Using U-statistics based processes to detect multiple change-points

Markus Orasch*

Department of Statistics, Probability Theory and Actuarial Mathematics, Technische Universität Wien, 1040 Vienna, Austria

Abstract

We study the asymptotic behavior of U-statistics based processes which can be used to detect (multiple or structural) changes in the distribution of independent observations as well as independent vectors. For these processes we prove invariance principles. We give an application to test for (multiple) changes in the mean or in the variance.

AMS Subject Classifications: Primary 62G20 (Asymptotic properties); secondary 62G10 (Non-parametric hypothesis testing)

Keywords: Consistency; Multiple change-points; U-statistics; Weak approximations; Wiener process

1 Introduction

The problem of abrupt parameter changes arises in many situations of daily life, as well as in a variety of experimental and mathematical sciences. For instance, in archaeology, in econometrics, in epidemiology, in nuclear physics, in medicine, or in quality control.

In practice, usually a (large) set of data is observed. Then a statistical test should determine whether there was a change in the data or not. This set of data may be modeled by saying that we observe independent random variables over a special period of time, hence via a random process. Then we wish to detect whether a change could have occurred in the distribution that governs this random process as time goes by.

We wish to study such phenomena in terms of special stochastic processes based on U-statistics. It is needless to say that there are many other ways to study such phenomena. The construction of these processes is such that statistical tests can be based on them for detecting possible changes in their distribution.

As mentioned by Csörgő and Horváth (1988a), change-point problems have originally arisen in the context of quality control, where one typically observes the output of a production line and would wish to signal deviations from an acceptable output level while observing the data. When one observes such a random process sequentially and stops observing at a random time of detecting change, then one speaks of a sequential procedure (e. g., stop a production line, if a specified percentage of the output is not good). Otherwise, one usually observes a change in a chronically ordered finite sequence for the sake of determining possible change(s) during the data collection (e. g., check whether a production line produced reasonable output or not). Most such fixed sample size non-sequential procedures are described in terms of asymptotic results ('infinite' sample size, i. e., $n \to \infty$).

*Research done while at Carleton University, supported partially by an NSERC Canada Grand of Miklós Csörgő at Carleton University, Ottawa. Also supported by the Austrian Science Foundation (FWF) under grant SFB#010 ('Adaptive Information Systems and Modelling in Economics and Management Science').

Depending on whether the distribution of the data is assumed to be known we use either parametric or non-parametric models. For parametric models we refer to a survey of Csörgő and Horváth (1997) at the end of the first Chapter of their book. For non-parametric cases we refer, for example, to Brodsky and Darkhovsky (1993), Csörgő and Horváth (1988a, 1988b, 1997), Erasmus and Lombard (1988), Szyszkowicz (1992, 1998) as well as to their bibliographies.

We now state the problem of testing for a change in the data in a more mathematical way. Suppose we would like to test the null hypothesis

H_0 : $X_1, \ldots, X_n$ *are independent identically distributed random variables*

against the alternative that there is at most one change-point in the sequence $X_1, \ldots, X_n$, namely that we have

$H_A^{(1)}$: $X_1, \ldots, X_n$ *are independent random variables and there is an integer* τ, $1 \le \tau < n$, *such that* $\mathbb{P}\{X_1 \le t\} = \cdots = \mathbb{P}\{X_\tau \le t\}$, $\mathbb{P}\{X_{\tau+1} \le t\} = \cdots = \mathbb{P}\{X_n \le t\}$ *for all* t *and* $\mathbb{P}\{X_\tau \le t_0\} \ne \mathbb{P}\{X_{\tau+1} \le t_0\}$ *for some* t_0.

This means that we are testing, for having n independent random variables belonging to the same distribution, versus the first τ ones belonging to the same distribution and the last $n-\tau$ ones to a different one. Therefore, for each $1 \le k < n$, we will compare the first k observations to the last $n-k$ ones by using a bivariate function $h(x,y)$ that is often called the kernel function in the literature on U-statistics (cf., for example, Koroljuk and Borovskich (1994) and Serfling (1980, Chapter 5)).

Tests for at most one change-point which are based on processes of U-statistics were first studied by Csörgő and Horváth (1988b, 1997). They investigate the asymptotic properties (as $n \to \infty$) of the U-statistics based process

$$Z_k = \sum_{1 \le i \le k} \sum_{k < j \le n} h(X_i, X_j), \qquad 1 \le k < n, \tag{1.1}$$

where the kernel $h(x,y)$ is either symmetric, i. e.,

$$h(x,y) = h(y,x), \qquad \text{for all } x, y \in \mathbf{R}, \tag{1.2}$$

or antisymmetric, i. e.,

$$h(x,y) = -h(y,x), \qquad \text{for all } x, y \in \mathbf{R}. \tag{1.3}$$

Typical choices for (1.2) are xy, $\frac{x^2+y^2}{2}$, $\frac{(x-y)^2}{2}$ (sample variance), $|x-y|$ (Gini's mean difference), or $\text{sign}(x+y)$ (Wilcoxon's one-sample statistic) and those for (1.3) are $(x-y)$ or $\text{sign}(x-y)$. In particular, we will use $\frac{(x-y)^2}{2}$ to detect changes in the variance, $(x-y)$ to detect changes in the mean, and $\frac{x^2+y^2}{2}$ to detect changes in the mean and/or variance.

Csörgő and Horváth (1988b, 1997) give various asymptotic distributions of the U-statistics based process $\{Z_k, 1 \le k < n,\}$ under the null hypothesis H_0 and the alternative $H_A^{(1)}$ for symmetric and antisymmetric kernels. They also give tests that can be used to consistently reject H_0 against $H_A^{(1)}$.

However, instead of the alternative $H_A^{(1)}$ of at most one change, we can also consider a more general one that allows at most s change-points ($s \in \mathbf{N}$) in the sequence $X_1, \ldots, X_n$, namely that we have

$H_A^{(s)}$: $X_1, \ldots, X_n$ *are independent random variables and there exist* s, $1 \le s < n$, *integers* $\tau_1 = \tau_1(n)$, $\tau_2 = \tau_2(n)$, $\ldots$, $\tau_s = \tau_s(n)$, $1 \le \tau_1 \le \tau_2 \le \cdots \le \tau_s < n$, *such that* $\mathbb{P}\{X_1 \le t\} = \cdots = \mathbb{P}\{X_{\tau_1} \le t\}$, $\mathbb{P}\{X_{\tau_1+1} \le t\} = \cdots = \mathbb{P}\{X_{\tau_2} \le t\}$, $\ldots$, $\mathbb{P}\{X_{\tau_s+1} \le t\} = \cdots = \mathbb{P}\{X_n \le t\}$ *for all* t *and* $\mathbb{P}\{X_{\tau_i} \le t_0\} \ne \mathbb{P}\{X_{\tau_i+1} \le t_0\}$ *for some* t_0 *and for all* $1 \le i \le s$.

We note that the alternative $H_A^{(s)}$ allows us to consider random variables $X_1, X_2, \ldots, X_n$ with s changes in the distribution, which do not necessarily result in $(s+1)$ different distributions.

Since we are interested in testing for at most s, $1 \le s < n$, changes it seems natural to define a sequence of stochastic processes depending on s parameters. To construct such processes we split the given sample $X_1, \ldots, X_n$ into $(s+1)$ blocks and compare each of the blocks with the others via the kernel function $h(x, y)$ as in (1.2) or (1.3) respectively. Since, out of s, we always compare two blocks with each other, we have $\binom{s+1}{2}$ different possibilities to do so.

Therefore, for the problem in hand, we define a sequence of s-time parameter stochastic processes as follows:

$$\begin{aligned}
Z_{k_1,k_2,\ldots,k_s} &:= \sum_{i=1}^{k_1}\Big(\sum_{j=k_1+1}^{k_2} h(X_i,X_j) + \cdots + \sum_{j=k_s+1}^{n} h(X_i,X_j)\Big) \\
&\quad + \sum_{i=k_1+1}^{k_2}\Big(\sum_{j=k_2+1}^{k_3} h(X_i,X_j) + \cdots + \sum_{j=k_s+1}^{n} h(X_i,X_j)\Big) + \ldots \\
&\quad + \sum_{i=k_{s-1}+1}^{k_s}\sum_{j=k_s+1}^{n} h(X_i,X_j) \\
&= \sum_{m=1}^{s}\sum_{i=k_{m-1}+1}^{k_m}\sum_{l=m}^{s}\sum_{j=k_l+1}^{k_{l+1}} h(X_i,X_j), \\
&\qquad\qquad 1 \le k_1 < k_2 < \cdots < k_s < n,
\end{aligned} \tag{1.4}$$

where $k_0 := 0$ and $k_{s+1} := n$. In this way we compare the $(s+1)$ blocks $(X_1, \ldots, X_{k_1})$, $(X_{k_1+1}, \ldots, X_{k_2})$, $\ldots$, $(X_{k_s+1}, \ldots, X_n)$ with each other, where the k_i's, $1 \le i \le s$, vary from i to $n+i-1-s$.

When testing for at most one change, this process reduces to

$$\begin{aligned}
Z_{k_1} &= \sum_{m=1}^{1}\sum_{i=k_{m-1}+1}^{k_m}\sum_{l=1}^{1}\sum_{j=k_l+1}^{k_{l+1}} h(X_i,X_j) \\
&= \sum_{i=1}^{k_1}\sum_{j=k_1+1}^{n} h(X_i,X_j), \qquad 1 \le k_1 < n,
\end{aligned}$$

which is the stochastic process from (1.1) introduced by Csörgő and Horváth (1988b, 1997). When testing for at most two changes the process in (1.4) reduces to

$$\begin{aligned}
Z_{k_1,k_2} &= \sum_{m=1}^{2}\sum_{i=k_{m-1}+1}^{k_m}\sum_{l=m}^{2}\sum_{j=k_l+1}^{k_{l+1}} h(X_i,X_j) \\
&= \sum_{i=k_0+1}^{k_1}\Big(\sum_{j=k_1+1}^{k_2} h(X_i,X_j) + \sum_{j=k_2+1}^{k_3} h(X_i,X_j)\Big) + \sum_{i=k_1+1}^{k_2}\sum_{j=k_2+1}^{k_3} h(X_i,X_j) \\
&= \sum_{i=1}^{k_1}\sum_{j=k_1+1}^{n} h(X_i,X_j) + \sum_{i=k_1+1}^{k_2}\sum_{j=k_2+1}^{n} h(X_i,X_j), \qquad 1 \le k_1 < k_2 < n.
\end{aligned}$$

We study $Z_{k_1,k_2,\ldots,k_s}$, $1 \le k_1 < k_2 < \cdots < k_s < n$, under the null hypothesis H_0 in Section 2, and under the alternative hypothesis $H_A^{(s)}$ in Section 3 when the kernels are both, symmetric and antisymmetric. Since many practical situations can be modeled via random variables, the main part of this paper focuses on them.

However, we note that throughout this paper we can exchange the random variables $X_1, \ldots, X_n$ by random vectors $\mathbf{X_1}, \ldots, \mathbf{X_n}$, where $\mathbf{X_i} \in \mathbf{R}^d$, $1 \leq i \leq n$. The following theorems hold in this case as well, but for proofs we need to apply the corresponding results by Koroljuk and Borovskich (1994, Chapter 5) who investigate U-statistics for random vectors. Random vectors in the theory of U-statistic processes were, for example, also used by Gombay and Horváth (1999).

2 Asymptotics under H_0

Throughout this paper let $k_0 := 0$, $k_{s+1} := n$, $t_0 := 0$, $t_{s+1} := 1$ and $k_i := [(n+1)t_i]$, $1 \leq i \leq s$. Furthermore, let us assume that

$$\mathbb{E}h^2(X_1, X_2) \quad < \quad \infty, \tag{2.1}$$

and define $\theta := \mathbb{E}h(X_1, X_2)$ and $\tilde{h}(t) := \mathbb{E}\{h(t, X_2) - \theta\}$. In addition, let us also assume that

$$\tilde{\sigma}^2 := \mathbb{E}\tilde{h}^2(X_1) > 0, \tag{2.2}$$

which, as a consequence of (2.1), is finite as well. Since $\mathbb{E}Z_{k_1,k_2,\ldots,k_s} = \sum_{i=1}^{s}(k_{i+1} - k_i)k_i\theta$, we define the centralized process

$$U_{k_1,k_2,\ldots,k_s} \quad = \quad Z_{k_1,k_2,\ldots,k_s} - \sum_{i=1}^{s}(k_{i+1} - k_i)k_i\theta, \quad 1 \leq k_1 < k_2 < \cdots < k_s < n,$$

where the kernel function h in $Z_{k_1,k_2,\ldots,k_s}$ is symmetric. Since $\theta = 0$ for antisymmetric kernels, we define in this case

$$\bar{U}_{k_1,k_2,\ldots,k_s} \quad = \quad Z_{k_1,k_2,\ldots,k_s}, \quad 1 \leq k_1 < k_2 < \cdots < k_s < n.$$

Although $U_{k_1,k_2,\ldots,k_s}$ itself is not a U-statistic, we can write it as the sum of $(s+2)$ U-statistics (with missing normalizing factor), and thus for $1 \leq k_1 < k_2 < \cdots < k_s < n$ we get that

$$U_{k_1,k_2,\ldots,k_s} \quad = \quad U_n^{(s+2)} - (U_{k_1}^{(1)} + U_{k_1,k_2}^{(2)} + \cdots + U_{k_s,n}^{(s+1)}), \tag{2.3}$$

where

$$\begin{aligned}
U_{k_1}^{(1)} &= \sum_{1 \leq i < j \leq k_1} h(X_i, X_j) - \binom{k_1}{2}\theta, \\
U_{k_1,k_2}^{(2)} &= \sum_{k_1 < i < j \leq k_2} h(X_i, X_j) - \binom{k_2 - k_1}{2}\theta, \\
&\vdots \\
U_{k_s,n}^{(s+1)} &= \sum_{k_s < i < j \leq n} h(X_i, X_j) - \binom{n - k_s}{2}\theta, \\
U_n^{(s+2)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2}\theta.
\end{aligned}$$

Consequently, $\bar{U}_{k_1,k_2,\ldots,k_s}$ can be written in a similar way when we put $\theta = 0$ in (2.3). For further use we also define the sequence of s-time parameter stochastic processes

$$U_n(t_1, t_2, \ldots, t_s) \quad = \quad \frac{1}{\tilde{\sigma}n^{3/2}} U_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]}, \quad 0 < t_1 < t_2 < \cdots < t_s < 1,$$

and

$$\bar{U}_n(t_1, t_2, \ldots, t_s) \quad = \quad \frac{1}{\tilde{\sigma}n^{3/2}} \bar{U}_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]}, \quad 0 < t_1 < t_2 < \cdots < t_s < 1.$$

Furthermore, we note that the upper index ^{sym} first used in Theorem 2.1 symbolizes that the herein defined Gaussian process corresponds to the case where we have a symmetric kernel. In case of an antisymmetric kernel, as in Theorem 2.3, we will use the upper index ^{a} instead.

Using the previous assumptions and notations we arrive at the following theorem, when the kernel is symmetric.

Theorem 2.1 *Assume that H_0, (1.2), (2.1) and (2.2) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n^{sym}(t_1, t_2, \ldots, t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\}_{n \in \mathbf{N}}$ such that, as $n \to \infty$,*

$$\sup_{0<t_1<t_2<\cdots<t_s<1} |U_n(t_1, t_2, \ldots, t_s) - \Gamma_n^{sym}(t_1, t_2, \ldots, t_s)| = o_P(1),$$

and, for each n,

$$\{\Gamma_n^{sym}(t_1, t_2, \ldots, t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^{sym}(t_1, t_2, \ldots, t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\}, \tag{2.4}$$

where the Gaussian process Γ^{sym} is defined via a linear combination of a standard Wiener processes W as follows:

$$\Gamma^{sym}(t_1, t_2, \ldots, t_s) = \sum_{i=1}^{s} (t_{i+1} + t_{i-1} - 2t_i) W(t_i) + t_s W(1), \quad 0 \le t_1 \le t_2 \cdots \le t_s \le 1, \tag{2.5}$$

where $t_0 := 0$ and $t_{s+1} := 1$.

Proof. As a consequence of Hall (1979, Theorem 1) we have

$$\max_{1 \le k_1 \le k_2 \le \cdots \le k_s \le n} \left| U_{k_1}^{(1)} - k_1 \sum_{i=1}^{k_1} \tilde{h}(X_i) \right| = O_P(n),$$

$$\max_{1 \le k_1 \le k_2 \le \cdots \le k_s \le n} \left| U_{k_1,k_2}^{(2)} - (k_2 - k_1) \sum_{i=k_1+1}^{k_2} \tilde{h}(X_i) \right| = O_P(n),$$

$$\vdots$$

$$\max_{1 \le k_1 \le k_2 \le \cdots \le k_s \le n} \left| U_{k_s,n}^{(s+1)} - (n - k_s) \sum_{i=k_s+1}^{n} \tilde{h}(X_i) \right| = O_P(n),$$

$$\left| U_n^{(s+2)} - n \sum_{i=1}^{n} \tilde{h}(X_i) \right| = O_P(n).$$

Hence, via observing that

$$n \sum_{i=1}^{n} \tilde{h}(X_i) - \Big(k_1 \sum_{i=1}^{k_1} \tilde{h}(X_i) + (k_2 - k_1) \sum_{i=k_1+1}^{k_2} \tilde{h}(X_i) + \cdots + (n - k_s) \sum_{i=k_s+1}^{n} \tilde{h}(X_i) \Big)$$
$$= \sum_{i=1}^{s} (k_{i+1} + k_{i-1} - 2k_i) \sum_{j=1}^{k_i} \tilde{h}(X_j) + k_s \sum_{i=1}^{n} \tilde{h}(X_i),$$

we have that

$$\max_{1 \le k_1 \le k_2 \le \cdots \le k_s \le n} \left| U_{k_1,k_2,\ldots,k_s} - \Big(\sum_{i=1}^{s} (k_{i+1} + k_{i-1} - 2k_i) \sum_{j=1}^{k_i} \tilde{h}(X_j) + k_s \sum_{i=1}^{n} \tilde{h}(X_i) \Big) \right| = O_P(n). \tag{2.6}$$

We note at this point that while the U-statistics in the latter statements are sums of identically, but not necessarily independent, distributed random variables, they are approximated by sums of the independent and identically distributed random variables $\tilde{h}(X_1), \ldots, \tilde{h}(X_n)$.

Furthermore, we use the fact that we can define a Wiener process $\{W(t), 0 \le t < \infty\}$ such that (cf. Csörgő and Révész (1981, Theorem S.2.2.1 by Major (1979) combined with (S.2.2.2))), as $n \to \infty$,

$$\frac{1}{n^{1/2}} \sup_{0 \le t \le 1} \Big| \frac{1}{\tilde{\sigma}} \sum_{i=1}^{[(n+1)t]} \tilde{h}(X_i) - W(nt) \Big| \quad = \quad o_P(1).$$

Hence, bounding above $\sup_{0<t_1<t_2<\cdots<t_s<1}$ by $\sup_{0<t_j<1}$, $1 \le j \le s$, then we also have, as $n \to \infty$,

$$\frac{1}{n^{1/2}} \sup_{0<t_1<t_2<\cdots<t_s<1} \Big| \frac{1}{\tilde{\sigma}} \sum_{i=1}^{[(n+1)t_j]} \tilde{h}(X_i) - W(nt_j) \Big| \quad = \quad o_P(1), \text{ for } j = 1, 2, \ldots, s. \tag{2.7}$$

Thus the result follows from (2.6) combined with (2.7). Obviously (2.5) is a Gaussian process and (2.4) holds true as well when we replace, for example, $W(\cdot)$ in (2.5) by $\frac{W(n\cdot)}{\sqrt{n}}$ to define $\Gamma_n^{sym}(\cdot)$. □

Clearly, in case of at most one change, namely when $s = 1$, Theorem 2.1 reduces to Theorem 2.1 of Csörgő and Horváth (1988b). In this case $\Gamma^{sym}(t_1) = (1 - 2t_1)W(t_1) + t_1 W(1)$, $0 \le t_1 \le 1$. When testing for at most two changes, namely $s = 2$, then $\Gamma^{sym}(t_1, t_2) = (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2 W(1)$, $0 \le t_1 \le t_2 \le 1$.

As an immediate consequence it is easy to see that the Gaussian process from (2.5) satisfies the following:

Lemma 2.1 *For $0 \le t_1 \le t_2 \cdots \le t_s \le 1$ we have*

$$\begin{aligned}
\mathbb{E}\Gamma^{sym}(t_1, t_2, \ldots, t_s) &= 0, \\
\mathbb{E}\Big(\Gamma^{sym}(t_1, t_2, \ldots, t_s)\Big)^2 &= \sum_{i=0}^{s} (1 - t_{i+1} + t_i)^2 (t_{i+1} - t_i), \\
\mathbb{E}\Gamma^{sym}(t_1, t_2, \ldots, t_2)\Gamma^{sym}(r_1, r_2, \ldots, r_s) &= \sum_{i=1}^{s}\sum_{j=1}^{s} (t_{i+1} - 2t_i + t_{i-1})(r_{j+1} - 2r_j \\
+ \ r_{j-1})(t_i \wedge r_j) + 2t_s\Big(\sum_{i=1}^{s-1} r_{i+1}r_i - \sum_{i=1}^{s} r_i^2\Big) &+ 2r_s\Big(\sum_{i=1}^{s-1} t_{i+1}t_i - \sum_{i=1}^{s} t_i^2\Big) + 3t_s r_s,
\end{aligned}$$

where $t_0 = r_0 := 0$ and $t_{s+1} = r_{s+1} := 1$.

Theorem 2.1 can now be used to test for multiple change-points. This is due to the fact that under H_0

$$\sup_{0<t_1<t_2<\cdots<t_s<1} |U_n(t_1, t_2, \ldots, t_s)| \quad \xrightarrow{\mathcal{D}} \quad \sup_{0<t_1<t_2<\cdots<t_s<1} |\Gamma^{sym}(t_1, t_2, \ldots, t_s)|. \tag{2.8}$$

Hence, this consequence of Theorem 2.1 allows us to produce tables for the limiting distribution in (2.8) and reject the null hypothesis of no-change for large values of the supremum of $|U_n(t_1,\, t_2,\, \ldots,\, t_s)|$.

Since Theorem 2.1 holds for any integer s, $1 \le s < n$, we are in the position to construct statistics which converge under H_0 in distribution to the sup-functional of the corresponding s-time parameter Gaussian process. Hence, we can combine statistics for different s and arrive at the following theorem.

Theorem 2.2 *Assume that H_0, (1.2), (2.1) and (2.2) hold. Then we can define sequences of Gaussian processes $\big\{\Gamma_n^{sym}(t_1),\ 0 \le t_1 \le 1\big\}$, $\big\{\Gamma_n^{sym}(t_1, t_2),\ 0 \le t_1 < t_2 \le 1\big\}$, $\ldots$, $\big\{\Gamma_n^{sym}(t_1, t_2, \ldots, t_s),\ 0 \le t_1 <$*

$t_2 < \cdots < t_s \le 1,\ 1 \le s < n$ } *such that with the sup-Euclidean norm we have, as* $n \to \infty$,

$$\Big(U_n(t_1), U_n(t_1,t_2), \ldots, U_n(t_1,t_2,\ldots,t_s)\Big) \xrightarrow{\mathcal{D}} \Big(\Gamma^{sym}(t_1), \Gamma^{sym}(t_1,t_2), \ldots, \Gamma^{sym}(t_1,t_2,\ldots,t_s)\Big),$$

where for each n *and* $1 \le i \le s$

$$\{\Gamma_n^{sym}(t_1,t_2,\ldots,t_i), 0 \le t_1 \le t_2 \cdots \le t_i \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^{sym}(t_1,t_2,\ldots,t_i), 0 \le t_1 \le t_2 \cdots \le t_i \le 1\}.$$

Proof. Let $(x_1,x_2,\ldots,x_s)^T$ be an s-dimensional vector $\in (0,1)^s \subset \mathbf{R}^s$ and define the sup-Euclidean norm

$$\| (x_1,x_2,\ldots,x_s)^T \| = \sup_{0<x_1<x_2<\cdots<x_s<1} \sqrt{x_1^2 + x_2^2 + \cdots + x_s^2}.$$

Then with Theorem 2.1

$$\begin{aligned} &\left\| \Big(U_n(t_1) - \Gamma_n^{sym}(t_1), \ldots, U_n(t_1,\ldots,t_s) - \Gamma_n^{sym}(t_1,\ldots,t_s)\Big)^T \right\| \\ &\le \sqrt{s} \sup_{0<t_1<t_2<\cdots<t_s<1} \left\{ \max_{1\le i\le s} \Big| U_n(t_1,\ldots,t_i) - \Gamma_n^{sym}(t_1,\ldots,t_i) \Big| \right\} = o_P(1). \end{aligned}$$

This implies that we have componentwise convergence in distribution of the U_n's to the appropriate Γ^{sym}'s using the appropriate norm. □

Similarly, as in case of symmetric kernels we can proceed in case of antisymmetric ones. However, we note that the corresponding limiting results are different.

Theorem 2.3 *Assume that* H_0, (1.3), (2.1) *and* (2.2) *hold. Then we can define a sequence of Gaussian processes* $\{\Gamma_n^a(t_1,t_2,\ldots,t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\}_{n\in\mathbf{N}}$ *such that, as* $n \to \infty$,

$$\sup_{0<t_1<t_2<\cdots<t_s<1} |\bar{U}_n(t_1,t_2,\ldots,t_s) - \Gamma_n^a(t_1,t_2,\ldots,t_s)| = o_P(1),$$

and, for each n,

$$\{\Gamma_n^a(t_1,t_2,\ldots,t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1,t_2,\ldots,t_s), 0 \le t_1 \le t_2 \cdots \le t_s \le 1\},$$

where the Gaussian process Γ^a *is defined via a linear combination of a standard Wiener process* W *as follows:*

$$\Gamma^a(t_1,t_2,\ldots,t_s) = \sum_{i=1}^{s} (t_{i+1} - t_{i-1}) W(t_i) - t_s W(1), \qquad 0 \le t_1 \le t_2 \cdots \le t_s \le 1, \tag{2.9}$$

where $t_0 := 0$ *and* $t_{s+1} := 1$.

Proof. The proof is similar to that of Theorem 2.1. Instead of Theorem 1 of Hall (1979) we use Theorem

2.1 of Janson and Wichura (1983), which implies

$$\begin{aligned}
\max_{1\le k_1\le k_2\le\cdots\le k_s\le n}\Big|\bar U^{(1)}_{k_1}-\sum_{i=1}^{k_1}(k_1-2i+1)\tilde h(X_i)\Big| &= O_P(n),\\
\max_{1\le k_1\le k_2\le\cdots\le k_s\le n}\Big|\bar U^{(2)}_{k_1,k_2}-\sum_{i=k_1+1}^{k_2}(k_2+k_1-2i+1)\tilde h(X_i)\Big| &= O_P(n),\\
&\ \vdots\\
\max_{1\le k_1\le k_2\le\cdots\le k_s\le n}\Big|\bar U^{(s+1)}_{k_s,n}-\sum_{i=k_s+1}^{n}(n+k_s-2i+1)\tilde h(X_i)\Big| &= O_P(n),\\
\Big|\bar U^{(s+2)}_{n}-\sum_{i=1}^{n}(n-2i+1)\tilde h(X_i)\Big| &= O_P(n).
\end{aligned}$$

Hence, via observing that

$$\begin{aligned}
&\sum_{i=1}^{n}(n-2i+1)\tilde h(X_i)-\Big(\sum_{i=1}^{k_1}(k_1-2i+1)\tilde h(X_i)+\sum_{i=k_1+1}^{k_2}(k_2+k_1-2i+1)\tilde h(X_i)\\
&+\cdots+\sum_{i=k_s+1}^{n}(n+k_s-2i+1)\tilde h(X_i)\Big)\\
&= \sum_{i=1}^{s}(k_{i+1}-k_{i-1})\sum_{j=1}^{k_i}\tilde h(X_j)-k_s\sum_{i=1}^{n}\tilde h(X_i),
\end{aligned}$$

we have that

$$\begin{aligned}
&\max_{1\le k_1\le k_2\le\cdots\le k_s\le n}\Big|\bar U_{k_1,k_2,\ldots,k_s}-\Big(\sum_{i=1}^{s}(k_{i+1}-k_{i-1})\sum_{j=1}^{k_i}\tilde h(X_j)-k_s\sum_{i=1}^{n}\tilde h(X_i)\Big)\Big|\\
&= O_P(n). \qquad (2.10)
\end{aligned}$$

Now (2.10) combined with (2.7) yields the desired result. □

Clearly, in case of at most one change, namely $s=1$, $\{\Gamma^a(t_1), 0\le t_1\le 1\}$ is a Brownian bridge and Theorem 2.3 reduces to Theorem 4.1 of Csörgő and Horváth (1988b). When testing for at most two changes, namely when $s=2$, then $\Gamma^a(t_1,t_2)=t_2W(t_1)+(1-t_1)W(t_2)-t_2W(1)$, $0\le t_1\le t_2\le 1$.

We mention that the Gaussian processes Γ^a from (2.9) and Γ^{sym} from (2.5) are different and for $0\le t_1\le t_2\le\cdots\le t_s\le 1$ their relationship is as follows:

$$\Gamma^a(t_1,t_2,\ldots,t_s) = \Gamma^{sym}(t_1,t_2,\ldots,t_s)+2\Big(\sum_{i=1}^{s}(t_i-t_{i-1})W(t_i)-t_sW(1)\Big).$$

Furthermore we note that the Gaussian process from (2.9) satisfies the following:

Lemma 2.2 *For $0\le t_1\le t_2\cdots\le t_s\le 1$ we have*

$$\begin{aligned}
I\!\!E\Gamma^a(t_1,t_2,\ldots,t_s) &= 0,\\
I\!\!E\Big(\Gamma^a(t_1,t_2,\ldots,t_s)\Big)^2 &= \sum_{i=0}^{s}(1-t_{i+1}-t_i)^2(t_{i+1}-t_i),\\
I\!\!E\Gamma^a(t_1,t_2,\ldots,t_2)\Gamma^a(r_1,r_2,\ldots,r_s) &= \sum_{i=1}^{s}\sum_{j=1}^{s}(t_{i+1}-t_{i-1})(r_{j+1}-r_{j-1})(t_i\wedge r_j)\\
&- t_sr_s,
\end{aligned}$$

where $t_0 = r_0 := 0$ *and* $t_{s+1} = r_{s+1} := 1$.

The computation of the covariance is straight forward. We note that

$$\sum_{i=1}^{s}(r_{i+1} - r_{i-1})r_i \quad = \quad r_s.$$

When using antisymmetric kernels then Theorem 2.3 implies that under H_0

$$\sup_{0<t_1<t_2<\cdots<t_s<1} |\bar{U}_n(t_1, t_2, \ldots, t_s)| \quad \xrightarrow{\mathcal{D}} \quad \sup_{0<t_1<t_2<\cdots<t_s<1} |\Gamma^a(t_1, t_2, \ldots, t_s)|. \tag{2.11}$$

Moreover, this allows us to produce tables for the limiting distribution in (2.11) and reject the null hypothesis of no-change for large values of the supremum of $|\bar{U}_n(t_1, t_2, \ldots, t_s)|$.

Similarly to Theorem 2.2 we get the following when combining statistics used to test for a different number of changes s, $1 \le s < n$.

Theorem 2.4 *Assume that* H_0, (1.3), (2.1) *and* (2.2) *hold. Then we can define sequences of Gaussian processes* $\{\Gamma_n^a(t_1),\ 0 \le t_1 \le 1\}$, $\{\Gamma_n^a(t_1, t_2),\ 0 \le t_1 < t_2 \le 1\}$, $\ldots$, $\{\Gamma_n^a(t_1, t_2, \ldots, t_s),\ 0 \le t_1 < t_2 < \ldots < t_s \le 1,\ 1 \le s < n\}$ *such that with the sup-Euclidean norm we have componentwise convergence in distribution, namely, as* $n \to \infty$,

$$\Big(\bar{U}_n(t_1), \bar{U}_n(t_1, t_2), \ldots, \bar{U}_n(t_1, t_2, \ldots, t_s)\Big) \\ \xrightarrow{\mathcal{D}} \Big(\Gamma^a(t_1), \Gamma^a(t_1, t_2), \ldots, \Gamma^a(t_1, t_2, \ldots, t_s)\Big),$$

where for each n *and* $1 \le i \le s$

$$\{\Gamma_n^a(t_1, t_2, \ldots, t_i), 0 \le t_1 \le t_2 \cdots \le t_i \le 1\} \\ \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2, \ldots, t_i), 0 \le t_1 \le t_2 \cdots \le t_i \le 1\}.$$

The proof of this theorem goes along the lines of the proof of Theorem 2.2 via using Theorem 2.3 instead of Theorem 2.1.

3 Asymptotics under $H_A^{(s)}$

In this section we study the limiting behavior of the proper normalized stochastic process $Z_{k_1,k_2,\ldots,k_s}$ from (1.4) under the alternative of at most s changes. Its limiting function will depend on the location of the change-points $\tau_1 = [n\lambda_1], \tau_2 = [n\lambda_2], \ldots, \tau_s = [n\lambda_s]$, $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_s < 1$. However, it will involve many variables, since we have to handle all possible combinations of t_i's and λ_i's. In particular the limiting function consists of $\binom{2s}{2}$ different parts, since there are $\binom{2s}{2}$ possibilities to choose $t_1, \ldots, t_s$ given the postulated change-points $\lambda_1, \ldots, \lambda_s$.

Let $F_1(t) = \mathbb{P}\{X_{\tau_1} \le t\}$, $F_2(t) = \mathbb{P}\{X_{\tau_1+1} \le t\}$, $F_3(t) = \mathbb{P}\{X_{\tau_2+1} \le t\}$, $\ldots$, $F_{s+1}(t) = \mathbb{P}\{X_{\tau_s+1} \le t\}$ be the respective distribution functions of the observations before the first, between the first and second, between the second and third, $\ldots$, and after the s-th change respectively, and put

$$\mathbb{E}h(X_i, X_j) =: \theta_{q+1,r+1},$$

where $\tau_q < i \le \tau_{q+1}$ and $\tau_r < j \le \tau_{r+1}$ for all $0 \le q \le r \le s$, and where we define $\tau_0 := 1$ and $\tau_{s+1} := n$. We also assume under $H_A^{(s)}$ the existence of first moments, namely

$$|\mathbb{E}h(X_i, X_j)| \quad < \quad \infty, \qquad 1 \le i < j \le n. \tag{3.1}$$

Furthermore, we put

$$\tau_i \quad := \quad [n\lambda_i], \quad 0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_s < 1,$$

For the sake of strong laws, we use a weaker assumption than a finite second moment of h, namely that for random variables from different distributions the following holds:

$$\mathbb{E}\big\{|h(X_{\tau_m}, X_{\tau_l+1})| \log^+\big(|h(X_{\tau_m}, X_{\tau_l+1})|\big)\big\} < \infty, \qquad 1 \le m \le l \le s, \tag{3.2}$$

where $\log^+ x = \log(x \vee 1)$. Since we do not know in advance the location of the s possible change-points, we define for technical purposes

$$l(z) \quad := \quad \begin{cases} 1, & 0 < z \le \lambda_1, \\ 2, & \lambda_1 < z \le \lambda_2, \\ \vdots & \\ s, & \lambda_{s-1} < z \le \lambda_s, \\ s+1, & \lambda_s < z \le 1, \end{cases}$$

which will be used to remind ourselves of the location (either before the first or between the first and second or between the second and third or ... or after the last change-point) of a block of r.v.'s which does not contain any change-point.

When looking at the definition of $Z_{k_1,k_2,\ldots,k_s}$, $1 \le k_1 < k_2 \cdots < k_s < n$ in (1.4), we see that we may split it into many double sums, where each of these double sums is of the form

$$\sum_{i=a+1}^{b} \sum_{j=c+1}^{d} h(X_i, X_j), \tag{3.3}$$

where $0 \le a < b \le c < d \le n$ and $a, b, c, d \in \mathbf{N}$. These double sums may be associated with comparing the two blocks $(X_{a+1}, \ldots, X_b)$ and $(X_{c+1}, \ldots, X_d)$ with each other. In the case of testing for s changes, we have to compare each of the blocks $(X_1, \ldots, X_{k_1})$, $(X_{k_1+1}, \ldots, X_{k_2})$, ..., $(X_{k_s+1}, \ldots, X_n)$ with each other. Of course, each of these $s+1$ blocks may contain up to s change-points, if, for example, the others have none. Hence, we consider double sums as in (3.3), but we split each of the sums into $s+1$ sums to avoid summing over two blocks with a possible change-point inside, namely

$$\begin{aligned} \sum_{i=a+1}^{b} \sum_{j=c+1}^{d} h(X_i, X_j) \quad = \quad & \Big(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^{\gamma_2} + \cdots + \sum_{i=\gamma_{s-1}+1}^{\gamma_s} + \sum_{i=\gamma_s+1}^{b} \Big) \Big(\sum_{j=c+1}^{\gamma_{s+1}} \\ & + \sum_{j=\gamma_{s+1}+1}^{\gamma_{s+2}} + \cdots + \sum_{j=\gamma_{2s-1}+1}^{\gamma_{2s}} + \sum_{j=\gamma_{2s}+1}^{d} \Big) h(X_i, X_j), \end{aligned} \tag{3.4}$$

where $0 \le a < \gamma_1 \le \gamma_2 \le \cdots \le \gamma_s \le b \le c < \gamma_{s+1} \le \gamma_{s+2} \le \cdots \le \gamma_{2s} \le d \le n$.

Consequently, we are comparing blocks with each other that do not have a change inside. When using (3.4), the double sum in (3.3) is split into $(s+1)^2$ double sums. Again we emphasize that there are at most s changes in total which implies that some of the new small blocks may be viewed as bigger ones, since there is no change inside the blocks, nor in between them.

As an example, consider the case where we have $(s-2)$ changes in block $(X_{a+1}, \ldots, X_b)$ and 2 changes in block $(X_{c+1}, \ldots, X_d)$ then (3.4) reduces to

$$\begin{aligned} \sum_{i=a+1}^{b} \sum_{j=c+1}^{d} h(X_i, X_j) \quad = \quad & \Big(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^{\gamma_2} + \cdots + \sum_{i=\gamma_{s-1}+1}^{\gamma_{s-2}} + \sum_{i=\gamma_{s-2}+1}^{b} \Big) \Big(\sum_{j=c+1}^{\gamma_{s+1}} \\ & + \sum_{j=\gamma_{s+1}+1}^{\gamma_{s+2}} + \sum_{j=\gamma_{s+2}+1}^{d} \Big) h(X_i, X_j), \end{aligned}$$

where the blocks $(X_{a+1},\ldots,X_{\gamma_1})$, $(X_{\gamma_1+1},\ldots,X_{\gamma_2})$, …, $(X_{\gamma_{s-2}+1},\ldots,X_b)$ and $(X_{c+1},\ldots,X_{\gamma_{s+1}})$, $(X_{\gamma_{s+1}+1},\ldots,X_{\gamma_{s+2}})$, …, $(X_{\gamma_{s+2}+1},\ldots,X_d)$ do not contain a change-point.

Each of the $(s+1)^2$ different double sums in (3.4) is of the form

$$\sum_{i=R_1+1}^{R_2}\sum_{j=R_3+1}^{R_4} h(X_i,X_j),$$

where $1 \le R_1 = R_1(n) := [nr_1] < R_2 = R_2(n) := [nr_2] \le R_3 = R_3(n) := [nr_3] < R_4 = R_4(n) := [nr_4] \le n$ are chosen properly according to the double sums in (3.4). Furthermore, we can prove the following lemma.

Lemma 3.1 *Assume that* (1.2) *or* (1.3), (3.1) *and* (3.2) *hold. In addition assume that either* H_0 *or* $H_A^{(s)}$ *holds and that the two blocks of independent random variables* $(X_{[nr_1]+1},\ldots,X_{[nr_2]+1})$ *and* $(X_{[nr_3]+1},\ldots,X_{[nr_4]+1})$, $1 \le [nr_1] < [nr_2] \le [nr_3] < [nr_4] \le n$ *have distribution function* F *and* G, *respectively (We note that the case of* $F \equiv G$ *is included as well). Then, as* $n \to \infty$,

$$\frac{1}{([nr_2]-[nr_1])([nr_4]-[nr_3])}\sum_{i=[nr_1]+1}^{[nr_2]}\sum_{j=[nr_3]+1}^{[nr_4]} h(X_i,X_j) \xrightarrow{P} \theta_{l(r_2),l(r_4)}. \tag{3.5}$$

Proof. This is an immediate consequence of Theorem 1 for generalized two sample U-statistics by Sen (1977), and Hoeffding's Strong Law of Large Numbers (1961) for two samples from the same distribution.

Note that the double sum in (3.5) may be associated with comparing the two blocks $(X_{R_1+1},\ldots,X_{R_2})$ and $(X_{R_3+1},\ldots,X_{R_4})$, where $[nr_i] =: R_i$, $1 \le i \le 4$, with each other, where both do not have any changes inside. If both belong to different distributions, then Theorem 1 of Sen (1977) applies, and (3.5) follows immediately. On the other hand, if both belong to the same distribution, then we may consider the two blocks as one large block. We do this by deleting everything between these two blocks. We therefore consider the block of i.i.d. r.v.'s $Y = (Y_{R_1+1} := X_{R_1+1},\ Y_{R_1+2} := X_{R_1+2},\ldots,Y_{R_2} := X_{R_2}, Y_{R_2+1} := X_{[R_3-(R_3-R_2)]+1},\ Y_{R_2+2} := X_{[R_3-(R_3-R_2)]+2},\ldots,Y_{[R_4-(R_3-R_2)]} := X_{[R_4-(R_3-R_2)]})$. We may then use Hoeffding's SLLN, but first we have to write the double sum in (3.5) in terms of U-statistics (with missing normalizing factor). Hence, we now write

$$\begin{aligned}
\sum_{i=R_1+1}^{R_2}\sum_{j=R_3+1}^{R_4} h(X_i,X_j) &= \sum_{i=R_1+1}^{R_2}\sum_{j=R_2+1}^{R_4-(R_3-R_2)} h(Y_i,Y_j)\\
&= \sum_{R_1<i<j\le R_4-(R_3-R_2)} h(Y_i,Y_j) - \sum_{R_1<i<j\le R_2} h(Y_i,Y_j)\\
&\quad - \sum_{R_2<i<j\le R_4-(R_3-R_2)} h(Y_i,Y_j),\\
&=: A_n^{(1)} - A_n^{(2)} - A_n^{(3)}
\end{aligned}$$

and

$$\sum_{R_2<i<j\le R_4-(R_3-R_2)} h(Y_i,Y_j) \overset{\mathcal{D}}{=} \sum_{R_1<i<j\le R_4-(R_3-R_2)-(R_2-R_1)} h(Y_i,Y_j),$$

for each fixed n. Using now Hoeffding's SLLN, we get the following convergence results for the U-statistics $A_n^{(1)}$, $A_n^{(2)}$ and $A_n^{(3)}$, as $n \to \infty$,

$$\begin{aligned}
\frac{A_n^{(1)}}{n^2} &\xrightarrow{P} \frac{(r_4-(r_3-r_2)-r_1)^2}{2}\theta_{l(r_2),l(r_4-(r_3-r_2))},\\
\frac{A_n^{(2)}}{n^2} &\xrightarrow{P} \frac{(r_2-r_1)^2}{2}\theta_{l(r_2),l(r_4-(r_3-r_2))},\\
\frac{A_n^{(3)}}{n^2} &\xrightarrow{P} \frac{(r_4-(r_3-r_2)-(r_2-r_1)-r_1)^2}{2}\theta_{l(r_2),l(r_4-(r_3-r_2))}.
\end{aligned}$$

Hence, as $n \to \infty$,

$$\frac{A_n^{(1)} - A_n^{(2)} - A_n^{(3)}}{n^2} \xrightarrow{P} (r_4 - r_3)(r_2 - r_1)\theta_{l(r_2),l(r_4-(r_3-r_2))}$$

and, therefore, (3.5) also holds, if two different blocks have the same distribution. Note that $\theta_{l(r_2),l(r_4)} = \theta_{l(r_2),l(r_4-(r_3-r_2))}$ when using the block of i.i.d. r.v.'s Y.□

Similarly to (3.4), define now

$$\begin{aligned} &S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \ldots, x_{s+1}, x_{s+4}, x_{s+5}, \ldots, x_{2s+3}) \\ &:= \sum_{i=[(n+1)x_1]+1}^{[(n+1)x_{s+2}]} \sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{2s+4}]} h(X_i, X_j) \\ &= \Big(\sum_{i=[(n+1)x_1]+1}^{[(n+1)x_2]} + \sum_{i=[(n+1)x_2]+1}^{[(n+1)x_3]} + \cdots + \sum_{i=[(n+1)x_{s+1}]+1}^{[(n+1)x_{s+2}]} \Big)\Big(\sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{s+4}]} \\ &\quad + \sum_{j=[(n+1)x_{s+4}]+1}^{[(n+1)x_{s+5}]} + \cdots + \sum_{j=[(n+1)x_{2s+2}]+1}^{[(n+1)x_{2s+3}]} + \sum_{j=[(n+1)x_{2s+3}]+1}^{[(n+1)x_{2s+4}]} \Big) h(X_i, X_j) \\ &= \sum_{p=1}^{s+1} \sum_{q=s+3}^{2s+3} \sum_{i=[(n+1)x_p]+1}^{[(n+1)x_{p+1}]} \sum_{j=[(n+1)x_q]+1}^{[(n+1)x_{q+1}]} h(X_i, X_j), \end{aligned}$$

where $0 \le x_1 < x_2 \le x_3 \le \cdots \le x_{s+2} \le x_{s+3} < x_{s+4} \le x_{s+5} \le \cdots \le x_{2s+4} < 1$. Moreover, since this is a sum of many double sums, (3.5) can be applied many times and as $n \to \infty$ we have the following convergence in probability result, namely

$$\begin{aligned} &\frac{1}{n^2} S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \ldots, x_{s+1}, x_{s+4}, x_{s+5}, \ldots, x_{2s+3}) \\ &\xrightarrow{P} \eta(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \ldots, x_{s+1}, x_{s+4}, x_{s+5}, \ldots, x_{2s+3}) \\ &\quad := \sum_{q=s+3}^{2s+3} (x_{q+1} - x_q) \sum_{p=1}^{s+1} (x_{p+1} - x_p)\theta_{l(x_{p+1}),l(x_{q+1})}. \end{aligned} \tag{3.6}$$

We note that if $\theta_{l(x_{r+1}),l(x_{q+1})} = \theta$ for all possible choices of r and q then $\eta(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \ldots, x_{s+1}, x_{s+4}, x_{s+5}, \ldots, x_{2s+3}) = (x_{2s+4} - x_{s+3})\,(x_{s+2} - x_1)\theta$. It should be clear that $x_2, \ldots, x_{s+1}$ and $x_{s+4}, \ldots, x_{2s+3}$ are dummy variables used to obtain the limiting result in (3.6).

We are now in the position to go back to the definition of $Z_{k_1,k_2,\ldots,k_s}$, $1 \le k_1 < k_2 \cdots < k_s < n$ in (1.4), and write it in terms of double sums $S(\cdot)$ as above. Thus, for $0 < t_1 < \cdots < t_s < 1$, we obtain

$$\begin{aligned} Z_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]} = \sum_{j=1}^{s} \sum_{i=j}^{s} S(&t_{j-1}, t_j, t_i, t_{i+1}, x_2, x_3, \ldots, x_{s+1}, \\ &x_{s+4}, x_{s+5}, \ldots, x_{2s+3}), \end{aligned}$$

where we define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$.

Since we do not know the location of the change-points $[n\lambda_1], [n\lambda_2], \ldots, [n\lambda_s]$, where $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_s < 1$, we define the following functions a_i, $1 \le i \le (s+1)s$, which will be used to derive a formula for the limiting function of (1.4) that will allow us to handle all possible combinations of $\lambda_1, \ldots, \lambda_s$ and

$t_1, \ldots, t_s$. Hence, we define

$$a_i \quad := \quad \begin{cases} \lambda_{i-m\cdot s}, & t_m < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{i-m\cdot s} \le t_{m+1}, \\ \lambda_{(i+1)-m\cdot s}, & \lambda_1 \le t_m < \lambda_2 \le \lambda_3 \le \cdots \le \lambda_{(i+1)-m\cdot s} \le t_{m+1}, \\ \vdots & \\ \lambda_s, & \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{s-1} \le t_m < \lambda_s \le t_{m+1}, \\ c_i, & \text{otherwise} \end{cases} \tag{3.7}$$

with $m := ((i-1) \mod s)$, $t_0 := 0$, $t_1 := 1$, $0 < a_1 \le a_2 \le \cdots \le a_{s\cdot s+s} < 1$ and $c_i \in [0,1]$, $1 \le i \le (s+1)s$. We note that the c_i's have to be chosen such that $(a_i)_{i=1}^{(s+1)s}$ is an increasing sequence and $c_i \neq \lambda_j$, $1 \le j \le s$.

We need to define these a_i's, $1 \le i \le (s+1)s$, since there are exactly $(s+1)s$ possibilities to place s change-points in $(s+1)$ blocks. Moreover, exactly s of the latter a_i's will change to one of the values λ_j, $1 \le j \le s$, while the other s^2 a_i 's will get the value c_i and they will drop out of the limiting function $u_{\lambda_1,\lambda_2,\ldots,\lambda_s}(t_1, t_2, \ldots, t_s)$ as defined below. The latter arguments yield the following theorem.

Theorem 3.1 *Assume that* (1.2), (3.1), (3.2), *and* $H_A^{(s)}$ *hold. Define* $t_0 := 0$ *and* $t_{s+1} := \frac{n}{n+1}$. *If* $\tau_i = \tau_i(n) := [n\lambda_i]$, $i = 1, 2, \ldots, s$, $0 < \lambda_1 \le \lambda_2 \le \cdots \le \lambda_s < 1$, *then, as* $n \to \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]} \xrightarrow{P} u_{\lambda_1,\lambda_2,\ldots,\lambda_s}(t_1, t_2, \ldots, t_s),$$

where for $0 < t_1 < t_2 < \cdots < t_s < 1$

$$u_{\lambda_1,\lambda_2,\ldots,\lambda_s}(t_1, t_2, \ldots, t_s) \quad = \quad \sum_{j=1}^{s} \sum_{i=j}^{s} \eta(t_{j-1}, t_j, t_i, t_{i+1}, a_{s(j-1)+1}, a_{s(j-1)+2}, \ldots, \\ a_{s(j-1)+s}, a_{si+1}, a_{si+2}, \ldots, a_{si+s}),$$

and η *and the* a_i *'s,* $1 \le i \le s(s+1)$, *are defined in* (3.6) *and* (3.7), *respectively.*

In particular, when we replace assumption (1.2) in Theorem 3.1 by (1.3), then the theorem is the same except that $\theta_{1,1} = \cdots = \theta_{s+1,s+1} = 0$ and $\theta_{q+1,r+1} = 0$, whenever the two blocks of random variables $X_{\tau_q+1}, \ldots, X_{\tau_{q+1}}$ and $X_{\tau_r+1}, \ldots, X_{\tau_{r+1}}$ have the same distribution.

We observe that if $\theta_{i,j} = \theta$ for all possible choices of $1 \le i \le j \le s+1$, then $u_{\lambda_1,\lambda_2,\ldots,\lambda_s}(t_1, t_2, \ldots, t_s) = \sum_{j=1}^{s} (t_{j+1} - t_j) t_j \theta$, which is the limit ($n \to \infty$) of the expected value of $\frac{1}{n^2} Z_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]}$. Moreover, it follows that under the null hypothesis H_0 of no-change $\frac{1}{n^2} U_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]} = 0$. Assuming that $\theta_{i,j} = \theta = 0$ for all possible choices of $1 \le i \le j \le n$, the sequence

$$T_n \quad = \quad \frac{1}{n^{3/2}\tilde{\sigma}} \sup_{0<t_1<t_2<\cdots<t_s<1} \left| U_{[(n+1)t_1],[(n+1)t_2],\ldots,[(n+1)t_s]} \right|, \quad n \in \mathbf{N},$$

is not consistent against any class of alternatives. On the other hand, if at least one $\theta_{i,j}$ is not equal to 0 and we use T_n, then

$$\mathbb{P}\{H_0 \text{ is rejected when using } T_n | H_A^{(s)} \text{ is true}\} \underset{n\to\infty}{\longrightarrow} 1.$$

This implies that the limits of the sequence $\{T_n\}_{n\in\mathbf{N}}$ are different in probability under H_0 and $H_A^{(s)}$, and hence we have consistency of $\{T_n\}_{n\in\mathbf{N}}$.

The limiting function $u_{\lambda_1,\lambda_2,\ldots,\lambda_s}(t_1, t_2, \ldots, t_s)$ in Theorem 3.1 looks quite complicated, hence we give some examples when $s = 1$ and $s = 2$ respectively.

Example 3.1 Let $s = 1$, $\lambda := \lambda_1$ and $t := t_1$. We observe that

$$a_1 \quad := \quad \begin{cases} \lambda, & 0 < \lambda \leq t, \\ c_1, & \text{otherwise}, \end{cases}$$

and

$$a_2 \quad := \quad \begin{cases} \lambda, & t < \lambda \leq 1, \\ c_2, & \text{otherwise}. \end{cases}$$

When $0 < t \leq \lambda$ then we have by Theorem 3.1 that

$$\begin{aligned} \eta(0,t,t,1,c_1,\lambda) &= (\lambda - t)(c_1\theta_{l(c_1),l(\lambda)} + (t-c_1)\theta_{l(t),l(\lambda)}) \\ &+ (1-\lambda)(c_1\theta_{l(c_1),l(1)} + (t-c_1)\theta_{l(t),l(1)}), \end{aligned}$$

where $l(t) = l(c_1) = l(\lambda) = 1$ and $l(1) = 2$, and when $\lambda \leq t < 1$ then

$$\begin{aligned} \eta(0,t,t,1,\lambda,c_2) &= (c_2 - t)(\lambda\theta_{l(\lambda),l(c_2)} + (t-\lambda)\theta_{l(t),l(c_2)}) \\ &+ (1-c_2)(\lambda\theta_{l(\lambda),l(1)} + (t-\lambda)\theta_{l(t),l(1)}), \end{aligned}$$

where $l(\lambda) = 1$ and $l(c_2) = l(t) = l(1) = 2$. Consequently, we have that

$$\frac{1}{n^2} Z_{[(n+1)t]} \quad \xrightarrow{P} \quad u_\lambda(t) \quad = \begin{cases} t(\lambda - t)\theta_{1,1} + t(1-\lambda)\theta_{1,2}, & 0 < t \leq \lambda, \\ (t-\lambda)(1-t)\theta_{2,2} + (1-t)\lambda\theta_{1,2}, & \lambda \leq t < 1, \end{cases}$$

and under H_0 it converges in probability to zero. This coincides with Theorem 3.1 by Csörgő and Horváth (1988b) where the case of at most one change-point is investigated.□

Example 3.2 Let $s = 2$. We observe that

$$a_1 \quad := \quad \begin{cases} \lambda_1, & 0 < \lambda_1 \le t_1, \\ c_1, & \text{otherwise}, \end{cases}$$

$$a_2 \quad := \quad \begin{cases} \lambda_2, & 0 < \lambda_1 \le \lambda_2 \le t_1, \\ c_2, & \text{otherwise}, \end{cases}$$

$$a_3 \quad := \quad \begin{cases} \lambda_1, & t_1 < \lambda_1 \le t_2, \\ \lambda_2, & \lambda_1 \le t_1 < \lambda_2 \le t_2, \\ c_3, & \text{otherwise}, \end{cases}$$

$$a_4 \quad := \quad \begin{cases} \lambda_2, & t_1 < \lambda_1 \le \lambda_2 \le t_2, \\ c_4, & \text{otherwise}, \end{cases}$$

$$a_5 \quad := \quad \begin{cases} \lambda_1, & t_2 < \lambda_1 \le 1, \\ \lambda_2, & \lambda_1 \le t_2 < \lambda_2 \le 1, \\ c_5, & \text{otherwise}, \end{cases}$$

and

$$a_6 \quad := \quad \begin{cases} \lambda_2, & t_2 < \lambda_1 \le \lambda_2 \le 1, \\ c_6, & \text{otherwise}. \end{cases}$$

Then by Theorem 3.1 we have under $H_A^{(s)}$ that

$$\begin{aligned} \frac{1}{n^2} Z_{[(n+1)t_1],[(n+1)t_2]} &\xrightarrow{P} u_{\lambda_1,\lambda_2} \\ &= \eta(0, t_1, t_1, t_2, a_1, a_2, a_3, a_4) + \eta(0, t_1, t_2, 1, a_1, a_2, a_5, a_6) \\ &\quad + \eta(t_1, t_2, t_2, 1, a_3, a_4, a_5, a_6) = \end{aligned}$$

$$
\begin{cases}
\Big((t_2-t_1)t_1+(\lambda_1-t_2)t_2\Big)\theta_{1,1}+(\lambda_2-\lambda_1)t_2\theta_{1,2} & \\
\quad +\ (1-\lambda_2)t_2\theta_{1,3}, & 0<t_1<t_2\le\lambda_1\le\lambda_2<1,\\
(\lambda_1-t_1)t_1\theta_{1,1}+\Big((t_2-\lambda_1)t_1+(\lambda_2-t_2)\lambda_1\Big)\theta_{1,2} & \\
\quad +\ (1-\lambda_2)\lambda_1\theta_{1,3}+(\lambda_2-t_2)(t_2-\lambda_1)\theta_{2,2} & \\
\quad +\ (1-\lambda_2)(t_2-\lambda_1)\theta_{2,3}, & 0<t_1\le\lambda_1<t_2\le\lambda_2<1,\\
(\lambda_1-t_1)t_1\theta_{1,1}+(\lambda_2-\lambda_1)t_1\theta_{1,2}+\Big((1-\lambda_2)t_1 & \\
\quad +\ (\lambda_1-t_1)(1-t_2)\Big)\theta_{1,3}+(\lambda_2-\lambda_1)(1-t_2)\theta_{2,3} & \\
\quad +\ (t_2-\lambda_2)(1-t_2)\theta_{3,3}, & 0<t_1\le\lambda_1\le\lambda_2<t_2<1,\\
(\lambda_2-t_1)\lambda_1\theta_{1,2}+\Big((\lambda_2-t_1)(t_1-\lambda_1) & \\
\quad +\ (t_2-t_1)(\lambda_2-t_2)\Big)\theta_{2,2}+(1-\lambda_2)(t_2-\lambda_1)\theta_{2,3} & \\
\quad +\ (1-\lambda_2)\lambda_1\theta_{1,3}, & 0<\lambda_1<t_1<t_2\le\lambda_2<1,\\
(\lambda_2-t_1)\lambda_1\theta_{1,2}+(1-\lambda_2)\lambda_1\theta_{1,3}+\Big((1-\lambda_2)(t_1-\lambda_1) & \\
\quad +\ (\lambda_2-t_1)(1-t_2)\Big)\theta_{2,3}+(t_1-\lambda_1)(\lambda_2-t_1)\theta_{2,2} & \\
\quad +\ (t_2-\lambda_2)(1-t_2)\theta_{3,3}, & 0<\lambda_1<t_1\le\lambda_2<t_2<1,\\
(1-t_1)\lambda_1\theta_{1,3}+(1-t_1)(\lambda_2-\lambda_1)\theta_{2,3} & \\
\quad +\ \Big((1-t_1)(t_1-\lambda_2)+(1-t_2)(t_2-t_1)\Big)\theta_{3,3}, & 0<\lambda_1\le\lambda_2<t_1<t_2<1,
\end{cases}
$$

and under H_0 it converges in probability to zero.□

4 Testing for changes in the mean

We are to test the *no-change in the mean* null-hypothesis

> H_0 *: $X_1,\ldots,X_n$ are independent identically distributed random variables with $\mathbb{E}X_i=\mu$ and $0<\sigma^2=VarX_i<\infty$, $1\le i\le n$,*

against the *at most s changes in the mean* alternative

> $H_A^{(s)}$ *: $X_1,\ldots,X_n$ are independent random variables and there are s integers $\tau_1,\tau_2,\ldots,\tau_s$, $1\le\tau_1\le\tau_2\le\cdots\le\tau_s<n$, such that $\mathbb{E}X_1=\cdots=\mathbb{E}X_{\tau_1}\ne\mathbb{E}X_{\tau_1+1}=\cdots=\mathbb{E}X_{\tau_2}$, $\mathbb{E}X_{\tau_1+1}=\cdots=\mathbb{E}X_{\tau_2}\ne\mathbb{E}X_{\tau_2+1}=\cdots=\mathbb{E}X_{\tau_3}$, $\ldots$, $\mathbb{E}X_{\tau_{s-1}+1}=\cdots=\mathbb{E}X_{\tau_s}\ne\mathbb{E}X_{\tau_s+1}=\cdots=\mathbb{E}X_n$, and $0<\sigma^2=VarX_i<\infty$, $1\le i\le n$.*

Let $h(x,y)=x-y$, then after some calculations (1.4) can be written as

$$
Z_{k_1,k_2,\ldots,k_s} \;=\; \sum_{i=1}^{s}(k_{i+1}-k_{i-1})S(k_i)-k_sS(n),
$$

where $k_0:=0$, $k_{s+1}:=n$ and $S(k):=\sum_{i=1}^{k}X_i$. We note that the same process was obtained by Orasch (1999b) using a geometrical argument instead of using U-statistics based processes. Furthermore, with $\tilde\sigma:=\sigma$, Theorem 2.3 yields, as $n\to\infty$,

$$
\sup_{0<t_1<t_2<\cdots<t_s<1}\frac{1}{n^{3/2}\sigma}\Big|Z_{k_1,k_2,\ldots,k_s}\Big| \;\xrightarrow{\mathcal{D}}\; \sup_{0<t_1<t_2<\cdots<t_s<1}\Big|\Gamma^a(t_1,t_2,\ldots,t_s)\Big|,
$$

with $t_0:=0$ and $t_{s+1}:=1$, and where the limiting process $\Gamma^a(t_1,\,t_2,\,\ldots,\,t_s)$ is defined in (2.9).

5 Testing for changes in the variance

We are to test the *no-change in the variance* hypothesis

> H_0 : *$X_1, \ldots, X_n$ are independent identically distributed random variables with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = Var X_i < \infty$, $1 \le i \le n$,*

against the *at most s changes in the variance* alternative

> $H_A^{(s)}$: *$X_1, \ldots, X_n$ are independent random variables and there exist s integers $\tau_1, \tau_2, \ldots, \tau_s$, $1 \le \tau_1 \le \tau_2 \le \ldots \le \tau_s < n$, such that $Var X_1 = \ldots = Var X_{\tau_1} \ne Var X_{\tau_1+1} = \ldots = Var X_{\tau_2}$, $Var X_{\tau_1+1} = \ldots = Var X_{\tau_2} \ne Var X_{\tau_2+1} = \ldots = Var X_{\tau_3}, \ldots, Var X_{\tau_{s-1}+1} = \ldots = Var X_{\tau_s} \ne Var X_{\tau_s+1} = \ldots = Var X_n$, $0 < Var X_1, Var X_{\tau_1+1}, Var X_{\tau_2+1}, \ldots, Var X_{\tau_s+1} < \infty$, and $\mathbb{E}X_1 = \cdots = \mathbb{E}X_n = \mu$.*

Let $h(x,y) = \frac{1}{2}(x-y)^2$, then after some algebraic manipulations, $Z_{k_1,k_2,\ldots,k_s}$ in (1.4) can be written as

$$\begin{aligned} Z_{k_1,k_2,\ldots,k_s} &= \frac{1}{2}\Big(\sum_{i=1}^{s}(k_{i+1}+k_{i-1}-2k_i)R(k_i) + k_s R(n) \\ &\quad - 2\sum_{i=1}^{s}(S(k_i)-S(k_{i-1}))(S(n)-S(k_i))\Big), \end{aligned}$$

where $k_0 := 0$, $k_{s+1} := n$, $S(k) := \sum_{i=1}^{k} X_i$ and $R(k) := \sum_{i=1}^{k} X_i^2$. Assume that $\mathbb{E}h^2(X_1,X_2) = \frac{1}{2}\mathbb{E}(X_1-\mu)^4 + 3(\sigma^2)^2$ is finite and $\tilde{\sigma}^2 = \frac{1}{4}Var((X_1-\mu)^2)$ is positive. Then we have by Theorem 2.1, as $n \to \infty$, that under H_0

$$\begin{aligned} &\sup_{0<t_1<t_2<\cdots<t_s<1} \frac{1}{n^{3/2}\tilde{\sigma}} \Big| Z_{k_1,k_2,\ldots,k_s} - \sigma^2 \sum_{i=1}^{s}(k_{i+1}-k_i)k_i \Big| \\ &\xrightarrow{\mathcal{D}} \sup_{0<t_1<t_2<\cdots<t_s<1} \Big| \sum_{i=1}^{s}(t_{i+1}+t_{i-1}-2t_i)W(t_i) + t_s W(1) \Big|, \end{aligned}$$

with $t_0 := 0$ and $t_{s+1} := 1$, and where the limiting process is the same as $\Gamma^{sym}(t_1, t_2, \ldots, t_s)$ from (2.5).

6 Testing for changes in mean and/or variance

We can base tests on the statistics considered in the previous sections to test for changes in the mean or changes in the variance, separately only. Frequently, it is of interest to be able to test for changes in both the mean and the variance. It turns out that this is not an easy task in general. Based on U-statistics-type processes, here we propose a statistic where tests can be based on, that will test for changes either in the mean or the variance or both. Unfortunately, in the following setup we are not able to distinguish between changes in both and changes in only one of them. Nevertheless, this test can be used when both depend on each other, namely the mean changes if and only if the variance changes. In case of independent normal variables Gombay and Horváth (1997) proposed an estimator for testing one single change in the mean and/or variance using the likelihood ratio test.

We are to test the *no-change in the mean and variance* hypothesis

> H_0 : *$X_1, \ldots, X_n$ are independent identically distributed random variables with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = Var X_i < \infty$, $1 \le i \le n$,*

against the *at most s changes in the mean and/or variance* alternative

$H_A^{(s)}$: *$X_1, \ldots, X_n$ are independent random variables and there exist s integers τ_1, τ_2, $\ldots$, τ_s, $1 \le \tau_1 \le \tau_2 \le \ldots \le \tau_s < n$, such that $\mathbb{E}X_1 = \ldots = \mathbb{E}X_{\tau_1} \ne \mathbb{E}X_{\tau_1+1} = \ldots = \mathbb{E}X_{\tau_2}$ and/or $VarX_1 = \ldots = VarX_{\tau_1} \ne VarX_{\tau_1+1} = \ldots = VarX_{\tau_2}$, $\mathbb{E}X_{\tau_1+1} = \ldots = \mathbb{E}X_{\tau_2} \ne \mathbb{E}X_{\tau_2+1} = \ldots = \mathbb{E}X_{\tau_3}$ and/or $VarX_{\tau_1+1} = \ldots = VarX_{\tau_2} \ne VarX_{\tau_2+1} = \ldots = VarX_{\tau_3}$, $\ldots$, $\mathbb{E}X_{\tau_{s-1}+1} = \ldots = \mathbb{E}X_{\tau_s} \ne \mathbb{E}X_{\tau_s+1} = \ldots = \mathbb{E}X_{\tau_n}$ and/or $VarX_{\tau_{s-1}+1} = \ldots = VarX_{\tau_s} \ne VarX_{\tau_s+1} = \ldots = VarX_n$, and $0 < VarX_1$, $VarX_{\tau_1+1}$, $\ldots$, $VarX_{\tau_s+1} < \infty$.*

Since under H_0

$$\mathbb{E}X_i^2 = VarX_i + (\mathbb{E}X_i)^2, \qquad 1 \le i \le n,$$

it is reasonable to consider symmetric kernels of the form

$$h(x,y) = \frac{x^2+y^2}{2}.$$

Consequently, under H_0, $h(X_i, X_j)$ is an unbiased estimator for $\theta = \mu^2 + \sigma^2$. It is obvious that changes in μ or σ^2 or in both will change θ. By using this kernel function, we can not distinguish which parameter changed. Nevertheless it may be used to detect, if there were any changes at all in any one, or in both of these parameters.

To apply our theory on U-statistic based processes, we assume that under H_0

$$\mathbb{E}h^2(X_i, X_j) = \frac{1}{2}Var(X_i^2) + (\mu^2+\sigma^2)^2 < \infty,$$

and put

$$\begin{aligned} \tilde{h}(t) &:= \mathbb{E}\{h(t, X_2) - (\mu^2+\sigma^2)\} \\ &= \frac{1}{2}(t^2 - (\mu^2+\sigma^2)). \end{aligned}$$

Furthermore, we also assume that under H_0

$$\tilde{\sigma}^2 := \mathbb{E}\tilde{h}^2(X_1) = \frac{1}{4}Var(X_i^2)$$

is positive and finite. Hence, after some algebraic manipulations, $Z_{k_1,k_2,\ldots,k_s}$ in (1.4) with the kernel from above can be written as

$$Z_{k_1,k_2,\ldots,k_s} = \sum_{i=1}^{s}(k_{i+1} + k_{i-1} - 2k_i)R(k_i) + k_sR(n),$$

where $k_0 := 0$, $k_{s+1} := n$ and $R(k) := \sum_{i=1}^{k} X_i^2$. By Theorem 2.1 we have that under H_0, as $n \to \infty$,

$$\begin{aligned} &\sup_{0<t_1<t_2<\cdots<t_s<1} \frac{1}{n^{3/2}\tilde{\sigma}} \Big| Z_{k_1,k_2,\ldots,k_s} - (\mu^2+\sigma^2)\sum_{i=1}^{s}(k_{i+1}-k_i)k_i \Big| \\ &\xrightarrow{\mathcal{D}} \sup_{0<t_1<t_2<\cdots<t_s<1} \Big| \sum_{i=1}^{s}(t_{i+1}+t_{i-1}-2t_i)W(t_i) + t_sW(1)\Big|, \end{aligned}$$

with $t_0 := 0$ and $t_{s+1} := 1$, and where the limiting process is the same as $\Gamma^{sym}(t_1, t_2, \ldots, t_s)$ from (2.5).

Acknowledgments

I am very grateful to Professor Miklós Csörgő for several discussions and his supervision of my PhD thesis Orasch (1999a) on which the present paper is based. In addition I would like to thank Professor Lajos Horváth, University of Utah, for pointing out that results in this paper I originally obtained for random variables also hold for random vectors.

References

[1] Brodsky, B.E. and Darkhovsky, B.S. (1993). *Nonparametric Methods in Change-Point Problems.* Kluwer, Dordrecht.

[2] Csörgő, M. and Horváth, L. (1988a). Nonparametric methods for changepoint problems. In: P.R. Krishnaiah and C.R. Rao, Eds., *Handbook of Statistics*, Vol. 7 Elsevier, Amsterdam, 403-425.

[3] Csörgő, M. and Horváth, L. (1988b). Invariance principles for changepoint problems. *Journal of Multivariate Analysis* **27**, 151-168.

[4] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-point Analysis.* John Wiley, Chichester.

[5] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[6] Erasmus, C.M. and Lombard, F. (1988). Asymptotic distributions of quadratic forms occurring in changepoint problems. *Canad. J. Statist* **16**, 259-268.

[7] Gombay, E. and Horváth, L. (1997). An application of the likelihood method to change-point detection. *Environmetrics* **8**, 459-467.

[8] Gombay, E. and Horváth, L. (1999). Rates of convergence for U-statistic processes and their bootstrapped versions. Preprint.

[9] Hall, P. (1979). On the invariance principle for U-statistics. *Stochastic Process. Appl.* **9**, 163-174.

[10] Hoeffding, W. (1961). The strong law of large numbers for U-statistics. Univ. of North Carolina Institute of Statistics Mimeo Series, No. 302.

[11] Janson, S. and Wichura, M.J. (1983). Invariance principles for stochastic area and related stochastic integrals. *Stochastic Process. Appl.* **16**, 71-84.

[12] Koroljuk, V.S. and Borovskich, Yu.V. (1994). *Theory of U-Statistics.* Kluwer Academic Publishers, Dordrecht.

[13] Major, P. (1979). An improvement of Strassen's invariance principle. *Ann. Probab.* **7**, 55-61.

[14] Orasch, M. (1999a). Multiple Change-points with an Application to Financial Modeling. PhD Thesis,*Carleton University*, Ottawa, Canada.

[15] Orasch, M. (1999b). Testing multiple changes in the mean. To appear in *Bolyai Society Mathematical Studies*, Volume 11: Limit Theorems in Probability and Statistics, Janos Bolyai Mathematical Society, Budapest.

[16] Sen, P.K. (1977). Almost sure convergence of generalized U-statistics. *Ann. Probab.* **5**, 287-290.

[17] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley, Chichester.

[18] Szyszkowicz, B. (1992). Weak Convergence of Stochastic Processes in Weighted Metrics and their Applications to Contiguous Changepoint Analysis. PhD Thesis, *Carleton University*, Ottawa, Canada.

[19] Szyszkowicz, B. (1998). Weighted sequential empirical type processes with applications to changepoint problems. In: N. Balakrishnan and C.R. Rao, Eds., *Handbook of Statistics* Vol. 16, 573-630.