

# A procedure for the change point problem in parametric models based on $\phi$ -divergence test-statistics

Batsidis, A.<sup>1</sup>, Martín, N.<sup>2\*</sup>, Pardo, L.<sup>3</sup> and Zografos, K.<sup>1</sup>

<sup>1</sup>Dept. Mathematics, University of Ioannina, Greece

<sup>2</sup>Dept. Statistics, Carlos III University of Madrid, Spain

<sup>3</sup>Dept. Statistics and O.R., Complutense University of Madrid, Spain

July 6, 2011

## Abstract

This paper studies the change point problem for a general parametric, univariate or multivariate family of distributions. An information theoretic procedure is developed which is based on general divergence measures for testing the hypothesis of the existence of a change. For comparing the accuracy of the new test-statistic a simulation study is performed for the special case of a univariate discrete model. Finally, the procedure proposed in this paper is illustrated through a classical change-point example.

**MSC:** *primary* 62F03; 62F05; *secondary* 62H15

**Keywords:** Change point; Information criterion; Divergence; Wald test-statistic; General distributions.

## 1 Introduction

The change point problem has been considered and studied by several authors the last five decades. Change point analysis is a statistical tool for determining whether a change has taken place at a point of a sequence of observations, such that the observations are described by one distribution up to that point and by another distribution after that point. Change-point analysis concerns with the detection and estimation of the point at which the distribution changes. One change point problem or multiple change points problem have been studied in the literature, depending on whether one or more change points are observed in a sequence of random variables. Several methods, parametric or non-parametric, have been developed to approach the solution of this problem while the range of applications of change point analysis is broad. Applications can be encountered in many areas such as statistical quality control, public health, medicine, finance, biomedical signal processing, meteorology, seismology, etc. The monograph by Chen and Gupta (2000) summarizes recent developments in parametric change-point analysis.

Typical situations encountered in the literature of parametric multiple change points analysis are as follows: Let  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$  be  $K$  independent  $d$ -variate observations ( $d \in \mathbb{N}$ ) and let  $(\mathcal{X}^{(d)}, \beta_{\mathcal{X}}, P_{\theta})_{\theta \in \Theta}$

---

\*Corresponding author, E-mail: nirian.martin@uc3m.es.

the statistical space associated with the random variable (r.v.)  $\mathbf{X}_i$ ,  $i = 1, \dots, K$ . The probability density function with respect to a  $\sigma$ -finite measure  $\mu$  given by  $f_{\boldsymbol{\theta}_i}(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\theta}_i) = \frac{dP_{\boldsymbol{\theta}_i}}{d\mu}$ ,  $\boldsymbol{\theta}_i \in \mathbb{R}^m$ ,  $i = 1, \dots, K$ ,  $\mathbf{x} \in \mathbb{R}^d$ . For simplicity,  $\mu$  is either the Lebesgue measure or a counting measure. We adopt in the sequel the formulation of the multiple change point problem as it appeared in Srivastava and Worsley (1986) and Chen and Gupta (2000, 2004). Based on these authors, suppose that adjacent observations are grouped in  $q$  groups, so that  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_{k_1}$ , are in the first group,  $\mathbf{X}_{k_1+1}, \mathbf{X}_{k_1+2}, \dots, \mathbf{X}_{k_2}$ , are in the second group and we continue in a similar manner until  $\mathbf{X}_{k_{q-1}+1}, \mathbf{X}_{k_{q-1}+2}, \dots, \mathbf{X}_{k_q} = \mathbf{X}_K$  are in the  $q$ -th group.

Consider the model for changes in the parameters. This is formulated as a problem of testing the following hypotheses,

$$H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \dots = \boldsymbol{\theta}_K \quad (= \boldsymbol{\theta}_0, \boldsymbol{\theta}_0 \text{ unknown}), \quad (1)$$

versus the alternative

$$H_1 : \boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_{k_1} \neq \boldsymbol{\theta}_{k_1+1} = \dots = \boldsymbol{\theta}_{k_2} \neq \dots \neq \boldsymbol{\theta}_{k_{q-1}+1} = \dots = \boldsymbol{\theta}_{k_q} = \boldsymbol{\theta}_K,$$

where  $q$ ,  $1 \leq q \leq K$ , is the unknown number of changes and  $k_1, k_2, \dots, k_q$  are the unknown positions of the change points. The above hypotheses can be equivalently stated in the form

$$H_0 : \mathbf{X}_i \text{ are described by } f_{\boldsymbol{\theta}_0}, i = 1, \dots, K \text{ and } \boldsymbol{\theta}_0 \text{ unknown}, \quad (2)$$

versus the alternative

$$H_1 : \mathbf{X}_{k_j+1}, \mathbf{X}_{k_j+2}, \dots, \mathbf{X}_{k_{j+1}}, j = 0, \dots, q - 1 \text{ are described by } f_{\boldsymbol{\theta}_{j+1}},$$

with  $\mathbf{X}_{k_q} = \mathbf{X}_K$ .

There is an extensive bibliography on the subject and several methods to search for the change point problem have appeared in the literature. Among them, the generalized likelihood ratio test, Bayesian solution of the problem, information criterion approaches, cumulative sum method, etc. Based on these methods, several papers discuss the change-point problems in specific probabilistic models, like the univariate and multivariate normal distribution, the gamma model and the exponential model. For instance, Sen and Srivastava (1980) focused on the single change-point problem. Moreover, they consider that within each section, the distributions are the same, while the distribution in a section is different from that in the preceding and the following section in mean vector or covariance matrix. For an exposition of these methods and their application to specific distributions we refer to the monograph or the survey paper by Chen and Gupta (2000, 2001) and the references appeared therein.

It has been proposed in these and other treatments (cf., for instance, Vostrikova (1981)), that in order to study the multiple change point problem, which is formulated by (1) or (2), we just need to test the single change point hypothesis and then to repeat the procedure for each subsequence. Hence, we turn to the testing of (2) against the alternative,

$$H_1 : \mathbf{X}_i \equiv f_{\boldsymbol{\theta}_0}, i = 1, \dots, \kappa \quad \text{and} \quad \mathbf{X}_i \equiv f_{\boldsymbol{\theta}_1}, i = \kappa + 1, \dots, K, \quad (3)$$

where the symbol  $\equiv$  is used to denote that the observations on the left follow the parametric density on the right. In (3),  $\kappa$  represents the position a single change point, which is supposed to be unknown. A general description of this technique in the detection of the changes is summarized in the following

steps by Chen and Gupta (2001). First we test for no change point versus one change point, that is, we test the null hypothesis given by (2) versus the alternative given by (3) and equivalently stated by  $H_1: \boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_\kappa \neq \boldsymbol{\theta}_{\kappa+1} = \dots = \boldsymbol{\theta}_K$ . Here,  $\kappa$  is the unknown location of the single change point. If  $H_0$  is not rejected, then the procedure is finished and there is no change point. If  $H_0$  is rejected, then there is a change point and we continue with the step 2. In the second step we test separately the two subsequences before and after the change point found in the first step for a change. In the sequel, we repeat these two steps until no further subsequences have change points. At the end of the procedure, the collection of change point locations found by the previous steps constitute the set of the change points.

The subject of change point analysis is twofold. On the one hand to detect if there is one or more changes in a sequence of observation. The second aspect of change point analysis is the estimation of the number of changes and their corresponding locations. In this paper we will develop an information theoretic procedure which is based on divergence, in order to study the change point problem. The measures background is a general parametric, univariate or multivariate family of distributions. We describe formally the framework and the problem in Section 2, and the main results are presented in Section 3. In Section 4 we focus our interest on a specific distribution, the binomial distribution and a simulation study is performed in order to compare the accuracy the new test-statistic with some pre-existing test-statistics. In the final Section 5, the general results of this paper are illustrated by means of the well-known Lisdisfarne scribes data set.

## 2 Information theoretic procedure

Consider now the single change point problem, that is the problem of testing the pair of hypotheses

$$H_0 : \mathbf{X}_i \equiv f_{\boldsymbol{\theta}_0}, i = 1, \dots, K \quad (4a)$$

$$H_1 : \mathbf{X}_i \equiv f_{\boldsymbol{\theta}_0}, i = 1, \dots, \kappa \text{ and } \mathbf{X}_i \equiv f_{\boldsymbol{\theta}_1}, i = \kappa + 1, \dots, K, \quad (4b)$$

which are presented by (2) and (3), respectively. In the above formulation,  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_1$  are unknown. Since  $\kappa$  is the unknown location of the single change point, we will consider all the candidate points  $k \in \{1, \dots, K - 1\}$ . Let  $\hat{\boldsymbol{\theta}}_{0,k}^{(K)}$  denotes the maximum likelihood estimator (MLE) of  $\boldsymbol{\theta}_0$  which is based on the random sample  $\mathbf{X}_1, \dots, \mathbf{X}_k$  from  $f_{\boldsymbol{\theta}_0}$  and let  $\hat{\boldsymbol{\theta}}_{1,k}^{(K)}$  denotes the m.l.e. of  $\boldsymbol{\theta}_1$  which is based on the random sample  $\mathbf{X}_{k+1}, \dots, \mathbf{X}_K$  from  $f_{\boldsymbol{\theta}_1}$ . If the hypothesis  $H_1$  is true, then there is a difference between the probabilistic models  $f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}$  and  $f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}$ , which cause a large value for a measure of the distance between  $f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}$  and  $f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}$ . Given that the  $\phi$ -divergence is a broad family of distance measures between probability distributions, the  $\phi$ -divergence between  $f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}$  and  $f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}$  is large if  $H_1$  is true and hence it can be used in order to decide if the candidate point  $k$  in (4b) is a change point ( $\kappa = k$ ). Taking into account that the m.l.e.  $\hat{\boldsymbol{\theta}}_{0,k}^{(K)}$  and  $\hat{\boldsymbol{\theta}}_{1,k}^{(K)}$  of  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_1$ , respectively, depend on the candidate change point  $k$ , we will adopt the following notation for the  $\phi$ -divergence between  $f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}$  and  $f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}$ ,

$$D_\phi^{(k)} = D_\phi^{(k)} \left( f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}, f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}} \right) = \int_{\mathcal{X}^{(d)}} f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}(\mathbf{x}) \phi \left( \frac{f_{\hat{\boldsymbol{\theta}}_{0,k}^{(K)}}(\mathbf{x})}{f_{\hat{\boldsymbol{\theta}}_{1,k}^{(K)}}(\mathbf{x})} \right) d\mu(\mathbf{x}), \quad (5)$$

provided that the convex function  $\phi$  satisfies some additional conditions (see page 408 in Pardo (2006)) which ensure the existence of the above integral. Moreover, we consider convex functions  $\phi$  which satisfy  $\phi(1) = 0$  and  $\phi''(1) \neq 0$ . Large values of  $D_\phi^{(k)}$  support the existence of a change point and therefore large values of  $D_\phi^{(k)}$  suggest rejection of the null hypothesis  $H_0$ . Hence  $D_\phi^{(k)}$  can be used as a test statistic for testing the hypotheses (4a). Then, motivated by the fact that large values of  $D_\phi^{(k)}$  are in favor of  $H_1$ , a test for testing the existence of a single change point, that is the hypotheses (4a), should be based on the  $\phi$ -divergence test statistic,

$$T_\phi^{(K)} = \max_{k \in \{1, \dots, K-1\}} T_\phi^{(K)}(k), \quad (6)$$

where

$$T_\phi^{(K)}(k) = \frac{k(K-k)}{K} \frac{2}{\phi''(1)} D_\phi^{(k)} \left( f_{\hat{\theta}_{0,k}^{(K)}}, f_{\hat{\theta}_{1,k}^{(K)}} \right). \quad (7)$$

Moreover, the unknown position of the change point  $\kappa$  is estimated by  $\hat{\kappa}_\phi$  such that

$$\hat{\kappa}_\phi = \arg \max_{k \in \{1, \dots, K-1\}} T_\phi^{(K)}(k) = \arg \max_{k \in \{1, \dots, K-1\}} \frac{k(K-k)}{K} D_\phi^{(k)} \left( f_{\hat{\theta}_{0,k}^{(K)}}, f_{\hat{\theta}_{1,k}^{(K)}} \right). \quad (8)$$

Based on the above discussion,  $H_0$  in (4a) is rejected for  $T_\phi^{(K)} > c$ , where  $c$  is a constant to be determined by the null distribution of  $T_\phi^{(K)}$ . Hence, in order to use  $T_\phi^{(K)}$  of (6) for testing hypotheses (4a), it is necessary the knowledge of the distribution of  $T_\phi^{(K)}$ , under  $H_0$ .

There are two important reasons why working directly with test-statistics  $T_\phi^{(K)}$ , defined in (6), is avoided, on one hand, its asymptotic distribution  $\sup_{t \in (0,1)} \frac{1}{t(1-t)} \left\| \mathbf{W}_0^{(m)}(t) \right\|^2$ , is not an easy to handle random variable (see for instance Theorem 1.2 and 1.3 in Gombay and Horváth (1989)) and on the other hand, in practice cases such that  $\kappa \in \{1, K-1\}$  are very difficult to detect. Let  $N(\epsilon)$  be the set all possible integers  $k \in \{1, \dots, K-1\}$  such that  $\frac{k}{K} \in [\epsilon, 1-\epsilon]$ , with  $\epsilon > 0$ , small enough. We shall modify (6) to be maximized in  $N(\epsilon)$ , i.e.

$$\epsilon T_\phi^{(K)} = \max_{k \in N(\epsilon)} T_\phi^{(K)}(k), \quad (9)$$

and in the same manner (8) becomes

$$\epsilon \hat{\kappa}_\phi = \arg \max_{k \in N(\epsilon)} T_\phi^{(K)}(k) = \arg \max_{k \in N(\epsilon)} \frac{k(K-k)}{K} D_\phi^{(k)} \left( f_{\hat{\theta}_{0,k}^{(K)}}, f_{\hat{\theta}_{1,k}^{(K)}} \right). \quad (10)$$

### 3 Main result

In order to get the asymptotic distribution of the family of tests statistics  $T_\phi^{(K)}$ , given in (6), we shall assume the usual regularity assumptions for the multiparameter Central Limit Theorem (see for instance Theorem 5.2.2. in Sen and Singer (1993)):

(i) The parameter space,  $\Theta$ , is either  $\mathbb{R}^m$  or a rectangle in  $\mathbb{R}^m$ .

(ii) For all  $\boldsymbol{\theta} \neq \boldsymbol{\theta}' \in \Theta \subset \mathbb{R}^m$ ,

$$\mu \left( \{\mathbf{x} \in \mathcal{X}^{(d)} : f_{\boldsymbol{\theta}}(\mathbf{x}) \neq f_{\boldsymbol{\theta}'}(\mathbf{x})\} \right) > 0.$$

(iii) For  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)^T$ ,

$$\frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}}(\mathbf{x}) \quad \text{and} \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_{\boldsymbol{\theta}}(\mathbf{x}), \quad i, j \in \{1, \dots, m\},$$

exist almost everywhere and are such that

$$\left| \frac{\partial}{\partial \theta_i} f_{\boldsymbol{\theta}}(\mathbf{x}) \right| \leq H_i(\mathbf{x}) \quad \text{and} \quad \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} f_{\boldsymbol{\theta}}(\mathbf{x}) \right| \leq G_{ij}(\mathbf{x}), \quad i, j \in \{1, \dots, m\},$$

where

$$\int_{\mathcal{X}^{(d)}} H_i(\mathbf{x}) d\mu(\mathbf{x}) < \infty \quad \text{and} \quad \int_{\mathcal{X}^{(d)}} G_{ij}(\mathbf{x}) d\mu(\mathbf{x}) < \infty, \quad i, j \in \{1, \dots, m\}.$$

(iv) Denoting  $\ell(\mathbf{x}; \boldsymbol{\theta}) = \log f_{\boldsymbol{\theta}}(\mathbf{x})$ ,

$$\frac{\partial}{\partial \theta_i} \ell(\mathbf{x}; \boldsymbol{\theta}) \quad \text{and} \quad \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\mathbf{x}; \boldsymbol{\theta}), \quad i, j \in \{1, \dots, m\},$$

exist almost everywhere and are such that the Fisher information matrix is finite and positive definite. In addition,  $\lim_{\delta \rightarrow 0} \psi(\delta) = 0$  where

$$\psi(\delta) = E_{\boldsymbol{\theta}} \left[ \sup_{\{\mathbf{h}: \|\mathbf{h}\| \leq \delta\}} \left\| \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell(\mathbf{x}; \boldsymbol{\theta} + \mathbf{h}) - \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell(\mathbf{x}; \boldsymbol{\theta}) \right\| \right],$$

with  $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell(\mathbf{x}; \boldsymbol{\theta}) = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \ell(\mathbf{x}; \boldsymbol{\theta}) \right)_{i,j \in \{1, \dots, m\}}$  and  $\|\bullet\|$  is the Euclidean norm.

**Theorem 1** Under  $H_0$  in (4a) and the previous regularity assumptions, (i)-(iv), the asymptotic distribution of (9) is given by

$${}_{\epsilon} T_{\phi}^{(K)} \xrightarrow[K \rightarrow \infty]{\mathcal{L}} \mathcal{T}_{m, \epsilon} \quad (11)$$

where  $m = \dim(\Theta)$ ,

$$\mathcal{T}_{m, \epsilon} = \sup_{t \in [\epsilon, 1-\epsilon]} \frac{1}{t(1-t)} \left\| \mathbf{W}_0^{(m)}(t) \right\|^2, \quad (12)$$

with  $\mathbf{W}_0^{(m)}(t) = \{(W_{0,1}(t), \dots, W_{0,m}(t))^T\}_{t \in [0,1]}$ , being an  $m$ -dimensional vector of independent Brownian bridges and  $\left\| \mathbf{W}_0^{(m)}(t) \right\|^2 = \sum_{i=1}^m W_{0,i}^2(t)$ .

**Proof.** According to the properties of the MLEs we know that

$$\begin{aligned} \sqrt{k} \left( \widehat{\boldsymbol{\theta}}_{0,k}^{(K)} - \boldsymbol{\theta}_0 \right) &\xrightarrow[k \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, \mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)^{-1} \right), \\ \sqrt{K-k} \left( \widehat{\boldsymbol{\theta}}_{1,k}^{(K)} - \boldsymbol{\theta}_1 \right) &\xrightarrow[(K-k) \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, \mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_1)^{-1} \right), \end{aligned}$$

where for  $\boldsymbol{\theta} \in \Theta$ , such that  $m = \dim \Theta$ ,  $\mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}) = \left( -E \left[ \frac{\partial^2}{\partial \theta_i \partial \theta_i} \log f_{\boldsymbol{\theta}}(\mathbf{X}_1) \right] \right)_{i,j \in \{1, \dots, m\}}$ , is the information matrix. If we consider that  $\lambda_k^{(K)} = \lim_{K \rightarrow \infty} \frac{k}{K}$ , then

$$\begin{aligned} \sqrt{\frac{k(K-k)}{K}} \left( \widehat{\boldsymbol{\theta}}_{0,k}^{(K)} - \boldsymbol{\theta}_0 \right) &\xrightarrow{K \rightarrow \infty} \mathcal{N} \left( 0, (1 - \lambda_k^{(K)}) \mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)^{-1} \right), \\ \sqrt{\frac{k(K-k)}{K}} \left( \widehat{\boldsymbol{\theta}}_{1,k}^{(K)} - \boldsymbol{\theta}_1 \right) &\xrightarrow{K \rightarrow \infty} \mathcal{N} \left( 0, \lambda_k^{(K)} \mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_1)^{-1} \right). \end{aligned}$$

This means that under  $\mathcal{H}_0$ , i.e.  $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ ,

$$\sqrt{\frac{k(K-k)}{K}} \left( \widehat{\boldsymbol{\theta}}_{0,k}^{(K)} - \widehat{\boldsymbol{\theta}}_{1,k}^{(K)} \right) \xrightarrow{K \rightarrow \infty} \mathcal{N} \left( 0, \mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)^{-1} \right),$$

and hence we can construct a Wald-type test-statistic as follows

$$Q_k^{(K)} = \frac{k(K-k)}{K} \left( \widehat{\boldsymbol{\theta}}_{0,k}^{(K)} - \widehat{\boldsymbol{\theta}}_{1,k}^{(K)} \right)^T \widehat{\mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)} \left( \widehat{\boldsymbol{\theta}}_{0,k}^{(K)} - \widehat{\boldsymbol{\theta}}_{1,k}^{(K)} \right), \quad (13)$$

where  $\widehat{\mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)}$  is any consistent estimator of  $\mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)$ . From Theorem 1 in Hawkins (1987) we know that

$$\max_{k \in N(\epsilon)} Q_k^{(K)} \xrightarrow{K \rightarrow \infty} \mathcal{T}_{m,\epsilon}$$

In addition from Pardo (2006), page 443, we have

$$T_{\phi}^{(K)}(k) = \frac{k(K-k)}{K} \frac{2}{\phi''(1)} D_{\phi}(f_{\widehat{\boldsymbol{\theta}}_{0,k}^{(K)}}, f_{\widehat{\boldsymbol{\theta}}_{1,k}^{(K)}}) = Q_k^{(K)} + o_P(1)$$

where

$$D_{\phi}(f_{\widehat{\boldsymbol{\theta}}_{0,k}^{(K)}}, f_{\widehat{\boldsymbol{\theta}}_{1,k}^{(K)}}) = \int f_{\widehat{\boldsymbol{\theta}}_{1,k}^{(K)}}(x) \phi \left( \frac{f_{\widehat{\boldsymbol{\theta}}_{0,k}^{(K)}}(x)}{f_{\widehat{\boldsymbol{\theta}}_{1,k}^{(K)}}(x)} \right) dx.$$

With both results we conclude (11). ■

**Remark 2** If we compare (13) with formula (2.3) in Hawkins (1987), both apparently are not equivalent because in our case  $\frac{k(K-k)}{K}$  appears rather than  $k(K-k)$  of formula (2.3). This difference is associated with the way of understanding Fisher Information matrix, in fact our Wald test-statistic coincide with the empirical stochastic process denoted by  $\widetilde{Q}_K(t)$  at the beginning of Section 3 in Hawkins (1987).

**Remark 3** The probability distribution function of random variable  $\mathcal{T}_{m,\epsilon}$ , for  $\epsilon > 0$ , given in (12), can be found in Sen (1981, page 397) and De Long (1981). The computation of the probability distribution function is complex, however it is possible to approximate the p-value of the test in which the distribution of  $\mathcal{T}_{m,\epsilon}$  is considered under the null hypothesis. In Estrella (2003), for instance,

$$\widetilde{p\text{-value}}(x, \epsilon) = \frac{1}{\Gamma\left(\frac{m}{2}\right)} \left(\frac{x}{2}\right)^{\frac{m}{2}} \exp\left\{-\frac{x}{2}\right\} \left(\log\left(\frac{(1-\epsilon)^2}{\epsilon^2}\right) \left(1 - \frac{m}{x}\right) + \frac{2}{x}\right), \quad (14)$$

with  $\Gamma(t)$  being the Gamma function, is proposed as an approximation of

$$\begin{aligned} p\text{-value}(x, \epsilon) &= \Pr(\mathcal{T}_{m, \epsilon} > x) = \Pr\left(\sup_{s \in (1, (1-\epsilon)^2/\epsilon^2)} \frac{1}{\sqrt{s}} \|\mathbf{W}_0^{(m)}(s)\| > \sqrt{x}\right) \\ &= \frac{1}{\Gamma\left(\frac{m}{2}\right)} \left(\frac{x}{2}\right)^{\frac{m}{2}} \exp\left\{-\frac{x}{2}\right\} \left(\log\left(\frac{(1-\epsilon)^2}{\epsilon^2}\right) \left(1 - \frac{m}{x}\right) + \frac{2}{x} + O\left(\frac{1}{x^2}\right)\right). \end{aligned}$$

When calibrating the approximation for the univariate parameter ( $m = 1$ ), we can take into account that the exact quantiles of order  $(1 - \alpha) \in \{0.90, 0.95, 0.99\}$  for  $\epsilon = 0.05$ , are 8.31, 9.90 and 13.45 respectively, i.e.  $p\text{-value}(8.31, 0.05) = 0.1$ ,  $p\text{-value}(9.90, 0.05) = 0.05$ ,  $p\text{-value}(13.45, 0.05) = 0.01$ . If we use (14) with  $\epsilon = 0.05$  and the aforementioned quantiles, we obtain  $p\text{-value}(8.31, 0.05) = 9.7789 \times 10^{-2}$ ,  $p\text{-value}(9.90, 0.05) = 4.8868 \times 10^{-2}$ ,  $p\text{-value}(13.45, 0.05) = 9.8358 \times 10^{-3}$ . We can see that in particular,  $p\text{-value}(x, 0.05)$  approximates very well  $p\text{-value}(x, 0.05)$  when  $x$  is the quantile of order  $1 - \alpha = 0.99$ , which is in practice of major interest.

## 4 Simulation Study

In this section we are going to focus on the change point analysis for a particular discrete probability model, the binomial model. For this special case we will give an explicit expression for divergence based test-statistics. The accuracy will be compared by simulation with respect to pre-existing test-statistics. In this context, suppose we are dealing with a sequence of independent r.v.'s  $\tilde{X}_i \sim \mathcal{B}\text{in}(n_i, \theta_i)$ ,  $i = 1, \dots, K$ , for which we are interested in testing (1). In order to do that we are going to consider a sequence of independent Bernoulli r.v.'s  $X_{ih} \sim \mathcal{B}\text{er}(\theta_i)$ ,  $i = 1, \dots, K$ ,  $h = 1, \dots, n_i$ , whose probability mass function (p.m.f.) is given by  $p_{\theta_i}(x) = \theta_i^x (1 - \theta_i)^{1-x}$ ,  $x \in \{0, 1\}$ , and  $p_{\theta_i}(x) = 0$ ,  $x \notin \{0, 1\}$ . If we denote the cumulative steps between consecutive Binomial r.v.'s by

$$N_k = \sum_{i=1}^k n_i.$$

the change points are located at  $\{1, 2, \dots, N_K - 1, N_K\}$  for  $X_{ih}$  and at  $\{N_k\}_{k=1}^K$  for  $\tilde{X}_i$ . Hence,  $X_{ih}$  is the only sequence of r.v.'s which are strictly identically distributed, but the change points of interest are located in  $\{N_k\}_{k=1}^K \subset \{1, 2, \dots, N_K - 1, N_K\}$ . This means that we can construct the test-statistic by considering a sequence of i.i.d. r.v.'s but in addition we restrict the set of possible change points to  $\{N_k\}_{k=1}^K$ , rather than one step change points. When the change point is located at  $N_k$ , the MLEs of  $\theta_0$  and  $\theta_1$  are given by

$$\begin{aligned} \hat{\theta}_{0,k}^{(K)} &= \frac{Y_k}{N_k}, & \hat{\theta}_{1,k}^{(K)} &= \frac{Y_K - Y_k}{N_K - N_k}, \\ Y_k &= \sum_{i=1}^k \tilde{X}_i = \sum_{i=1}^k \sum_{h=1}^{n_i} X_{ih}. \end{aligned}$$

The likelihood ratio test-statistic is given by  $S^{(K)} = \max_{k \in \{1, \dots, K\}} S_k^{(K)}$ , where

$$S_k^{(K)} = 2 \left[ N_k \left( \widehat{\theta}_{0,k}^{(K)} \log \left( \frac{\widehat{\theta}_{0,k}^{(K)}}{\widehat{\theta}_{0,K}^{(K)}} \right) + (1 - \widehat{\theta}_{0,k}^{(K)}) \log \left( \frac{1 - \widehat{\theta}_{0,k}^{(K)}}{\widehat{\theta}_{0,K}^{(K)}} \right) \right) \right. \\ \left. + (N_K - N_k) \left( \widehat{\theta}_{1,k}^{(K)} \log \left( \frac{\widehat{\theta}_{1,k}^{(K)}}{\widehat{\theta}_{0,K}^{(K)}} \right) + (1 - \widehat{\theta}_{1,k}^{(K)}) \log \left( \frac{1 - \widehat{\theta}_{1,k}^{(K)}}{\widehat{\theta}_{0,K}^{(K)}} \right) \right) \right] \quad (15)$$

Two important papers which cover  $S^{(K)}$  are Worsley (1983), and Horváth (1989). The expression they gave for  $S_k^{(K)}$  is not exactly the same, but it is equivalent to (15) (see formula (3.22) in Horváth and Serbinowska (1995)). Horváth (1989) found that the asymptotic distribution for a kind of normalization of  $S^{(K)}$  based on the Darling-Erdős formula

$$G^{(K)} = \sqrt{2 \log N_K S^{(K)}} - 2 \log N_K - \frac{1}{2} \log \log N_K + \frac{1}{2} \log \pi,$$

is asymptotically equal to a Extreme Value random variable with parameters  $\mu = \log 2$  and  $\beta = 1$ . In addition, in Theorem 1.2 of Horváth and Serbinowska (1995), a modified version of the likelihood ratio test-statistic was given,  $\widetilde{S}^{(K)} = \max_{k \in \{1, \dots, K\}} \widetilde{S}_k^{(K)}$ , where

$$\widetilde{S}_k^{(K)} = \frac{N_k(N_K - N_k)}{N_K^2} S_k^{(K)}.$$

The asymptotic distribution of  $\widetilde{S}^{(K)}$  is the supremum in  $(0, 1)$  of a standard univariate Brownian bridge (its probability distribution function is tabulated in Kiefer (1959)). We consider the version of the Wald test-statistic  ${}^\epsilon Q^{(K)} = \max_{k \in N(\epsilon)} Q_k^{(K)}$ , with

$$Q_k^{(K)} = \frac{N_k(N_K - N_k)}{N_K} (\widehat{\theta}_{0,k}^{(K)} - \widehat{\theta}_{1,k}^{(K)})^2 \widehat{\mathbf{I}}_{\mathcal{F}}(\boldsymbol{\theta}_0),$$

where the consistent estimator of  $\mathbf{I}_{\mathcal{F}}(\boldsymbol{\theta}_0)$  is given by

$$\widehat{\mathbf{I}}_{\mathcal{F}}(\boldsymbol{\theta}_0) = \frac{N_k}{N_K} \mathbf{I}_{\mathcal{F}}(\widehat{\theta}_{0,k}^{(K)}) + \frac{N_K - N_k}{N_K} \mathbf{I}_{\mathcal{F}}(\widehat{\theta}_{1,k}^{(K)}) = \frac{N_k}{N_K} \frac{1}{\widehat{\theta}_{0,k}^{(K)} (1 - \widehat{\theta}_{0,k}^{(K)})} + \frac{N_K - N_k}{N_K} \frac{1}{\widehat{\theta}_{1,k}^{(K)} (1 - \widehat{\theta}_{1,k}^{(K)})}.$$

Finally, in order to give an explicit expression for divergence based test-statistics we are going to focus on a family of divergences, power divergences (see Read and Cressie (1988)), for which  $\phi_\lambda(x) = \frac{1}{\lambda(1+\lambda)} (x^{\lambda+1} - x - \lambda(x-1))$ , if  $\lambda(1+\lambda) \neq 0$  and  $\phi_\lambda(x) = \lim_{\ell \rightarrow \lambda} \phi_\ell(x)$ , if  $\lambda(1+\lambda) = 0$ , that is for each  $\lambda \in \mathbb{R}$  we obtain a different divergence measure between the p.m.f.s  $p_{\theta_0}$  and  $p_{\theta_1}$ ,

$$D_\lambda(p_{\theta_0}, p_{\theta_1}) = \frac{1}{\lambda(1+\lambda)} \left( \frac{\theta_0^{\lambda+1}}{\theta_1^\lambda} + \frac{(1-\theta_0)^{\lambda+1}}{(1-\theta_1)^\lambda} - 1 \right), \text{ if } \lambda(1+\lambda) \neq 0.$$

When  $\lambda = 0$  the power divergence coincides with the so called Kullback divergence

$$D_0(p_{\theta_0}, p_{\theta_1}) = D_{\text{Kull}}(p_{\theta_0}, p_{\theta_1}) = \left( \theta_0 \log \left( \frac{\theta_0}{\theta_1} \right) + (1 - \theta_0) \log \left( \frac{1 - \theta_0}{1 - \theta_1} \right) \right),$$



and when  $\lambda = -1$  the power divergence coincides with the modified Kullback divergence  $D_{-1}(p_{\theta_0}, p_{\theta_1}) = D_{\text{Kull}}(p_{\theta_1}, p_{\theta_0})$ . Hence, the shape of the power-divergence based test-statistics is  ${}^\epsilon T_\lambda^{(K)} = \max_{k \in N(\epsilon)} T_\lambda(\hat{\theta}_{0,k}^{(K)}, \hat{\theta}_{1,k}^{(K)})$ , where

$$T_\lambda(\hat{\theta}_{0,k}^{(K)}, \hat{\theta}_{1,k}^{(K)}) = 2 \frac{N_k(N_K - N_k)}{N_K} D_\lambda \left( p_{\hat{\theta}_{0,k}^{(K)}}, p_{\hat{\theta}_{1,k}^{(K)}} \right),$$

that is

$$T_\lambda(\hat{\theta}_{0,k}^{(K)}, \hat{\theta}_{1,k}^{(K)}) = \frac{N_k(N_K - N_k)}{N_K} \frac{2}{\lambda(1+\lambda)} \left( \frac{(\hat{\theta}_{0,k}^{(K)})^{\lambda+1}}{(\hat{\theta}_{1,k}^{(K)})^\lambda} + \frac{(1 - \hat{\theta}_{0,k}^{(K)})^{\lambda+1}}{(1 - \hat{\theta}_{1,k}^{(K)})^\lambda} - 1 \right), \text{ for } \lambda(1+\lambda) \neq 0, \quad (16)$$

and

$$T_0(\hat{\theta}_{0,k}^{(K)}, \hat{\theta}_{1,k}^{(K)}) = 2 \frac{N_k(N_K - N_k)}{N_K} \left( \hat{\theta}_{0,k}^{(K)} \log \left( \frac{\hat{\theta}_{0,k}^{(K)}}{\hat{\theta}_{1,k}^{(K)}} \right) + (1 - \hat{\theta}_{0,k}^{(K)}) \log \left( \frac{1 - \hat{\theta}_{0,k}^{(K)}}{1 - \hat{\theta}_{1,k}^{(K)}} \right) \right). \quad (17)$$

Assuming that there is a monotone, continuous function  $g$  such that  $g(0) = 0$  and

$$\lim_{K \rightarrow \infty} \max_{k \in N(\epsilon)} \left| \frac{N_k(N_K - N_k)}{N_K} - g \left( \frac{k(K - k)}{K} \right) \right| = 0,$$

the asymptotic distribution of  ${}^\epsilon Q^{(K)}$  and  ${}^\epsilon T_\lambda^{(K)}$ , for all  $\lambda \in \mathbb{R}$ , is the supremum in  $[\epsilon, 1 - \epsilon]$  of the univariate tied-down Bessel process, i.e. (12) with  $m = 1$ . This assumption is very similar to the assumption given in Horváth and Serbinowska (1995) for the asymptotic distribution of  $\tilde{S}^{(K)}$ .

A simulation study is performed in order to compare the accuracy of the proposed power divergence type test with respect to pre-existing test-statistics. In this context we apply test-statistics  $\tilde{S}^{(K)}$ ,  $G^{(K)}$ ,  ${}^{0.05}T_0^{(K)}$ ,  ${}^{0.05}T_1^{(K)}$ ,  ${}^{0.05}T_2^{(K)}$ ,  ${}^{0.05}Q^{(K)}$  with 5000 replication. The design is essentially the same as the study performed in Horváth and Serbinowska (1995):  $\theta_0 = 0.5$ ; three possible values of  $K$  and nominal sizes  $\alpha$  are considered; apart from the quantiles of order  $1 - \alpha$ ,  $x_{1-\alpha}$ , the exact sizes  $\hat{\alpha}$  are calculated. With  $K = \infty$ , it is understood that  $x_{1-\alpha}$  is the asymptotic quantile associated to the corresponding test-statistic. Taking into account that the maximization for obtaining  ${}^\epsilon T_1^{(K)}$ ,  ${}^\epsilon T_2^{(K)}$ ,  ${}^\epsilon Q^{(K)}$ , with  $\epsilon = 0.05$  is over all possible integers  $k \in N(\epsilon)$ , we removed  $k \in \{1, \dots, K - 1\}$  when  $k < \epsilon K$  or  $k > (1 - \epsilon)K$ .

Looking at the results given in Table 1, the worst approximation of  $\alpha$  is obtained with  $G^{(K)}$ . The Wald test-statistic  ${}^{0.05}Q^{(K)}$  is a good competitor for the test-statistic introduced in Horváth and Serbinowska (1995),  $\tilde{S}^{(K)}$ . All the exact sizes underestimate the nominal size, which means that the best approximation is obtained with the greatest value of  $\hat{\alpha}$ , hardly ever obtained with the power-divergence based test-statistic with  $\lambda = 2$ ,  ${}^{0.05}T_2^{(K)}$ .

Table 1: Exact simulated sizes.

|                      |      | $K = 64$     |                | $K = 300$      |                | $K = 500$      |                | $K = \infty$   |
|----------------------|------|--------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                      |      | $1 - \alpha$ | $x_{1-\alpha}$ | $\hat{\alpha}$ | $x_{1-\alpha}$ | $\hat{\alpha}$ | $x_{1-\alpha}$ | $\hat{\alpha}$ |
| $\tilde{S}^{(K)}$    | 0.90 | 1.302        | 0.0664         | 1.386          | 0.0786         | 1.420          | 0.0860         | 1.498          |
|                      | 0.95 | 1.619        | 0.0318         | 1.710          | 0.0372         | 1.740          | 0.0400         | 1.844          |
|                      | 0.99 | 2.595        | 0.0094         | 2.484          | 0.0072         | 2.531          | 0.0074         | 2.649          |
| $G^{(K)}$            | 0.90 | 1.707        | 0.0208         | 1.939          | 0.0260         | 2.011          | 0.0288         | 2.943          |
|                      | 0.95 | 2.277        | 0.0076         | 2.431          | 0.0094         | 2.555          | 0.0118         | 3.663          |
|                      | 0.99 | 3.394        | 0.0002         | 3.653          | 0.0002         | 3.796          | 0.0000         | 5.293          |
| ${}^{0.05}T_0^{(K)}$ | 0.90 | 7.351        | 0.0654         | 7.881          | 0.0834         | 7.801          | 0.0824         | 8.31           |
|                      | 0.95 | 8.968        | 0.0340         | 9.458          | 0.0412         | 9.514          | 0.0430         | 9.90           |
|                      | 0.99 | 12.730       | 0.0086         | 13.143         | 0.0094         | 12.981         | 0.0078         | 13.45          |
| ${}^{0.05}T_1^{(K)}$ | 0.90 | 7.374        | 0.0664         | 7.884          | 0.0832         | 7.809          | 0.0828         | 8.31           |
|                      | 0.95 | 9.007        | 0.0352         | 9.464          | 0.0412         | 9.509          | 0.0432         | 9.90           |
|                      | 0.99 | 12.851       | 0.0088         | 13.128         | 0.0094         | 12.981         | 0.0078         | 13.45          |
| ${}^{0.05}T_2^{(K)}$ | 0.90 | 7.432        | 0.0688         | 7.911          | 0.0840         | 7.818          | 0.0834         | 8.31           |
|                      | 0.95 | 9.141        | 0.0370         | 9.495          | 0.0416         | 9.519          | 0.0434         | 9.90           |
|                      | 0.99 | 13.061       | 0.0094         | 13.129         | 0.0094         | 13.015         | 0.0084         | 13.45          |
| ${}^{0.05}Q^{(K)}$   | 0.90 | 7.311        | 0.0642         | 7.871          | 0.0822         | 7.800          | 0.0824         | 8.31           |
|                      | 0.95 | 8.934        | 0.0334         | 9.441          | 0.0408         | 9.508          | 0.0430         | 9.90           |
|                      | 0.99 | 12.742       | 0.0084         | 13.135         | 0.0094         | 12.973         | 0.0078         | 13.45          |

## 5 Numerical Example: Lindisfarne Scribes problem

The Lindisfarne Gospels are presumed to be the work of a monk named Eadfrith, who became Bishop of Lindisfarne in year 698. In the 10th century an Old English translation of the Gospels was made for one or more scribes. Several statisticians have been devoted to studying the problem of the number of scribes who participated in the translation of the Gospels. Such a problem is known as the ‘‘Lindisfarne Scribes problem’’.

In the framework of the model that is followed in the simulation study, the Lindisfarne Gospels are considered to be divided into  $K = 64$  consecutive sections (see Ross (1950) for more details). It is supposed that each section could have been translated by one scribe and the same scribe is associated only with consecutive sections. Since the present indicative in Old English verbs admitted several variants in its spelling, the custom of using these variants can be used as a key factor useful to identifying different translators. Based on the data given in Table 2, it is counted  $n_i$  as the total of observed frequencies that the third singular or second plural appears in each section  $i = 1, \dots, 64$ , and the observation  $x_i$  (coming from r.v.  $X_i$ ) represents how many times ending  $-s$  appear in these verbs. Note that either the third singular or second plural admit two endings,  $-s$  and  $-\delta$ , and hence if we want to know how many times ending  $-\delta$  appear in these verbs, the observations are obtained as  $n_i - x_i$ ,  $i = 1, \dots, K$ . It is assumed that the custom of using both endings for each scribe is different and for this reason our interest is to find the consecutive changes in the probability structure of both endings.

Table 2: Data of the Lindisfarne’s problem

|       |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |
|-------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| $i$   | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| $x_i$ | 12 | 29 | 31 | 21 | 14 | 41 | 49 | 30 | 39 | 35 | 26 | 32 | 30 | 17 | 19 | 33 |
| $n_i$ | 21 | 39 | 44 | 25 | 19 | 66 | 62 | 34 | 47 | 47 | 29 | 33 | 38 | 21 | 21 | 36 |
| $i$   | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 |
| $x_i$ | 36 | 28 | 10 | 2  | 8  | 12 | 5  | 3  | 14 | 13 | 21 | 19 | 29 | 16 | 16 | 5  |
| $n_i$ | 40 | 33 | 25 | 5  | 23 | 28 | 20 | 28 | 20 | 23 | 41 | 32 | 39 | 28 | 21 | 24 |
| $i$   | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 |
| $x_i$ | 3  | 1  | 6  | 1  | 10 | 5  | 2  | 10 | 5  | 14 | 8  | 10 | 9  | 13 | 6  | 8  |
| $n_i$ | 30 | 15 | 23 | 5  | 35 | 30 | 14 | 56 | 51 | 62 | 45 | 55 | 42 | 27 | 36 | 31 |
| $i$   | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 |
| $x_i$ | 2  | 11 | 8  | 3  | 19 | 17 | 12 | 15 | 15 | 12 | 21 | 40 | 30 | 4  | 3  | 6  |
| $n_i$ | 9  | 26 | 38 | 29 | 55 | 37 | 45 | 47 | 44 | 45 | 33 | 65 | 85 | 13 | 9  | 16 |

Since the proposed test-statistics are valid for single change-point detection, now we are going to describe the algorithm based on the binary segmentation procedure. In order to make a sequence of hypothesis testing, it is convenient to use  $\alpha = 0.01$  if we want to get a not very large upper bound for the global significance level according to the Bonferroni’s inequality. Suppose that the power-divergence based test-statistics with  $\lambda = 2$ ,  $\epsilon = 0.05$ ,  ${}^{0.05}T_2^{(K)}$ , is our focus of interest. The algorithm based on the binary segmentation procedure (Vostrikova (1981)) is described in Figures 1-2. We consider  $N(\epsilon) = \{3, \dots, 61\}$  as change point candidates in Step 1, i.e. we have initially taken  $\{1, \dots, K - 1\}$  but we have removed all candidates  $k$  such that  $k < K\epsilon$  or  $k > K\epsilon$ . Once the values of  $T_2(\hat{\theta}_{0,k}^{(K)}, \hat{\theta}_{1,k}^{(K)})$  are

obtained for each candidate belonging to  $k \in N(\epsilon)$ , we select its maximum argument,  $k = 31$ , which is accepted as change-point because the  $p$ -value is less than 0.1. The  $p$ -values are calculated by following (14). From now we have to investigate how to divide  $[1, 31]$  into segments (Step 2) and  $[32, 64]$ . We will continue until all candidates have  $p$ -values greater than 0.1. After 12 steps it is concluded that the Lindisfarne Gospels could have been written by seven scribes because the obtained segments are  $[1, 10]$ ,  $[11, 18]$ ,  $[19, 23]$ ,  $\{24\}$ ,  $[25, 31]$ ,  $[32, 52]$ ,  $[53, 64]$ . This conclusion differs a little bit from the conclusion obtained in Horváth and Serbinowska (1995), because the number of scribes they proposed was one less and the locations of the change points are not exactly the same.

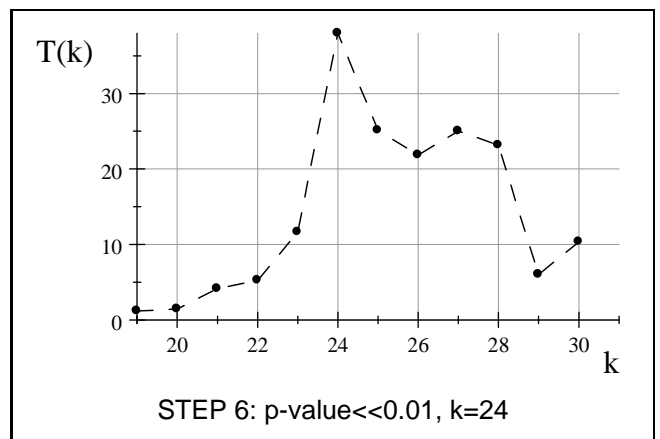
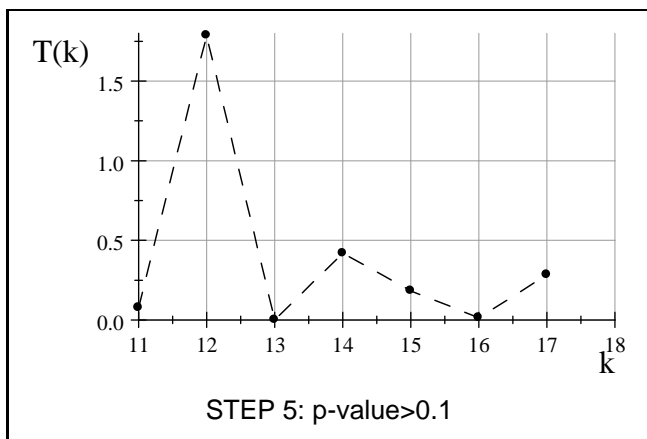
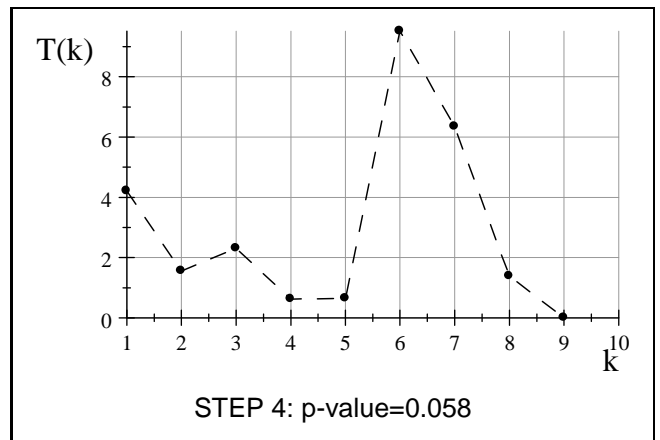
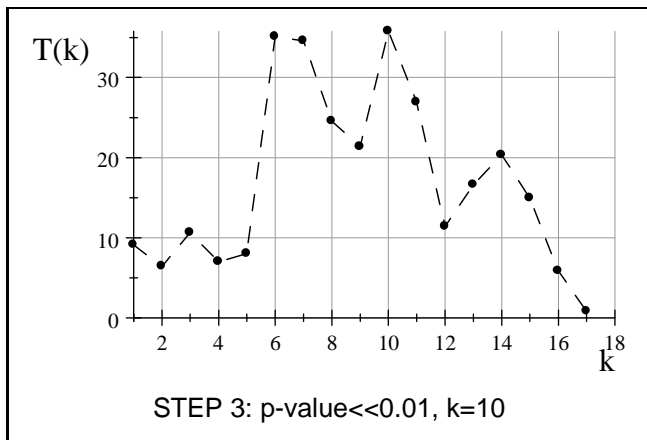
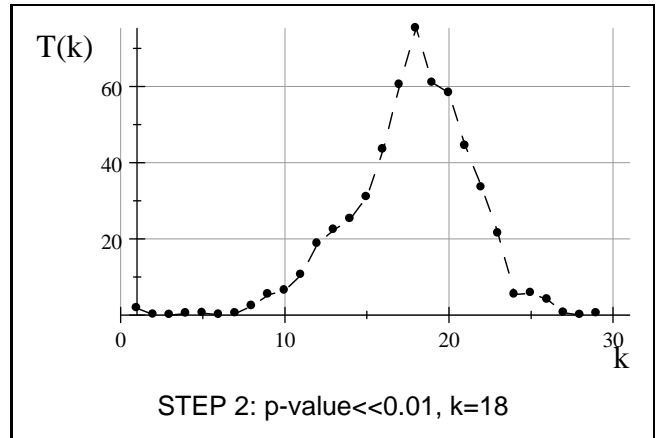
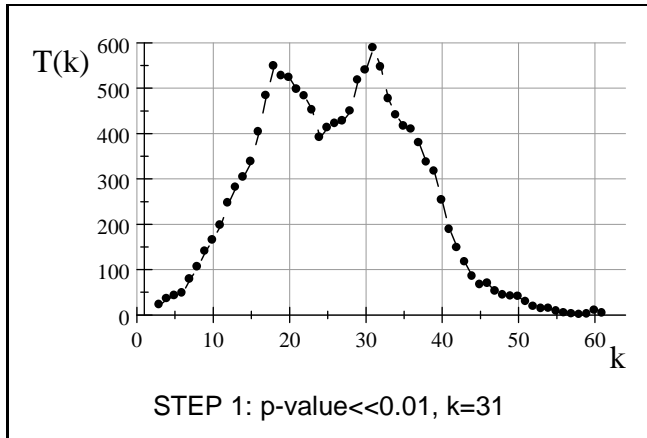


Figure 1: Binary segmentation procedure for the Lindisfarne's problem (part I)

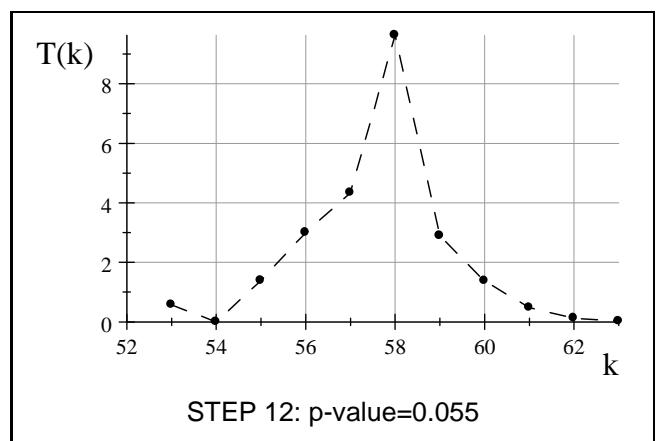
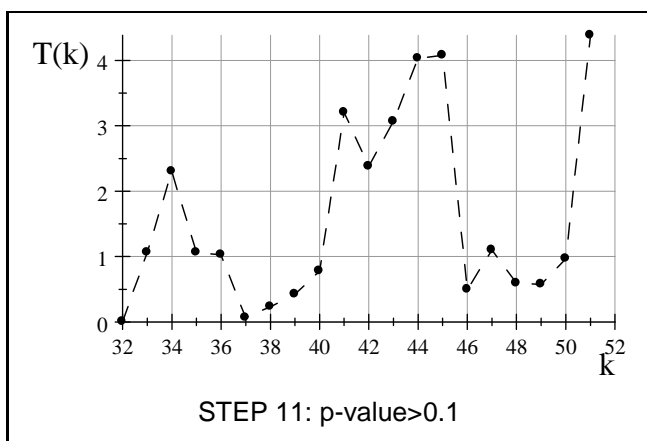
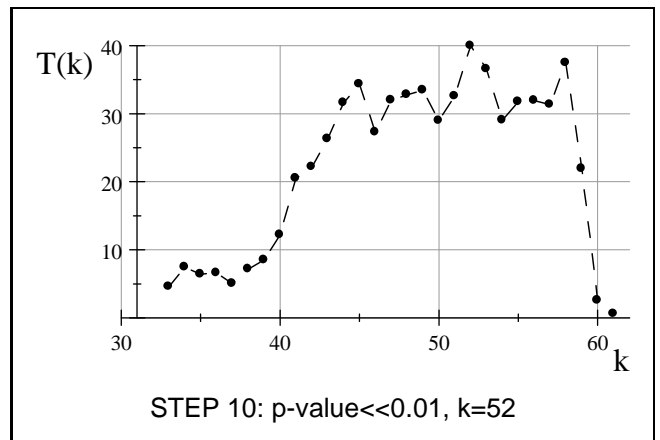
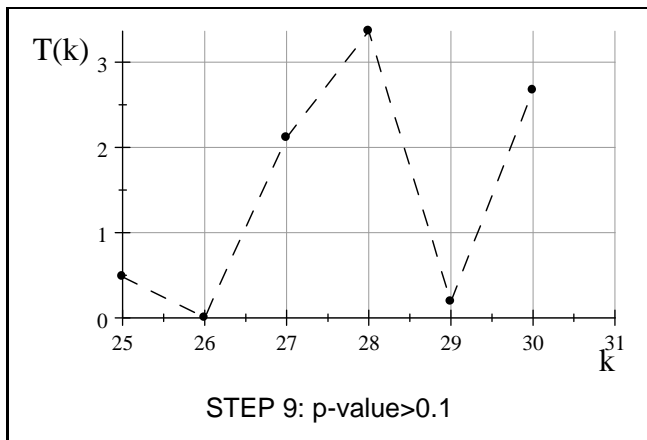
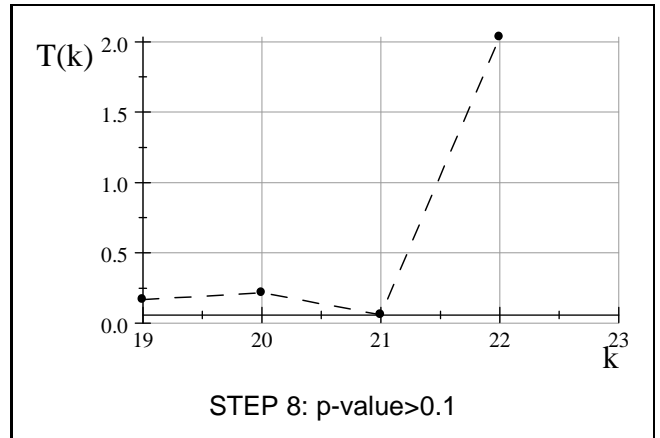
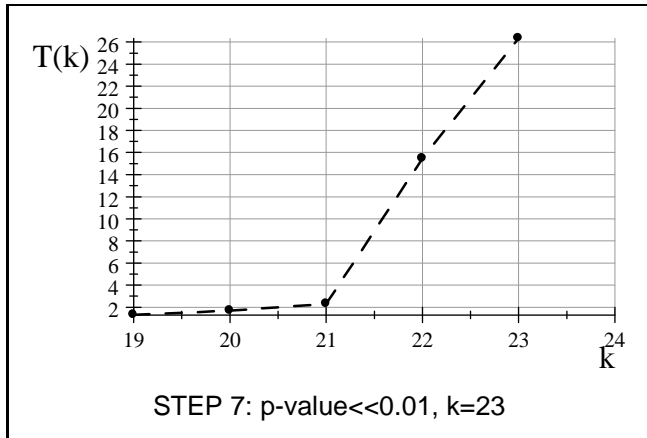


Figure 2: Binary segmentation procedure for the Lindisfarne's problem (part II)

## References

- [1] Chen, J. and Gupta, A. K. (1995). Likelihood procedure for testing change points hypothesis for multivariate Gaussian model. *Random Operators and Stochastic Equations* **3**, 235–244.
- [2] Chen, J. and Gupta, A. K. (2000). *Parametric statistical change point analysis*. Birkhauser Boston, Inc., Boston, MA.
- [3] Chen, J. and Gupta, A. K. (2001). On change point detection and estimation. *Comm. Statist. Simulation Comput.* **30**, 665–697.
- [4] Chen, J. and Gupta, A. K. (2004). Statistical inference on covariance change points in Gaussian model. *Statistics*, **38**, 17–28.
- [5] Csörgö, M. and Horváth, L. (1998). *Limit Theorems in Change-Point Analysis*. Wiley, New York.
- [6] De Long, D.M. (1981). Crossing probabilities for a square root boundary by a Bessel process . *Communications in Statistics–Theory and Methods*, **10**, 2197–2213.
- [7] Estrella , A. (2003): Critical Values And P Values Of Bessel Process Distributions: Computation And Application To Structural Break Tests. *Econometric Theory*, **19**, 2003, 1128–1143.
- [8] Gombay, E. and Horváth, L. (1996). On the Rate of Approximations for Maximum Likelihood Tests in Change-point Models. *Journal of Multivariate Analysis*, **56**, 120–152.
- [9] Hawkins, D. L. (1987). A Test for a Change Point in a Parametric Model Based on a Maximal Wald-Type Statistic. *Sankhyā: The Indian Journal of Statistics, Series A*, **49**, 368-376.
- [10] Horváth, L. (1989). The limit distributions of the likelihood ratio and cumulative sum tests for a change in binomial probability. *Journal of Multivariate Analysis*, **31**, 148–159.
- [11] Horváth, L. and Serbinowska, M. (1995). Testing for Changes in Multinomial Observations: the Lindisfarne Scribes problem. *Scandinavian Journal of Statistics*, **22**, 371–384.
- [12] Kiefer (1959). K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-V. Mises Tests. *Annals of Mathematical Statistics*, **30**, 420–447.
- [13] Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC, Boca Raton.
- [14] Read, T. and Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data*. Springer, New York.
- [15] Ross, A.S.C. (1950). Philological probability problems. *Journal of the Royal Statistical Society – Series B*, **12**, 19–59.
- [16] Sen, A. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference*. Wiley, New York

- [17] Sen, A. K. and Srivastava, M. S. (1980). On tests for detecting change in the multivariate mean. Tech. Report No. 3, University of Toronto.
- [18] Sen, A. K. and Singer, J. M. (2003). Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall, New York.
- [19] Srivastava, M. S. and Worsley, K. J. (1986). Likelihood ratio tests for a change in the multivariate normal mean. *Journal of American Statistical Association*, **81**, 199–204.
- [20] Vostrikova, L. Ju. (1981). Detecting disorder in multidimensional random processes. *Soviet Math Dokl.* **24**, 55–59.
- [21] Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, **70**, 455–464.