Marie Hušková; Aleš Slabý

Permutation tests for multiple changes

Kybernetika, Vol. 37 (2001), No. 5, [605]--622

Persistent URL: http://dml.cz/dmlcz/135430

KYBERNETIKA — VOLUME 37 (2001), NUMBER 5, PAGES 605–622

PERMUTATION TESTS FOR MULTIPLE CHANGES[1]

MARIE HUŠKOVÁ AND ALEŠ SLABÝ

Approximations to the critical values for tests for multiple changes in location models are obtained through permutation tests principle. Theoretical results say that the approximations based on the limit distribution and the permutation distribution of the test statistics behave in the same way in the limit. However, the results of simulation study show that the permutation tests behave considerably better than the corresponding tests based on the asymptotic critical value.

1. INTRODUCTION

Consider the model for multiple changes in the location model:

$$X_i = \sum_{j=0}^{q} \mu_j I\{m_j < i \le m_{j+1}\} + e_i, \quad i = 1, \dots, n \tag{1.1}$$

where $0 = m_0 < m_1 \le \ldots \le m_{q+1} = n$ and $\mu_0, \ldots, \mu_q$, are unknown parameters fulfilling $\mu_j \ne \mu_{j+1}$, $j = 0, \ldots, q-1$; q can be known or unknown. The observations $X_1, \ldots, X_n$ are obtained in some time ordered points $t_1 < \cdots < t_n$. The error terms $e_1, \ldots, e_n$ are assumed to follow the assumption:

$$\begin{aligned} &e_1, \ldots, e_n \text{ are independent identically distributed,} \\ &Ee_i = 0,\ 0 < \operatorname{var} e_i < \infty,\ E|e_i|^{2+\Delta} < \infty \text{ with some } \Delta > 0. \end{aligned} \tag{1.2}$$

In this context, the values $m_1, \ldots, m_q$ are change points and the respective differences $\mu_{j+1} - \mu_j$, $j = 0, \ldots, q-1$, are magnitudes of the changes.

We are testing the null hypothesis

$$H_0 : m_1 = \ldots = m_q = n \tag{1.3}$$

against the alternative that at least one change has occurred:

$$H_1 : m_1 \le \ldots \le m_q, \text{ where at least one inequality is strict,} \tag{1.4}$$

[1]Partially supported by Grant 201/00/0769 of the Grant Agency of the Czech Republic and by the MSM 113200008 Project.

where q can be or need not be specified.

There is a number of tests available for this testing problem, for information see, e. g., Csörgő and Horváth [5] and Horváth and Kokoszka [10]. One of the main problems is to find reasonable approximations to the critical values. Typically, approximations based on limit behavior of the test statistics under the null hypothesis are used. However, the convergence to the limit distributions of the test statistics for the change point problem is rather slow and therefore these approximations are reasonable only for very large sample sizes and, usually, the resulting tests are conservative otherwise. Csörgő and Horváth [5], among others, pointed out this fact and proposed an improvement. This is based on asymptotic arguments combined with a proper trimming.

In the present paper we focus on the test statistics generated by a kernel function K. Their limit behavior under the null hypothesis will be derived. Particular attention will be paid to the related permutation tests. An approximation to the critical values through the bootstrap method will be also discussed.

We assume that the kernel K satisfies either assumptions (K.1) and (K.2) or (K.1) and (K.3):

(K.1) K is a non-negative symmetric function such that

$$K(t) = 0, \quad t \notin [-1, 1], \quad \int_{-1}^{1} K(t)\,\mathrm{d}t > 0.$$

(K.2) The second derivative $K^{(2)}$ exists and is Lipschitz of order $\beta \geq \frac{1}{2}$ on $(0, 1)$, one sided second order derivatives exist at 0 and 1 and $K(0) + K(1) > 0$.

(K.3) The second derivative $K^{(2)}$ exists and is Lipschitz of order 1 on $(0, 1)$, one sided second order derivatives exist at 0 and 1 and $K(0) + K(1) = 0$.

Notice that the set of assumptions (K.1) and (K.2) covers the situation where at least one of $K(0)$ and $K(1)$ are nonzero while the set (K.1)and (K.3) corresponds to the case when $K(0) = K(1) = 0$. Inside of the interval (0,1) and $(-1, 0)$ the kernels are assumed to be smoothed. Both sets of assumptions imply that K is bounded and therefore $\int_{-1}^{1} K(u)\,\mathrm{d}u < \infty$.

The test procedure also depends on G which is related to the bandwidth in the area of nonparametric regression. We assume that $G = G(n)$ satisfies, as $n \to \infty$,

$$\frac{G}{n} \log(n/G) \to 0, \quad \frac{G}{n^{2/(2+\Delta)}} \to \infty \tag{1.5}$$

which means that G tends to infinity together with n but not too fast.

We consider the test statistic

$$\begin{aligned} T_n(G) \;\; = \;\; & \max_{G<k<n-G} \frac{1}{\sqrt{2\sum_{i=1}^{G} K^2(i/G)}} \frac{1}{\sigma_n} \times \\ & \times \left| \sum_{i=k-G+1}^{k} X_i K\left(\frac{k-i+1}{G}\right) - \sum_{i=k+1}^{k+G} X_i K\left(\frac{k-i}{G}\right) \right| \end{aligned} \tag{1.6}$$

where

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2, \tag{1.7}$$

which is an estimator of $var\, e_i$ defined in (1.2).

Motivation for this test statistic comes from the area of nonparametric regression estimation. Notice that

$$N_n^-(k/G) = \frac{1}{G}\sum_{i=k-G+1}^{k} X_i\, K\left(\frac{k-i+1}{G}\right) \tag{1.8}$$

and

$$N_n^+(k/G) = \frac{1}{G}\sum_{i=k+1}^{k+G} X_i\, K\left(\frac{k-i}{G}\right) \tag{1.9}$$

are kernel type estimators of the expectation $E\, X_k$ based on the observations $X_{k-G+1}, \ldots, X_k$ (i. e., G observations till time point t_k) and $X_{k+1}, \ldots, X_{k+G}$ (i. e., G observations after time point t_k), respectively. Then $T_n(G)$ can be expressed as

$$T_n(G) = \max_{G<k<n-G} \frac{G}{\sqrt{2\sum_{i=1}^{G} K^2(i/G)}}\frac{1}{\sigma_n}\left|N_n^-(k/G) - N_n^+(k/G)\right|. \tag{1.10}$$

Clearly, large values of $T_n(G)$ indicate that at least one change has occurred. Possible approximation to the critical value follows from Theorem 2.1 below where the limit distribution of $T_n(G)$ under H_0 is stated. The test based on $T_n(G)$ is sensitive w.r.t. to a wide spectrum of alternatives. Moreover, the differences $N_n^+(k/G) - N_n^-(k/G)$, $k = G+1, \ldots, n-G$, can be helpful to identify the change points m_j.

The limit behavior of $T_n(G)$ under the null hypothesis H_0 is studied, the consistency of the tests based on $T_n(G)$ is proved and various modifications of this test statistic are discussed in Section 2. The permutation tests related to $T_n(G)$ are developed and investigated in Section 3. Section 4 contains results of a simulation study.

2. LIMIT PROPERTIES OF $T_n(G)$

Here we derive the limit behavior of $T_n(G)$ under the null hypothesis (no change), prove the consistency of the related test and discuss possible modifications and extensions. The main assertion of this section states:

Theorem 2.1. (no change) Let $X_1, \ldots, X_n$ be i.i.d. random variables with nonzero variance and $E\,|X_i|^{2+\Delta} < \infty$ with some $\Delta > 0$. Let (1.5) be satisfied.

(i) If the kernel K satisfies (K.1), (K.2) then, as $n \to \infty$,

$$P\Big(\sqrt{2\log(n/G)}\, T_n(G) \le y + 2\log(n/G) + \frac{1}{2}\log\log(n/G)$$
$$+ \log\frac{2K^2(0) + K^2(1)}{2\int_0^1 K^2(t)\,\mathrm{d}t} - \frac{1}{2}\log(\pi)\Big) \to \exp\{-2\exp\{-y\}\}, \quad y \in R^1,$$

where $T_n(G)$ is defined in (1.6).

(ii) If the kernel K satisfies (K.1), (K.3) then, as $n \to \infty$,

$$P\left(\sqrt{2\log(n/G)}\,T_n(G) \le y + 2\log(n/G) + \frac{1}{2}\log\frac{\int_0^1 (K'(t))^2\,\mathrm{d}t}{4\int_0^1 K^2(t)\,\mathrm{d}t} - \log(\pi)\right)$$
$$\to \exp\{-2\exp\{-y\}\}, \quad y \in R^1.$$

Proof. The proof is divided into three steps. In the first one we show that it is sufficient to study the the limit behavior of $T_n(G)$ for X_i being i.i.d. with $N(0,1)$ distribution. Then we prove that properly standardized $N_n^-(k/G) - N_n^+(k/G)$, $k = 1,\dots,n$, defined by (1.8) through (1.9), converge to a Gaussian process and, finally, applying the results on the extremes of Gaussian processes we get the desired results.

Without loss of generality we may assume that X_i have zero mean and unit variance.

Denoting

$$M_n(j) = \sum_{i=1}^{j} X_i, \quad j = 1,\dots,n, \tag{2.1}$$

we find that

$$\sum_{i=k+1}^{k+G} X_i K\left(\frac{k-i}{G}\right) = \big(M_n(k+G) - M_n(k)\big)K(0) \tag{2.2}$$
$$+\sum_{j=1}^{G}\big(M_n(k+G) - M_n(k+j-1)\big)\big(K(-j/G) - K(-(j-1)/G)\big)$$

and a similar expression for $\sum_{i=k-G+1}^{k} X_i K\big(\frac{k-i+1}{G}\big)$ can be derived via the partialsums $M_n(k)$. By arguments of Einmahl [7] there are Wiener processes $\{W_n(t),\ 0 \le t < \infty\}$, $n = 1,\dots$, such that, as $n \to \infty$,

$$\max_{1\le k\le n} k^{1/(2+\Delta)}\big|M_n(k) - W_n(k)\big| = O_p(1) \tag{2.3}$$

that immediately implies

$$\max_{G<k\le n-G}\max_{0\le i\le G}\big|\big(M_n(k+G) - M_n(k+i)\big) - \big(W_n(k+G) - W_n(k+i)\big)\big|$$
$$= O_P\big(G^{1/(2+\Delta)}\big). \tag{2.4}$$

By Theorem 1.2.1 of Csörgő and Révész [4] we have

$$\sup_{0\le t\le n-G}\sup\nolimits_{0\le s\le G}\big|W_n(t+s) - W_n(t)\big|$$
$$= O_P\big((G\log(n/G))^{1/2}\big) + O_P\big((G\log\log n))^{1/2}\big). \tag{2.5}$$

Then by the assumptions the kernel K has finite variation and then combining (2.1) through (2.5) we observe that

$$T_n(G) = O_p\big(\sqrt{\log(n/G)}\big)$$

and moreover, that it suffices to derive the limit distribution of $T_n(G)$ for the case $X_1, \dots, X_n$ being i.i.d. with $N(0,1)$ distribution.

Hence it remains to derive the limit behavior of

$$\max_{G<k<n-G} \big|L_n(k)\big|$$

where

$$\begin{aligned} L_n(k) \quad &= \quad \frac{1}{\sqrt{2\sum_{i=1}^{G} K^2(i/G)}} \times \\ &\times \left(\sum_{i=k-G+1}^{k} X_{Ni} K\left(\frac{k-i+1}{G}\right) - \sum_{i=k+1}^{k+G} X_{Ni} K\left(\frac{k-i}{G}\right) \right) \end{aligned} \tag{2.6}$$

with $X_{N1}, \dots, X_{Nn}$ being i.i.d. with $N(0,1)$. Notice that $L_n(k)$, $k = G, \dots, n-G$, is a stationary $2G$-dependent sequence of random variables with distribution with zero mean, unit variance and covariances for $G \le k_1 < k_2 \le n-G$, $k_2 - k_1 \le G$

$$\begin{aligned} \mathrm{cov}\{L_n(k_1),\, L_n(k_2)\} &= \frac{1}{2\sum_{i=1}^{G} K^2(i/G)} \left(\sum_{i=k_2-G+1}^{k_1} K\left(\frac{k_1-i+1}{G}\right) K\left(\frac{k_2-i+1}{G}\right) \right. \\ &+ \left. \sum_{i=k_2+1}^{k_1+G} K\left(\frac{k_1-i}{G}\right) K\left(\frac{k_2-i}{G}\right) - \sum_{i=k_1+1}^{k_2} K\left(\frac{k_1-i}{G}\right) K\left(\frac{k_2-i+1}{G}\right) \right). \end{aligned}$$

Define the process

$$Y_n(t) = L_n\big([Gt]\big), \quad 1 \le t \le n/G - 1.$$

Using standard arguments one can show that

$$\{Y_n(t), 1 \le t \le n/G - 1\} \to \{Y(t), 1 \le t < \infty\}$$

where

$$Y(t) = \left(\int_t^{t+1} K(y-t)\, \mathrm{d}W(y) - \int_{t-1}^{t} K(y-t)\, \mathrm{d}W(y) \right) \frac{1}{\sqrt{2\int_0^1 K^2(u)\, \mathrm{d}u}}, \quad t \ge 1,$$

with $\{W(t), 0 \le t < \infty\}$ being a Wiener process. The process $\{Y(t); 1 \le t < \infty\}$ is a stationary Gaussian process with unit variance and the autocorrelation function $\rho(v) = cov(Y(t+v), Y(t))$. It has the property

$$\rho(v) = 0, \quad v > 2,$$

and for $1 > v > 0$

$$\begin{aligned}\rho(v) \;=\; 1+ & \Bigg(\int_v^{1-v} K(z)(K(z+v) - K(z-v) - 2K(z))\,\mathrm{d}z \\ & + \int_0^v K(z)(K(z+v) - K(z-v) - 2K(z))\,\mathrm{d}z \\ & + \int_{1-v}^1 K(z)(K(z-v) - 2K(z))\,\mathrm{d}z \Bigg) \left(2\int_0^1 K^2(z)\,\mathrm{d}z \right)^{-1}.\end{aligned}$$

To finish the proof we have to check the behavior of the covariance for $v \to 0$. The limit depends on the assumptions on the kernel K. Under the assumptions (K.1), (K.2) we receive that for $v \to 0+$

$$\rho(v) = 1 - v\frac{2K^2(0) + K^2(1)}{2\int_0^1 K^2(z)\,\mathrm{d}z} + o(v)$$

and under assumptions (K.1), (K.3) we obtain that for $v \to 0+$

$$\rho(v) = 1 - v^2\frac{\int_0^1 (K'(z))^2\,\mathrm{d}z}{2\int_0^1 K^2(z)\,\mathrm{d}z} + o(v^2).$$

Then applying Theorem 12.3.5 in Leadbetter et al [13] we receive the desired assertions. □

Remark 2.1. Going through the paper by Einmahl [7] we find that (2.3) holds true even for triangular array, i. e., (2.3) remains true if in the definition of $M_n(k/G)$ (see (2.1)) $X_1, \dots, X_n$ are replaced by $X_{1n}, \dots, X_{nn}$ that are i.i.d. with zero mean, unit variance and $E\,|X_{1n}|^{2+\Delta} \le D_2 > 0$, $n \ge n_0$, where $D_2 > 0$ does not depend on n.

Remark 2.2. Notice that $T_n(G)$ will not change if X_i is replaced by the residual $X_i - \overline{X}_n$, $i = 1, \dots, n$, where $\overline{X}_n = \sum_{i=1}^n X_i/n$. So that $T_n(G)$ can be rewritten as a functional of these residuals. Then one can develop corresponding M- and R-test statistics. They are obtained from $T_n(G)$ just replacing the residual $X_i - \overline{X}_n$ and the estimator σ_n^2 by their M- or R-counterparts. It can be shown that under the null hypothesis and under some assumptions on scores and score function they have the same limit behavior as $T_n(G)$.

Remark 2.3. The critical region of the test (1.3) versus (1.4) based on $T_n(G)$ on the level α has the form

$$T_n(G) > z_{1-\alpha,n}(G) \tag{2.7}$$

where $z_{1-\alpha,n}(G)$ is the $100(1-\alpha)\,\%$ quantile of the distribution of $T_n(G)$ under H_0. Theorem 2.1 provides an approximation to $z_{1-\alpha,n}(G)$ and implies that

$$z_{1-\alpha,n}(G) = 2\sqrt{\log\frac{n}{G}}(1 + o(1)). \tag{2.8}$$

Next, we study the limit behavior of $T_n(G)$ under alternatives. Namely, we assume that in the model (1.1) with $\mu_0, \dots, \mu_q$, $\mu_i \neq \mu_{i+1}$ and q are fixed (not dependent on n and that m_i, $i = 0, \dots, q$ increase together with n, namely,

$$m_i/n \to \kappa_i, \quad i = 0, \dots, q+1, \quad 0 = \kappa_0 < \kappa_1 < \dots < \kappa_{q+1} = 1. \tag{2.9}$$

Theorem 2.2 Let $X_1, \dots, X_n$ follow the model (1.1) with $\mu_0, \dots, \mu_q$, $\mu_i \neq \mu_{i+1}$ and q fixed. Let (1.5) and (2.9) be satisfied. Then the test with the critical region (2.7) is consistent.

Proof. Standard tools give

$$\begin{aligned} T_n(G) = & \max_{i=0,\dots,q-1} \frac{|\mu_{i+1} - \mu_i|}{\sqrt{\sigma^2 + \sum_{i=0}^q (\mu_i - \overline{\mu})^2 (\kappa_{i+1} - \kappa_i)}} \frac{\sqrt{G} \int_0^1 K(u)\,du}{\sqrt{2 \int_0^1 K^2(u)\,du}} \\ & + O_P\left(\sqrt{\log(n/G)}\right) \end{aligned} \tag{2.10}$$

where $\overline{\mu} = \sum_{i=0}^q \mu_i(\kappa_{i+1} - \kappa_i)$. This together with (2.9) and the assumption (1.5) implies the consistency. □

Remark 2.4. Studying the limit behavior of the test based on $T_n(G)$ under various alternatives we find that it is sensitive w.r.t. a wide spectrum of alternatives (multiple abrupt changes, gradual changes).

Remark 2.5. The statistics $N_n^-(k/G) - N_n^+(k/G)$, $k = G, \dots, n-G$ can be used to estimate the change points $m_1, \dots, m_q$, for details see Grabowsky et al [9].

Next, we study the R-type (rank based) version of the test statistics of $T_n(G)$. It will appear to be extremely useful when studying the permutation tests in Section 3.

The rank based version of $T_n(G)$ is defined as

$$\begin{aligned} T_{n,Q}(G) = & \max_{G<k<n-G} \frac{1}{\sqrt{2\sum_{i=1}^G K^2(i/G)}} \frac{1}{\tau_n} \times \\ & \times \left| \sum_{i=k-G+1}^{k} a_n(Q_i) K\left(\frac{k-i+1}{G}\right) - \sum_{i=k+1}^{k+G} a_n(Q_i) K\left(\frac{k-i}{G}\right) \right| \end{aligned} \tag{2.11}$$

where $Q_1, \dots, Q_n$ are the ranks of $X_1, \dots, X_n$; $a_n(1), \dots, a_n(n)$ are scores and

$$\tau_n^2 = \frac{1}{n-1} \sum_{i=1}^n \left(a_n(i) - \overline{a}_n\right)^2, \quad \overline{a}_n = \frac{1}{n} \sum_{i=1}^n a_n(i). \tag{2.12}$$

Theorem 2.3. Let $X_1, \ldots, X_n$ be i.i.d. random variables with continuous distribution function. Let the scores $a_n(1), \ldots, a_n(n)$ have the properties

$$\tau_n^2 \geq D_1, \quad n \geq n_0 \tag{2.13}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \left|a_n(i) - \overline{a}_n\right|^{2+\Delta_1} \leq D_2, \quad n \geq n_0 \tag{2.14}$$

with some positive D_1, D_2, n_0 and Δ_1. Then the assertion of Theorem 2.1 remains true if $T_n(G)$ is replaced by $T_{n,Q}(G)$.

P r o o f. To prove the assertion it is sufficient to show that $T_{n,Q}(G)$ is close to $T_n(G)$ with X_i replaced by suitable random variables Z_i that are i.i.d. and are fulfilling the assumptions of Theorem 2.1.

Notice that the ranks $Q_1, \ldots, Q_n$ can be viewed as the ranks corresponding to the random sample $U_1, \ldots, U_n$ from uniform distribution on $(0,1)$, where $U_i = F(X_i)$, $i = 1, \ldots, n$ and F is the distribution of X_i under the null hypothesis. We introduce the simple linear rank statistics

$$S_k = \sum_{i=1}^{k} \left(a_n(Q_i) - \overline{a}_n\right), \quad k = 1, \ldots, n, \tag{2.15}$$

and the accompanying partial sums of i.i.d. random variables

$$Z_k = \sum_{i=1}^{k} \left(a_n(1 + [nU_i]) - \overline{a}_n\right), \quad k = 1, \ldots, n. \tag{2.16}$$

Direct calculation gives

$$ES_k = EZ_k = 0, \quad k = 1, \ldots, n, \tag{2.17}$$

$$\mathrm{var} S_k = \frac{n}{n-1}\mathrm{var}\left\{Z_k - \frac{k}{n}Z_n\right\} = \frac{k(n-k)}{n}\tau_n^2, \quad k = 1, \ldots, n, \tag{2.18}$$

and

$$E\left|a_n(1 + [nU_i]) - \overline{a}_n\right|^{2+\Delta_1} \leq D_2. \tag{2.19}$$

Then the assumptions of Theorem 3 in Hušková [11] are satisfied and this theorem implies that, as $n \to \infty$,

$$\max_{G \leq k \leq n-G} \frac{1}{\sqrt{G}} \left|S_k - \left(Z_k - \frac{k}{n}Z_n\right)\right| = O_P(n^{-v}) \tag{2.20}$$

with some $v > 0$. Replacing X_i by $\big(a_n(Q_i) - (a_n([U_i n] + 1) - Z_n/n)\big)$ in (2.2) we get

$$\begin{aligned}
&\sum_{i=k-G+1}^{k} \Big(a_n(Q_i) - \big(a_n([nU_i] + 1) - Z_n/n\big)\Big)\, K\left(\frac{k-i+1}{G}\right) \\
&= \sum_{j=1}^{G} \Big(\big(S_{k+G} - S_{k+j-1}\big) - \big(Z_{k+G} - Z_{k+j-1} - (G-j+1)Z_n/n\big)\Big) \times \\
&\qquad \times \big(K(-j/G) - K(-(j-1)/G)\big) \\
&\quad + \big((S_{k+G} - S_k) - (Z_{k+G} - Z_k - GZ_n/n)\big)\, K(0).
\end{aligned}$$

After few standard steps these relations together with (2.21) imply that, as $n \to \infty$,

$$\begin{aligned}
&\max_{G<k\le n-G} \frac{1}{\sqrt{G}} \left| \sum_{i=k-G+1}^{k} \Big(a_n(Q_i) - (a_n([nU_i]+1) - Z_n/n)\Big) K\left(\frac{k-i+1}{G}\right) \right| \\
&= o_P\big((\log(n/G))^{-1/2}\big)
\end{aligned} \tag{2.21}$$

and similarly, as $n \to \infty$,

$$\begin{aligned}
&\max_{G<k\le n-G} \frac{1}{\sqrt{G}} \left| \sum_{i=k+1}^{k+G} \Big(a_n(Q_i) - (a_n([nU_i]+1) - Z_n/n)\Big) K\left(\frac{k-i}{G}\right) \right| \\
&= o_P\big((\log(n/G))^{-1/2}\big).
\end{aligned} \tag{2.22}$$

The random variables $a_n([nU_i] + 1)$, $i = 1, \dots, n$ satisfy the assumptions of Theorem 2.1 and Z_n/n cancels from the corresponding statistic $T_n(G)$, so that Theorem 2.1 holds true if X_i is replaced by $a_n([nU_i]+1) - Z_n/n$ for $i = 1 \dots, n$. Then the assertion of our theorem can be concluded if Remark 2.1 is taken into account and (2.22) and (2.23) are applied. □

The problem of the choice of the kernel K and of the limit behavior under alternatives will be considered in a different paper.

3. PERMUTATION TESTS

Here we describe the permutation test related to the statistic $T_n(G)$ and study its limit performance.

Elements of theory of permutation tests together with relevant references can be found in the books by Lehmann [12] and by Good [6] among others. Rather general remarks on the permutation tests in change point analysis can be found in Romano [14]. Antoch and Hušková [3] investigated permutation tests for at most one change in location model.

The permutation distribution of $T_n(G)$ can be described as the conditional distribution, given $X_1, \dots, X_n$, of

$$T_n(\boldsymbol{R}, G) = \max_{G<k<n-G} \frac{1}{\sqrt{2\sum_{i=1}^{G} K^2(i/G)}} \frac{1}{\sigma_n} \tag{3.1}$$

$$n \times \Big| \sum_{i=k-G+1}^{k} X_{R_i} K(\frac{k-i+1}{G}) - \sum_{i=k+1}^{k+G} X_{R_i} K(\frac{k-i}{G}) \Big|,$$

where $\boldsymbol{R} = (R_1, \dots, R_n)$ is random permutation of $(1, \dots, n)$ independent of $X_1, \dots$ $\dots, X_n$. This permutation distribution $F_n(.\,; \boldsymbol{X})$ can be expressed as

$$F_n(x\,; \boldsymbol{X}) = \frac{1}{n!} \#\{\boldsymbol{r} \in \mathcal{R}_n;\ T_n(\boldsymbol{r}, G) \le x\}, \quad x \in R^1, \tag{3.2}$$

where $\mathcal{R}_n$ is the set of all permutations of $(1, \dots, n)$ and $\#A$ denotes the cardinality of a set A. Denoting by $x_{1-\alpha,n}(G, \boldsymbol{X})$ the corresponding $100(1-\alpha)\%$ quantile the critical region of the permutation test based on $T_n(G)$ with the level α has the form

$$T_n(G) \ge x_{1-\alpha,n}(G, \boldsymbol{X}). \tag{3.3}$$

Computational aspects of the critical values $x_{1-\alpha,n}(G, \boldsymbol{X})$ are discussed in Appendix.

Next, we derive the limit distribution of the permutation distribution $F_n(x\,; \boldsymbol{X})$ which is the main result of the paper.

Theorem 3.1. Let the observations $X_1, \dots, X_n$ follow the model (1.1). Let the assumptions (1.2), (1.5) be satisfied, let $q \le D_3$

$$|\mu_j| \le D_4 > 0, \quad j = 0, \dots, q, \tag{3.4}$$

with some $D_3 > 0$ and $D_4 > 0$ be satisfied. Let (1.5) be satisfied.

(i) If the kernel K satisfies (K.1), (K.2) then, as $n \to \infty$,

$$\begin{aligned} &P\Big(\sqrt{2\log(n/G)}\, T_n(\boldsymbol{R}, G) \le y + 2\log(n/G) + \frac{1}{2}\log\log(n/G) \\ &+ \log\left(\frac{2K^2(0) + K^2(1)}{2\int_0^1 K^2(t)\,dt}\right) - \frac{1}{2}\log(\pi) \Big| X_1, \dots, X_n\Big) \\ &\to \exp\{-2\exp\{-y\}\}, \quad [P]\text{-a.s.}, \end{aligned}$$

where $P(A|X_1, \dots, X_n)$ denotes the conditional probability of an event A given $X_1, \dots, X_n$ and $[P]$–a.s. denotes probability measure generated by random variables X_i's.

(ii) If the kernel K satisfies (K.1), (K.3) then, as $n \to \infty$,

$$\begin{aligned} &P\Big(\sqrt{2\log(n/G)}\, T_n(\boldsymbol{R}, G) \le y + 2\log(n/G) + \frac{1}{2}\log\left(\frac{\int_0^1 (K'(t))^2\,dt}{4\int_0^1 K^2(t)\,dt}\right) \\ &- \log(\pi) \Big| X_1, \dots, X_n\Big) \to \exp\{-2\exp\{-y\}\}, \quad [P]\text{-a.s.} \end{aligned}$$

Proof. We apply Theorem 2.3 with $a_n(i) = X_i$, $i = 1\dots,n$. Hence the proof of our theorem reduces to the verification of the assumptions of Theorem 2.2.

Clearly,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(e_i - \overline{e}_n)^2 + 2\frac{1}{n}\sum_{j=0}^{q}(\mu_j - \overline{\mu}_n)\sum_{i=m_j+1}^{m_{j+1}} e_i$$
$$+\frac{1}{n}\sum_{j=0}^{q}(\mu_j - \overline{\mu}_n)^2(m_{j+1} - m_j),$$

where

$$\overline{\mu}_n = \frac{1}{n}\sum_{j=0}^{q}\mu_j.$$

Classical strong law of large numbers and few standard steps imply that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \ge \operatorname{var} e_1 > 0, \quad [P]\text{-a.s.}$$

Further, by the Minkowski inequality

$$\Big(\frac{1}{n}\sum_{i=1}^{n}|X_i - \overline{X}_n|^{2+\Delta}\Big)^{1/(2+\Delta)} \le \Big(\frac{1}{n}\sum_{i=1}^{n}|e_i - \overline{e}_n|^{2+\Delta}\Big)^{1/(2+\Delta)}$$
$$+\Big(\frac{1}{n}\sum_{j=0}^{q}|\mu_j - \overline{\mu}_n|^{2+\Delta}(m_{j+1} - m_j)\Big)^{1/(2+\Delta)}$$

which together with the strong law of large numbers and the assumptions implies

$$\limsup_{n\to\infty}\Big(\frac{1}{n}\sum_{i=1}^{n}|X_i - \overline{X}_n|^{2+\Delta}\Big)^{1/(2+\Delta)}$$
$$\le 2\big((E\,|e_1|^{2+\Delta})^{1/(2+\Delta)} + D_4^{1/(1+\Delta)}\big), \quad [P]\text{-a.s.}$$

Hence the assumptions of Theorem 2.3 are satisfied and our theorem follows. □

Remark 3.1. Notice that the assumptions of Theorem 3.1 covers both the null hypothesis and alternatives. Moreover, the limit permutation distribution is the same in both cases and does not depend on the original observations $X_1,\dots,X_n$. This means that the critical value for the permutation test provides an approximation to the critical value to the test based on $T_n(G)$.

Under the null hypothesis the permutation principle provides the exact critical values, otherwise it does reasonable approximation. The resulting tests are consistent for a large variety of alternatives. The behavior of the power function will be treated in a separate paper.

4. SIMULATIONS

To investigate behaviour of $T_n(\boldsymbol{R}, G)$ under various alternatives and kernel choices we carried out a comprehensive simulation study. Also, the study illustrates how far the limit critical values can depart from the exact ones.

We simulated values of $T_n(\boldsymbol{R}, G)$ for $G = \sqrt{n}$ and the following six types of kernels.

(i) $K_1(x) = 1$

(ii) $K_2(x) = 1 - |x|$

(iii) $K_3(x) = 1 - x^2$

(iv) $K_4(x) = |x|(1 - |x|)$

(v) $K_5(x) = |x|$

(vi) $K_6(x) = x^2$.

The kernel K_4 satisfies condition (K.3) while the rest satisfy condition (K.2). Specifically, $K(0) > 0$ and $K(1) = 0$ hold for kernels K_2, K_3 while $K(0) = 0$ and $K(1) > 0$ hold for kernels K_5, K_6. Finally, values $K(0)$ and $K(1)$ are both positive for kernel K_1. The choice of K_1 relates to the classical moving sum statistic, which has been widely studied in literature. More information about these kernels can be found in the Appendix.

All the six statistics above were studied under six different change-point hypotheses introduced below.

H_0: $q = 0$ (no change)

H_1: $q = 1$, $m_1 = n/2$, $\mu_1 - \mu_0 = 1$

H_2: $q = 1$, $m_1 = n/2$, $\mu_1 - \mu_0 = 2$

H_3: $q = 2$, $m_1 = n/3$, $m_2 = 2n/3$, $\mu_1 - \mu_0 = 1$, $\mu_2 - \mu_1 = 1$

H_4: $q = 2$, $m_1 = n/3$, $m_2 = 2n/3$, $\mu_1 - \mu_0 = 1$, $\mu_2 - \mu_1 = -1$

H_5: $q = 2$, $m_1 = n/3$, $m_2 = 2n/3$, $\mu_1 - \mu_0 = 1$, $\mu_2 - \mu_1 = -2$.

We used samples from the following three standardized error distributions: normal, Laplace, t_4. It implies 18 different underlying probability model setups. The sample sizes were chosen as $n = 100, 200$. All that resulted in 36 combinations studied for each kernel (i) through (vi).

For each combination of change-point hypothesis, error distribution and sample size we proceeded as follows:

(1) Observations $X_1, \ldots, X_n$ following model (1.1) with the (fixed) combination parameters are generated.

(2) A random permutation $\boldsymbol{r} = (r_1, \ldots, r_n)$ of $(1, \ldots, n)$ is generated.

(3) $T_n(\boldsymbol{R}, G)$ with $\boldsymbol{R} = \boldsymbol{r}$ is calculated and stored for each kernel (i) through (vi).

(4) The steps (2) through (3) are repeated 10,000 times.

(5) Sample quantiles corresponding to those 10,000 simulated values of $T_n(\boldsymbol{R}, G)$ are computed for each kernel (i) through (vi).

The results for $n = 100$ are summarized in Table 2. Table 1 contains asymptotic quantiles corresponding to Theorem 3.1.

As the case of $n = 200$ basically incites the same conclusions as the ones discussed below the corresponding tables for $n = 200$ are omitted. If we compare Table 1 and Table 2 it is evident that asymptotic critical values are conservative in contrast with the permutation procedure. The differences given by Table 1 and Table 2 varies between 0.4 and 1.3. The size of the differences increases with size $1 - \alpha$. Further, choices of kernel (ii) and (vi) lead to the largest differences while the choice of kernel (iv) lead to the smallest differences. These observations are independent of error distribution.

Recall that Theorem 3.1 shows the limit behaviour of $T_n(\boldsymbol{R}, G)$ is independent of alternative. The results in Table 2 matches this fact though we can see slightly different sizes subject to error distribution. Anyway, the differences are much smaller in contrast with deviations from the corresponding asymptotic quantiles. It may be explained in the following way.

To derive limit behaviour of $T_n(\boldsymbol{R}, G)$ an approximation by certain i.i.d. normally distributed variables is employed. Convergence of $T_n(\boldsymbol{R}, G)$ based on these variables to some standardized Gaussian process is then used and the limit distribution is obtained via extreme value theory for Gaussian processes. The simulation results lead to a supposition that the convergence to those i.i.d. normally distributed variables is much faster than the convergence to the Gaussian process.

The higher sample quantiles (95%, 99%) are slightly larger for Laplace and t_4 errors than for normal errors. The relation between Laplace and t_4 distribution may depend on the alternative. For instance, under H_0 the sample quantiles are larger in case of Laplace distribution while under H_5 it is the other way around no matter what kernel or probability we choose. On the other hand there is no such clear relation under H_4.

Table 1. Asymptotic quantiles for different kernels ($n = 100$).

$1-\alpha$	90%	95%	99%
K_1	3.634	3.970	4.729
K_2	3.957	4.293	5.052
K_3	3.738	4.074	4.833
K_4	3.358	3.693	4.453
K_5	3.634	3.970	4.729
K_6	3.872	4.208	4.967

Table 2. Sample quantiles for different setups of the simulation ($n = 100$).

errors		Normal			Laplace			t_4		
$1-\alpha$		90 %	95 %	99 %	90 %	95 %	99 %	90 %	95 %	99 %
K_1	H_0	3.042	3.260	3.714	3.062	3.297	3.750	3.058	3.291	3.738
	H_1	3.037	3.235	3.665	3.014	3.232	3.654	3.032	3.238	3.679
	H_2	3.061	3.258	3.678	3.053	3.270	3.680	3.034	3.248	3.661
	H_3	3.032	3.236	3.651	3.027	3.238	3.652	3.027	3.244	3.680
	H_4	3.039	3.239	3.673	3.036	3.267	3.708	3.059	3.275	3.748
	H_5	3.032	3.250	3.678	3.041	3.262	3.726	3.073	3.289	3.756
K_2	H_0	3.131	3.337	3.750	3.305	3.540	4.029	3.198	3.410	3.860
	H_1	3.113	3.298	3.677	3.140	3.345	3.759	3.135	3.341	3.743
	H_2	3.109	3.316	3.691	3.159	3.364	3.757	3.104	3.295	3.700
	H_3	3.110	3.296	3.648	3.105	3.298	3.701	3.179	3.381	3.764
	H_4	3.122	3.309	3.673	3.184	3.398	3.855	3.191	3.394	3.829
	H_5	3.113	3.317	3.694	3.177	3.379	3.815	3.237	3.461	3.874
K_3	H_0	3.105	3.307	3.723	3.192	3.429	3.899	3.138	3.351	3.794
	H_1	3.097	3.273	3.667	3.096	3.308	3.712	3.102	3.305	3.717
	H_2	3.092	3.301	3.678	3.129	3.328	3.729	3.089	3.276	3.709
	H_3	3.081	3.275	3.679	3.074	3.271	3.671	3.125	3.329	3.735
	H_4	3.092	3.288	3.633	3.129	3.348	3.790	3.141	3.349	3.789
	H_5	3.095	3.308	3.675	3.129	3.322	3.757	3.176	3.400	3.836
K_4	H_0	2.943	3.173	3.647	3.019	3.262	3.736	2.984	3.206	3.668
	H_1	2.935	3.147	3.565	2.931	3.143	3.575	2.931	3.157	3.635
	H_2	2.942	3.147	3.602	2.949	3.177	3.640	2.928	3.149	3.591
	H_3	2.918	3.146	3.575	2.926	3.148	3.590	2.939	3.161	3.638
	H_4	2.927	3.145	3.546	2.974	3.198	3.657	2.972	3.219	3.687
	H_5	2.937	3.165	3.573	2.957	3.193	3.653	2.997	3.241	3.701
K_5	H_0	3.036	3.236	3.669	3.213	3.465	3.960	3.095	3.321	3.812
	H_1	3.010	3.220	3.626	3.032	3.250	3.654	3.036	3.260	3.686
	H_2	3.019	3.228	3.660	3.068	3.276	3.671	2.992	3.185	3.603
	H_3	3.007	3.219	3.596	3.024	3.236	3.656	3.063	3.275	3.716
	H_4	3.028	3.243	3.652	3.072	3.313	3.796	3.077	3.291	3.746
	H_5	3.015	3.220	3.618	3.069	3.297	3.754	3.118	3.347	3.812
K_6	H_0	3.093	3.299	3.697	3.431	3.683	4.149	3.209	3.432	3.895
	H_1	3.049	3.240	3.610	3.097	3.315	3.740	3.110	3.307	3.725
	H_2	3.038	3.235	3.642	3.142	3.361	3.761	3.020	3.203	3.569
	H_3	3.037	3.229	3.586	3.081	3.282	3.657	3.155	3.359	3.814
	H_4	3.071	3.272	3.680	3.184	3.403	3.877	3.172	3.385	3.805
	H_5	3.052	3.247	3.629	3.175	3.394	3.806	3.243	3.468	3.956

Comparing the asymptotic quantiles (Table 1) and the sample quantiles (Table 2) with their counterparts in the paper Antoch and Hušková [3], where the problem of one change point is treated, we see very similar patterns, i. e. the empirical critical values are substantially smaller then the corresponding asymptotic ones and the empirical critical values are not almost influenced by the amount of change, location of change point(s) and the underlying distribution.

APPENDIX: COMPUTATIONAL PROCEDURES

This section contains quantities which are subjective to particular choice of kernel K and which we used for the purpose of the simulation. First introduce a useful notation.

$$
\begin{aligned}
\theta(K) &= \begin{cases} \log \dfrac{2K^2(0)+K^2(1)}{2\int_0^1 K^2(t)\,\mathrm{d}t} - \dfrac{1}{2}\log(\pi) & \text{under (K.2)}, \\[2ex] \dfrac{1}{2}\log \dfrac{\int_0^1 (K'(t))^2\,\mathrm{d}t}{4\int_0^1 K^2(t)\,\mathrm{d}t} - \log(\pi) & \text{under (K.3)}, \end{cases} \\
V(K) &= 2\sum_{i=1}^{G} K^2(\tfrac{i}{G}), \\
M_k(K) &= \sum_{i=0}^{G-1} Y_{k+i}\, K\left(\tfrac{G-i}{G}\right) - \sum_{i=0}^{G-1} Y_{k+G+i}\, K\left(\tfrac{i+1}{G}\right), \quad k = 1, \dots, n-2G+1
\end{aligned}
$$

where, in the simulation, $Y_k = X_{R_k}$. Then

$$T_n(G) = \sigma_n^{-1} V(K)^{-1/2} \max_{1 \le k \le n-2G+1} |M_k(K)|.$$

We introduce computational procedures for $M_k(K)$ optimized subject to computational time for particular choices of K. They are generally based on summing differences and their speed increase merely with n independently of G. Typically, it involves at most bn arithmetic operations (b varies according to K up to 10 within our choices) whereas the speed of explicit formulae is $4G(n-2G)$ arithmetic operations. This independence is the cause the procedures based on differences to be better than corresponding explicit formulae.

To be specific we consider

$$
\begin{aligned}
D_0 &= M_1 \\
D_k &= M_{k+1} - M_k, \quad k = 1, \dots, n-2G,
\end{aligned}
$$

which implies that

$$M_k(K) = \sum_{i=0}^{k-1} D_i(K), \quad k = 1, \dots, n-2G+1.$$

In cases of quadratic-type kernels we have to difference twice. We typically end with a formula containing a simple moving sum. The moving sums can be easily obtained as differences of respective cumulative sums. The formulae below may seem rather complicated but when properly programmed they result in very fast procedures. The summary of formulae for particular kernels used in the simulation follows.

(i) $K(x) = 1$

$$\theta(K) = -\frac{1}{2}\log\frac{4\pi}{9}, \qquad V(K) = 2G$$

$$M_k(K) = \sum_{i=0}^{G-1} Y_{k+i} - \sum_{i=0}^{G-1} Y_{k+G+i}$$

(ii) $K(x) = 1 - |x|$

$$\theta(K) = -\frac{1}{2}\log\frac{\pi}{9}, \qquad V(K) = \frac{(G-1)(2G-1)}{3G}$$

$$\begin{aligned} GD_0(K) &= \sum_{i=1}^{G-1} i\,Y_{i+1} - \sum_{i=1}^{G-1}(G-i)Y_{G+i} \\ GD_k(K) &= (2G-1)Y_{k+G} - \sum_{i=1}^{2G-1} Y_{k+i}, \quad k = 1, \ldots n - 2G \end{aligned}$$

(iii) $K(x) = 1 - x^2$

$$\theta(K) = -\frac{1}{2}\log\frac{64\pi}{225}, \qquad V(K) = \frac{(G-1)(16G^3 + G^2 + G + 1)}{15G^3}$$

$$\begin{aligned} G^2 D_0(K) &= \sum_{i=1}^{G-1}(G^2 - (G-i)^2)Y_{i+1} - \sum_{i=1}^{G-1}(G^2 - i^2)Y_{i+G} \\ G^2 D_k(K) &= (2G^2 - 1)Y_{k+G} + \sum_{i=0}^{k-1} A_i(K), \quad k = 1, \ldots, n - 2G \\ A_0(K) &= -\left(\sum_{i=1}^{G-1} (2(G-i)+1)Y_{i+1} + \sum_{i=1}^{G}(2i-1)Y_{i+G}\right) \\ A_k(K) &= (2G-1)(Y_{k+2G} - Y_{k+1}) + 2\left(\sum_{i=2}^{G} Y_{k+i} - \sum_{i=G+1}^{2G-1} Y_{k+i}\right) \end{aligned}$$

(iv) $K(x) = |x|(1 - |x|)$

$$\theta(K) = -\log\left(\sqrt{\frac{2}{5}}\,\pi\right), \qquad V(K) = \frac{(G-1)(G+1)(G^2+1)}{15G^3}$$

$$
\begin{aligned}
G^2 D_0(K) &= \sum_{i=1}^{G-1} i\,(G-i)(Y_{i+1} - Y_{i+G}) \\
G^2 D_k(K) &= \sum_{i=0}^{k-1} A_i(K), \quad k = 1, \dots, n-2G \\
A_0(K) &= \sum_{i=1}^{G-1} (G-2i+1)(Y_{i+G} - Y_{i+1}) \\
A_k(K) &= (G-1)(Y_{k+1} + Y_{k+G+1} - Y_{k+G} - Y_{k+2G}) \\
&\quad +2\left(\sum_{i=1}^{G-1} Y_{k+G+i} - \sum_{i=1}^{G-1} Y_{k+1+i}\right)
\end{aligned}
$$

(v) $K(x) = |x|$

$$
\theta(K) = -\frac{1}{2}\log\frac{4\pi}{9}, \qquad V(K) = \frac{(G+1)(2G+1)}{3G}
$$

$$
\begin{aligned}
GD_0(K) &= \sum_{i=0}^{G-1} (G-i)Y_{i+1} - \sum_{i=0}^{G-1} i\,Y_{i+G} \\
GD_k(K) &= Y_{k+G} - G(Y_k + Y_{k+2G}) + \sum_{i=1}^{2G-1} Y_{k+i}, \quad k = 1, \dots n-2G
\end{aligned}
$$

(vi) $K(x) = x^2$

$$
\theta(K) = -\frac{1}{2}\log\frac{4\pi}{25}, \qquad V(K) = \frac{(G+1)(2G+1)(3G^2+3G-1)}{15G^3}
$$

$$
\begin{aligned}
G^2 D_0(K) &= \sum_{i=0}^{G-1} (G-i)^2 Y_{i+1} - \sum_{i=1}^{G} i^2 Y_{i+G} \\
G^2 D_k(K) &= Y_{k+G} - G^2(Y_k + Y_{k+2G}) + \sum_{i=0}^{k-1} A_i(K), \quad k = 1, \dots, n-2G \\
A_0(K), &\qquad A_k(K) \quad \text{as in case (iii).}
\end{aligned}
$$

ACKNOWLEDGEMENT

The authors wish to express their sincere thanks to the referees for very careful reading and helpful remarks on the manuscript.

(Received November 13, 2000.)

REFERENCES

[1] J. Antoch and M. Hušková: Tests and estimators for epidemic alternatives. In: Tatra Mountains Mathematical Publications (A. Pázman and V. Witkovský, eds.), 1996, vol. 7, pp. 311–330.

[2] J. Antoch, M. Hušková, and D. Jarušková: Change point detection. In: Lecture Notes of the 5th IASC Summer School (Lauro et al, eds.), ISI, Voorburg 2000, pp. 1–75.

[3] J. Antoch and M. Hušková: Permutation tests for change point analysis. Statist. Probab. Lett. *53* (2001), 37–46.

[4] M. Csörgő and P. Révész: Strong Approximations in Probability and Statistics. Academic Press, New York 1981.

[5] M. Csörgő and L. Horváth: Limit Theorems in Change-point Analysis. Wiley, New York 1997.

[6] P. Good: Permutation Tests. Second edition. Springer–Verlag, New York 2000.

[7] U. Einmahl: A useful estimate in the multidimensional invariance principle. Probab. Theory Related Fields *76* (1987), 81–101.

[8] I. Grabovsky, L. Horváth, and M. Hušková: Limit theorems for kernel-type estimators for the time of change. J. Statist. Plann. Inference *89* (2000), 25–56.

[9] L. Horváth: Detecting changes in linear regressions. Statistics *26* (1995), 189–208.

[10] L. Horváth and P. Kokoszka: Change-point Detection with Non-parametric Regression. Report No. 313, Laboratory for Research in Statistics and Probability, Carlton University 1997.

[11] M. Hušková: Limit theorems for rank statistics. Statist. Probab. Lett. *32* (1997), 45–55.

[12] E. L. Lehmann: Theory of Point Estimation. Wadworth & Brooks/Cole, Calif. 1991.

[13] M. R. Leadbetter, L. Lindgren, and H. Rootzen: Extremes and Related Properties of Random Sequences and Processes. Springer–Verlag, New York 1983.

[14] J. P. Romano: Bootstrap and randomization tests of some nonparametric hypotheses. Ann. Statist. *17* (1989), 141–159.

Prof. Dr. Marie Hušková, Department of Probability and Statistics, Charles University, Sokolovská 83, 186 00 Praha 8 and Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: huskova@karlin.mff.cuni.cz

Mgr. Aleš Slabý, Department of Probability and Statistics, Charles University and ČSOB, Na Příkopě 14, 115 20 Praha 1. Czech Republic.
e-mail: aslaby@csob.cz