \Im_{n-1}.
\end{aligned}$$

On the other hand, R'_n is a submartingale, and therefore $E_{-\infty} R'_n \Big| \Im_{n-1} = R'_{n-1} + \eta_{n-1}^{(S)}$. With this we conclude that $E_{-\infty} \varsigma_n \Big| \Im_{n-1} = \varsigma_{n-1}$.

From the basic definitions and Lemma 3.1 (in a similar manner to the proof of Proposition 3.1) it follows that

$$\begin{aligned}
\langle R'\rangle_n^{(M)} - \langle R'\rangle_{n-1}^{(M)} &= \varsigma_n - \varsigma_{n-1} + W_n \left(R_n - R_{n-1} - \prod_{i=1}^{n-1} \frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)} \right) \\
&= \varsigma_n - \varsigma_{n-1} + W_n \left(\langle R\rangle_n^{(M)} - \langle R\rangle_{n-1}^{(M)} \right).
\end{aligned}$$

Hence

$$\varsigma_n - \varsigma_{n-1} = (a_n - W_n) \left(\langle R\rangle_n^{(M)} - \langle R\rangle_{n-1}^{(M)} \right), \ (a_n - W_n) \in \Im_{n-1}.$$

This completes the proof of Proposition 3.2. ■

Proof of Proposition 3.5. By (3.1), R_n is the positive H_0-submartingale. Therefore, $\alpha \leq E_{-\infty}R_n/C$. By virtue of the recursion

$$
\begin{aligned}
E_{-\infty}R_n &= E_{-\infty}E_{-\infty}R_n\Big|\Im_{n-1} = E_{-\infty}R_{n-1} + E_{-\infty}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)} \\
&= E_{-\infty}R_{n-1} + E_{-\infty}E_{-\infty}\prod_{i=1}^{n-1}\frac{f_{z\,i}(z_i;\gamma_{n-1})}{f_x(z_i)}\Big|\gamma_{n-1} \\
&= E_{-\infty}R_{n-1} + e_{n-1},
\end{aligned}
$$

we have $\alpha \leq \sum_{m=1}^{n} e_{m-1}/C$. Moreover, using a maximal inequality for nonnegative submartingales (Bhattacharya, 2005) leads to

$$
\alpha \leq P_{-\infty}\left\{R_1 > \frac{C}{2}\right\} + \frac{2E_{-\infty}R_nI\left\{R_n > \frac{C}{2}\right\}}{C}
$$

$$
= P_{-\infty}\left\{R_1 > \frac{C}{2}\right\} + \frac{2\sum_{m=1}^{n}E_{-\infty}\Lambda_n(\gamma'_{m-1})I\left\{R_n > \frac{C}{2}\right\}}{C}
$$

$$
= P_{-\infty}\left\{R_1 > \frac{C}{2}\right\} + \frac{2}{C}\sum_{m=1}^{n}E_{-\infty}E_{-\infty}\Lambda_n(\gamma'_{m-1})I\left\{R_n > \frac{C}{2}\right\}\Big|\gamma'_{m-1}
$$

$$
= P_{-\infty}\left\{R_1 > \frac{C}{2}\right\}
$$

$$
+\frac{2}{C}\sum_{m=1}^{n}E_{-\infty}\int\frac{f(z_1,\ldots,z_n|\text{ under } H_1, \text{ where } d=\gamma'_{m-1})}{f_x(z_1,\ldots,z_n)}f_x(z_1,\ldots,z_n)
$$

$$
\times I\left\{R_n > \frac{C}{2}\right\} = P_{-\infty}\left\{R_1 > \frac{C}{2}\right\} + \frac{2}{C}\sum_{m=1}^{n}E_{-\infty}E_{\gamma'_{m-1}}I\left\{R_n > \frac{C}{2}\right\}.
$$

This completes the proof of Proposition 3.5. ■

References

Bhattacharya, P. K. (2005). A maximal inequality for nonnegative submartingales. *Statistics & Probability Letters* **72**, 11-12.

Brostrom, G. (1997). A martingale approach to the changepoint problem. *J. Amer. Statist. Assoc.* **92**, 1177-1183.

Cooper, G. S., Savitz D. A., Millikan, R. and Chiu K.T. (2002). Organochlorine exposure and age at natural menopause. *Epidemiology* **13**, 729-733.

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.

Finkelstein, M. M. and Verma, D. K. (2001). Exposure estimation in the presence of nondetectable values: another look. *American Industrial Hygiene Association Journal* **62**, 195-198.

Garel, B. (2005). Asymptotic theory of the likelihood ratio test for the identification of a mixture. *J. Statist. Plann. Inference* **131**, 271-296.

Gombay, E. and Horvath, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Process. Appl.* **50**, 161-171.

Gordon, L. and Pollak, M. (1995). A robust surveillance scheme for stochastically ordered alternatives. *Ann. Statist.* **23**, 1350-1375.

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

Helsel D. (2005). *Nondetects and data analysis: Statistics for censored environmental data.* Hoboken, New Jersey: John Wiley & Sons, Inc.

Hornung, R. W. and Reed, L. (1990). Estimation of Average Concentration in the Presence of Nondetectable Values. *Appl. Occup. Environ. Hyg.* **5**, 46-51.

Keith, L. H., Crummett, W., Deegan, J., Libby, R. A., Taylor, J. K. and Wentler, G. (1983). Principles of environmental analysis. *Anal. Chem.* **55**, 2210-2218.

Krieger, A. M., Pollak, M. and 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. R. Statist. Soc. B.* 57 **4**, 613-658.

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

Lorden, G. and Pollak, M. (2005). Nonanticipating estimation applied to sequential analysis and changepoint Detection. *Ann. Statist.* **33**, 1422-1454.

Page, E. S. (1954). Continuous inspectation schemes. *Biometrika* **41**, 100-114.

Lubin, J. H., Colt, J. S., Camann D., Davis, S., Cerhan, J. R., Severson, R. K., Bernstein, L. and Hartge P. (2004). Epidemiologic evaluation of measurement data in the presence of detection limits. *Environ Health Perspect.* **112**, 1691-1696.

Pettitt, A. N. (1980). A simple cumulative sum type statistic for the change-point with zero-one observations. *Biometrika* **67**, 79-84.

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.

Ritov, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18**, 1464-1469.

Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41**, 1397-1409.

Robbins, H. and Siegmund, D. (1973). *A class of stopping rules for testing parametric hypotheses.* Proc. Sixth Berkeley Symp., Math. Statist. Prob. 4, Univ. of Calif. Press.

Schisterman, E. F., Vexler, A., Whitcomb, B. W. and Liu, A. (2006). The limitations due to exposure detection limits for regression models. *Amer. J. Epidemiol.,* **163**, 374-383.

Sen, A. and 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., Liu, A. and Schisterman, E. (2006). Efficient design and analysis of biospecimens with measurements subject to detection limit. *Biometrical Journal* **48**, 780-791.

Vexler, A., Liu, A. and Pollak, M. (2006). Transformation of change-point detection methods into a Shiryayev-Roberts form. Technical report, Department of Biostatistics, The New York State University at Buffalo.

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

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. (1998). On the average run length to false alarm in surveillance problems which possess an invariance structure. *Ann. Statist.* **26**, 1198-1214.