OFF – LINE STATISTICAL PROCESS CONTROL

Jaromír Antoch, Marie Hušková and Daniela Jarušková

Charles University of Prague, Department of Statistics, Sokolovská 83, CZ – 186 75 Praha 8, Czech Republic; jaromir.antoch@karlin.mff.cuni.cz. marie.huskova@karlin.mff.cuni.cz
Czech Technical University, Department of Mathematics, Thákurova 7, CZ – 166 29 Praha 6, Czech Republic; daniela.jaruskova@fsv.cvut.cz

Abstract: First part of this paper deals with tests on the stability of statistical models. The problem is formulated in terms of testing the null hypothesis H against the alternative hypothesis A. The null hypothesis H claims that the model remains the same during the whole observational period, usually it means that the parameters of the model do not change. The alternative hypothesis A claims that, at an unknown time point, the model changes, which means that some of the parameters of the model are subject to a change. In case we reject the null hypothesis H, i.e. when we decide that there is a change in the model, we concentrate on a number of questions that arise:

- when has the model changed;
- is there just one change or are there more changes;
- what is the total number of changes etc.

The time moment when the model has changed is usually called *change point*. Aside testing for a change, our interest is to estimate change point(s) in different models. The least squares, M-, R- and MOSUM estimators are introduced and studied. Of course, we also estimate other parameters of the model(s), show approximations to the distributions of the change point estimators and show that the estimators of the change points are usually closely related to some of the test statistics treated in the first part.

Three types of confidence intervals are developed, one based on the limit distribution of the (point) estimators of m and two based on the bootstrap methods. All three methods are suitable for local changes while only the bootstrap constructions apply also to fixed changes.

The test statistics described below are typically certain functionals of partial sums of independent, identically distributed variables and their distribution is very complex. Therefore, we present selected limit results that form the basis for establishing the limit distribution of considered test statistics, functionals of partial sums and change point estimators.

Several Matlab codes, that implement selected methods described below and include detailed description and links to the previous sections, are presented to illustrate the possibilities of the studied methods. Complete Matlab codes are available from the authors on request.

2000 *Mathematics Subject Classification.* 62-01, 62F03, 62F05, 62F10, 62F12, 62F15, 62F35, 62F40.

Key words and phrases. Change point problem; $L_1-, L_2-, ML-, M-$ and R-estimators; Bayesian, maximumum-type and trimmed maximum-type statistics; MOSUM; location model, linear regression, abrupt and gradual changes, multiple changes; AR and ARMA processes; hypotheses testing, approximate critical values, Bonferroni inequality; Brownian bridge, Gaussian and Wiener processes; bootstrap.

Acknowledgement. The work was partially supported by grants GAČR 201/00/0769, MSM 113200008 and MSM210000001. We are grateful for valuable comments we got from Prof. J. Steinebach.

Contents

Part I. Testing

1. Introduction

In the scope of mathematical statistics the decision whether observed series remained stationary or whether a change of a specific kind occurred is usually based on hypotheses testing. *The null hypothesis claims that the process is stationary while the alternative hypothesis claims that the process is non-stationary and the stationarity was violated in a specific way.*

We start with the simplest situation that arises if we assume that a certain characteristic (e.g. that of a manufacturing process) varies around a certain constant μ_0 given by the production design. We suppose that at the beginning the process is *in control*. However, it can happen that due to a failure of the production device, e.g., the observed characteristic suddenly starts to vary around another *out-of control* constant $\mu_1 \neq \mu_0$. The same sudden failure may cause a change of the variance as well, but it is also possible that the variance remains the same. Moreover, sometimes we can even suppose, because of our long experience with the production process, that the variance σ^2 is known.

In this simple case, with starting value μ_0 and variance σ^2 known both, one can standardize the observations to obtain the standardized variables Y_i, $i = 1, \dots, n$, which have at the beginning a zero mean and a unit variance, and to test the following null hypothesis H against the alternative A, i.e.,

$$\begin{aligned} &H : Y_i = e_i, && i = 1, \dots, n, \\ &A : \exists\, \mathsf{m} \in \{0, \dots, n-1\} \quad \text{such that} && \\ &\qquad Y_i = e_i, && i = 1, \dots, \mathsf{m}, \\ &\qquad Y_i = \mu + e_i, && i = \mathsf{m}+1, \dots, n, \end{aligned} \tag{I.1}$$

where $\mu \neq 0$ and e_i are independent identically distributed (iid) random variables (errors). *The quantity* m *is called change point.*

2. Methods for deriving test statistics

The decision rule for rejecting the null hypothesis H is based on test statistics. Two basic methods may be applied to derive them, namely, the *maximum likelihood method* and the *pseudo-Bayes method*. We demonstrate on the testing problem (I.1) how both these approaches can be applied. For simplicity we suppose that e_i are independent and distributed according to the standard normal distribution $N\big(0, 1\big)$ with the density $\phi(x)$ and the distribution function $\Phi(x)$.

2.1. Likelihood ratio method

♣ Let us suppose for a moment that the change point m is known and put $\mathsf{m} = k$. If $\mu \neq 0$, the log-likelihood ratio for testing H against A is

$$\Lambda_k = \sup_\mu \log \frac{\prod_{i=1}^n f_A(Y_i)}{\prod_{i=1}^n f_H(Y_i)} = \sup_\mu \log \frac{\prod_{i=1}^k \phi(Y_i) \prod_{i=k+1}^n \phi(Y_i - \mu)}{\prod_{i=1}^n \phi(Y_i)}$$
$$= \sup_\mu \left\{ -\frac{1}{2} \sum_{i=k+1}^n (Y_i - \mu)^2 + \frac{1}{2} \sum_{i=k+1}^n Y_i^2 \right\} = \frac{1}{2(n-k)} \left(\sum_{i=k+1}^n Y_i \right)^2 .$$

The null hypothesis H is rejected when $\Lambda_k > C_\alpha$, which can be equivalently expressed as

$$\left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^n Y_i \right| > \sqrt{2C_\alpha},$$

where C_α is a constant chosen so as to correspond to the fixed significance level α. In other words, the log-likelihood ratio is a function of the average of the second part of the series of observations $\{Y_i\}$.

To simplify the notation, we put

$$\overline{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i \quad \text{and} \quad \overline{Y}_k^o = \frac{1}{n-k} \sum_{i=k+1}^n Y_i \,. \tag{I.2}$$

Notice, that $\overline{Y}_k^o$ is the least squares estimator of the unknown constant μ and $\sqrt{n-k}\ \overline{Y}_k^o$ has a standard normal distribution $N(0,1)$.

- For the one-sided alternative with $\mu > 0$, we obtain the test statistic

$$\frac{1}{\sqrt{n-k}} \sum_{i=k+1}^n Y_i,$$

while for the two-sided alternative with $\mu \neq 0$, we apply its absolute value

$$\left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^n Y_i \right| .$$

♣ When the change point m is unknown (so that both μ and m are unknown), we have to take the supremum of the log-likelihood ratio with respect to both of them, i.e.,

$$\max_{0\le k\le n-1} \sup_{\mu} \log \frac{\prod_{i=1}^{k} \phi(Y_i) \prod_{i=k+1}^{n} \phi(Y_i-\mu)}{\prod_{i=1}^{n} \phi(Y_i)}$$
$$= \max_{0\le k\le n-1} \frac{1}{2(n-k)} \left(\sum_{i=k+1}^{n} Y_i \right)^2,$$

and the test statistics usually applied for the case with the unknown change point m are of the form

$$\max_{0\le k\le n-1} \left\{ \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right\}, \tag{I.3}$$

and

$$\max_{0\le k\le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\}. \tag{I.4}$$

These statistics belong to the so-called *maximum-type statistics.*

• If we consider the two-sided alternative with $\mu \neq 0$, the null hypothesis H is rejected if for a suitably chosen constant $C_{1\alpha}$

$$\max_{0\le k\le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\} > C_{1\alpha},$$

where $C_{1\alpha}$ is a constant chosen so as to correspond to the fixed significance level α. This rule is reasonable because it rejects H if for at least one k, $0 \le k \le n-1$,

$$\left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| > C_{1\alpha}\,.$$

• Similarly, for the one-sided alternative with $\mu > 0$, the null hypothesis is rejected if

$$\max_{0\le k\le n-1} \left\{ \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right\} > C_{2\alpha}\,,$$

where $C_{2\alpha}$ is again an appropriately chosen constant.

2.2. Pseudo-Bayesian method

The method is based on the assumption that the unknown change point m and unknown magnitude of shift $\boldsymbol{\mu}$ are random variables such that the prior distribution of m is uniform, i.e. $P(\mathsf{m} = i) = 1/n, i = 0, \ldots, n-1$, and $\boldsymbol{\mu}$ is distributed according to $N(0, \gamma^2)$. Since the density of $Y_1, \ldots, Y_n$ given $\boldsymbol{\mu} = \mu$ and $\mathsf{m} = k$ is normal, i.e.,

$$f\Big(y_1, \ldots, y_n \mid \boldsymbol{\mu} = \mu, \mathsf{m} = k\Big) = \prod_{i=1}^{k} \phi(y_i) \prod_{i=k+1}^{n} \phi(y_i - \mu),$$

the unconditional density can be expressed as

$$f(y_1,\dots,y_n) = \sum_{k=1}^{n} \frac{1}{n} \int_{-\infty}^{\infty} \prod_{i=1}^{k} \phi(y_i) \prod_{i=k+1}^{n} \phi(y_i-\mu) \frac{1}{\gamma\sqrt{2\pi}} \exp\Big\{-\mu^2/2\gamma^2\Big\}\, d\mu =$$

$$= \prod_{i=1}^{n} \phi(y_i) \left(\sum_{k=1}^{n} \frac{1}{n} \int_{-\infty}^{\infty} \frac{1}{\gamma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} \left(-2\mu \sum_{i=k+1}^{n} y_i + (n-k)\mu^2 + \frac{\mu^2}{\gamma^2} \right) \right\} d\mu \right)$$

and the corresponding likelihood ratio has the form

$$\begin{aligned}\widetilde{\Lambda} &= \frac{f_A(Y_1,\dots,Y_n)}{f_H(Y_1,\dots,Y_n)} \\ &= \sum_{k=1}^{n} \frac{1}{n} \sqrt{\frac{1}{1+(n-k)\gamma^2}}\ \exp\left\{ \frac{\gamma^2}{2\big(1+(n-k)\gamma\big)^2} \Big(\sum\nolimits_{i=k+1}^{n} Y_i\Big)^2 \right\}.\end{aligned}$$

• Letting $\gamma \to 0$ and applying Taylor expansion, the likelihood ratio $\widetilde{\Lambda}$ is, for the two-sided alternative, equivalent to the test statistic

$$\sum_{k=1}^{n} \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{i=k+1}^{n} Y_i \right)^2 . \tag{I.5}$$

The obtained statistic belongs to the so called *sum-type* test statistics. For details see Gardner (1969).

• For the one-sided alternative with $\mu > 0$ (μ is assumed to follow a 50% truncated normal distribution) we may analogously derive the sum-type test statistic

$$\sum_{k=1}^{n} \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{i=k+1}^{n} Y_i \right) . \tag{I.6}$$

For details see Chernoff and Zacks (1964) and Kander and Zacks (1966).

2.3. Critical values

For the decision about rejection of the null hypothesis H we need to know critical values of the suggested test statistics. It means to know their distributions under H.

• Let us start with the test statistic

$$\max_{0\le k\le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\} . \tag{I.7}$$

Assuming that $\{Y_i\}$ are iid with the standard normal distribution $N(0,1)$, then all statistics

$$\frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i, \quad k = 0,\dots,n-1, \tag{I.8}$$

have a $N(0,1)$ distribution. If m is known and equal to k, one would reject H at the significance level α if

$$\left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| > u_{1-\alpha/2},$$

where $u_{1-\alpha/2}$ is the $100(1-\alpha/2)\%$ quantile of $N(0,1)$. Clearly, the statistic (I.7) is a stochastically larger variable than any of the absolute values of statistics (I.8), i.e., $\forall x \in \mathcal{R}^1$ and $0 \le k \le n-1$

$$P\left(\max_{0\le k\le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\} > x \right) \ge P\left(\left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| > x \right).$$

Therefore, the $100\,\alpha\%$-quantile of the distribution of statistic (I.7) is larger than $u_{1-\alpha/2}$. Analyzing the same data set, we reject the null hypothesis in the case when the change point is known much more often than in the case of an unknown change point.

• To find the exact distribution of (I.4) means to find the distribution of the maximum of absolute values of standardized normal variables that are (unfortunately) correlated. The correlation coefficients are

$$corr\left(\frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i, \frac{1}{\sqrt{n-l}} \sum_{i=l+1}^{n} Y_i \right) = \sqrt{\frac{n-l}{n-k}}, \quad k \le l.$$

Theoretically, it should not be a problem to find the distribution of (I.4). However, in practice the distribution is so complex, that its quantiles (desired critical values) may be computed only for small values of n, see Hawkins (1977).

• Sometimes, the approximate critical values may be quite satisfactory for practical use. To find approximate critical values we can use a very simple idea by applying the Bonferroni inequality as follows:

$$P\left(\max_{0\le k\le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\} > C \right) = P\left(\bigcup_{k=0}^{n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| > C \right\} \right)$$
$$\le \sum_{k=0}^{n-1} P\left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| > C \right\} = n\,P\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \right| > C \right).$$

Hence, the $100\big(1-\alpha/(2n)\big)\%$-quantile of the standard normal distribution $N(0,1)$ may serve as an upper estimate of the critical value at the significance level α for the problem (I.1) applying the test statistic (I.4). The approximate critical values obtained in this way are good enough for small samples (for small values of n), but they are too conservative for n large.

• Therefore, for n large, the asymptotic behavior of the studied test statistic (I.4) is of interest. It can be proved, applying the law of iterated logarithm, that

$$\max_{0\le k\le n-1}\left\{\left|\frac{1}{\sqrt{n-k}}\sum\nolimits_{i=k+1}^{n} Y_i\right|\right\}\longrightarrow\infty \quad \text{almost surely as} \quad n\to\infty.$$

It follows that the limit distribution of (I.4) does not exist and that critical values increase to infinity as $n\to\infty$. The problem is caused by the behavior of the sequence $\left\{(n-k)^{-1/2}\sum_{i=k+1}^{n} Y_i, i = 1,\ldots,n-1\right\}$ near to its end. Here the averages, whose departures from zero are studied, are calculated only for "a small" number of observations and it can happen, with a large probability, that at least one of them attains a rather large value.

Therefore, some authors suggest to use, instead of the statistics (I.3) and (I.4), the *trimmed maximum-type* test statistics

$$\max_{0\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\frac{1}{\sqrt{n-k}}\sum\nolimits_{i=k+1}^{n} Y_i\right\} \tag{I.9}$$

and

$$\max_{0\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\left|\frac{1}{\sqrt{n-k}}\sum\nolimits_{i=k+1}^{n} Y_i\right|\right\}, \tag{I.10}$$

where β is a small positive constant less than one and $\lfloor x\rfloor$ denotes the integer part of x. The advantage of the statistics (I.9) and (I.10) is that they are bounded in probability. The *trimming off* a $100\beta\%$ portion of the sample (upper time points) means that one assumes that the change did not occur during this time period. Notice that, typically, $\beta\in[0.01,0.1]$. The decision "*How much to trim off?*" depends on the subjective decision of the statistician and his/her a priori knowledge of the problem. If the statistician decides to trim off only a very small portion of the time points or no time points (observations) at all, he/she pays for it by a loss of the power of his/her test as the critical values depend rather strongly on the value of β.

Tables 1. – 6. below contain critical values for statistics (I.3) – (I.4) and (I.9) – (I.10) obtained by simulations.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.297	2.604	2.874	3.192	3.412
$\beta = 0.01$	2.269	2.581	2.854	3.177	3.398
$\beta = 0.05$	2.190	2.510	2.792	3.118	3.347
$\beta = 0.10$	2.119	2.446	2.731	3.068	3.302

Table 1. Simulated critical values of the over-all maximum-type test statistic (I.3) and the corresponding trimmed maximum-type test statistic (I.9) for different trimming portions β, $n = 100$.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.603	2.874	3.117	3.410	3.613
$\beta = 0.01$	2.581	2.853	3.102	3.395	3.597
$\beta = 0.05$	2.510	2.792	3.040	3.344	3.555
$\beta = 0.10$	2.447	2.732	2.991	3.299	3.512

Table 2. Simulated critical values of the over-all maximum-type test statistic (I.4) and the corresponding trimmed maximum-type test statistic (I.10) for different trimming portions β, $n = 100$.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.455	2.754	3.018	3.328	3.542
$\beta = 0.01$	2.387	2.696	2.969	3.286	3.508
$\beta = 0.05$	2.270	2.589	2.867	3.189	3.418
$\beta = 0.10$	2.186	2.515	2.801	3.130	3.363

Table 3. Simulated critical values of the over-all maximum-type test statistic (I.3) and the corresponding trimmed maximum-type test statistic (I.9) for different trimming portions β, $n = 500$.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.753	3.016	3.255	3.540	3.737
$\beta = 0.01$	2.694	2.967	3.211	3.503	3.707
$\beta = 0.05$	2.589	2.869	3.117	3.420	3.629
$\beta = 0.10$	2.515	2.800	3.055	3.362	3.573

Table 4. Simulated critical values of the over-all maximum-type test statistic (I.4) and the corresponding trimmed maximum-type test statistic (I.10) for different trimming portions β, $n = 500$.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.510	2.805	3.066	3.373	3.585
$\beta = 0.01$	2.416	2.721	2.991	3.311	3.530
$\beta = 0.05$	2.289	2.610	2.886	3.214	3.442
$\beta = 0.10$	2.205	2.533	2.817	3.149	3.373

Table 5. Simulated critical values of the over-all maximum-type test statistic (I.3) and the corresponding trimmed maximum-type test statistic (I.9) for different trimming portions β, $n = 1\,000$.

	significance level				
	10%	5%	2.5%	1%	0.5%
over-all	2.804	3.066	3.301	3.586	3.785
$\beta = 0.01$	2.723	2.994	3.238	3.531	3.735
$\beta = 0.05$	2.609	2.887	3.142	3.446	3.651
$\beta = 0.10$	2.531	2.815	3.071	3.374	3.590

Table 6. Simulated critical values of the over-all maximum-type test statistic (I.4) and the corresponding trimmed maximum-type test statistic (I.10) for different trimming portions β, $n = 1\,000$.

• For large n, the approximate critical values can be calculated from the asymptotic behavior of the probabilities under H, because we have $\forall x \in \mathcal{R}^1$:

$$P\left(\max_{0\le k\le n-1}\left\{\left|\frac{1}{\sqrt{n-k}}\sum_{i=k+1}^{n} Y_i\right|\right\} > \frac{x+b_n}{a_n}\right) \approx 1-\exp\left\{-e^{-x}\right\}, \quad \text{(I.11)}$$

where

$$a_n = \sqrt{2\log\log n} \quad \text{and} \quad b_n = 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log \pi,$$

and

$$P\left(\max_{0\le k\le \lfloor(1-\beta)\,n\rfloor}\left\{\left|\frac{1}{\sqrt{n-k}}\sum_{i=k+1}^{n} Y_i\right|\right\} > x\right) \approx 2\big(1-\Phi(x)\big) + x\phi(x)\log\frac{1}{\beta}\cdot \quad \text{(I.12)}$$

The approximation (I.11) was derived by Darling and Erdős (1956).

Both formulas (I.11) and (I.12) were derived from the approximation of the maximum of the sequence

$$\left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\}$$

by the maximum of the random process

$$\left\{ \frac{|\,W(1-t)\,|}{\sqrt{1-t}} \right\},$$

where $\{W(t), t \geq 0\}$ denotes the Wiener process, see Part III Sections 11 and 12.

Notice that the distribution of

$$\max_{0 \leq t \leq 1-\beta} \frac{|\,W(1-t)\,|}{\sqrt{1-t}} \quad \text{is the same as the distribution of} \quad \max_{\beta \leq t \leq 1} \frac{|\,W(t)\,|}{\sqrt{t}} \cdot$$

The approximation (I.12), as well as the approximations of the same type stated later on, are derived for x large and therefore they can be useful for finding critical values. For small x their validity fails.

The critical values obtained from (I.11) are not very good because they are too conservative. The approximation (I.12) gives more accurate critical values but the choice of β affects them significantly. The critical values obtained from (I.12) get better as β gets larger.

• For the one-sided alternatives with $\mu > 0$, the approximations (I.11) and (I.12) have to be adapted as follows, i.e., $\forall x \in \mathcal{R}^1$

$$P\left(\max_{0 \leq k \leq n-1} \left\{ \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right\} > \frac{x + b_n}{a_n} \right) \approx 1 - \exp\left\{ -\frac{1}{2} e^{-x} \right\}, \quad \text{(I.13)}$$

and

$$P\left(\max_{0 \leq k \leq \lfloor (1-\beta)\, n \rfloor} \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i > x \right) \approx \big(1 - \Phi(x)\big) + \frac{1}{2} x \phi(x) \log \frac{1}{\beta} \cdot \quad \text{(I.14)}$$

• Now we come to the sum-type statistics obtained from the pseudo-Bayes method. For n large, the approximate critical values of (I.5) derived for the two-sided alternative can be calculated from the convergence

$$\sum_{k=1}^{n} \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{i=k+1}^{n} Y_i \right)^2 \xrightarrow{\mathcal{D}} \int_0^1 W^2(t)\, dt = \sum_{j=1}^{\infty} \frac{4}{(2j+1)^2 \pi^2}\, \chi_j^2(1), \quad \text{(I.15)}$$

where $\{\chi_j^2(1)\}$ are independent variables distributed according to the χ^2 distribution with one degree of freedom. The distribution of $\int_0^1 W^2(t)\, dt$ was studied by MacNeill (1978) who obtained the following selected $100(1-\alpha)\%$ quantiles

$$P\left(\int_0^1 W^2(t)\,dt < x\right) = 1 - \alpha,$$

which can serve as the $100\alpha\%$ critical values presented in the following table

α	0.1	0.05	0.025	0.01
x	1.196	1.656	2.134	2.788

Table 7. $100\alpha\%$ critical values for the sum-type statistic calculated according to (I.15).

• The calculation of critical values of (I.6) derived for the one-sided alternatives is very simple as the sum-type test statistic

$$\sum_{k=1}^{n} \frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=k+1}^{n} Y_i\right) = \frac{1}{n}\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(k-1)Y_k$$

is normally distributed as

$$N\left(0, \frac{1}{3} - \frac{1}{2n} + \frac{1}{6n^2}\right).$$

3. Detection of changes in the parameters of a normal distribution

In this section we will, in more details, concentrate on the detection of changes in the parameters of a normal distribution. More general situations will be considered in Section 4.

3.1. Change in mean with unknown starting value

It (very) often happens that, even before the change point, the mean (the starting value) is unknown. In this case we test the following null hypothesis H against the alternative A:

$$\begin{aligned} &H: Y_i = \mu + e_i, && i = 1, \dots, n, \\ &A: \exists\, \mathsf{m} \in \{1, \dots, n-1\} \ \text{ such that} && \\ &\quad Y_i = \mu + e_i, && i = 1, \dots, \mathsf{m}, \\ &\quad Y_i = \mu + \delta + e_i, && i = \mathsf{m}+1, \dots, n,\ \delta \neq 0. \end{aligned} \tag{I.16}$$

We suppose again that the variables e_i are iid with the normal distribution $N(0, \sigma^2)$, σ^2 known. Then the maximum-type test statistics have the form

$$\max_{1 \le k \le n-1} \left\{ \frac{1}{\sigma} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \left(Y_i - \overline{Y}_n\right) \right| \right\} \tag{I.17}$$

and

$$\max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta)\, n \rfloor} \left\{ \frac{1}{\sigma} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \left(Y_i - \overline{Y}_n\right) \right| \right\}, \tag{I.18}$$

while the sum-type test statistic has the form

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}\frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right)^2. \tag{I.19}$$

Notice that $\sum_{i=1}^{k}\big(Y_i-\overline{Y}_n\big) = -\sum_{i=k+1}^{n}\big(Y_i-\overline{Y}_n\big)$ and that we deal with the partial sums of residuals $S_k=\sum_{i=1}^{k}\big(Y_i-\overline{Y}_n\big)$ under H instead of with partial sums of observations.

• To find approximate critical values, we may proceed in the same way as in the case with the known starting value. For n small we can use the Bonferroni inequality, and for n large we may apply the limit behavior of the studied probabilities, i.e.,

$$\begin{aligned}P\Bigg(\max_{1\le k\le n-1}\left\{\frac{1}{\sigma}\left|\sqrt{\frac{n}{k(n-k)}}\ \sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\} > \frac{x+b_n}{a_n}\Bigg)\approx\\ \approx 1-\exp\big\{-2e^{-x}\big\},\quad x\in\mathcal{R}^1,\end{aligned} \tag{I.20}$$

$$\begin{aligned}P\Bigg(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\frac{1}{\sigma}\left|\sqrt{\frac{n}{k(n-k)}}\ \sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\} > x\Bigg)\approx\\ \approx 2\big(1-\Phi(x)\big)+2x\phi(x)\log\frac{1-\beta}{\beta}\cdot\end{aligned} \tag{I.21}$$

The formulas (I.20) and (I.21) are derived from the approximation of the sequence

$$\left\{\frac{1}{\sigma}\left|\sqrt{\frac{n}{k(n-k)}}\ \sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\}\quad\text{by the process}\quad\left\{\frac{|\,B(t)\,|}{\sqrt{t(1-t)}}\right\},$$

see Part III Sections 11 and 12. The approximation (I.20) can be found in Yao and Davis (1986).

For normally distributed random variables, a better approximation suggested by James et al. (1987) may be used:

$$\begin{aligned}P\Bigg(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\left|\frac{1}{\sigma}\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\} > x\Bigg)\approx\\ \approx 2\big(1-\Phi(x)\big)+2x\,\phi(x)\int_{x\sqrt{\beta/(n(1-\beta))}}^{x\sqrt{(1-\beta)/(n\beta)}}\frac{1}{y}\,\nu\Big(y+\frac{x^2}{n\,y}\Big)\,dy,\end{aligned} \tag{I.22}$$

where

$$\nu(y)=\frac{2}{y^2}\exp\left\{-2\sum_{n=1}^{\infty}\frac{1}{n}\Phi\Big(-\frac{1}{2}y\sqrt{n}\Big)\right\}\approx\exp\big\{-0.583\,y\big\},\quad y>0. \tag{I.23}$$

Tables 8. – 11. below contain critical values for statistics (I.17) – (I.18) obtained by simulations when σ is known and equal to one. Tables 8'. – 11'.

contain asymptotic critical values calculated using (I.20) – (I.21) for comparison purposes.

n	(I.17)	(I.18)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	2.709	2.709	2.639	2.558
100	2.809	2.783	2.703	2.627
200	2.892	2.855	2.763	2.682
300	2.931	2.884	2.780	2.694
500	2.973	2.916	2.804	2.714

Table 8. Simulated 10 % critical values for statistics (I.17) and (I.18).

(I.20)	$n = 100$	$n = 300$	$n = 500$
	3.226	3.285	3.310
(I.21)	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
	3.082	2.920	2.810

Table 8'. Asymptotic 10 % critical values for statistics (I.20) and (I.21).

n	(I.17)	(I.18)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	2.960	2.960	2.900	2.823
100	3.065	3.040	2.965	2.900
200	3.143	3.110	3.030	2.950
300	3.176	3.135	3.042	2.967
500	3.218	3.169	3.068	2.983

Table 9. Simulated 5 % critical values for statistics (I.17) and (I.18).

(I.20)	$n = 100$	$n = 300$	$n = 500$
	3.637	3.671	3.686
(I.21)	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
	3.320	3.173	3.074

Table 9'. Asymptotic 5 % critical values for statistics (I.20) and (I.21).

n	(I.17)	(I.18)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	3.200	3.200	3.144	3.070
100	3.294	3.275	3.203	3.146
200	3.371	3.340	3.262	3.194
300	3.410	3.370	3.286	3.212
500	3.440	3.397	3.301	3.229

Table 10. Simulated 2.5 % critical values for statistics (I.17) and (I.18).

(I.20)	$n = 100$	$n = 300$	$n = 500$
	4.041	4.049	4.056
(I.21)	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
	3.541	3.404	3.312

Table 10'. Asymptotic 2.5 % critical values for statistics (I.20) and (I.21).

n	(I.17)	(I.18)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	3.486	3.486	3.441	3.370
100	3.563	3.546	3.490	3.436
200	3.649	3.623	3.548	3.481
300	3.684	3.649	3.581	3.510
500	3.703	3.664	3.587	3.518

Table 11. Simulated 1 % critical values for statistics (I.17) and (I.18).

(I.20)	$n = 100$	$n = 300$	$n = 500$
	4.570	4.545	4.539
(I.21)	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
	3.809	3.684	3.600

Table 11'. Asymptotic 1 % critical values for statistics (I.20) and (I.21).

The approximate critical values for the sum-type test statistic may be obtained from the convergence

$$\frac{1}{\sigma^2}\sum_{k=1}^{n}\frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{k}\big(Y_i-\overline{Y}_n\big)\right)^2 \xrightarrow{\mathcal{D}} \int_0^1 B^2(t)\,dt = \sum_{j=1}^{\infty}\frac{1}{j^2\pi^2}\,\chi_j^2(1), \quad \text{(I.24)}$$

where $\big\{B(t), 0 \le t \le 1\big\}$ is the Brownian bridge and $\big\{\chi_j^2(1)\big\}$ are independent variables distributed according to the χ^2 distribution with one degree of freedom, see, e.g., Csörgő and Horváth (1997). It follows from Theorem 2.1 of Anderson and Darling (1952), see also Anderson and Darling (1954) and Kiefer (1960), that $\forall x \in \mathcal{R}^1$

$$P\left(\int_0^1 B^2(t)\,dt > x\right) = \quad \text{(I.25)}$$

$$1-\frac{\sqrt{2}}{\pi^{3/2}\sqrt{x}}\sum_{j=0}^{\infty}\frac{\Gamma\big(j+1/2\big)}{\Gamma\big(j+1\big)}\sqrt{2j+\frac{1}{2}}\,\exp\Big\{-\frac{(4j+1)^2}{16x}\Big\}\,K_{1/4}\left(\frac{(4j+1)^2}{16x}\right),$$

where $K_{1/4}(\cdot)$ denotes a modified Bessel function of the second type (see, e.g., function `besselk(nu,z)` in Matlab or function `BesselK` in Mathematica).

Now, let us turn to the case when the variance is unknown. In all statistics (I.17), (I.18) and (I.19) one can replace the unknown σ^2 by its estimate

$$\frac{1}{n}\sum_{i=1}^{n}\Big(Y_i-\overline{Y}_n\Big)^2 \quad \text{(I.26)}$$

and the limit distribution remains the same under H. However, in the case of the maximum-type test statistics, one gets a more powerful test if one applies the test statistics

$$\max_{1\le k\le n-1}\left\{\left|\frac{1}{s_k}\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\} \quad \text{(I.27)}$$

and

$$\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\left|\frac{1}{s_k}\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\Big(Y_i-\overline{Y}_n\Big)\right|\right\}, \quad \text{(I.28)}$$

where

$$s_k^2 = \frac{1}{n-2}\left(\sum_{i=1}^{k}\Big(Y_i - \overline{Y}_k\Big)^2 + \sum_{i=k+1}^{n}\Big(Y_i - \overline{Y}_k^{\,o}\Big)^2\right). \tag{I.29}$$

It is interesting to realize that

$$\sqrt{\frac{n}{k(n-k)}}\ \sum_{i=1}^{k}\big(Y_i - \overline{Y}_n\big) = \sqrt{\frac{k(n-k)}{n}}\ \Big(\overline{Y}_k - \overline{Y}_k^{\,o}\Big),$$

so that all variables, from which the maximum is taken, are the usual two-sample t-statistics that are distributed, under H, according to the t-distribution with $n-2$ degrees of freedom. This fact may be used when the critical values are approximated by the Bonferroni inequality.

Instead of statistic (I.27), we may also use the statistic

$$\max_{1\le k\le n-1}\left\{\left|\frac{1}{\widehat{\sigma}_n}\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\Big(Y_i - \overline{Y}_n\Big)\right|\right\}, \tag{I.30}$$

where

$$\widehat{\sigma}_n^2 = \min_{2\le k\le n-2}\left\{\frac{1}{n}\left(\sum_{i=1}^{k}\Big(Y_i - \overline{Y}_k\Big)^2 + \sum_{i=k+1}^{n}\Big(Y_i - \overline{Y}_k^{\,o}\Big)^2\right)\right\}. \tag{I.31}$$

The statistics (I.27) and (I.30) are asymptotically the same.

For n large, the approximations (I.20) and (I.21) hold also for the statistics (I.27) and (I.30). Aside that, for normally distributed random variables Y_i and for large values x, James et al. (1987) suggested to use the approximation

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\left|\frac{1}{s_k}\sqrt{\frac{n}{k(n-k)}}\sum_{i=1}^{k}\Big(Y_i - \overline{Y}_n\Big)\right|\right\} > x\right) \approx$$

$$\sqrt{\frac{2n}{\pi}}\int_{\frac{x}{\sqrt{n}}}^{1}\Big(1-y^2\Big)^{\frac{n-4}{2}}dy + \frac{x\sqrt{2}}{\sqrt{\pi}}\Big(1-\frac{x^2}{n}\Big)^{\frac{n-4}{2}}\int_x^x \frac{\overline{\frac{1-\beta}{(n-x^2)\beta}}}{\frac{\beta}{(n-x^2)(1-\beta)}}\ \frac{1}{y}\nu\Big(y + \frac{x^2}{(n-x^2)y}\Big)\,dy.$$

Tables 12. – 15. contain critical values obtained by simulations for statistics (I.27) and (I.28) in the case that σ^2 was not known and was estimated using (I.29).

n	(I.27)	(I.28)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	2.857	2.856	2.775	2.683
100	2.891	2.864	2.778	2.694
200	2.934	2.893	2.801	2.715
300	2.961	2.914	2.805	2.718
500	2.993	2.931	2.820	2.728

Table 12. Simulated 10 % critical values for statistics (I.27) and (I.28).

n	(I.27)	(I.28)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	3.157	3.157	3.079	2.992
100	3.164	3.139	3.061	2.984
200	3.196	3.159	3.071	2.992
300	3.213	3.172	3.076	2.994
500	3.241	3.189	3.088	3.003

Table 13. Simulated 5 % critical values for statistics (I.27) and (I.28).

n	(I.27)	(I.28)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	3.421	3.421	3.359	3.279
100	3.402	3.383	3.311	3.248
200	3.428	3.391	3.314	3.245
300	3.452	3.409	3.324	3.246
500	3.462	3.420	3.324	3.251

Table 14. Simulated 2.5 % critical values for statistics (I.27) and (I.28).

n	(I.27)	(I.28)		
		$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$
50	3.747	3.747	3.695	3.625
100	3.696	3.678	3.615	3.558
200	3.719	3.690	3.611	3.558
300	3.737	3.700	3.628	3.555
500	3.735	3.700	3.617	3.547

Table 15. Simulated 1 % critical values for statistics (I.27) and (I.28).

3.2. Change in variance

We suppose that the observations $Y_1, \dots, Y_n$ are independent normally distributed with a known mean μ and unknown variances. Supposing that the mean remains the same, the problem of the detection of a change in variance can be formulated as the following testing problem, i.e. we test the null hypothesis H against the alternative A:

$$\begin{aligned} &H : Y_1 \dots, Y_n \sim N\big(\mu, \sigma^2\big) \\ &A : \exists\, \mathsf{m} \in \{1, \dots, n-1\} \text{ such that} \\ &\quad Y_1, \dots, Y_{\mathsf{m}} \quad\ \sim N\big(\mu, \sigma_1^2\big), \\ &\quad Y_{\mathsf{m}+1}, \dots, Y_n \sim N\big(\mu, \sigma_2^2\big), \end{aligned} \tag{I.32}$$

where $\sigma_1^2 \neq \sigma_2^2$.

It is evident that the variables $\big\{\big(Y_i - \mu\big)^2, i = 1, \dots, n\big\}$ follow a gamma distribution with the shape parameter $\alpha = 1/2$ and the scale parameter $\beta = 2\sigma^2$. It means that our problem leads to a special case of the detection of change in the scale parameter of random variables distributed according to the gamma law with a known and constant shape parameter.

The maximum-type test statistics obtained by the maximum likelihood approach have the form

$$\max_{1 \le k \le n-1} \Big\{ \big| Z_k \big| \Big\} \quad \text{and} \quad \max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta)\, n \rfloor} \Big\{ \big| Z_k \big| \Big\}, \tag{I.33}$$

where

$$\begin{aligned} Z_k^2 &= n \log\left(\frac{1}{n} \sum\nolimits_{i=1}^{n} \big(Y_i - \mu\big)^2\right) - k \log\left(\frac{1}{k} \sum\nolimits_{i=1}^{k} \big(Y_i - \mu\big)^2\right) - \\ &\quad - (n-k) \log\left(\frac{1}{n-k} \sum\nolimits_{i=k+1}^{n} \big(Y_i - \mu\big)^2\right) = \\ &= k \log \frac{k}{n} \frac{\sum_{i=1}^{n} \big(Y_i - \mu\big)^2}{\sum_{i=1}^{k} \big(Y_i - \mu\big)^2} + (n-k) \log \frac{n-k}{n} \frac{\sum_{i=1}^{n} \big(Y_i - \mu\big)^2}{\sum_{i=k+1}^{n} \big(Y_i - \mu\big)^2}. \end{aligned} \tag{I.34}$$

For smaller n, the approximate critical values may be obtained by the Bonferroni inequality realizing that

$$P\left(Z_k^2 < t\right) = P\left(a_k(t) < \frac{\sum_{i=1}^k \left(Y_i-\mu\right)^2}{\sum_{i=1}^n \left(Y_i-\mu\right)^2} < b_k(t)\right),$$

where $a_k(t)$ and $b_k(t)$ are the solutions of the equation

$$-k\log\left(x\frac{n}{k}\right) - (n-k)\log\left((1-x)\frac{n}{n-k}\right) = t.$$

Clearly, the random variable $\sum_{i=1}^k \left(Y_i-\mu\right)^2 \Big/ \sum_{i=1}^n \left(Y_i-\mu\right)^2$ is distributed according to the beta distribution with parameters $k/2$ and $(n-k)/2$. This fact may be used if the approximate critical values are calculated by the Bonferroni inequality, for details see Worsley (1986).

For n large, the limit behavior of the studied probabilities is the same as before, i.e.

$$P\left(\max_{1\le k\le n-1}\left\{\left| Z_k \right|\right\} > \frac{x+b_n}{a_n}\right) \approx 1-\exp\left\{-2e^{-x}\right\}, \quad x\in\mathcal{R}^1, \qquad \text{(I.35)}$$

and

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)\,n\rfloor}\left\{\left| Z_k \right|\right\} > x\right) \approx \left(1-\Phi(x)\right) + 2x\phi(x)\log\frac{1-\beta}{\beta}\cdot \qquad \text{(I.36)}$$

Notice that (I.35) was derived in Gombay and Horváth (1990).

Denote for a while $S_\mu^2 = n^{-1}\sum_{j=1}^n \left(Y_j-\mu\right)^2$. Then the sum-type test statistic has the form

$$\sum_{k=1}^n \frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^k \frac{\left(Y_i-\mu\right)^2 - S_\mu^2}{\sqrt{2}S_\mu^2}\right)^2$$

and for large n the critical values may be computed from the approximation

$$\begin{aligned} P\Bigg(\sum_{k=1}^n \frac{1}{n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^k \frac{\left(Y_i-\mu\right)^2 - S_\mu^2}{\sqrt{2}S_\mu^2}\right)^2 > x\Bigg) &\approx P\left(\int_0^1 B^2(t)\,dt > x\right) \\ &= P\left(\sum_{j=1}^\infty \frac{1}{j^2\pi^2}\,\chi_j^2(1) > x\right) \qquad \text{(I.37)} \end{aligned}$$

For derivation of (I.37) see MacNeill (1974); see also (I.25) and Csörgő and Horváth (1997).

Let us remark that in case the mean μ is unknown, it can be replaced by its estimator $\overline{Y}_n$ and the approximations (I.35) – (I.37) are still valid.

	$\alpha = 0.10$		$\alpha = 0.05$		$\alpha = 0.01$	
n	Simul.	Bonf.	Simul.	Bonf.	Simul.	Bonf.
20	2.679	2.907	2.953	3.129	3.474	3.596
50	2.816	3.153	3.077	3.357	3.605	3.791
100	2.898	3.332	3.152	3.524	3.657	3.938

Table 16. Critical values for over-all (non-trimmed) maximum-type statistic (I.33) calculated using simulations and Bonferroni inequality.

3.3. Change in mean and/or variance

Sometimes it can happen that the change may occur either in one parameter or in both (simultaneously) . Then we test the null hypothesis H against the alternative A:

$$\begin{aligned} &H : Y_1 \dots, Y_n \sim N\big(\mu, \sigma^2\big) \\ &A : \exists\, \mathsf{m} \in \{2, \dots, n-2\} \text{ such that} \\ &\quad Y_1, \dots, Y_{\mathsf{m}} \quad \sim N\big(\mu_1, \sigma_1^2\big), \\ &\quad Y_{\mathsf{m}+1}, \dots, Y_n \sim N\big(\mu_2, \sigma_2^2\big), \end{aligned} \tag{I.38}$$

where $\big(\mu_1, \sigma_1^2\big) \neq \big(\mu_2, \sigma_2^2\big)$. The test statistic derived by likelihood ratio approach has the form

$$\max_{1 \le k \le n-1} \Big\{ \big| \widetilde{Z}_k \big| \Big\} \quad \text{and} \quad \max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta)\, n \rfloor} \Big\{ \big| \widetilde{Z}_k \big| \Big\}, \tag{I.39}$$

where

$$\begin{aligned} \widetilde{Z}_k^2 = &\, n \log \left(\frac{1}{n} \sum_{i=1}^{n} \big(Y_i - \overline{Y}_n\big)^2 \right) - k \log \left(\frac{1}{k} \sum_{i=1}^{k} \big(Y_i - \overline{Y}_k\big)^2 \right) \\ &- (n-k) \log \left(\frac{1}{n-k} \sum_{i=k+1}^{n} \big(Y_i - \overline{Y}_k^{\,o}\big)^2 \right). \end{aligned} \tag{I.40}$$

For n large, the approximate critical values may be computed from the following approximations, i.e.,

$$P \left(\max_{1 \le k \le n-1} \Big\{ \big| \widetilde{Z}_k \big| \Big\} > \frac{x + b_{n,2}}{a_n} \right) \approx 1 - \exp\big\{ -2e^{-x} \big\}, \quad x \in \mathcal{R}^1, \tag{I.41}$$

where

$$a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_{n,2} = 2 \log \log n + \log \log \log n,$$

and

$$P \left(\max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta)\, n \rfloor} \Big\{ \big| \widetilde{Z}_k \big| \Big\} > x \right) \approx e^{-x^2/2} + e^{-x^2/2} x^2 \log \frac{1-\beta}{\beta} \cdot \tag{I.42}$$

The formulas (I.41) and (I.42) follow from the approximation of the sequence of $\{\widetilde{Z}_k^2\}$ by the process $\{(B_1^2(t) + B_2^2(t))/(t(1-t))\}$, see Part III Sections 11 and 12. The approximation (I.41) is due to Horváth (1993), for (I.42) see Albin (1990). Unfortunately, we do not know how to derive the sum-type test statistic in this case.

4. Detection of changes in a location model (general type of errors)

This section deals with detection of changes in a location model when the distribution of errors is not necessarily normal. Three basic situations, i.e. change in location, change in scale and change in location and/or scale are considered. We concentrate on the maximum-type test statistics.

4.1. Changes in location

We consider the same problem as in 3.1., however we will not assume that the error terms have a normal distribution. Of course, the situation when e_i are normally distributed represents a special case.

More precisely, we are interested in testing the null hypothesis H against the alternative A:

$$\begin{aligned} H &: Y_i = \mu + e_i, && i = 1, \ldots, n, && \text{(I.43)} \\ A &: \exists\, \mathsf{m} \in \{1, \ldots, n-1\} \quad \text{such that} \\ & \quad Y_i = \mu + e_i, && i = 1, \ldots, \mathsf{m}, \\ & \quad Y_i = \mu + \delta + e_i, && i = \mathsf{m}+1, \ldots, n, \ \delta \neq 0, \end{aligned}$$

where μ, $\delta \neq 0$ and m are parameters, $e_1, \ldots, e_n$ are iid random variables with distribution function F, zero mean, nonzero variance σ^2 and with $\mathsf{E}\,|e_i|^{2+\Delta} < \infty$ for some $\Delta > 0$.

If the distribution of the errors is known up to some parameters and only one parameter is subject to a change, we can apply the likelihood ratio method (see Section 2.1. for the description of the general principle), which means that we can derive $\max_{1 \le k < n} \Lambda_k$ in such a case. Under quite mild assumptions on the distribution of e_i one can use the same approximations to the distribution of the corresponding test statistic, see Csörgő and Horváth (1997) for details. Of course, the approximation is usually poorer than in the case of a normal distribution in Section 3.1..

If our knowledge about the distribution of the error terms e_i is poor or their distribution is too complicated, we can still use the test statistics described in Section 3 (see (I.17) – (I.19)) with the same approximation to the distribution of the test statistics. However, the quality of these approximations is (much) worse than in the case of a normal distribution.

Nevertheless, it can be shown that, under quite mild assumptions, these tests are consistent, i.e. they are sensitive to the considered alternatives. We are not going to prove it, but we demonstrate this on the expectations and variances of the partial sums

$$S_j = \sum_{i=1}^{j} \Big(Y_i - \overline{Y}_n\Big), \quad j = 1, \ldots, n, \tag{I.44}$$

under both the null hypothesis H and the alternative A. $\big($Notice that most of the tests considered in Section 3 were based on $S_j\big)$.

$$\begin{aligned} \mathsf{E}_H\, S_j &= 0, && 1 \le j \le n,\\ \mathsf{E}_A\, S_j &= \begin{cases} -j\delta\,\dfrac{n-\mathsf{m}}{n}, & 1 \le j \le \mathsf{m},\\ -\mathsf{m}\delta\,\dfrac{n-j}{n}, & \mathsf{m} < j \le n, \end{cases}\\ var_H\, S_j = var_A\, S_j &= \sigma^2\,\frac{j(n-j)}{n}, && 1 \le j \le n. \end{aligned}$$

It is seen that the expectation under the null hypothesis H is zero while the expectation under A is nonzero with the extreme reached for $j = \mathsf{m}$, i.e.

$$\max_{1 \le j < n} \big|\mathsf{E}_A S_j\big| = \big|\mathsf{E}_A S_{\mathsf{m}}\big| = \big|\delta\big|\,\frac{\mathsf{m}(n-\mathsf{m})}{n}.$$

We should also remark that by the central limit theorem S_k has approximately $\big($for k large$\big)$ normal distribution $N(E\,S_k,\, var\,S_k)$ both under H and A.

To test H against A, other procedures were also developed. Similarly, as in the two-sample problem, there are various robust and nonparametric test procedures. Let us mention M-tests, rank-based R-tests, Kolmogorov-Smirnov type tests among others.

Noticing that the two-sample problem coincides with our problem if k is known, we can develop M- and R- test procedures, as well as Kolmogorov-Smirnov type tests for our problem along the lines of the two-sample tests.

4.1.1. M-test procedures

We consider again the testing problem (I.43) assuming that e_i have a common distribution function F that is symmetric around zero, i.e. generally there is no need to have zero mean and a finite variance.

We remind the definition of M-estimators with the score function ψ, that is usually assumed to be monotone and skew symmetric, i.e. $\psi(x) = -\psi(x)\ \forall x \in \mathcal{R}^1$. The M-estimator $\widehat{\mu}_{n,M}(\psi)$ of μ is defined as any solution of the equation

$$\sum_{i=1}^{n} \psi\Big(Y_i - t\Big) = 0.$$

The basic theory of M-procedures was developed by Huber, for details see his book Huber (1981), and further studied in a number of papers and books.

The test procedures for the change point problem are based on the partial sums

$$S_{k,M}(\psi) = \sum_{i=1}^{k} \psi\Big(Y_i - \widehat{\mu}_{n,M}(\psi)\Big), \quad k = 1, \ldots, n-1. \tag{I.45}$$

The role of the scale plays $\sigma(\psi)$ defined by

$$\sigma^2(\psi) = \int_{-\infty}^{\infty} \psi^2(x)\, dF(x),$$

which can be estimated for example by the estimator

$$\widehat{\sigma}^2_{n,M}(\psi) = \tag{I.46}$$

$$= \min_{2 \le k \le n-2} \left\{ \frac{1}{n} \left(\sum_{i=1}^{k} \psi^2\Big(Y_i - \widehat{\mu}_{k,M}(\psi)\Big) + \sum_{i=k+1}^{n} \psi^2\Big(Y_i - \widehat{\mu}^o_{k,M}(\psi)\Big) \right) \right\},$$

where $\widehat{\mu}_{k,M}(\psi)$ and $\widehat{\mu}^o_{k,M}(\psi)$ are the M-estimators calculated from $Y_1, \ldots, Y_k$ and $Y_{k+1}, \ldots, Y_n$, respectively.

Then, the maximum-type M-test statistic is defined as

$$\max_{1 \le k < n} \left\{ \frac{1}{\widehat{\sigma}_{n,M}(\psi)} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \psi\Big(Y_i - \widehat{\mu}_{n,M}(\psi)\Big) \right| \right\}. \tag{I.47}$$

The quantities $\psi\big(Y_i - \widehat{\mu}_{n,M}(\psi)\big)$, $i = 1, \ldots, n$, play the role of robust residuals. It can be easily checked that for $\psi(x) = x$, $x \in \mathcal{R}^1$, statistic (I.47) reduces to (I.27). Similarly, as in Section 3 we can approximate under certain assumptions the distribution of (I.47) under the null hypothesis using

$$P\left(\max_{1 \le k < n} \left\{ \frac{1}{\widehat{\sigma}_{n,M}(\psi)} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \psi\Big(Y_i - \widehat{\mu}_{n,M}(\psi)\Big) \right| \right\} > \frac{x + b_n}{a_n} \right) \approx$$
$$\approx 1 - \exp\{-2e^{-x}\}, \quad x \in \mathcal{R}^1, \tag{I.48}$$

where

$$a_n = \sqrt{2 \log\, \log\, n} \quad \text{and} \quad b_n = 2 \log\, \log\, n + \frac{1}{2} \log \log\, \log\, n - \frac{1}{2} \log\, \pi.$$

Typical ψ-functions are summarized in Table 17. For detailed discussion concerning the choice of parameters A, B, C and D see, e.g., Huber (1981), Hampel et al. (1986) or Antoch and Víšek (1992). Notice that, letting A, B, C and $D \to \infty$ gives in all four cases $\psi(x) = x$, i.e. the classical least squares estimator.

If we choose $\psi(x) = -f'(x)/f(x), x \in \mathcal{R}^1$, where f and f' are, respectively, density and its derivative, then under some assumptions the resulting maximum-type test statistic is asymptotically equivalent to that obtained by the likelihood ratio method.

	$\varrho(x)$	$\psi(x)$	
Fair	$A^2\Big(\frac{\lvert x\rvert}{A} - \ln\big(1+\frac{\lvert x\rvert}{A}\big)\Big)$	$\frac{x}{1+\lvert x\rvert/A}$	$x \in \mathcal{R}_1$
Huber	$\frac{x^2}{2}$ $B\mid x\mid - \frac{B^2}{2}$	x $B\,\mathrm{sign}\,(x)$	$\mid x\mid \le B$ $\mid x\mid > B$
Tukey	$\frac{C^2}{6}\Big(1-\big(1-(\frac{x}{C})^2\big)^3\Big)$ $\frac{C^2}{6}$	$x\big(1-(\frac{x}{C})^2\big)^2$ 0	$\mid x\mid \le C$ $\mid x\mid > C$
Welsh	$\frac{D^2}{2}\Big(1-\exp\big\{-(\frac{x}{D})^2\big\}\Big)$	$x\exp\{-(\frac{x}{D})^2\}$	$x \in \mathcal{R}_1$

Table 17. Typical ψ - functions generating M - estimators.

Concerning the choice of the score function ψ, statisticians often use

$$\psi(x) = \begin{cases} x, & |x| \le K, \\ K\,\mathrm{sign}(x), & |x| \ge K, \end{cases}$$

the so-called *Huber function*, where K is a suitably chosen constant. Letting $K \to \infty$ one has the classical least squares estimator while, for $K \to 0$ the so-called L_1-norm estimator is attained. Huber's ψ-function was proposed as a function leading to the estimators not influenced by outliers as, e.g., least squares estimator is. The use of Huber's ψ-function is wise because the procedures considered in Section 4.1. are sensitive with respect to outliers which could be wrongly detected as change points.

We mention here the important case with $\psi(x) = \mathrm{sign}(x)$, $x \in \mathcal{R}^1$, the so called L_1 test procedure. The test statistic is very easy to calculate in this case. We assume that the distribution of the error terms e_i is symmetric around zero and has the density positive in the neighborhood of zero.

Then the test statistic has the form

$$\max_{1\le k<n} \sqrt{\frac{n}{k(n-k)}}\, \left| \sum_{i=1}^{k} \mathrm{sign}\left(Y_i - \widetilde{\mu}_n\right) \right|,$$

where $\widetilde{\mu}_n$ is the sample median based on all observations. This means that the partial sums are just the differences of the number of the observations Y_i above and below the sample median $\widetilde{\mu}_n$.

4.1.2. R-test procedures

Now we turn to the rank-based test procedures. The test statistics are based on the simple linear rank statistics

$$S_{k,R} = \sum_{i=1}^{k} \Big(a_n(R_i) - \overline{a}_n\Big), \quad k = 1, \ldots, n, \tag{I.49}$$

where $R_1, \ldots, R_n$ is the vector of ranks corresponding to the observations $Y_1, \ldots, Y_n$, $a_n(1), \ldots, a_n(n)$ are scores the and $\overline{a}_n = \frac{1}{n}\sum_{i=1}^n a_n(i)$.

Here we need the continuity of the distribution function F of the observations, no other assumption on F is needed. Typical choices of the scores are, e.g.,

(1) Wilcoxon scores $a_n(i) = i/(n+1),\ i = 1, \ldots, n;$
(2) van der Waerden scores $a_n(i) = \Phi^{-1}\big(i/(n+1)\big),\ i = 1, \ldots, n$, etc.

The role of the scale is played by

$$\sigma_{n,R}^2 = \frac{1}{n-1}\sum_{i=1}^n \Big(a_n(i) - \overline{a}_n\Big)^2, \tag{I.50}$$

which needs not be estimated, while the role of residuals by $\big\{a_n(R_i) - \overline{a}_n,\ i = 1, \ldots, n\big\}$.

The maximum-type R-test statistic is defined as

$$\max_{1 \le k < n} \left\{ \frac{1}{\sigma_{n,R}} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k \Big(a_n(R_i) - \overline{a}_n\Big) \right| \right\}. \tag{I.51}$$

Similarly, as in the previous section, we can approximate under certain assumptions the distribution of (I.51) under the null hypothesis using

$$P\left(\max_{1 \le k < n} \left\{ \frac{1}{\sigma_{n,R}(\psi)} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k \Big(a_n(R_i) - \overline{a}_n\Big) \right| \right\} > \frac{x + b_n}{a_n} \right) \approx$$
$$\approx 1 - \exp\{-2e^{-x}\}, \quad x \in \mathcal{R}^1, \tag{I.52}$$

where

$$a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_n = 2 \log \log n + \frac{1}{2} \log\log \log n - \frac{1}{2} \log \pi.$$

However, this approximation is reasonable only when n is large. Nevertheless, under the null hypothesis H, the distribution of the R-test statistics does not depend on F. Therefore, we can get the approximation to the distribution of the R-test statistics via simulations, i.e., we can simulate the samples for example from the uniform distribution $R(0, 1)$. The basic advantage is that these simulated critical values give good approximations even for n small. One can also use various modifications developed along the lines of Section 3.

Example: One of the most often used rank-based tests is the test with the Wilcoxon scores. Corresponding (change point) R-test, which is easy to calculate, behaves reasonably well for a broad spectrum of the distributions F. Here

$$\sigma_{n,R}^2 = \frac{n}{12(n+1)} \approx \frac{1}{12},$$

hence the test statistic (I.51) becomes

$$\max_{1 \le k < n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{\sqrt{12}}{n} \left| \sum_{i=1}^k \left(R_i - \frac{n+1}{2}\right) \right| \right\}.$$

REMARK: There were developed also U-test statistics and Kolmogorov-Smirnov type test statistics for our problem. For more information we refer to the book Csörgő and Horváth (1997).

4.1.3. MOSUM type test statistics

Now, we introduce two different classes of test statistics for our problem. They are based on the moving sums (MOSUM) of statistics S_k introduced in (I.44) and defined as

$$\max_{G<k\le n} \frac{1}{\sqrt{G}}\frac{1}{\widehat{\sigma}_n}\left|S_k - S_{k-G}\right| = \max_{G<k\le n}\frac{1}{\sqrt{G}}\frac{1}{\widehat{\sigma}_n}\Big|\sum_{i=k-G+1}^{k}\big(Y_i-\overline{Y}_n\big)\Big| \tag{I.53}$$

and

$$\max_{G<k\le n-G}\frac{1}{\sqrt{2G}}\frac{1}{\widehat{\sigma}_n}\left|S_{k+G}-2S_k+S_{k-G}\right| = \tag{I.54}$$
$$\max_{G<k\le n-G}\frac{1}{\sqrt{2G}}\frac{1}{\widehat{\sigma}_n}\Big|\sum_{i=k+1}^{k+G}Y_i - \sum_{i=k-G+1}^{k}Y_i\Big|,$$

where $\widehat{\sigma}_n$ is any consistent estimator of σ, e.g. that defined by (I.31) and G/n small. Typically we choose $G/n \sim 0.05$ or 0.10.

Notice that (I.53) is the first order difference (the first order derivative) of S_k's while (I.54) corresponds to the second order difference (the second order derivative) of S_k's.

If n is large and G/n is small (e.g. $G/n \sim 0.1$ or 0.05) we can use the following approximation to the distributions of (I.53) and (I.54) under the null hypothesis H, i.e.

$$P\left(\max_{G<k\le n}\left\{\frac{1}{\sqrt{G}}\frac{1}{\widehat{\sigma}_n}\left|S_k-S_{k-G}\right|\right\} > \frac{x+b_{n,G}}{a_{n,G}}\right)\approx 1-\exp\left\{-2e^{-x}\right\},\ x\in\mathcal{R}^1,$$

and

$$P\left(\max_{G<k\le n-G}\left\{\frac{1}{\sqrt{2G}}\frac{1}{\widehat{\sigma}_n}\left|S_{k+G}-2S_k+S_{k-G}\right|\right\} > \frac{x+b_{n,G}-\log\frac{2}{3}}{a_{n,G}}\right)\approx$$
$$\approx 1-\exp\left\{-2e^{-x}\right\},\quad x\in\mathcal{R}^1,$$

where

$$a_{n,G} = \sqrt{2\log\frac{n}{G}} \quad\text{and}\quad b_{n,G} = 2\log\frac{n}{G}+\frac{1}{2}\log\log\frac{n}{G}-\frac{1}{2}\log\pi.$$

Large values of these test statistics indicate that at least one change has occurred. Using the above approximations we get the approximation to the critical value with level α. Namely, we obtain the critical regions with approximate level α

$$\max_{G<k\leq n}\left\{\frac{1}{\sqrt{G}}\,\frac{1}{\widehat{\sigma}_n}\,\Big|S_k - S_{k-G}\Big|\right\} > \\ > \frac{-\log\log\frac{1}{\sqrt{1-\alpha}} + 2\log\,\frac{n}{G} + \frac{1}{2}\log\log\,\frac{n}{G} - \frac{1}{2}\log\pi}{\sqrt{2\log\,\frac{n}{G}}}$$

and

$$\max_{G<k\leq n-G}\left\{\frac{1}{\sqrt{2G}}\,\frac{1}{\widehat{\sigma}_n}\,\Big|S_{k+G} - 2S_k + S_{k-G}\Big|\right\} > \\ > \frac{-\log\log\frac{1}{\sqrt{1-\alpha}} + 2\log\,\frac{n}{G} + \frac{1}{2}\log\log\,\frac{n}{G} - \frac{1}{2}\log\pi - \log\frac{2}{3}}{\sqrt{2\log\,\frac{n}{G}}}.$$

To get a simple picture on the sensitivity of the test statistics (I.53) and (I.54) with respect to alternatives, we present the expectations and variances of the moving sums both under the null hypothesis H and alternative A:

$$\mathsf{E}_H\big(S_j - S_{j-G}\big) = 0, \qquad G < j \leq n,$$

$$\mathsf{E}_A\big(S_j - S_{j-G}\big) = \begin{cases} -\delta G\dfrac{n-\mathsf{m}}{n}, & G < j \leq \mathsf{m}, \\ \delta\left(j - \mathsf{m} - G\dfrac{n-\mathsf{m}}{n}\right), & \mathsf{m} < j \leq \mathsf{m} + G, \\ \delta\, G\,\dfrac{\mathsf{m}}{n}, & \mathsf{m} + G < j \leq n. \end{cases}$$

$$\mathsf{E}_H\Big(S_{j+G} - 2S_j + S_{j-G}\Big) = 0, \qquad G < j \leq n - G,$$

$$\mathsf{E}_A(S_{j+G} - 2S_j + S_{j-G}) = \begin{cases} 0, & G < j \leq \mathsf{m} - G, \\ \delta\big(j + G - \mathsf{m}\big), & \mathsf{m} - G < j \leq \mathsf{m}, \\ \delta\big(\mathsf{m} + G - j\big), & \mathsf{m} < j \leq \mathsf{m} + G, \\ 0, & \mathsf{m} + G < j \leq n - G, \end{cases}$$

$$var_H\big(S_j - S_{j-G}\big) = var_A\big(S_j - S_{j-G}\big) = \sigma^2\frac{G(n-G)}{n}, \quad G < j \leq n,$$

$$var_H\big(S_{j+G} - 2S_j + S_{j-G}\big) = var_A\big(S_{j+G} - 2S_j + S_{j-G}\big) = 2\sigma^2 G,\ G < j \leq n - G.$$

We see that the expectations under H are zero, while under alternatives the expectations of $S_j - S_{j-G}$ are nonzero. Concerning $S_{j+G} - 2S_j + S_{j-G}$, their expectations (under A) are nonzero only for j close to the change point m $\big(|j - \mathsf{m}| \leq G\big)$. Hence, both test statistics are sensitive with respect to the considered alternatives. For this particular reason the test statistic (I.54) is suitable if we expect more changes and it is useful as a diagnostic tool.

♣ Similarly, the MOSUM type M-test statistics can be proposed, just replacing in (I.53) and (I.54) S_k and $\widehat{\sigma}_n$ by their M-type counterparts $S_{k,M}\big(\psi\big)$ and $\widehat{\sigma}_{n,M}\big(\psi\big)$ defined by (I.45) and (I.46), i.e., one can use

$$\max_{G<k\leq n}\left\{\frac{1}{\sqrt{G}}\,\frac{1}{\widehat{\sigma}_{n,M}\big(\psi\big)}\,\Big|\,S_{k,M}\big(\psi\big) - S_{k-G,M}\big(\psi\big)\,\Big|\right\} \tag{I.55}$$

and

$$\max_{G<k\le n-G}\left\{\frac{1}{\sqrt{2G}}\,\frac{1}{\widehat{\sigma}_{n,M}(\psi)}\,\Big|\,S_{k+G,M}(\psi)-2S_{k,M}(\psi)+S_{k-G,M}(\psi)\,\Big|\right\}. \tag{I.56}$$

♣ The MOSUM R-type statistics can be derived quite analogously. Namely we use, instead of (I.53) and (I.54), the test statistics

$$\max_{G<k\le n}\left\{\frac{1}{\sqrt{G}}\,\frac{1}{\sigma_{n,R}}\,\Big|\,S_{k,R}-S_{k-G,R}\,\Big|\right\} \tag{I.57}$$

and

$$\max_{G<k\le n-G}\left\{\frac{1}{\sqrt{2G}}\,\frac{1}{\sigma_{n,R}}\,\Big|\,S_{k+G,R}-2S_{k,R}+S_{k-G,R}\,\Big|\right\}, \tag{I.58}$$

where $S_{k,R}$ and $\sigma^2_{n,R}$ were introduced in (I.49) and (I.50).

4.2. Change in scale

We consider the same problem as in Section 3.2., however, we will not assume that the error terms have a normal distribution. Supposing that the mean remains the same, the problem of the detection of a change in variance can be formulated as the following testing problem, i.e. we test the null hypothesis H against the alternative A:

$$\begin{aligned}
&H: Y_i=\mu+e_i, && i=1,\ldots,n, \\
&A: \exists\,\mathsf{m}\in\{1,\ldots,n-1\}\quad\text{such that} && \\
&\qquad Y_i=\mu+e_i, && i=1,\ldots,\mathsf{m}, \\
&\qquad Y_i=\mu+(1+h)e_i, && i=\mathsf{m}+1,\ldots,n,
\end{aligned} \tag{I.59}$$

where $\mu\neq 0, h\neq 0,-1$ and m are parameters (realize that only the scale can change, not the mean), $e_1,\ldots,e_n$ are iid random variables with distribution function F, zero mean, nonzero variance σ^2 and $\mathsf{E}\,e_i^{4+\Delta}<\infty$ for some $\Delta>0$. Of course, the situation with e_i normally distributed as $N(0,\sigma^2)$ is a special case.

If the distribution of the error terms is known, we can apply the maximum likelihood principle in order to get a test procedure. Nevertheless, here we construct another test procedure. Testing H against A means that we are looking for a change in the expected values of $\{(Y_i-\mu)^2,\ i=1,\ldots,n\}$, and, since μ is assumed unknown, we replace it by its estimator $\overline{Y}_n$. In this light we propose to use tests for H against A based on the partial sums

$$S_{k,sc}=\sum_{i=1}^{k}\left(\left(Y_i-\overline{Y}_n\right)^2-\frac{1}{n}\sum\nolimits_{j=1}^{n}\left(Y_j-\overline{Y}_n\right)^2\right),\quad k=1,\ldots,n. \tag{I.60}$$

The differences

$$\left(Y_i-\overline{Y}_n\right)^2-\frac{1}{n}\sum\nolimits_{j=1}^{n}\left(Y_j-\overline{Y}_n\right)^2,\quad i=1,\ldots,n,$$

play the role of the residuals. The maximum-type test statistics are constructed in the same way as in the previous section, i.e. we use the test statistics

$$\max_{1\le k<n}\left\{\left|\sqrt{\frac{n}{k(n-k)}}\frac{1}{\widehat{\sigma}_{n,sc}}\sum_{i=1}^{k}\left(\left(Y_i-\overline{Y}_n\right)^2-\frac{1}{n}\sum_{j=1}^{n}\left(Y_j-\overline{Y}_n\right)^2\right)\right|\right\}, \tag{I.61}$$

where $\widehat{\sigma}^2_{n,sc}$ is an estimator of $var\left(Y_1-\mu\right)^2$, for example

$$\begin{aligned}\widehat{\sigma}^2_{n,sc}=\min_{1\le k<n}&\left\{\frac{1}{n}\left(\sum_{i=1}^{k}\left[\left(Y_i-\overline{Y}_k\right)^2-\frac{1}{k}\sum\nolimits_{j=1}^{k}\left(Y_j-\overline{Y}_k\right)^2\right]^2+\right.\right.\\&\left.\left.+\sum_{i=k+1}^{n}\left[\left(Y_i-\overline{Y}^o_k\right)^2-\frac{1}{n-k}\sum\nolimits_{j=k+1}^{n}\left(Y_j-\overline{Y}^o_k\right)^2\right]^2\right)\right\}.\end{aligned} \tag{I.62}$$

Under H the distribution of the test statistics (I.61) can be approximated similarly as in the analogous maximum-type test statistics, i.e. for large n the critical values may be computed from the approximation

$$\begin{aligned}P\Bigg(\max_{1\le k<n}&\left\{\left|\sqrt{\frac{n}{k(n-k)}}\frac{1}{\widehat{\sigma}_{n,sc}}\sum_{i=1}^{k}\left[\left(Y_i-\overline{Y}_n\right)^2-\frac{1}{n}\sum_{j=1}^{n}\left(Y_j-\overline{Y}_n\right)^2\right]\right|\right\}>\\&>\frac{x+b_n}{a_n}\Bigg)\approx 1-\exp\left\{-e^{-x}\right\},\quad x\in\mathcal{R}^1,\end{aligned}$$

where

$$a_n=\sqrt{2\log\,\log\,n}\quad\text{and}\quad b_n=2\log\,\log\,n+\frac{1}{2}\log\log\,\log\,n-\frac{1}{2}\log\,\pi.$$

♣ Concerning robust and nonparametric test procedures, they can be developed again along the lines in the previous section. We present here only the test based on an L_1-procedure that is really useful in practice. Concerning the error terms, we assume that they have symmetric distribution function F (no need of finiteness of any moment). The test statistic is of the form

$$\max_{1\le k<n}\sqrt{\frac{n}{k(n-k)}}\ \sum_{i=1}^{k}\text{sign}\left(\left|\,Y_i-\widetilde{\mu}_n\,\right|-\widetilde{\nu}_n\right),$$

where $\widetilde{\mu}$ and $\widetilde{\nu}_n$ are sample median based on $Y_1,\ldots,Y_n$ and the sample median of $\left|\,Y_i-\widetilde{\mu}_n\,\right|,\ i=1,\ldots,n$. This is a very simple test statistics. Again the same approximation to its distribution under H can be used.

Remark: We could also develop MOSUM type and Bayesian test statistic here. In fact, it suffices to replace the partial sums of residuals S_k and the scale estimator $\widehat{\sigma}$ by $S_{k,sc}$ and $\widehat{\sigma}_{k,sc}$, respectively. The approximations to the distribution of these new test statistics are the same as above.

4.3. Change in location and/or scale

Now we consider the same situation as in the previous section, however, aside the scale also the mean can change. This can be formulated as the following testing problem, i.e., we test the null hypothesis H against the alternative A:

$$\begin{aligned} H: Y_i &= \mu + e_i, & i &= 1,\dots,n, \\ A: \exists\, \mathsf{m} &\in \{1,\dots,n-1\} \quad \text{such that} \\ Y_i &= \mu + e_i, & i &= 1,\dots,\mathsf{m}, \\ Y_i &= \mu + \delta + (1+h)e_i, & i &= \mathsf{m}+1,\dots,n, \end{aligned} \tag{I.63}$$

where μ, $(\delta, h) \neq (0,0), h \neq -1$ and m are parameters, $e_1,\dots,e_n$ are iid random variables with distribution function F, zero mean, nonzero variance σ^2 and $\mathsf{E}\, e_i^{4+\Delta} < \infty$ for some $\Delta > 0$. Again, the situation with e_i having the normal distribution $N(0,\sigma^2)$ is a special case.

The basic test procedures are based either on (I.40) or on the quadratic form of the partial sums S_k and $S_{k,sc}$ introduced in (I.44) and (I.60). More precisely, the corresponding maximum-type statistic is defined as

$$\max_{1\le k<n} \left\{ \sqrt{\frac{n}{k(n-k)}} \sqrt{\frac{S_k^2}{\widehat{\sigma}_n^2} + \frac{S_{k,sc}^2}{\widehat{\sigma}_{n,sc}^2}} \right\}, \tag{I.64}$$

where $\widehat{\sigma}_n^2$ and $\widehat{\sigma}_{n,sc}^2$ were introduced in (I.31) and (I.62).

Comparing the test statistics with their counterparts for a change in location only and for a change in variance, we see that the test statistic (I.64) is the maximum over m of the roots of quadratic forms of the terms for testing a change in location and a change in scale separately. It can be shown, assuming $\mathsf{E}\, e_i^3 = 0$, that

$$cov\,(S_k, S_{k,sc}) = 0, \quad k = 1,\dots,n.$$

Further, under H and for k large it can be shown that

$$\frac{n}{k(n-k)} \left(\frac{S_k^2}{\widehat{\sigma}_n^2} + \frac{S_{k,sc}^2}{\widehat{\sigma}_{n,sc}^2} \right) \tag{I.65}$$

has approximately a χ^2-distribution with two degrees of freedom.

Moreover, it can be shown that under H

$$P\left(\max_{1\le k<n} \left\{ \sqrt{\frac{n}{k(n-k)}} \sqrt{\frac{S_k^2}{\widehat{\sigma}_n^2} + \frac{S_{k,sc}^2}{\widehat{\sigma}_{n,sc}^2}} \right\} > \frac{x + b_{n,2}}{a_n} \right) \approx \\ \approx 1 - \exp\left\{ -2e^{-x} \right\}, \quad x \in \mathcal{R}^1, \tag{I.66}$$

where

$$a_n = \sqrt{2 \log\log n} \quad \text{and} \quad b_{n,2} = 2\log\log n + \log\log\log n.$$

This approximation can be again used to get the approximation to the critical values. Concerning test statistic (I.40) under H_0 and under additional assumptions $\mathsf{E}\, e_i = 0, \mathsf{E}\, e_i^3 = 0$ and $0 < \mathsf{E}\, e_i^4 = 3\, var\ e_i < \infty$, the approximation (I.41) is still reasonable. For details see Horváth (1995).

Notice that introduction of the L_1 test procedures is very simple (we have to assume only that the error terms have a symmetric distribution around zero and a positive continuous density in a neighborhood of zero). The test procedure is based on the statistic

$$\max_{1\le k<n} \left\{ \sqrt{\frac{n}{k(n-k)}}\ \sqrt{\sum_{i=1}^{k} \Big(\operatorname{sign}\big(Y_i - \widetilde{\mu}_n\big)\Big)^2 + \sum_{i=1}^{k} \Big(\operatorname{sign}\big(\,|\,Y_i - \widetilde{\mu}_n\,| - \widetilde{\nu}_n\big)\Big)^2} \right\},$$

where $\widetilde{\mu}$ and $\widetilde{\nu}_n$ are the sample median based on $Y_1, \ldots, Y_n$ and the sample median of $\big|\,Y_i - \widetilde{\mu}_n\,|,\ i = 1, \ldots, n$. Under H, its distribution can be approximated in the same way as in (I.66).

However, as soon as we can assume that the error terms have the first four moments (almost) the same as the double exponential distribution, then it is more convenient to use the test statistic

$$\max_{1\le k<n} \left\{ \sqrt{-2\log \frac{\widetilde{\sigma}_k^k (\widetilde{\sigma}_k^o)^{n-k}}{\widetilde{\sigma}_n^n}} \right\},$$

where

$$\widetilde{\sigma}_k = \frac{1}{k}\sum_{i=1}^{k} \Big|\, Y_i - \widetilde{\mu}_k \,\Big| \quad \text{and} \quad \widetilde{\sigma}_k^o = \frac{1}{n-k}\sum_{i=k+1}^{n} \Big|\, Y_i - \widetilde{\mu}_k^o \,\Big|$$

with $\widetilde{\mu}_k$ and $\widetilde{\mu}_k^o$ being medians based on $X_1, \ldots, X_k$ and $X_{k+1}, \ldots, X_n$, respectively. Under H, the distribution can be approximated again as in (I.66).

4.4. Change in mean with unknown starting value – dependent observations

Consider the case of testing null hypothesis H against the alternative A:

$$\begin{aligned}
&H : Y_i = \mu + e_i, && i = 1, \ldots, n, && \text{(I.67)}\\
&A : \exists\, \mathsf{m} \in \{1, \ldots, n-1\} \quad \text{such that} && && \\
&\qquad Y_i = \mu + e_i, && i = 1, \ldots, \mathsf{m}, && \\
&\qquad Y_i = \mu + \delta + e_i, && i = \mathsf{m}+1, \ldots, n,\ \delta \neq 0, &&
\end{aligned}$$

where μ, $\delta \neq 0$ and m are parameters, and where the variables e_i are not anymore independent, but form a stationary sequence.

Here we can again apply the tests described in Section 3.1. when σ is replaced by a proper estimator or, in other words, a different standardization is needed. We will discuss it further. More precisely, we assume that in (I.67) the error terms e_i form a linear process satisfying

$$e_i = \sum_{t=0}^{\infty} w_t \epsilon_{i-t}, \quad i = 1, 2, \ldots,$$

where ϵ_i are iid random variables, $\mathsf{E}\,\epsilon_i = 0$, $var\,\epsilon_i = \sigma^2 > 0$, $\mathsf{E}\,|\epsilon_i|^{2+\triangle} < \infty$ for some $\triangle > 0$ and the weights $\{w_j\}_{j=0}^{\infty}$ satisfy

$$\sum_{j=0}^{\infty} j\,|w_j| < \infty, \quad \sum_{j=0}^{\infty} w_j \neq 0.$$

Notice that AR and ARMA processes fulfill these assumptions.

In procedures described in Section 3.1. σ^2 has to be replaced by

$$\sigma_0^2 = \sigma^2 \Big(\sum_{j=0}^{\infty} w_j \Big)^2.$$

For example, if the sequence e_n is an AR(1) sequence with the coefficient $\rho \in (-1, 1)$, then σ^2 in (II.9) has to be replaced by $\sigma^2/(1-\rho)$ or by its estimator.

Generally, σ_0^2 can be estimated, e.g., by

$$\widehat{\sigma}_{0,n}^2(L) = \widehat{R}(0) + 2 \sum_{k=1}^{L} \left(1 - \frac{k}{L} \right) \widehat{R}(k), \quad L < n,$$

where for $k \geq 0$

$$\widehat{R}(k) = \frac{1}{n} \left\{ \sum_{t=1}^{\widehat{\mathsf{m}}-k} \left(Y_t - \overline{Y}_{\widehat{\mathsf{m}}} \right) \left(Y_{t+k} - \overline{Y}_{\widehat{\mathsf{m}}} \right) + \sum_{t=\widehat{\mathsf{m}}+1}^{n-k} \left(Y_t - \overline{Y}_{\widehat{\mathsf{m}}}^{o} \right) \left(Y_{t+k} - \overline{Y}_{\widehat{\mathsf{m}}}^{o} \right) \right\}.$$

If we have more information about the type of dependency, it is advisable to use the estimator specific for the particular model since the above mentioned estimator $\widehat{\sigma}_{0,n}^2(L)$ behaves quite poorly for small and moderate sample sizes. Further discussions the on choice of L together with other theoretical results as well as results from a simulation study can be found in Antoch et al. (1997).

There exists an extensive literature about these procedures, mostly in econometrically oriented papers.

5. Change in simple linear regression

The basic change point detection problem in regression concerns the decision whether a relationship among some variables changed during the observation time. The simplest case that can be treated is that of the simple linear regression when only the relationship between two variables is studied with the explained variable depending linearly on the explanatory one. Even in such a simple setting we can consider many different situations, e.g.:

- either one or both parameters (intercept, slope) can change;
- either the starting parameters before the change point are known or they are unknown;

- either the continuity of the regression function at the change point is assumed or there can be discontinuity etc.

Analogously as above, we can consider for the testing problem both maximum-type statistics as well as sum-type statistics. Nevertheless, we will deal here with the maximum-type statistics only. Provided the number of observations is large, the decision whether to reject the null hypothesis may be based on the asymptotic distribution of the test statistic under H. However, we have to admit that, for many problems that appear in application, the limit behavior of test statistics is not known because the limit behavior depends on the values attained by the explanatory variable. In what follows, we treat several special cases for two basic situations. In the first one we suppose that the design points $\{x_i\}$ are randomly chosen, e.g. they represent a realization of a sequence of independent random variables, or more generally a realization of a stationary ARMA sequence. In the second case the design points are equally spaced, i.e. $x_i = i/n$, $i = 1, \ldots, n$.

5.1. Change in intercept – random design

We consider the problem of testing the null hypothesis H against the alternative A, i.e.

$$\begin{aligned} H &: Y_i = a + bx_i + e_i, && i = 1, \ldots, n, \\ A &: \exists\, \mathsf{m} \in \{2, \ldots, n-2\} \quad \text{such that} && \\ &\quad Y_i = a + bx_i + e_i, && i = 1, \ldots, \mathsf{m}, \\ &\quad Y_i = a^o + bx_i + e_i, && i = \mathsf{m}+1, \ldots, n, \end{aligned} \tag{I.68}$$

where $a \neq a^o$. Here, and in the whole section about the changes in linear regression, the random errors e_i are supposed to be iid with $\mathsf{E}\, e_i = 0, \mathsf{E}\, e_i^2 = \sigma^2$ and $\mathsf{E}|e_i|^{2+\Delta} < \infty$ for some $\Delta > 0$ and independent of x_i. Moreover, the variance σ^2 is supposed to be known. If it is unknown, it can be replaced by its usual estimator

$$\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \big(Y_i - \widehat{a} - \widehat{b}x_i\big)^2,$$

where $\widehat{a}$ and $\widehat{b}$ are the least squares estimators of the parameters a and b calculated under H.

The maximum-type test statistics have the form

$$\max_{2 \le k \le n-2} \Big\{ |\, U_k| \Big\} \quad \text{and} \quad \max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta) n \rfloor} \big\{ |\, U_k| \big\}, \tag{I.69}$$

where

$$U_k = \frac{1}{\sigma} \sqrt{\frac{k}{1 - k/n}}\, \frac{\overline{Y}_k - \overline{Y}_n - \big(\overline{x}_k - \overline{x}_n\big)\widehat{b}}{\sqrt{1 - k\big(\overline{x}_k - \overline{x}_n\big)^2 \big(1 - k/n\big) Q_{xx}^{-1}}}$$

Here $Q_{xx}^{-1} = \sum_{i=1}^{n} \big(x_i - \overline{x}_n\big)^2$.

Notice that the statistics $\{U_k\}$ are standardized partial sums of regression residuals, i.e.

$$U_k = \frac{n^{-1/2}\sum_{i=1}^{k}\big(Y_i - \widehat{Y}_i\big)}{\sqrt{var\Big(n^{-1/2}\sum_{i=1}^{k}\big(Y_i - \widehat{Y}_i\big)\Big)}},$$

where $\widehat{Y}_i = \widehat{a} + \widehat{b}x_i$. To find critical values of the test statistic (I.69), the Bonferroni inequality may be used if n is moderate whereas, for n large, the asymptotic behavior of the proposed test is applied.

Supposing that $\{x_i\}$ represent a realization of a stationary ARMA sequence, it follows that, as k and $n-k$ tend to ∞,

$$\frac{1}{k}\sum_{i=1}^{k} x_i \xrightarrow{\mathcal{P}} \zeta,\ \frac{1}{n-k}\sum_{i=k+1}^{n} x_i \xrightarrow{\mathcal{P}} \zeta,\ \frac{1}{k}\sum_{i=1}^{k} x_i^2 \xrightarrow{\mathcal{P}} \eta,\ \frac{1}{n-k}\sum_{i=k+1}^{n} x_i^2 \xrightarrow{\mathcal{P}} \eta \tag{I.70}$$

for some $\zeta \in \mathcal{R}_1$ and $\eta > 0$. Moreover, under additional assumptions about the rate of convergence of the above averages, for more information see Csörgő and Horváth (1997), we have

$$P\left(\max_{2\le k\le n-2}\Big\{|\,U_k|\Big\} > \frac{x+b_n}{a_n}\right) \approx 1-\exp\big\{-2e^{-x}\big\},\quad x\in\mathcal{R}^1, \tag{I.71}$$

where

$$a_n = \sqrt{2\,\log\,\log\,n} \quad \text{and} \quad b_n = 2\,\log\,\log\,n + \frac{1}{2}\log\,\log\,\log\,n - \frac{1}{2}\log\pi,$$

and

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\Big\{|\,U_k|\Big\} > x\right) \approx 2\big(1-\Phi(x)\big) + 2x\phi(x)\log\frac{1-\beta}{\beta}\cdot \tag{I.72}$$

5.2. Change in intercept – equally spaced design

We consider again the problem (I.68) with $x_i = i/n$, $i = 1,\dots,n$, and the maximum-type statistics (I.69). Under H, the limit behavior of (I.69) is again given by (I.71), while

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\Big\{|\,U_k|\Big\} > x\right) \approx \tag{I.73}$$
$$\approx 2\big(1-\Phi(x)\big) + 2x\phi(x)\int_{\beta}^{1-\beta}\frac{1}{t(1-t)\big(1-3t(1-t)\big)}\,dt.$$

For deatils see Kim and Siegmund (1989).

5.3. Change in both intercept and slope – random design

We test the null hypothesis H against the alternative A in the form

$$\begin{aligned} H &: Y_i = a + bx_i + e_i, && i = 1,\dots,n, \\ A &: \exists\, \mathsf{m} \in \{2,\dots,n-2\} \quad \text{such that} && \\ &\quad Y_i = a + bx_i + e_i, && i = 1,\dots,\mathsf{m}, \\ &\quad Y_i = a^o + b^o x_i + e_i, && i = \mathsf{m}+1,\dots,n, \end{aligned} \tag{I.74}$$

where $a \neq a^o$ and/or $b \neq b^o$.

Let us denote

$$\boldsymbol{X}_k = \begin{pmatrix} 1 & x_1 \\ & \dots \\ 1 & x_k \end{pmatrix}, \qquad \boldsymbol{X}_k^o = \begin{pmatrix} 1 & x_{k+1} \\ & \dots \\ 1 & x_n \end{pmatrix}$$

and

$$\chi^2(k) = \frac{1}{\sigma^2}\big(\widehat{a} - \widehat{a}^o, \widehat{b} - \widehat{b}^o\big)\Big(\big(\boldsymbol{X}_k'\boldsymbol{X}_k\big)^{-1} + \big(\boldsymbol{X}_k^{o\prime}\boldsymbol{X}_k^o\big)^{-1}\Big)^{-1}\big(\widehat{a} - \widehat{a}^o, \widehat{b} - \widehat{b}^o\big)'$$
$$= \frac{1}{\sigma^2}\left(\frac{nk\big(\overline{Y}_k - \overline{Y}_n\big)^2}{n-k} + \frac{Q_{xy}^2(k)}{Q_{xx}(k)} + \frac{Q_{xy}^{o2}(k)}{Q_{xx}^o(k)} - \frac{Q_{xy}^2(n)}{Q_{xx}(n)}\right),$$

where $\widehat{a}$, $\widehat{a}^o, \widehat{b}$ and $\widehat{b}^o$ are the least squares estimators of corresponding quantities under A and

$$Q_{xx}(k) = \sum_{i=1}^{k}\big(x_i - \overline{x}_k\big)\big(x_i - \overline{x}_k\big), \qquad Q_{xy}(k) = \sum_{i=1}^{k}\big(x_i - \overline{x}_k\big)\big(Y_i - \overline{Y}_k\big),$$
$$Q_{xx}^o(k) = \sum_{i=k+1}^{n}\big(x_i - \overline{x}_k^o\big)\big(x_i - \overline{x}_k^o\big), \quad Q_{xy}^o(k) = \sum_{i=k+1}^{n}\big(x_i - \overline{x}_k^o\big)\big(Y_i - \overline{Y}_k^o\big).$$

The maximum-type test statistics are of the form

$$\max_{2\le k\le n-2}\Big\{\chi_k^2\Big\} \quad \text{and} \quad \max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\Big\{\chi_k^2\Big\}. \tag{I.75}$$

It holds that

$$P\left(\max_{2\le k\le n-2}\Big\{\chi_k^2\Big\} > \Big(\frac{x + b_{n,2}}{a_n}\Big)^2\right) \approx 1 - \exp\big\{-2e^{-x}\big\}, \quad x \in \mathcal{R}^1, \tag{I.76}$$

where

$$a_n = \sqrt{2\,\log\,\log\,n} \quad \text{and} \quad b_{n,2} = 2\,\log\,\log\,n + \frac{1}{2}\log\,\log\,\log\,n,$$

and

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\Big\{\chi_k^2\Big\} > x^2\right) \approx e^{-x^2/2} + x^2 e^{-x^2/2}\log\frac{1-\beta}{\beta}\cdot \tag{I.77}$$

The approximation (I.76) can be found in Csörgő and Horváth (1997), see also Albin (1990).

5.4. Change in both intercept and slope – equally spaced design
We consider the problem (I.74) with $x_i = i/n,\ i = 1,\dots,n$. Then for the maximum-type test statistics (I.75) hold

$$P\left(\max_{2\le k\le n-2}\left\{\chi_k^2\right\} > \left(\frac{x+b_{n,2}}{a_n}\right)^2\right) \approx 1-\exp\left\{-2e^{-x}\right\}, \quad x\in\mathcal{R}^1,$$

where

$$a_n = \sqrt{2\log\log n} \quad \text{and} \quad b_{n,2} = 2\log\log n + \frac{1}{2}\log\log\log n,$$

and

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\left\{\chi_k^2\right\} > x^2\right) \approx e^{-x^2/2} + 2x^2 e^{-x^2/2}\log\frac{1-\beta}{\beta}\cdot$$

For more details see Albin et al. (2000).

5.5. Appearance of gradual trend – equally spaced design
Sometimes it is important to decide whether all observations have the same mean and whether at some unknown time point a gradual (continuous) trend appeared. When the appearing trend is supposed to be linear, we are to test the null hypothesis H against the alternative A:

$$\begin{aligned} &H: Y_i = a + e_i, && i = 1,\dots,n, \\ &A: \exists\,\mathsf{m}\in\{1,\dots,n-1\} \quad \text{such that} && \\ &\qquad Y_i = a + e_i, && i = 1,\dots,\mathsf{m}, \\ &\qquad Y_i = a + b\cdot\frac{i-\mathsf{m}}{n} + e_i, && i = \mathsf{m}+1,\dots,n, \end{aligned} \tag{I.78}$$

where $b \neq 0$.

The maximum-type statistics have the form

$$\max_{1\le k<n}\left\{\frac{|\,\widehat{b}_k|}{\sqrt{var\,\widehat{b}_k}}\right\} \quad \text{and} \quad \max_{1\le k\le\lfloor(1-\beta)n\rfloor}\left\{\frac{|\,\widehat{b}_k|}{\sqrt{var\,\widehat{b}_k}}\right\}, \tag{I.79}$$

where $\widehat{b}_k$ is the least squares estimator of b under A supposing that the change occurred at the moment k. Notice that

$$\frac{|\,\widehat{b}_k|}{\sqrt{var\,\widehat{b}_k}} = \frac{\left|\dfrac{1}{\sigma}\dfrac{1}{\sqrt{n}}\sum_{i=k+1}^{n}\left(Y_i-\overline{Y}_n\right)\dfrac{i-k}{n}\right|}{\sqrt{\frac{(n-k)(n-k+1)(n-k+1/2)}{3n^3} - \frac{(n-k)^2(n-k+1)^2}{4n^4}}}.$$

For n large, critical values can be attained using the approximation

$$P\left(\max_{1\le k\le n-1}\left\{\frac{|\,\widehat{b}_k|}{\sqrt{var\,\widehat{b}_k}}\right\} > \frac{x+b_{n,3}}{a_n}\right) \approx 1-\exp\left\{-2e^{-x}\right\}, \quad x\in\mathcal{R}^1, \tag{I.80}$$

where

$$a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_{n,3} = 2 \log \log n + \log \frac{\sqrt{3}}{4\pi},$$

and

$$P\left(\max_{1 \le k \le \lfloor (1-\beta)n \rfloor} \left\{ \frac{|\,\widehat{b}_k|}{\sqrt{var\, \widehat{b}_k}} \right\} > x \right) \approx \tag{I.81}$$

$$\approx 2\big(1 - \Phi(x)\big) + 2 \frac{1}{\sqrt{\pi}}\, \phi(x) \int_0^{1-\beta} \frac{\sqrt{6t}}{(1-t)(1+3t)}\, dt$$

$$= 2\big(1 - \Phi(x)\big) + \frac{1}{\sqrt{\pi}}\, \phi(x) \left(\sqrt{\frac{3}{2}} \log \frac{1 + \sqrt{1-\beta}}{1 - \sqrt{1-\beta}} - \sqrt{2} \arctan \sqrt{3(1-\beta)} \right).$$

The approximation (I.80) can be found in Jarušková (1998a), for (I.81) see Jarušková (1997).

5.6. Continuous change in linear regression

In this section we suppose that the regression slope may continuously change and thus test the hypothesis H against the alternative A:

$$\begin{aligned} H &: Y_i = a + b \cdot \frac{i}{n} + e_i, && i = 1, \ldots, n, \\ A &: \exists\, \mathsf{m} \in \{2, \ldots, n-2\} \quad \text{such that} \\ &\quad Y_i = a + b \cdot \frac{i}{n} + e_i, && i = 1, \ldots, \mathsf{m}, \\ &\quad Y_i = a + b \cdot \frac{i}{n} + c \cdot \frac{i - \mathsf{m}}{n} + e_i, && i = \mathsf{m}+1, \ldots, n, \end{aligned} \tag{I.82}$$

where $c \neq 0$.

The maximum-type test statistics are of the form

$$\max_{2 \le k \le n-2} \left\{ \frac{|\,\widehat{c}_k|}{\sqrt{var\, \widehat{c}_k}} \right\} \quad \text{and} \quad \max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta)n \rfloor} \left\{ \frac{|\,\widehat{c}_k|}{\sqrt{var\, \widehat{c}_k}} \right\}, \tag{I.83}$$

where $\widehat{c}_k$ is the least squares estimator of c under A supposing that the change occurred at the moment k. It holds that

$$P\left(\max_{2 \le k \le n-2} \left\{ \frac{|\,\widehat{c}_k|}{\sqrt{var\, \widehat{c}_k}} \right\} > \frac{x + b_{n,3}}{a_n} \right) \approx 1 - \exp\big\{ -4e^{-x} \big\}, \quad x \in \mathcal{R}^1, \tag{I.84}$$

where

$$a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_{n,3} = 2 \log \log n + \log \frac{\sqrt{3}}{4\pi},$$

and

$$P\left(\max_{\lfloor\beta n\rfloor\le k\le\lfloor(1-\beta)n\rfloor}\left\{\frac{|\,\widehat{c}_k|}{\sqrt{var\,\widehat{c}_k}}\right\}>x\right)\approx$$

$$\approx 2\big(1-\Phi(x)\big)+2\,\frac{1}{\sqrt{\pi}}\,\phi(x)\int_{\beta}^{1-\beta}\sqrt{\frac{3}{8}}\,\frac{1}{t(1-t)}\,dt \tag{I.85}$$

$$=2\big(1-\Phi(x)\big)+\sqrt{\frac{6}{\pi}}\,\phi(x)\,\log\frac{1-\beta}{\beta}\cdot$$

6. Nonparametric and robust procedures in simple regression

The M- and R-procedures can be constructed along the lines of the procedures described in Sections 5.1. – 5.6.. One has only to replace the least squares estimators and corresponding residuals by their M- and R-counterparts.

6.1. M-test procedures

We start with the M-test procedures. Consider the problem of testing the null hypothesis H against the alternative A:

$$\begin{aligned} &H: Y_i = a + bx_i + e_i, && i=1,\dots,n, \\ &A: \exists\, \mathsf{m} \in \{2,\dots,n-2\} \quad \text{such that} && \\ &\qquad Y_i = a + bx_i + e_i, && i=1,\dots,\mathsf{m}, \\ &\qquad Y_i = \mu^o + b^o x_i + e_i, && i=\mathsf{m}+1,\dots,n, \end{aligned} \tag{I.86}$$

where $\big(a,b\big)\neq\big(a^o,b^o\big)$. The random errors e_i are supposed to be iid with symmetric distribution function, the moment assumptions need not to be satisfied and the errors be independent of $\{x_i\}$.

We assume that ψ is a monotone and skew symmetric $\big(\psi(x)=-\psi(x)\ \forall x\in\mathcal{R}^1\big)$ score function. For information on the choice of ψ see Section 4.1.1. Then the M-estimators $\widehat{a}_{n,M}(\psi)$ and $\widehat{b}_{n,M}(\psi)$ based on all n observations can be defined as the solution of the equations

$$\sum_{i=1}^{n}\psi\big(Y_i-a-bx_i\big)=0 \quad \text{and} \quad \sum_{i=1}^{n}x_i\psi\big(Y_i-a-bx_i\big)=0,$$

the corresponding M-residuals as

$$\widehat{e}_{i,M}(\psi)=\psi\Big(Y_i-\widehat{a}_{n,M}(\psi)-x_i\widehat{b}_{n,M}(\psi)\Big),\quad i=1,\dots,n,$$

and the scale estimator can be introduced, e.g., as

$$\widehat{\sigma}^2_{n,M}(\psi) = \min_{2\le k<n-2} \frac{1}{n}\Bigg\{ \sum_{i=1}^{k} \psi^2\Big(Y_i - \widehat{a}_{n,M}(\psi) - x_i\widehat{b}_{n,M}(\psi)\Big) + \sum_{i=k+1}^{n} \psi^2\Big(Y_i - \widehat{a}^o_{n,M}(\psi) - x_i\widehat{b}^o_{n,M}(\psi)\Big)\Bigg\},$$

where $\widehat{a}^o_{n,M}(\psi)$ and $\widehat{b}^o_{n,M}(\psi)$ are the M-estimators of a and b based on $Y_{m+1},\ldots,Y_n$.

The maximum-type test statistic for the case $b = b^o$ (change in the intercept a only) has the form

$$\max_{1\le k<n} \left\{ \frac{1}{\widehat{\sigma}_{n,M}(\psi)} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \widehat{e}_{i,M}(\psi) \right| \right\}. \tag{I.87}$$

Under mild assumptions on ψ, on the error distribution F, on design points (satisfying (I.70)) and if H holds true, we can use the same approximation to the distribution of (I.87) as in previous case, i.e.

$$P\left(\max_{1\le k<n} \left\{ \frac{1}{\widehat{\sigma}_{n,M}(\psi)} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \widehat{e}_{i,M}(\psi) \right| \right\} > \frac{x+b_n}{a_n} \right) \approx \approx 1 - \exp\big\{-2e^{-x}\big\}, \quad x \in \mathcal{R}^1, \tag{I.88}$$

where

$$a_n = \sqrt{2\log\log n} \quad \text{and} \quad b_n = 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log \pi.$$

Next, we consider the testing problem (I.86) when both parameters can change. Let the error terms e_i have a symmetric distribution function. Then the M-test statistic is of the form

$$\max_{2\le k\le n-2} \Big\{ \widetilde{\chi}^2(k) \Big\}, \tag{I.89}$$

where

$$\widetilde{\chi}^2(k) = \frac{1}{\widehat{\sigma}^2_{n,M}(\psi)} \left\{ \frac{n}{k(n-k)} \Big(\sum_{i=1}^{k} \widehat{e}_{i,M}(\psi)\Big)^2 + \frac{\Big[\sum_{i=1}^{n}(x_i-\overline{x}_n)^2\Big]\cdot\Big[\sum_{i=1}^{k}(x_i-\overline{x}_n)\widehat{e}_{i,M}(\psi)\Big]^2}{\Big[\sum_{i=1}^{k}(x_i-\overline{x}_n)^2\Big]\cdot\Big[\sum_{i=k+1}^{n}(x_i-\overline{x}_n)^2\Big]} \right\}.$$

The approximation of the distribution of the statistic (I.89) analogous as in (I.76) applies here too, i.e., we can use

$$P\left(\max_{2\le k\le n-2} \Big\{ \widetilde{\chi}^2(k) \Big\} > \Big(\frac{x+b_{n,2}}{a_n}\Big)^2 \right) \approx 1 - \exp\big\{-e^{-x}\big\}, \quad x \in \mathcal{R}^1, \tag{I.90}$$

where

$$a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_{n,2} = 2 \log \log n + \frac{1}{2} \log \log \log n.$$

6.2. R-test procedures

We consider the testing problem (I.86) with the error terms having continuous distribution function F. Here we need an estimator of the slope parameter b. We can use either some R-estimator if we want to have purely rank procedure or, which is more often the case, we can use any consistent estimator (least squares, L_1-norm etc.). We denote this estimator by $\widetilde{b}_n$. Then we calculate the ranks $R_1(\widetilde{b}_n), \dots, R_n(\widetilde{b}_n)$ of $Y_1 - x_1\widetilde{b}_n, \dots, Y_n - x_n\widetilde{b}_n$. Of course, in practice we estimate both parameters a and b. However, it is clear that the ranks of $\big(Y_i - x_i\widetilde{b}_n\big)$'s do not depend on the estimate of the intercept.

The maximum-type R-statistic for testing problem (I.86) is defined as

$$\max_{1 \le k < n} \Big\{ \widetilde{\widetilde{\chi}}^{2}(k) \Big\}, \tag{I.91}$$

where

$$\widetilde{\widetilde{\chi}}^{2}(k) = \frac{1}{\widehat{\sigma}^2_{n,R}(\psi)} \left\{ \frac{n}{k(n-k)} \Big(\sum\nolimits_{i=1}^{k} \Big[a\big(R_i(\widetilde{b}_n)\big) - \overline{a}_n \Big] \Big)^2 + \right.$$
$$\left. + \frac{\Big[\sum_{i=1}^{n} (x_i - \overline{x}_n)^2 \Big] \cdot \Big[\sum_{i=1}^{k} (x_i - \overline{x}_n) a\big(R_i(\widehat{b}_n)\big) \Big]^2}{\Big[\sum_{i=1}^{k} (x_i - \overline{x}_n)^2 \Big] \cdot \Big[\sum_{i=k+1}^{n} (x_i - \overline{x}_n)^2 \Big]} \right\}$$

with

$$\widehat{\sigma}^2_{n,R} = \frac{1}{n-1} \sum_{i=1}^{n} \Big(a_n(i) - \overline{a}_n \Big)^2.$$

Unfortunately, see Section 4.1.2., this rank test is not distribution free under H as it was the case in the location model. Under the same assumptions on the design matrix as in Section 5.1., the distribution of (I.91) can be under H approximated as in (I.90), where the statistic (I.89) is replaced by statistic (I.91).

The R-procedures for problems in Sections 5.1. – 5.6. can be constructed along the same line. However, they are not very attractive. Nevertheless, the R-test statistic for the testing problem (I.78), where the error terms e_i have a continuous distribution function F, is quite appealing because under H it is distribution free. More precisely, the R-test statistic has the form

$$\max_{1 \le k < n} \left\{ \frac{1}{\widehat{\sigma}_{n,R}} \left| \sum_{i=k+1}^{n} (i-k)\big(a(R_i) - \overline{a}_n\big) \right| \left(\frac{(n-k)^3}{3} - \frac{(n-k)^4}{4n} \right)^{-1/2} \right\}, \tag{I.92}$$

where $R_1, \dots, R_n$ are the ranks of $Y_1, \dots, Y_n$.

Under H, the distribution of (I.92) does not depend on the error distribution F. Therefore, similarly as in the case of the rank-based procedures for changes in location described earlier, the approximation as in (I.90) (where (I.89) is replaced by (I.92)) remains valid for our test statistic, too. Aside that, we can obtain the approximation also through simulations.

7. Change in general linear regression

We briefly survey procedures for detection changes in general linear models. We consider the testing problem:

$$\begin{aligned} &H : Y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i, && i = 1, \dots, n, && \text{(I.93)} \\ &A : \exists\, \mathsf{m} \in \{0, \dots, n-1\} \quad \text{such that} \\ &\quad Y_i = \boldsymbol{x}_i\prime\boldsymbol{\beta} + e_i, && i = 1, \dots, \mathsf{m}, \\ &\qquad = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_i'\boldsymbol{\delta} + e_i, && i = \mathsf{m}+1, \dots, n, \end{aligned}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)' \neq \mathbf{0}$ are parameters, $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are $p-$dimensional vectors; $e_1, \dots, e_n$ are random errors that are iid with zero mean, nonzero variance σ^2 and $E|e_i|^{2+\Delta} < \infty$ for some $\Delta > 0$. Concerning the sets of assumptions on the design points $\boldsymbol{x}_i$, $i = 1 \dots, n$ slightly stronger then for classical tests on the parameters $\boldsymbol{\beta}$ are needed. Here are some possibilities:

- $\boldsymbol{x}_i = \boldsymbol{g}(i/n) = (g_1(i/n), \dots, g_p(i/n))'$, $i = 1, \dots, n$, where $g_1, \dots, g_p$ are continuous functions on $[0,1]$ such that $\int_a^b \boldsymbol{g}(\boldsymbol{x})\, dx$ is a positive definite matrix for each $0 \leq a < b \leq 1$.
- There are fixed p linearly independent p-dimensional vectors $\boldsymbol{x}_1^0, \dots,$ $\boldsymbol{x}_p^0$ and the design points are chosen according to some rule from this p vectors.
- The design points can be a realization of AR or ARMA sequences fulfilling certain moment assumptions.

Notice that as a special case one gets the testing problems considered in Section 5.

Similarly as in the previous sections the test procedures are developed using either likelihood ratio approach or Bayesian one (or their modifications). We present likelihood ratio type test statistics only.

If one assumes that e_i's have $N(0, \sigma^2)$ distribution the likelihood ratio type test statistic is

$$V_{n0} = \max_{p<k<n-p} \left\{ -2 \log \frac{\frac{k}{n}\widehat{\sigma}_{1,k}^2 + \frac{n-k}{n}\widehat{\sigma}_{k+1,n}^2}{\widehat{\sigma}_{k+1,n}^2} \right\},$$

where

$$\widehat{\sigma}_{1,k}^2 = \frac{1}{k}\sum_{i=1}^{k} \left(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{1,k}\right)^2, \quad \widehat{\sigma}_{k+1,n}^2 = \frac{1}{n-k}\sum_{i=k+1}^{n} \left(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{k+1,n}\right)^2,$$

$\widehat{\boldsymbol{\beta}}_{1,k}$ and $\widehat{\boldsymbol{\beta}}_{k+1,n}$ are the least squares estimators of $\boldsymbol{\beta}$ based on $Y_1, \dots, Y_k$ and $Y_{k+1}, \dots, Y_n$, respectively.

A slight modification leads to the test statistics

$$V_n = \max_{p \le k < n} \Big\{ \frac{1}{\widetilde{\sigma}_n^2} \frac{n}{(n-k)k} \boldsymbol{S}_k' (\boldsymbol{X}_n' \boldsymbol{X}_n)^{-1} \boldsymbol{S}_k \Big\},$$

where $\boldsymbol{X}_n = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)'$, $\boldsymbol{S}_k = \sum_{i=1}^k x_i \big(Y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}_{1,n}\big)$, $k = 1, \ldots, n$ and, $\widetilde{\sigma}_n^2$ is an estimator of σ^2; for instance

$$\widetilde{\sigma}_n^2 = \frac{1}{n-p} \min_{p<k<n-p} \Big\{ k \widehat{\sigma}_{1,k}^2 + (n-k) \widehat{\sigma}_{k+1,n}^2 \Big\}$$

can be used.

Under the null hypothesis H and mild assumptions we have the following approximations

$$P\Big(V_{n0} > \Big(\frac{x + b_{n,p}}{a_n}\Big)^2\Big) \approx 1 - \exp\{-2e^{-x}\}, \quad x \in \mathcal{R}^1$$

and

$$P\Big(V_n > \Big(\frac{x + b_{n,p}}{a_n}\Big)^2\Big) \approx 1 - \exp\{-2e^{-x}\}, \quad x \in \mathcal{R}^1,$$

where

$$b_{n,p} = 2 \log\log n + \frac{p}{2} \log\log\log n - log\, \Gamma\,(p/2) \quad \text{and} \quad a_n = \sqrt{2 \log\log n}.$$

Similarly as in the previous sections these approximations can be used to get an approximation to the critical values.

In econometrically oriented literature one often meets the test procedures based on partial sums of residuals, namely the following test statistic is used

$$\max_{1 \le k < n} \Big\{ \frac{1}{\sqrt{n} \widetilde{\sigma}_n} \Big| \sum_{i=1}^k (Y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}_{1,n}) \Big| \Big\}$$

This test statistic is much simpler than V_n and V_{n0} but it is sensitive only to some $\delta \neq 0$.

Under the null hypothesis H and mild assumptions (e_i can be even dependent)

$$P\Big(\max_{1 \le k < n} \frac{1}{\sqrt{n} \widetilde{\sigma}_n} \Big| \sum_{i=1}^k \big(Y_i - \boldsymbol{x}_i' \widehat{\boldsymbol{\beta}}_{1,n}\big) \Big| \le x \Big) \approx P\Big(\sup_{0<t<1} |B(t)| \le x \Big) =$$

$$= 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} e^{-2j^2 x^2}, \quad x \in \mathcal{R}^1,$$

where $\{B(t);\, t \in (0,1)\}$ is a Brownian bridge and $\widetilde{\sigma}_n$ is a proper standardization. More information can be found e.g. in Jandhyala and MacNeill (1991), Jandhyala and MacNeill (1992), MacNeill (1978), Horváth (1995), Kim and Siegmund (1989) or Ploberger (1992).

Part II. Estimators of changes

Part I deals with tests on the stability of statistical models. The problem was formulated in terms of testing the null hypothesis H against the alternative hypothesis A. The null hypothesis H claims that the model remains the same during the whole observational period, usually it means that the parameters of the model do not change. The alternative hypothesis A claims that, at an unknown time point, the model changes, which means that some of the parameters of the model are subject to a change. In case we reject the null hypothesis H, i.e. we decide that there is a change in the model, a number of questions arise:

- when has the model changed;
- is there just one change or are there more changes;
- what is the total number of changes.

The time moment when the model has changed is usually called *change point*. In this part our primary interest is to *estimate the change point(s)* in various models. Of course, we also estimate other parameters of the models. We will show approximations to the distributions of the change point estimators and finally construct the interval estimators for the change points. As we will see below, the estimators of the change points are usually closely related to some of the test statistics treated in Part I.

8. Change in location

We assume that the observations $Y_1, \dots, Y_n$ follow the model

$$Y_i = \begin{cases} \mu + \sigma e_i, & i = 1, 2, \dots, \mathsf{m}, \\ \mu + \delta + \sigma e_i, & i = \mathsf{m}+1, \dots, n, \end{cases} \tag{II.1}$$

where μ, $\sigma^2 > 0$, $\delta \neq 0$ and $\mathsf{m}(< n)$ are parameters, and $e_1, \dots, e_n$ are iid random variables with zero mean, unit variance and with $\mathsf{E}\,|e_i|^{2+\Delta} < \infty$ with some $\Delta > 0$. The distribution of e_i will be denoted by F.

It is supposed that the observations $Y_1, \dots, Y_n$ are obtained at the ordered time moments $t_1 < \cdots < t_n$. Hence, this model says that there is just one change after the m^{th} observation, and the corresponding time moment is called the *change point*. More exactly, the change occurs in the time period $(t_{\mathsf{m}}, t_{\mathsf{m}+1}]$. However, we will work here only with integer valued t_{m}.

We assume the change point m satisfies

$$\mathsf{m} = \lfloor n\gamma \rfloor, \quad \gamma \in (0,1), \tag{II.2}$$

where $\lfloor a \rfloor$ denotes the integer part of a. This means that the change can occur neither at the very beginning nor at the very end of the observational period. To estimate the unknown parameters we can apply some of the general methods, e.g. the maximum likelihood method, however the distribution function F of the error terms e_i has to be known.

• Another possibility is to estimate the unknown parameters m, μ and δ by the least squares (LS) method, that gives a simple solution. The least squares estimators $\widehat{\mathsf{m}}_{LS}$, $\widehat{\mu}_{LS}$ and $\widehat{\delta}_{LS}$ of the parameters m, μ and δ are defined as solutions of the minimization problem

$$\min\Big\{ \sum_{i=1}^{k} \big(Y_i - \mu\big)^2 + \sum_{i=k+1}^{n} \big(Y_i - \mu - \delta\big)^2; \qquad k \in \big\{1,\dots,n-1\big\},\ \mu \in \mathcal{R}^1,\ \delta \in \mathcal{R}^1 \Big\}, \tag{II.3}$$

In other words, the parameters m, μ and δ are estimated in such a way that the sum of squares of residuals is minimal.

Direct calculation gives the explicit expressions for the estimators $\widehat{\mu}_{LS}$ and $\widehat{\delta}_{LS}$:

$$\widehat{\mu}_{LS} = \overline{Y}_{\widehat{\mathsf{m}}_{LS}}, \tag{II.4}$$

$$\widehat{\delta}_{LS} = \overline{Y}^{o}_{\widehat{\mathsf{m}}_{LS}} - \overline{Y}_{\widehat{\mathsf{m}}_{LS}}, \tag{II.5}$$

while $\widehat{\mathsf{m}}_{LS}$ is a solution of the maximization problem

$$\max\left\{ \sqrt{\frac{n}{k(n-k)}}\,\big|S_k\big|;\ k \in \big\{1,\dots,n-1\big\} \right\},$$

where $\overline{Y}_k$ and $\overline{Y}^{o}_k$ are defined in (I.2). We recall the definition of the partial sums S_k:

$$S_k = \sum_{i=1}^{k} \big(Y_i - \overline{Y}_n\big), \quad k = 1,\dots,n. \tag{II.6}$$

Notice that if the error terms e_i have $N(0,1)$ distribution, the least squares estimators coincide with the maximum likelihood estimators.

The solution of the maximization problem need not to be unique, therefore we make the following convention.

Convention. We will typically treat a random variable V defined as a solution of the maximization problem

$$\max\Big\{W(t);\ t \in T\Big\},$$

where $\big\{W(t);\, t \in T\big\}$ is a random process or a sequence of random variables. If the solution is not uniquely determined, we always take the smallest solution and write shortly

$$V = \arg\max\ \Big\{W(t);\ t \in T\Big\}.$$

Using this convention, the estimator $\widehat{\mathsf{m}}_{LS}$ is defined as

$$\widehat{\mathsf{m}}_{LS} = \arg\max\ \left\{ \sqrt{\frac{n}{k(n-k)}}\,\big|S_k\big|;\ k \in \big\{1,\dots,n-1\big\} \right\}. \tag{II.7}$$

Moreover, the estimator $\widehat{\mathsf{m}}_{LS}$ can be equivalently defined as

$$\widehat{\mathsf{m}}_{LS} = \arg\max \left\{ \frac{k(n-k)}{n} \big(\overline{Y}_k - \overline{Y}_k^{\,o} \big)^2;\ k \in \{1, \dots, n-1\} \right\}.$$

• Another type of estimators of m is based on moving sums (MOSUM). These estimators will be denoted by $\widehat{\mathsf{m}}_{LS}(G)$ and defined as

$$\widehat{\mathsf{m}}_{LS}(G) = \arg\max \left\{ \big| S_{k+G} - 2S_k + S_{k-G} \big|;\ k \in \{G+1, \dots, n-G\} \right\}. \quad \text{(II.8)}$$

The bandwidth G is usually chosen much smaller than n, typically $g = 0.1n$ or $0.05n$. Some tuning has to be done for the real data.

The connection of the estimators $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ with the test statistics (I.17) and (I.54) is clearly visible. Calculating the tests statistics we usually calculate the corresponding estimators as a “free” by-product.

To get an idea about the behavior of the estimators, we present the expectations and variances of the partial sums of residuals and the moving sums:

$$\mathsf{E}\, S_k = \begin{cases} -\delta k \dfrac{n-\mathsf{m}}{n}, & 1 \le k \le \mathsf{m}, \\ -\delta (n-k) \dfrac{\mathsf{m}}{n}, & \mathsf{m} < k \le n, \end{cases}$$

$$var\ S_k = \sigma^2 \frac{k(n-k)}{n}, \qquad 1 \le k \le n,$$

$$\mathsf{E}\left(S_{k+G} - 2S_k + S_{k-G}\right) = \begin{cases} 0, & G < k \le \mathsf{m} - G, \\ \delta\big(k + G - \mathsf{m}\big), & \mathsf{m} - G < k \le \mathsf{m}, \\ \delta\big(\mathsf{m} + G - k\big), & \mathsf{m} < k \le \mathsf{m} + G, \\ 0, & \mathsf{m} + G < k \le n - G, \end{cases}$$

$$var\left(S_{k+G} - 2S_k + S_{k-G}\right) = 2\sigma^2 G.$$

In both cases the maximum of the absolute values of the expectations is reached for $k = \mathsf{m}$. Moreover, if the random parts of $(S_k - \mathsf{E} S_k)\sqrt{\frac{n}{k(n-k)}}$ and $\big(S_{k+G} - 2S_k + S_{k-G}\big) - \mathsf{E}\big(S_{k+G} - 2S_k + S_{k-G}\big)$ are stochastically smaller than the expectations, then the estimators $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ are “close” to the change point. More exactly, under certain assumptions, as $n \to \infty$,

$$\frac{\widehat{\mathsf{m}}_{LS} - \mathsf{m}}{n} \xrightarrow{\mathcal{P}} 0 \quad \text{and} \quad \frac{\widehat{\mathsf{m}}_{LS}(G) - \mathsf{m}}{n} \xrightarrow{\mathcal{P}} 0,$$

i.e., the estimators are consistent.For details see Antoch and Hušková (1998).

Finding the exact distribution of either type of estimators is a very complex problem (similarly as in the testing problem); up to now no exact solution has been found. The limit distribution of these estimators has been treated and this provides a reasonable approximation to the exact distribution.

If $\delta \neq 0$ can be assumed small with respect to the number of observations n, mathematically expressed as $\delta \equiv \delta_n \to 0$ as $n \to \infty$, we speak about a *local change*. In such a case we have the following approximation, i.e.:

$$P\left(\frac{\delta^2}{\sigma^2}\left(\widehat{\mathsf{m}}_{LS} - \mathsf{m}\right) \leq x\right) \approx P(V \leq x), \quad x \in \mathcal{R}^1, \tag{II.9}$$

where the random variable V is defined by

$$V = \arg\max \left\{W(s) - |s|/2\,;\ s \in \mathcal{R}^1\right\}. \tag{II.10}$$

Here $\{W(s); s \in \mathcal{R}^1\}$ is a two-sided standard Wiener process, i.e.,

$$W(s) = \begin{cases} W_1(-s), & s < 0, \\ W_2(s), & s \geq 0, \end{cases}$$

where $\{W_1(t);\ t \in [0, \infty)\}$ and $\{W_2(t);\ t \in [0, \infty)\}$ are independent standard Wiener processes. This approximation is "distribution free", which means that it does not depend on the distribution of the error terms e_i.

Concerning the MOSUM type estimators, it holds

$$P\left(\frac{2\delta^2}{3\sigma^2}\left(\widehat{\mathsf{m}}_{LS}(G) - \mathsf{m}\right) \leq x\right) \approx P(V \leq x), \quad x \in \mathcal{R}^1, \tag{II.11}$$

with V given by (II.10).

The parameter σ^2 can be estimated, e.g., by

$$\widehat{\sigma}_n^2 = \frac{1}{n-2}\left\{\sum_{i=1}^{\widehat{\mathsf{m}}}\left(Y_i - \overline{Y}_{\widehat{m}}\right)^2 + \sum_{i=\widehat{\mathsf{m}}+1}^{n}\left(Y_i - \overline{Y}^{o}_{\widehat{\mathsf{m}}}\right)^2\right\}, \tag{II.12}$$

where $\widehat{\mathsf{m}}$ is any of the above estimators of m. The approximations described in (II.9) and (II.11) remain valid if δ and σ are replaced by their estimators (II.5) and (II.12), respectively.

The distribution of the random variable V is known. It was independently derived by several authors. Stryhn (1996) showed that random variable V has the distribution function

$$P(V \leq x) = \begin{cases} 1 + \sqrt{\frac{x}{2\pi}}e^{-x/8} - \frac{1}{2}(x+5)\Phi(-\frac{1}{2}\sqrt{x}) + \frac{3}{2}e^{x}\Phi(-\frac{3}{2}\sqrt{x}), & x \geq 0, \\ 1 - P(V \leq -x), & x < 0. \end{cases} \tag{II.13}$$

This result can be used for construction of a confidence interval for the change point m. The tables needed for this task can be found in the book of Csörgő and Horváth (1997); here we give only a few useful quantiles.

β	0.9	0.95	0.975	0.99	0.995
v_β	4.696	7.687	11.033	15.868	19.767

Table 18. Selected quantiles of V, $P\left(V \leq v_\beta\right) = \beta$.

8.1. Dependent observations

Now, we shortly discuss the problem of estimation of m when in the model (II.1) the error terms are dependent. More precisely, we consider the model (II.1) with m fulfilling (II.2), where the error terms e_i form a linear process satisfying

$$e_i = \sum_{t=0}^{\infty} w_t \epsilon_{i-t}, \quad i = 1, 2, \dots,$$

where ϵ_i are iid random variables, $\mathsf{E}\,\epsilon_i = 0$, $var\,\epsilon_i = \sigma^2 > 0$, $\mathsf{E}\,|\epsilon_i|^{2+\triangle} < \infty$ for some $\triangle > 0$ and the weights $\{w_j\}_{j=0}^{\infty}$ satisfy

$$\sum_{j=0}^{\infty} j\,|w_j| < \infty, \quad \sum_{j=0}^{\infty} w_j \neq 0.$$

Notice that AR and ARMA processes fulfill these assumptions.

The maximum likelihood method can be applied here, however, we need to know the distribution of the error terms e_i. It appears that we can apply the estimators $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ defined by (II.7) and (II.8), respectively. Their distributions can be approximated as in the case of independent errors, i.e. (II.9) and (II.11) apply, however, *different standardization is needed.* Namely, σ^2 has to be replaced by

$$\sigma_0^2 = \sigma^2 \Big(\sum_{j=0}^{\infty} w_j \Big)^2.$$

For example, if the sequence e_n is an AR(1) sequence with the coefficient $\rho \in (-1, 1)$, then σ^2 in (II.9) has to be replaced by $\sigma^2/(1-\rho)$ or by its estimator.

Generally, σ_0^2 can be estimated, e.g., by

$$\widehat{\sigma}_{0,n}^2(L) = \widehat{R}(0) + 2 \sum_{k=1}^{L} \left(1 - \frac{k}{L} \right) \widehat{R}(k), \quad L < n,$$

where for $k \geq 0$

$$\widehat{R}(k) = \frac{1}{n} \left\{ \sum_{t=1}^{\widehat{\mathsf{m}}-k} \left(Y_t - \overline{Y}_{\widehat{\mathsf{m}}} \right) \left(Y_{t+k} - \overline{Y}_{\widehat{\mathsf{m}}} \right) + \sum_{t=\widehat{\mathsf{m}}+1}^{n-k} \left(Y_t - \overline{Y}_{\widehat{\mathsf{m}}}^{o} \right) \left(Y_{t+k} - \overline{Y}_{\widehat{\mathsf{m}}}^{o} \right) \right\}.$$

If we have more information about the type of dependency, it is advisable to use the estimator specific for the particular model since the above mentioned estimator $\widehat{\sigma}_{0,n}^2(L)$ behaves quite poorly for small and moderate sample sizes. Further discussions the on choice of L together with other theoretical results as well as results from a simulation study can be found in Antoch et al. (1997).

♣ Now, we turn to the *estimators based on M-statistics* and *R-statistics.* Analogously as in the testing problem H against A, we can construct the estimators of the change point m replacing in the least squares type estimators

(II.7) and (II.8) the partial sums S_k by their M-counterparts $S_{k,M}(\psi)$ or their R-counterparts $S_{k,R}$, cf Part I.

8.2. M-estimators of the change point

We assume that the observations $Y_1, \ldots, Y_n$ follow the model (II.1) with m fulfilling (II.2) and with the error term e_i having a common distribution F that is symmetric about zero (no need of zero mean and a finite variance). The M- estimator $\widehat{\mu}_n(\psi)$ of μ generated by the score function ψ is defined as the solution of the equation

$$\sum_{i=1}^{n} \psi(Y_i - t) = 0.$$

The score function ψ is assumed to be monotone and skew symmetric, i.e. $\psi(x) = -\psi(-x),\ \forall x \in \mathcal{R}^1$.

The M-estimator $\widehat{\mathsf{m}}_M(\psi)$ of m is defined as

$$\widehat{\mathsf{m}}_M(\psi) = \arg\max \left\{ \sqrt{\frac{n}{k(n-k)}}\ \big|S_{k,M}(\psi)\big|\,;\ k \in \big\{1, \ldots, n-1\big\} \right\}, \quad \text{(II.14)}$$

where

$$S_{k,M}(\psi) = \sum_{i=1}^{k} \psi\Big(Y_i - \widehat{\mu}_n(\psi)\Big), \quad k = 1, \ldots, n. \quad \text{(II.15)}$$

Under some mild assumptions on ψ and F we have the following approximation to the distribution of these errors:

$$P\left(\frac{\delta^2 \lambda^2(\psi, F)}{\int \psi^2(x)\, dF(x)} \Big(\widehat{\mathsf{m}}_M(\psi) - \mathsf{m}\Big) \le x \right) \approx P(V \le x), \quad x \in \mathcal{R}^1, \quad \text{(II.16)}$$

where $\lambda(\psi, F)$ is the value of the derivative, in t, of the function $\int -\psi(x - t)dF(x)$ at $t = 0$. This means that the approximation to the distribution of $\widehat{\mathsf{m}}_M(\psi)$ is the same as that to the distribution of $\widehat{\mathsf{m}}_{LS}$ or $\widehat{\mathsf{m}}_{LS}(G)$ up to the multiplicative constant that depends on the choice of ψ and on the distribution of the error terms. The unknown constants can be replaced by suitable estimators, e.g., by:

$$\widehat{\delta}_{n,M}(\psi) = \widehat{\mu}^{o}_{\widehat{\mathsf{m}}_M}(\psi) - \widehat{\mu}_{\widehat{\mathsf{m}}_M}(\psi),$$

$$\widehat{\sigma}^2_{n,M}(\psi) = \frac{1}{n-2} \left\{ \sum_{i=1}^{\widehat{\mathsf{m}}_M} \psi^2\big(Y_i - \widehat{\mu}_{\widehat{\mathsf{m}}_M}(\psi)\big) + \sum_{i=\widehat{\mathsf{m}}_M+1}^{n} \psi^2\big(Y_i - \widehat{\mu}^{o}_{\widehat{\mathsf{m}}_M}(\psi)\big) \right\},$$

$$\begin{aligned}\widehat{\lambda}_{n,M}(\psi) = & \frac{1}{2nc_n} \Bigg\{ \sum_{i=1}^{\widehat{\mathsf{m}}_M} \Big(\psi\big(Y_i - \widehat{\mu}_{\widehat{\mathsf{m}}_M}(\psi) + c_n\big) - \psi\big(Y_i - \widehat{\mu}_{\widehat{\mathsf{m}}_M}(\psi) - c_n\big)\Big) \\ & + \sum_{i=\widehat{\mathsf{m}}_M+1}^{n} \Big(\psi\big(Y_i - \widehat{\mu}^{o}_{\widehat{\mathsf{m}}_M}(\psi) + c_n\big) - \psi\big(Y_i - \widehat{\mu}^{o}_{\widehat{\mathsf{m}}_M}(\psi) - c_n\big)\Big) \Bigg\},\end{aligned}$$

where $\widehat{\mu}_{\widehat{\mathsf{m}}_M}(\psi)$ and $\widehat{\mu}^o_{\widehat{\mathsf{m}}_M}(\psi)$ are M-estimators based on $Y_1, \ldots, Y_{\widehat{\mathsf{m}}_M}$ and $Y_{\widehat{\mathsf{m}}_M+1}, \ldots, Y_n$ and $\{c_n\}_n$ is a sequence of positive numbers tending to 0 not faster than $n^{-1/2}$.

Motivated by the definition of the least squares estimators in (II.3), we can also define other M-estimators. Let ρ be a non-negative convex function on $\mathcal{R}^1$ and let ψ be its derivative. Then the M-estimators $\widetilde{\mathsf{m}}_M(\rho), \widetilde{\delta}_M(\rho)$ and $\widetilde{\mu}_M(\rho)$ are defined as a solution of the minimization problem

$$\min\left\{\sum_{i=1}^{k} \rho\big(Y_i - \mu\big) + \sum_{i=k+1}^{n} \rho\big(Y_i - \mu - \delta\big); \mu \in \mathcal{R}^1, \delta \in \mathcal{R}^1, k \in \{1, \ldots, n-1\}\right\}$$

The distribution of the estimator $\widetilde{\mathsf{m}}_M(\rho)$ can be approximated in the same way as that of $\widehat{\mathsf{m}}_M(\psi)$, i.e. (II.16) remains true if $\widehat{\mathsf{m}}_M(\psi)$ is replaced by $\widetilde{\mathsf{m}}_M(\rho)$.

Typical choices of the score functions are discussed in subsection 4.1.1. of the present paper or in some books or papers devoted to the $M-$ estimators.

Next, we shortly discuss the particular choice $\psi(x) = \operatorname{sign} x$, $x \in \mathcal{R}^1$. This is usually called L_1-procedure and the corresponding estimators are called L_1-estimators. This is due to the fact that the sample median based on $Y_1, \ldots, Y_n$ is a solution of the minimization problem

$$\min\left\{\sum_{i=1}^{n} |Y_i - t|\,;\ t \in \mathcal{R}^1\right\}.$$

We should recall that the sample median minimizes the sum of the L_1-distances, while the least squares estimator minimizes the sum of the L_2-distances. We denote this L_1-estimator of the change point m by $\widehat{\mathsf{m}}_{L_1}$ and we receive the approximation to its distribution in the following form:

$$P\left(\frac{\widehat{\mathsf{m}}_{L_1} - \mathsf{m}}{4f^2(0)} \le x\right) \approx P\big(V \le x\big), \quad x \in \mathcal{R}^1,$$

where $f(0)$ is the density of the error term at the point 0. If the density is unknown, we have to replace it by an estimator.

There is also another possibility to define the L_1 estimators that is more transparent, however, more difficult to calculate. Namely, we define $\widetilde{\mathsf{m}}_{L_1}, \widetilde{\delta}_{L_1}$ and $\widetilde{\mu}_{L_1}$ as a solution of the minimization problem

$$\min\left\{\sum_{i=1}^{k} |Y_i - \mu| + \sum_{i=k+1}^{n} |Y_i - \mu + \delta|\,;\ \mu \in \mathcal{R}^1, \delta \in \mathcal{R}^1, k \in \{1, \ldots, n-1\}\right\}.$$

The M-estimators of a change in other models can be introduced along the lines explained above.

8.3. R-estimators of the change point

We assume the model (II.1) with m fulfilling (II.2) and with the error terms having absolutely continuous density f with the Fisher information

$$0 < I(f) = \int_{-\infty}^{\infty} \frac{(f'(x))^2}{f(x)} dx < \infty,$$

where f' is the derivative of f.

The rank-based estimator of m is based on the partial sums

$$S_{k,R} = \sum_{i=1}^{k} \Big(a(R_i) - \overline{a}_n \Big), \quad k = 1, \ldots, n, \tag{II.17}$$

where $R_1, \ldots, R_n$ are the ranks corresponding to $Y_1, \ldots, Y_n$ and

$$\overline{a}_n = n^{-1} \sum_{i=1}^{n} a(i).$$

The scores $a(1), \ldots, a(n)$ are usually related to a function ξ defined on the interval $(0,1)$ such that

$$0 < \int_0^1 \big| \xi(u) \big|^{2+\Delta} du < \infty$$

with some $\Delta > 0$; typically,

$$a(i) = \xi\big(i/(n+1)\big), \quad i = 1, \ldots, n. \tag{II.18}$$

In this section we will work with these scores; for other possibilities consult books on rank-based procedures.

The R-estimator of m is defined as

$$\widehat{\mathsf{m}}_R = \arg\max \left\{ \sqrt{\frac{n}{k(n-k)}} \, |S_{k,R}| \, ; \; k \in \big\{1, \ldots, n-1\big\} \right\} \cdot \tag{II.19}$$

We have the following approximation to the distribution of $\widehat{\mathsf{m}}_R$:

$$P\left(\frac{\delta^2 b_R^2(\xi)}{\sigma_R^2} \Big(\widehat{\mathsf{m}}_R - \mathsf{m} \Big) \le x \right) \approx P(V \le x), \quad x \in \mathcal{R}^1,$$

where

$$\sigma_R^2 = \frac{1}{n-1} \sum_{i=1}^{n} \big(a(R_i) - \overline{a}_n \big)^2 \approx \int_0^1 \xi^2(u) du - \Big(\int_0^1 \xi(u) \, du \Big)^2$$

and

$$b_R^2(\xi) = \Big(\int_{-\infty}^{\infty} \xi\big(F(x)\big) f'(x) \, dx \Big)^2.$$

If F is unknown, we replace $b_R^2(\xi)$ by an estimator, however, as this is going beyond the scope of this paper, the interested reader should consult some advanced material on the rank-based procedures.

Comparing the rank-based tests discussed in Part I with rank-based estimators just introduced, we see that the estimators are not anymore distribution-free even asymptotically and, moreover, we have stronger assumptions on the scores than in case of tests.

The MOSUM type M- and R-estimators can be introduced analogously and the same holds for the approximations to their distributions.

8.4. Multiple changes

Consider the location model with multiple changes

$$Y_i = \mu_j + e_i, \quad \lfloor n\gamma_{j-1} \rfloor < i \le \lfloor n\gamma_j \rfloor, \; j = 1, \ldots, q+1, \tag{II.20}$$

where $\mu_j \in \mathcal{R}^1$, $\mu_j \neq \mu_{j+1}$, $j = 1, \ldots, q$, $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_{q+1} = 1$ and $e_1, \ldots, e_n$ are iid random variables with zero mean, nonzero variance $var\ e_i = \sigma^2$ and $\mathsf{E}\,|e_i|^{2+\Delta} < \infty$ for some $\Delta > 0$. The change points are $\lfloor n\gamma_j \rfloor$, $j = 1, \ldots, q$. Their number q can be known or unknown, however an upper bound q_o for q is supposed to be known.

We should point out, that if we test the null hypothesis

$$H : Y_1, \ldots, Y_n \quad \text{are iid with mean } \mu$$

against the alternative A corresponding to the multiple changes described in (II.20), where only the upper bound q_0 for the number of possible changes is known, we can apply any of the test procedures developed for the alternative with one change only. These tests are consistent. We could construct, e.g., the likelihood ratio test, however this brings big problems with finding an approximation to the critical values. A Bayesian approach is studied, e.g., in Chib (1999). We describe three types of estimators for the number of changes. They also provide the estimators for the change points.

The first one was proposed by Yao (1988). This is a modification of the Schwarz' criterion for determining the dimension of regression. The estimator $\widehat{q}$ of the number of change q is defined as a solution of the minimization problem

$$\min \left\{ \frac{n}{2} \log \widetilde{\sigma}_q^2 + q \log n \;\middle|\; q = 1, \ldots, q_0 \right\}, \tag{II.21}$$

where q_0 is the possible largest number of changes and

$$\widetilde{\sigma}_q^2 = \min \left\{ \sum_{j=1}^{q+1} \sum_{i=n_{j-1}+1}^{n_j} \big(Y_i - \mu_j\big)^2 \;\middle|\; \mu_j \in \mathcal{R}^1,\, j = 1, \ldots, q,\, 0 = n_0 < n_1 < \cdots < n_q < n_{q+1} = n \right\}.$$

The estimators of the parameters $n_j = \lfloor n\gamma_j \rfloor$ and μ_j are then obtained as a solution of the minimization problem

$$\min \left\{ \sum_{j=1}^{\widehat{q}+1} \sum_{i=n_{j-1}+1}^{n_j} \big(Y_i - \mu_j\big)^2 \;\middle|\; \mu_j \in \mathcal{R}^1,\, j = 1, \ldots, \widehat{q}+1,\, 0 = n_0 < n_1 < \cdots < n_{\widehat{q}} < n_{\widehat{q}+1} = n \right\}.$$

Computational difficulties are evident and therefore some modifications were proposed. Notice, however, that if the observations follow a normal distribution, then these estimators are consistent and coincide with the maximum likelihood estimators.

The second and the third method are related to the methods for one change only developed, and treated, in previous sections. Thus the estimators $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ introduced in (II.7) and (II.8) can be used to estimate multiple changes, too.

Direct calculation gives

$$\mathsf{E}S_k = \sum_{j=1}^{v} \Big(\lfloor n\gamma_j\rfloor - \lfloor n\gamma_{j-1}\rfloor\Big)\big(\mu_j - \overline{\mu}\big) + \Big(k - \lfloor n\gamma_v\rfloor\Big)\big(\mu_v - \overline{\mu}\big),$$

$$\lfloor n\gamma_{v-1}\rfloor < k \le \lfloor n\gamma_v\rfloor,\ v = 1,\dots,q+1,$$

where

$$\overline{\mu} = \frac{1}{n}\sum_{j=1}^{q+1}\Big(\lfloor n\gamma_j\rfloor - \lfloor n\gamma_{j-1}\rfloor\Big)\mu_j$$

and

$$\mathsf{E}\Big(S_{k+G} - 2S_k + S_{k-G}\Big) = \tag{II.22}$$

$$= \begin{cases} 0, & G < |k - \lfloor n\gamma_j\rfloor|, \\ \big(\mu_j - \mu_{j-1}\big)\big(\lfloor n\gamma_j\rfloor - k + G\big), & \lfloor n\gamma_j\rfloor - G < k \le \lfloor n\gamma_j\rfloor, \\ \big(\mu_j - \mu_{j-1}\big)\big(k - \lfloor n\gamma_j\rfloor + G\big), & \lfloor n\gamma_j\rfloor < k \le \lfloor n\gamma_j\rfloor + G, \end{cases}$$

for $j = 1,\dots,q$. The extremes in both expectations can be reached only for $k = \lfloor n\gamma_j\rfloor$, $j = 1,\dots,q$. This leads to the following estimation procedures.

At first, find

$$\widetilde{\mathsf{m}}^{(1)} = \arg\max \left\{ \sqrt{\frac{n}{k(n-k)}}\ \big|S_k\big|;\ k \in \big\{1,\dots,n-1\big\} \right\}, \tag{II.23}$$

where S_k is given by (II.6). At second divide observations into two groups $Y_1,\dots,Y_{\widetilde{\mathsf{m}}^{(1)}}$ and $Y_{\widetilde{\mathsf{m}}^{(1)}+1},\dots,Y_n$ and find the estimator in each group using (II.7). The whole procedure is repeated for each subset until "a (constant) mean is obtained", i.e. until there is a statistically significant difference in the means of groups. We apply the test based on

$$\max_{k\in D}\left\{ \sqrt{\frac{n}{k(n-k)}}\ \frac{1}{\widehat{\sigma}_n}\ \left|\sum_{i\in D}\Big(Y_i - \frac{1}{\#D}\sum_{j\in D} Y_j\Big)\right| \right\},$$

where D denotes the indexes corresponding to the particular group, $\#D$ its cardinality and $\widehat{\sigma}_n^2$ is a suitable estimator of σ^2. Critical values can be determined from the approximation (I.20) and the level α_n has to be chosen such that, as $n \to \infty$, $\alpha_n \to 0$. Then, this procedure consistently estimates all change points and also the number of changes. This type of procedure was developed by Vostrikova (1981).

Motivated by (II.22) and by the behavior of $\widehat{\mathsf{m}}_{LS}(G)$, we propose the following estimation procedure. We find all pairs of indices v_j, w_j, $j = 1,\dots,\widehat{q}$, such that $w_j - v_j \ge G/2$, $j = 1,\dots,\widehat{q}$ and such that for $k = v_j,\dots,w_j$

$$\begin{aligned}\left|S_{k+G} - 2S_k + S_{k-G}\right| &\le D(n;G,\alpha_n),\\ \left|S_{v_j-1+G} - 2S_{v_j-1} + S_{v_j-1-G}\right| &\ge D(n;G,\alpha_n),\\ \left|S_{w_j+1+G} - 2S_{w_j+1} + S_{w_j+1-G}\right| &\ge D(n;G,\alpha_n),\end{aligned}$$

where

$$D(n;G,\alpha_n) = \sigma_n\sqrt{2G}\,\frac{2\log\dfrac{n}{G} + \dfrac{1}{2}\log\log\dfrac{n}{G} - \dfrac{1}{2}\log\dfrac{4\pi}{9} - \log\log\dfrac{1}{1-\alpha_n}}{\sqrt{2\log(n/G)}}$$

and $\alpha_n \to 0$ as $n \to \infty$. This means that we find all segments where a certain critical value is exceeded. This critical value corresponds to test whether a change occurred or not based on

$$\max_{G<k<n_G}\left\{\left|S_{k+G} - 2S_k + S_{k-G}\right|\right\}$$

with asymptotic level α_n. The estimator of the number of change points is $\widehat{q}$.

Then $\widehat{q}$, estimate of the number of change points, and the index $k \in [v_j, w_j]$ for which the maximum over the set $\{v_j, \ldots, w_j\}$ is reached, can serve as the estimator of one of the change points.

These two procedures, particularly the second one, are quite simple and can certainly serve as a simple diagnostic tool. There is a problem to find a reasonable estimator of the variance σ^2. For that, we need an upper bound, say q_o, of the number q of change points. Then we can estimate σ^2 as follows:

$$\begin{aligned}\sigma_n^2 =& \frac{1}{n-q_o-1}\min_{1\le k_1<k_2<\cdots<k_{q_o}\le n}\left\{\sum_{i=1}^{k_1}\left(Y_i - \overline{Y}_{0,k_1}\right)^2 + \right.\\ &\left. + \sum_{i=k_1+1}^{k_2}\left(Y_i - \overline{Y}_{k_1,k_2}\right)^2 + \cdots + \sum_{i=k_{q_o}+1}^{n}\left(Y_i - \overline{Y}_{k_{q_o},n}\right)^2\right\},\end{aligned}$$

where

$$\overline{Y}_{k,s} = \frac{1}{s-k}\sum_{i=k+1}^{s} Y_i.$$

There exist more sophisticated procedures for detection and identification of multiple changes, however, they need stronger assumptions, for details see, e.g., Yao (1988).

Particular attention has been paid to the *epidemic alternatives* corresponding to the model (II.20) with $q = 2$ and $\mu_1 = \mu_3 \ne \mu_2$. For more information see, among others, Bhattacharya and Brockwell (1976), Yao (1993), Hušková (1995), Antoch and Hušková (1996).

8.5. Change in location or/and scale

The present section deals with changes in location or/and scale in the location model:

$$Y_i = \begin{cases} \mu + \sigma e_i, & i = 1, \dots, \mathsf{m}, \\ \mu + \delta + (\sigma + h) e_i, & i = \mathsf{m}+1, \dots, n, \end{cases} \tag{II.24}$$

where m, μ, $(\delta, h) \neq (0,0)$ and $\sigma^2 > 0$ are parameters, and $e_1, \dots, e_n$ are iid random variables with symmetric distribution around zero, unit variance and $\mathsf{E}\,|e_i|^{4+\Delta} < \infty$ with some $\Delta > 0$. The change point m fulfills (II.2).

If the distribution function F of e_i is known, we can try to derive the maximum likelihood estimators (MLE). For example, if $e_1, \dots, e_n$ are iid with $N(0,1)$ distribution, direct calculation gives the MLE $\widehat{\mathsf{m}}_{MLE}$ of the change point m in the form

$$\widehat{\mathsf{m}}_{MLE} = \arg\max \left\{ \frac{\widetilde{\sigma}_n^n}{\widetilde{\sigma}_k^k (\widetilde{\sigma}_k^o)^{n-k}};\ k \in \{1, \dots, n\} \right\},$$

where for $k = 1, \dots, n-1$,

$$\widetilde{\sigma}_k^2 = \frac{1}{k} \sum_{i=1}^{k} \big(Y_i - \overline{Y}_k\big)^2 \quad \text{and} \quad \widetilde{\sigma}_k^{o2} = \frac{1}{n-k} \sum_{i=k+1}^{n} \big(Y_i - \overline{Y}_k^o\big)^2,$$

i.e. $\widehat{\sigma}_k^2$ and $\widehat{\sigma}_k^{o2}$ are estimators of σ^2 based on $Y_1, \dots, Y_k$ and $Y_{k+1}, \dots, Y_n$, respectively.

The distribution of $\widehat{\mathsf{m}}_{MLE}$ can be approximated using, $x \in \mathcal{R}^1$,

$$P\left(\frac{\frac{\delta^2}{\sigma^2} + \frac{1}{2}\left(\frac{h^2}{\sigma^2}\right)^2}{\frac{\delta^2}{\sigma^2} + \frac{1}{2}\left(\frac{h^2}{\sigma^2}\right)^2 var\, e_1^2} \Big(\widehat{\mathsf{m}}_{MLE} - \mathsf{m}\Big) \le x \right) \approx P(V \le x), \tag{II.25}$$

where the random variable V is defined by (II.10).

Despite the fact, that the estimator $\widehat{\mathsf{m}}_{MLE}$ was developed for observations with normal distribution, it can be used for other underlying distributions, too. The same holds for the approximation (II.25). Moreover, the approximation (II.25) can be used if we replace δ^2, σ^2 and $var\, e_1^2$ by suitable estimators.

Notice, that the estimator $\widehat{\mathsf{m}}_{LS}$ can be used also when we a priori know that the change occurs only in the location (i.e. $h = 0$) or only in the scale σ (i.e. $\delta = 0$); in the approximation, we replace h and δ, respectively, by zero.

Next, we propose an estimator of m that works for a quite general situation described in the model (II.24). The estimator is based on the pairs of the partial sums S_k and $S_{k,sc}$, $k = 1, \dots, n$, where S_k is defined by (II.6) and

$$S_{k,sc} = \sum_{i=1}^{k} \left(\big(Y_i - \overline{Y}_n\big)^2 - \frac{1}{n} \sum_{j=1}^{n} \big(Y_j - \overline{Y}_n\big)^2 \right), \quad k = 1, \dots, n. \tag{II.26}$$

The estimator itself has the form

$$\widetilde{\mathsf{m}}_{SC} = \arg\max\left\{\frac{n}{k(n-k)}\Big(\frac{S_k^2}{\widehat{\sigma}_k^2}+\frac{S_{k,sc}^2}{\widehat{\sigma}_{k,sc}^2}\Big);\ k\in\big\{1,\dots,n-1\big\}\right\}, \qquad \text{(II.27)}$$

where $\widehat{\sigma}_k^2$ and $\widehat{\sigma}_{k,sc}^2$ are, respectively, estimators of σ and $var\,(Y_1-\mu)^2$. We can use, e.g., the estimators

$$\widehat{\sigma}_k^2 = \frac{1}{n-2}\left\{\sum_{i=1}^{k}\big(Y_i-\overline{Y}_k\big)^2+\sum_{i=k+1}^{n}\big(Y_i-\overline{Y}_k^{\,o}\big)^2\right\}$$

and

$$\widehat{\sigma}_{k,sc}^2 = \frac{1}{n-2}\Bigg(\sum_{i=1}^{k}\Big(\big(Y_i-\overline{Y}_k\big)^2-\frac{1}{k}\sum_{j=1}^{k}\big(Y_j-\overline{Y}_k\big)^2\Big)+ \\ +\sum_{i=k+1}^{n}\Big(\big(Y_i-\overline{Y}_k^{\,o}\big)^2-\frac{1}{n-k}\sum_{j=k+1}^{k}\big(Y_j-\overline{Y}_k^{\,o}\big)^2\Big)\Bigg).$$

The approximation to the distribution is of a similar type as above, i.e.:

$$P\left(\frac{\frac{\delta^2}{\sigma^2}+\frac{1}{2}\Big(\frac{h^2}{\sigma^2}\Big)^2}{\frac{\delta^2}{\sigma^2}+\frac{1}{2}\Big(\frac{h^2}{\sigma^2}\Big)^2 var\,e_1^2}\Big(\widetilde{\mathsf{m}}_{SC}-\mathsf{m}\Big)\le x\right)\approx P\big(V\le x\big), \quad x\in\mathcal{R}^1, \qquad \text{(II.28)}$$

where the random variable V is defined by (II.10).

Finally, we present an L_1-estimator that is asymptotically equivalent with (II.27). This estimator has the form

$$\arg\max\left\{\frac{n}{k(n-k)}\sum_{i=1}^{k}\Big(\operatorname{sign}\big(Y_i-\widetilde{\mu}_n\big)\Big)^2+\sum_{i=1}^{k}\Big(\operatorname{sign}\big(|Y_i-\widetilde{\mu}_n|-\widetilde{\nu}_n\big)\Big)^2;\right. \\ \left. k\in\big\{1,..,n-1\big\}\right\}, \qquad \text{(II.29)}$$

where $\widetilde{\mu}_n$ is the sample median of $Y_1,\dots,Y_n$ and $\widetilde{\nu}_n$ is the sample median of $|Y_1-\widetilde{\mu}_n|,\dots,|Y_n-\widetilde{\nu}_n|$. This estimator is very simple to calculate. The approximation to the distribution is the same as in (II.28).

8.6. Gradual changes

We assume the observations $Y_1,\dots,Y_n$ follow the model:

$$Y_i=\begin{cases}\mu+\sigma e_i, & i=1,\dots,\mathsf{m},\\ \mu+\delta\frac{i-\mathsf{m}}{n}+\sigma e_i, & i=\mathsf{m}+1,\dots,n,\end{cases} \qquad \text{(II.30)}$$

where μ, $\sigma^2>0$, $\delta\neq 0$ and $\mathsf{m}\,(<n)$ are unknown parameters, $e_1,\dots,e_n$ are iid random errors with zero mean, nonzero variance σ^2 and $\mathsf{E}\,|e_i|^{2+\Delta}<\infty$ with some $\Delta>0$. Again, m is the change point which is assumed to fulfill

(II.2). This type of change is called *gradual*. The least squares method leads to the following estimator of m, i.e.:

$$\widetilde{\mathsf{m}}_{gr} = \arg\max\left\{ \frac{\left(\sum_{i=k+1}^{n}\left(Y_i - \overline{Y}_n\right)\dfrac{i-k}{n}\right)^2}{\sum_{i=k+1}^{n}\dfrac{(i-k)^2}{n^2} - \dfrac{1}{n}\left(\sum_{i=k+1}^{n}\dfrac{i-k}{n}\right)^2};\ k \in \{1, \ldots, n-1\} \right\},$$

and the parameters μ and δ are estimated by

$$\widehat{\mu}_n = \overline{Y}_n + \frac{\widehat{\delta}_n}{n}\sum_{i=\widehat{\mathsf{m}}_{gr}+1}^{n}\frac{i-\widehat{\mathsf{m}}_{gr}}{n}$$

and

$$\widehat{\delta}_n = \frac{\sum_{i=\widehat{\mathsf{m}}_{gr}+1}^{n}\left(Y_i - \overline{Y}_n\right)\dfrac{i-\widehat{\mathsf{m}}_{gr}+1}{n}}{\sum_{i=\widehat{\mathsf{m}}_{gr}+1}^{n}\left(\dfrac{i-\widehat{\mathsf{m}}_{gr}}{n}\right)^2 - \left(\sum_{i=\widehat{\mathsf{m}}_{gr}+1}^{n}\dfrac{i-\widehat{\mathsf{m}}_{gr}}{n}\right)^2}.$$

The distribution of this estimator is approximately normal, more exactly,

$$P\left(\sqrt{\frac{\theta(1-\theta)}{1+3\theta}}\,\frac{\delta}{\sqrt{n}\sigma}\left(\widetilde{\mathsf{m}}_{gr} - \mathsf{m}\right) \le x\right) \approx \Phi(x), \quad x \in \mathcal{R}^1.$$

Notice that the approximation to the distribution of the estimator $\widetilde{\mathsf{m}}_{gr}$, in this particular case, completely differs from the approximations we had so far. This is due to the fact that we have a gradual (continuous) change while all other considered changes are abrupt (jump).

9. Change in regression

9.1. Change point estimators in regression models

We assume that the observations $Y_1, \ldots, Y_n$ follow the regression model:

$$Y_i = \begin{cases} \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma e_i, & i = 1, \ldots, \mathsf{m}, \\ \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_i'\boldsymbol{\delta} + \sigma e_i, & i = \mathsf{m}+1, \ldots, n, \end{cases}$$

where $\boldsymbol{\beta}$, $\sigma^2 > 0$, $\boldsymbol{\delta} \neq \mathbf{0}$, $\mathsf{m}(< n)$, are parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are p-dimensional vectors and $e_1, \ldots, e_n$ are iid random variables with zero mean, nonzero variance σ^2, and with $\mathsf{E}\,|e_i|^{4+\Delta} < \infty$ with some $\Delta > 0$. The change point m fulfills (II.2).

Similarly as in the testing problem, there are many variants of the assumptions on the design points $\boldsymbol{x}_i$, $i = 1, \ldots, n$. Generally, in the usual linear regression setup the assumptions are stronger than in the estimation problem. We give here only three possible sets of assumptions:

- $\boldsymbol{x}_i = \boldsymbol{g}\big(i/n\big) = \big(g_1(i/n), \dots, g_p(i/n)\big)'$, $i = 1, \dots, n$, where $g_1, \dots, g_p$ are continuous functions on $[0,1]$ such that $\int_a^b \boldsymbol{g}(\boldsymbol{x})\,dx$ is a positive definite matrix for each $0 \le a < b \le 1$.
- There are fixed p linearly independent p-dimensional vectors $\boldsymbol{x}_1^0, \dots,$ $\boldsymbol{x}_p^0$ and the design points are chosen according to some rule from this p vectors.
- The design points can be a realization of AR or ARMA sequences fulfilling certain moment assumptions.

The estimator is defined as follows:

$$\widehat{\mathsf{m}}_{regr} = \arg\min \left\{ \sum_{i=1}^{k} \big(Y_i - \boldsymbol{x}'\widehat{\boldsymbol{\beta}}_k\big)^2 + \sum_{i=k+1}^{n} \big(Y_i - \boldsymbol{x}'\widehat{\boldsymbol{\beta}}_k^{\,o}\big)^2;\ k = 1, \dots, n \right\},$$

where $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{\boldsymbol{\beta}}_k^{\,o}$ are least squares estimators of $\boldsymbol{\beta}$ based on $Y_1, \dots, Y_k$ and $Y_{k+1}, \dots, Y_n$, respectively. Equivalently, the estimators can be also defined as

$$\widehat{\mathsf{m}}_{regr} = \arg\max \left\{ \sum_{i=1}^{k} \boldsymbol{x}_i'\big(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n\big) \cdot \boldsymbol{C}_k^{-1}\boldsymbol{C}_n(\boldsymbol{C}_n - \boldsymbol{C}_k) \cdot \right.$$
$$\left. \cdot \sum_{i=1}^{k} \boldsymbol{x}_i'\big(Y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_n\big);\ k \in \{1, \dots, n\} \right\},$$

where $\boldsymbol{C}_n = \sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i'$. Here $\sum_{i=1}^k \boldsymbol{x}_i'\big(Y_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}_n\big), i = 1, \dots, n$, are vectors of sums of partial residuals.

The approximation to the distribution of the estimators

$$P\left(\frac{1}{\sigma^2}\frac{1}{2k_n} \sum_{i=\widehat{\mathsf{m}}-k_n+1}^{\widehat{\mathsf{m}}+k_n} \big(\boldsymbol{x}_i'\boldsymbol{\delta}\big)^2 \left(\widehat{\mathsf{m}}_{regr} - \mathsf{m}\right) \le x \right) \approx P(V \le x), \quad x \in \mathcal{R}^1,$$

with $k_n \to \infty$, $k_n/n \to 0$ and the random variable V is defined by (II.10).

The readers interested in this problem should consult some more advanced text, e.g., Csörgő and Horváth (1997), Horváth and Kokozska (1997) and Bai (1994).

9.2. Change in regression parameters and/or scale

We assume that the observations $Y_1, \dots, Y_n$ follow the regression model:

$$Y_i = \begin{cases} \boldsymbol{x}_i'\boldsymbol{\beta} + \sigma e_i, & i = 1, \dots, \mathsf{m}, \\ \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_i'\boldsymbol{\delta} + (\sigma + h)e_i, & i = \mathsf{m}+1, \dots, n, \end{cases}$$

where $\boldsymbol{\beta}$, $\sigma^2 > 0$, $(\boldsymbol{\delta}, h) \ne \boldsymbol{0}$, $\mathsf{m}(< n)$, are parameters, and $e_1, \dots, e_n$ are iid random variables with unit variance, and with $\mathsf{E}\,|e_i|^{4+\Delta} < \infty$ with some $\Delta > 0$, the distribution of e_i is supposed to be symmetric around 0. The assumptions on the design points are the same as in the previous section and the change point m fulfills (II.2).

The estimator is defined as

$$\widetilde{\mathsf{m}}_{regr} = \arg\max \left\{ \frac{\widetilde{\sigma}_n^n}{\widetilde{\sigma}_k^k \widetilde{\sigma}_k^{o(n-k)}};\ k \in \{1, \ldots, n-1\} \right\},$$

where

$$\widetilde{\sigma}_k^2 = \frac{1}{k} \sum_{i=1}^{k} \left(Y_i - \boldsymbol{x}'\widehat{\boldsymbol{\beta}}_k\right)^2, \quad k = 1, \ldots, n-1,$$

$$\widetilde{\sigma}_k^{o2} = \frac{1}{n-k} \sum_{i=k+1}^{n} \left(Y_i - \boldsymbol{x}'\widehat{\boldsymbol{\beta}}_k^0\right)^2, \quad k = 1, \ldots, n-1,$$

and $\widehat{\boldsymbol{\beta}}_k$ and $\widehat{\boldsymbol{\beta}}_k^o$ are the least squares estimators of $\boldsymbol{\beta}$ based on $Y_1, \ldots, Y_k$ and $Y_{k+1}, \ldots, Y_n$. The approximation to the distribution of $\widetilde{\mathsf{m}}_{regr}$ is $\forall x \in \mathcal{R}^1$

$$P\left(\left(\frac{1}{\sigma^2}\frac{1}{2k_n} \sum_{i=\widehat{\mathsf{m}}-k_n+1}^{\widehat{\mathsf{m}}+k_n} \left(\boldsymbol{x}_i'\boldsymbol{\delta}\right)^2 + \frac{1}{2}\left(\frac{h}{\sigma^2}\right)^2\right)^2 \left(\widetilde{\mathsf{m}}_{regr} - \mathsf{m}\right) \le x\right) \approx P\left(V \le x\right),$$

where the random variable V is defined by (II.10).

10. Confidence intervals

We focus on the confidence intervals based on the estimator $\widehat{\mathsf{m}}_{LS}$, defined by (II.7), of the change points in the location model studied in Section 8. However, we attempt to keep in mind generalization to other estimators and models as well.

Three types of confidence intervals will be developed, one based on the limit distribution of the (point) estimator(s) of m and two based on the bootstrap methods. All three methods are suitable for local changes while only the bootstrap constructions apply also to fixed changes.

Asymptotic approach

Using the approximation to the distribution of $\widehat{\mathsf{m}}_{LS}$ described in (II.9), we get the $100(1-\alpha)\%$ approximate confidence interval

$$\left(\widehat{\mathsf{m}}_{LS} - v_{1-\alpha/2}\frac{\widehat{\sigma}_n^2}{\widehat{\delta}_n^2},\ \widehat{\mathsf{m}}_{LS} + v_{1-\alpha/2}\frac{\widehat{\sigma}_n^2}{\widehat{\delta}_n^2}\right), \qquad \text{(II.31)}$$

where $\widehat{\mathsf{m}}_{LS}$, $\widehat{\delta}_n$ and $\widehat{\sigma}_n^2$ are defined in (II.7), (II.5) and (II.12), respectively, and $v_{1-\alpha/2}$ is the quantile corresponding to the random variable V defined in (II.10). (Selected quantiles are in Table 18.)

Now, we turn to the bootstrap approximations to the confidence intervals based on the estimator $\widehat{\mathsf{m}}_{LS}$ of m.

Bootstrap sampling scheme I.

Take two independent samples $Y_1^*, \ldots, Y_{\widehat{\mathsf{m}}}^*$ and $Y_{\widehat{\mathsf{m}}+1}^*, \ldots, Y_n^*$ from the empirical cumulative distribution function of $Y_1, \ldots, Y_{\widehat{\mathsf{m}}}$ and $Y_{\widehat{\mathsf{m}}+1}, \ldots, Y_n$, respectively. Then the bootstrap estimator $\widehat{\mathsf{m}}_{LS}^*$ (corresponding to the estimator $\widehat{\mathsf{m}}_{LS}$) is defined as

$$\widehat{\mathsf{m}}^*_{LS} = \arg\max \left\{ \sqrt{\frac{n}{k(n-k)}} \left| \sum_{i=1}^{k} \left(Y_i^* - \overline{Y}_n^* \right) \right|; \ k \in \{1, \ldots, n-1\} \right\}. \tag{II.32}$$

This means, that the bootstrap estimator is calculated exactly as $\widehat{\mathsf{m}}$, however, $Y_i, \ i = 1, \ldots, n$ are replaced by their bootstrap counterparts $Y_i^*, \ i = 1, \ldots, n$.

We can also modify the procedure taking the maximum only over the set $\{k : |k - \widehat{\mathsf{m}}| \le D_n\}$, with $\{D_n\}$ fulfilling, as $n \to \infty$,

$$D_n \le \min(\mathsf{m}, n - \mathsf{m}) \quad \text{and} \quad D_n \delta_n^2 \longrightarrow \infty. \tag{II.33}$$

If the amount of the change $\delta_n \equiv \delta \neq 0$ is fixed, $\{D_n\}$ can be chosen to tend to infinity arbitrary slowly, while in the case of local changes $\{D_n\}$ has to tend to infinity faster than ${\delta_n}^{-2}$.

The modified bootstrap estimator $\widehat{\widehat{\mathsf{m}}}^*_{LS}$ related to $\widehat{\mathsf{m}}_{LS}$ is defined as

$$\widehat{\widehat{\mathsf{m}}}^*_{LS} = \arg\max \left\{ \sqrt{\frac{n}{k(n-k)}} \left| \sum_{i=1}^{k} \left(Y_i^* - \overline{Y}_n^* \right) \right|; \ \left|k - \widehat{\mathsf{m}}_{LS}\right| \le D_n \right\}, \tag{II.34}$$

where $\{D_n\}$ fulfills (II.33).

The bootstrap estimators $\widehat{\mathsf{m}}^*_{LS}(G)$ and $\widehat{\widehat{\mathsf{m}}}^*_{LS}(G)$ related to $\widehat{\mathsf{m}}_{LS}(G)$ are defined accordingly.

Bootstrap sampling scheme II.

Define the estimated residuals

$$\widetilde{e}_i = \begin{cases} Y_i - \overline{Y}_{\widehat{\mathsf{m}}}, & i = 1, \ldots, \widehat{\mathsf{m}}, \\ Y_i - \overline{Y}^o_{\widehat{\mathsf{m}}}, & i = \widehat{\mathsf{m}} + 1, \ldots, n, \end{cases}$$

and the centered residuals

$$\widehat{e}_i = \widetilde{e}_i - \frac{1}{n} \sum_{j=1}^{n} \widetilde{e}_j, \quad i = 1, \ldots, n,$$

where $\overline{Y}_{\widehat{\mathsf{m}}}$ and $\overline{Y}^o_{\widehat{\mathsf{m}}}$ are defined by (I.2), respectively, with $\mathsf{m} = \widehat{\mathsf{m}}$.

Take $e_1^{**}, \ldots, e_n^{**}$ iid from the empirical cdf of $\widehat{e}_1, \ldots, \widehat{e}_n$ and consider the bootstrap observations

$$Y_i^{**} = \begin{cases} \widehat{Y}_{\widehat{\mathsf{m}}} + e_i^{**}, & i = 1, \ldots, \widehat{\mathsf{m}}, \\ \widehat{Y}^o_{\widehat{\mathsf{m}}} + e_i^{**}, & i = \widehat{\mathsf{m}} + 1, \ldots, n. \end{cases}$$

Then we proceed as in the bootstrap sampling scheme I, i.e. we apply (II.32) and (II.34) with Y_i^* replaced by Y_i^{**}.

Both these bootstrap schemes provide a reasonable approximation to the confidence intervals for the change point m, typically they provide a better approximation than the one based on the approximation (II.31).

Part III. Selected limit properties

11. Selected limit theorems for test statistics

The test statistics which appeared in the text are mostly functionals of partial sums of iid variables e_i. Their distributions are very complex. However, as we suppose that the number of observations n is large, their asymptotic distributions are of interest. The theory used for obtaining the limit distributions is based on the Donsker invariance principle and the theory of strong approximations. Therefore, the limit distributions do not depend on the assumption of the normality of e_i. Usually it is sufficient to suppose that the random errors e_i are iid satisfying $\mathsf{E}\, e_i = 0, \mathsf{E}\, e_i^2 = 1$ and $\mathsf{E}\, |e_i|^{2+\delta} < \infty$ for some $\delta > 0$.

The over-all maximum-type statistics go to infinity as $n \to \infty$ almost surely. The speed of the convergence of their critical values can be traced by approximating these statistics by the maximum of certain processes over an increasing interval, e.g.,

$$\left| \max_{1\le k\le n} \frac{|\sum_{i=1}^k e_i|}{\sqrt{k}} - \sup_{1/n\le t\le 1} \frac{|W(t)|}{\sqrt{t}} \right| = o_P\left(\frac{1}{\sqrt{2\log\log n}}\right),$$

$$\left| \max_{1\le k\le n-1} \left| \left(\tfrac{k}{n}\left(1-\tfrac{k}{n}\right)\right)^{-\eta} \frac{1}{\sqrt{n}} \sum_{i=1}^k \left(e_i - \overline{e}_n\right) \right| - \sup_{1/n\le t\le 1-1/n} \frac{|B(t)|}{\left(t(1-t)\right)^{\eta}} \right| =$$

$$= o_P\left(\frac{1}{\sqrt{2\log\log n}}\right)$$

$\forall \eta \in [0, 1/2]$, and

$$\left| \max_{1\le k\le n} \frac{\left|\frac{1}{\sqrt{n}} \sum_{i=1}^k \frac{k-i}{n} e_i\right|}{\sqrt{\frac{1}{n}\sum_{i=1}^k \left(\frac{k-i}{n}\right)^2}} - \sup_{1/n\le t\le 1} \frac{|\int_0^t (t-x)\, dW(x)|}{\sqrt{t^3/3}} \right| = o_P\left(\frac{1}{\sqrt{2\log\log n}}\right)$$

Similar approximations may be derived for a multidimensional case, for details see the book Csörgő and Horváth (1997).

Next we present the limit behavior of some of over-all maximum type statistics. The type of limits are known as extreme value theorems. Under the assumptions formulated at the beginning of the section and with $a_n = \sqrt{2\log\log n}$ we have $\forall y \in \mathcal{R}^1$, as $n \to \infty$,

$$P\left(a_n \max_{1\le k<n} \frac{1}{\sqrt{k}} \sum_{i=1}^k e_i \le y + 2\log\log n + \frac{\log\log\log n}{2} - \frac{\log \pi}{2}\right) \to \exp\left\{\frac{e^{-y}}{2}\right\},$$

$$P\left(a_n \max_{1\le k<n} \frac{1}{\sqrt{k}} \left|\sum_{i=1}^k e_i\right| \le y + 2\log\log n + \frac{\log\log\log n}{2} - \frac{\log \pi}{2}\right) \to \exp\left\{e^{-y}\right\},$$

$$P\left(a_n \max_{1\le k<n} \sqrt{\frac{n}{k(n-k)}}\left|\sum_{i=1}^{k}(e_i-\overline{e}_n)\right| \le \right.$$
$$\left. \le y + 2\log\log n + \frac{1}{2}\log\log\log n - \frac{1}{2}\log\pi\right) \to \exp\left\{2e^{-y}\right\},$$

$$P\left(a_n \max_{1\le k<n} \frac{\left|\sum_{i=1}^{k}(k-i)e_i\right|}{\sqrt{\sum_{i=1}^{k}(k-i)^2}} \le y + 2\log\log n - \frac{1}{2}\log(4\pi/3)\right) \to \exp\{2e^{-y}\},$$

$$P\Big(a_n \max_{1\le k<n} \frac{\left|\sum_{i=1}^{k}(k-i)(e_i-\overline{e}_n)\right|}{\sqrt{\sum_{i=1}^{k}(k-i)^2 - (\sum_{i=1}^{k}(k-i))^2/n}} \le$$
$$\left. \le y + 2\log\log n - \frac{1}{2}\log(4\pi/3)\right) \to \exp\left\{2e^{-y}\right\}\cdot$$

Some of the above results can be extended to the multivariate case. Particularly, if $\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_n$ are iid p-dimensional random vectors with zero mean, positive variance matrix $\boldsymbol{\Sigma}$ and $\mathsf{E}|Y_{ij}|^{2+\Delta} < \infty$ with some $\Delta > 0$, $i = 1,\ldots,n$, $j = 1,\ldots,p$, then, for any $y \in \mathcal{R}^1$,

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

as $n \to \infty$, where

$$\boldsymbol{S}_k = \left(S_{1,k},\ldots,S_{p,k}\right)', \quad \boldsymbol{S}_{j,k} = \sum_{i=1}^{k}\left(Y_{ij}-\overline{Y}_j\right), \quad k=1,\ldots,n,\ j=1,\ldots,p.$$

For maximum-type statistic of moving sums under the assumptions (denoting $G = G_n$)

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

we have for any $y \in \mathcal{R}^1$, as $n \to \infty$,

$$P\left(\sqrt{2\log\frac{n}{G}} \max_{G<k<n-G} \left\{\frac{1}{\sqrt{G}}\frac{1}{\sigma}\left|S_{k+G}-2S_k+S_{k-G}\right|\right\} \le \right.$$
$$\left. \le y + 2\log\frac{n}{G} + \frac{1}{2}\log\log\frac{n}{G} - \frac{1}{2}\log\frac{4\pi}{9}\right) \to \exp\left\{-2\exp\{-y\}\right\}.$$

For a trimmed maximum-type statistic as well as for a sum-type statistic we obtain the limit distributions by approximating them, respectively, by the maximum of some processes over a fixed interval, e.g.,

$$\sum_{k=1}^{n} \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{k} e_i \right)^2 \xrightarrow{\mathcal{D}} \int_0^1 W^2(t)\, dt,$$

$$\max_{\lfloor \beta n \rfloor \le k \le n} \frac{|\sum_{i=1}^{k} e_i|}{\sqrt{k}} \xrightarrow{\mathcal{D}} \max_{\beta \le t \le 1} \frac{|W(t)|}{\sqrt{t}}, \forall \beta \in (0,1),$$

$$\max_{\lfloor \beta n \rfloor \le k \le n} \frac{\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \frac{k-i}{n} e_i \right|}{\sqrt{\frac{1}{n} \sum_{i=1}^{k} \left(\frac{k-i}{n}\right)^2}} \xrightarrow{\mathcal{D}} \max_{\beta \le t \le 1} \frac{|\int_0^t (t-x)\, dW(x)|}{\sqrt{t^3/3}}, \forall \beta \in (0,1),$$

$$\max_{\lfloor \beta n \rfloor \le k \le \lfloor (1-\beta) n \rfloor} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^{k} \left(e_i - \overline{e}_n\right) \right| \xrightarrow{\mathcal{D}}$$

$$\max_{\beta \le t \le (1-\beta)} \frac{|B(t)|}{\sqrt{t(1-t)}}, \forall \beta \in (0,1/2),$$

and

$$\max_{1 \le k \le n-1} \left| \left(\frac{k}{n}\left(1 - \frac{k}{n}\right) \right)^{-\eta} \frac{1}{\sqrt{n}} \sum_{i=1}^{k} \left(e_i - \overline{e}_n\right) \right| \xrightarrow{\mathcal{D}}$$

$$\sup_{1/n \le t \le 1 - 1/n} \frac{|B(t)|}{\left(t(1-t)\right)^{\eta}}, \forall \eta \in [0,1/2).$$

Here $\xrightarrow{\mathcal{D}}$ denotes the convergence in distibution, $\{W(t),\ t \in (0,1)$ and $\{B(t),\ t \in (0,1)$ are Wiener process and Brownian bridge, respectively. The random variables on the rhs are functionals of Wiener process and Brownian bridge. The exact form of their distribution functions are usually explicitely unknown. Some approximations were developed, selected details are described bellow. For more information see Billingsley (1968) and Csörgő and Horváth (1993).

12. Properties of maximum of one-dimensional Gaussian processes

The limit distributions of the maximum-type statistics are given by distributions of a maximum of a Gaussian process either over a fixed or an increasing interval. As we wish to approximate the upper quantiles of our test statistics, we are interested in the high level exceedence probabilities of the considered limit processes. More specifically, approximate critical values can be found applying theorems for the exceedence probabilities of stationary Gaussian processes, see Theorem 12.2.9 and 12.3.5 of Leadbetter et al. (1983).

Let $\{\xi(t), t \geq 0\}$ be a zero-mean standardised stationary Gaussian process with the covariance function $\rho_\xi(\tau)$ that satisfies

$$\lim_{\tau\to\infty} \rho_\xi(\tau) \log\tau = 0 \tag{III.1}$$

and has the following expansion at zero, i.e.,

$$\rho_\xi(\tau) = 1 - C|\tau|^\alpha + o(|\tau|^\alpha) \quad \text{as} \quad \tau \to 0,$$

where $\alpha = 1$ or $\alpha = 2$. Then it holds:

1) for T being fixed and the exceedence level $u \to \infty$

$$\lim_{u\to\infty} \frac{P\big(\max_{0\leq t\leq T} \xi(t) > u\big)}{u^{2/\alpha-1}\phi(u)} \;=\; TC^{1/\alpha}H_\alpha; \tag{III.2}$$

2) for T tending to infinity and

$$x_T = \sqrt{2\log T} + \frac{1}{\sqrt{2\log T}}\left(x + \frac{2-\alpha}{2\alpha}\log\log T + \log\left(C^{1/\alpha}H_\alpha \frac{2^{(2-\alpha)/2\alpha}}{\sqrt{2\pi}}\right)\right),$$

$$\lim_{T\to\infty} P\left(\max_{0\leq t\leq T} \xi(t) > x_T\right) = 1 - \exp\left\{-e^{-x}\right\},$$

$x \in \mathcal{R}^1, H_1 = 1, H_2 = 1/\sqrt{\pi}$.

If we are interested in the maximum of absolute values of studied processes then, under the same conditions as above, the following properties hold:

1) for T being fixed and the exceedence level $u \to \infty$

$$\lim_{u\to\infty} \frac{P\big(\max_{0\leq t\leq T} |\xi(t)| > u\big)}{u^{2/\alpha-1}\phi(u)} \;=\; 2TC^{1/\alpha}H_\alpha, \tag{III.3}$$

2) for T tending to infinity and x_T as above

$$\lim_{T\to\infty} P\left(\max_{0\leq t\leq T} |\xi(t)| > x_T\right) = 1 - \exp\left\{-2e^{-x}\right\}. \tag{III.4}$$

Remark: For T fixed and u large, the following approximations seem to be more accurate than (III.2) and (III.3), i.e.

$$P\left(\max_{0\leq t\leq T}\; \xi(t) > u\right) \approx \big(1 - \Phi(u)\big) \;+\; u^{2/\alpha-1}\phi(u)TC^{1/\alpha}H_\alpha$$

and

$$P\left(\max_{0\leq t\leq T}\; \big|\xi(t)\big| > u\right) \approx 2\big(1 - \Phi(u)\big) \;+\; 2u^{2/\alpha-1}\phi(u)TC^{1/\alpha}H_\alpha.$$

Example 1: Consider the process

$$\left\{Y(t) = \frac{\int_0^t x^p\, dW(x)}{\sqrt{t^{2p+1}/(2p+1)}},\; t \in (0,1]\right\}, \quad p = 0, 1, \ldots,$$

and the process $\{U(t) = Y(e^{-t}),\, t \in [0,\infty)\}$. The covariance function of the process $\{Y(t),\, t \in (0,1]\}$ is $R_Y(t,s) = (t/s)^{p+1}, 0 < t \leq s \leq 1$. It follows that $\{U(t), t \in [0,\infty)\}$ is a zero-mean standardized Gaussian process with the covariance function

$$R_U(\tau) = \mathsf{E}\, U(t)U(t+\tau) = \exp\left\{-\frac{2p+1}{2}\tau\right\}, \quad \tau \geq 0,$$

that satisfies (III.1) and

$$R_u(\tau) = 1 - \frac{2p+1}{2}|\tau| + o(\tau) \quad \text{as} \quad \tau \to 0.$$

It follows that

$$P\left(\max_{\beta \leq t \leq 1} |Y(t)| > u\right) = P\left(\max_{0 \leq t \leq -\log \beta} |U(t)| > u\right)$$
$$\approx 2(-\log \beta)\, \frac{2p+1}{2}\, u\phi(u).$$

for small β and large u. Particularly, for $p = 0$ we get

$$P\left(\max_{\beta \leq t \leq 1} \frac{|\, W(t)\,|}{\sqrt{t}} > u\right) \approx u\phi(u) \log \frac{1}{\beta}$$

A better approximation can be found in the subsection 13.

Example 2: Consider the process

$$\left\{YC(t) = \frac{\int_0^t (t-x)^p \, dW(x)}{\sqrt{t^{2p+1}/(2p+1)}}, t \in (0,1]\right\}, \quad p = 1, 2, \ldots,$$

with the covariance function

$$R_{YC}(t,s) = (2p+1) \sum_{i=0}^{p} \binom{p}{i} \frac{1}{p+i+1} \left(\frac{t}{s}\right)^{i+\frac{1}{2}} \left(1 - \frac{t}{s}\right)^{p-i}, \ 0 < t \leq s \leq 1.$$

The process $\{UC(t) = YC(e^{-t}),\ t \in [0,\infty)\}$ is a zero-mean standardized stationary Gaussian process with the covariance function

$$R_{UC}(t) = \big(2p+1\big) \sum_{i=0}^{p} \binom{p}{i} \frac{1}{p+i+1} e^{-(i+1/2)\tau} \left(1 - e^{-\tau}\right)^{p-i}, \quad \tau \geq 0$$

satisfying (III.1) and

$$R_{UC}(\tau) = 1 - \frac{2p+1}{8(2p-1)}\tau^2 + o(\tau^2) \quad \text{as} \quad \tau \to 0.$$

It follows that

$$P\left(\max_{\beta \leq t \leq 1} |YC(t)| > u\right) = P\left(\max_{0 \leq t \leq -\log \beta} |UC(t)| > u\right) \approx$$
$$\approx 2(-\log \beta) \sqrt{\frac{2p+1}{8(2p-1)}}\, \frac{1}{\pi}\, \phi(u).$$

The former theory for stationary Gaussian processes were generalized for locally stationary Gaussian processes by Hüsler (1990), Hüsler (1993) and

Bräker (1993). Let $\{\xi(t), t \geq 0\}$ be a zero-mean locally stationary Gaussian process with the covariance function $\rho_\xi(t,s)$ that satisfies

$$\lim_{\tau\to\infty}\Big(\sup\big\{\rho_\xi(t,s), |\,t-s\,| > \tau\big\}\Big)\log\tau = 0$$

and has the following expansion at zero, i.e.

$$\rho_\xi\big(t, t+\tau\big) = 1 - C(t)|\,\tau\,|^\alpha + o(|\,\tau\,|^\alpha) \quad \text{as} \quad \tau \to 0,$$

where $\alpha = 1$ or $\alpha = 2$ and $0 < \inf C(t) \leq \sup C(t) < \infty$. Then

1) for T being fixed and the exceedence level $u \to \infty$

$$\lim_{u\to\infty}\frac{P\big(\max_{0\leq t\leq T}\xi(t) > u\big)}{u^{2/\alpha-1}\phi(u)} = H_\alpha\int_0^T\big(C(t)\big)^{1/\alpha}\,dt;$$

2) for $T \to \infty$ and

$$x_T = \sqrt{2\log T} + \frac{1}{\sqrt{2\log T}}\left(x + \frac{2-\alpha}{2\alpha}\log\log T + \log\left(C_T^\star H_\alpha\frac{2^{(2-\alpha)/2\alpha}}{\sqrt{2\pi}}\right)\right),$$

where $C_T^\star = T^{-1}\int_0^T\big(C(t)\big)^{1/\alpha}\,dt$,

$$\lim_{T\to\infty}P\left(\max_{0\leq t\leq T}\xi(t) > x_T\right) = 1 - \exp\left\{-e^{-x}\right\}, \quad x \in \mathcal{R}^1.$$

13. Exceedence level properties of Wiener process and Brownian bridge

The standardized Wiener process $\{W(t)/\sqrt{t}, \beta \leq t \leq 1\}$ and standardized Brownian bridge process

$$\left\{\frac{B(t)}{\sqrt{t(1-t)}} = \frac{W(t) - tW(1)}{\sqrt{t(1-t)}},\ \beta \leq t \leq 1-\beta\right\}$$

belong, for every $0 < \beta < 1/2$, to locally stationary random processes with $\alpha = 1$, so that the described exceedence theory for locally stationary random processes may be applied to get high level exceedence probabilities that we applied in our text. However, for these two, a better approximation is known, i.e., for small β and large x

$$P\left(\sup_{\beta\leq t\leq 1}\frac{|\,W(t)\,|}{\sqrt{t}} > x\right) \approx x\phi(x)\left((1 - \frac{1}{x^2})\log\frac{1}{\beta} + \frac{4}{x^2}\right)$$

and

$$P\left(\sup_{\beta\leq t\leq 1}\frac{|\,B(t)\,|}{\sqrt{t(1-t)}} > x\right) \approx 2x\phi(x)\left((1 - \frac{1}{x^2})\log\frac{1-\beta}{\beta} + \frac{2}{x^2}\right)\cdot$$

For the multivariate Wiener process $\boldsymbol{W}(t) = \{W_1(t), \ldots, W_d(t)\}$ with independent components and the multivariate Brownian bridge process $\boldsymbol{B}(t) = \{B_1(t), \ldots, B_d(t)\}$, the following approximations hold, i.e., for small β and large x

$$P\Big(\sup_{\beta\le t\le 1}\frac{W_1^2(t)+\dots+W_d^2(t)}{t}>x^2\Big)\approx\frac{x^d e^{-x^2/2}}{2^{d/2}\Gamma(d/2)}\Big(\Big(1-\frac{d}{x^2}\Big)\log\frac{1}{\beta}+\frac{4}{x^2}\Big)$$

and

$$P\Big(\sup_{\beta\le t\le (1-\beta)}\frac{B_1^2(t)+\dots+B_d^2(t)}{t(1-t)}>x^2\Big)\approx\frac{2x^d e^{-x^2/2}}{2^{d/2}\Gamma(d/2)}\Big(\Big(1-\frac{d}{x^2}\Big)\log\frac{1-\beta}{\beta}+\frac{2}{x^2}\Big)$$

14. Selected limit properties of of change point estimators

In this section we formulate some important theoretical results on behavior of change point estimators described in Section 8.

If random variables $Y_1,\dots,Y_n$ follow the model (II.1) with $\delta_n\equiv\delta\neq 0$ fixed, $\mathsf{m}=\lfloor n\gamma\rfloor, \gamma\in(0,1)$ and $e_1,\dots,e_n$ be iid random variables with zero mean, nonzero variance σ^2 and $\mathsf{E}\,|e_i|^{2+\Delta}<\infty$ with some $\Delta>0$. We deal with $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ defined in (II.7) and (II.8).

Then $\widehat{\mathsf{m}}_{LS}-\mathsf{m}$, as $n\to\infty$, has the limit distribution as

$$\arg\max\ \Big\{\delta W_I(j)-\delta^2|j|/2;\ j=0,\pm1,\pm2,\dots\Big\},$$

where

$$W_I(j)=\begin{cases}0, & j=0,\\ \sum_{i=j}^{0}e_i, & j=-1,-2,\dots,\\ -\sum_{i=1}^{j}e_i, & j=1,2,\dots\end{cases}$$

This assertion says that the approximation depends on the distribution of the error terms e_i, which sometimes can create problems.

However, this is not the case with local changes i.e. when $\delta\equiv\delta_n\to 0$ as $n\to\infty$. Particularly, if

$$\delta_n\to 0\quad\text{and}\qquad |\delta_n|\sqrt{n}/\sqrt{\log\log n}\ \to\infty. \tag{III.5}$$

then, as $n\to\infty$, $\delta_n^2\sigma^{-2}\Big(\widehat{\mathsf{m}}_{LS}-\mathsf{m}\Big)$ has the same distribution as V, where V is a random variable defined by (II.10). The distribution function of V is known and it has the form (II.13).

If, moreover,

$$G/n\to 0\quad\text{and}\quad G^{-1}n^{2/(2+\Delta)}\log n\to 0\quad\text{as } n\to\infty, \tag{III.6}$$

then, as $n\to\infty$, $2\delta_n^2\big(3\sigma^2\big)^{-1}\Big(\widehat{\mathsf{m}}_{LS}(G)-\mathsf{m}\Big)$ has the same limit distribution as V, where V is a random variable defined by (II.10).

In case $\delta=0$ (no change) $\widehat{\mathsf{m}}_{LS}$ and $\widehat{\mathsf{m}}_{LS}(G)$ have different limit distributions. Particularly, we have $\forall\varepsilon\in(0,1/2]$

$$P\Big(\widehat{\mathsf{m}}_{LS}<n\varepsilon\Big)\approx 1/2,\qquad P\Big(\widehat{\mathsf{m}}_{LS}>n(1-\varepsilon)\Big)\approx 1/2$$

and $\widehat{\mathsf{m}}_{LS}(G)$ has approximately uniform distribution on $\big\{G+1,\dots,n-G\big)$.

Part IV. Selected Matlab codes

We decided to offer the reader the possibility to use the methods described in details in this paper. Therefore, we prepared Matlab codes covering them. As an example, several of these macros including detailed description and links to the previous sections are included here. They cover the following models, i.e. detection of:

(1) change in mean with unknown starting value, maximum – type test statistic;
(2) change in mean with unknown starting value, sum – type test statistic;
(3) change in variance;
(4) change in mean and/or variance.

More codes are available from the authors.

Respective macros were written and tested in Matlab version 6.0.0.88 (R12 of September 22, 2000. However, we believe that they can be run under much older versions of the Matlab as well.

Each user should take into account that the critical values are not implemented into the codes. The reason is that, for each method, there exist several possibilities how to calculate them using either asymptotic results, Bonferroni inequality or simulations. Despite the fact that the authors personally prefer, for most situations, the simulated critical values, we leave the choice on the reader. For each of the methods, this paper reports all the possibilities.

As an example we used the following three data sets:

Nile data

This data set is probably one of the most frequently used in the change point setup, for details see, e.g., Cobb (1978) or Hinkley and Schechtman (1987). The data correspond to the annual flows (in billions of cubic meters) in the Nile river at Aswan (Egypt) during the years 1871 – 1970.

Simulated data I.

The data used as the testing example for the subroutine `m3_2.m` were generated as follows:

1) $X_1, \ldots, X_{20}$ follow standard normal distribution $N(0, 1)$;
2) $X_{21}, \ldots, X_{50}$ follow a normal distribution $N(0, 3^2)$;

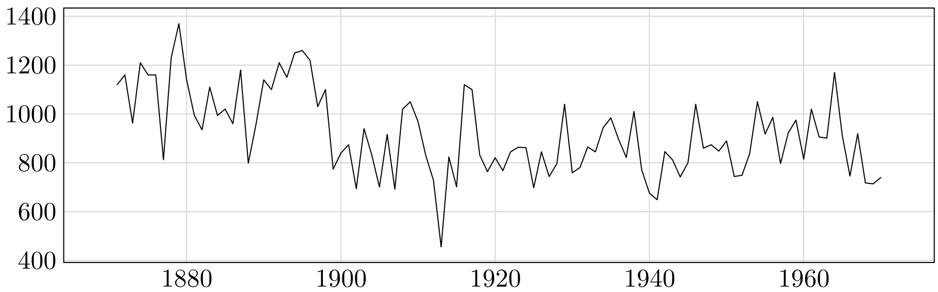
Simulated data II.

The data used as the testing example for the subroutine `m3_3.m` were generated as follows:

1) $X_1, \ldots, X_{20}$ follow standard normal distribution $N(0, 1)$;
2) $X_{21}, \ldots, X_{50}$ follow a normal distribution $N(4, 3^2)$;

year	flow	year	flow	year	flow	year	flow
1871	1120	1896	1220	1921	768	1946	1040
1872	1160	1897	1030	1922	845	1947	860
1873	963	1898	1100	1923	864	1948	874
1874	1210	1899	774	1924	862	1949	848
1875	1160	1900	840	1925	698	1950	890
1876	1160	1901	874	1926	845	1951	744
1877	813	1902	694	1927	744	1952	749
1878	1230	1903	940	1928	796	1953	838
1879	1370	1904	833	1929	1040	1954	1050
1880	1140	1905	701	1930	759	1955	918
1881	995	1906	916	1931	781	1956	986
1882	935	1907	692	1932	865	1957	797
1883	1110	1908	1020	1933	845	1958	923
1884	994	1909	1050	1934	944	1959	975
1885	1020	1910	969	1935	984	1960	815
1886	960	1911	831	1936	897	1961	1020
1887	1180	1912	729	1937	822	1962	906
1888	799	1913	456	1938	1010	1963	901
1889	958	1914	824	1939	771	1964	1170
1890	1140	1915	702	1940	676	1965	912
1891	1100	1916	1120	1941	649	1966	746
1892	1210	1917	1100	1942	846	1967	919
1893	1150	1918	832	1943	812	1968	718
1894	1250	1919	764	1944	742	1969	714
1895	1260	1920	821	1945	801	1970	740

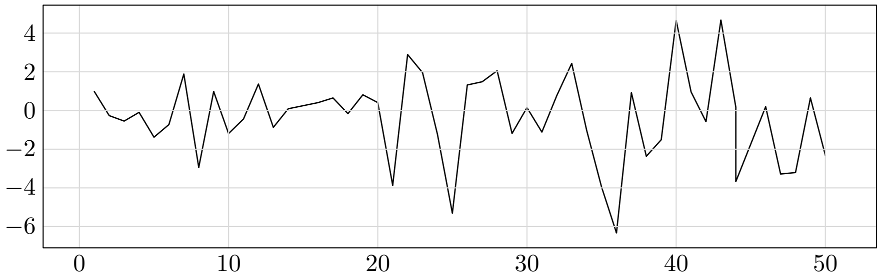
The annual flows (in billions of cubic meters) in river Nile at Aswan (Egypt) during the years 1871−1970. Data are used as the example in subroutines `m3_1_be.m`. and `m3_1_le.m`.



Nile data.

i	value	i	value	i	value	i	value	i	value
1	0.979	11	-0.438	21	-3.870	31	-1.116	41	0.962
2	-0.265	12	1.366	22	2.891	32	0.770	42	-0.578
3	-0.548	13	-0.872	23	1.957	33	2.432	43	4.673
4	-0.096	14	0.089	24	-1.232	34	-1.020	44	0.169
5	-1.380	15	0.247	25	-5.308	35	-3.955	45	-3.673
6	-0.728	16	0.407	26	1.318	36	-6.327	46	0.195
7	1.886	17	0.648	27	1.485	37	0.919	47	-3.284
8	-2.941	18	-0.164	28	2.053	38	-2.366	48	-3.208
9	0.980	19	0.811	29	-1.187	39	-1.511	49	0.649
10	-1.191	20	0.408	30	0.147	40	4.659	50	-2.317

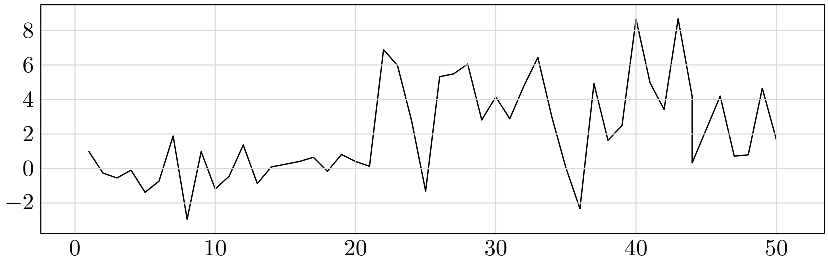
Simulated data used as the example in subroutine `m3_2.m`.



Simulated data I.

i	value	i	value	i	value	i	value	i	value
1	0.979	11	-0.438	21	0.129	31	2.883	41	4.962
2	-0.265	12	1.366	22	6.891	32	4.770	42	3.422
3	-0.548	13	-0.872	23	5.957	33	6.432	43	8.673
4	-0.096	14	0.089	24	2.767	34	2.979	44	4.169
5	-1.380	15	0.247	25	-1.308	35	0.044	45	0.326
6	-0.728	16	0.407	26	5.318	36	-2.327	46	4.195
7	1.886	17	0.648	27	5.485	37	4.919	47	0.715
8	-2.941	18	-0.164	28	6.053	38	1.633	48	0.791
9	0.980	19	0.811	29	2.812	39	2.488	49	4.649
10	-1.191	20	0.408	30	4.147	40	8.659	50	1.682

Simulated data used as the example in subroutine `m3_3.m`.



Simulated data II.

15. Change in mean with unknown starting value, max-type test statistics

```
function Results = m3_1_le(data, sigma2, trimming)

%%% SUBROUTINE m3_1.m
%%%
%%% PURPOSE To test the null hypothesis H against the alternative
%%%         A in the model:
```

$$
\begin{aligned}
&H : Y_i = \mu + e_i, && i = 1, \ldots, n,\\
&A : \exists\, \mathsf{m} \in \{1, \ldots, n-1\} \quad \text{such that} && \\
&\quad Y_i = \mu + e_i, && i = 1, \ldots, \mathsf{m},\\
&\quad Y_i = \mu + \delta + e_i, && i = \mathsf{m}+1, \ldots, n,\ \delta \neq 0.
\end{aligned} \tag{I.16}
$$

We suppose that the variables e_i are iid with the normal distribution $N\big(0, \sigma^2\big)$, σ^2 either known or unknown. If σ^2 is known, we apply statistics (I.17) or (I.18). On the other hand, if σ^2 is unknown, we apply statistics (I.27) or (I.28) with σ^2 estimated using (I.29).

```
%%%
%%% CALLING SEQUENCE
%%%
%%% function   Results = m3_1(data, sigma2, trimming)
%%%
%%% INPUT
%%%
%%% data            ...  data series to be tested for the presence
%%%                      of change point in model (I.16)
%%%
%%% sigma2          ...  variance of the error term
%%%
%%%    > 0          ...  user-supplied value for ‘sigma^2’ is used
%%%    <=0          ...  ‘sigma^2’ is estimated using (I.29)
%%%
%%% trimming        ...  trimming proportion (default: beta = 0.05)
%%%
%%%   in (0,1/2)    ...  trimmed statistic (I.18) or (I.28) is used
```

```
%%%   otherwise    ...  non-trimmed statistic (I.17) or (I.27) is used
%%%
%%% OUTPUT
%%%
%%% Results         :
%%%
%%%    chp_place   ... place of maximum of test statistics
%%%    ts_value    ... value of maximum of test statistics
%%%    sigma2_used ... value of 's_k^2 (sigma^2)' used for calculation
%%%    test_stats  ... values of test statistics (I.17) or (I.27),
%%%                    respectively (I.18) or (I.28)
%%%    sigma2_stats .. values of statistics (I.29)
%%%    trimming_used . value of 'trimming' used for calculation
%%%
%%% EXTERNAL PROCEDURES CALLED    none
%%%
%%% EXAMPLE
%%%
%%%    nile                            <- data are read in
%%%    Results = m3_1_le(nil, 0, 0.05) <- calculation
%%%
%%%    Results.chp_place    = 28      -> change point place
%%%    Results.ts_value     =  8.7143 -> value of test statistic (I.17)
%%%                                      or (I.18)
%%%    Results.sigma2_used  = 16293   -> 'sigma^2' used
%%%    Results.test_stats   = 1.1943, 1.8867, ...
%%%                                   -> values of test statistics
%%%    Results.sigma2_stats = 28503, 27904, ...
%%%                              -> values of estimators of 'sigma^2'
%%%    Results.trimming_used = 0.05
%%%
%%% CONTROL of PARAMETERS

if nargin   == 2, trimming  = 0;   end
if nargin   == 1, sigma2    = 0;   trimming = 0;  end
if trimming  < 0| trimming >= 0.5, trimming = 0;  end
if sigma2    < 0, sigma2    = 0;   end

%%% COMPUTATION

%%% CALCULATION OF NECESSARY CONSTANTS

data = data(:);
n    = length(data);
nm1  = n - 1;
nm2  = n - 2;
sk2  = zeros(nm1,1);
n1   = 1;
```

```
n2   = nm1;
if   trimming > 0
     n1 = round(n*trimming);
     n2 = n - n1;
     if  n1 < 1,   n1 = 1;   end
     if  n2 > nm1, n2 = nm1; end
end
konst1 = sqrt(n ./ ([1:nm1].*[nm1:-1:1]))';

%%% CALCULATION OF TEST STATISTICS

junk  = cumsum(data - mean(data));
stats = abs(konst1 .* junk(1:nm1));

if sigma2 == 0
   for i = 1:nm1
           data1  = data(1:i);
           data2  = data(i+1:n);
           sk2(i) = (sum((data1 - mean(data1)).^2)  + ...
                    sum((data2  - mean(data2)).^2)) / nm2;
   end
  else
   sk2 = sigma2;
end
test_stats = abs(stats./sqrt(sk2));

[ts_value, chp_place] = max(test_stats(n1:n2));
chp_place = chp_place + n1 - 1;

%%% STORING THE RESULTS

Results.chp_place = chp_place;
Results.ts_value  = ts_value;
if sigma2 > 0
     Results.sigma2_used = sigma2;
   else
     Results.sigma2_used = sk2(chp_place);
end
Results.test_stats    = test_stats;
Results.sigma2_stats  = sk2;
Results.trimming_used = trimming;

%%% PLOTTING (RELEVANT FOR MATLAB USERS ONLY)

plot([n1:n2], test_stats(n1:n2))
axis([0 n floor(min(test_stats(n1:n2))) ceil(max(test_stats(n1:n2)))])
xlabel('Index')
if sigma2 > 0  & trimming == 0
```

```
   ylabel('Test statistics (I.17)')
end
if sigma2 > 0 & trimming > 0
   ylabel('Test statistics (I.18)')
end
if sigma2 == 0  & trimming == 0
   ylabel('Test statistics (I.27)')
end
if sigma2 == 0  & trimming > 0
   ylabel('Test statistics (I.28)')
end
title('MODEL (I.16)')
grid on

if sigma2 == 0
   figure
   plot([n1:n2], sk2(n1:n2))
   xlabel('Index')
   ylabel('Values s_k^2 calculated using (I.29)')
   title('MODEL (I.16)')
   grid on
   axis([0 n floor(min(sk2)) ceil(max(sk2))])
end
```

16. Change in mean with unknown starting value, sum-type test statistics

```
function Results = m3_1_be(data, sigma2)

%%% SUBROUTINE m3_1_be.m
%%%
%%% PURPOSE To test the null hypothesis H against the alternative
%%%         A in the model:
```

$$
\begin{aligned}
&H : Y_i = \mu + e_i, && i = 1, \ldots, n,\\
&A : \exists\, \mathsf{m} \in \{1, \ldots, n-1\} \quad \text{such that} &&\\
&\qquad Y_i = \mu + e_i, && i = 1, \ldots, \mathsf{m},\\
&\qquad Y_i = \mu + \delta + e_i, && i = \mathsf{m}+1, \ldots, n,\ \delta \neq 0.
\end{aligned}
$$

We suppose that the variables e_i are iid with the normal distribution $N(0, \sigma^2)$, σ^2 either known or unknown. If σ^2 is known, we apply Bayesian-type statistic (I.19). On the other hand, if σ^2 is unknown or supplied value of σ^2 is ≤ 0, we apply Bayesian-type statistic (I.19) with σ^2 estimated using (I.26).

```
%%% CALLING SEQUENCE   Results = m3_1_be(data, sigma2)
%%%
%%% INPUT
```

```
%%%
%%% data            ...  data series to be tested for the presence
%%%                      of change point
%%%
%%% sigma2          ...  variance of the error term
%%%    > 0          ...  statistic (I.19) is used with user-supplied
%%%                      value for 'sigma^2'
%%%    <=0          ...  statistic (I.19) is used and 'sigma^2'
%%%    or missing        estimated using (I.26)
%%%
%%% OUTPUT
%%%
%%% Results          :
%%%    test_stat    ...  value of the test statistic (I.19)
%%%    sigma2       ...  user-supplied value for 'sigma^2' or value of
%%%                      the statistic (I.26)
%%%
%%% EXTERNAL PROCEDURES CALLED  none
%%%
%%% EXAMPLE
%%%    nile                          <- data about Nile are read in
%%%    Results = m3_1_be(nil, 0)     <- calculation
%%%
%%%    Results.test_stat = 2.5276     -> value of test statistic (I.19)
%%%    Results.sigma2    = 2.8340e+04 -> value of statistic (I.26)
%%%
%%% CONTROL OF PARAMETERS

if nargin == 1, sigma2 = 0; end

%%% CALCULATION OF NECESSARY CONSTANTS

data = data(:);
n    = length(data);
nm1  = n - 1;

%%% CALCULATION OF THE TEST STATISTIC

junk = cumsum(data - mean(data));
stat = sum(junk.^2)/(n^2);

if sigma2 <= 0, sigma2 = sum((data - mean(data)).^2)/n; end
test_stat  = stat/sigma2;

%%% STORING THE RESULTS

Results.test_stat = test_stat;
Results.sigma2    = sigma2;
```

17. Change in variance

```
function   Results = m3_2(data, trimming, a)

%%% SUBROUTINE m3_2.m
%%%
%%% PURPOSE
```

We suppose that the observations $Y_1, \dots, Y_n$ are independent normally distributed with a known mean μ and unknown variances. Supposing that the mean remains the same, the problem of the detection of a change in variance can be formulated as the following testing problem, i.e. we test the null hypothesis H against the alternative A:

$$\begin{aligned} &H : Y_1 \dots, Y_n \sim N\big(\mu, \sigma^2\big) \\ &A : \exists\, \mathsf{m} \in \{1, \dots, n-1\} \text{ such that} \\ &\qquad Y_1, \dots, Y_{\mathsf{m}} \quad\ \sim N\big(\mu, \sigma_1^2\big), \\ &\qquad Y_{\mathsf{m}+1}, \dots, Y_n \sim N\big(\mu, \sigma_2^2\big), \end{aligned} \tag{I.32}$$

where $\sigma_1^2 \neq \sigma_2^2$.

```
%%% CALLING SEQUENCE   Results = m3_2(data, trimming, a)
%%%
%%% INPUT
%%%
%%% data            ...  data series to be tested for the presence of
%%%                      change point in variance in model (I.32)
%%%
%%% trimming        ...  trimming proportion (default: beta = 0.05)
%%%
%%%   if in (0,1/2) ..  trimmed statistic (I.33) is used
%%%   otherwise     ...  non-trimmed statistic (I.33) is used
%%%
%%% a               ...  (expected) mean of observations
%%%
%%%  specified      ...  user-supplied value of 'a'
%%%  not specified ...   'a' will be estimated by the mean of the data
%%%
%%% OUTPUT
%%%
%%% Results              :
%%%
%%%    chp_place        ...  place of the maximum of test statistics
%%%    ts_value         ...  value of the maximum of test statistics
%%%    stats_zk         ...  values of statistics $Z_k$ from (I.34)
%%%    trimming_used    ...  value of 'trimming' used for calculation
%%%    a_used           ...  value of 'a' used for calculation
%%%
%%% EXTERNAL PROCEDURES CALLED    none
```

```
%%% EXAMPLE
%%%    randn('seed',1);                     <- seeding RNG
%%%    data = [randn(20,1); 3*randn(30,1)]  <- data are generated
%%%    Results = m3_2(data, 0.05, 0)        <- calculation
%%%
%%%    Results.chp_place     = 20     -> change point place
%%%    Results.ts_value      = 4.1267 -> value of test statistic (I.17)
%%%                                      or (I.18)
%%%    Results.stats_zk      = 0.9174, 1.6637, ...
%%%                             -> values of statistics Z_k
%%%    Results.a_used        = 0
%%%    Results.trimming_used = 0.9174, 1.6637, ...
%%%
%%% CONTROL OF PARAMETERS

if nargin == 1,   trimming = 0; a = mean(data);  end
if nargin == 2,   a = mean(data);   end
if trimming < 0 | trimming >= 0.5, trimming = 0; end

%%% CALCULATION OF NECESSARY CONSTANTS

data = data(:);
n    = length(data);
nm1  = n - 1;
zk2  = zeros(n,1);
n1   = 1;
n2   = nm1;
if   trimming > 0
     n1 = round(n*trimming);
     n2 = n - n1;
     if  n1 < 1,   n1 = 1;   end
     if  n2 > nm1, n2 = nm1; end
end

%%% CALCULATION OF TEST STATISTICS

junk  = n*log(sum((data - a).^2)/n);
for k = n1:n2
        data1  = data(1:k);
        data2  = data(k+1:n);
        junk1  = sum((data1 - a).^2)/k;
        junk2  = sum((data2 - a).^2)/(n-k);
        zk2(k) = junk - k*log(junk1) - (n-k)*log(junk2);
end
stats_zk = sqrt(zk2);

[ts_value, chp_place] = max(stats_zk(n1:n2));
chp_place = chp_place + n1 - 1;
```

```
%%% STORING THE RESULTS

Results.chp_place = chp_place;
Results.ts_value  = ts_value;
Results.stats_zk  = stats_zk;
Results.a_used    = a;
Results.trimming_used = trimming;

%%% PLOTTING (RELEVANT FOR MATLAB USERS ONLY)

plot([n1:n2],stats_zk(n1:n2));
axis([0 n floor(min(stats_zk(n1:n2))) ceil(max(stats_zk))])
xlabel('Index')
ylabel('Test statistics (I.33)')
title('MODEL (I.32)')
grid on
```

18. Change in mean and/or variance

```
function   Results = m3_3(data, trimming)

%%% SUBROUTINE m3_3.m
%%%
%%% PURPOSE
```

Sometimes it can happen that the change may occur either in one parameter or in both (simultaneously) . Then we test the null hypothesis H against the alternative A:

$$\begin{aligned} &H : Y_1 \dots, Y_n \sim N\big(\mu, \sigma^2\big) \\ &A : \exists\, \mathsf{m} \in \{2, \dots, n-2\} \text{ such that} \\ &\qquad Y_1, \dots, Y_{\mathsf{m}} \quad \sim N\big(\mu_1, \sigma_1^2\big), \\ &\qquad Y_{\mathsf{m}+1}, \dots, Y_n \sim N\big(\mu_2, \sigma_2^2\big), \end{aligned} \tag{I.38}$$

where $\big(\mu_1, \sigma_1^2\big) \neq \big(\mu_2, \sigma_2^2\big)$.

```
%%% CALLING SEQUENCE   Results = m3_3(data, trimming)
%%%
%%% INPUT
%%%
%%% data            ...  data series to be tested for the presence of
%%%                      change point in variance in model (I.38)
%%%
%%% trimming        ...  trimming proportion (default: beta = 0.05)
%%%
%%%   if in (0,1/2) ..  trimmed statistic (I.39) is used
%%%   otherwise     ...  non-trimmed statistic (I.39) is used
```

```
%%%
%%% OUTPUT
%%%
%%% Results              :
%%%
%%%    chp_place       ...  place of the maximum of test statistics
%%%    ts_value        ...  value of the maximum of test statistics
%%%    stats_zk        ...  values of statistics sqrt of (I.41)
%%%    trimming_used   ...  value of 'trimming' used for calculation
%%%    a1_used         ...  value of 'a1' used for calculation
%%%    a2_used         ...  value of 'a2' used for calculation
%%%
%%% EXTERNAL PROCEDURES CALLED    none
%%%
%%% EXAMPLE
%%%    randn('seed',1);                       <- seeding RNG
%%%    data = [randn(20,1); 4+3*randn(30,1)]  <- data are generated
%%%
%%%    Results = m3_3(data, 0.05)             <- calculation
%%%
%%%    Results.chp_place     = 21     -> change point place
%%%    Results.ts_value      =  6.639 -> value of test statistic (I.39)
%%%    Results.stats_zk      =  0, 2.212, ...
%%%                                   -> values of statistics Z_k
%%%    Results.trimming_used =  0.05
%%%    Results.a1_used       = -0.0321
%%%    Results.a2_used       =  3.5931
%%%
%%% CONTROL OF PARAMETERS

if nargin == 1,   trimming = 0; end
if trimming < 0 | trimming >= 0.5, trimming = 0; end

%%% CALCULATION OF NECESSARY CONSTANTS

data = data(:);
n    = length(data);
zk2  = zeros(n,1);
n1   = 2;
n2   = n - 2;
nm2  = n - 2;
if   trimming > 0
     n1   = round(n*trimming);
     n2   = n - n1;
     if  n1 < 2,   n1 = 2;   end
     if  n2 > nm2, n2 = nm2; end
end
```

```
%%% CALCULATION OF TEST STATISTICS

junk = n*log(sum((data - mean(data)).^2)/n);
for k = n1:n2
        data1  = data(1:k);
        data2  = data(k+1:n);
        junk1  = sum((data1 - mean(data1)).^2)/k;
        junk2  = sum((data2 - mean(data2)).^2)/(n-k);
        zk2(k) = junk - k*log(junk1) - (n-k)*log(junk2);
end
stats_zk = sqrt(zk2);

[ts_value, chp_place] = max(stats_zk(n1:n2));
chp_place = chp_place + n1 - 1;

%%% STORING THE RESULTS

Results.chp_place = chp_place;
Results.ts_value  = ts_value;
Results.stats_zk  = stats_zk;
Results.trimming_used = trimming;
Results.a1_used    = mean(data(1:chp_place));
Results.a2_used    = mean(data(chp_place+1:n));

%%% PLOTTING (RELEVANT FOR MATLAB USERS ONLY)

plot([n1:n2],stats_zk(n1:n2));
axis([0 n floor(min(stats_zk(n1:n2))) ceil(max(stats_zk))])
xlabel('Index')
ylabel('Test statistics (I.39)')
title('MODEL (I.38)')
grid on
```

References

Selected books and proceedings related to the problem

Basseville M. and Benveniste A. (eds.) (1986). *Detection of Abrupt Changes in Signals and Dynamic Systems*. Springer Verlag, Heidelberg.

Basseville M. and Nikiforov I. V. (1993). *Detection of Abrupt Changes. Theory and Applications*. Prentice Hall, London.

Broemling L. and Tsurumi H. (1987). *Econometrics and Structural Change over Time*. M. Dekker, New York.

Brodsky B. S. and Darkhovsky B. E. (1993). *Nonparametric Methods in Change Point Problem*. Kluwer, Dordrecht.

Carlstein E., Müller H. G. and Siegmund D. (eds.) (1994). *Change Point Problem*. IMS Lecture Notes – Monograph Series **23**.

Csörgő M. and Horváth L. (1997). *Limit Theorems in Change Point Analysis*. J. Wiley, New York.

Hackl P. (1980). *Testing the Constancy of Regression Relationship over Time*. Vandenhoeck and Ruprecht, Göttingen.

Hackl P. (ed.) (1989). *Statistical Analysis and Forecasting of Economic Structural Change*. Springer Verlag, New York.

Hackl P. and Westlund A. H. (eds.) (1991). *Economic Structural Changes. Analysis and Forecastings*. IIASA & Springer Verlag, Heidelberg.

Selected articles in journals and proceedings; additional monographs

Albin J. M. P. (1990). *On extremal theory for stationary processes*. Ann. Probab. **18**, 92 – 108.

Albin J. M. P., Carlsson B. and Jarušková D. (2000). *On the asymptotic distribution of a statistic for test of appearance of linear trend*. Extremes (submitted).

Anderson T. W. (1971). *The Statistical Analysis of Time Series*. J. Wiley, New York.

Anderson T. W. and Darling D. A. (1952). *Asymptotic theory of certain goodness of fit criteria based on stochastic processes*. Ann. Math. Statist. **23**, 193 – 212.

Anderson T. W. and Darling D. A. (1954). *A test of goodness of fit*. J. American Statistical Association **49**, 765 – 769.

Antoch J. and Ekblom H. (1993). *Recursive robust regression – computational aspects and comparison*. Computational Statistics and Data Analysis **19**, 115 – 128.

Antoch J. and Hušková M. (1993). *Change point problem*. In: Computational Aspects of Model Choice, Antoch J. (ed.), Physica - Verlag, Heidelberg, 11 – 38.

Antoch J. and Hušková M. (1994). *Procedures for detection of multiple changes in series of independent observations*. In: 5th Prague Symp. on Asymptotic Statistics, Hušková M. and Mandl P. (eds.), Physica - Verlag, Heidelberg, 3 – 20.

Antoch J. and Hušková M. (1996). *Tests and estimators for epidemic alternatives*. In: Probastat'94, Witkovsky V. and Pázman A. (eds.), Tatra Mountains Math. Publ. **7**, 316 – 329.

Antoch J. and Hušková M. (1998). *Estimators of changes*. In: Nonparametrics, Asymptotics and Time Series, Ghosh S. (ed.), M. Dekker, New York, 533 – 578.

Antoch J. and Hušková M. (2000). *Bayesian type estimators for change point*. J. Stat. Plan. Infer. **91**, 195 – 208.

Antoch J. and Hušková M. (2000). *Bayesian like R- and M-estimators of change points*. Discussiones Mathematicae **20**, 115 – 134.

Antoch J. and Hušková M. (2001). *Permutation tests in change point analysis*. Statist. Probab. Letters (to appear).

Antoch J., Hušková M. and Jarušková D. (2000). In: *Lecture Notes of the 5th IASC Summer School*, Lauro N.C. et al eds., ISI, Voorburg, 1 – 75.

Antoch J., Hušková M. and Prášková Z. (1997). *Effect of dependence on statistics for determination of change.* J. Stat. Plan. Infer. **60**, 291 – 310.

Antoch J., Hušková M. and Veraverbeke N. (1995). *Change point problem and bootstrap.* J. Nonparametric Statistics **5**, 123 – 144.

Antoch J. and Víšek J. Á. (1992). *Robust estimation in linear model and its computational aspects.* Computational Aspects of Model Choice, Antoch J. (ed.), Physica – Verlag, Heidelberg, 39 – 104.

Bai J. (1991). *On the partial sums of residuals in autoregressive and moving average models.* J. Time Series Analysis **14**, 247 – 259.

Bai J. (1994). *Least squares estimation of a shift in linear processes.* J. Time Series Analysis **15**, 453 – 472.

Berman S. M. (1982). *Sojourns and extremes of stationary processes.* Ann. Probab. **10**, 1 – 46.

Bhattacharya P. K. and Brockwell P. J. (1976). *The minimum of additive process with applications to signal, estimation and storage theory.* Z. Wahrsch. Verw. Gebiete **37**, 51 – 75.

Billingsley P. (1968). *Convergence of Probability Measures.* J. Wiley, New York.

Bräker H. U. (1993). *High boundary excursions of locally stationary Gaussian processes.* Dissertation Thesis, University of Bern.

Brown R. L., Durbin J. and Evans J. M. (1975). *Techniques for testing the constancy of regression relationships over time (with discussion).* JRSS **B 37**, 149 – 192.

Carlstein E. (1988). *Nonparametric estimation of a change point.* Ann. Statist. **16**, 188 – 197.

Chen X. (1988). *Inference in a simple change point problem.* Scientia Sinica **A 31**, 654 – 667.

Chernoff H. and Zacks S. (1964). *Estimating the current mean of normal distribution which is subjected to changes in time.* Ann. Math. Statist. **35**, 999 – 1018.

Chib J. (1999). *Bayesian estimation and comparison of multiple change point models.* Manuscript.

Chow Y. S. and Teicher H. (1988). *Probability Theory.* Springer Verlag, New York.

Chung-Bow Lee (1995). *Estimation of the number of change points in exponential families distributions.* Preprint.

Chung-Bow Lee (1995). *Estimation of the number of change points in normal random variables.* Preprint.

Csörgő M. and Horváth L. (1988). *Nonparametric methods for the change point problem.* In: Handbook of Statistics **7**, Krishnaiah P. R. and Rao C. R. (eds.), North Holland, Amsterdam, 403 – 425.

Chow Yu. S. and Teicher H. (1988). *Probability Theory.* Springer Verlag, Heidelberg.

Cobb G. W. (1978). *The problem of the Nile. Conditional solution to a change point problem.* Biometrika **65**, 243 – 251.

Cramér H. and Leadbetter M. R. (1967). *Stationary and Related Stochastic Processes.* J. Wiley, New York.

Csörgő M. and Horváth L. (1993). *Weighted Approximations in Probability and Statistics.* J. Wiley, New York.

Darkhovsky B. E. (1976). *A nonparametric method for a posteriori detection of the "disorder" time of a sequence of independent random variables.* Theory Probab. Appl. **21**, 178 – 183.

Darling D. A. and Erdős P. (1956). *A limit theorem for the maximum of normalized sums of independent random variables.* Duke Math. J. **23**, 143 – 155.

Davies R. B. (1987). *Hypothesis testing when a nuisance parameter is present only under the alternative.* Biometrika **74**, 33 – 43.

Davis R. A., Huang D. and Yao Y. C. (1995). *Testing for a change in the parameter values and order of an autoregressive model.* Ann. Statist. **23**, 282 – 304.

De Long D. (1981). *Crossing probabilities for a square root boundary by a Bessel process.* Commun. in Statist. – Theory and Method. **A10**, 2191 – 2213.

Deshayes J. and Picard D. (1986). *Off-line statistical analysis of change point models using nonparametric and likelihood methods.* In: Lecture Notes in Control and Information Sciences **77**, Basseville M. et al. (eds.), Springer Verlag, New York, 103 – 168.

Dümbgen L. (1991). *The asymptotic behavior of some nonparametric change point estimators.* Ann. Statist. **19**, 1471 – 1495.

Ferger D. (1994a). *Change point estimators in case of small disorders.* J. Stat. Plan. Infer. **40**, 33 – 49.

Ferger D. (1994b). *On the rate of almost sure convergence of Dümbgen change point estimators.* Statist. Probab. Letters **25**, 27 – 31.

Ferger D. (1994c). *Asymptotic distribution theory of change point estimators and confidence intervals based on bootstrap approximation.* Math. Meth. Statist. **3**, 362 – 378.

Ferger D. (1995). *Nonparametric tests for nonstandard change point problems.* Ann. Statist. **23**, 1848 – 1861.

Ferger D. and Stute W. (1992). *Convergence of change point estimators.* Stoch. Proc. Appl. **42**, 345 – 351.

Gardner L. A. (1969). *On detecting changes in the mean of normal variates.* Ann. Math. Statist. **40**, 116 – 126.

Gombay E. and Horváth L. (1990). *Asymptotic distributions of maximum likelihood tests for change in the mean.* Biometrika **77**, 411 – 414.

Gombay E. and Horváth L. (1994). *Limit theorems for change in linear regression.* J. Multivariate Analysis **48**, 43 – 69.

Gombay E., Horváth L. and Hušková M. (1996). *Estimators and tests for change in variances.* Statistics & Decisions **14**, 145 – 159.

Gombay E. and Horváth L. (1996). *Approximations for the time of change and the power function in change point models.* J. Stat. Plan. Infer. **52**, 43 – 66.

Gombay E. and Horváth L. (1996). *On the rate of approximations for maximum likelihood tests in change point models.* J. Mult. Anal. **56**, 120 – 152.

Gombay E. and Hušková M. (1998). *Rank based estimators of the change point.* J. Stat. Plan. Infer. **67**, 137 – 154.

Hampel F. R., Ronchetti E. M., Rousseeuw P. and Stahel W. A. (1981). *Robust Statistics. The Approach Based on Influence Functions.* J. Wiley, New York.

Hawkins D. H. (1977). *Testing a sequence of observations for a shift in location.* J. American Statistical Association **72**, 180 – 186.

Hinkley D. V. (1969). *Inference about the intersection in two-phase regression.* Biometrika **56**, 495 – 504.

Hinkley D. V. (1970a). *Inference about the change point in a sequence of random variables.* Biometrika **57**, 1 – 17.

Hinkley D. V. and Hinkley E. A. (1970). *Inference about the change point in a sequence of binomial variables.* Biometrika **57**, 477 – 488.

Hinkley D. and Schechtman E. (1987). *Conditional bootstrap methods in the mean-shift model.* Biometrika **74**, 85 – 93.

Horváth L. (1993). *The maximum likelihood method for testing changes in the parameters of normal observations.* Ann. Statist. **21**, 671 – 680.

Horváth L. (1995). *Detecting changes in linear regression.* Statistics **26**, 189 – 208.

Horváth L. and Gombay E. (1994). *Limit theorems for change in linear regression.* J. Multivariate Analysis **48**, 43 – 69.

Horváth L., Hušková M. and Serbinovska M. (1997). *Estimators for the time of change in linear models.* Statistics **27**, 109 – 130.

Horváth L. and Kokozska P. (1997). *The effect of long-range dependence on change point estimators.* J. Stat. Plan. Infer. **64**, 57 – 81.

Horváth L. and Qi-Man Shao (1996). *Darling-Erdős type theorems for sums of Gaussian variables with long range memory.* Stoch. Proc. Appl. **63**, 117 – 137.

Huber P. J. (1981). *Robust Statistics.* J. Wiley, New York.

Hüsler J. (1990). *Extreme values and high boundary crossings of locally stationary Gaussian processes.* Ann. Probability **18**, 1141 – 1158.

Hüsler J. (1993). *A note on extreme values of locally stationary Gaussian processes.* J. Stat. Plan. Infer. **45**, 203 – 213.

Hušková M. (1994). *Nonrecursive procedures for detecting change in simple linear regression models.* In: New Progress in Probability and Statistics, Puri M. L. and Vilaplana J. P. (eds.), 105 – 126.

Hušková M. (1995). *Estimators for epidemic alternatives.* Comment. Math. Univ. Carolinae **36**, 279 – 291.

Hušková M. and Sen P. K. (1989). *Nonparametric tests for shift and change in regression at an unknown time point.* In: Statistical Analysis and Forecasting of Economic Structural Change, Hackl P. (ed.), Springer Verlag, New York, 71 – 85.

James B., James K. L. and Siegmund D. (1987). *Tests for change points.* Biometrika **74**, 71 – 84.

Jandhyala V. K. and MacNeill I. B. (1990). *Residual partial sum limit process for regression models with applications to detecting parameter changes at unknown times.* Stoch. Proc. Appl. **33**, 309 – 323.

Jandhyala V. K. and MacNeill I. B. (1991). *Tests for parameter changes at unknown times in linear regression models.* J. Stat. Plan. Infer. **27**, 291 – 316.

Jandhyala V. K. and MacNeill I. B. (1992). *On testing for the constancy of regression coefficients under random walk and change point alternatives.* Econometric Theory **8**, 501 – 517.

Jarušková D. (1996). *Change point detection in meteorological measurement.* Monthly Weather Review **124**, 1535 – 1543.

Jarušková D. (1997). *Some problems with application of change point detection methods to environmental data.* Environmetrics **8**, 469 – 483.

Jarušková D. (1998a). *Change point estimator in gradually changing sequences.* Comment. Math. Univ. Carolinae **39**, 551 – 561.

Jarušková D. (1998b). *Testing appearance of linear trend.* J. Stat. Plan. Infer. **70**, 263 – 276.

Jarušková D. (1999). *Testing appearance of polynomial trend.* Extremes **2**, 25 – 37.

Jarušková D. and Antoch J. (1994). *Detection of change in variance.* In: 5th Prague Symposium on Asymptotic Statistics, Mandl P. and Hušková M. (eds.), Physica-Verlag, Heidelberg, 297 – 301.

Jones P. D., Wigley M. L. and Briffa K. R. (1994). *Global and hemispheric temperature anomalies - land and marine instrumental records.* In: Trends'93: A Compendium of Data on Global Change, Boden T. A. et al. (eds.), ORNL/CDIAC - 65, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn., USA, 603 – 608.

Kander Z. and Zacks S. (1966). *Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points.* Ann. Math. Statist. **37**, 1196 – 1210.

Kiefer J. (1960). *K-sample analogues of the Kolmogorov-Smirnov's and Cramér-von Mises tests.* Ann. Math. Statist. **30**, 420 – 447.

Kim Hyune-Ju (1995). *Detection of a change in longitudinal data.* In: Int. Conference on Statistical Methods and Statistical Computing for Quality and Productivity Improvement, 748 – 754.

Kim Hyune-Ju and Siegmund D. (1989). *The likelihood ratio test for a change point in simple linear regression.* Biometrika **76**, 409 – 423.

Knowles M., Siegmund D. and Zhang H. (1991). *Confidence regions in semilinear regression.* Biometrika **78**, 15 – 31.

Leadbetter M. R., Lindgren G. and Rootzén H. (1983). *Extremes and related properties of random sequences and processes.* Springer Verlag, Heidelberg.

MacNeill I. B. (1974). *Tests for change of parameter at unknown time and distribution of some related functionals of Brownian motion.* Ann. Statist. **2**, 950 – 962.

MacNeill I. B. (1978). *Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times.* Ann. Statist. **6**, 422 – 433.

Michels P. and Trenkler G. (1990). *Testing the stability of regression coefficients using generalized recursive residuals.* Austr. J. Statist. **22**, 293 – 312.

Nicholson S. E. (1994). *Century-scale series of standardised annual departure of African rainfall.* In: Trends'93: A Compendium of Data on Global Change, Boden T. A. et al. (eds.), ORNL/CDIAC - 65, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn., USA, 952 – 962.

Ploberger W. and Krämer W. (1992). *The CUSUM test with OLS residuals.* Econometrica **60**, 217 – 285.

Ritov Y. (1990). *Asymptotic efficient estimation of the change point with unknown distribution.* Ann. Statist. **18**, 1829 – 1839.

Schechtman E. (1983). *A conservative nonparametric distribution-free confidence bound for the shift in the change point problem.* Comment. Statist. **A12**, 2455 – 2464.

Sen P. K. (1982). *Asymptotic theory of some tests for constancy of regression relationships over time.* Statistics **13**, 21 – 31.

Sen P. K. (1984). *Recursive M-tests for constancy of multivariate regression relationships over time.* Sequential Analysis **3**, 191 – 211.

Shepp L. A. (1979). *The joint density of the maximum and its location for a Wiener process with drift.* J. Appl. Probab. **16**, 423 – 427.

Shorack G. R. and Wellner J. A. (1986). *Empirical Processes with Applications to Statistics.* J. Wiley, New York.

Siegmund D. (1986). *Boundary crossing probabilities and statistical applications.* Ann. Statist. **14**, 361 – 404.

Siegmund D. (1988). *Confidence sets in change point problems.* Int. Statist. Review **56**, 31 – 48.

Siegmund D. and Zhang H. (1994). *Confidence region in broken line regression.* In: Change point problems, IMS Lecture Notes – Monograph Series **23**, Carlstein E. et al. (eds.). 292 – 316.

Steinebach J. and Eastwood V. R. (1996). *Extreme value asymptotics for multivariate renewal processes.* J. Multivariate Analysis **56**, 284 – 302.

Stryhn H. (1996). *The location of maximum of asymmetric two-sided Brownian motion with triangular drift.* Statist. Probab. Letters **29**, 279 – 284.

Tang S.M. and Mac Neill I.B. (1993). *The effect of serial correlation on tests for parameter change at unknown time.* Ann. Statist. **21**, 552 – 575.

Vostrikova L. Y. (1981). *Detection of "disorder" in multidimensional random processes.* Sov. Math. Dokl. **24**, 55 – 59.

Wolfe D. A. and Schechtman E. (1984). *Nonparametric procedures for the change point problem.* J. Stat. Plan. Infer. **9**, 389 – 396.

Worsley K. J. (1983). *Testing for a two-phase multiple regression.* Technometrics **25**, 35 – 42.

Worsley K. J. (1986). *Confidence regions and tests for a change point in a sequence of exponential family random variables.* Biometrika **73**, 91 – 104.

Yao Q. (1993). *Tests for change points with epidemic alternatives.* Biometrika **80**, 179 – 191.

Yao Y.-C. (1988). *Estimating number of change points via Schwarz criterion.* Statist. Probab. Letters **6**, 181 – 189.

Yao Y.-C. (1990). *On asymptotic behavior of a class of nonparametric tests.* Statist. Probab. Letters **19**, 173 – 177.

Yao Y.-C. (1987). *Approximating the distribution of maximum likelihood estimate of the change point in a sequence of independent random variables.* Ann. Statist. **15**, 1321 – 1328.

Yao Yi-Ching and Davis R. A. (1986). *The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates.* Sankhya **48**, 339 – 353.

Zacks S. (1983). *Survey of classical and Bayesian approach to the change point problem. Fixed sample and sequential procedures of testing and sequential procedures of testing and estimation.* In: Recent Advances in Statistics, Papers in Honor of Herman Chernoff's Sixties Birthday, Rizvi M. H. (ed.), Academic Press, New York, 245 – 269.