REVSTAT – Statistical Journal
Volume 15, Number 3, July 2017, 355–372

TWO-SAMPLE GRADUAL CHANGE ANALYSIS

Authors: Zdeněk Hlávka
– Department of Statistics, Faculty of Mathematics and Physics, Charles University, Czech Republic
hlavka@karlin.mff.cuni.cz

Marie Hušková
– Department of Statistics, Faculty of Mathematics and Physics, Charles University, Czech Republic
huskova@karlin.mff.cuni.cz

Received: May 2015 Revised: July 2015 Accepted: August 2015

Abstract:

- We propose two-sample gradual change analysis motivated by gender differences observed in a real data set containing jumping speeds of 432 girls and 364 boys aged 6 to 19 years. Looking at this data set from the point of view of change-point analysis is more natural and it leads to more precise estimators than application of standard two-sample t-test in each age group. Apart of establishing the asymptotic distribution of the proposed two-sample change-point estimator, we also investigate its small sample properties in a simulation study.

Key-Words:

- *change point; gradual change; multiple comparison; two-sample test; wild bootstrap.*

AMS Subject Classification:

- 62F10, 62F25, 62F40, 62F03.

1. INTRODUCTION

In Table 5 and Figure 3, we present summary statistics of jumping speeds observed in a sample of 432 girls and 364 boys between 6 and 19 years measured by Leonardo Mechanograph Ground Reaction Force Plate (Šumník *et al.*, 2013). In this data set, one is naturally interested in investigating the location of the unknown *change point*: looking at the p-values of two-sample t-tests calculated for each of the thirteen age categories, it seems that jumping speeds for boys and girls are about the same from 6 to 10 years and boys' jumping speeds are clearly higher from 13 years on.

Unfortunately, applying the two-sample t-test thirteen times cannot be recommended without multiple testing corrections. Therefore, Table 5 contains also the p-values adjusted for multiple comparisons using Bonferroni and Benjamini–Hochberg (BH) method. The conclusions based on these two multiple comparison methods are similar although the Bonferroni method controls the family-wise error rate while the Benjamini–Hochberg method (Benjamini and Hochberg, 1995) controls the false discovery rate. It is interesting that also other standard multiple comparisons methods (Holm, 1979; Hommel, 1988; Hochberg, 1988; Benjamini and Yekutieli, 2001) implemented in the function `p.adjust()` in R (R Core Team, 2015) detect statistically significant differences at the same age category (13 years and above) while statistically significant differences are not detected for the two most "suspicious" un-adjusted p-values (0.061 and 0.047 for 11 and 12 years, respectively).

In Section 2, we study this two-sample testing problem from the point of view of change-point analysis using a simple model of gradual change (Hušková, 1999) so that instead of many independent two-sample t-tests we only estimate a single change-point. In Sections 3 and 4, we investigate the asymptotic properties of the proposed estimators under various assumptions (motivated by the application to the jumping speeds data set) and we show that the wild bootstrap provides both confidence intervals and p-values controlling the overall significance level.

Section 5 contains a small simulation study to check the behavior for finite sample situations. The jumping speeds data set is analyzed in Section 6 and we will see that the change-point approach detects statistically significant differences earlier (i.e., for younger children) than the two-sample t-tests. A short summary is given in Section 7.

2. PROCEDURES

We assume that our observations fall into two distinct subgroups that are further split into n distinct ordered categories and that the n_{ji} observations in the j-th subgroup and i-th category are summarized by their sample mean $\overline{Y}_{ji}$ and sample variance $\widehat{\sigma}^2_{ji}$, $j \in \{1,2\}$, $i = 1,...,n$. Under additional assumptions one could naturally apply n independent two-sample t-tests in order to compare the two subgroups within each category and use some of the multiple test procedures as discussed above.

However, we propose another approach based on ideas of the change point analysis. Particularly, motivated by the above data set on the jumping speeds, we introduce a simple *two sample model with gradual changes*:

- (A1) Observations Y_{jik} $(j = 1,2;\ k = 1,...,n_{ji})$ are obtained at time i $(i = 1,...,n)$.
- (A2) All observations are independent.
- (A3) $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = \mu + \delta((i-k_0)/n)_+$ $(i = 1,...,n)$, where μ, δ are unknown parameters and $k_0 = n\theta_0$ for some $\theta_0 \in (0,1)$.
- (A4) $\mathrm{Var}(Y_{jik}) = \sigma^2_{ji} > 0$ $(j = 1,2;\ i = 1,...,n;\ k = 1,...,n_{ji})$.

We use the notation $\overline{Y}_{ji} = \sum_{k=1}^{n_{ji}} Y_{jik}/n_{ji}$, $a_+ = \max(a,0)$ with k_0 denoting the unknown location of the change point, μ the unknown expectation of difference before the change, and δ_n the slope (speed) of the gradual change after k_0. Notice that, generally, variances of the single observations need not be the same.

Assumptions (A1)–(A4) are motivated by the application in Section 6: particularly, in this case, Assumption (A2) is satisfied since we observe only one measurement per subject. In other applications, Assumptions (A2) and (A3) may require some modifications to cover panel (longitudinal) data or time series. Also the trend after the change may not necessarily be linear; more generally, it can be some nondecreasing function strictly increasing after the change point.

We propose to estimate the unknown parameters by the least squares method. In the following, we deal separately with the homoscedastic case (Section 3) and the heteroscedastic case (Section 4).

3. HOMOSCEDASTIC CASE

Here we deal with a *two sample homoscedastic model with gradual changes* assuming additionally:

(A4*) $\operatorname{Var}(\overline{Y}_{1i} - \overline{Y}_{2i}) = \sigma^2/m \ \ (i = 1, ..., n)$, where $\sigma^2 > 0$ is an unknown parameter and m can depend on n.

One-sample homoscedastic models with various gradual changes were studied by a number of authors, e.g., Hinkley (1971); Feder (1975); Shaban (1980); Jarušková (1998); Hušková (1999); Hušková and Steinebach (2000, 2002). They constructed procedures for testing the null hypothesis *no change* versus the alternative *there is a change*, derived the least squares estimators, and studied its limit behavior for $n \to \infty$. We use the same method for our problem.

The least squares estimators $\widehat{\mu}$, $\widehat{\delta}$, $\widehat{k}_\mu$ are defined as minimizers of the sum of squares $\sum_{i=1}^n \{\overline{Y}_{1i} - \overline{Y}_{2i} - a - d((i-k)/n)_+\}^2$ with respect to a, d, k. Denoting $x_{ik} = ((i-k)/n)_+$ and $\bar{x}_k = \sum_{i=1}^n x_{ik}/n$, direct calculations give:

$$\begin{aligned}
\widehat{k}_\mu &= \arg\max_{k\in(1,n)} \left[\frac{\{\sum_{i=1}^n (x_{ik} - \overline{x}_k)\,(\overline{Y}_{1i} - \overline{Y}_{2i})\}^2}{\sum_{i=1}^n (x_{ik} - \overline{x}_k)^2} \right], \\
\widehat{\delta}_\mu &= \frac{\sum_{i=1}^n (x_{i\widehat{k}} - \overline{x}_{\widehat{k}})\,(\overline{Y}_{1i} - \overline{Y}_{2i})}{\sum_{i=1}^n (x_{i\widehat{k}} - \overline{x}_{\widehat{k}})^2}, \\
\widehat{\mu} &= \frac{1}{n}\sum_{i=1}^n (\overline{Y}_{1i} - \overline{Y}_{2i}) - \widehat{\delta}_\mu \overline{x}_{\widehat{k}}\,.
\end{aligned} \tag{3.1}$$

Assuming additionally that $\mu = 0$, the least squares estimators are:

$$\begin{aligned}
\widehat{k}_0 &= \arg\max_{k\in(1,n)} \left[\frac{\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})\}^2}{\sum_{i=1}^n x_{ik}^2} \right], \\
\widehat{\delta}_0 &= \frac{\sum_{i=1}^n x_{i\widehat{k}}(\overline{Y}_{1i} - \overline{Y}_{2i})}{\sum_{i=1}^n x_{i\widehat{k}}^2}\,.
\end{aligned} \tag{3.2}$$

Unfortunately, there are no explicit expressions for $\widehat{k}_\mu$ and $\widehat{k}_0$ and these estimators have to be found as a solution of an optimization problem. The properties of these estimators can be studied either through asymptotics (if n is large enough) or through a simulation study. We start with asymptotics and simulations are presented in Section 5.

Following the proofs in Jarušková (1998) and Hušková (1998, 1999) we get that in our homosecastic setup ((A1)–(A3) and (A4*)) for $n \to \infty$

$$(nm)^{1/2} \frac{\delta}{\sigma} \left\{ \frac{\theta_0(1-\theta_0)}{1+3\theta_0} \right\}^{1/2} \frac{\widehat{k}_\mu - k_0}{n} \xrightarrow{\mathcal{D}} N(0,1)$$

and

$$(nm)^{1/2} \frac{(1-\theta_0)^{3/2}}{\sigma} \left(\frac{1+3\theta_0}{12} \right)^{1/2} (\widehat{\delta}_\mu - \delta) \xrightarrow{\mathcal{D}} N(0,1) ,$$

where $N(0,1)$ denotes the standard normal distribution and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution. Both assertions hold true both for m fixed and $m \to \infty$ together with $n \to \infty$. The limit properties remain true even if δ depends on n and tends to 0 for $n \to \infty$ but no faster than $n^{-1/2} \log\log n$. The above results also imply consistency:

$$(nm)^{1/2}\, \delta(\widehat{k}_\mu - k_0)/n = O_P(1) \qquad \text{and} \qquad (nm)^{1/2}\, (\widehat{\delta}_\mu - \delta) = O_P(1) \,.$$

Quite analogously when $\mu = 0$ we get that the limit distributions of

$$(nm)^{1/2} \frac{\delta}{\sigma} \left(\frac{1-\theta_0}{4} \right)^{1/2} \frac{\widehat{k}_0 - k_0}{n} \qquad \text{and} \qquad (nm)^{1/2} \frac{(1-\theta_0)^{3/2}}{3^{1/2}\,\sigma} (\widehat{\delta}_0 - \delta)$$

are standard normal $N(0,1)$.

In Figure 1, the asymptotic distributions of $\widehat{k}_\mu$ and $\widehat{k}_0$ for nine distinct values of k_0 are compared to histograms obtained by 1000 Monte Carlo simulations. Very good approximations via the limit distribution are evident for $k_0 \leq 15$ and, as expected, they are slightly worse but still acceptable for $k_0 > 15$. The assumption $\mu = 0$ visibly improves the precision of $\widehat{k}_0$ for smaller values of k_0.

In case the trend in the means is not linear after the change (as in (A3)) but nondecreasing with strict monotonicity after the change point, the proposed change point estimators may be biased (Hušková and Steinebach, 2002).

Under the assumption of homoscedasticity, we may combine the estimators $\widehat{\sigma}^2_{ji}$ observed in each category into the standard pooled estimator $\widehat{\sigma}^2_{\text{pooled}}$ of the variance σ^2. The assumption that $\text{Var}(\overline{Y}_{1i} - \overline{Y}_{2i})$ does not depend on i is rather restrictive and a more general case of variances will be studied in the next section.

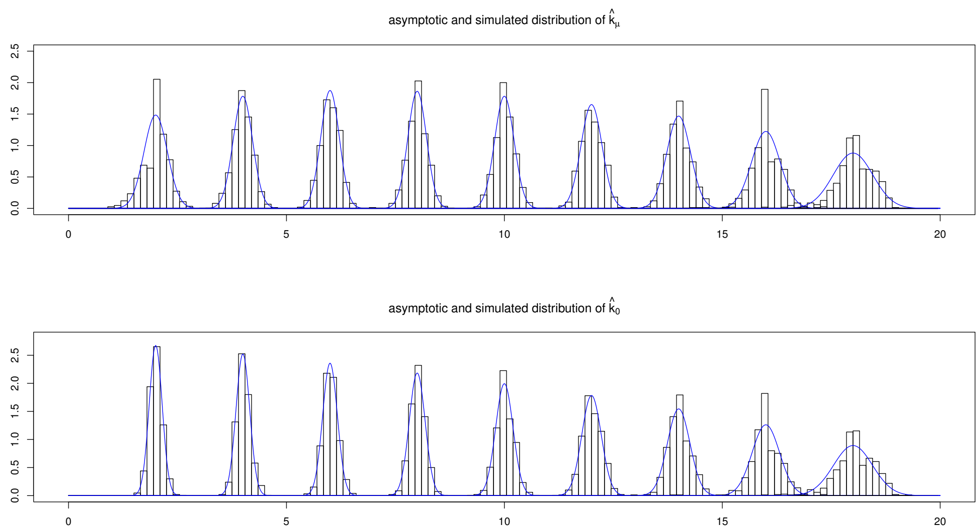

Figure 1: Densities of asymptotic distributions and histograms of 1000 simulated values of $\widehat{k}_\mu$ (upper plot) and $\widehat{k}_0$ (lower plot) in the homoscedastic case for $n = 20$, $\sigma^2 = 1$, $m = 20$, $\mu = 0$, $\delta = 1$, and $k_0 \in \{2, 4, ..., 18\}$.

4. HETEROSCEDASTIC CASE

4.1. Change-point estimators

Let us assume (A1)–(A4) with $\mu = 0$. We may still use the estimators introduced in the previous section: they still have the same limit distributions but with different standardizations. Denoting by $\tau_i^2 = \mathrm{Var}(\overline{Y}_{1i} - \overline{Y}_{2i}) = \sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i}$, we define the estimator $\widehat{k}_0(\tau^2)$ taking also the heteroscedasticity into account:

$$\widehat{k}_0(\tau^2) = \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})/\tau_i^2\right\}^2}{\sum_{i=1}^n x_{ik}^2/\tau_i^2} \right] .$$

In practice, the unknown true variances τ_i^2 are replaced by $\widehat{\tau}_i^2 = \widehat{\sigma}_{1i}^2/m_{1i} + \widehat{\sigma}_{2i}^2/m_{2i}$ leading to the change-point estimator:

$$(4.1) \quad \widehat{k}_0(\widehat{\tau}^2) = \arg\max_{k\in(1,n)} \left[\frac{\left\{\sum_{i=1}^n x_{ik}(\overline{Y}_{1i} - \overline{Y}_{2i})/\widehat{\tau}_i^2\right\}^2}{\sum_{i=1}^n x_{ik}^2/\widehat{\tau}_i^2} \right] = \arg\max_{k\in(1,n)} T_{2,\widehat{\tau}^2}(k) .$$

Concerning properties of these estimators under Assumptions (A1)–(A4) with the additional assumption

$$(4.2) \qquad \tau_-^2/n \le \tau_i^2 \le \tau_+^2/n , \qquad i = 1, ..., n ,$$

for some $0 < \tau_-^2 \leq \tau_+^2 < \infty$, the asymptotic distribution remains normal with zero mean but the asymptotic variance has a more complicated structure and we do not give here explicit formulas. This can again be proved along the lines of the proofs in Hušková (1999). To get approximation for the distribution of the estimator $\widehat{k}_0(\widehat{\tau}^2)$, a proper version of the wild bootstrap provides a reasonable approximation. The algorithm is described below.

4.2. Bootstrap approximation for the distribution of $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$

For simplicity, we will write $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$. Under Assumptions (A1)–(A4) and (4.2), the observed sample mean differences $D_i = \overline{Y}_{1i} - \overline{Y}_{2i}$ have zero mean and standard deviation $\tau_i = (\sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i})^{1/2}$. The distribution of $\widehat{k} = \widehat{k}_0(\widehat{\tau}^2)$ can be approximated by the wild bootstrap (Shao and Tu, 1995):

Algorithm 1. Bootstrap algorithm

Estimate parameters δ and k_0.
Calculate *fitted values* $\widehat{D}_i = \widehat{\delta}_0((i - \widehat{k})/n)_+$ $(i = 1, ..., n)$.
For $b = 1$ to $b = B$
  Generate $D_i^* = \widehat{D}_i + \widehat{\tau}_i \varepsilon_i^*$ $(i = 1, ..., n)$, where $\varepsilon_i^* \sim N(0, 1)$ are independent.
  Calculate the change-point estimator $\widehat{k}_b^*$ from the bootstrap sample $D_1^*, ..., D_n^*$.
Calculate the empirical quantile q_α^* from $\widehat{k}_1^* - \widehat{k}, ..., \widehat{k}_B^* - \widehat{k}$ for prechosen $\alpha \in (0, 1)$.

The empirical bootstrap quantiles q_α^* provide approximations for the true quantiles q_α of $\widehat{k} - k_0$, particularly it can be proved:

$$1 - \alpha \;=\; P(\widehat{k} - k_0 > q_\alpha) \;=\; P(k_0 < \widehat{k} - q_\alpha) \;=\; P(k_0 < \widehat{k} - q_\alpha^*) + o_P(1)$$

and, therefore, $\widehat{k} - q_\alpha^*$ can be used as an upper bound of an asymptotic one-sided $(1 - \alpha)$ confidence interval for k_0.

Remark 4.1. As a complementary problem, we can test hypotheses concerning the change-point location, i.e., the null hypothesis $H_0 \colon k_0 \geq k_1$ against $H_1 \colon k_0 < k_1$ for some given k_1. Denoting by K a random variable with the same distribution as $\widehat{k} - k_0$ and defining the p-value as $P(K < \widehat{k} - k_1)$ (we reject H_0 for small values of $\widehat{k}$), we obtain that, for large B, $\sum_{b=1}^{B} I(\widehat{k}_b^* - \widehat{k} < \widehat{k} - k_1)/B$ is a reasonable approximation of the p-value.

Remark 4.2. The null hypothesis of no-change can easily be tested by bootstrapping the test statistic $T_{2,\widehat{\tau}^2}(k)$ because, under the null hypothesis of no-change, we can easily generate the bootstrap samples $D_i^* = \widehat{\tau}_i \varepsilon_i^*$.

5. SIMULATIONS

5.1. Setup of the simulation study

In this section, we investigate small sample properties of the proposed asymptotic tests and confidence intervals in various setups. We consider the model of gradual change (A3). In each step of the simulation we proceed as follows:

Algorithm 2. Simulation study

Set n and the change-point $\theta_0 = k_0/n$.
Set variances σ_{1i}^2 and σ_{2i}^2 and numbers of observations n_{1i} and n_{2i} $(i = 1, ..., n)$.
Calculate variances $\tau_i^2 = \sigma_{1i}^2/n_{1i} + \sigma_{2i}^2/n_{2i}$ $(i = 1, ..., n)$.
For $s = 1$ to $s = S$
  For $i = 1$ to $i = n$
    Generate $D_i = \overline{Y}_{1i} - \overline{Y}_{2i}$ from $N((i - k_0)_+, \tau_i^2)$.
    Generate $\widehat{\tau}_i^2$ from $\sigma_{1i}^2 \chi_{n_{1i}-1}^2/\{n_{1i}(n_{1i} - 1)\} + \sigma_{2i}^2 \chi_{n_{2i}-1}^2/\{n_{2i}(n_{2i} - 1)\}$.
  Calculate $\widehat{k}_0^{(s)}$ applying one of the change-point estimators $\widehat{k}$ defined by (3.1), (3.2), or (4.1).
  Calculate the 95% confidence interval for k_0 using Algorithm 1.
Calculate the bias and the mean squared error of the simulated $\widehat{k}_0^{(s)}$ $(s = 1, ..., S)$.
Calculate the empirical coverage probability.

Simulations for the homoscedastic case are reported in Section 5.2 while the heteroscedastic case is investigated in Section 5.3. In Section 5.4, we comment on some practical problems caused by rounding effects.

5.2. Homoscedastic case

Under homoscedasticity, we may utilize the asymptotic normality of $\widehat{k}_\mu$ and $\widehat{k}_0$ with σ^2 estimated by the "pooled" estimator $\widehat{\sigma}^2_{\text{pooled}}$.

A pilot simulation study, not presented here, with $n \in \{10, 20\}$ and $m \in \{20, 40\}$, comparing the empirical distributions of $\widehat{k}_\mu$ and $\widehat{k}_0$ suggests that both estimators are generally reasonably good but exhibit large mean squared error and negative bias for k_0 close to n. The mean squared error of $\widehat{k}_\mu$ is larger than the mean squared error of $\widehat{k}_0$ for small values of k_0. This observation corresponds to the asymptotic variances derived in Section 3, see also Figure 1. The coverage probabilities were close to the nominal values unless k_0 was very large (for both estimators) or very small (only for $\widehat{k}_\mu$). The coverage probabilities of the confidence intervals based on σ^2 and its estimator $\widehat{\sigma}^2_{\text{pooled}}$ were very similar.

The worse behavior $\widehat{k}_\mu$ for small k_0 seems to result from the additional uncertainty caused by estimating the parameter μ. This leads to the corrected estimator $\widehat{k}_\mu^{\rm corr} = \widehat{k}_\mu - \widehat{\mu}/\widehat{\delta}_\mu$ that will also be considered in further simulations.

In Table 1, we investigate the empirical coverage probabilities of one-sided 95% bootstrap confidence intervals calculated with and without homoscedasticity assumptions (homoscedasticity assumptions are applicable only because the number of observations in each category is constant). Under homoscedasticity assumptions, we estimate the common variance by $\widehat{\sigma}^2_{\rm pooled}$ and this variance estimator is also used in the bootstrap. More generally, we can also proceed without assuming homoscedasticity and follow Algorithm 1 from Section 4.2 using all $2n$ sample variances $\widehat{\sigma}^2_{ji}$.

Results in Table 1 confirm that coverage probabilities are rather small if the change occurs close to n. The heteroscedastic version works well even in the homoscedastic setup.

Table 1: Coverage probabilities (in %) of one-sided 95% confidence intervals of four change point estimators in the homoscedastic case (1000 simulations, $B = 1000$). The confidence intervals are based on bootstrapping utilizing either the pooled variance estimator $\widehat{\sigma}^2_{\rm pooled}$ (homoscedastic version) or $2n$ sample variances $\widehat{\sigma}^2_{ji}$ (heteroscedastic version).

		θ_0	$\widehat{\sigma}^2_{\rm pooled}$				$\widehat{\sigma}^2_{ji}$			
			$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\rm corr}$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\rm corr}$	$\widehat{k}_0(\widehat{\tau}^2)$
$n = 10$	$n_{ji} = 10$	0.1	88.8	95.2	93.3	94.0	89.5	93.9	93.7	94.5
		0.2	92.4	95.9	92.8	94.9	91.7	94.8	94.3	95.6
		0.4	94.1	92.3	92.9	91.9	95.5	92.4	93.3	92.0
		0.6	92.8	93.2	92.6	92.4	93.9	89.9	92.2	90.2
		0.8	90.4	90.5	90.0	90.8	89.6	87.7	89.2	89.1
		0.9	78.1	78.3	79.2	78.4	80.1	74.8	76.4	76.0
	$n_{ji} = 20$	0.1	93.9	92.0	94.4	92.1	95.1	92.8	94.6	93.0
		0.2	96.3	92.8	95.6	92.4	95.4	92.7	95.8	94.7
		0.4	93.5	92.0	92.3	91.1	92.7	91.6	92.0	90.8
		0.6	87.6	90.1	90.1	89.8	89.3	87.1	88.9	88.4
		0.8	88.8	89.0	89.7	87.8	89.9	86.9	87.2	88.4
		0.9	72.1	70.4	74.9	70.1	72.4	70.9	72.6	70.3
$n = 20$	$n_{ji} = 10$	0.1	96.5	94.3	94.0	94.9	94.9	93.5	95.8	93.1
		0.2	97.1	94.1	95.0	93.7	96.9	93.0	95.0	93.6
		0.4	94.0	93.7	93.8	93.9	94.4	92.1	93.0	92.8
		0.6	93.2	90.9	92.7	92.7	91.9	92.1	91.7	91.8
		0.8	94.8	95.6	94.3	95.3	93.8	94.5	92.5	93.7
		0.9	84.1	84.3	84.8	84.0	83.1	81.8	84.4	80.9
	$n_{ji} = 20$	0.1	97.3	95.0	94.4	94.9	97.0	93.5	93.4	95.3
		0.2	95.1	94.3	94.1	94.3	93.5	93.9	94.1	94.0
		0.4	93.0	93.1	93.1	92.9	93.6	93.6	93.1	94.7
		0.6	91.9	90.7	92.8	92.8	91.7	93.6	92.0	91.4
		0.8	93.2	91.9	91.3	90.7	91.8	92.5	91.5	89.3
		0.9	79.5	81.4	83.4	79.3	82.3	80.4	82.4	82.4

5.3. Heteroscedastic case

Real life is typically heteroscedastic and therefore we pay more attention to such situations. In Table 2, we investigate the behaviour of the proposed method in several artificial heteroscedastic situations caused both by different variances and numbers of observations in the observed categories.

Table 2: Coverage percentages (in %) of 95% bootstrap confidence intervals based on 4 change point estimators for $n \in \{10, 20\}$, $n_{ji} \equiv 10$, and $\sigma^2 = 1$ based on 1000 bootstrap replicates and 1000 simulations. The first four columns are obtained from the homoscedastic version of the bootstrap scheme using the pooled variance estimator $\widehat{\sigma}^2_{\text{pooled}}$.

	θ_0	$n = 10$								$n = 20$			
		$\widehat{\sigma}^2_{\text{pooled}}$				$\widehat{\sigma}^2_{ji}$				$\widehat{\sigma}^2_{ji}$			
		$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\text{corr}}$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\text{corr}}$	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_\mu$	$\widehat{k}_0$	$\widehat{k}_\mu^{\text{corr}}$	$\widehat{k}_0(\widehat{\tau}^2)$
H01	0.1	65.1	66.8	58.0	67.9	88.1	92.5	92.5	93.2	97.5	93.1	95.1	93.8
	0.4	69.9	68.9	67.7	70.7	96.3	94.7	95.7	92.2	94.1	93.3	94.4	95.1
	0.8	74.7	71.2	72.2	68.5	88.5	86.5	86.7	88.4	95.9	94.9	95.2	91.4
	0.9	84.1	82.8	77.9	80.5	78.6	77.4	78.8	78.1	77.5	80.4	80.8	78.9
H02	0.1	60.5	63.2	50.6	65.6	83.7	94.2	92.5	93.9	95.3	93.0	95.0	93.1
	0.4	65.9	65.4	63.9	69.6	90.6	88.4	91.0	93.0	93.6	92.5	92.8	93.2
	0.8	69.4	69.5	74.1	72.9	90.4	89.2	91.0	86.3	89.6	88.4	90.7	90.4
	0.9	80.0	78.2	77.2	83.7	76.8	71.9	75.6	75.8	82.6	84.2	82.3	81.9
H10	0.1	90.0	93.9	92.4	94.0	88.8	94.6	93.4	92.8	92.6	94.0	93.1	95.3
	0.4	97.1	99.6	99.7	99.7	92.7	91.5	94.3	91.1	94.6	94.9	94.9	93.3
	0.8	89.0	93.6	93.3	92.2	91.6	92.5	90.7	89.0	95.3	91.4	94.8	90.5
	0.9	94.1	87.5	87.5	68.5	92.4	88.8	86.1	73.4	87.5	86.6	89.1	82.3
H11	0.1	65.6	65.4	55.6	69.9	89.9	93.1	91.8	93.1	90.4	95.5	91.7	94.2
	0.4	71.0	67.2	67.4	70.8	92.4	94.8	93.2	91.9	93.7	92.4	93.1	93.3
	0.8	79.0	78.7	76.1	71.0	92.2	84.3	90.7	89.8	96.8	95.9	96.4	89.3
	0.9	95.5	92.5	84.6	73.1	93.1	90.3	87.5	66.2	88.4	86.8	86.3	80.1
H12	0.1	58.9	63.9	49.9	64.0	88.3	95.6	92.0	94.2	88.8	94.4	92.4	93.1
	0.4	70.3	68.3	65.3	69.5	90.0	87.1	88.6	91.5	94.5	93.0	94.7	92.9
	0.8	79.6	77.5	78.8	72.2	91.1	90.1	92.7	87.9	93.9	88.5	91.3	88.8
	0.9	93.2	88.3	81.7	78.0	91.3	85.2	85.7	74.3	88.6	87.3	90.9	82.2
H20	0.1	80.9	92.4	91.0	99.4	82.1	91.5	88.0	98.5	97.0	97.2	98.9	99.3
	0.4	93.8	94.6	93.2	99.8	93.0	89.6	89.9	93.7	97.4	95.4	96.6	99.2
	0.8	78.0	75.5	79.1	74.1	78.2	77.4	76.2	73.8	93.2	90.6	91.1	94.1
	0.9	90.6	86.2	84.8	78.3	90.0	86.5	83.3	77.5	79.0	78.4	83.7	74.4
H21	0.1	58.8	63.8	48.8	66.9	76.6	88.7	80.7	94.5	87.8	87.4	87.0	93.3
	0.4	60.6	62.8	61.9	65.2	83.9	82.0	83.3	88.4	85.0	85.4	84.2	90.8
	0.8	73.3	70.7	68.0	69.3	79.9	79.4	76.9	79.8	84.7	84.8	84.3	88.1
	0.9	93.3	88.7	81.8	82.5	87.6	85.4	81.3	78.9	81.1	81.9	79.7	77.4
H22	0.1	49.8	58.4	39.4	69.1	61.6	84.4	73.7	94.8	79.0	83.2	80.4	93.0
	0.4	57.1	61.2	56.0	72.5	74.1	68.2	75.6	90.2	78.0	81.9	81.0	93.3
	0.8	72.6	67.5	72.1	67.6	82.6	79.0	74.9	73.5	76.7	75.4	75.5	89.4
	0.9	92.2	86.6	79.2	79.9	90.3	83.7	83.2	82.1	83.9	78.7	81.9	76.8

Here, we consider altogether 8 heteroscedastic situations obtained by considering two simple models for nonconstant variances and two simple models for nonconstant numbers of observations. The simulation setups (H01,...,H22) are summarized in the following table:

	Nr. of observations (n_{ji})		
	$n_{ji} = m$	$m\{1+3I(i \text{ odd})\}/2$	$m\{1+3I(i > n/2)\}/2$
σ_{ji} constant ($\sigma_{ji} = \sigma$)		H01	H02
$\sigma_{ji} = \sigma(1 + 2I(i > k_0))$	H10	H11	H12
$\sigma_{ji} = \sigma(1 + 2I(i \text{ even}))$	H20	H21	H22

As expected, Table 2 shows that bootstrap using the pooled estimator of variance does not lead to reliable results in the heteroscedastic setup. The confidence intervals based on the heteroscedastic estimator $\widehat{k}_0(\widehat{\tau}^2)$ provide reasonable coverage probabilities for all scenarios as long as k_0 is not too close to n.

5.4. Rounding effects

In the jumping speeds example, children aged i to $i+1$ years are included in the i-th age category. We use summary statistics observed in these age categories and we have to keep in mind that the i-th observed sample mean and sample standard deviation correspond to the marginal distribution of jumping speeds for all children aged from i to $i+1$ years.

Assuming that $E(Y_1 \mid \text{Age}{=}x) = E(Y_2 \mid \text{Age}{=}x) + \delta((x - k_0)/n)_+$ for $x \in (1, n+1)$ and that the age distribution in both groups is the same, it follows that $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = 0$, for $i \le \lfloor k_0 \rfloor$, and the true $E(\overline{Y}_{1i} - \overline{Y}_{2i}) = \delta(i - k_0)/n$ for $i \ge \lceil k_0 \rceil$. Hence, for sample means based on categorization of continuous explanatory variable, the model (A3) is valid only if k_0 is a natural number.

Denoting $i_0 = \lfloor k_0 \rfloor$ and $d_0 = k_0 - i_0$, we may calculate the true expectation of the mean differences $E(\overline{Y}_{1i_0} - \overline{Y}_{2i_0}) = E\big\{\delta((X - k_0)/n)_+\big\}$ under the above assumptions (with uniform distribution of the explanatory variable X in the i_0-th age category and for $d_0 > 0$):

$$E(\overline{Y}_{1i_0} - \overline{Y}_{2i_0}) = \frac{\delta}{n} \int_{i_0+d_0}^{i_0+1} (x - k_0)\, dx = \frac{\delta}{n} \int_0^{1-d_0} x\, dx = \frac{\delta(1-d_0)^2}{2n} .$$

In Figure 2, we plot the theoretical expectation for various values of k_0. Obviously, whenever k_0 is not a natural number, the i_0-th sample mean can be "somewhat larger than it should be".

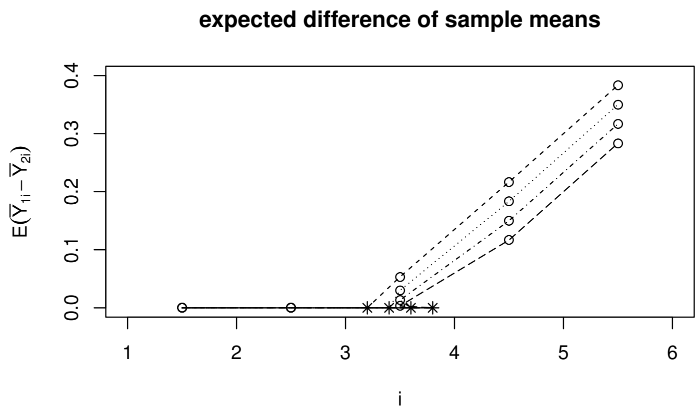

Figure 2: Expectations of mean differences for five categories for change points $k_0 \in (3.2, 3.4, 3.6, 3.8)$. Each line connects the expectations (denoted by circles) corresponding to given changepoint (denoted by star).

In practice, it is more natural to define the i-th category by values of the explanatory variable $x \in (i-0.5, i+0.5)$ and this notation is also in accordance with the theoretical part of this paper. Therefore, we define the bias corrected estimator $\widehat{k}_0^{\mathrm{bc}}$ by using $x_{ik}^{\mathrm{bc}} = I(i > \lceil k-0.5 \rceil)\,(i-k)/n + I(i = \lceil k-0.5 \rceil)\,(\lceil k-0.5 \rceil - k + 0.5)/(2n)$ instead of x_{ik} in (3.2).

Results of a small simulation study comparing the behavior of $\widehat{k}_0$ and $\widehat{k}_0^{\mathrm{bc}}$ in a homoscedastic case are given in Table 3. As expected, the empirical bias of the bias corrected estimator $\widehat{k}_0^{\mathrm{bc}}$ tends to be somewhat smaller. The effect of the rounding bias on the coverage probabilities based on $\widehat{k}_0$ is most clearly visible for $n = 20$, $n_{ji} \equiv 20$, and k_0 lying in the center of the category (i.e., for $k_0 = 14$, 15, and 16).

6. JUMPING SPEEDS

In order to analyze the real data set given in Table 5, it is important to understand the meaning of the row-labels. The various labels and its meanings are summarized in Table 4. In the theoretical part of this paper, we were using the "Index scale" given in the first column. For practical considerations, it is important to notice that $k = 1$ actually corresponds to children aged approximately 6.5 years.

In order to calculate the estimators $\widehat{k}_0(\widehat{\tau}^2)$ and $\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$, we maximize the function $T_{2+\widehat{\tau}^2}(k)$ and its bias corrected version, $T_{2+\widehat{\tau}^2}^{\mathrm{bc}}(k)$, plotted in Figure 3.

Table 3: Empirical mean squared error (MSE), bias and coverage probabilities of 95% confidence intervals (in %) for $\widehat{k}_0$ and $\widehat{k}_0^{\mathrm{bc}}$, 1000 simulations with 1000 bootstrap replicates.

			$\widehat{k}_0$			$\widehat{k}_0^{\mathrm{bc}}$		
		θ_0	MSE	bias	coverage	MSE	bias	coverage
$n=10$	$n_{ji}\equiv 10$	0.20	0.124	0.003	93.6%	0.112	−0.010	94.6%
		0.22	0.113	−0.001	95.4%	0.113	−0.016	95.3%
		0.25	0.125	−0.018	91.5%	0.123	−0.010	92.9%
		0.28	0.122	0.012	91.2%	0.133	0.011	90.3%
		0.30	0.116	0.008	93.9%	0.131	0.001	95.0%
		0.70	0.432	−0.109	92.7%	0.365	−0.041	93.3%
		0.72	0.466	−0.099	94.0%	0.498	−0.065	91.7%
		0.75	0.678	−0.151	89.6%	0.726	−0.138	88.4%
		0.78	0.936	−0.226	86.5%	0.971	−0.123	91.5%
		0.80	1.160	−0.263	91.3%	1.255	−0.224	91.8%
	$n_{ji}\equiv 20$	0.20	0.053	−0.011	94.1%	0.052	0.002	93.2%
		0.22	0.054	−0.013	95.4%	0.050	0.003	98.4%
		0.25	0.053	−0.011	95.8%	0.060	−0.017	95.4%
		0.28	0.059	0.001	92.8%	0.054	−0.009	95.2%
		0.30	0.063	0.000	92.6%	0.060	0.011	92.8%
		0.70	0.177	−0.056	86.9%	0.173	0.004	96.3%
		0.72	0.194	−0.073	94.1%	0.191	−0.024	95.6%
		0.75	0.246	−0.072	94.6%	0.230	−0.052	91.7%
		0.78	0.319	−0.116	88.1%	0.302	−0.034	93.2%
		0.80	0.399	−0.097	88.0%	0.506	−0.092	96.9%
$n=20$	$n_{ji}\equiv 10$	0.20	0.054	−0.005	93.6%	0.050	−0.004	95.7%
		0.22	0.056	−0.005	94.5%	0.057	−0.003	95.3%
		0.25	0.056	−0.016	94.6%	0.050	0.003	95.4%
		0.28	0.061	0.002	94.5%	0.055	−0.009	94.6%
		0.30	0.063	−0.004	93.1%	0.060	−0.012	93.9%
		0.70	0.150	−0.040	93.7%	0.158	0.001	94.7%
		0.72	0.175	−0.035	93.9%	0.163	−0.033	94.5%
		0.75	0.200	−0.051	93.3%	0.178	−0.023	96.7%
		0.78	0.220	−0.043	94.0%	0.229	−0.026	90.8%
		0.80	0.263	−0.050	94.7%	0.276	−0.022	93.8%
	$n_{ji}\equiv 20$	0.20	0.027	−0.006	94.0%	0.026	−0.004	93.3%
		0.22	0.026	−0.008	94.3%	0.024	−0.010	98.0%
		0.25	0.030	−0.006	93.3%	0.029	0.005	93.6%
		0.28	0.026	0.005	95.8%	0.030	0.001	95.9%
		0.30	0.033	−0.015	93.9%	0.031	−0.004	94.1%
		0.70	0.078	−0.024	90.7%	0.072	−0.000	94.1%
		0.72	0.089	−0.032	96.7%	0.075	−0.001	95.9%
		0.75	0.093	−0.037	90.6%	0.095	−0.006	93.4%
		0.78	0.112	−0.046	95.3%	0.103	−0.012	94.3%
		0.80	0.114	−0.040	90.0%	0.111	−0.012	95.7%

Table 4: Meaning of row labels in the jumping speeds example.

Index (k)	Label	Meaning	Interpretation	$\overline{Y_1}$ ($\widehat{\sigma}_1$)	$\overline{Y_2}$ ($\widehat{\sigma}_2$)
1	6	6–7 years	∼ 6.5 years	1.89 (0.17)	1.87 (0.18)
2	7	7–8 years	∼ 7.5 years	2.00 (0.21)	1.98 (0.20)
⋮	⋮	⋮	⋮	⋮	⋮
13	18	18–19 years	∼ 18.5 years	2.33 (0.17)	2.87 (0.10)

In both plots, the estimator $\widehat{k} = 5$ (on the "Index scale") corresponds to the estimated change point $\widehat{k}^{\mathsf{age}} = 10.5$ years.

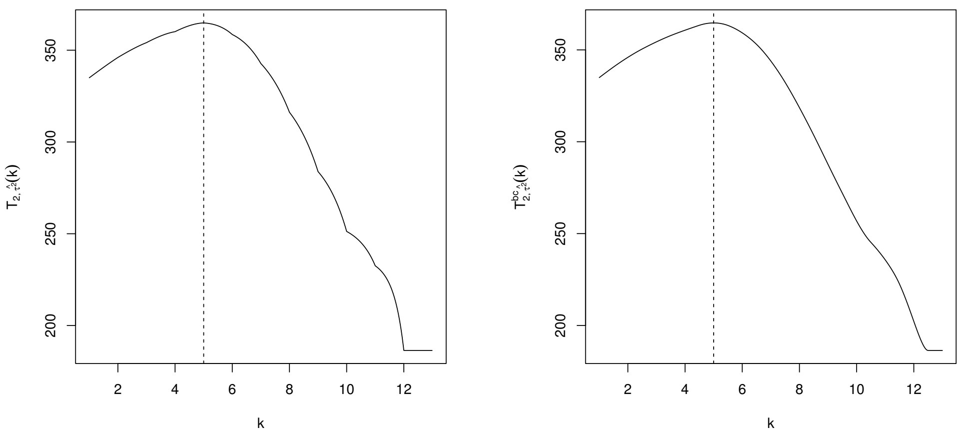

Figure 3: Function $T_{2+\widehat{\tau}^2}(k)$ and its bias corrected version $T^{\mathrm{bc}}_{2+\widehat{\tau}^2}(k)$ for the jumping speed data. The vertical dashed lines denote the estimates $\widehat{k}$ and $\widehat{k}^{\mathrm{bc}}$.

Applying the bootstrap algorithm described in Section 4.2, we obtain that the upper limit of the one-sided 95% confidence interval based on $\widehat{k}_0(\widehat{\tau}^2)$ is $5.72 + 5.5 = 11.22$ years. Applying the bias correction from Section 5.4, we obtain the one-sided 95% confidence interval $(-\infty, 11.26)$ years.

For both estimators, the test of the null hypothesis "no changepoint before 12 years" is actually carried out by testing the index $k_1 = 12 - 5.5$ (see Remark 4.1 and Table 4). The p-values corresponding to the change-point tests of the null hypothesis $H_0\colon k_0 \geq k_1$ against $H_1\colon k_0 < k_1$ for $k_1 \in \{0.5, ..., 12.5\}$ are given in Table 5. Since each test concerns the age $k_1 + 5.5$ years, it seems more natural to shift the lines with these p-values in order to point out the difference between the two-sample t-test (comparing the marginal means in i-th age category, i.e., for approximately $i + 0.5$ years) and the change-point approach (testing whether there is a significant difference for children aged precisely i years).

Table 5: Observed mean jumping speeds and standard deviations for boys and girls in 13 age categories. P-values of the two-sample t-test in each age category, its Bonferroni and Benjamini–Hochberg (BH) adjustments and p-values of the test for change point location based on $\widehat{k}_0(\widehat{\tau}^2)$ and $\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$.

Age cat.	girls		boys		p-values					Age
	$\overline{Y}_1$ $(\widehat{\sigma}_1)$	n_1	$\overline{Y}_2$ $(\widehat{\sigma}_2)$	n_2	t-test	Bonferroni	BH	$\widehat{k}_0(\widehat{\tau}^2)$	$\widehat{k}_0^{\mathrm{bc}}(\widehat{\tau}^2)$	
6–7	1.89 (0.17)	33	1.87 (0.18)	19	0.780	1.000	0.780	1.000	1.000	6
7–8	2.00 (0.21)	43	1.98 (0.20)	38	0.646	1.000	0.763	1.000	1.000	7
8–9	2.01 (0.21)	33	2.06 (0.21)	38	0.369	1.000	0.479	1.000	1.000	8
9–10	2.06 (0.18)	42	2.14 (0.18)	29	0.081.	1.000	0.117	0.999	0.997	9
10–11	2.19 (0.22)	42	2.17 (0.19)	45	0.713	1.000	0.773	0.861	0.846	10
11–12	2.23 (0.15)	30	2.31 (0.23)	37	0.062.	0.800	0.100	0.113	0.117	11
12–13	2.26 (0.13)	41	2.35 (0.23)	40	0.047*	0.615	0.088.	0.003**	0.003**	12
13–14	2.30 (0.22)	32	2.53 (0.21)	36	0.000***	0.001***	0.000***	0.000***	0.000***	13
14–15	2.28 (0.23)	31	2.66 (0.19)	20	0.000***	0.000***	0.000***	0.000***	0.000***	14
15–16	2.37 (0.17)	29	2.72 (0.22)	26	0.000***	0.000***	0.000***	0.000***	0.000***	15
16–17	2.33 (0.19)	17	2.83 (0.28)	9	0.001***	0.006**	0.001**	0.000***	0.000***	16
17–18	2.35 (0.18)	25	2.76 (0.16)	13	0.000***	0.000***	0.000***	0.000***	0.000***	17
18–19	2.33 (0.17)	34	2.87 (0.10)	14	0.000***	0.000***	0.000***	0.000***	0.000***	18

We conclude that the estimated change-point is 10.5 years (with 95% confidence interval $(-\infty, 11.26)$) while the two-sample t-tests *without multiple testing correction* show statistically significant difference only after 12 years (in the age category 12 to 13 years).

In order to verify the validity of Assumption (A3), we plot the data set and the resulting least squares fit in Figure 4.

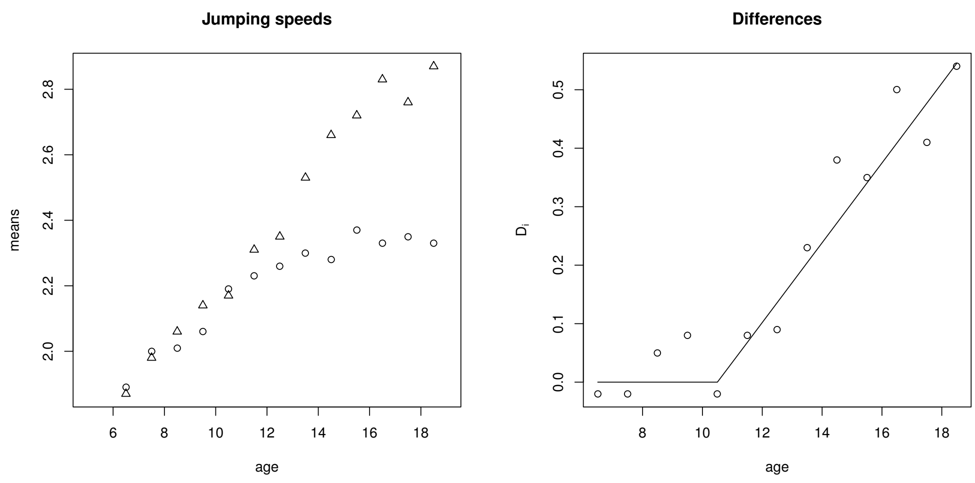

Figure 4: Observed sample means of jumping speed for boys (△) and girls (◯) in thirteen age categories. The right plot shows the observed differences D_i and the least squares fit.

7. SUMMARY AND OUTLOOK

A rigorous approach to multiple hypotheses testing is needed in many real-life situations. Typically, a Bonferroni-type adjustment increases all p-values in order to control either the family-wise error rate or the false discovery rate. However, the structure of the observed data often calls for a more appropriate and powerful solution. Using gender-specific growth curves as a motivation, we proposed a simple two-sample gradual change model in order to develop bootstrap-based tests and confidence intervals for the location of the unknown change-point. In this way, many two-sample t-tests can be replaced by a single test concerning only the change-point. Therefore, adjustments for multiple hypotheses testing become unnecessary.

In practice, the linearity assumption may not be fulfilled. This problem can be solved in a simple way, e.g., by using a finer grid to investigate only a small neighborhood of the suspected change point.

Obviously, the proposed method is applicable also to different sample characteristics. For example, we could investigate a two-sample gradual change in the slope using a table of estimated slopes (and estimates of their standard deviations) in each age category. Such a test would correspond to a model of quadratic change for the original observations.

Depending on further applications, various extensions of the proposed methodology to more general setups may be considered, e.g., dependent observations and more general changes than a linear trend. Also some aspects of nonparametric regression can be utilized if one can analyze the original data set instead of only sample means and sample standard deviations observed in n ordered categories.

ACKNOWLEDGMENTS

The research was supported by GAČR 15-09663S and IAP P7/06 StUDyS network. We thank Zdeněk Šumník (University Hospital in Motol and 2nd Faculty of Medicine, Charles University) for the permission to use the jumping speeds data set. We are grateful to the Editor and an anonymous referee for a number of useful remarks that lead to a considerable improvement in presentation.

REFERENCES

[1] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57**(1), 289–300.

[2] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency, *Ann. Statist.*, **29**(4), 1165–1188.

[3] Feder, P.I. (1975). On asymptotic distribution theory in segmented regression problems–identified case, *Ann. Statist.*, **3**(1), 49–83.

[4] Hinkley, D.V. (1971). Inference in two-phase regression, *J. Amer. Statist. Assoc.*, **66**(336), 736–743.

[5] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance, *Biometrika*, **75**(4), 800–802.

[6] Holm, S. (1979). A simple sequentially rejective multiple test procedure, *Scand. J. Stat.*, **6**(2), 65–70.

[7] Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test, *Biometrika*, **75**(2), 383–386.

[8] Hušková, M. (1998). Estimators in the location model with gradual changes, *Comment. Math. Univ. Carolin.*, **39**(1), 147–157.

[9] Hušková, M. (1999). Gradual changes versus abrupt changes, *J. Statist. Plann. Inference*, **76**(1), 109–125.

[10] Hušková, M. and Steinebach, J. (2000). Limit theorems for a class of tests of gradual changes, *J. Statist. Plann. Inference*, **89**(1–2), 57–77.

[11] Hušková, M. and Steinebach, J. (2002). Asymptotic tests for gradual changes, *Statist. Decisions*, **20**(2), 137–151.

[12] Jarušková, D. (1998). Testing appearance of linear trend, *J. Statist. Plann. Inference*, **70**(2), 263–276.

[13] R Core Team (2015). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

[14] Shaban, S.A. (1980). Change point problem and 2-phase regression: an annotated bibliography, *Int. Stat. Rev.*, **48**(1), 83–93.

[15] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer.

[16] Šumník, Z.; Matysková, J.; Hlávka, Z.; Durdilová, L.; Souček, O. and Zemková, D. (2013). Reference data for jumping mechanography in healthy children and adolescents aged 6–18 years, *J. Musculoskelet. Neuronal. Interact.*, **13**(3), 259–273.