ELSEVIER

Statistics & Probability Letters 53 (2001) 37–46

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Permutation tests in change point analysis

Jaromír Antoch[1], Marie Hušková[*,2]

Department of Statistics, Charles University, Sokolovska 83, 18600 Praha, Czech Republic

Received August 2000; received in revised form November 2000

Abstract

The critical values for various tests for changes in location model are obtained through the use of permutation tests principle. Theoretical results show that in the limit these new "permutation tests" behave in the same way as the "classical tests" stemming from both maximum likelihood and Bayes principles. However, the results of the simulation study show that the permutation tests behave considerably better than the corresponding classical tests if measured by the critical values attained.

MSC: 62G20; 62E20; 62F07; 65C60

Keywords: Change(s) in location model; Permutation tests; Monte Carlo

1. Introduction

For simplicity, consider the location model with a change after an unknown time point m, i.e.

$$X_i = \mu + \delta_n I\{i > m\} + e_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where $1 \leqslant m \leqslant n$, μ and $\delta_n \neq 0$ are unknown parameters and $I\{A\}$ denotes the indicator of a set A. Assume, moreover, that

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

We are interested in the testing problem

$$\mathrm{H}_0 : m = n \quad \text{against} \quad \mathrm{H}_1 : m < n. \tag{1.3}$$

[*] Corresponding author. Tel.: +420-2-2191-3277; fax: +420-2-23-23-316.
E-mail address: huskova@karlin.mff.cuni.cz (M. Hušková).

[1] URL:http://www.karlin.mff.cuni.cz/~antoch

[2] URL:http://www.karlin.mff.cuni.cz/~huskova

0167-7152/01/$ - see front matter
PII: S0167-7152(01)00009-8

A number of test procedures for this problem using various principles has been developed; for recent references see, the book of Csörgő and Horváth (1997) or the survey paper of Antoch et al. (2000). It is well known that one of the main problems in the change point analysis is to find reasonable approximations for the critical values. Typically, approximations based on the 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. Therefore, these limit approximations are reasonable only for very large sample sizes and lead to conservative tests otherwise. Gombay and Horváth (1996), cf. also Csörgő and Horváth (1997), among others, pointed out this fact and proposed an improvement based on asymptotic arguments combined with a proper trimming.

In this paper, it is shown that the permutation tests can also be used to get asymptotically correct approximations for critical values. The simulation study shows that the permutation tests provide reasonable approximations to the critical values even in the case of small and moderate sample sizes. Basic information on general principles of permutation tests can be found, e.g., in Lehmann (1991) and Good (2000).

From a computational point of view, one needs a reasonable computer because the test statistic has to be calculated for a large number of permutations. On the other hand, the implementation of the basic idea is quite easy and straightforward.

The permutation test suggested below motivated us to develop along the same lines a variety of permutation tests related to other test statistics used in change point analysis. This means that the same principle can be applied to other test statistics (including M-tests) for the testing problem (1.3) in the model (1.1) as well as to the case of multiple changes in location models. The crucial point is that the test statistic must be expressible through the partial sums of residuals and that under H_0 these residuals are exchangeable random variables.

Asymptotic properties of our permutation test are investigated in Section 2. Section 3 deals with the tests for a change in location and/or in scale. The problem of how to evaluate the permutation distribution is discussed in Section 4. Here, we also summarize the outcome of a simulation study which supports the idea of the permutation based simulation. In the appendix, we collect some results on the asymptotics of rank statistics used in the proofs.

2. Permutation tests and their properties

We apply the permutation arguments to the test based on the statistic

$$T_{n1} = \max_{1<k<n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\hat{\sigma}_n} \left| \sum_{i=1}^{k} (X_i - \bar{X}_n) \right| \right\}, \tag{2.1}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2. \tag{2.2}$$

This test statistic is closely related to the likelihood ratio test when the error terms e_is have normal distribution. The large values indicate that the null hypothesis is not true and therefore the null hypothesis H_0 is rejected for large values of T_{n1}.

The permutation distribution of T_{n1} can be described as the conditional distribution (given $X_1,\ldots,X_n$) of

$$T_{n1}(\boldsymbol{R}) = \max_{1<k<n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\hat{\sigma}_n} \left| \sum_{i=1}^{k} (X_{R_i} - \bar{X}_n) \right| \right\}, \tag{2.3}$$

where $\boldsymbol{R}=(R_1,\ldots,R_n)$ is a random permutation of $(1,\ldots,n)$. This permutation distribution, denoted $F_P(x;T_{n1})$, can be expressed as

$$F_P(x;T_{n1})=\frac{1}{n!}\#\{\boldsymbol{r}\in\mathscr{R}_n;T_{n1}(\boldsymbol{r})\leqslant x\},\tag{2.4}$$

where $\mathscr{R}_n$ is the set of all permutations of $\{1,\ldots,n\}$ and $\#\{A\}$ denotes the cardinality of set A. Denote by $x_{1-\alpha,n}$ the $100(1-\alpha)\%$ quantile of the permutation distribution $F_P(.;T_{n1})$. Then the critical region with the level α of the permutation test based on T_{n1} has the form

$$T_{n1}\geqslant x_{1-\alpha,n}.\tag{2.5}$$

The permutation test can be described as follows:

(1) we calculate T_{n1} according to (2.1) and the quantile $x_{1-\alpha,n}$;
(2) the null hypothesis is rejected if (2.5) holds true.

Now we will study the permutation distribution function of T_{n1}; more precisely, we derive the conditional limit distribution of $T_{n1}(\boldsymbol{R})$ given $X_1,\ldots,X_n$. It is important to realize that $T_{n1}(\boldsymbol{R})$ given $X_1,\ldots,X_n$ can be viewed as a functional of a simple linear rank statistic and theorems on rank statistics for change point can be applied. The main assertion of this section states:

Theorem 2.1. *Let the observations $X_1,\ldots,X_n$ follow the model* (1.1), *the assumptions* (1.2) *be satisfied and let $|\delta_n|\leqslant D_0$ with some $D_0>0$. If $n\to\infty$, then for all $-\infty<y<\infty$ we have*

$$P(\sqrt{2\log\log n}\,T_{n1}(\boldsymbol{R})\leqslant y+2\log\log n+\tfrac{1}{2}\log\log\log n-\tfrac{1}{2}\log\pi\,|X_1,\ldots,X_n)$$

$$\to\exp\{-2\exp\{-y\}\},\quad [P]\text{-}a.s.$$

Proof. We apply Theorem A.1 with $a_{n1}(i)=X_i$, $i=1,\ldots,n$. Towards this, we have to check whether the assumptions of Theorem A.1 are satisfied. In other words, it is sufficient to check whether (A.2) and (A.3) are satisfied in our case. Clearly,

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\frac{1}{n}\sum_{i=1}^{n}(e_i-\bar{e}_n)^2+2\delta_n\frac{1}{n}\sum_{i=m+1}^{n}(e_i-\bar{e}_n)+\delta_n^2\frac{m(n-m)}{n^2}.$$

The classical strong law of large numbers implies that

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

Similarly, we find that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|X_i-\bar{X}_n|^{2+\Delta}\leqslant D_1(E|X_1|^{2+\Delta}+D_0^{2+\Delta}),\quad [P]\text{-a.s.},$$

with some $D_1>0$. These relations ensure that the assumptions of Theorem A.1 are satisfied and the assertion of our theorem follows. $\square$

Remark 2.1. Notice that the assumptions of Theorem 2.1 cover H_0 and local as well as fixed alternatives. Moreover, the conditional limit distribution does not depend on $X_1, \ldots, X_n$, so that the conditional and unconditional limit distribution of $T_{n1}(\boldsymbol{R})$ is the same, both under the null hypothesis and fixed alternative. One should recall that under H_0 the distributions of T_{n1} and $T_{n1}(\boldsymbol{R})$ coincide.

Remark 2.2. Checking the proof, one can easily find that the assertion of Theorem 2.1 remains true even if there are more than one but finite number of changes.

Remark 2.3. The important issue is that the permutation distribution provides the exact critical values ($x_{1-\alpha,n}$) for our testing problem under the null hypothesis. Moreover, under certain conditions on δ_n (cf. Csörgő and Horváth, 1997) the statistics T_{n1} tends to infinity so fast that we will reject H_0 with probability one under the alternative if the critical value $x_{1-\alpha,n}$ is used.

The permutation tests principle can be straightforwardly applied to other test statistics for the testing problem H_0 versus H_1. Namely, it can be applied to the test statistics based on various functionals of the partial sums $\sum_{i=1}^{k}(X_i - \bar{X}_n)$, $k = 1, \ldots, n$; for survey of these test statistics see, e.g., Csörgő and Horváth (1997). Here are just two examples, i.e.

$$T_{n2}(q) = \max_{1<k<n} \left\{ \frac{1}{q(k/n)} \frac{|\sum_{i=1}^{k}(X_i - \bar{X}_n)|}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}} \right\}$$

and

$$T_{n3}(w) = \sum_{k=1}^{n} \left(\frac{1}{w(k/n)} \frac{\left|\sum_{i=1}^{k}(X_i - \bar{X}_n)\right|}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}} \right)^2,$$

where $q(.)$ and $w(.)$ are properly chosen weight functions. However, we should remark that the explicit form of the corresponding limit distributions is known only for some particular choices of the weight functions $q(.)$ and $w(.)$, e.g., $q(t) = 1$, $t \in (0,1)$ and $w(t) = (t(1-t))^{3/2}$, $t \in (0,1)$. For details, see again Csörgő and Horváth (1997), p. 82. Therefore, it is highly desirable to have an approach on how to get approximations of the critical values for more complicated test statistics.

Another possibility is a generalization to the M-test statistics relevant to our problem. To recall, the M-test counterparts of T_{n1} can be obtained through replacing the partial sums of residuals $\sum_{i=1}^{k}(X_i - \bar{X}_n)$, $k = 1, \ldots, n$, and the sums of the squared residuals $\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ by their M-counterparts, i.e. by $\sum_{i=1}^{k} \psi(X_i - \hat{\mu}(\psi))$, $k = 1, \ldots, n$, and $\sum_{i=1}^{n} \psi^2(X_i - \hat{\mu}(\psi))$, respectively, where ψ is a score function and $\hat{\mu}(\psi)$ is the M-estimator of μ corresponding to the model (1.1) with $m = n$. For these permutation tests, one can prove similar results as those stated in Theorem 2.1 using the results of Hušková (1997b, c). In other words, the limit distribution of the permutation variant of the robust M-test statistics is the same as the limit of the original test statistic under the null hypothesis.

3. Tests for a change in location and/or in scale

Here, we shortly discuss the permutation version of a test for a change in location and/or in scale. More precisely, we consider the location model with a change after an unknown time point m in location and/or in scale, i.e.

$$X_i = \mu + \delta_n I\{i > m\} + (1 + \eta_n I\{i > m\})e_i, \quad i = 1, \dots, n, \tag{3.1}$$

where $m(\leqslant n)$, μ and $(\delta_n, \eta_n)' \neq (0,0)'$ are unknown parameters. Assume, moreover, that

$$e_1, \dots, e_n \text{ are iid rvs with symmetric distribution function,}$$
$$0 < \operatorname{var} e_1 < \infty, \quad E|e_1|^{4+\Delta} < \infty \text{ with some } \Delta > 0, \tag{3.2}$$

and

$$\lim_{n\to\infty} \delta_n \to 0, \quad \limsup_{n\to\infty} \eta_n^2 \leqslant D_2 \tag{3.3}$$

with some $D_2 > 0$.

As a test statistic for the testing problem (1.3), we consider the max type test statistics of the form

$$T_{n4} = \max_{1<k<n} \sqrt{\frac{n^2}{k(n-k)}} \left\{ \frac{[\sum_{i=1}^k (X_i - \bar{X}_n)]^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} + \frac{[\sum_{i=1}^k [(X_i - \bar{X}_n)^2 - \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)^2]]^2}{\sum_{i=1}^n [(X_i - \bar{X}_n)^2 - \frac{1}{n}\sum_{j=1}^n \left(X_j - \bar{X}_n\right)^2]^2} \right\}^{1/2}.$$

The corresponding permutation distribution of T_{n4} can be described as the conditional distribution, given $X_1, \dots, X_n$, of

$$T_{n4}(\boldsymbol{R}) = \max_{1<k<n} \sqrt{\frac{n^2}{k(n-k)}} \left\{ \frac{[\sum_{i=1}^k (X_{R_i} - \bar{X}_n)]^2}{\sum_{i=1}^n (X_i - \bar{X}_n)^2} + \frac{[\sum_{i=1}^k [(X_{R_i} - \bar{X}_n)^2 - \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)^2]]^2}{\sum_{i=1}^n \left[(X_i - \bar{X}_n)^2 - \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)^2\right]^2} \right\}^{1/2},$$

where $\boldsymbol{R} = (R_1, \dots, R_n)$ is a random permutation of $(1, \dots, n)$. Notice that given $X_1, \dots, X_n$ statistic, $T_{n4}(\boldsymbol{R})$ can be expressed as

$$\max_{1<k<n} \sqrt{\frac{n^2}{k(n-k)}} \sqrt{\frac{[\sum_{i=1}^k (a_{n1}(R_i) - \bar{a}_{n1})]^2}{\sum_{i=1}^n (a_{n1}(i) - \bar{a}_{n1})^2} + \frac{[\sum_{i=1}^k (a_{n2}(R_i) - \bar{a}_{n2})]^2}{\sum_{i=1}^n (a_{n2}(i) - \bar{a}_{n2})^2}},$$

where

$$a_{n1}(i) = X_i \quad \text{and} \quad a_{n2}(i) = (X_i - \bar{X}_n)^2, \quad i = 1, \dots, n.$$

With this choice of scores, one has

$$\operatorname{var}\left\{\sum_{i=1}^k a_{n1}(R_i)\,|\,X_1, \dots, X_n\right\} = \frac{k(n-k)}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

$$\operatorname{var}\left\{\sum_{i=1}^k a_{n2}(R_i)\,|\,X_1, \dots, X_n\right\} = \frac{k(n-k)}{n(n-1)} \sum_{i=1}^n \left[(X_i - \bar{X}_n)^2 - \frac{1}{n}\sum_{j=1}^n (X_j - \bar{X}_n)^2\right]^2$$

and

$$\operatorname{cov}\left\{\left(\sum_{i=1}^{k} a_{n1}(R_i), \sum_{i=1}^{k} a_{n2}(R_i)\right)\Bigg| X_1,\dots,X_n\right\} = \frac{k(n-k)}{n(n-1)}\left[\sum_{i=1}^{n}(X_i-\bar{X}_n)^3\right],$$

where $\operatorname{var}\{.\,|X_1,\dots,X_n\}$ and $\operatorname{cov}\{(.,.)\,|\,X_1,\dots,X_n\}$ denote conditional variance and covariance, respectively.

It can be easily checked that under the assumptions in (3.1)–(3.3), as $n\to\infty$,

$$\limsup_{n\to\infty}\frac{1}{n-1}\sum_{i=1}^{n}\left|(X_i-\bar{X}_n)^2-\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2\right|^{2+\Delta/2} \leqslant D_3E|X_1|^{4+\Delta}+D_4, \quad [P]\text{-a.s.},$$

$$\liminf_{n\to\infty}\frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-\bar{X}_n)^2-\frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2\right]^{2} \geqslant \operatorname{var}\{e_1^2\}, \quad [P]\text{-a.s.}$$

and

$$\frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i-\bar{X}_n)^3\right]\to 0, \quad [P]\text{-a.s.}$$

with some $D_3>0,\ D_4>0$.

Thus, if we combine these facts with Theorem A.2 we obtain the following result on the limit permutation distribution after a few standard steps.

Theorem 3.1. *Let the observations* $X_1,\dots,X_n$ *follow the model* (3.1) *and the assumptions* (1.2) *and* (3.2) *and* (3.3) *be satisfied. Then, as* $n\to\infty, \forall y\in\mathscr{R}_1$

$$P(\sqrt{2\log\log n}\,T_{n4}(\boldsymbol{R}) \leqslant y+2\log\log n+\log\log\log n|X_1,\dots,X_n)$$

$$\to \exp\{-2\exp\{-y\}\}, \quad [P]\text{-a.s.}$$

4. Simulations

To compare the behavior of the above described procedures, i.e. empirical distribution of statistics T_{n1} and $T_{n1}(\boldsymbol{R})$, we prepared (among others) the simulation experiment in which we generated data from the model (1.1), where we used:

- $n=80,120,200$;
- $m=n/4, n/2$ and $3n/4$;
- $\delta_n=0,1,2,3,4,5$;
- normally and Laplace distributed errors with the variance equal to one.

This resulted in 108 combinations of the above parameters. *Note that the choice* $\delta_n=0$ *corresponds to the situation under the null hypothesis* H_0, *i.e. no change*.

For each combination (e.g. $n=80$, $m=n/2$ and $\delta_n=5$, and normally distributed errors) we proceeded as follows:

Table 1
Empirical critical values for different setups of the simulation

			Normally distributed errors				Laplace distributed errors			
n	m	δ	10%	5%	2.5%	1%	10%	5%	2.5%	1%
80	20	0	2.765	2.966	3.179	3.417	2.842	3.217	3.343	3.530
80	20	1	2.726	2.966	3.173	3.438	2.844	3.220	3.242	3.481
80	20	2	2.740	2.988	3.213	3.482	2.824	3.002	3.231	3.510
80	20	3	2.747	3.012	3.251	3.511	2.756	3.001	3.247	3.522
80	20	4	2.760	3.027	3.261	3.528	2.761	3.015	3.267	3.537
80	20	5	2.770	3.035	3.271	3.535	2.768	3.024	3.268	3.552
80	40	1	2.723	2.952	3.178	3.406	2.838	3.197	3.349	3.501
80	40	2	2.714	2.949	3.162	3.398	2.810	2.965	3.172	3.419
80	40	3	2.692	2.930	3.136	3.376	2.703	2.924	3.139	3.373
80	40	4	2.682	2.910	3.123	3.366	2.681	2.908	3.120	3.367
80	40	5	2.670	2.903	3.113	3.366	2.672	2.900	3.111	3.361
80	60	1	2.728	2.947	3.161	3.400	2.843	2.995	3.226	3.521
80	60	2	2.713	2.958	3.172	3.413	2.808	2.975	3.208	3.495
80	60	3	2.721	2.968	3.192	3.436	2.810	2.990	3.206	3.490
80	60	4	2.732	2.984	3.203	3.473	2.744	2.998	3.229	3.504
80	60	5	2.746	2.996	3.219	3.492	2.749	3.004	3.244	3.521
120	30	0	2.840	3.059	3.261	3.503	2.972	3.339	3.610	3.680
120	30	1	2.847	3.107	3.253	3.532	2.968	3.245	3.682	3.847
120	30	2	2.799	3.046	3.265	3.512	2.870	3.192	3.627	3.637
120	30	3	2.803	3.051	3.273	3.541	2.851	3.181	3.337	3.604
120	30	4	2.807	3.067	3.295	3.569	2.855	3.091	3.322	3.607
120	30	5	2.813	3.074	3.306	3.589	2.863	3.078	3.324	3.601
120	60	1	2.849	3.125	3.254	3.517	2.962	3.299	3.529	3.646
120	60	2	2.816	3.017	3.226	3.462	2.883	3.059	3.267	3.527
120	60	3	2.758	2.992	3.194	3.446	2.770	3.003	3.207	3.450
120	60	4	2.743	2.974	3.179	3.436	2.750	2.974	3.183	3.431
120	60	5	2.735	2.966	3.172	3.430	2.735	2.965	3.170	3.430
120	90	1	2.848	3.158	3.486	3.542	2.961	3.237	3.435	3.660
120	90	2	2.840	3.157	3.446	3.575	2.879	3.192	3.314	3.630
120	90	3	2.841	3.162	3.293	3.565	2.874	3.080	3.314	3.616
120	90	4	2.850	3.064	3.301	3.564	2.829	3.081	3.317	3.602
120	90	5	2.857	3.071	3.318	3.566	2.827	3.089	3.316	3.603
200	50	0	2.896	3.116	3.340	3.615	2.953	3.142	3.371	3.652
200	50	1	2.898	3.118	3.350	3.598	2.997	3.280	3.406	3.669
200	50	2	2.871	3.116	3.341	3.620	2.946	3.284	3.389	3.651
200	50	3	2.875	3.121	3.354	3.623	2.930	3.203	3.375	3.656
200	50	4	2.877	3.127	3.356	3.637	2.933	3.142	3.376	3.676
200	50	5	2.881	3.131	3.368	3.641	2.873	3.141	3.375	3.675
200	100	1	2.895	3.113	3.329	3.601	2.982	3.206	3.368	3.632
200	100	2	2.848	3.084	3.303	3.557	2.922	3.099	3.324	3.581
200	100	3	2.825	3.057	3.273	3.537	2.833	3.072	3.284	3.539
200	100	4	2.808	3.044	3.258	3.522	2.813	3.050	3.265	3.521
200	100	5	2.801	3.036	3.249	3.515	2.801	3.038	3.254	3.511
200	150	1	2.904	3.117	3.346	3.633	2.914	3.121	3.346	3.615
200	150	2	2.899	3.133	3.371	3.650	2.853	3.097	3.322	3.575
200	150	3	2.905	3.137	3.384	3.674	2.849	3.091	3.312	3.582
200	150	4	2.880	3.149	3.382	3.677	2.850	3.096	3.319	3.605
200	150	5	2.885	3.157	3.382	3.680	2.857	3.106	3.331	3.628

(1) "original" observations $X_1, \dots, X_n$ are simulated using chosen (fixed) combination of parameters;
(2) a random permutation $\boldsymbol{r} = (r_1, \dots, r_n)$ of $(1, \dots, n)$ is generated;
(3) $T_{n1}(\boldsymbol{R})$ with $\boldsymbol{R} = \boldsymbol{r}$ is calculated and its value stored;
(4) steps (2) and (3) are repeated 10^5 times;
(5) empirical distribution function corresponding to these 10^5 values of $T_{n1}(\boldsymbol{R})$ is formed;
(6) empirical quantiles related to the above empirical distribution function are computed and used as the estimators of the "true quantiles" $x_{1-\alpha,n}$.

Recall that in order to get the "complete" permutation distribution of $T_{n1}(\boldsymbol{R})$, it would be necessary to calculate $T_{n1}(\boldsymbol{r})$ for all $n!$ permutations $\boldsymbol{r}$ being unfeasible.[3] Instead, $T_{n1}(\boldsymbol{r})$ was calculated for a large number (10^5) of randomly simulated permutations. In spite of the fact that $10^5 \ll n!$ for $n = 80, 120$ and 200, the results of our simulations show that the values of empirical quantiles stabilized already for smaller number of repetitions of the permutation step than used in our simulations.More precisely, empirical quantiles calculated for 2×10^4 permutation cycles are already very close to those presented in Table 1.

It is also important to note that we considered two versions of the basic situation for our simulations, i.e.:

(i) we used the same stream of errors for different values of δ when values of n, m and the type of the errors were fixed, i.e. we used same seed for the random numbers generator;
(ii) for each setup of n, m, δ and the type of the errors we used different stream of errors, i.e. we used different seeds for the random numbers generator in each of 108 simulations.

In Table 1, the results for the first situation are presented. However, the situation for the second approach is quite close to the first one.

Selected empirical quantiles, i.e. 90%, 95%, 97.5% and 99%, of the permutation distribution of $T_{n1}(\boldsymbol{R})$ corresponding to 10%, 5%, 2.5% and 1% empirical critical values, are summarized in Table 1. From this table it is evident that:

(i) Values of the empirical quantiles obtained through the permutation principle are very stable when the value of the jump is increased from $\delta_n = 0$ corresponding to the null hypothesis of no change to $\delta_n = 5$, corresponding to the evident change. Practically, this means that the procedure is very stable.
(ii) Values of the empirical quantiles obtained through the permutation principle are slightly larger for the Laplace distributed errors, however, the difference is practically negligible due to the fact that the difference between the corresponding quantiles is typically smaller than 0.1 for $n = 80$ and smaller than 0.05 for $n = 120$ and $n = 200$.
(iii) There is no influence of the location of the true change point m, at least in our setting.

For comparison, we calculated also the asymptotic quantiles according to the Theorem 2.1 and summarized them in Table 2. It is evident that these asymptotic critical values are very conservative contrary to the permutation test procedure.

Table 2
Asymptotic critical values

n	10%	5%	2.5%	1%
80	3.212	3.631	4.041	4.579
120	3.236	3.643	4.042	4.564
200	3.265	3.659	4.045	4.551

[3] It is important to realize that for computers it will forever hold that $2^{100} = \infty$.

For the simulation, we used Matlab v.5.3 running on 300 MHz Pentium II-powered notebook with 160 MB memory. The time necessary for one simulation experiment was between 400 and 1200 seconds including the calculation and drawing ten (10) complicated PostScript figures describing the results graphically, i.e. histograms, empirical distribution functions etc. We have analogous experience for the more complicated models, too. This demonstrates that the procedure is practically usable.

Acknowledgements

The authors acknowledge the valuable remarks and comments from the referee and colleagues. The work was supported by Grants GAČR 201/00/0769 and MSM 113200008.

Appendix A. Results on rank statistics

Consider the vector of the simple linear rank statistics

$$\boldsymbol{S}_k = (S_{k1}, S_{k2})', \quad S_{kj} = \sum_{i=1}^{k} (a_{nj}(R_i) - \bar{a}_{nj}), \quad k = 1, \ldots, n, \ j = 1, 2, \tag{A.1}$$

where $(R_1, \ldots, R_n)$ is a random permutation of $(1, \ldots, n)$, $a_{nj}(i)$, $i = 1, \ldots, n$, $j = 1, 2$, are scores and $\bar{a}_{nj} = n^{-1} \sum_{i=1}^{n} a_{nj}(i)$, $j = 1, 2$. Note that the random permutation $(R_1, \ldots, R_n)$ can be viewed as the vector of ranks corresponding to a random sample $(U_1, \ldots, U_n)$ from the uniform distribution on (0,1).

We consider two sets of conditions imposed on the scores. The first set concerns only $a_{nj}(i)$, $i = 1, \ldots, n$, and the other one concerns $\boldsymbol{a}_n(i) = (a_{n1}(i), a_{n2}(i))'$, $i = 1, \ldots, n$. We assume that $a_{nj}(i)$ satisfies $j = 1, 2$,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (a_{nj}(R_i) - \bar{a}_{nj})^2 \geqslant D_5 \tag{A.2}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |a_{nj}(R_i) - \bar{a}_{nj}|^{2+\Delta_1} \leqslant D_6 \tag{A.3}$$

with some positive D_5, D_6 and Δ_1.

The assumptions on the scores $a_{nj}(i)$ are slightly stronger, namely, we assume that

$$\liminf_{n\to\infty} \min_{\boldsymbol{d}\neq\boldsymbol{0}} \frac{\boldsymbol{d}'\boldsymbol{\Sigma}_n\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{d}} \geqslant D_7 \tag{A.4}$$

and

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{a}_n(i) - \bar{\boldsymbol{a}}_n\|^{2+\Delta_2} \leqslant D_8 \tag{A.5}$$

with some positive D_7, D_8 and Δ_2. Here, $\|.\|$ denotes the Euclidean norm,

$$\boldsymbol{\Sigma}_n = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{a}_n(i) - \bar{\boldsymbol{a}}_n)(\boldsymbol{a}_n(i) - \bar{\boldsymbol{a}}_n)'$$

and

$$\sigma_{nj}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (a_{nj}(i) - \bar{a}_{nj})^2, \quad j = 1, 2.$$

Theorem A.1. *Let* $(R_1,\dots,R_n)$ *be the the random permutations of* $(1,\dots,n)$ *and let assumptions* (A.2) *and* (A.3) *for* $j=1,2$, *be satisfied. Then, as* $n\to\infty$, *it holds* $\forall y\in\mathscr{R}_1$ *and* $j=1,2$ *that*

$$P\left(\sqrt{2\log\log n}\max_{1\leqslant k<n}\left\{\sqrt{\frac{n}{k(n-k)}}\sigma_{nj}^{-1}|S_{kj}|\right\}\leqslant y+2\log\log n+\tfrac{1}{2}\log\log\log n-\tfrac{1}{2}\log\pi\right)$$
$$\to\exp\{-2\exp\{-y\}\}.$$

Proof. The assertion follows from the Theorem 1 in Hušková (1997a). □

We also need an analog of Theorem A.1 concerning the behavior of quadratic forms of $\boldsymbol{S}_k$, which is formulated next:

Theorem A.2. *Let* $(R_1,\dots,R_n)$ *be the random permutations of* $(1,\dots,n)$ *and let assumptions* (A.4) *and* (A.5) *be satisfied. Then, as* $n\to\infty$, *it holds* $\forall y\in\mathscr{R}_1$ *that*

$$P\left(\sqrt{2\log\log n}\max_{1\leqslant k<n}\sqrt{\frac{n}{k(n-k)}\boldsymbol{S}_k'\boldsymbol{\Sigma}_n^{-1}\boldsymbol{S}_k}\leqslant y+2\log\log n+\log\log\log n\right)$$
$$\to\exp\{-2\exp\{-y\}\}.$$

Proof. Can be found in Hušková (1997b). □

References

Antoch, J., Hušková, M., Jarušková, D., 2000. Change point detection. In: Lauro, N.C. et al. (Eds.), Lecture Notes of the Fifth IASC Summer School. ISI, Voorburg, pp. 1–75.

Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. Wiley, New York.

Gombay, E., Horváth, L., 1996. On the rate of approximations for maximum likelihood test in change-point models. J. Multivariate Anal. 56, 120–152.

Good, P., 2000. Permutation Tests, 2nd Edition. Springer Verlag, New York.

Hušková, M., 1997a. Limit theorems for rank statistics. Statist. Probab. Lett. 32, 45–55.

Hušková, M., 1997b. Multivariate rank statistics processes and change point analysis. In: Ahmed, S.E., Ahsanullah, M., Sinha, B.K. (Eds.), Applied Statistical Sciences III. Nova Science Publishers, New York, pp. 83–96.

Hušková, M., 1997c. Limit theorems for M-processes via rank statistics processes. in: Balakrishnan, N. (Ed.), Advances in Combinatorial Methods with Applications to Probability and Statistics, pp. 521–534.

Lehmann, E.L., 1991. Theory of Point Estimation. Wadsworth & Brooks/Cole, California.