Communications in Statistics—Simulation and Computation®, 39: 1–22, 2010

ISSN: 0361-0918 print/1532-4141 online
DOI: 10.1080/03610911003663881

Taylor & Francis
Taylor & Francis Group

Retrospective Change Point Detection: From Parametric to Distribution Free Policies

GREGORY GUREVICH[1] AND ALBERT VEXLER[2]

[1]Department of Industrial Engineering and Management, Shamoon College of Engineering, Beer-Sheva, Israel
[2]Department of Biostatistics, The State University of New York at Buffalo, Buffalo, New York

The literature displays change point detection problems in the context of one of the key issues that belong to testing statistical hypotheses. The main focus in this article is to review recent retrospective change point policies and propose new relevant procedures. Commonly applied practical quality control purposes have declared statements of the change point problems. Various biostatistical and engineering applications cause consideration of an extended form of the change point problem. In this article, we consider parametric and distribution free generalized change point detection policies, attending to different contexts of optimality and robustness of the procedures. We conducted a broad Monte Carlo study to compare various parametric and nonparametric tests, also investigating a sensitivity of the change point detection policies with respect to assumptions required for correct executions of the procedures. An example based on real biomarker measurements is provided to judge our conclusions.

Keywords Change point; CUSUM; Entropy; Likelihood ratio; Most powerful; Nonparametric likelihood; Nonparametric tests; Optimal testing; Robustness; Shiryayev–Roberts.

Mathematics Subject Classification ■. AQ1

1. Introduction

In this article, we aim to introduce and examine different tests for a change in the distribution of independent observations $X_1, X_2, \ldots, X_n$ with the fixed sample size n. In the formal context of hypotheses testing, we state the problem to test for

$$\begin{aligned} &H_0, \text{the null: } X_1, X_2, \ldots, X_n \sim F_0 \quad \text{versus} \\ &H_1, \text{the alternative: } X_i \sim F_1, X_j \sim F_2, \quad i = 1, \ldots, v-1, \;\; j = v, \ldots, n, \end{aligned} \tag{1}$$

Received April 7, 2009; Accepted January 27, 2010
Address correspondence to Albert Vexler, Department of Biostatistics, The State University of New York at Buffalo, Farber Hall, Room 246, Buffalo, NY 14214, USA; E-mail: avexler@buffalo.edu

where F_0, F_1, F_2 are distribution functions that correspond to density functions f_0, f_1, f_2. The unknown parameter ν, $1 \le \nu \le n$ is called a *change point*. Following certain applied aspects of quality control studies, the literature assumes commonly the function F_1 is equal to F_0 (e.g., Gombay and Horvath, 1994; James et al., 1987). We state a general case when F_1 can be different from the null distribution F_0. Various biostatistical and engineering studies can motivate to eliminate the constraint $F_1 = F_0$ (e.g., Vexler et al., 2009b).

In accordance with the statistical literature, we can investigate the problem (1) in parametric or nonparametric forms, depending on assumptions made on the distribution functions F_0, F_1, and F_2. In the parametric case of (1), we assume the distribution functions F_0, F_1, and F_2 have known forms that can contain certain unknown parameters (e.g., Gombay and Horvath, 1994; Gurevich, 2007; James et al., 1987; Vexler, 2006; Vexler and Gurevich, 2009a).

In the nonparametric case of (1), the functions F_0, F_1, F_2 are assumed to be completely unknown (e.g., Ferger, 1994; Gombay, 2000, 2001; Gurevich, 2006; Wolfe and Schechtman, 1984; Zou et al., 2007).

In this article, we review, develop, and compare different policies for the problem (1), in both the parametric and nonparametric cases, attending to different contexts of optimality of tests. In Secs. 2 and 3 we present the parametric and nonparametric methods, correspondingly. Section 4 displays a Monte Carlo study to compare the powers of parametric and nonparametric change point tests, analyzing sensitivity (robustness) of the change point policies with respect to assumptions that are required for correct executions of the procedures. Section 5 provides an example based on real biomarker measurements that judges reviewed change point detection policies in practice. We state our conclusion in Sec. 6.

2. Parametric Methods

The parametric case of testing the change point problem (1) has been dealt with extensively in both the theoretical and applied literature (e.g., Chernoff and Zacks, 1964; Csorgo and Horvath, 1997; Gombay and Horvath, 1994; Gurevich, 2007; Gurevich and Vexler, 2005; James et al., 1987; Kander and Zacks, 1966; Sen and Srivastava, 1975). Chernoff and Zacks (1964) considered the problem (1) based on normally distributed observations with $F_0 = F_1 = N(\theta_0, 1)$, $F_2 = N(\theta, 1)$, where θ_0 and $\theta > \theta_0$ are unknown. In this case, a noninformative uniform prior for ν was assumed and a Bayesian test statistic was proposed. Kander and Zacks (1966) extended Chernoff and Zacks's (1964) results to a case based on data from the one-parameter exponential family. Sen and Srivastava (1975) used the maximum likelihood technique to present a test statistic. James et al. (1987) proposed, in the context of (1), decision rules based on likelihood ratios and recursive residuals. This change point literature concluded that there is not a globally (with respect to values of ν, under H_1) preferable test for (1). It turned out that Chernoff and Zacks's (1964) test has a larger power than that of tests based on the likelihood ratio or recursive residuals when ν is around $n/2$, but this property is reversed if the change point ν is close to the edges; i.e., when $\nu \approx n$ or $\nu \approx 1$.

Because the change point ν is unknown, the maximum likelihood estimation of ν and unknown parameters of F_0, F_1, F_2 can be applied, and then the likelihood ratio tests can be defined to test for (1). The resulting tests have a CUSUM type structure that is well addressed in the literature (e.g., Gombay and Horvath, 1994;

Gurevich, 2007). Gombay and Horvath (1994), Lai (1995), Gurevich and Vexler (2005, 2006), and Gurevich (2007) proved that the CUSUM approach implies very powerful parametric change point policies.

When, in accordance with the statement (1), observations can have the density functions f_0, f_1, or f_2 based on likelihood ratio, which are assumed to be completely known, the CUSUM statistic has the form of

$$\Delta_n = \max_{1 \le k \le n} \Lambda_k^n, \tag{2}$$

where the likelihood ratios

$$\Lambda_k^n = \frac{\prod_{i=1}^{k-1} f_1(X_i) \prod_{i=k}^{n} f_2(X_i)}{\prod_{i=1}^{n} f_0(X_i)}, \prod_{i=1}^{0} f_1(X_i) = 1.$$

(Here, as mentioned above, the maximum likelihood estimator $\hat{\nu} = \arg\max_{1 \le k \le n} \Lambda_k^n$ of the unknown parameter ν was used, modifying the most powerful test statistic Λ_ν^n to have the maximum likelihood ratio form Δ_n; for details, see, e.g., Vexler and Gurevich, 2009a. Under certain assumptions, one can show the estimator $\hat{\nu}$ is consistent; e.g., Borovkov, 1999; Gurevich and Vexler, 2005; Pollak and Tartakovsky, 2009; Tartakovsky et al., 2009.)

The null hypothesis H_0 of (1) is proposed to be rejected for large values of the CUSUM test statistic. It is clear that, when the density functions f_0, f_1, and f_2 have forms with unknown parameters, one needs to estimate the unknown parameters and then an approximated CUSUM type test statistic can be defined. For example, Gombay and Horvath (1994) considered the following situation; that is,

$$f_0(x) = f(x;\theta_0), \quad f_1(x) = f(x;\theta_1), \quad f_2(x) = f(x;\theta_2), \tag{3}$$

where the vector parameters $\theta_i \in \Theta \subseteq R^d$, $i = 0, 1, 2$ are unknown, $\theta_1 \ne \theta_2$. Gombay and Horvath suggested rejecting the hypothesis H_0 of (1), if, for a fixed test threshold AQ2 $C_1 > 0$,

$$\max_{1 < k \le n} \Lambda_k^{*^n} > C_1, \tag{4}$$

where

$$\Lambda_k^{*^n} = \frac{\sup_{\theta_1 \in \Theta} \prod_{i=1}^{k-1} f(X_i;\theta_1) \sup_{\theta_2 \in \Theta} \prod_{i=k}^{n} f(X_i;\theta_2)}{\sup_{\theta_0 \in \Theta} \prod_{i=1}^{n} f(X_i;\theta_0)}.$$

To control the type I error of CUSUM-type tests, evaluation of the null distribution of the corresponding CUSUM test statistics is commonly required. Therefore, investigators need to use a simulation study or complex asymptotic ($n \to \infty$) propositions to approximate the type I error of CUSUM type tests. Gombay and Horvath (1994, 1996) obtained the asymptotic null distribution of the statistic (4) and the rate of convergence of this approximation. Note that there are no results that show a nonasymptotic optimality of CUSUM-type tests in a general setting of (1).

Alternatively, and in contrast to the CUSUM method, Vexler (2006, 2008) as well as Vexler and Gurevich (2009a) proposed to use retrospectively the Shiryayev–Roberts (SR) approach that is well accepted for developing optimal sequential change point detection procedures (e.g., Krieger et al., 2003; Pollak, 1985; Pollak and Tartakovsky, 2009; Tartakovsky et al., 2009). In the context of this method, the authors suggested avoiding estimating the change point v location, when tests for (1) are considered. When f_0, f_1, and f_2 of (1) are completely known, the SR test statistic has the form of

$$R_n = \sum_{k=1}^{n} \Lambda_k^n, \tag{5}$$

where Λ_k^n is stated in (2). The hypothesis H_0 of (1) is proposed to be rejected if

$$R_n > C_2, \tag{6}$$

for a fixed test threshold $C_2 > 0$.

The change point literature concludes, generally speaking, there are no uniformly most powerful tests for (1) (e.g., James et al., 1987). Vexler and Gurevich (2009a) proved the following proposition, pointing out an optimal nonasymptotic property of the decision rule (6). To formulate this proposition, we define probability measures $P_{H_0}(A) = \Pr(A \mid H_0)$ and $P_{v=k}(A) = \Pr(A \mid H_1, v = k)$, where A is a random event, $k = 1, \ldots, n$.

Proposition 1. *The policy* (6) *is average most powerful*; *i.e., for any decision rule* δ *based on* $\{X_i, i = 1, \ldots, n\}$ *we have*

$$\frac{1}{n}\sum_{k=1}^{n} P_{v=k}(R_n > C) \geq \frac{1}{n}\sum_{k=1}^{n} P_{v=k}(\delta \text{ rejects } H_0), \tag{7}$$

when the significance level of tests is fixed to be $\alpha = P_{H_0}(\delta \text{ rejects } H_0)$.

Remark. While assuming conditions (3), we propose to modify the SR test statistic (5) to have the appropriate form

$$\sum_{k=2}^{n} \Lambda_k^{*^n}, \tag{8}$$

where the ratios $\Lambda_k^{*^n}$ are denoted in (4). (The stated problem (1) with $f_0 \neq f_1$ has not been well addressed in the parametric change point literature. The retrospective test based on (8) is not examined well in the literature, in general cases, even when $f_0 = f_1$.)

3. Nonparametric Methods

When the problem (1) is stated nonparametrically, there is no universal powerful methodology (e.g., as the likelihood methods mentioned in Sec. 2) for this subject. In this case, common components of nonparametric change point detection policies have been proposed to be based on signs and/or ranks and/or U statistics (e.g., Csorgo and Horvath, 1997; Ferger, 1994; Gombay, 2000, 2001; Gurevich, 2006; Wolfe and Schechtman, 1984). Sen and Srivastava (1975) focused on the problem (1) with the unknown distributions $F_0(x) = F_1(x)$, $F_2(x) = F_1(x - \beta)$, $\beta > 0$. The authors suggested to reject H_0, for large values of the statistic

$$D_1 = \max_{2 \le k \le n} \left\{ \left[U_{k-1,n-k+1} - \big((k-1)(n-k+1)\big)/2 \right] \Big/ \left[(k-1)(n-k+1)(n+1)/12 \right]^{\frac{1}{2}} \right\}, \tag{9}$$

where $U_{k-1,n-k+1}$ is the Mann–Whitney statistic for two samples of size $k-1$ and $n-k+1$. Setting the problem (1) in a similar manner to Sen and Srivastava (1975), Pettitt (1979) used the statistic

$$K = \max_{2 \le k \le n} \left\{ -\sum_{i=1}^{k-1} \sum_{j=k}^{n} Q_{ij} \right\}, \quad Q_{ij} = \operatorname{sign}(X_i - X_j) = \begin{cases} 1 & X_i > X_j \\ 0 & X_i = X_j \\ -1 & X_i < X_j \end{cases} \tag{10}$$

to propose a change point detection policy. Wolfe and Schechtman (1984) showed that the statistic K can be presented as $2\max_{2 \le k \le n} \left\{ U_{k-1,n-k+1} - (k-1)(n-k+1)/2 \right\}$. Then, the statistics K and D_1 have a similar structure. In this case, Csorgo and Horvath (1988) modified the statistics D_1 and K to have the form of

$$D_2 = \sqrt{3} \max_{2 \le k \le n} \frac{U_k}{[(k-1)(n-k+2)n]^{\frac{1}{2}}}, \tag{11}$$

where $U_k = -\sum_{1 \le i \le k-1} \sum_{k \le j \le n} \operatorname{sign}(X_i - X_j)$. This modification was introduced to evaluate asymptotically ($n \to \infty$) the type I error of the corresponding to the statistic (11) test that requires rejecting H_0, if $D_2 > C_3$, where C_3 is a test threshold. When the two-sided statement $F_0(x) = F_1(x)$, $F_2(x) = F_1(x - \beta)$, $\beta \neq 0$ is assumed, the absolute values of the statistics (9)–(11) should be considered to construct the tests for the two-sided alternative. Gurevich (2006) analyzed the problem (1), when $F_0 = F_1$ is unknown and the post-change distribution function F_2 is stochastically larger than the pre-change distribution function F_1. In contrast to the methods above, in this case, a test statistic is suggested to be based on the likelihood ratio of the ranks of observations, assuming that flat prior information regarding the pre- and post-change distribution functions F_1 and F_2 is available. Especially, the technique of Gurevich (2006) considers the rank-based likelihood ratios

$$\Lambda_k^n(\rho) = \frac{f_{H_1:\nu=k}[\rho(1,n), \rho(2,n), \ldots, \rho(n,n)]}{f_{H_0}[\rho(1,n), \rho(2,n), \ldots, \rho(n,n)]}, \quad \rho(j,n) = \sum_{k=1}^{n} I_{\{X_k \le X_j\}}$$

(here, $f(\cdot)$ denotes a joint density, $I_{\{A\}}$ is the indicator function of an event A) to be main components of the CUSUM-type test statistic. To simplify these likelihood ratios, presenting analytical forms, Gurevich (2006) invited a method by Gordon and Pollak (1995) that was proposed to create a robust sequential surveillance scheme for stochastically ordered alternatives. This approach proposes pretending that observations follow a distribution from an exponential family, say, Ψ, under the null hypothesis of (1), whereas, under the alternative H_1, the observations are distributed corresponding to a mixture of distribution functions that belong also to Ψ and depend on a set of parameters. By virtue of the probabilistic characteristics of the rank statistics and the maximum likelihood methodology, the parameters can be reasonably derived to maximize $E_{X\sim F_2}\{\log\lfloor f_1^p(F_0^{p^{-1}}[F_0(X)])/f_0^p(F_0^{p^{-1}}[F_0(X)])\rfloor\}$ based on a guess regarding the pre- and post-change distributions F_0, F_2 (here, f_0^p, f_1^p denote the pretended pre- and post-change density functions; F^{-1} corresponds to the inverse function of F). In this case, the null hypothesis H_0 is proposed to be rejected if the test statistic

$$D = \max\left\{\max_{\frac{n}{2}+1\le k\le n}\Lambda_k^n(\rho_n, \underline{X}),\ \max_{\frac{n}{2}+1\le k\le n}\Lambda_k^n(\rho_n^*, \underline{Z})\right\} > C_D, \tag{12}$$

for a fixed test threshold $C_D > 0$, where

$$\Lambda_k^n(\rho_n, \underline{X}) = \sum_{m=0}^{n}\lambda_{k,m}^n(\rho_n, \underline{X}), \tag{13}$$

$\lambda_{k,m}^n(\rho_n, \underline{X}) = \binom{n}{m}\left(\frac{1}{2}\right)^n\left(\frac{p\alpha}{q\beta}\right)^{U_k(m,n)}\quad(2q\beta)^{n+1-k}\prod_{i=1}^m\left(1+\frac{V_k(i,n)}{i}(\beta-1)\right)^{-1}\quad\prod_{i=m+1}^n\left(1+\frac{U_k(i-1,n)}{n+1-i}(\alpha-1)\right)^{-1}$, $U_k(m,n) = \sum_{j=k}^n I_{\{\rho(j,n)>m\}}$, $\rho(i,n) = \sum_{j=1}^n I_{\{X_j\le X_i\}}$ is the rank of observation X_i, $V_k(m,n) = (n+1-k) - U_k(m,n)$; p, q, α, β are some positive parameters, $q = 1-p$; $Z_i = -X_{n-i+1}$, $i = 1, \ldots, n$,

$$\Lambda_k^n(\rho_n^*, \underline{Z}) = \sum_{m=0}^{n}\lambda_{k,m}^n(\rho_n^*, \underline{Z}), \tag{14}$$

$$\lambda_{k,m}^n(\rho_n^*, \underline{Z}) = \binom{n}{m}\left(\frac{1}{2}\right)^n\left(\frac{p^*\alpha^*}{q^*\beta^*}\right)^{U_k^*(m,n)}(2q^*\beta^*)^{n+1-k}\prod_{i=1}^{m}\left(1+\frac{V_k^*(i,n)}{i}(\beta^*-1)\right)^{-1}$$
$$\times\prod_{i=m+1}^{n}\left(1+\frac{U_k^*(i-1,n)}{n+1-i}(\alpha^*-1)\right)^{-1},$$

$U_k^*(m,n) = \sum_{j=k}^n I_{\{\rho^*(j,n)>m\}}$, $\rho^*(i,n) = \sum_{j=1}^n I_{\{Z_j\le Z_i\}}$ is the rank of the observation Z_i, $V_k^*(m,n) = (n+1-k) - U_k^*(m,n)$, p^*, q^*, α^*, β^* are some positive parameters, $q^* = 1-p^*$. To obtain optimal values of p, α, β, p^*, α^*, β^*, corresponding to a maximum power of the test (12), Gurevich (2006) proposed to utilize different suppositions regarding F_1 and F_2. For example, when we suspect the pre- and post-change distribution functions are close to $N(0,1)$ and $N(1,1)$, respectively, the optimal values of the parameters are

$$p = p^* \approx 0.8413,\quad \alpha = \alpha^* \approx 0.531,\quad \beta = \beta^* \approx 1.703. \tag{15}$$

The proposed procedure possesses robustness of validity, because it is based on ranks. Thus, the procedure with near-optimal parameters obtained for specific alternatives is assumed to be a powerful change point detection policy for various alternatives (Gurevich, 2006; Gordon and Pollak, 1995). In this article, Sec. 4 will present Monte Carlo simulations to confirm this proposition.

To evaluate the p-values of the test (12), Gurevich (2006) derived asymptotically a distribution-free upper bound for the type I error of the policy (12).

We extend the policy of Gurevich (2006) to allow for cases when a stochastic order between the distribution functions F_1 and F_2 cannot be assumed to be known. Note that, when, under H_1, $X_\nu, \ldots, X_n \sim F_2$ are stochastically larger than $X_1, \ldots, X_{\nu-1}$, we can construct a test using similar schemes to those that are mentioned above (see the notations (12)–(15)). However, if, under the alternative H_1, observations $X_\nu, \ldots, X_n \sim F_2$ are stochastically smaller than $X_1, \ldots, X_{\nu-1}$ we transform the sample $\{X_1, \ldots, X_n\}$ to be presented in the form of $\{-X_1, \ldots, -X_n\}$. Then, $-X_1, \ldots, -X_{\nu-1} \sim F_0^*$; $-X_\nu, \ldots, -X_n \sim F_2^*$, where $F_0^*(x) = 1 - F_0(-x)$, $F_2^*(x) = 1 - F_2(-x)$. That is, $-X_\nu, \ldots, -X_n$ are stochastically larger than $-X_1, \ldots, -X_{\nu-1}$. For example, if $F_0 = N(0, 1)$ and $F_2 = N(-1, 1)$, then

$$F_0^* = N(0, 1) \prec F_2^* = N(1, 1). \tag{16}$$

Rewrite the definitions (13) and (14), utilizing $-X_i$, $i = 1, \ldots, n$, instead of X_i, $i = 1, \ldots, n$, we have

$$\Lambda 1_k^n(\rho 1_n, -\underline{X}) = \sum_{m=0}^{n} \lambda_{k,m}^n(\rho 1_n, -\underline{X}), \tag{17}$$

$\lambda 1_{k,m}^n(\rho 1_n, -\underline{X}) = \binom{n}{m}\left(\frac{1}{2}\right)^n\left(\frac{p_1\alpha_1}{q_1\beta_1}\right)^{U1_k(m,n)}$ $(2q_1\beta_1)^{n+1-k} \prod_{i=1}^{m}\left(1 + \frac{V1_k(i,n)}{i}(\beta_1 - 1)\right)^{-1}$ $\prod_{i=m+1}^{n}\left(1 + \frac{U1_k(i-1,n)}{n+1-i}(\alpha_1 - 1)\right)^{-1} U1_k(m, n) = \sum_{j=k}^{n} I_{\{\rho 1(j,n) > m\}}$, $\rho 1(i, n) = \sum_{j=1}^{n} I_{\{-X_j \le -X_i\}}$, $V1_k(m, n) = (n + 1 - k) - U1_k(m, n)$, $p_1, q_1, \alpha_1, \beta_1$ are certain positive parameters, $q_1 = 1 - p_1$, $-Z_i = X_{n-i+1}$, $i = 1, 2, \ldots, n$,

$$\Lambda 1_k^n(\rho 1_n^*, -\underline{Z}) = \sum_{m=0}^{n} \lambda 1_{k,m}^n(\rho 1_n^*, -\underline{Z}), \tag{18}$$

$$\lambda 1_{k,m}^n(\rho 1_n^*, -\underline{Z}) = \binom{n}{m}\left(\frac{1}{2}\right)^n\left(\frac{p_1^*\alpha_1^*}{q_1^*\beta_1^*}\right)^{U1_k^*(m,n)}(2q_1^*\beta_1^*)^{n+1-k}\prod_{i=1}^{m}\left(1 + \frac{V1_k^*(i, n)}{i}(\beta_1^* - 1)\right)^{-1}$$
$$\times \prod_{i=m+1}^{n}\left(1 + \frac{U1_k^*(i-1, n)}{n+1-i}(\alpha_1^* - 1)\right)^{-1}$$
$$U1_k^*(m, n) = \sum_{j=k}^{n} I_{\{\rho 1^*(j,n) > m\}}, \quad \rho 1^*(i, n) = \sum_{j=1}^{n} I_{\{-Z_j \le -Z_i\}}, \quad (i = 1, 2, \ldots, n),$$

$V1_k^*(m, n) = (n + 1 - k) - U1_k^*(m, n)$, p_1^*, q_1^*, α_1^*, β_1^* are certain positive parameters, $q_1^* = 1 - p_1^*$. The proposed policy is to reject H_0 if

$$DD = \max\left\{\max_{\frac{n}{2}+1 \le k \le n}\left(\frac{1}{2}\Lambda_k^n(\rho_n, \underline{X}) + \frac{1}{2}\Lambda_k^n(\rho_n^*, \underline{Z})\right),\right.$$
$$\left.\max_{\frac{n}{2}+1 \le k \le n}\left(\frac{1}{2}\Lambda 1_k^n(\rho 1_n, -\underline{X}) + \frac{1}{2}\Lambda 1_k^n(\rho 1_n^*, -\underline{Z})\right)\right\} > C_{DD}, \tag{19}$$

for a fixed test threshold $C_{DD} > 0$. Optimal values of the parameters p_1, α_1, β_1, p_1^*, α_1^*, β_1^* can be defined via the method presented by Gurevich (2006), when suspected representatives of the pre- and post-change distributions are used. To control the type I error of the test (19), we present the next proposition.

Proposition 2. *Set up $p\alpha \geq q\beta$, $p_1\alpha_1 \geq q_1\beta_1$, $p^*\alpha^* \geq q^*\beta^*$, $p_1^*\alpha_1^* \geq q_1^*\beta_1^*$. Then, for all $C_{DD} > 0$,*

$$\lim_{n\to\infty} \sup_{F_0} P_{H_0}(DD > C_{DD}) \leq \frac{2}{C_{DD}}. \tag{20}$$

The proof scheme of Proposition 2 *is technically based on that of Theorem* 2.1 *presented by Gurevich* (2006). *Proposition* 2 *provides a distribution free upper bound for the significance level of the test* (19). *That is, for large values of the sample size n, we can use $2/C_{DD}$ to approximate the significance level of the policy* (19). *In Sec.* 4, *we Monte Carlo study the accuracy of this approximation based on data with different sample sizes.*

Remark. One can show that if $F_0 = F_1$ and F_2 are expected to be close to $N(0, 1)$ and $N(\mu, 1)$ with $|\mu| = 1$, then the optimal values of the parameters of (19) are

$$p = p_1 = p^* = p_1^* \approx 0.8413, \quad \alpha = \alpha_1 = \alpha^* = \alpha_1^* \approx 0.531, \quad \beta = \beta_1 = \beta^* = \beta_1^* \approx 1.703. \tag{21}$$

Thus, we can conclude that the well-addressed nonparametric tests for (1) in the literature have been initially defined in cases with stochastically ordered one-sided alternatives. Then, practical applications have required modifying the tests to be adjusted for two-sided alternatives. Note also that, commonly, the change point literature has paid attention to different comparisons between the powers of change point policies, when the pre- and post-change distributions (F_1 and F_2, respectively) have the same form with different parameters (e.g., Ferger, 1994; Gurevich, 2006; Wolfe and Schechtman, 1984).

Alternatively to traditional change point detection schemes' constructions, Vexler and Gurevich (2009b) proposed to approximate nonparametrically the likelihood ratio's components of the parametric CUSUM (2) and SR (5) test statistics. Toward this end, principles of the empirical likelihood methodology (e.g., Owen, 2001; Vexler et al., 2009a) were proposed to be applied. To approximate likelihood ratios, Vexler and Gurevich (2009b) used and extended a nonparametric methodology proposed by Vexler and Gurevich (2010), considering the likelihood ratio $\prod_{i=1}^{k-1} \frac{f_1(X_i)}{f_0(X_i)} \prod_{i=k}^{n} \frac{f_2(X_i)}{f_0(X_i)}$ from (2) as a product of n unknown parameters that should be maximum likelihood estimated, under constraints having forms of empirical approximations to $\int f_1 du = 1$ and $\int f_2 du = 1$. That is, for example, the Lagrange multiplier method provides values of f_{1r}/f_{0r}, $r = 1, \ldots, k-1$ to approximate the ratios $f_1(X_{(r:k-1)})/f_0(X_{(r:k-1)})$, $r = 1, \ldots, k-1$ that are present in the likelihood ratio $\prod_{r=1}^{k-1} \frac{f_1(X_{(k:k-1)})}{f_0(X_{(k:k-1)})} = \prod_{r=1}^{k-1} \frac{f_1(X_r)}{f_0(X_r)}$, where $X_{(r:k-1)}$ is the r-order statistic based on $X_1, \ldots, X_{k-1}$. Toward this end, f_{1r}/f_{0r}, $r = 1, \ldots, k-1$ can be chosen to maximize the corresponding log likelihood function provided that an empirical approximation to $\int f_1 du = 1$ is satisfied; i.e., f_{1r}/f_{0r}, $r = 1, \ldots, k-1$ can be derived from the equation

$$\frac{\partial}{\partial(f_{1r}/f_{0r})}\left[\sum_{i=1}^{k-1} \log \frac{f_{1i}}{f_{0i}} + \lambda\left(1 - \sum_{i=1}^{k-1} \frac{f_{1i}}{f_{0i}} \Delta_{i,k-1}\right)\right] = 0,$$

where λ is the Lagrange multiplier; $\sum_{i=1}^{k-1} \frac{f_{1i}}{f_{0i}} \Delta_{i,k-1}$ is assumed to be equal to 1, because $\Delta_{i,k-1}$ must be defined to approximate empirically $\int f_1 du$ using

$\sum_{i=1}^{k-1} \frac{f_{1i}}{f_{0i}} \Delta_{i,k-1}$. In a similar manner to the $\prod_{r=1}^{k-1} \frac{f_1(X_r)}{f_0(X_r)}$ approximation, the likelihood ratio $\prod_{i=k}^{n} \frac{f_2(X_i)}{f_0(X_i)}$ can be evaluated. Thus, one can show that Λ_k^n from (2) can be approximated by

$$\widetilde{\Lambda}_k^n = \min_{1 \le m \le (k-1)^{1-\delta}} \left(\prod_{i=1}^{k-1} \frac{2m}{(k-1)(F_{0n}(Z_{(i+m)}) - F_{0n}(Z_{(i-m)}))} \right) \times \min_{1 \le r \le (n-k+1)^{1-\delta}} \left(\prod_{j=1}^{n-k+1} \frac{2r}{(n-k+1)(F_{0n}(Y_{(j+r)}) - F_{0n}(Y_{(j-r)}))} \right), \tag{22}$$

where $0 < \delta < 1$, $Z_i = X_i$, $i = 1, \dots, k-1$, $Y_j = X_{k-1+j}$, $j = 1, \dots, n-k+1$; $Z_{(l)}$ is the order statistic based on $Z_1, \dots, Z_k$, $Z_{(l)} = Z_{(1)}$, if $l \le 1$, and $Z_{(l)} = Z_{(k-1)}$, if $l \ge k-1$; $Y_{(l)}$ is the order statistic based on $Y_1, \dots, Y_{n-k+1}$, $Y_{(l)} = Y_{(1)}$, if $l \le 1$, and $Y_{(l)} = Y_{(n-k+1)}$, if $l \ge n-k+1$; $F_{0n}(x) = n^{-1} \sum_{i=1}^{n} I_{\{X_i \le x\}} = n^{-1}\big(\sum_{i=1}^{k-1} I_{\{Z_i \le x\}} + \sum_{j=1}^{n-k+1} I_{\{Y_j \le x\}} \big)$ is the empirical distribution function that estimates $F_0(x)$.

Vexler and Gurevich (2009b) proved that the approximate likelihood ratio (22) is an entropy-based likelihood ratio. Entropy-based methods are well developed in the context of tests for goodness of fit (e.g., Vasicek, 1976). The statistics $\widetilde{\Lambda}_k^n$ used instead of Λ_k^n in the structures of the CUSUM and SR statistics provide the very powerful nonparametric test statistics

$$\widetilde{\Delta}_n = \max_{2 \le k \le n} \widetilde{\Lambda}_k^n, \tag{23}$$

$$\widetilde{R}_n = \sum_{k=2}^{n} \widetilde{\Lambda}_k^n. \tag{24}$$

In the next section, we compare numerically the considered tests. We will also Monte Carlo study how parametric assumptions mentioned in Sec. 2 are robust, in the context of the parametric change point detection schemes, compared with the nonparametric policies.

4. Monte Carlo Study

To examine the change point policies, we begin with definitions of notations presented in Table 1. These notations will be utilized in this section. T1

To conduct the Monte Carlo simulations below, for each distribution set with different sample sizes, we generated 10,000 times corresponding data. The Monte Carlo powers of the nonparametric tests were evaluated at the level of significance 0.05 that was fixed experimentally by choosing special values of the test thresholds. (Under the null hypothesis of (1), the baseline distribution functions of the nonparametric test statistics of *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR* do not depend on data distributions, only tables of critical values of the tests are required for their implementation.) The 95% critical values of the parametric tests *PCUS* and *PSR* were calculated using the assumption that under the null hypothesis we observe data from $f_0(x) = f_{N(0,1)}(x)$, because this assumption was used theoretically to construct the structure of the tests (i.e., when we apply *PCUS* and *PSR*, we believe (quasi-correctly) the parametric assumption).

Table 1
The notations and their descriptions that are utilized in Sec. 4

Notation	Description
RDD	Nonparametric test (19), where the parameters of the statistic *DD* are defined by (21);
RAK	Nonparametric test based on the absolute value of the statistic (10); i.e., we assume rejecting H_0, if $\lvert K\rvert > C_K$;
RAD_1	Nonparametric test based on the absolute value of the statistics (9); i.e., reject H_0 if $\lvert D_1\rvert > C_{D_1}$;
NPCUS	Nonparametric test based on the statistic (23); i.e., to reject H_0 if $\widetilde{\Delta}_n > C_{\widetilde{\Delta}}$;
NPSR	Nonparametric test based on the statistic (24); i.e., to reject H_0, if $\widetilde{R}_n > C_{\widetilde{R}}$;
PCUS	Parametric test (4) based on the adjusted CUSUM statistic, when a tester believes that $f_0(x) = f_{N(\mu,\sigma^2)}(x)$, $f_1(x) = f_{N(\mu+\beta_1\sigma,\sigma^2)}(x)$, $f_2(x) = f_{N(\mu+\beta_2\sigma,\sigma^2)}(x)$ with unknown μ, σ, β_i, $i = 1, 2$, $\beta_1 \neq \beta_2$; i.e., *PCUS* declares rejection of H_0, if $\max_{1<k\leq n} \Lambda_k^{*n} = \max_{2\leq k\leq n} \left(\left\{(k-1)S_n/n - S_{k-1}\right\}/\left\{(k-1)[1-(k-1)/n]\right\}^{\frac{1}{2}}\right) > C_1$, $S_k = \sum_{i=1}^{k} X_i$;
PSR	Parametric test (8) based on the adjusted Shiryayev–Roberts type statistic, when a tester believes that $f_0(x) = f_{N(\mu,\sigma^2)}(x)$, $f_1(x) = f_{N(\mu+\beta_1\sigma,\sigma^2)}(x)$, $f_2(x) = f_{N(\mu+\beta_2\sigma,\sigma^2)}(x)$ with unknown μ, σ, β_i, $i = 1, 2$, $\beta_1 \neq \beta_2$; i.e., *PSR* rejects H_0 if $\sum_{k=2}^{n} \left\{(k-1)S_n/n - S_{k-1}\right\}/\left\{(k-1)(1-(k-1)/n)\right\}^{\frac{1}{2}} > C_{SR}$.

Table 2 reports a Monte Carlo comparison of the powers of the nonparametric tests $RDD, RAK, RAD_1, NPCUS$, and *NPSR* when the actual pre- and post-change distributions are $N(\mu, \sigma^2)$ and $N(\mu + \beta\sigma, \sigma^2)$, $\beta \neq 0$, respectively. (Note that computations of the statistics of $RDD, RAK, RAD_1, NPCUS$, and *NPSR* do not depend on the parameters μ and σ.) In this case, it is clear that the tests *PCUS* and *PSR* are created utilizing the correct information regarding the actual pre- and post-change distributions and hence these tests are expected to be very powerful. Therefore, we calculated and presented the Monte Carlo powers of the parametric tests in Table 2 to judge the nonparametric procedures. (The statistics of the parametric tests do not depend on the parameter μ but do depend on the parameter σ.) T2

The power functions of the tests are symmetric with respect to values of $v - 1$ regarding $v - 1 = n/2$. We note also that the powers of these tests do not depend on a sign of β, for all fixed v. In accordance with Table 2, the parametric *PSR* test is mostly more powerful (not just more powerful in average; see Proposition 1) than the well accepted *PCUS* test in the literature. However, when the change point location v is relatively very close to 1, *PCUS* is weakly superior to *PSR*. The proposed nonparametric procedure *RDD* is very efficient, especially when v is not relatively large; however, we should note that the parameters setting (21) used in *RDD* is appropriate to the cases considered in Table 2 (e.g., when $n = 70$, $v - 1 = 10$, $\beta = 1$, *RDD* demonstrated the power that is comparable with that of the parametric tests). Generally speaking, all the tests displayed good power properties in Table 1. Here, RAD_1 and *RAK* are known to be especially very efficient when

Table 2
The Monte Carlo powers of the nonparametric tests *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR*, and the parametric tests *PCUS*, *PSR*, when the actual pre- and post-change distributions are $F_0 = F_1 = N(0, 1)$ and $F_2 = N(\beta, 1)$, correspondingly. Observations with the subscript $\nu - 1$ are the last observations before the change. The significance level is $\alpha = 0.05$

n	β	$\nu - 1$	*RDD*	*RAK*	RAD_1	*NPCUS*	*NPSR*	*PCUS*	*PSR*
20	0.8	10	0.283	0.309	0.260	0.215	0.223	0.280	0.331
		5	0.175	0.162	0.177	0.144	0.140	0.205	0.220
		3	0.096	0.080	0.098	0.084	0.085	0.144	0.134
	1.0	10	0.411	0.452	0.383	0.309	0.322	0.421	0.486
		5	0.257	0.229	0.257	0.201	0.199	0.311	0.326
		3	0.125	0.093	0.128	0.107	0.108	0.206	0.187
	1.2	10	0.558	0.602	0.525	0.428	0.447	0.582	0.638
		5	0.355	0.310	0.355	0.279	0.277	0.440	0.454
		3	0.159	0.109	0.164	0.138	0.139	0.291	0.252
40	0.8	20	0.518	0.570	0.487	0.368	0.401	0.513	0.606
		10	0.376	0.329	0.357	0.228	0.241	0.385	0.422
		5	0.170	0.099	0.174	0.106	0.104	0.221	0.205
	1.0	20	0.720	0.769	0.688	0.553	0.592	0.727	0.797
		10	0.545	0.490	0.526	0.358	0.374	0.576	0.609
		5	0.255	0.128	0.263	0.147	0.145	0.332	0.290
	1.2	20	0.873	0.898	0.852	0.732	0.765	0.884	0.921
		10	0.723	0.656	0.694	0.513	0.529	0.763	0.776
		5	0.358	0.162	0.367	0.209	0.198	0.473	0.400
70	0.8	35	0.750	0.847	0.766	0.614	0.638	0.771	0.835
		20	0.659	0.675	0.648	0.453	0.460	0.667	0.707
		10	0.417	0.235	0.382	0.182	0.172	0.404	0.362
	1.0	35	0.925	0.960	0.929	0.831	0.846	0.937	0.957
		20	0.855	0.869	0.856	0.682	0.693	0.873	0.891
		10	0.612	0.361	0.571	0.291	0.276	0.616	0.544
	1.2	35	0.986	0.993	0.986	0.953	0.958	0.990	0.994
		20	0.958	0.966	0.960	0.864	0.866	0.967	0.969
		10	0.781	0.505	0.748	0.448	0.427	0.794	0.723

the change β in a location of observations is in effect (Wolfe and Schechtman, 1984), whereas tests *NPCUS* and *NPSR* are developed to attend to various complex alternatives.

In Table 3, we present results of the Monte Carlo comparison between the powers of the tests *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR*, *PCUS*, *PSR*, when the actual pre- and post-change distributions are $T_{(3)}(\eta)$ and $T_{(3)}(\eta + \beta)$, $\beta \neq 0$, respectively, where $T_{(3)}(\eta)$ denotes a noncentral Student's distribution with three degree of freedom and the parameter of noncentrality η. None of the test-statistics depend on the parameter η.

In fact, the powers of the parametric tests are incorrect, because the actual Type I error of this tests is not 0.05 (the corresponding critical values were

Table 3
The Monte Carlo powers of the tests *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR*, *PCUS*, *PSR*, when the actual distributions are $F_0 = F_1 = T_{(3)}(0)$ and $F_2 = T_{(3)}(\beta)$. Observations with the subscript $v-1$ are the last observations before the change. The significance level is fixed to be $\alpha = 0.05$

n	β	$v-1$	*RDD*	*RAK*	RAD_1	*NPCUS*	*NPSR*	*PCUS*	*PSR*
20	0.8	10	0.262	0.290	0.241	0.193	0.201	0.666	0.627
		5	0.173	0.158	0.174	0.139	0.134	0.618	0.561
		3	0.098	0.081	0.099	0.090	0.088	0.565	0.483
	1.0	10	0.371	0.413	0.352	0.280	0.292	0.756	0.731
		5	0.241	0.219	0.243	0.192	0.190	0.711	0.657
		3	0.121	0.095	0.126	0.111	0.113	0.640	0.564
	1.2	10	0.489	0.543	0.466	0.383	0.400	0.845	0.824
		5	0.325	0.288	0.326	0.258	0.263	0.791	0.749
		3	0.156	0.112	0.160	0.136	0.141	0.719	0.639
40	0.8	20	0.460	0.535	0.451	0.332	0.363	0.854	0.826
		10	0.329	0.301	0.325	0.207	0.217	0.794	0.747
		5	0.153	0.095	0.164	0.099	0.099	0.694	0.603
	1.0	20	0.642	0.728	0.645	0.490	0.524	0.931	0.919
		10	0.490	0.445	0.479	0.314	0.333	0.886	0.848
		5	0.225	0.123	0.241	0.135	0.132	0.784	0.698
	1.2	20	0.800	0.866	0.805	0.653	0.685	0.974	0.968
		10	0.653	0.600	0.638	0.446	0.471	0.950	0.924
		5	0.319	0.154	0.339	0.190	0.183	0.865	0.789
70	0.8	35	0.663	0.796	0.714	0.534	0.554	0.950	0.940
		20	0.577	0.624	0.599	0.399	0.406	0.924	0.896
		10	0.359	0.217	0.352	0.174	0.166	0.838	0.742
	1.0	35	0.856	0.938	0.889	0.755	0.770	0.988	0.984
		20	0.780	0.816	0.800	0.601	0.610	0.975	0.963
		10	0.534	0.320	0.529	0.266	0.255	0.921	0.851
	1.2	35	0.956	0.986	0.971	0.903	0.912	0.998	0.997
		20	0.914	0.931	0.924	0.776	0.786	0.995	0.990
		10	0.703	0.450	0.689	0.391	0.380	0.970	0.927

still chosen under the conjecture $f_0(x) = f_{N(\mu_0,1)}(x)$; see the paragraph above the description for Table 2). Table 6 presents the actual Type I error of *PCUS* and *PSR* that are not close to the expected level 0.05, in the case of $F_0 = T_{(3)}(0)$. However, with respect to the risk Pr{to reject $H_0|H_1$}–Pr{to reject $H_0|H_0$}, the rule *PSR* seems to be more preferable than *PCUS*. Note that, in this experiment, the power functions of the tests are symmetric around $v - 1 = n/2$ and the powers of the tests do not depend on a sign of β for a fixed v. Although the parameters setting (21) used in *RDD* does not allow for the situations of Table 3, the test *RDD* is shown to be reasonable to be applied, especially when values of v are expected to be relatively close to 1 or n.

Table 4 Monte Carlo compares between the powers of the tests *RDD*, T4
RAK, RAD_1, *NPCUS*, *NPSR*, *PCUS*, *PSR*, when the actual pre- and post-change distributions are $N(0, 1)$ and $Unif(a, b)$, $a < b$, respectively.

In Table 4, the parametric decision rule *PSR* is shown to be more preferable than the test *PCUS*. Table 4 considers a setting of the statement (1), when the post-change distribution has a form that is different from a pre-change distribution function's form. In this case, we observe the proposed *NPCUS* and *NPSR* procedures are very efficient. Especially when the variance of observations was changed from 1 to 0.083 ($N(0,1) \mapsto Unif[0,1]$), the powers of *NPCUS* and *NPSR* are frequently superior to the powers multiplied by more than two of the other nonparametric tests. In accordance with Table 4, the powers of the proposed nonparametric tests *RDD* are significantly superior to those of the parametric tests *PCUS* and *PSR*.

Table 4

The Monte Carlo powers of the tests *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR*, *PCUS*, *PSR*, when the actual distributions are $F_0 = F_1 = N(0,1)$ and $F_2 = Unif[a,b]$. Observations with the subscript $v-1$ are the last observations before the change. The significance level of the tests is fixed to be $\alpha = 0.05$

n	a	b	$v-1$	*RDD*	*RAK*	RAD_1	*NPCUS*	*NPSR*	*PCUS*	*PSR*
20	0.0	2.0	17	0.131	0.108	0.134	0.123	0.123	0.113	0.133
			15	0.310	0.307	0.333	0.276	0.263	0.212	0.271
			10	0.552	0.617	0.542	0.512	0.538	0.338	0.439
			5	0.450	0.385	0.413	0.352	0.367	0.228	0.246
			3	0.251	0.111	0.246	0.232	0.238	0.134	0.100
	0.0	1.0	17	0.063	0.074	0.070	0.100	0.100	0.034	0.033
			15	0.103	0.134	0.122	0.287	0.267	0.036	0.041
			10	0.210	0.278	0.247	0.402	0.465	0.049	0.062
			5	0.259	0.185	0.238	0.236	0.257	0.040	0.031
			3	0.197	0.081	0.197	0.196	0.202	0.030	0.010
40	0.0	2.0	35	0.235	0.161	0.318	0.181	0.165	0.214	0.231
			30	0.633	0.687	0.727	0.645	0.632	0.529	0.607
			20	0.847	0.908	0.865	0.928	0.938	0.706	0.812
			10	0.773	0.657	0.721	0.702	0.753	0.500	0.552
			5	0.526	0.179	0.444	0.340	0.348	0.251	0.184
	0.0	1.0	35	0.069	0.086	0.093	0.252	0.226	0.039	0.053
			30	0.151	0.242	0.235	0.855	0.835	0.057	0.098
			20	0.318	0.455	0.407	0.970	0.979	0.104	0.173
			10	0.367	0.282	0.365	0.539	0.673	0.077	0.078
			5	0.310	0.103	0.267	0.243	0.263	0.045	0.015
70	0.0	2.0	60	0.668	0.509	0.783	0.537	0.468	0.590	0.539
			50	0.946	0.976	0.971	0.993	0.991	0.910	0.934
			35	0.979	0.993	0.986	1.000	1.000	0.949	0.976
			20	0.959	0.944	0.942	0.986	0.990	0.860	0.897
			10	0.861	0.520	0.752	0.673	0.720	0.555	0.479
	0.0	1.0	60	0.113	0.166	0.200	0.861	0.781	0.057	0.095
			50	0.310	0.544	0.479	1.000	1.000	0.128	0.232
			35	0.469	0.698	0.614	1.000	1.000	0.215	0.353
			20	0.528	0.520	0.549	0.990	0.996	0.179	0.218
			10	0.484	0.216	0.399	0.527	0.659	0.091	0.049

Table 5

The Monte Carlo powers (at $\alpha = 0.05$) of the tests *RDD*, *RAK*, RAD_1, *NPCUS*, *NPSR*, *PCUS*, *PSR*, $PCUS^*$, PSR^* when the actual distributions are $F_0 = N(0, 1)$, $F_1 = N(0, 0.5^2)$, $F_2 = N(0, 1.5^2)$ (design (a)); and $F_0 = N(0.7, 1/12)$, $F_1 = Exp(1)$, $F_2 = Unif(0, 1)$ (design (b)). Observations with the subscript $\nu - 1$ are the last observations before the change

n	$\nu - 1$	*RDD*	*RAK*	RAD_1	*NPCUS*	*NPSR*	*PCUS*	*PSR*	$PCUS^*$	PSR^*
Design (a)										
20	17	0.100	0.058	0.100	0.098	0.103	0.108	0.047	0.301	0.336
	15	0.118	0.074	0.112	0.121	0.131	0.139	0.071	0.378	0.436
	10	0.085	0.079	0.087	0.206	0.252	0.185	0.095	0.341	0.414
	5	0.056	0.059	0.057	0.203	0.196	0.237	0.135	0.138	0.159
	3	0.051	0.053	0.047	0.086	0.087	0.270	0.169	0.093	0.099
40	35	0.138	0.057	0.128	0.111	0.121	0.131	0.039	0.581	0.630
	30	0.132	0.073	0.132	0.221	0.314	0.172	0.070	0.796	0.850
	20	0.082	0.081	0.100	0.723	0.789	0.222	0.100	0.846	0.905
	10	0.054	0.061	0.062	0.590	0.585	0.272	0.145	0.467	0.561
	5	0.048	0.058	0.050	0.181	0.168	0.313	0.196	0.152	0.176
70	60	0.199	0.062	0.148	0.215	0.279	0.164	0.045	0.911	0.917
	50	0.162	0.078	0.143	0.762	0.848	0.205	0.077	0.985	0.991
	35	0.094	0.085	0.105	0.988	0.990	0.244	0.100	0.994	0.997
	20	0.058	0.068	0.067	0.962	0.957	0.284	0.134	0.954	0.977
	10	0.045	0.052	0.049	0.541	0.479	0.320	0.177	0.505	0.576
Design (b)										
20	17	0.059	0.066	0.061	0.069	0.069	0.063	0.052	0.952	0.968
	15	0.076	0.090	0.082	0.113	0.111	0.070	0.066	0.961	0.975
	10	0.147	0.152	0.135	0.171	0.191	0.089	0.086	0.955	0.968
	5	0.144	0.109	0.140	0.127	0.136	0.083	0.055	0.892	0.917
	3	0.114	0.064	0.111	0.104	0.108	0.065	0.032	0.815	0.851
40	35	0.065	0.065	0.070	0.104	0.100	0.066	0.066	0.999	1.000
	30	0.124	0.138	0.128	0.299	0.297	0.090	0.114	0.992	1.000
	20	0.242	0.257	0.222	0.556	0.621	0.136	0.181	0.999	0.999
	10	0.260	0.160	0.209	0.244	0.314	0.123	0.099	0.995	0.997
	5	0.188	0.073	0.155	0.131	0.143	0.084	0.033	0.985	0.989
70	60	0.105	0.110	0.117	0.237	0.204	0.083	0.099	1.000	1.000
	50	0.240	0.287	0.246	0.817	0.789	0.149	0.228	1.000	1.000
	35	0.372	0.432	0.354	0.951	0.959	0.226	0.337	1.000	1.000
	20	0.433	0.309	0.329	0.691	0.787	0.195	0.212	1.000	1.000
	10	0.385	0.131	0.228	0.241	0.298	0.126	0.063	1.000	1.000

(The test *RDD* was created pretending the alternative distributions have forms that belong to exponential families. Moreover, the test statistic of *RDD* contains parameters with optimal values obtained assuming F_1 and F_2 are expected to be close to $N(0, 1)$ and $N(\mu, 1)$.) Thus, in this case, the power property of *RDD* is demonstrated to be more robust than that of the parametric tests *PCUS* and *PSR*, when the assumptions regarding distributions F_1 and F_2 are incorrect.

Table 5 depicts the Monte Carlo comparison between the powers of the nonparametric and parametric decision rules corresponding to the designs $F_0 = N(0, 1)$, $F_1 = N(0, 0.5^2)$, $F_2 = N(0, 1.5^2)$ (the design (a) of the table) and $F_0 = N(0.7, 1/12)$, $F_1 = Exp(1)$, $F_2 = Unif(0, 1)$ (the design (b) of the table). Taking into account the designs of Table 5 and utilizing the methods of Sec. 2, we constructed the correct parametric tests $PCUS^*$ and PSR^* based on the test statistics $\max_{1<k\leq n} \widetilde{\Lambda}_k^{*n}$ and $\sum_{k=2}^{n} \widetilde{\Lambda}_k^{*n}$, respectively, with

$$\widetilde{\Lambda}_k^{*n} = \frac{\sup_{\theta_1\in\Theta} \prod_{i=1}^{k-1} f_1(X_i; \theta_1) \sup_{\theta_2\in\Theta} \prod_{i=k}^{n} f_2(X_i; \theta_2)}{\sup_{\theta_0\in\Theta} \prod_{i=1}^{n} f_0(X_i; \theta_0)},$$

where

$$\widetilde{\Lambda}_k^{*n} = \begin{cases} S_0^n \left(S_1^{k-1} S_2^{n-k+1}\right)^{-1} & \text{if } 2 < k < n \\ 0 & \text{if } k = 2, n \end{cases}, \quad S_0 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}, \quad \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$S_1 = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k-1}\left(X_i - \overline{X}_{k-1}\right)^2}, \quad \overline{X}_{k-1} = \frac{1}{k-1}\sum_{i=1}^{k-1} X_i,$$

$$S_2 = \sqrt{\frac{1}{n-k+1}\sum_{i=k}^{n}\left(X_i - \overline{X}^*_{n-k+1}\right)^2},$$

$\overline{X}^*_{n-k+1} = \frac{1}{n-k+1}\sum_{i=k}^{n} X_i$, for the design (a);

$$\widetilde{\Lambda}_k^{*n} = \frac{\left(\overline{X}_{k-1} e\right)^{-(k-1)} (\max_{k\leq i\leq n} X_i - \min_{k\leq i\leq n} X_i)^{-(n-k+1)} I_{\{\min_{1\leq i<k} X_i > 0\}}}{\left(2\pi e S_0^2\right)^{-n/2}},$$

for the design (b). Here the correct information regarding parametric forms of the distributions F_0, F_1, and F_2 are applied.

Table 5 demonstrates the parametric tests to be very efficient, provided that correct parametric forms of the null and alternative distributions F_0, F_1, and F_2 are known. In these examples, the proposed test PSR^* is superior to the classical test $PCUS^*$ based on the CUSUM-type statistic. The parametric tests *PCUS* and *PSR* have weak powers even for $n = 40, 70$. That is to say, the parametric tests for the change point problem do not have robust power properties. Note that, perhaps, the problem of developing goodness-of-fit tests under the regime of (1) does not have simple solutions. The nonparametric tests *RDD*, *RAK*, and RAD_1 are shown to be inefficient (these tests are close to being biased, in certain cases). This is partly because the tests *RDD*, *RAK*, RAD_1 were proposed assuming the stochastically ordered alternatives. The proposed *NPCUS* and *NPSR* procedures are superior to the rest of the considered nonparametric tests in terms of their powers, in almost all Monte Carlo experiments conducted to present Table 5. The simulations related to design (b) of Table 5 confirm that the proposed nonparametric test *RDD* is significantly more robust to the assumptions on distributions F_0, F_1, and F_2 than the parametric tests *PCUS* and *PSR*.

To investigate a sensitivity of the parametric change point policies with respect to assumptions required for correct executions of the procedures, we conducted Monte Carlo experiments. The outputs of these experiments are shown in Table 6.

Generally speaking, we must note that these parametric tests are not robust, in the context of the Type I error control. When the degree of freedom of a t-distribution is greater than 25, the actual type I error is comparable with the expected 0.05. However, it is clear: testers should be very accurate when choosing parametric forms of the null distribution of (1). This issue has not been well addressed in the literature. In this context, the proposed decision rule *PSR* is also shown to be more preferable than the classical retrospective detection scheme *PCUS*.

Table 6
The actual Monte Carlo type I errors of the parametric tests *PCUS* and *PSR* with the critical values that correspond to the conjecture $f_0(x) = f_{N(0,1)}(x)$ for the different null distributions and sample sizes n

Null distribution	n	Type I error of the *PCUS*	Type I error of the *PSR*
$T_{(2)}(0)$	20	0.626	0.529
	40	0.744	0.636
	70	0.820	0.689
$T_{(3)}(0)$	20	0.402	0.316
	40	0.472	0.360
	70	0.534	0.373
$T_{(10)}(0)$	20	0.119	0.093
	40	0.126	0.095
	70	0.135	0.096
$T_{(15)}(0)$	20	0.088	0.072
	40	0.096	0.077
	70	0.102	0.080
$T_{(25)}(0)$	20	0.072	0.064
	40	0.075	0.068
	70	0.075	0.069
$LogNorm(1, 1)$	20	0.970	0.951
	40	0.995	0.988
	70	1.000	1.000
$Unif[0, 1]$	20	< 0.005	< 0.005
	40	< 0.005	< 0.005
	70	< 0.005	< 0.005
$Exp(1)$	20	0.098	0.067
	40	0.101	0.062
	70	0.105	0.056
$Norm(0, 0.5^2)$	20	< 0.005	< 0.005
	40	< 0.005	< 0.005
	70	< 0.005	< 0.005
$Norm(0, 1.5^2)$	20	0.393	0.281
	40	0.446	0.312
	70	0.473	0.298

Table 7
The Monte Carlo levels of significance of the test *RDD* with the threshold $C_{DD} = 2/\alpha = 40$

Sample size	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 70$	$n = 100$	$n = 150$
Level of significance	0.000	0.005	0.021	0.032	0.043	0.042	0.037

The biases between the actual and expected the type I errors depend on the sample sizes n, but these dependencies are not strong.

Proposition 2 provides asymptotically ($n \to \infty$) the distribution free upper bound for the Type I error of the test *RDD*. To examine the accuracy of this result, we determine a critical value for the test *RDD*, taking into account the inequality (20), in the form of $2/\alpha$, for finite sample sizes. We present in Table 7 the T7 Monte Carlo Type I errors of the test *RDD* with the test threshold $C_{DD} = 2/\alpha = 40$, expecting to obtain the nominal significance level that is close to $\alpha = 0.05$.

In this case, we do not recommend to utilize Proposition 2 for obtaining the critical values of the test *RDD*, when $n \leq 20$; however, when $n \geq 30$ the asymptotic result (20) can be used in practice.

Remark. This article focuses on testing the hypothesis of (1). Gurevich and Vexler (2005, 2006) showed that, in general, a process of estimation of the change point

Table 8
The Monte Carlo means and standard deviations of the estimator $\hat{\nu}$, when $F_1 = Exp(1)$, $F_2 = Unif(0, 1)$, for different sample sizes n and values of ν

	$n = 40$					
ν	∞ (No change)	21	36			
Mean	21.048	20.835	25.617			
STD	10.534	4.770	10.411			
	$n = 70$					
ν	∞ (No change)	21	36			
Mean	35.779	24.176	35.309			
STD	16.681	8.154	4.059			
		$n = 100$				
ν	∞ (No change)	21	36	51		
Mean	51.337	28.296	36.126	49.955		
STD	21.963	14.027	5.127	3.531		
			$n = 120$			
ν	∞ (No change)	21	36	51	61	
Mean	60.856	30.869	36.440	50.180	59.846	
STD	25.213	17.459	6.046	3.639	3.389	
			$n = 150$			
ν	∞ (No change)	21	36	51	61	76
Mean	76.416	35.941	37.449	50.453	60.004	74.772
STD	29.391	23.726	8.472	4.126	3.376	3.079

v should be started if needed, provided that just the null hypothesis is rejected. When H_0 is rejected, the issue to estimate the unknown parameter v can be stated. Borovkov (1999) as well as Gurevich and Vexler (2005) investigated different parametric estimators of the change point v based on the likelihood ratios Λ_k^n or Λ_k^{*n}. Section 3 introduces the nonparametric tests *NPCUS* and *NPSR* based on the approximations $\widetilde{\Lambda}_k^n$ from (22) to the parametric likelihood ratio Λ_k^n. Thus, combining the material of Sec. 3 and the techniques of Borovkov (1999) and Gurevich and Vexler (2005), we can propose, e.g., the maximum nonparametric likelihood estimator

$$\hat{v} = \arg \max_{2 \le k \le n} \widetilde{\Lambda}_k^n$$

of v. Theoretical evaluations of $\hat{v}$ need substantial mathematical details that are beyond the scope of this article. To illustrate briefly the behavior of the estimator, we conducted the following experiments. Table 8 presents the Monte Carlo estimators of means and standard deviations of the estimator $\hat{v}$, when samples of Xs were drawn from $F_1 = Exp(1)$, $F_2 = Unif(0, 1)$, for different sample sizes n and values of v.

It seems from Table 8 that the estimator $\hat{v}$ is consistent when $v \to \infty$, $n - v \to \infty$ and can be recommended to be applied in practice.

5. A Data Example

In this section, we exemplify the proposed methods to evaluate a biomarker related to atherosclerotic coronary heart disease in the context of having potential discriminatory abilities for myocardial infarction (MI). We consider the biomarker called *Cholesterol* that measures sub-products of lipid peroxidation and has been proposed as a discriminating measurement between individuals with cardiovascular disease and healthy populations (for details see, e.g., Vexler et al., 2008a,b). A cohort of 799 men and women without myocardial infarction (say, MI $= 0$) and 143 individuals who recently survived an MI (say, MI $= 1$) were selected for the analyses to present data that contain measurements of cholesterol. Participants provided a 12-h fasting food specimen for biochemical analysis at baseline, and a number of parameters were examined from fresh blood samples.

5.1. *Type I Error Evaluation*

To evaluate the type I error of the proposed tests we apply a bootstrap-type procedure. The strategy was that a Bernoulli random variable $\tau(\Pr\{\tau = 0\} = \Pr\{\tau = 1\} = 1/2)$ was generated and then a sample of cholesterol measurements with size $n = 70$ was randomly selected from individuals with MI $= \tau$. We repeated this strategy 10,000 times calculating the frequencies of the event {Test statistic $>$ Theoretical 95% Critical value} based on measurements related to the status MI $= \tau$. In this evaluation we present situations when investigators do not know whether data correspond to healthy or diseased populations (i.e., F_0 of (1) is an unknown mixture distribution), and the biomarker has no discriminatory ability. The derived results are presented in Table 9. The parametric tests $PCUS^*$ and PSR^* were defined in Sec. 4 to present outputs of Table 5: design (a). (In many epidemiological studies,

Table 9
The bootstrap-type evaluation of the ability of cholesterol to discriminate groups of populations for myocardial infarction (MI)

	NPCUS	*NPSR*	*RDD*	*RAK*	RAD_1	$PCUS^*$	PSR^*
Type I error evaluation	0.058	0.055	0.051	0.056	0.054	0.151	0.163
$v-1=30$	0.703	0.742	0.541	0.598	0.515	0.339	0.464
$v-1=50$	0.489	0.475	0.366	0.448	0.404	0.239	0.318

cholesterol measurements are shown to be normally distributed; e.g., Vexler et al., 2008a,b.) It is clear; Table 9 does not recommend applying the parametric tests T9
to the study that evaluates the cholesterol biomarker related to atherosclerotic coronary heart disease.

5.2. *Power Evaluation*

In certain situations, to analyze objectively the discriminatory ability of a biomarker, investigators observe groups of individuals with $\mathrm{MI}=0$ and $\mathrm{MI}=1$ when groups sizes are unknown (i.e., we do not know when we finish observing measurements from population with $\mathrm{MI}=0$ and start to survey measurements from individuals with $\mathrm{MI}=1$). Note that, corresponding to the paragraph above, we consider the problem (1), when $F_0 \neq F_1$. We sampled first $v-1$ observations from the population with the status $\mathrm{MI}=0$ and $70-v+1$ observations from the population with $\mathrm{MI}=1$. We repeated this sampling 10,000 times, obtaining decisions of the tests, for $v-1=30$ and 50. Thus, we evaluated the powers of the tests. In accordance with Table 9, the cholesterol biomarker can clearly discriminate the populations. The proposed nonparametric tests *NPCUS* and *NPSR* can be highly recommended to be applied to different epidemiological studies that evaluate the cholesterol biomarker related to atherosclerotic coronary heart disease.

6. Conclusion

The main aims of this article were to review, develop, and compare various techniques applied to create decision rules for the retrospective change point detection issue. We concentrated on presenting general ideas regarding the change point tests' constructions. Thus, although we concern ourselves the relatively simple statement of the problem (1) with independent observations, in a similar manner to considerations mentioned in this article, complex models (including regressions, autoregressions, etc., see, e.g., Vexler, 2008) can be tested for different change points' occurrences in data distributions.

Commonly, the theoretical change point literature has introduced retrospective change point detection problems based on statements that are typical of modified sequential quality control issues. In this context, while considering (1), investigators have declared that $F_1 = F_0$. In this article, we pointed out that the retrospective statement of the change point problem cannot require assuming that $F_1 = F_0$, in a general context. This leads to extend forms of the known parametric and distribution free change point detection policies.

We examined various parametric and nonparametric tests for the change point problem (1), attending to different contexts of optimality and robustness of the procedures.

We showed that the recently developed (in the retrospective context) parametric Shiryayev–Roberts policy is appropriate to replace the classical CUSUM scheme in many practical applications, because the SR rule demonstrates the optimal property, efficiency, and robustness, compared with the CUSUM test. However, future studies are needed to investigate theoretically different sorts of SR-type tests.

We indicate that the parametric change point detection policies are very sensitive to the null distribution assumptions, in the context of a type I error control. Generally speaking, the parametric policies examined in Sec. 4 have constructions that are based on sums of independent random variables. (Section 4 evaluated the normally distributed data-based CUSUM procedure that is well addressed in the theoretical change point literature.) Thus, one would expect that, at least, these policies are robust when instead of assumed baseline normal distributions, t-distributed observations are in effect. We cannot confirm this property. This issue has not been well addressed in the literature, and hence future studies are required. Perhaps the problem of developing goodness-of-fit tests under the regime of (1) does not have simple solutions. We introduced the nonparametric procedure (19) that possesses robustness of validity, because it is based on ranks. A near-optimal property of (19) can be obtained for specific alternatives. However, in contrast to the parametric tests, because (19) utilizes rank statistics, the proposed nonparametric procedure has been shown to be a powerful change point detection policy when a guess related to the alternatives is incorrect. It is interesting to note that we observed that a small change in expected and assumed distribution forms can lead the parametric tests being helpless in the context of the type I error control, whereas in the same conditions the test (19) demonstrated powerful characteristics.

The proposed nonparametric tests are shown to be very efficient under various alternative hypotheses.

Section 3 of this article introduced the nonparametric methodology for approximating the likelihood ratios. This method can be applied to construct nonparametric estimators of the unknown change point parameter, provided that the null hypothesis of (1) is rejected. Toward this end, relevant parametric techniques (e.g., Borovkov, 1999; Gurevich and Vexler, 2005) can be approximated in the nonparametric manner. Our limited simulation results have shown that this approach is reasonable to investigate intensively and applied in practice.

Thus, we believe that the outputs of this manuscript have great potential to be applied in practice and induce investigators to study the retrospective change point issues.

Acknowledgments

We are grateful to the editor and two anonymous referees for their helpful comments that clearly improved this article. This work was partially supported by the Internal Funding Program of the Shamoon College of Engineering (SCE).

References

Borovkov, A. A. (1999). Asymptotically optimal solutions in the change-point problem. *Theory of Probability and Its Applications* 43:539–561.

Chernoff, H., Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics* 35:99–1018.

Csorgo, M., Horvath, L. (1988). Invariance principles for change-point problems. *Journal of Multivariate Analysis* 27:151–168.

Csorgo, M., Horvath, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: John Wiley.

Ferger, D. (1994). On the power of nonparametric changepoint-tests. *Metrika* 41:277–292.

Gombay, E. (2000). U-statistics for sequential change detection. *Metrika* 52:113–145.

Gombay, E. (2001). U-statistics for change under alternatives. *Journal of Multivariate Analysis* 78:139–158.

Gombay, E., Horvath, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and Their Applications* 50:161–171.

Gombay, E., Horvath, L. (1996). On the rate of approximations for maximum likelihood tests in change-point models. *Journal of Multivariate Analysis* 56:120–152.

Gordon, L., Pollak, M. (1995). A robust surveillance scheme for stochastically ordered alternatives. *Annals of Statistics* 23:1350–1375.

Gurevich, G. (2006). Nonparametric AMOC changepoint tests for stochastically ordered alternatives. *Communications in Statistics – Theory and Methods* 35:887–903.

Gurevich, G. (2007). Retrospective parametric tests for homogeneity of data. *Communications in Statistics – Theory and Methods* 36:2841–2862.

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

Gurevich, G., Vexler, A. (2006). Guaranteed maximum likelihood splitting tests of a linear regression model. *Statistics* 40:465–484.

James, B., James, K. L., Siegmund, D. (1987). Tests for a change-point. *Biometrika* 74:71–83.

Kander, Z., Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Annals of Mathematical Statistics* 37:1196–1210.

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

Lai, T. L. (1995). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society, Series B* 57:1–33.

Owen, A. B. (2001). *Empirical Likelihood*. Boca Raton, FL: Chapman and Hall/CRC.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Applied Statistics* 28:126–135.

Pollak, M. (1985). Optimal detection of a change in distribution. *Annals of Statistics* 13:206–227.

Pollak, M., Tartakovsky, A. G. (2009). On optimality properties of the Shiryaev–Roberts procedure. *Statistica Sinica*. AQ3

Sen, A., Srivastava, M. S. (1975). On tests for detecting change in mean. *Annals of Statistics* 3:98–108.

Tartakovsky, A. G., Polunchenko, A. S., Moustakides, G. V. (2009). Design and comparison of Shiryaev–Roberts and CUSUM-type change-point detection procedures. *Proceedings of the Second International Workshop on Sequential Methodologies (IWSM09).* AQ3

Vasicek, O. (1976). A test for normality based on sample entropy. *Journal of the Royal Statistical Society, Series B* 38:54–59.

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

Vexler, A. (2008). Martingale type statistics applied to change points detection. *Communications in Statistics – Theory and Methods* 37:1207–1224.

Vexler, A., Gurevich, G. (2009a). Average most powerful tests for a segmented regression. *Communications in Statistics – Theory and Methods* 38:2214–2231.

Vexler, A., Gurevich, G. (2009b). Entropy-based empirical likelihood ratio change point detection policies. *Communications in Statistics – Simulation and Computation*. AQ3

Vexler, A., Gurevich, G. (2010). Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. *Computational Statistics and Data Analysis* 54:531–545.

Vexler, A., Schisterman, E. F., Liu, A. (2008a). Estimation of ROC based on stably distributed biomarkers subject to measurement error and pooling mixtures. *Statistics in Medicine* 27:280–296.

Vexler, A., Liu, A., Eliseeva, E., Schisterman, E. F. (2008b). Maximum likelihood ratio tests for comparing the discriminatory ability of biomarkers subject to limit of detection. *Biometrics* 64: 895–903.

Vexler, A., Liu, S., Kang, L., Hutson, A. D. (2009a). Modifications of the empirical likelihood interval estimation with improved coverage probabilities. *Communications in Statistics – Simulation and Computation* 38:2171–2183.

Vexler, A., Wu, C., Liu, A., Whitcomb, B. W., Schisterman, E. F. (2009b). An extension of a change point problem. *Statistics* 43:213–225.

Wolfe, D. A., Schechtman, E. (1984). Nonparametric statistical procedures for the changepoint problem. *Journal of Statistical Planning and Inference* 9:389–396.

Zou, C., Liu, Y., Qin, P., Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics & Probability Letters* 77:374–382.