



Asymptotically optimal methods of change-point detection for composite hypotheses

Boris Brodsky^a, Boris Darkhovsky^{b,*}

^a*Higher School of Economics, State University, Moscow, Russia*

^b*Institute for Systems Analysis RAS, Prospekt 60-letya Oktyabrya, 9, Moscow 117312, Russia*

Received 9 August 2003; accepted 6 January 2004

Abstract

In this paper the problem of change-point detection for the case of composite hypotheses is considered. We assume that the distribution functions of observations before and after an unknown change-point belong to some parametric family. The true value of the parameter of this family is unknown but belongs to two disjoint sets for observations before and after the change-point, respectively. A new criterion for the quality of change-point detection is introduced. Modifications of generalized CUSUM and GRSh (Girshick–Rubin–Shiryayev) methods are considered and their characteristics are analyzed. Comparing these characteristics with an a priori boundary for the quality of change-point detection we establish asymptotic optimality of these methods when the family of distributions before the change-point consists of one element.

© 2004 Elsevier B.V. All rights reserved.

MSC: Primary 62C25; Secondary 62L15

Keywords: Change-point problem; Composite hypotheses

1. Introduction

The mathematical formulation of the change-point problem as a certain extremal problem was first considered in the report of Kolmogorov and Shiryayev at the 6th All-Union Conference on probability and mathematical statistics (Vilnius, 1960; see the reference comment in Shiryayev (1976)). The book of Shiryayev (1976) contains solutions of this problem both for discrete and continuous time in the Bayesian and extremal formulations for the case of the a priori known density function (d.f.) of a change-point. Earlier in works of Page (1954)

* Corresponding author. Tel.: +7-95-1354322; fax: +7-95-9382209.

E-mail address: darbor@isa.ru (B. Darkhovsky).

and Girshik and Rubin (1952) certain heuristical (for that time) methods for solving sequential change-point problems were proposed. Lorden (1971) found an asymptotically optimal method of sequential change-point detection which minimizes an average delay time in detection given an upper boundary for the average time before a “false alarm” without a priori assumptions about the d.f. of a change-point. Pollak (1985) proved that the method of Girshick and Rubin can be obtained as a certain limit from the method proposed by Shiryaev. He also demonstrated that this method is asymptotically optimal in the sense of Lorden’s criterion. Moustakides (1986) proved that Page’s CUSUM procedure is strictly optimal (not only asymptotically) in Lorden’s formulation of sequential change-point detection problem. For continuous time this result was recently obtained by Shiryaev (1996).

Lai (1995, 2000) and Lai and Shan (1999) considered the change-point problem in the more general situation of dependent random variables. In Lai (1995) it was shown that the Neyman–Pearson type procedure with the “moving window” of observations is asymptotically optimal. In Lai and Shan (1999) different boundaries for the “false alarm” probability were introduced instead of upper boundaries for the average time before the “false alarm” and the asymptotic optimality of Page’s CUSUM method was proved. Lai (1998, 2000) developed information-theoretic bounds for sequential multihypothesis testing and fault detection in stochastic systems.

Different modifications and generalizations of the CUSUM method can be found in Basseville and Nikiforov (1993) and Nikiforov (1995).

The case of composite hypotheses for the problem of sequential change-point detection is the most interesting for applications. Siegmund and Venkatraman (1995) considered a special variant of this problem when the mathematical expectation of observed Gaussian random variables changes from zero to δ or $-\delta$. For this problem, the asymptotic optimality in Lorden’s sense of the generalized likelihood ratio statistic was established. The general form of this statistic can be found in Lai (2001). The detailed review of the literature on this topic can be also found there.

To the best of our knowledge, the sequential change-point detection problem in the general context of composite hypotheses, when the d.f. of observations not only *after* the change-point but also *before* it is unknown and belongs to a certain family of distributions, was not considered in the literature. However, this problem arises in many applications, and in particular, in change-point detection problems for dynamical systems. In our paper, we consider this general situation of composite hypotheses.

In the sequel we introduce and analyze a *new criterion* of effectiveness of sequential change-point detection which is different from all known in literature. This criterion is based upon the a priori inequality proved by Brodsky and Darkhovsky (1990) (see also Brodsky and Darkhovsky, 2000). In our opinion, this criterion corresponds very well to the intuitive requirements of effectiveness of sequential change-point detection. In this respect it is no worse than other commonly used criteria in sequential change-point problems. At the same time, this criterion enables us to find asymptotically optimal methods of change-point detection in a general enough situation via the a priori low boundary analogous to the Rao–Cramer boundary in estimation problems.

In this paper, we have two major objectives:

- (1) to propose a new criterion of quality for sequential change-point detection methods.

- (2) To consider the general change-point detection problem with a change of one composite hypothesis to another composite hypothesis. It seems to us that this statement of change-point detection problem was not fully considered in the literature. We also think that the methodology of our approach based upon a priori inequalities in sequential change-point problems is ideologically close to Lai (1998, 2000) .

In Section 2, we explain our approach to the analysis of effectiveness of change-point detection methods and formulate the problem for composite hypotheses. In Section 3 modifications of the generalized CUSUM and Girshick–Rubin–Shiryayev (GRSh) methods are considered and a priori estimates of their quality with respect to (w.r.t.) a new criterion are established. This is the main result of our paper. From this result we conclude that the generalized CUSUM and GRSh methods are *asymptotically optimal* in the case when the family of distributions *before* a change-point consists of one element. The a priori estimate of the quality of change-point detection enables us to consider an effective detection procedure in the general situation for a stationary regime of data collection. In Section 4, we give the proof of the main result, in Section 5 some experimental results are presented.

2. A priori estimate for the quality of change-point detection. Formulation of the problem

First, let us introduce necessary notations and formulate main assumptions. Let $\theta = (\theta_0, \theta_1), \theta \in \Theta, \theta_0 \in \Theta_0, \theta_1 \in \Theta_1$, where Θ is a certain parametric set which belongs to some open set U in the finite dimensional space, $\Theta = \Theta_0 \cup \Theta_1, \Theta_0 \cap \Theta_1 = \emptyset$. We observe a sequence of independent random vectors $\{\zeta_k\}_{k=1}^\infty$ with the d.f. w.r.t. some σ -finite measure μ equal to $f(x, \theta_0), \theta_0 \in \Theta_0$ before an unknown change-point, and $f(x, \theta_1), \theta_1 \in \Theta_1$ —after this change-point. The d.f. f is known and defined for all parameter values from U . In what follows we denote by $\mathbf{P}_{m,\theta}(\mathbf{E}_{m,\theta})$ the measure (mathematical expectation) corresponding to a sequence $\{\zeta_k\}_{k=1}^\infty$ with the change-point at the instant m and the fixed value of the parameter $\theta = (\theta_0, \theta_1)$ (so the density function of observations ζ_n is equal to $f(x, \theta_0)$ if $n \leq m$ and $f(x, \theta_1)$ if $n > m$). Symbols $\mathbf{P}_{\infty,\theta}(\mathbf{E}_{\infty,\theta})$ correspond to an observed sequence without change-points.

We assume that the following conditions are satisfied:

- 2.1. Θ is a compact set;
- 2.2. $\mu\{x : f(x, \theta_1) \neq f(x, \theta_2)\} > 0$ if $\theta_1 \neq \theta_2$;
- 2.3. For any $\theta \in \Theta$, the value

$$I(\theta) \stackrel{\text{def}}{=} \int \ln \frac{f(x, \theta_1)}{f(x, \theta_0)} f(x, \theta_1) \mu(dx),$$

is defined and $\inf_{\theta \in \Theta} I(\theta) > 0$;

- 2.4. For any $\theta \in \Theta, \theta^* \in \Theta$ $\mathbf{P}_{\infty,\theta^*}$ -distribution of the random variable $\eta(\theta) = \ln(f(\cdot, \theta_1)/f(\cdot, \theta_0))$ satisfies the uniform (w.r. t. $\theta \in \Theta$) Cramer condition

$$\sup_{\theta \in \Theta} \mathbf{E}_{\infty,\theta^*} \exp\{t\eta(\theta)\} < \infty \quad \text{for } |t| < H(\theta^*),$$

where $\inf_{\theta^* \in \Theta} H(\theta^*) > 0$,

2.5. For any $\theta_0^* \in \Theta$ the function

$$\kappa(t, \theta, \theta_0^*) \stackrel{\text{def}}{=} \ln \int \left(\frac{f(x, \theta_1)}{f(x, \theta_0)} \right)^t f(x, \theta_0^*) \mu(dx),$$

has only two zeros: 0 and $t^*(\theta, \theta_0^*) > 0$ and $\inf_{\theta \in \Theta} t^*(\theta, \theta_0^*) > 0$.

Suppose a is a certain change-point detection method and $d^a(n)$ is its decision function such that $d^a(n) = 1$ ($d^a(n) = 0$) corresponds to the decision about the presence (absence) of a change at the instant n , $\tau^a \stackrel{\text{def}}{=} \min\{n : d^a(n) = 1\}$ is the stopping time w.r.t. the natural flow of σ -algebras generated by observations.

Brodsky and Darkhovskiy (2000) showed that all known methods of change-point detection contain a certain “large parameter” N such that the normalized by N delay time in change-point detection tends to some deterministic limit a.s. as $N \rightarrow \infty$ and the “false alarm” probability tends exponentially to zero as $N \rightarrow \infty$. Taking this into account, in what follows we add the index N to the values $d^a(n)$ and τ^a . In Section 3, we give precise definitions of the large parameter N and the decision functions $d_N^a(n)$ for the CUSUM and GRSh methods.

Denote the probability of the error decision by

$$\alpha_N^a(\theta) \stackrel{\text{def}}{=} \sup_n \mathbf{P}_{\infty, \theta} \{d_N^a(n) = 1\}.$$

For any fixed value of θ , a method of detection a and the large parameter N , let us consider the following value:

$$\mathcal{K}_N^a(\theta, m) = \frac{\mathbf{E}_{m, \theta}(\tau_N^a - m)^+}{|\ln \alpha_N^a(\theta)|}.$$

Brodsky and Darkhovskiy (2000) showed that for all known change-point detection methods, there exists the following limit not depending on m

$$\lim_{N \rightarrow \infty} \mathcal{K}_N^a(\theta, m) \stackrel{\text{def}}{=} \mathcal{K}^a(\theta) \tag{2.1}$$

and for any known method of detection a the following a priori inequality takes place

$$\mathcal{K}^a(\theta) \geq I^{-1}(\theta). \tag{2.2}$$

Brodsky and Darkhovskiy (2000) proved that (2.2) follows from the fact that for any m the value $(\tau_N^a - m)^+ / N$ tends $\mathbf{P}_{m, \theta}$ -a.s. as $N \rightarrow \infty$ to a certain deterministic limit (not depending on m), and the value $\alpha_N^a(\theta)$ tends to zero exponentially as $N \rightarrow \infty$. Moreover, Brodsky and Darkhovskiy (2000) proved that inequality (2.2) holds (for all known methods) in the general situation of dependent observations.

From inequality (2.2) it follows that the natural characteristic of quality for any method of change-point detection a (for any fixed θ) can be represented by the value $\mathcal{K}^a(\theta)$. The quality criterion of change-point detection then consists in minimization of this value. The sense of this criterion is as follows. It is easy to conclude (see also Darkhovskiy and Brodsky, 1987) that the average time before a “false alarm” $\mathbf{E}_{\infty, \theta} \tau_N^a$ has the asymptotic order (as $N \rightarrow \infty$) of $\{\alpha_N^a(\theta)\}^{-1}$. Therefore, the characteristic $\mathcal{K}^a(\theta)$ represents the limit ratio (in

an appropriate scale) of the average delay time in change-point detection to the average time before a “false alarm”. Evidently, this ratio is of the main interest for any change-point detection problem and we should try to make it as small as possible. A priori low estimate (2.2) for this characteristic enables us to call a method a^* asymptotically optimal (for a given θ) if the strict equality is attained for this method in inequality (2.2). This situation is essentially the same as in Rao–Cramer inequality which provides the low boundary for the quality of any estimate of an unknown parameter.

In our opinion, this approach to the definition of an optimal method of change-point detection strictly corresponds to the practical sense of this problem. Moreover, in many cases it is even more convenient than conventional definitions of change-point detection optimality criteria when a method is found which minimizes the average delay time in detection given an upper boundary for the average time before a “false alarm”.

For new methods of detection which we consider in this paper, it is a priori unknown whether the limit in (2.1) exists or not. However, for any method of change-point detection b with the large parameter N , we can always consider the value

$$\limsup_{N \rightarrow \infty} \mathcal{K}_N^b(\theta, m) \stackrel{\text{def}}{=} \bar{\mathcal{K}}^b(\theta, m).$$

We will show later that

$$\bar{\mathcal{K}}^b(\theta, m) \geq I^{-1}(\theta) \tag{2.3}$$

for any method of detection with a large parameter, and therefore the value $\bar{\mathcal{K}}^b(\theta, m)$ can represent the qualitative characteristic of change-point detection for any such method of detection.

Now we can formulate our problem: we wish to find such a method of change-point detection for which inequality (2.3) turns into a strict equality for any (unknown!) value of the parameter θ .

3. Main result

In the sequel we consider methods based upon two classic statistics: Page’s cumulative sums (CUSUM) and the quasi–Bayesian statistic of Girshick–Rubin–Shiryaev (GRSh).

Let us recall these stopping rules for any fixed value of θ . Define

$$\mathcal{R}_n(\theta) \stackrel{\text{def}}{=} \sum_{k=1}^n \prod_{i=k}^n \left\{ \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} \right\}, \quad \mathcal{L}_n(\theta) \stackrel{\text{def}}{=} \max_{1 \leq k \leq n} \sum_{i=k}^n \ln \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}.$$

Then the CUSUM rule takes the form

$$\tau^{\text{CS}}(\theta) = \inf\{n \geq 0 : \mathcal{L}_n(\theta) \geq C_{\text{CS}}\}$$

and the GRSh rule is

$$\tau^{\text{GRSh}}(\theta) \stackrel{\text{def}}{=} \{n : \mathcal{R}_n(\theta) \geq C_{\text{GRSh}}\},$$

where $C_{\text{CS}}, C_{\text{GRSh}}$ are certain boundaries (thresholds of detection).

Brodsky and Darkhovsky (2000) demonstrated that the large parameter N of the CUSUM method can be chosen as $N = C_{CS}$ and for the GRSh method as $N = \ln C_{GRSh}$, and both of these methods are asymptotically optimal and asymptotically equivalent to each other, i.e. (for any fixed θ)

$$\mathcal{K}^{CS}(\theta) = \mathcal{K}^{GRSh}(\theta) = I^{-1}(\theta). \tag{3.1}$$

Let $\Theta^p \subset \Theta$ be a fixed finite subset of $\#\Theta = p$. Let us consider the following stopping times

$$\begin{aligned} T^{ACS,p} &\stackrel{\text{def}}{=} \min \left\{ n : \max_{\theta \in \Theta^p} \mathcal{L}_n(\theta) \geq C_{CS} \right\}, \\ T^{AGRSh,p} &\stackrel{\text{def}}{=} \min \left\{ n : \max_{\theta \in \Theta^p} \mathcal{R}_n(\theta) \geq C_{GRSh} \right\} \end{aligned} \tag{3.2}$$

and let us further denote methods of detection based upon these stopping times as (ACS, p) and (AGRSh, p), respectively.

The large parameter N of the method (ACS, p) is equal to C_{CS} and for the method (AGRSh, p) $N = \ln C_{GRSh}$. The values $d_N^{ACS,p}(n)$, $d_N^{AGRSh,p}(n)$, $\alpha_N^{ACS,p}(\theta)$, $\alpha_N^{AGRSh,p}(\theta)$ are defined as usual for these methods.

The quality characteristics of the methods (ACS, p) and (AGRSh, p) are $\bar{\mathcal{K}}^{ACS,p}(\theta, m)$ and $\bar{\mathcal{K}}^{AGRSh,p}(\theta, m)$, respectively.

Now consider a sequence of expanding finite subsets Θ^p of the set Θ with an increasing number p of elements and such that each Θ^p is a finite $1/p$ -net for Θ as $p \rightarrow \infty$. This is possible because of a compactness of set Θ . Then a sequence of stopping times $T^{ACS,p}$, $T^{AGRSh,p}$ is monotonously decreasing and therefore for any ω there exist the limits

$$\begin{aligned} T^{ACS} &\stackrel{\text{def}}{=} \lim_{p \rightarrow \infty} T^{ACS,p}, \\ T^{AGRSh} &\stackrel{\text{def}}{=} \lim_{p \rightarrow \infty} T^{AGRSh,p}. \end{aligned}$$

The decision rule T^{ACS} is called *the adaptive CUSUM rule* and the rule T^{AGRSh} —*the adaptive GRSh rule*. We use the term “adaptive”, because (as we show it later) these decision rules are asymptotically optimal for any (unknown!) value of the parameter θ .

Methods of detection based upon these rules will be called *the adaptive CUSUM method* (ACS) and *the adaptive GRSh method* (AGRSh).

Since the adaptive CUSUM and GRSh methods are introduced by means of the monotonous limit procedure using growing numbers of the finite $1/p$ -net, it is natural to estimate their quality by the following characteristics:

$$\begin{aligned} \bar{\mathcal{K}}^{ACS}(\theta, m) &\stackrel{\text{def}}{=} \limsup_{p \rightarrow \infty} \bar{\mathcal{K}}^{ACS,p}(\theta, m), \\ \bar{\mathcal{K}}^{AGRSh}(\theta, m) &\stackrel{\text{def}}{=} \limsup_{p \rightarrow \infty} \bar{\mathcal{K}}^{AGRSh,p}(\theta, m). \end{aligned}$$

In the sequel we denote by $\theta^* = (\theta_0^*, \theta_1^*) \in \Theta$ a true (unknown) parametric point in the change-point problem.

Our main result in this paper is the following

Theorem 3.1. *Suppose conditions (2.1)–(2.5) are satisfied for any value of the parameter $\theta^* \in \Theta$. Then for the adaptive CUSUM and GRSh methods the following estimates hold true:*

$$\begin{aligned}
 I^{-1}(\theta^*) &\leq \bar{\mathcal{K}}^{\text{ACS}}(\theta^*, m) \leq I^{-1}(\theta^*) \frac{1}{\inf_{\theta \in \Theta} t^*(\theta, \theta_0^*)}, \\
 I^{-1}(\theta^*) &\leq \bar{\mathcal{K}}^{\text{AGRSh}}(\theta^*, m) \leq I^{-1}(\theta^*) \frac{1}{\inf_{\theta \in \Theta} t^*(\theta, \theta_0^*)}
 \end{aligned}
 \tag{3.3}$$

Corollary 3.1. *Suppose the set Θ_0 consists of a unique point θ_0^* . Then the adaptive CUSUM and GRSh methods are asymptotically optimal.*

Proof. By its definition, the function $t^*(\theta, \theta_0^*)$ depends only from the component θ_0^* of the pair $\theta^* = (\theta_0^*, \theta_1^*)$. That means if $\Theta_0 = \{\theta_0^*\}$ then

$$t^*(\cdot) = t^*(\theta_0^*, \theta_1, \theta_0^*).$$

So in order to determine the point t^* , it is necessary to find the roots of the equation

$$\ln \int \left(\frac{f(x, \theta_1)}{f(x, \theta_0^*)} \right)^t f(x, \theta_0^*) \mu(dx) = 0. \tag{3.4}$$

But (taking into account assumption 2.2) for $0 < t < 1$ by virtue of Hölder’s inequality,

$$\begin{aligned}
 &\ln \left(\int f^t(x, \theta_1) f^{1-t}(x, \theta_0^*) \mu(dx) \right) \\
 &< \left(t \ln \int f(x, \theta_1) \mu(dx) + (1 - t) \ln \int f(x, \theta_0^*) \mu(dx) \right) = 0
 \end{aligned}$$

and for $t > 1$ for any $s < t$ by virtue of Lyapunov inequality

$$\left(\int f^t(x, \theta_1) f^{1-t}(x, \theta_0^*) \mu(dx) \right)^{1/t} > \left(\int f^s(x, \theta_1) f^{1-s}(x, \theta_0^*) \mu(dx) \right)^{1/s}.$$

Put here $s = 1$ we obtain

$$\ln \left(\int (f^t(x, \theta_1) f^{1-t}(x, \theta_0^*)) \mu(dx) \right)^{1/t} > 0.$$

Therefore, a unique non-zero root of Eq. (3.4) is equal to 1, i.e.

$$t^*(\theta_0^*, \theta_1, \theta_0^*) \equiv 1. \tag{3.5}$$

Now the asymptotic optimality follows from (3.5) and (3.3). \square

If we suppose that the density $f(x, \theta)$ is smooth enough w.r.t. the parameter θ , then the function $t^*(\cdot, \theta^*)$ is continuous. In this case if the diameter of the set Θ_0 is small enough, then the right-hand side of (3.3) is also small and the considered stopping rules are close to the asymptotically optimal by virtue of Corollary 3.1. Hence we conclude that in rules (3.2) it is expedient to take the maximum by θ_0 on the set

$$\Theta_0(n) \stackrel{\text{def}}{=} \{\theta_0 \in \Theta_0 : \|\theta_0 - \hat{\theta}_0(n)\| \leq 3\sqrt{\text{Sp}(D_n)}\}, \tag{3.6}$$

where $\hat{\theta}_0(n)$ is any asymptotically normal estimate of the true parameter θ_0^* by the first n observations and D_n is the covariance matrix of this estimate. In other words, we need to use the following stopping rule for the CUSUM

$$\tilde{\tau}^{\text{ACS}} = \inf \left\{ n \geq 0 : \sup_{\theta_0 \in \Theta_0(n)} \sup_{\theta_1 \in \Theta_1} \mathcal{L}_n(\theta) \geq C_{\text{CS}} \right\} \tag{3.7}$$

and the analogous rule for the GRSh method.

In fact, if we consider the quasi-stationary process of data collection when any change-point is preceded by a long stationary period sometimes interrupted by “false alarms”, then a true parametric point will be covered by the set $\Theta_0(n)$ with the probability close to 1, and therefore such modified stopping rules will be close to the asymptotically optimal rule.

4. Proofs

We limit ourselves to the analysis of the adaptive CUSUM, since the analysis of the adaptive GRSh is analogous.

Suppose that a true (unknown) parametric point is $\theta^* = (\theta_0^*, \theta_1^*)$.

Let b be an arbitrary change-point detection method which depends on the large parameter N , T_N is the corresponding stopping time, and $d_N(n)$ is the decision function of this method.

Lemma 4.1. *For any method b with the large parameter N and any fixed m , the following inequalities hold:*

$$(a) \quad \frac{\mathbf{E}_{m, \theta^*}(T_N - m)^+}{|\ln(N\alpha_N(\theta^*))|} \geq I^{-1}(\theta^*), \tag{4.1a}$$

(b) if $\liminf_{N \rightarrow \infty} (|\ln \alpha_N(\theta^*)|/N) > 0$ then

$$\tilde{\mathcal{K}}^b(\theta^*, m) \stackrel{\text{def}}{=} \liminf_{N \rightarrow \infty} \frac{\mathbf{E}_{m, \theta^*}(T_N - m)^+}{|\ln \alpha_N(\theta^*)|} \geq I^{-1}(\theta^*). \tag{4.1b}$$

Proof. Let

$$\tilde{\alpha}_N(\theta^*) \stackrel{\text{def}}{=} \sup_{k \leq N} \mathbf{P}_{\infty, \theta^*}\{d_N(k) = 1\},$$

$$\tilde{\beta}_N(\theta^*) \stackrel{\text{def}}{=} \sup_{k \leq N} \mathbf{P}_{\infty, \theta^*}\{T_N = k\},$$

$$\tilde{T}_N \stackrel{\text{def}}{=} m + (T_N - m)^+ \wedge N.$$

Then for any fixed m

$$\begin{aligned}
 N\tilde{\beta}_N(\theta^*) &\geq \sum_{k=m+1}^{N+m} \mathbf{P}_{\infty, \theta^*}(T_N = k), \\
 &\sum_{k=m+1}^{N+m} \mathbf{E}_{m, \theta^*} \left\{ \mathbb{I}(T_N = k) \exp \left(- \sum_{i=m+1}^k \ln(f(x_i, \theta_1^*)/f(x_i, \theta_0^*)) \right) \right\} \\
 &= \mathbf{E}_{m, \theta^*} \left\{ \exp \left(- \sum_{i=m+1}^{\tilde{T}_N} \ln(f(x_i, \theta_1^*)/f(x_i, \theta_0^*)) \right) \right\} \\
 &\geq \exp \left\{ - \mathbf{E}_{m, \theta^*} \left(\sum_{i=m+1}^{\tilde{T}_N} \ln(f(x_i, \theta_1^*)/f(x_i, \theta_0^*)) \right) \right\}.
 \end{aligned}$$

Since $\mathbf{E}_{m, \theta^*} \tilde{T}_N < \infty$, due to Wald’s inequality we obtain

$$\mathbf{E}_{m, \theta^*} \left(\sum_{i=m+1}^{\tilde{T}_N} \ln(f(x_i, \theta_1^*)/f(x_i, \theta_0^*)) \right) = I(\theta^*) \mathbf{E}_{m, \theta^*}(\tilde{T}_N - m). \tag{4.2}$$

Now taking into account that $(T_N = k) \subset (d_N(k) = 1)$ and so

$$\tilde{\beta}_N(\theta^*) \leq \tilde{\alpha}_N(\theta^*)$$

and using (4.2), we obtain

$$\frac{\mathbf{E}_{m, \theta^*}(\tilde{T}_N - m)^+}{|\ln(N\tilde{\alpha}_N(\theta^*))|} \geq I^{-1}(\theta^*). \tag{4.3}$$

Since

$$\frac{\mathbf{E}_{m, \theta^*}(T_N - m)}{|\ln \alpha_N(\theta^*)|} \geq \frac{\mathbf{E}_{m, \theta^*}(\tilde{T}_N - m)^+}{|\ln \tilde{\alpha}_N(\theta^*)|},$$

we get from here and (4.3) relation (4.1a) and in case $\liminf_{N \rightarrow \infty} (|\ln \alpha_N(\theta^*)|)/N > 0$ we get (4.1b). \square

Lemma 4.2. *The following relation holds*

$$T^{\text{ACS}, p} = \min_{\theta \in \Theta^p} \tau^{\text{CS}}(\theta). \tag{4.4}$$

Proof. By definition, for any $\theta \in \Theta^p$ $T^{\text{ACS}, p} \leq \tau^{\text{CS}}(\theta)$ and therefore

$$T_p^{\text{ACS}} \leq \min_{\theta \in \Theta^p} \tau^{\text{CS}}(\theta). \tag{4.5}$$

From the other side, if $T^{\text{ACS},p} = n$ then $\max_{\theta \in \Theta^p} \mathcal{L}_n(\theta) \geq C_{\text{CS}}$ and so there exists an element $\theta^{(i)} \in \Theta^p$ such that $\mathcal{L}_n(\theta^{(i)}) \geq C_{\text{CS}}$. Therefore, for any elementary event ω there exists a number $1 \leq i(\omega) \leq p$ such that $T^{\text{ACS},p}(\omega) \geq \tau^{\text{CS}}(\theta^{(i(\omega))})$ and so

$$T^{\text{ACS},p} \geq \min_{\theta \in \Theta^p} \tau^{\text{CS}}(\theta). \tag{4.6}$$

The conclusion of the lemma follows from (4.5) and (4.6). \square

For any fixed $\theta \in \Theta$ put

$$y_n(\theta) \stackrel{\text{def}}{=} \ln \frac{f(x_n, \theta_1)}{f(x_n, \theta_0)}, \quad S_k(\theta) \stackrel{\text{def}}{=} \sum_{i=1}^k y_i(\theta).$$

For Page decision rule under fixed parameter θ we will use a designation $d_N^{\text{CS}}(n, \theta)$ (remember that the “large parameter” N in CUSUM is the threshold C_{CS}).

Then we have

$$\mathbf{P}_{\infty, \theta^*} \{d_N^{\text{CS}}(n, \theta) = 1\} = \mathbf{P}_{\infty, \theta^*} \left(\max_{k \leq n} S_k(\theta) \geq C_{\text{CS}} \right). \tag{4.7}$$

It follows from the fact that the observable sequence $\{y_n(\theta)\}$ is the sequence of i.i.d. under measure $\mathbf{P}_{\infty, \theta^*}$ and that is why we can get (4.7) using time transformation. Similarly

$$\sup_n \mathbf{P}_{\infty, \theta^*} \{d_N^{\text{ACS},p}(n) = 1\} = \sup_n \mathbf{P}_{\infty, \theta^*} \left\{ \max_{\theta \in \Theta^p} \max_{k \leq n} S_k(\theta) \geq C_{\text{CS}} \right\}. \tag{4.8}$$

Taking into account that

$$\left\{ \omega : \max_{k \leq n} \max_{\theta \in \Theta^p} S_k(\theta) \geq C_{\text{CS}} \right\} \subseteq \left\{ \omega : \max_{k \leq n+1} \max_{\theta \in \Theta^p} S_k(\theta) \geq C_{\text{CS}} \right\},$$

we obtain from (4.8)

$$\begin{aligned} \alpha_N^{\text{ACS},p}(\theta^*) &\stackrel{\text{def}}{=} \sup_n \mathbf{P}_{\infty, \theta^*} \{d_N^{\text{ACS},p}(n) = 1\} = \mathbf{P}_{\infty, \theta^*} \left\{ \sup_k \max_{\theta \in \Theta^p} S_k(\theta) \geq C_{\text{CS}} \right\}, \\ \alpha_N^{\text{CS}}(\theta, \theta^*) &\stackrel{\text{def}}{=} \sup_n \mathbf{P}_{\infty, \theta^*} \{d_N^{\text{CS}}(n, \theta) = 1\} = \mathbf{P}_{\infty, \theta^*} \left\{ \sup_k S_k(\theta) \geq C_{\text{CS}} \right\}. \end{aligned} \tag{4.9}$$

Lemma 4.3. *Under conditions (2.2), (2.4) and (2.5) the following relations hold:*

$$\lim_{N \rightarrow \infty} \frac{|\ln \alpha_N^{\text{ACS},p}(\theta^*)|}{N} \geq \inf_{\theta \in \Theta} t^*(\theta, \theta_0^*), \quad \lim_{N \rightarrow \infty} \frac{|\ln \alpha_N^{\text{CS}}(\theta, \theta^*)|}{N} = t^*(\theta, \theta_0^*).$$

Proof. The second equality was proved in Brodsky and Darkhovsky (2000, p. 261). Consider the first inequality. We obtain

$$\begin{aligned}
 & \mathbf{P}_{\infty, \theta^*} \left\{ \sup_k \max_{\theta \in \Theta^p} S_k(\theta) \geq N \right\} \\
 & \leq \mathbf{P}_{\infty, \theta^*} \left\{ \sum_{k=1}^{\infty} \sum_{\theta \in \Theta^p} \exp(S_k(\theta)) \geq \exp(N) \right\} \\
 & \leq \sum_{\theta^i \in \Theta^p} \mathbf{P}_{\infty, \theta^*} \left\{ \sum_{k=1}^{\infty} \exp(S_k(\theta^i)) \geq p^{-1} \exp(N) \right\} \\
 & \leq p \max_{\theta^i \in \Theta^p} \mathbf{P}_{\infty, \theta^*} \left\{ \sum_{k=1}^{\infty} \exp(S_k(\theta^i)) \geq p^{-1} \exp(N) \right\}. \tag{4.10}
 \end{aligned}$$

Brodsky and Darkhovsky (2000, p. 257) proved that the following relationship holds for given conditions (2.2), (2.4) and (2.5)

$$\lim_{u \rightarrow \infty} \frac{|\ln \mathbf{P}_{\infty, \theta^*} \{ \sum_{k=1}^{\infty} \exp(S_k(\theta)) > \exp(u) \}|}{u} = t^*(\theta, \theta_0^*). \tag{4.11}$$

Moreover, convergence in (4.11) is uniform w.r.t. $\theta \in \Theta$ due to condition (2.4). Taking this into account from (4.9)–(4.11) we obtain

$$\begin{aligned}
 & \lim_{N \rightarrow \infty} \frac{|\ln \alpha_N^{\text{ACS}, p}(\theta^*)|}{N} \\
 & \geq \lim_{N \rightarrow \infty} \min_{\theta^i \in \Theta^p} N^{-1} \left| \ln \mathbf{P}_{\infty, \theta^*} \left\{ \sum_{k=1}^{\infty} \exp(S_k(\theta^i)) \geq p^{-1} \exp(N) \right\} \right| \\
 & = \min_{\theta^i \in \Theta^p} \lim_{N \rightarrow \infty} N^{-1} \left| \ln \mathbf{P}_{\infty, \theta^*} \left\{ \sum_{k=1}^{\infty} \exp(S_k(\theta^i)) \geq p^{-1} \exp(N) \right\} \right| \\
 & = \min_{\theta^i \in \Theta^p} t^*(\theta, \theta_0^*) \geq \inf_{\theta \in \Theta} t^*(\theta, \theta_0^*) > 0. \tag{4.12}
 \end{aligned}$$

The lemma is proved. \square

Proof of the Theorem. From (4.4) it follows that for any m

$$\mathbf{E}_{m, \theta^*} (T^{\text{ACS}, p} - m)^+ \leq \mathbf{E}_{m, \theta^*} (\tau^{\text{CS}}(\theta^*) - m)^+.$$

Hence we obtain

$$\bar{\mathcal{K}}^{\text{ACS}, p}(\theta^*, m) \leq \lim_{N \rightarrow \infty} \frac{(N^{-1} \mathbf{E}_{m, \theta^*} (\tau^{\text{CS}}(\theta^*) - m)^+) (N^{-1} |\ln \alpha_N^{\text{CS}}(\theta^*, \theta^*)|)}{(N^{-1} |\ln \alpha_N^{\text{ACS}, p}(\theta^*)|) (N^{-1} |\ln \alpha_N^{\text{CS}}(\theta^*, \theta^*)|)}. \tag{4.13}$$

By virtue of (3.1)

$$\lim_{N \rightarrow \infty} \frac{(N^{-1} \mathbf{E}_{m, \theta^*} (\tau^{\text{CS}}(\theta^*) - m)^+)}{(N^{-1} |\ln \alpha_N^{\text{CS}}(\theta^*, \theta^*)|)} = \mathcal{K}^{\text{CS}}(\theta^*) = I^{-1}(\theta^*). \tag{4.14}$$

Table 1
(CUSUM $\theta_0 = 0.0$)

C _{CS}		4	5	6	7	Low bound	Upper bound
$\theta_0 = 0$	ET	185.8	436.1	1095.6	2939.1		
	$\ln ET$	5.22	6.08	7.00	7.99		
$\theta_1 = 1$	$E\tau$	7.47	9.22	11.30	12.98		
	$E\tau/\ln ET$	1.43	1.51	1.61	1.54	2.0	2.0
$\theta_1 = 2$	$E\tau$	2.58	3.06	3.53	4.06		
	$E\tau/\ln ET$	0.49	0.51	0.50	0.50	0.50	0.50
$\theta_1 = 3$	$E\tau$	1.51	1.69	1.90	2.11		
	$E\tau/\ln ET$	0.28	0.27	0.27	0.26	0.22	0.22

Table 2
(CUSUM $\theta_0 = 0.5$)

C _{CS}		9	12	15	20	Low bound	Upper bound
$\theta_0 = 0.5$	ET	101.0	106.6	249.5	428.4		
	$\ln ET$	4.62	5.08	5.52	6.06		
$\theta_1 = 1$	$E\tau$	17.32	23.30	29.34	39.47		
	$E\tau/\ln ET$	3.74	4.58	5.31	6.51	8.0	48.0
$\theta_1 = 2$	$E\tau$	5.06	6.54	7.99	10.54		
	$E\tau/\ln ET$	1.09	1.28	1.44	1.73	0.88	5.28
$\theta_1 = 3$	$E\tau$	2.53	3.24	3.92	5.01		
	$E\tau/\ln ET$	0.54	0.63	0.71	0.82	0.32	1.92

Taking into account (3.5) and the second equality in Lemma 4.3 we obtain

$$\lim_{N \rightarrow \infty} \frac{|\ln \alpha_N^{CS}(\theta^*, \theta^*)|}{N} \geq t^*(\theta^*, \theta_0^*) \equiv 1. \tag{4.15}$$

Therefore from (4.12) to (4.15) it follows that

$$\bar{\mathcal{K}}^{ACS,p}(\theta^*, m) \leq \frac{I^{-1}(\theta^*)}{\inf_{\theta \in \Theta} t^*(\theta, \theta_0^*)}. \quad \square \tag{4.16}$$

Estimate (4.16) holds for any stopping time $T^{ACS,p}$ and for any finite set Θ^P . Moreover, this estimate does not depend on Θ^P . Estimate (4.1b) holds for any method with a large parameter and condition (b) in Lemma 4.1, and in particular for the method (ACS, p) due

Table 3
(CUSUM: $\theta_0 = 0$)

C_{CS}		4	5	6	7	Low bound	Upper bound
$\theta_0 = 0$	ET	198.5	482.8	1184.7	3172.7		
	$\ln ET$	5.29	6.18	7.08	8.06		
$\theta_1 = 1$	$E\tau$	7.44	9.06	11.22	13.16		
	$E\tau/\ln ET$	1.41	1.47	1.58	1.63	2.0	2.0
$\theta_1 = 2$	$E\tau$	2.63	3.13	3.43	4.06		
	$E\tau/\ln ET$	0.49	0.50	0.47	0.50	0.50	0.50
$\theta_1 = 3$	$E\tau$	1.51	1.77	1.92	2.13		
	$E\tau/\ln ET$	0.28	0.29	0.27	0.26	0.22	0.22

Table 4
(CUSUM: $\theta_0 = 0.5$)

C_{CS}		4	5	6	7	Low bound	Upper bound
$\theta_0 = 0.5$	ET	189.7	450.0	1132.7	2815.5		
	$\ln ET$	5.25	6.11	7.03	7.94		
$\theta_1 = 1$	$E\tau$	20.79	28.94	35.53	46.19		
	$E\tau/\ln ET$	3.96	4.73	5.05	5.81	8.0	8.20
$\theta_1 = 2$	$E\tau$	4.10	4.94	5.74	6.53		
	$E\tau/\ln ET$	0.78	0.80	0.81	0.82	0.88	0.93
$\theta_1 = 3$	$E\tau$	1.93	1.99	2.55	2.89		
	$E\tau/\ln ET$	0.36	0.35	0.36	0.35	0.32	0.35

to (4.12). For this method it also does not depend on Θ^P . Therefore, taking into account the definition of $\bar{\mathcal{N}}^{ACS}(\theta^*, m)$, we obtain the required result from (4.1b) and (4.16).

5. Experimental results

In this section, we report some results of a small simulation study we performed in order to assess the accuracy of the lower and upper boundaries of a priori inequalities (3.3).

Data: The following data were analyzed. The Gaussian sequence was simulated with the d.f. $N(\theta, 1)$. Under the null hypothesis $H_0: \theta = \theta_0 \in [0, 1]$, under the alternative hypothesis $H_1: \theta = \theta_1 \in [1.2, 5]$.

Methods: CUSUM and GRSh methods of change-point detection were analyzed.

Table 5
(GRSh: $\theta_0 = 0$)

C_{GRSh}		100	300	500	1000	Low bound	Upper bound
$\theta_0 = 0$	ET	169.2	457.7	784.0	1421.6		
	$\ln ET$	5.13	6.13	6.66	7.26		
$\theta_1 = 1$	$E\tau$	7.09	9.28	10.27	11.37		
	$E\tau/\ln ET$	1.38	1.51	1.54	1.56	2.0	2.0
$\theta_1 = 2$	$E\tau$	2.72	3.26	3.45	3.76		
	$E\tau/\ln ET$	0.53	0.53	0.51	0.51	0.50	0.50
$\theta_1 = 3$	$E\tau$	1.63	1.89	1.98	2.12		
	$E\tau/\ln ET$	0.31	0.30	0.29	0.29	0.22	0.22

Table 6
(GRSh: $\theta_0 = 0.5$)

C_{GRSh}		300	500	1000	2000	Low bound	Upper bound
$\theta_0 = 0.5$	ET	299.0	529.4	1065.4	1977.4		
	$\ln ET$	5.70	6.27	6.97	7.59		
$\theta_1 = 1$	$E\tau$	22.96	27.95	32.65	37.95		
	$E\tau/\ln ET$	4.02	4.45	4.68	5.0	8.0	8.2
$\theta_1 = 2$	$E\tau$	5.14	5.45	5.97	6.68		
	$E\tau/\ln ET$	0.90	0.87	0.85	0.88	0.88	0.93
$\theta_1 = 3$	$E\tau$	2.40	2.64	2.75	2.94		
	$E\tau/\ln ET$	0.42	0.42	0.39	0.38	0.32	0.35

Results: (1) In the first test the adaptive CUSUM method (3.2) was used for detection of a change-point. We have taken different values of h_0 and h_1 . Results are reported in Tables 1 and 2. Each value in cells of these tables is obtained as the average of 5000 Monte Carlo trials. Here ET is the average time between “false alarms” (which is asymptotically equal to $1/p$, where p is the probability of the error decision, see Brodsky and Darkhovsky, 1987); $\ln ET$ is the logarithm of ET ; $E\tau$ is the average delay time in change-point detection; $E\tau/\ln ET$ is the ratio criterion of efficiency used in inequalities (3.3).

Results presented in Table 1 show that inequalities (3.3) are sharp in spite of both composite hypotheses H_0 and H_1 .

Table 2 shows that the adaptive CUSUM method is not asymptotically optimal in general due to the composite hypothesis H_0 .

(2) In the second series of experiments the methods CUSUM and GRSh with estimation of the parameter θ_0 before a change-point were analyzed. The estimate $\hat{\theta}_0$ was obtained as

the average from the first 500 observations. In this case the change-point $m = 500$. Results are reported in Tables 3–6.

Tables 3–6 show that the heuristic procedures (3.7) are close to asymptotically optimal rules.

These experimental results allow us to make the following conclusions. In some cases the adaptive CUSUM and GRSh rules are asymptotically optimal in spite of the fact that the set Θ_0 consists of more than one point. This indicates that estimates (3.3) are *sharp*. However, in general these adaptive rules are *not asymptotically optimal*.

Let us emphasize that efficiency of sequential change-point detection methods with estimation of the d.f. parameters before a change-point is essentially higher than the corresponding efficiency of methods without estimation of θ_0 . This is true for the case of composite hypotheses H_0 and H_1 .

Acknowledgements

The authors are grateful to the referee for many valuable comments.

Note added in Proof

Right before proofreading the authors have found the asymptotically optimal methods for the general case of composite hypotheses H_0 and H_1 , and in particular for the general multipoint set Θ_0 . One of these methods is as follows:

$$T \stackrel{\text{def}}{=} \min \left\{ n : \inf_{\theta_0 \in \Theta_0} \sup_{\theta_1 \in \Theta_1} \mathcal{L}_n(\theta) \geq C \right\}$$

We are going to consider the problem of change-point detection for composite hypotheses in a separate paper.

References

- Basseville, M., Nikiforov, I.V., 1993. Detection of Abrupt Changes: Theory and Applications. Prentice-Hall, Englewood Cliffs, NJ.
- Brodsky, B.E., Darkhovsky, B.S., 1990. Comparative study of some nonparametric methods of sequential change-point detection. *Teor. Veroyatnost. Primenen.* 35 (4), 655–668.
- Brodsky, B.E., Darkhovsky, B.S., 2000. Non-Parametric Statistical Diagnosis. Problems and Methods. Kluwer, Dordrecht.
- Darkhovsky, B.S., Brodsky, B.E., 1987. Nonparametric method of the quickest detection of a change in mean of a random sequence. *Teor. Veroyatnost. Primenen.* 32 (4), 703–711.
- Girshik, M.A., Rubin, H., 1952. A Bayes approach to a quality control model. *Ann. Math. Statist.* 23 (1), 114–125.
- Lai, T.L., 1995. Sequential change-point detection in quality control and dynamical systems (with discussion). *J. Roy. Statist. Ser. B* 57, 613–658.
- Lai, T.L., 1998. Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory* 44, 2917–2929.

- Lai, T.L., 2000. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *IEEE Trans. Inform. Theory* 46, 595–608.
- Lai, T.L., 2001. Sequential analysis: some classical problems and new challenges. *Statist. Sinica* 11 (2), 303–408.
- Lai, T.L., Shan, J.Z., 1999. Efficient recursive algorithms for detection of abrupt changes in signals and systems. *IEEE Trans. Automat. Control* 44, 952–966.
- Lorden, G., 1971. Procedures for reacting to a change in distribution. *Ann. Math. Statist.* 42, 1897–1908.
- Moustakides, G.V., 1986. Optimal stopping for detecting changes in distribution. *Ann. Statist.* 14, 1379–1387.
- Nikiforov, I.V., 1995. A generalized change detection problem. *IEEE Trans. Inform. Theory* 41, 171–181.
- Page, E.S., 1954. Continuous inspection schemes. *Biometrika* 41, 100–114.
- Pollak, M., 1985. Optimal detection of a change in distribution. *Ann. Statist.* 13, 206–227.
- Shiryayev, A.N., 1976. *Optimal Stopping Rules*. Nauka, Moscow.
- Shiryayev, A.N., 1996. Minimax optimality of the CUSUM method for the continuous time. *Adv. Math. Sci.* 310 (4), 173–174.
- Siegmund, D., Venkatraman, E.S., 1995. Using the generalized likelihood ratio statistics for sequential detection of a change-point. *Ann. Statist.* 23, 255–271.