ELSEVIER

Stochastic Processes and their Applications 51 (1994) 359–372

stochastic processes and their applications

# Nonparametric detection of changepoints for sequentially observed data

D. Ferger

*Mathematical Institute, Justus-Liebig University, Arndtstraße 2, D-35392 Giessen, Germany*

Received 31 March 1992; revised 31 January 1993

**Abstract**

Assume that independent data $X_1^n, \ldots, X_{k(n)}^n$ are observed sequentially in time, where $k(n) < \infty$ is a finite horizon. Suppose also that there exists $\theta \in (0,1]$ such that $X_1^n, \ldots, X_{[k(n)\theta]}^n$ have distribution $\nu_{1,n}$ and $X_{[k(n)\theta]+1}^n, \ldots, X_{k(n)}^n$ have distribution $\nu_{2,n}$. The distributions and the changepoint $\theta$ are unknown. Our aim is to react as soon as possible after the change has taken place. We propose a nonparametric stopping rule which attains a given probability of "false alarm" on the one hand and, on the other hand, is less than or equal to $k(n)\theta + O(\sqrt{k(n)})$ with probability one.

*Key words:* Sequential detection of a changepoint; Weak convergence of two-parameter stochastic processes; Martingale maximal-inequalities

## 0. Introduction and main results

We consider a triangular array $X_1^n, \ldots, X_{k(n)}^n$, $k(n) \in \mathbb{N}$, of independent random variables defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$, with values in a measurable space $(\mathscr{X}, \mathscr{B})$. Suppose there is a $\theta \in (0,1]$, such that $X_1^n, \ldots, X_{[k(n)\theta]}^n$ have distribution $\nu_{1,n}$ and $X_{[k(n)\theta]+1}^n, \ldots, X_{k(n)}^n$ have distribution $\nu_{2,n} \neq \nu_{1,n}$. The distributions as well as the changepoint $\theta$ are not known. Assume that the random variables $X_i^n$, $i = 1, \ldots, k(n)$, are observed sequentially. We are looking for a stopping rule satisfying:

(a) the probability to stop before $[k(n)\theta]$ (false alarm) should be controlled,

(b) the procedure should react as soon as possible after the changepoint has occurred.

In Bhattacharya and Frierson (1981) an example is given touching a typical problem of quality control: a machine produces items and as long as the production is correct, we do not want to interrupt the process. The machine is assumed to be adjusted at regular times. Within two successive adjustments we therefore take random samples. Denote by $X_i^n$ the value of the $i$th random sample after the $n$th adjustment. Then our model describes the situation in a suitable manner.


*SSDI* 0304-4149(93)E0056-K

Problems of fastest detection of the point where the stochastic mechanism of a random process changes have been extensively studied in literature. Many contributions to that subject require a certain amount of prior informations on the underlying distributions, e.g. when using the so-called Shewhart (1931) control charts one must be sure that the observed data come from a normal population. In opposite to this and other parametric approaches Page's (1954) cumulative sums (CUSUM) method is applicable in all cases where certain expectations have different signs. Lorden (1971) even proved optimality of the CUSUM procedure. However, the optimal method involves the densities of the data before the change and thereafter. But in practise typically such specific knowledge is not available and even some parametric assumptions cannot be justified. In these situations of minimal prior information one must rely on nonparametric methods. For instance, Bhattacharya and Frierson (1981) initiated the use of sequential ranks for detecting small disorders in a random sample that comes from a continuous distribution. Csörgő and Horváth (1987) introduced the weighted empirical distribution functions pertaining to these sequential ranks. Based on them they propose a sequential procedure for detecting a possible changepoint in a random sequence of continuous observations. There the probability of false alarm is fixed, whereas in the case of a change the process is stopped with probability one in a specified length of time. Recently, Brodskii and Darhovskii (1991) made a comparative analysis of several nonparametric methods embracing the CUSUM procedure of Page (1954), the Girshick–Rubin–Shiryaev algorithm (cf. Girshick and Rubin (1952), Shiryaev (1973)), the Shewhart (1931) method, the method of Darhovskii and Brodskii (1987) and the so-called exponential smoothing method. Here it should be mentioned that they replace the usual assumptions of independence by a weaker strong mixing condition.

Recall our example at the beginning. If the sampling costs are low, we dispose of a large number $k(n)$ of observations per row. Hence assume that $k(n) \to \infty$, as $n \to \infty$. Without loss of generality (w.l.o.g.) we may consider the case $k(n) = n$. We propose the following class of stopping times:

$$\tau_n \, \tau_n(c) = \inf\left\{ na \le k \le n\colon k^{-3/2} \max_{1 \le l \le k-1} \left| \sum_{i=l+1}^{k} \sum_{j=1}^{l} K(X_i^n, X_j^n) \right| > c \right\}$$

if $\{\,\cdots\,\} \neq \emptyset$ and $\tau_n = n + 1$ otherwise.

Here, $K\colon \mathscr{X} \times \mathscr{X} \to \mathbb{R}$ is a $\mathscr{B} \otimes \mathscr{B}$-measurable mapping (kernel), that is bounded and antisymmetric. The quantity $a \in (0, \theta)$ is assumed to be known and $c$ is a positive constant, which will be determined later. Concerning the quantity $a$, note that in our example, after each adjustment, it is realistic to assume that the production process is correct at least a certain number of time units. The motivation for the procedure is quite obvious: until time $[na]$ we know that no change has taken place. But after that time we start testing the hypothesis, whether our data come from the same distribution or not. We use a test statistic of Ferger (1991). As long as the test does not reject the hypothesis, it is reasonable to go on sampling. In Theorem 2.2 below we determine the critical value $c = c(\alpha)$ such that the error of false alarm, for each $\theta \in [a, 1]$, does not exceed a given value $0 < \alpha < 1$ at least asymptotically, i.e.

$$\lim_{n\to\infty} \mathbb{P}_\theta(\tau_n(c) \le [n\theta]) \le \alpha \quad \forall \theta \in [a, 1].$$

As already been mentioned, we also want to stop as soon as possible after the change has taken place. As to this, we will show that (cf. Theorem 2.4)

$$\lim_{n\to\infty} \mathbb{P}_\theta(\tau_n \le n\theta + C_0\sqrt{n}) = 1 \quad \forall \theta \in [a, 1],$$

where $C_0$ is a positive constant. Finally, we will prove that (cf. Theorem 2.3)

$$\lim_{n\to\infty} n^{-1}\mathbb{E}_1(\tau_n) \ge 1 - \alpha(1-a).$$

This means, in case there is no change, we will in the mean stop the process actually very late. In chapter three a simulation study will be reported on in which our theoretical results will be made clear visually.

Set

$$\xi_n(u,t) := [nu]^{-3/2} \sum_{i=[[nu]t]+1}^{[nu]} \sum_{j=1}^{[[nu]t]} K(X_i^n, X_j^n)$$

for $(u,t) \in [a,1] \times [0,1]$ and put

$$S_n(u) := \sup_{0\le t\le 1} |\xi_n(u,t)| \quad \text{for } u \in [a,1].$$

Then $S_n$ is a step function on $[a, 1]$ with jumps at the points $u = kn^{-1}$, $na \le k \le n$, and corresponding values $S_n(kn^{-1})$. Observe that

$$S_n\left(\frac{k}{n}\right) = \sup_{0\le t\le 1} \left| k^{-3/2} \sum_{i=[kt]+1}^{k} \sum_{j=1}^{[kt]} K(X_i^n, X_j^n)\right|$$

$$= k^{-3/2} \max_{1\le l\le k-1} \left| \sum_{i=l+1}^{k} \sum_{j=1}^{l} K(X_i^n, X_j^n)\right|.$$

We see that

$$\{\tau_n \le [n\theta]\} = \{S_n \text{ exceeds the boundary } c \text{ in } [a,\theta]\}$$

$$= \Omega \setminus \left\{ \sup_{a\le u\le \theta} \sup_{0\le t\le 1} |\xi_n(u,t)| \le c \right\},$$

so that the probability of false alarm is closely related to a boundary-crossing probability of $S_n$. We shall first show that the $\xi_n$ as random elements in the space $D([a,\theta]\times[0,1])$ converge in distribution.

## 1. An invariance principle for the stochastic processes $\xi_n$

Introduce the stochastic process

$$\hat{\xi}_n(u,t) := \sum_{l=1}^{[nu]} \mathbb{E}[\xi_n(u,t) \mid X_l^n] = \sum_{l=1}^{[nu]} \eta_{n,l}(u,t), \quad a \le u \le \theta,\ 0 \le t \le 1,$$

where

$$\eta_{n,l}(u,t) = \begin{cases} [nu]^{-3/2}([nu] - [[nu]t]) R_n \circ X_l^n, & 1 \le l \le [[nu]t], \\ -[nu]^{-3/2}[[nu]t] R_n \circ X_l^n, & [[nu]t] < l \le [nu] \end{cases}$$

and

$$R_n(y) = \int K(x,y) \nu_{1,n}(\mathrm{d}x).$$

Now, as a first step, we show that the two processes $\xi_n$ and $\hat{\xi}_n$ are asymptotically stochastically equivalent.

**Lemma 1.1.** *If $0 < a \le \theta$, then*

$$\lim_{n\to\infty} \mathbb{P}_\theta\left[ \sup_{a \le u \le \theta} \sup_{0 \le t \le 1} |\xi_n(u,t) - \hat{\xi}_n(u,t)| > \varepsilon \right] = 0 \quad \forall \varepsilon > 0.$$

Before proving Lemma 1.1, we provide two maximal inequalities.

**Lemma 1.2.** *Let $Z_1, \ldots, Z_n$ be i.i.d. random variables with values in a measurable space $(\mathscr{X}, \mathscr{B})$ and distribution $\nu$. Let $H$ be a kernel satisfying*:

$$h(y) = \int H(x,y)\nu(\mathrm{d}x) = 0 \quad \textit{for } \nu\textit{-almost all } y \in \mathscr{X}, \tag{1.1}$$

$$\tilde{h}(x) = \int H(x,y)\nu(\mathrm{d}y) = 0 \quad \textit{for } \nu\textit{-almost all } x \in \mathscr{X}, \tag{1.2}$$

$$\sigma^2 = \mathbb{E} H^2(Z_1, Z_2) < \infty. \tag{1.3}$$

*Then for all $\varepsilon > 0$*

$$\mathbb{P}\left[ \max_{1 \le k \le n} \max_{0 \le l \le k} \left| \sum_{l+1 \le j < i \le k} H(Z_i, Z_j) \right| > \varepsilon \right] \le 2\sigma^2 \varepsilon^{-2} n^2.$$

**Proof.** Denote

$$U_{k,l} = \sum_{l+1 \le j < i \le k} H(Z_i, Z_j).$$

Then

$$\mathbb{P}\left[ \max_{1 \le k \le n} \max_{0 \le l \le k} \left| \sum_{l+1 \le j < i \le k} H(Z_i, Z_j) \right| > \varepsilon \right] = \mathbb{P}\left( \max_{1 \le k \le n} S_k > \varepsilon^2 \right),$$

where $S_k = \max_{0 \le l \le k} U_{k,l}^2$. Now, put $\mathscr{F}_k = \sigma(Z_1, \ldots, Z_k)$, $1 \le k \le n$ and $\mathscr{F}_0 := \{\emptyset, \Omega\}$.

Observe that using (1.1) we obtain

$$\mathbb{E}[U_{k+1,l}|\mathscr{F}_k] = U_{k,l} \quad \forall 0 \le l \le k.$$

So, by Jensen's inequality

$$\mathbb{E}[S_{k+1}|\mathscr{F}_k] \ge \mathbb{E}[U^2_{k+1,l}|\mathscr{F}_k] \ge \mathbb{E}^2[U_{k+1,l}|\mathscr{F}_k] = U^2_{k,l} \quad \forall 0 \le l \le k,$$

which implies that

$$S_k = \max_{0 \le l \le k} U^2_{k,l} \le \mathbb{E}[S_{k+1}|\mathscr{F}_k].$$

In other words, $(S_k, \mathscr{F}_k)_{0 \le k \le n}$ is a sub-martingale. From Doob's inequality (cf. Chow and Teicher (1978), Theorem 8, p. 243]) we infer that

$$\mathbb{P}\left(\max_{1 \le k \le n} S_k > \varepsilon^2\right) \le \varepsilon^{-2}\,\mathbb{E}(S_n) = \varepsilon^{-2}\,\mathbb{E}\left(\max_{0 \le l \le n-1} |U_{n,l}|^2\right).$$

Use (1.2) to prove that $(U_{n,l}, \mathscr{G}_l)_{l=n-1,\dots,0}$ with $\mathscr{G}_l = \sigma(Z_{l+1}, \dots, Z_n)$ is a martingale (in reverse time). Hence $(|U_{n,l}|, \mathscr{G}_l)_{l=n-1,\dots,0}$ is a nonnegative submartingale. Thus Doob ensures

$$\mathbb{E}\left(\max_{0 \le l \le n-1} |U_{n,1}|^2\right) \le 4\mathbb{E}|U_{n,0}|^2 = 4 \sum_{1 \le j < i \le n} \mathbb{E}H^2(Z_i, Z_j) \le 2n\sigma^2$$

by (1.1)–(1.3). This proves the lemma. □

The following lemma has already been proved by Ferger (1994) Lemma 3.1.

**Lemma 1.3.** *Under the assumptions of Lemma* 1.2, *for all* $\varepsilon > 0$

$$\mathbb{P}\left[\sup_{1 \le l \le n}\left|\sum_{1 \le j < i \le l} H(Z_i, Z_j)\right| > \varepsilon\right] \le \tfrac{1}{2}\sigma^2\varepsilon^{-2}n^2.$$

**Proof of Lemma 1.1.** We have

$$\xi_n(u,t) - \hat{\xi}_n(u,t) = [nu]^{-3/2} \sum_{i=[[nu]t]+1}^{[nu]} \sum_{j=1}^{[[nu]t]} [K(X^n_i, X^n_j) - R_n \circ X^n_j + R_n \circ X^n_i],$$

where $R_n(y) = \int K(x,y)\nu_{1,n}(\mathrm{d}x)$. So, if we put $H_n(x,y) = K(x,y) - R_n(y) + R_n(x)$, we can conclude that

$$\begin{aligned}
&\sup_{a \le u \le \theta} \sup_{0 \le t \le 1} |\xi_n(u,t) - \hat{\xi}_n(u,t)| \\
&= \sup_{a \le u \le \theta} \sup_{0 \le t \le 1} [nu]^{-3/2}\left|\sum_{i=[[nu]t]+1}^{[nu]} \sum_{j=1}^{[[nu]t]} H_n(X^n_i, X^n_j)\right| \\
&= \max_{na \le k \le n\theta} \max_{0 \le l \le k} k^{-3/2}\left|\sum_{i=l+1}^{k} \sum_{j=1}^{l} H_n(X^n_i, X^n_j)\right|
\end{aligned}$$

$$\leq a^{-3/2} n^{-3/2} \max_{na \leq k \leq n\theta} \max_{0 \leq l \leq k} \left| \sum_{i=l+1}^{k} \sum_{j=1}^{l} H_n(X_i^n, X_j^n) \right|.$$

Because of

$$\sum_{i=l+1}^{k} \sum_{j=1}^{l} H_n(X_i^n, X_j^n) = \sum_{1 \leq j < i \leq k} H_n(X_i^n, X_j^n) - \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) - \sum_{l+1 \leq j < i \leq k} H_n(X_i^n, X_j^n)$$

we obtain with $C_0 = a^{-3/2}$ that

$$\sup_{a \leq u \leq \theta} \sup_{0 \leq t \leq 1} |\zeta_n(u, t) - \hat{\zeta}_n(u, t)|$$

$$\leq C_0 n^{-3/2} \max_{1 \leq k \leq n} \left| \sum_{1 \leq j < i \leq k} H_n(X_i^n, X_j^n) \right|$$

$$+ C_0 n^{-3/2} \max_{1 \leq k \leq n} \max_{0 \leq l \leq k} \left| \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) \right|$$

$$+ C_0 n^{-3/2} \max_{1 \leq k \leq n} \max_{0 \leq l \leq k} \left| \sum_{l+1 \leq j < i \leq k} H_n(X_i^n, X_j^n) \right|$$

$$\leq 2C_0 n^{-3/2} \max_{1 \leq l \leq n} \left| \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) \right|$$

$$+ C_0 n^{-3/2} \max_{1 \leq k \leq n} \max_{0 \leq l \leq k} \left| \sum_{l+1 \leq j < i \leq k} H_n(X_i^n, X_j^n) \right|,$$

since

$$\max_{1 \leq k \leq n} \max_{0 \leq l \leq k} \left| \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) \right| = \max_{0 \leq l \leq n} \left| \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) \right|.$$

Now, by Lemma 1.3, it follows that

$$\mathbb{P}\left[ 2C_0 n^{-3/2} \max_{1 \leq l \leq n} \left| \sum_{1 \leq j < i \leq l} H_n(X_i^n, X_j^n) \right| > \tfrac{1}{2}\varepsilon \right] \leq C_1 \varepsilon^{-2} n^{-1},$$

where $C_1$ is a constant not depending on $n$. Moreover, an application of Lemma 1.2 yields

$$\mathbb{P}\left[ C_0 n^{-3/2} \max_{1 \leq k \leq n} \max_{0 \leq l \leq k} \left| \sum_{l+1 \leq j < i \leq k} H_n(X_i^n, X_j^n) \right| > \tfrac{1}{2}\varepsilon \right] \leq C_2 \varepsilon^{-2} n^{-1},$$

where $C_2$ is a constant not depending on $n$. This proves the lemma. □

**Lemma 1.4.** *The sequence of processes* $(\hat{\xi}_n)_{n \geq 1}$ *is asymptotically C-tight, i.e. for all* $\varepsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}\left[ \sup_{\substack{|u-u'|, |t-t'| < \delta \\ a \leq u, u' \leq \theta, 0 \leq t, t' \leq 1}} \left| \hat{\xi}_n(u, t) - \hat{\xi}_n(u', t') \right| \geq \varepsilon \right] = 0 \tag{1.4}$$

*and*

$$\hat{\xi}_n(a, 0) \text{ is bounded in probability.} \tag{1.5}$$

**Proof.** Since $\hat{\xi}_n(a, 0) = 0 \; \forall \omega \in \Omega \; \forall n \geq 1$, (1.5) is trivially fulfilled. It remains to show (1.4). Set

$$v_q = q\delta, \quad 0 \leq q \leq \delta^{-1} \text{ (where w.l.o.g. } \delta^{-1} \in \mathbb{N})$$

and

$$u_r = a + r\delta, \; 0 \leq r \leq [(\theta - a)\delta^{-1}] \text{ and } u_{[(\theta-a)\delta^{-1}]+1} = \theta.$$

Moreover, put

$$s \equiv s_k \equiv s_{k,r}^{(m)} = u_{r-1} + k\delta m^{-1}, \quad 0 \leq k \leq m$$

$$t \equiv t_l \equiv t_{l,q}^{(m)} = v_{q-1} + l\delta m^{-1}, \quad 0 \leq l \leq m.$$

Following Billingsley's (1968) arguments on p. 56 in the one-dimensional case one can prove that it is enough to show: $\forall \varepsilon, \eta > 0 \;\; \exists 0 < \delta < 1$ such that $\forall 1 \leq r \leq [(\theta - a)\delta^{-1}] + 1$ and $\forall 1 \leq q \leq \delta^{-1} \; \exists n_0 = n_0(r, q, \delta), m_0 = m_0(r, q, \delta) \in \mathbb{N}$ so that

$$\mathbb{P}\left[ \sup_{0 \leq k, l \leq m} |\hat{\xi}_n(s_{k,r}^{(m)}, t_{l,q}^{(m)}) - \hat{\xi}_n(u_{r-1}, v_{q-1})| > \varepsilon \right] \leq \delta^2 \eta \quad \forall m \geq m_0 \;\; \forall n \geq n_0. \tag{1.6}$$

Now, we fix $r$ and $q$ and note that

$$|\hat{\xi}_n(s, t) - \hat{\xi}_n(u, v)| = \left| \sum_{i=1}^{[ns]} \eta_{n,i}(s, t) - \sum_{i=1}^{[nu]} \eta_{n,i}(u, v) \right|$$

$$\leq \left| \sum_{i=1}^{[ns]} [\eta_{n,i}(s, t) - \eta_{n,i}(u, v)] \right| + \left| \sum_{i=[nu]+1}^{[ns]} \eta_{n,i}(u, v) \right|.$$

It follows, that the left-hand side of (1.6) is less than or equal to

$$\mathbb{P}\left[ \sup_{0 \leq k, l \leq m} \left| \sum_{i=1}^{[ns_k]} [\eta_{n,i}(s_k, t_l) - \eta_{n,i}(u, v)] \right| > \tfrac{1}{2}\varepsilon \right]$$

$$+ \mathbb{P}\left[ \sup_{0 \leq k \leq m} \left| \sum_{i=[nu]+1}^{[ns_k]} \eta_{n,i}(u, v) \right| > \tfrac{1}{2}\varepsilon \right]. \tag{1.7}$$

The first summand in (1.7) is less than or equal to

$$\mathbb{P}\left[\sup_{0\le k,l\le m}\left|\sum_{i=1}^{[[nu]v]}[\eta_{n,i}(s_k,t_l)-\eta_{n,i}(u,v)]\right|>\tfrac{1}{6}\varepsilon\right]$$

$$+\mathbb{P}\left[\sup_{0\le k,l\le m}\left|\sum_{i=[[nu]v]+1}^{[[ns_k]t_l]}[\eta_{n,i}(s_k,t_l)-\eta_{n,i}(u,v)]\right|>\tfrac{1}{6}\varepsilon\right]$$

$$+\mathbb{P}\left[\sup_{0\le k,l\le m}\left|\sum_{i=[[ns_k]t_l]+1}^{[ns_k]}[\eta_{n,i}(s_k,t_l)-\eta_{n,i}(u,v)]\right|>\tfrac{1}{6}\varepsilon\right]. \tag{1.8}$$

Notice that

$$[\eta_{n,i}(s_k,t_l)-\eta_{n,i}(u,v)]=\{[ns_k]^{-3/2}([ns_k]-[[ns_k]t_l])$$
$$-[nu]^{-3/2}([nu]-[[nu]v])\}\,R_n\circ X_i^n.$$

Furthermore, one can show upon applying the Mean Value Theorem that $|\{\cdots\}|\le C_0n^{-1/2}\delta\ \forall 0\le k,l\le m$, if $n\ge n_0(\delta)$, where $C_0$ is a constant. Thus the first summand in (1.8) is less than or equal to

$$\mathbb{P}\left[C_0n^{-1/2}\delta\left|\sum_{i=1}^{[[nu]v]}R_n\circ X_i^n\right|>\tfrac{1}{6}\varepsilon\right]\le\delta^2\eta,$$

by an application of Hoeffding's (1963) inequality. Similarly we obtain that the third summand in (1.8) is less than or equal to

$$\mathbb{P}\left[C_1n^{-1/2}\delta\sup_{0\le k,l\le m}\left|\sum_{i=[[ns_k]t_l]}^{[ns_k]}R_n\circ X_i^n\right|>\tfrac{1}{6}\varepsilon\right]$$

$$\le\mathbb{P}\left[2C_1n^{-1/2}\delta\sup_{1\le k\le n}\left|\sum_{i=1}^{k}R_n\circ X_i^n\right|>\tfrac{1}{6}\varepsilon\right]\le\delta^2\eta,$$

utilizing Lévy's inequality [cf. Billingsley (1968), p. 69] and then Hoeffding's (1963) inequality. Since,

$$\eta_{n,i}(s_k,t_l)-\eta_{n,i}(u,v)$$
$$=\{[ns_k]^{-3/2}([ns_k]-[[ns_k]t_l]+[nu]^{-3/2}[[nu]v]\}\,R_n\circ X_i^n$$

and $|\{\cdots\}|\le C_2n^{-1/2}\ \forall 0\le k,\ l\le m$, the second summand of (1.8) is less than or equal to

$$\mathbb{P}\left[C_2n^{-1/2}\sup_{0\le k,l\le m}\left|\sum_{i=[[nu]v]+1}^{[[ns_k]t_l]}R_n\circ X_i^n\right|>\tfrac{1}{6}\varepsilon\right]$$

$$\le\mathbb{P}\left[\sup_{[[nu_{r-1}]v_{q-1}]<k\le[[nu_r]v_q]}\left|\sum_{i=[[nu_{r-1}]v]+1}^{k}R_n\circ X_i^n\right|>\tfrac{1}{6}\varepsilon n^{-1}C_2^{-1}\right]\le\delta^2\eta,$$

as a consequence of Lévy's and Hoeffding's inequality, respectively. The second summand of (1.7) is less than or equal to

$$\mathbb{P}\left[ C_3 n^{-1/2} \sup_{[nu_{r-1}]<k\le[nu_r]} \left| \sum_{i=[nu_{r-1}]+1}^{k} R_n \circ X_i^n \right| > \tfrac{1}{2}\varepsilon \right] \le \delta^2 \eta$$

upon arguing as above. The lemma is proved. □

**Lemma 1.5.** *If $K$ is a bounded, antisymmetric kernel and $s^2(n) = \int [\int K(x,y) \nu_{1,n}(\mathrm{d}x)]^2 \nu_{1,n}(\mathrm{d}y)$ converges to some quantity $\sigma^2$, then the finite dimensional distributions of $\hat{\xi}_n$ converge to those of a centered Gaussian process $\xi = \{\xi(u,t)$: $a \le u \le \theta$, $0 \le t \le 1\}$ with covariance function*

$$\operatorname{Cov}[\xi(u_1,t_1),\xi(u_2,t_2)] = \sigma^2(u_1u_2)^{-1/2}[u_1t_1(1-t_1)(1-t_2) - (u_2t_2-u_1t_1)t_1(1-t_2) + (u_1-u_2t_2)t_1t_2]. \quad (1.9)$$

*whenever $u_1 \le u_2$, $t_1 \le t_2$ and $t_2 \le u_1u_2^{-1}$. If $t_2 > u_1u_2^{-1}$ we obtain a similar expression.*

The proof of this lemma is a routine application of the CLT for triangular arrays in $\mathbb{R}^p$, $p \in \mathbb{N}$. Observe that for each fixed $u \in [a,\theta]$, $\operatorname{Cov}(\xi(u,t_1),\xi(u,t_2)) = \sigma^2 t_1(1-t_2)$. It follows that $\{\sigma^{-1}\xi(u,t): 0 \le t \le 1\}$ is a Brownian Bridge for each fixed $u$. Now, the following theorem is a consequence of the Lemmata 1.1, 1.4 and 1.5 [cf. Neuhaus (1971)].

**Theorem 1.6.** *Let $(\mathscr{X},\mathscr{B})$ be a measurable space. Assume that $K$ is a bounded, antisymmetric kernel. Also, let $s^2(n)$ converge to some $\sigma^2$ as $n \to \infty$. Then $\xi_n$ converges in distribution of $\xi$, where $\xi$ is a centered Gaussian process with a.s. continuous paths and covariance-function* (1.9).

## 2. The reacting procedure and its asymptotic properties

Notice that, up to the (in general unknown) constant $\sigma^2$, the distribution of the limit process $\xi$ does not depend on $(\nu_{1,n})_{n\in\mathbb{N}}$ and $K$. For this reason we estimate $\sigma^2$. Put

$$h(x,y,z) = K(x,z)K(y,z)$$

and

$$h_s(x,y,z) = \frac{1}{3!}\sum_{(u,v,w)\in P_3} h(u,v,w),$$

where $P_3$ is the set of all permutations of $(x,y,z)$. Define

$$\sigma_n^2 = \sigma_n^2(a) = \begin{bmatrix} [na] \\ 3 \end{bmatrix}^{-1} \sum_{1\le i<j<l\le na} h_s(X_i^n, X_j^n, X_l^n).$$

Following the proof of Lemma 2.13 of Ferger (1991) we obtain the following Lemma.

**Lemma 2.1.** *Let $\mathscr{X}$ be a separable metric space and let K be a bounded and continuous kernel. If $v_{1,n}$ converges weakly to some $v_1$, then with probability one*

$$\sigma_n^2 = \sigma_n^2(a) \to \sigma^2 = \int [\int K(x,y) v_1(\mathrm{d}x)]^2 v_1(\mathrm{d}y) \text{ as } n \to \infty .$$

Now, by Lemma 2.1, Theorem 1.6 and a Cramér–Slutzky argument it follows that the processes $\zeta_n = \sigma_n^{-1} \xi_n$ converge in distribution to a process $\zeta$, whose distribution is invariant with respect to the underlying distributions $(v_{1,n})_{n \in \mathbb{N}}$ and the kernel $K$. Therefore we consider the modified stopping rule

$$\tilde{\tau}_n := \inf \left\{ na \le k \le n \colon \sigma_n^{-1} k^{-3/2} \max_{1 \le l \le k-1} \left| \sum_{i=l+1}^{k} \sum_{j=1}^{l} K(X_i^n, X_j^n) \right| > c \right\}.$$

**Theorem 2.2.** *Let $\mathscr{X}$ be a separable metric space, K a bounded, continuous and anti-symmetric kernel and $c = c(\alpha)$ the $(1-\alpha)$-quantile of*

$$\sup_{a \le u \le 1} \sup_{0 \le t \le 1} |\zeta(u,t)|.$$

*If $v_{1,n}$ converges weakly to some $v_1$, then*

$$\lim_{n \to \infty} \mathbb{P}_\theta(\tilde{\tau}_n \le [n\theta]) \le \alpha \quad \forall a \le \theta \le 1.$$

**Proof.** Similar as before we have

$$\mathbb{P}_\theta(\tilde{\tau}_n \le [n\theta]) = 1 - \mathbb{P}\left( \sup_{0 \le u \le \theta} \sup_{0 \le t \le 1} |\zeta_n(u,t)| \le c \right).$$

From the Continuous Mapping Theorem we can conclude

$$\sup_{a \le u \le \theta} \sup_{0 \le t \le 1} |\zeta_n(u,t)| \xrightarrow{\mathscr{L}} \sup_{a \le u \le \theta} \sup_{0 \le t \le 1} |\zeta(u,t)| \quad \forall a \le \theta \le 1.$$

By the Corollary to Theorem 2 of Lifshits (1982) it follows that

$$\lim_{n \to \infty} \mathbb{P}_\theta(\tilde{\tau}_n \le [n\theta]) = 1 - \mathbb{P}\left( \sup_{a \le u \le \theta} \sup_{0 \le t \le 1} |\zeta(u,t)| \le c \right)$$

$$\le 1 - \mathbb{P}\left( \sup_{a \le u \le 1} \sup_{0 \le t \le 1} |\zeta(u,t)| \le c \right) = \alpha. \quad \square$$

Recall that the distribution of $\sup_{a \le u \le 1} \sup_{0 \le t \le 1} |\zeta(u,t)|$ does not depend on $(v_{1,n})_{n \in \mathbb{N}}$ and $K$, so we can estimate $c$ via Monte-Carlo simulation. Moreover, since

$$\sup_{0 \le t \le 1} |\zeta(u,t)| \stackrel{\mathscr{L}}{=} \sup_{0 \le t \le 1} |B^\circ(t)|$$

for each fixed $u \in [a,1]$, where $B^\circ$ denotes a Brownian Bridge, we can conclude that $c(\alpha) \ge x(1-\alpha)$ where $x(1-\alpha)$ is the $(1-\alpha)$-quantile of the Kolmogorov–Smirnov distribution. It is not hard to see that $c(\alpha) = c(\alpha, a)$ converges to $x(1-\alpha)$ as $a$ tends to one. So, if $a$ is close to one, $c(\alpha)$ is approximated by $x(1-\alpha)$.

Next, we investigate $\tilde{\tau}_n$ if no change has taken place, i.e if $\theta$ equals 1.

**Theorem 2.3.** *Under the assumptions of Theorem* 2.2

$$\lim_{n\to\infty} \mathbb{E}_1(n^{-1}\tilde{\tau}_n) \geq 1 - \alpha(1-a).$$

**Proof.** We have

$$\mathbb{E}_1(n^{-1}\tilde{\tau}_n) = \int_0^{(n+1)/n} \mathbb{P}_1(\tilde{\tau}_n > nx)\,\mathrm{d}x = a + \int_a^1 \mathbb{P}_1(\tilde{\tau}_n > nx)\,\mathrm{d}x + \int_1^{(n+1)/n} \mathbb{P}_1(\tilde{\tau}_n > nx)\,\mathrm{d}x$$

$$\geq a + \int_a^1 \mathbb{P}_1(\tilde{\tau}_n > n)\,\mathrm{d}x = a + (1-a)\,\mathbb{P}_1\Big(\sup_{a\leq u\leq 1}\ \sup_{0\leq t\leq 1} |\zeta_n(u,t)| \leq c(\alpha)\Big)$$

$$\to a + (1-a)(1-\alpha) = 1 - \alpha(1-a) \quad \text{as } n\to\infty. \qquad \square$$

**Theorem 2.4.** *Let $\mathscr{X}$ be a separable metric space, $K \in C^b(\mathscr{X}^2)$ antisymmetric and suppose that $\nu_{i,n}$ converges weakly to $\nu_i$ for $i = 1,2$. Moreover assume that $\lambda = \int\int K(x,y)\nu_1(\mathrm{d}y)\nu_2(\mathrm{d}x) \neq 0$. Then*

$$\lim_{n\to\infty} \mathbb{P}_\theta[\tilde{\tau}_n \leq n\theta + 2c\theta^{-1}\sigma_n|\lambda|^{-1}\sqrt{n}] = 1 \quad \forall\theta\in[a,1].$$

**Proof.** Set $l_n = 2c\theta^{-1}\sigma_n|\lambda|^{-1}\sqrt{n}$ and $k_n = n\theta + l_n$. Then

$$\mathbb{P}_\theta(\tilde{\tau}_n \leq k_n) = \mathbb{P}_\theta\left[\sup_{na\leq k\leq k_n} \sigma_n^{-1}k^{-3/2} \max_{1\leq l\leq k-1}\left|\sum_{i=l+1}^{k}\sum_{j=1}^{l} K(X_i^n, X_j^n)\right| > c\right]$$

$$\geq \mathbb{P}_\theta\left[\sigma_n^{-1}k_n^{-3/2}\left|\sum_{i=n\theta+1}^{k_n}\sum_{j=1}^{n\theta} K(X_i^n, X_j^n)\right| > c\right]$$

$$= \mathbb{P}_\theta\left[l_n n\theta k_n^{-3/2}\sigma_n^{-1}\left|\frac{1}{l_n n\theta}\sum_{i=n\theta+1}^{k_n}\sum_{j=1}^{n\theta}[K(X_i^n, X_j^n) - \lambda] + \lambda\right| > c\right]$$

$$\geq \mathbb{P}_\theta\left[l_n n\theta k_n^{-3/2}\sigma_n^{-1}\left|\left|\frac{1}{l_n n\theta}\sum_{i=n\theta+1}^{k_n}\sum_{j=1}^{n\theta}[K(X_i^n, X_j^n) - \lambda]\right| - |\lambda|\right| > c\right]$$

$$\geq 1 - \mathbb{P}_\theta\left[\left|\frac{1}{l_n n\theta}\sum_{i=n\theta+1}^{k_n}\sum_{j=1}^{n\theta}[K(X_i^n, X_j^n) - \lambda]\right| \geq |\lambda| - \frac{c\sigma_n k_n^{3/2}}{l_n n\theta}\right]$$

$$\geq 1 - \mathbb{P}_\theta\left[|\cdots| \geq |\lambda| - \frac{c\sigma_n\sqrt{n}}{l_n\theta}\right] \quad \text{since } k_n \leq n \text{ for } n \text{ large enough}$$

$$= 1 - \mathbb{P}_\theta(|\cdots| \geq \tfrac{1}{2}|\lambda|) \to 1 \quad \text{as } n\to\infty,$$

since $|\cdots| \to 0$ with probability one. In fact,

$$\frac{1}{l_n n\theta} \sum_{i=n\theta+1}^{k_n} \sum_{j=1}^{n\theta} K(X_i^n, X_j^n) = \int K(x, y) \mu_{2,n} \otimes \mu_{1,n} \, \mathrm{d}(x, y),$$

where

$$\mu_{2,n} = \frac{1}{l_n} \sum_{i=n\theta+1}^{k_n} \delta_{X_i^n} \quad \text{and} \quad \mu_{1,n} = \frac{1}{n\theta} \sum_{j=1}^{n\theta} \delta_{X_j^n}.$$

Using Varadarajan's (1958) arguments one shows that $\mu_{i,n}$ converges weakly to $v_i$ with probability one, $i = 1, 2$. Now apply Theorem 3.2 of Billingsley (1968) and the Portmanteau theorem to get the desired result. □

Note that we sometimes actually can replace $\sigma_n$ by the quantity $s(n)$, e.g., in the case $\mathscr{X} = \mathbb{R}$, $K(x, y) = \operatorname{sign}(x - y)$ and $v_{1,n}$ continuous for all $n \in \mathbb{N}$, $s^2(n) = 1/3$. So, in this situation we have a deterministic upper bound for the number of observations we will take (with high probability) after the changepoint. Moreover, the upper bound reflects the fact that the closer the alternative is to the hypothesis (measured by $\lambda$) the later we will stop.

## 3. Simulations

In a small sample simulation study we investigated our sequential procedure. Since we did not know the critical value $c(\alpha) = c(\alpha, a)$ explicitely, we had to approximate it by Monte–Carlo simulation. As pointed out in chapter two we can take w.l.o.g. $\mathscr{X} = (0, 1)$ and $v_{1,n}$ the uniform distribution. For the kernel we chose $K(x, y) = \operatorname{sign}(x - y)$, so that $s^2(n) = \frac{1}{3}$. From our theory we know that

$$S_n = \sqrt{3} \sup_{a \le u \le 1} \sup_{0 \le t \le 1} |\xi_n(u, t)| \xrightarrow{\mathscr{L}} \sup_{a \le u \le 1} \sup_{0 \le t \le 1} |\zeta(u, t)|.$$

For this reason we generated $m \in \mathbb{N}$ independent variables distributed as $S_n$ and took the $(1 - \alpha)$-quantile of the empirical distribution function of these variables as an approximation of $c$. For $m = 200$ and $n = 50$ we obtained the values presented in Table 1.

In our simulation study we fixed $a = 0.2$, $\alpha = 0.1$, $n = 50$ and $K(x, y) = \operatorname{sign}(x - y)$. We generated $m = 100$ independent variables $\tau_n^{(i)}$, $i = 1, \ldots, m$, distributed as $\tau_n$, where $X_1^n, \ldots, X_{[n\theta]}^n$ had distribution $v_1$ and the remaining $X_{[n\theta]+1}^n, \ldots, X_n^n$ had distribution $v_2$. By the SLLN the arithmetic means $m^{-1} \sum_{i=1}^m \tau_n^{(i)}$ and $m^{-1} \sum_{i=1}^m 1_{\{\tau_n^{(i)} \le [n\theta]\}}$ are reasonable estimators for $\mathbb{E}_\theta(\tau_n)$ and $\mathbb{P}_\theta(\tau_n \le [n\theta])$. If $v_1 = U(0, 1)$ and $v_2 = U(d, 1 + d)$, $0 < d \le 1$, $\lambda = 2d - d^2$. Tables 2 and 3 show the values for $d = 0.3$ ($\lambda = 0.51$) and $d = 0.4$ ($\lambda = 0.64$). If $v_i = \exp(\lambda_i)$, $\lambda_i > 0$, $i = 1, 2$, $\lambda = (\lambda_1 - \lambda_2)/(\lambda_1 + \lambda_2)$. Tables 4 and 5 show the values for $\lambda_1 = 1$, $\lambda_2 = 0.324503311$ ($\lambda = 0.51!$) and $\lambda_1 = 1$, $\lambda_2 = 0.219512195$ ($\lambda = 0.64!$).

Table 1

| | $\alpha$ | | | |
|---|---|---|---|---|
| $a$ | 0.1 | 0.05 | 0.025 | 0.01 |
| 0.1 | 1.422 | 1.489 | 1.581 | 1.730 |
| 0.2 | 1.422 | 1.489 | 1.581 | 1.730 |
| 0.3 | 1.422 | 1.489 | 1.581 | 1.730 |
| 0.4 | 1.413 | 1.476 | 1.581 | 1.730 |
| 0.5 | 1.408 | 1.473 | 1.581 | 1.730 |
| 0.6 | 1.346 | 1.455 | 1.531 | 1.666 |
| 0.7 | 1.299 | 1.405 | 1.508 | 1.615 |
| 0.8 | 1.232 | 1.376 | 1.457 | 1.615 |
| 0.9 | 1.156 | 1.303 | 1.417 | 1.616 |
| 1.0 | 1.107 | 1.235 | 1.323 | 1.538 |

Table 2

| $[n\theta]$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
|---|---|---|---|---|---|---|---|---|
| $\mathbb{E}_\theta(\tau_n)$ | 33.55 | 34.15 | 36.00 | 38.81 | 42.88 | 42.25 | 47.50 | 47.38 |
| $\mathbb{P}_\theta(\tau_n \le [n\theta])$ | 0 | 0.01 | 0.05 | 0.05 | 0.03 | 0.08 | 0.06 | 0.09 |

Table 3

| $[n\theta]$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
|---|---|---|---|---|---|---|---|---|
| $\mathbb{E}_\theta(\tau_n)$ | 25.23 | 26.30 | 30.31 | 33.99 | 39.35 | 43.11 | 46.57 | 46.51 |
| $\mathbb{P}_\theta(\tau_n \le [n\theta])$ | 0 | 0.03 | 0.06 | 0.10 | 0.04 | 0.14 | 0.09 | 0.14 |

Table 4

| $[n\theta]$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
|---|---|---|---|---|---|---|---|---|
| $\mathbb{E}_\theta(\tau_n)$ | 33.85 | 32.66 | 35.30 | 38.08 | 40.91 | 45.24 | 45.89 | 47.41 |
| $\mathbb{P}_\theta(\tau_n \le [n\theta])$ | 0 | 0.02 | 0.01 | 0.04 | 0.09 | 0.07 | 0.12 | 0.09 |

Table 5

| $[n\theta]$ | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
|---|---|---|---|---|---|---|---|---|
| $\mathbb{E}_\theta(\tau_n)$ | 23.79 | 24.90 | 29.45 | 34.17 | 38.94 | 43.05 | 47.12 | 46.16 |
| $\mathbb{P}_\theta(\tau_n \le [n\theta])$ | 0 | 0.02 | 0.04 | 0.05 | 0.09 | 0.06 | 0.07 | 0.13 |

## References

[1] P.K. Bhattacharya and D. Frierson Jr., A nonparametric control chart for detecting small disorders, Ann. Statist. 9 (1981) 544-554.
[2] P. Billingsley, Convergence of Probability Measures, (Wiley, New York, 1968).
[3] B.E. Brodskii and B.S. Darhovskii, Comparative analysis of some nonparametric methods of fastest detection of the moment of "disorder" of a random sequence, Theory Probab. Appl. 35 (1991) 639-652.
[4] Y.S. Chow and H. Teicher, Probability Theory, (Springer, New York, 1978).
[5] M. Csörgő and L. Horváth, Detecting change in a random sequence, J. Multivariate Anal. 23 (1987) 119-130.
[6] B.S. Darhovskii and B.E. Brodskii, Nonparametric method of the fastest detection of a change in the mean of a random sequence, Theory Probab. Appl. 32 (1987) 640-648.
[7] D. Ferger, Nonparametric changepoint - Detection based on U-statistics, Dissertation (University of Giessen, 1991).
[8] D. Ferger, Change-point estimators in case of small disorders, J. Statist. Planning Inference (1994), to appear.
[9] M.A. Girshick and H. Rubin, A Bayes approach to a quality control model, Ann. Math. Statist. 23 (1952) 114-125.
[10] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1963) 13-30.
[11] M.A. Lifshits, On the absolute continuity of distributions of functionals of random processes, Theory Probab. Appl. 27 (1982) 600-607.
[12] G. Lorden, Procedures for reacting to a change in distribution, Ann. Statist. 42 (1971) 1897-1908.
[13] G. Neuhaus, On weak convergence of stochastic processes with multidimensional time parameter, Ann. Math. Stat. 42 (1971) 1285-1295.
[14] E. Page, Continuous inspection schemes, Biometrika 42 (1954) 100-115.
[15] W.A. Shewhart, Economic Control of Quality of Manufactured Product (Van Nostrand Reinhold Co., New York, 1931).
[16] A.N. Shiryaev, Statistical Sequential Analysis (American Mathematical Society, Providence, RI, 1973).
[17] V.S. Varadarajan, Weak convergence of measures on separable metric spaces, Sankhya 19 (1958) 15-22.