UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 11-01
Statistics and Econometrics Series 01
January 2011

Departamento de Estadística
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624-98-49

# CHANGE POINT FOR MULTINOMIAL DATA USING PHI-DIVERGENCE TEST STATISTICS

**Apostolos Batsidis[1], Nirian Martín[2], Leandro Pardo[3], Konstantinos Zografos[4]**

## Abstract

We propose two families of maximally selected phi-divergence tests for studying change point locations when the unknown probability vectors of a sequence of multinomial random variables, with possibly different sizes, are piecewise constant. In addition, these test-statistics are valid to estimate the location of the change-point. Two variants of the first family are considered by following two versions of the Darling-Erdös' formula. Under the no changes null hypothesis, we derive their limit distributions, extreme value and Gaussian-type respectively. We pay special attention to the checking the accuracy of these limit distributions in case of finite sample sizes. In such a framework, a Monte Carlo analysis shows the possibility of improving the behaviour of the test-statistics based on the likelihood ratio and chi-square tests introduced in Horváth and Serbinowska (1995). The data of the classical Lindisfarne Scribes problem are used in order to apply the proposed test-statistics.

***Keywords:*** multinomial sampling, change-point, phi-divergence test-statistics

[1] Department of Mathematics, University of Ioannina, Greece. E-mail: abatsidis@uoi.gr

[2] Department of Statistics, Universidad Carlos III de Madrid, Spain. E-mail: nirian.martin@uc3m.es

[3] Department of Statistics and Operations Research, Universidad Complutense de Madrid, Spain. E-mail: lpardo@mat.ucm.es

[4] Department of Mathematics, University of Ioannina, Greece. E-mail: kzograf@uoi.gr

# Change Point for Multinomial Data using Phi-divergence Test Statistics

Batsidis, A.[1], Martín, N.[2,*], Pardo, L.[3] and Zografos, K.[1]

[1]Dep. Mathematics, University of Ioannina, Greece

[2]Dep. Statistics, Carlos III University of Madrid, Spain

[3]Dep. Statistics and O.R., Complutense University of Madrid, Spain

January 12, 2011

**Abstract**

We propose two families of maximally selected phi-divergence tests for studying change point locations when the unknown probability vectors of a sequence of multinomial random variables, with possibly different sizes, are piecewise constant. In addition, these test-statistics are valid to estimate the location of the change-point. Two variants of the first family are considered by following two versions of the Darling-Erdös' formula. Under the no changes null hypothesis, we derive their limit distributions, extreme value and Gaussian-type respectively. We pay special attention to checking the accuracy of these limit distributions in case of finite sample sizes. In such a framework, a Monte Carlo analysis shows the possibility of improving the behaviour of the test-statistics based on the likelihood ratio and chi-square tests introduced in Horváth and Serbinowska (1995). The data of the classical Lindisfarne Scribes problem are used in order to apply the proposed test-statistics.

**Keywords:** Multinomial sampling, Change-point, Phi-divergence test-statistics.

## 1 Introduction

Let $\boldsymbol{X}_1 = (X_{11}, ..., X_{1m})^T$,..., $\boldsymbol{X}_K = (X_{K1}, ..., X_{Km})^T$ be a sequence of independent multinomial random variables with parameters $(n_1, \boldsymbol{p}_1), ...., (n_K, \boldsymbol{p}_K)$, where $X_{i1} + ... + X_{im} = n_i$ are known integer values and $\boldsymbol{p}_i = (p_{i1}, ..., p_{im})^T$, with $p_{ij} > 0$, $p_{i1} + ... + p_{im} = 1$, are unknown probability vectors ($i \in \{1, ..., K\}$). We are interested in testing

$$\mathcal{H}_0(K) : \boldsymbol{p}_1 = ... = \boldsymbol{p}_K = \boldsymbol{p}, \tag{1}$$

against the change-point alternative:

$$\begin{aligned} \mathcal{H}_A(K) : \quad & \text{there is an unknown } \kappa \in \{1, ..., K-1\} \text{ such that} \\ & \boldsymbol{p}_1 = ... = \boldsymbol{p}_\kappa \neq \boldsymbol{p}_{\kappa+1} = ... = \boldsymbol{p}_K. \end{aligned} \tag{2}$$

The parametric problem of change points at unknown positions has been studied by many authors, see for instance, Hinkley and Hinkley (1970), Horváth (1989), Wolfe and Chen (1990), Horváth and Serbinowska (1995), Csörgo and Horváth (1997, Section 1.7.2), Chen and Gupta (2000, chapter 7), Hawkins (2001), and references therein.

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

The idea developed in the cited paper of Horváth and Serbinowska (1995) is to test one change-point temporarily located at $k$:

$$\mathcal{H}_0^{(k)}(K) : \boldsymbol{p}^{(k)}{=}\boldsymbol{q}^{(k)} \text{ versus } \mathcal{H}_A^{(k)}(K) : \boldsymbol{p}^{(k)} \neq \boldsymbol{q}^{(k)}, \tag{3}$$

where $k \in \{1, ..., K-1\}$ is fixed and $\boldsymbol{p}^{(k)}$ is the common distribution for the first $k$ populations, i.e., $\boldsymbol{p}^{(k)} = \boldsymbol{p}_1 = ... = \boldsymbol{p}_k$ and $\boldsymbol{q}^{(k)}$ is the common distribution for the remain populations, i.e., $\boldsymbol{q}^{(k)} = \boldsymbol{p}_{k+1} = ... = \boldsymbol{p}_K$. If we denote by $T_k^{(K)}$ the provisional test statistic considered for testing (3), it is clear that $T^{(K)} = \max_{k\in\{1,...,K-1\}} T_k^{(K)}$ will be the definite test statistic for testing the one change-point problem, $\mathcal{H}_0(K)$ given in (1) against $\mathcal{H}_A(K)$ given in (2). Once a change-point is identified, we can follow applying these test-statistics sequentially using the binary segmentation procedure proposed by Vostrikova (1981). This approach consists in splitting the sequenced data in subsequences separated by the identified change-points and using the test-statistics individually on these subsequences in order to test inside the hypothesis of equal probability vector.

The main purpose of this paper is to introduce a new families of test statistics for (1) against (2). Such families of test statistics contains as a particular case those that were were introduced and studied, for this problem, in Horváth and Serbinowska (1995).

In Section 2 we present a new family of test statistics for testing $\mathcal{H}_0(K)$ given in (1) against $\mathcal{H}_A(K)$ given in (2) based on phi-divergence measures. The main result is developed in Section 3 and finally in Sections 4 and 5 we present a simulation study and a numerical example respectively in order to clarify the results presented in this paper.

## 2 Phi-divergence test statistics and change-point for multinomial data

In order to present the family of phi-divergence test statistics for the change point in multinomial data it is necessary to introduce some notation. The maximum likelihood estimator of probability vector $\boldsymbol{p}_k^{(K)}$, where $k \in \{1, ..., K\}$, based on $\boldsymbol{X}_1 = (X_{11}, ..., X_{1,m})^T$,..., $\boldsymbol{X}_k = (X_{k1}, ..., X_{km})^T$, will be denoted by $\widehat{\boldsymbol{p}}_k = (\widehat{p}_{k1}, ..., \widehat{p}_{km})^T$, where

$$\widehat{p}_{kj} = \frac{Y_{kj}}{N_k}, \qquad j = 1, ..., m,\ k = 1, ...K,$$

being

$$Y_{kj} \equiv X_{1j} + ... + X_{kj} \qquad \text{and} \qquad N_k \equiv n_1 + ... + n_k.$$

On the other hand, the maximum likelihood estimator (MLE) of probability vector $\boldsymbol{q}^{(k)}$, where $k \in \{1, ..., K-1\}$, based on $\boldsymbol{X}_{k+1} = (X_{k+1,1}, ..., X_{k+1,m})^T$,..., $\boldsymbol{X}_K = (X_{K1}, ..., X_{Km})^T$, is given by $\widehat{\boldsymbol{q}}_k(K) = (\widehat{q}_{k1}(K), ..., \widehat{q}_{km}(K))^T$, where

$$\widehat{q}_{kj}(K) = \frac{Z_{kj}(K)}{M_k(K)}, \qquad j = 1, ..., m,\ k = 1, ...K-1,$$

being

$$Z_{kj}(K) \equiv X_{k+1,j} + ... + X_{Kj} = Y_{Kj} - Y_{kj} \qquad \text{and} \qquad M_k(K) \equiv n_{k+1} + ... + n_K = N_K - N_k.$$

Note that $\widehat{\boldsymbol{p}}_k$ and $\widehat{\boldsymbol{q}}_k(K)$ are the MLEs of $\boldsymbol{p}_i$ $i \in \{1, ..., k\}$ and $\boldsymbol{p}_i$ $i \in \{k+1, ..., K\}$ respectively under $\mathcal{H}_A^{(k)}(K)$, and $\widehat{\boldsymbol{p}}_K$ is the MLE of $\boldsymbol{p}_i$ $i \in \{1, ..., K\}$, under $\mathcal{H}_0^{(k)}(K)$. Hence, the likelihood ratio of (3) is

$$\Lambda_k(K) = \frac{\prod_{i=1}^{K} \widehat{p}_{K1}^{X_{i1}} \ldots\ldots \widehat{p}_{Km}^{X_{im}}}{\prod_{i=1}^{k} \widehat{p}_{k1}^{X_{i1}} \ldots\ldots \widehat{p}_{km}^{X_{im}} \prod_{i=k+1}^{K} \widehat{q}_{k1}^{X_{i1}}(K) \ldots\ldots \widehat{q}_{km}^{X_{im}}(K)}.$$

We reject $\mathcal{H}_0^{(k)}(K)$ given in (3), i.e. we accept that there is a change point, if the likelihood ratio test-statistic, $T_{k,0}^{(K)} \equiv -2 \log \Lambda_k(K)$, is large enough.

Since $k$ is not known, the test-statistic of interest based on the likelihood test-statistic is

$$Z_0^{(K)} \equiv \max_{k \in \{1,\ldots,K-1\}} T_{k,0}^{(K)} = \max_{k \in \{1,\ldots,K-1\}} \left(-2 \log \Lambda_k(K)\right).$$

We reject $H_0(K)$ given in (1), if $Z_0^{(K)}$ is large enough. Hence, under $\mathcal{H}_A(K)$ we must consider $\widehat{\kappa}(K) = \arg\max_{k \in \{1,\ldots,K-1\}} T_{k,0}^{(K)}$ as estimator of the location where the change occurs.

We can observe that the expression of $T_{k,0}$ can be written by

$$\begin{aligned}
T_{k,0}^{(K)} &= -2 \left( \sum_{i=1}^{K} \sum_{j=1}^{m} X_{ij} \log \widehat{p}_{Kj} - \sum_{i=1}^{k} \sum_{j=1}^{m} X_{ij} \log \widehat{p}_{kj} - \sum_{i=k+1}^{K} \sum_{j=1}^{m} X_{ij} \log \widehat{q}_{kj}(K) \right) \\
&= 2 \left( -\sum_{j=1}^{m} (Y_{kj} + Z_{kj}) \log \widehat{p}_{Kj} + \sum_{j=1}^{m} Y_{kj} \log \widehat{p}_{kj} + \sum_{j=1}^{m} Z_{kj} \log \widehat{q}_{kj}(K) \right) \\
&= 2 \left( \sum_{j=1}^{m} Y_{kj} \log \frac{\widehat{p}_{kj}}{\widehat{p}_{Kj}} + \sum_{j=1}^{m} Z_{kj} \log \frac{\widehat{q}_{kj}(K)}{\widehat{p}_{Kj}} \right) \\
&= 2 \left( N_k \sum_{j=1}^{m} \widehat{p}_{kj} \log \frac{\widehat{p}_{kj}}{\widehat{p}_{Kj}} + M_k \sum_{j=1}^{m} \widehat{q}_{kj} \log \frac{\widehat{q}_{kj}(K)}{\widehat{p}_{Kj}} \right).
\end{aligned}$$

If we consider $2m$-dimensional probability vectors

$$\widehat{\boldsymbol{p}}_{\mathcal{H}_A^{(k)}(K)} \equiv \left( \frac{N_k}{N_K} \widehat{\boldsymbol{p}}_k^T, \frac{M_k}{N_K} \widehat{\boldsymbol{q}}_k^T(K) \right)^T \qquad \text{and} \qquad \widehat{\boldsymbol{p}}_{\mathcal{H}_0^{(k)}(K)} \equiv \left( \frac{N_k}{N_K} \widehat{\boldsymbol{p}}_K^T, \frac{M_k}{N_K} \widehat{\boldsymbol{p}}_K^T \right)^T,$$

we can rewrite

$$\begin{aligned}
T_{k,0}^{(K)} &= 2N_K \left\{ \sum_{j=1}^{m} \frac{N_k}{N_K} \widehat{p}_{Kj} \phi\left( \frac{\widehat{p}_{kj}}{\widehat{p}_{Kj}} \right) + \sum_{j=1}^{m} \frac{M_k}{N_K} \widehat{p}_{Kj} \phi\left( \frac{\widehat{q}_{kj}(K)}{\widehat{p}_{Kj}} \right) \right\} \\
&= 2N_K D_{\text{Kullback}} \left( \widehat{\boldsymbol{p}}_{\mathcal{H}_A^{(k)}(K)}, \widehat{\boldsymbol{p}}_{\mathcal{H}_0^{(k)}(K)} \right),
\end{aligned} \tag{4}$$

where $\phi(x) = x \log x - x + 1$, i.e., test statistic $T_{k,0}^{(K)}$ can be expressed in terms of the Kullback-Leibler divergence between the probability vectors $\widehat{\boldsymbol{p}}_{\mathcal{H}_A^{(k)}(K)}$ and $\widehat{\boldsymbol{p}}_{\mathcal{H}_0^{(k)}(K)}$, multiplied by $2N_K$. Let $\phi : [0, \infty) \longrightarrow \mathbb{R}$ be a generic convex function such that at $x = 1$, $\phi(1) = 0$ at $x = 0$, $0\phi(0/0) = 0$ and $0\phi(p/0) = \lim_{u \to \infty} \phi(u)/u$ (for more details about $\phi$-divergence measures see Pardo (2006)). In this paper we shall consider, for testing $\mathcal{H}_0(K)$ given in (1) against $\mathcal{H}_A(K)$ given in (2), two family of test statistics:

- $Z_{\phi}^{(K)} \equiv \max_{k\in\{1,\dots,K-1\}} T_{k,\phi}^{(K)}$, where

$$T_{k,\phi}^{(K)} = \frac{2N_K}{\phi''(1)} D_{\phi}\left(\widehat{\boldsymbol{p}}_{\mathcal{H}_A^{(k)}(K)}, \widehat{\boldsymbol{p}}_{\mathcal{H}_0^{(k)}(K)}\right) \equiv \frac{2N_K}{\phi''(1)}\left(\sum_{j=1}^{m}\frac{N_k}{N_K}\widehat{p}_{Kj}\phi\left(\frac{\widehat{p}_{kj}}{\widehat{p}_{Kj}}\right) + \sum_{j=1}^{m}\frac{M_k}{N_K}\widehat{p}_{Kj}\phi\left(\frac{\widehat{q}_{kj}(K)}{\widehat{p}_{Kj}}\right)\right). \quad (5)$$

In addition, following the multivariate Darling–Erdös' formula, we shall consider two normalized variants of this family of test-statistics:

$$G_{\phi}(K) = \alpha(\log(K-1))\sqrt{Z_{\phi}^{(K)}} - \beta_{m-1}(\log(K-1)), \quad (6)$$

$$G'_{\phi}(K) = \alpha(\log N_K)\sqrt{Z_{\phi}^{(K)}} - \beta_{m-1}(\log N_K), \quad (7)$$

where $\alpha(x) \equiv (2\log x)^{1/2}$, $0 < x < \infty$, $\beta_d(x) = 2\log x + \frac{d}{2}\log x - \log\Gamma(d/2)$, $0 < x < \infty$, and $\Gamma(y) = \int_0^{\infty} t^{y-1}e^{-t}dt$, $1 \le y < \infty$, is the gamma function.

- $W_{\phi}(K) \equiv \max_{k\in\{1,\dots,K-1\}} \widetilde{T}_{k,\phi}^{(K)}$, where

$$\widetilde{T}_{k,\phi}^{(K)} = \frac{N_k M_k(K)}{N_K^2} T_{k,\phi}^{(K)}. \quad (8)$$

For each possible function $\phi$ a different kind of $\phi$-divergence based test-statistic for (3), $T_{k,\phi}^{(K)}$, is obtained. It is interesting to take into account that the so called Cressie-Read power-divergence test-statistics are a special subfamily depending on parameter $\lambda \in \mathbb{R}$ where $\phi_{\lambda}(x) = (x^{\lambda+1} - x - \lambda(x-1))/(\lambda(1+\lambda))$, for $\lambda(\lambda+1) \neq 0$ and $\phi_{\lambda}(x) = \lim_{\upsilon\to\lambda}\phi_{\upsilon}(x)$, for $\lambda(\lambda+1) = 0$ (see Read and Cressie (1988) for more details). That is, the family of Cressie-Read power-divergence test-statistics for (3) is given by

$$T_{k,\phi_\lambda}^{(K)} = \begin{cases} \dfrac{2}{\lambda(\lambda+1)}\left(\left(\dfrac{N_k}{N_K}\right)^{\lambda+1}\displaystyle\sum_{j=1}^{m}\frac{Y_{Kj}^{\lambda+1}}{Y_{kj}^{\lambda}} + \left(\dfrac{M_k(K)}{N_K}\right)^{\lambda+1}\sum_{j=1}^{m}\frac{Y_{Kj}^{\lambda+1}}{Z_{kj}^{\lambda}(K)} - N_K\right), & \lambda(\lambda+1)\neq 0,\\ 2\left(\displaystyle\sum_{j=1}^{m}Y_{kj}\log\left(\frac{N_K Y_{kj}}{N_k Y_{Kj}}\right) + \sum_{j=1}^{m}Z_{kj}(K)\log\left(\frac{N_K Z_{kj}(K)}{M_k(K)Y_{Kj}}\right)\right), & \lambda = 0,\\ 2\left(\dfrac{N_k}{N_K}\displaystyle\sum_{j=1}^{m}Y_{Kj}\log\left(\frac{N_k Y_{Kj}}{N_K Y_{kj}}\right) + \frac{M_k(K)}{N_K}\sum_{j=1}^{m}Y_{Kj}\log\left(\frac{N_K Y_{Kj}}{N_K Z_{kj}(K)}\right)\right), & \lambda = -1. \end{cases} \quad (9)$$

In particular, with $\lambda = 0$, the likelihood ratio test-statistic $T_{k,\phi_0}^{(K)} = T_{k,0}^{(K)} = -2\log\Lambda_k(K)$ and with $\lambda = 1$, the chi-square test-statistic

$$T_{k,\phi_1}^{(K)} = \left(\frac{N_k}{N_K}\right)^2\sum_{j=1}^{m}\frac{Y_{Kj}^2}{Y_{kj}} + \left(\frac{M_k(K)}{N_K}\right)^2\sum_{j=1}^{m}\frac{Y_{Kj}^2}{Z_{kj}(K)} - N_K, \quad (10)$$

are obtained. Note that $G'_{\phi_\lambda}(K)$ and $W_{\phi_\lambda}(K)$, with $\lambda \in \{0,1\}$, were analyzed in Horváth and Serbinowska (1995), and thus we have covered them inside a broad family of test-statistics. All the members of the normalized family of test-statistics $G_{\phi_\lambda}(K)$, for any $\lambda \in \mathbb{R}$, are totally new.

Note that for the same $\phi$ function the estimator of the location of the change point under $\mathcal{H}_A(K)$, $\widehat{\kappa}_{\phi}(K) = \arg\max_{k\in\{1,\dots,K-1\}} T_{k,\phi}^{(K)}$, is the same either for $G_{\phi}(K)$ or $G'_{\phi}(K)$, but it does not necessarily coincide with the location of the change point for $W_{\phi}(K)$, $\widetilde{\kappa}_{\phi}(K) = \arg\max_{k\in\{1,\dots,K-1\}} \widetilde{T}_{k,\phi}^{(K)}$.

The following section is devoted to find the asymptotic distribution of $G_{\phi}(K)$, $G'_{\phi}(K)$ and $W_{\phi}(K)$.

# 3 Main results

In this section we are going to provide the asymptotic distributions of the test-statistics we have just proposed. Their proofs follow a different line in comparison with Horvárth and Serbinowska (1995), where some of these results were proven.

**Lemma 1** *If we assume that for each $K$ there exists an unknown constant $\lambda_k^{(K)}$, such that $\lambda_k^{(K)} = \lim_{N_k\to\infty} \frac{N_k}{N_K}$, then the asymptotic distribution of $T_{k,\phi}^{(K)}$ is the same, that is as $N_k \to \infty$*

$$T_{k,\phi}^{(K)} = T_{k,0}^{(K)} + o_P(1).$$

**Proof.** It is a particular case of Theorem 2.1 in Pardo et al. (1999) with $\nu = 2$. ■

The following Lemma shows that the chi-square test-statistic $T_{k,\phi_1}^{(K)}$ have different expression, and in particular coincides with expression (1.3) given in Horváth and Sebinowska.

**Lemma 2** *The chi-square test-statistic given in (10) have these alternative expressions*

$$T_{k,\phi_1}^{(K)} = N_k \sum_{j=1}^{m} \frac{1}{\widehat{p}_{Kj}} \left(\widehat{p}_{kj} - \widehat{p}_{Kj}\right) + M_k \sum_{j=1}^{m} \frac{1}{\widehat{p}_{Kj}} \left(\widehat{q}_{kj}(K) - \widehat{p}_{Kj}\right) \tag{11}$$

$$T_{k,\phi_1}^{(K)} = \sum_{j=1}^{m} \frac{\left(Y_{kj} - Y_{Kj}\frac{N_k}{N_K}\right)^2}{Y_{Kj}\frac{M_k N_k}{N_K^2}} \tag{12}$$

**Proof.** Formula (11) is a direct application of function $\phi(x) = \frac{1}{2}(x-1)^2$ in (5). From formula (11) we shall derive (12): $T_{k,\phi_1}^{(K)} = A_k + B_k$, with

$$\begin{aligned} A_k &= N_k \sum_{j=1}^{m} \frac{1}{\widehat{p}_{Kj}} \left(\widehat{p}_{kj} - \widehat{p}_{Kj}\right) = N_k \sum_{j=1}^{m} \frac{N_K}{N_k^2 N_K^2} \frac{1}{Y_{Kj}} \left(Y_{kj} N_K - Y_{Kj} N_k\right)^2 \\ &= N_k \sum_{j=1}^{m} \frac{N_K}{N_k^2 N_K^2} N_K^2 \frac{1}{Y_{Kj}} \left(Y_{kj} - Y_{Kj}\frac{N_k}{N_K}\right)^2 = \sum_{j=1}^{m} \frac{1}{N_k} \frac{N_K}{Y_{Kj}} \left(Y_{kj} - Y_{Kj}\frac{N_k}{N_K}\right)^2, \end{aligned}$$

$$\begin{aligned} B_k &= M_k \sum_{j=1}^{m} \frac{1}{\widehat{p}_{K,j}} \left(\widehat{q}_{k,j}(K) - \widehat{p}_{K,j}\right)^2 = M_k \sum_{j=1}^{m} \frac{N_K}{Y_{Kj}} \left(\frac{Y_{Kj} - Y_{kj}}{M_k} - \frac{Y_{Kj}}{N_K}\right)^2 \\ &= M_k \sum_{j=1}^{m} \frac{N_K}{Y_{Kj}} \left(Y_{Kj}\left(\frac{1}{M_k} - \frac{1}{N_K}\right) - \frac{Y_{kj}}{M_k}\right)^2 = M_k \sum_{j=1}^{m} \frac{N_K}{Y_{Kj}} \left(Y_{Kj}\frac{N_k}{M_k N_K} - \frac{Y_{kj}}{M_k}\right)^2 \\ &= \sum_{j=1}^{m} \frac{1}{M_k} \frac{N_K}{Y_{Kj}} \left(Y_{kj} - Y_{Kj}\frac{N_k}{N_K}\right)^2. \end{aligned}$$

Therefore, because $T_{k,\phi_1}^{(K)} = \sum_{j=1}^{m} \left(\frac{1}{N_k} + \frac{1}{M_k}\right) \frac{N_K}{Y_{Kj}} \left(Y_{kj} - Y_{Kj}\frac{N_k}{N_K}\right)^2$, and $\frac{1}{N_k} + \frac{1}{M_k} = \frac{N_K}{N_k M_k}$, we can see that (11) and (12) are equal. ■

Let $\mathcal{G}$ be the Extreme Value distribution with parameters $\mu = \log 2$, $\beta = 1$, i.e.

$$\Pr(\mathcal{G} \le x) = \exp\{-e^{-(x-\mu)/\beta}\}, \tag{13}$$

and if we consider $\boldsymbol{W}_0^{(m-1)}(t) = \{(W_{0,1}(t), ..., W_{0,m-1}(t))\}_{t\in[0,1]}$, $i = 1, ..., m-1$ being an $(m-1)$-dimensional vector of independent Brownian bridges we define

$$\mathcal{W} \equiv \sup_{t\in[0,1]} \left\| \boldsymbol{W}_0^{(m-1)}(t) \right\|^2, \tag{14}$$

where $\|\bullet\|$ is the Euclidean norm, i.e. $\left\|\boldsymbol{W}_0^{(m-1)}(t)\right\|^2 = \sum_{i=1}^{m-1} W_{0,i}^2(t)$.

In the following three lemmas we are going to focuss directly on the test statistics based on likelihood ratio test. In some cases we are going to deal also with the chi-square test-statistics, indirectly, because it is much easier to work with it.

**Lemma 3** *If $\mathcal{H}_0(K)$, given in (1), holds then we have*

1. *For $n_i = 1$, $i = 1, ..., K$,*
$$\alpha(\log K)\sqrt{Z_0^{(K)}} - \beta_{m-1}(\log K) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{G},$$
2. *For $n_i = n$, $i = 1, ..., K$,*
$$\alpha(\log(nK))\sqrt{Z_0^{(K)}} - \beta_{m-1}(\log(nK)) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{G},$$
3. *For general values of $n_i$, $i = 1, ..., K$,*
$$\alpha(\log(N_K))\sqrt{Z_0^{(K)}} - \beta_{m-1}(\log(N_K)) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{G},$$
*when there is a monotone, continuous function $g$ such that $g(0) = 0$ and*
$$\lim_{K\to\infty} \max_{k\in\{1,...,K\}} \left|\frac{N_k}{N_K} - g\left(\frac{k}{K}\right)\right| = 0 \text{ and } \lim_{x\to 0^+} \frac{\log(\log x)}{\log(\log g(x))} = 1. \qquad (15)$$

**Proof.** Part 1 is a direct aplication of Theorem 2.1 in Gombay and Horváth. Note that the dimension of unknown parameters is $m-1$. For part 2 we shall apply Theorem 2.1,

$$\alpha(\log(K))\sqrt{Z_0^{(K)}} - \beta_{m-1}(\log(K)) \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{G}$$

but also an additional remark is needed: in virtue of Slatsky's Theorem $\alpha(\log(K))$ and $\beta_{m-1}(\log(K))$ can be replaced for $\alpha(h(K))$ and $\beta_{m-1}(h(K))$ such that

$$\lim_{K\to\infty} \frac{\alpha(\log(K))}{\alpha(h(K))} = 1 \qquad \text{and} \qquad \lim_{K\to\infty} \frac{\beta_{m-1}(\log(K))}{\beta_{m-1}(h(K))} = 1. \qquad (16)$$

The first condition is verified because $h(K) = N_K = N_1 g(\frac{1}{K})$ and $\lim_{K\to\infty} \frac{\alpha(\log(K))}{\alpha(h(K))} = \lim_{x\to 0^+} \frac{\log(\log x)}{\log(\log g(x))} = 1$ (second one is very similar). With respect to part 3 $\alpha(\log(K))$ and $\beta_{m-1}(\log(K))$ can be replaced by $\alpha(\log(N_K))$ and $\beta_{m-1}(\log(N_K))$ for the same reason as in part 2. Now, we would like to approximate $Z_0^{(K)}$ by $\max_{t\in[0,1]} \frac{1}{t(1-t)} \left\|\boldsymbol{W}_0^{(m-1)}(t)\right\|^2$, with $\boldsymbol{W}_0^{(m-1)}(t)$ being an $(m-1)$-dimensional vector of independent Brownian bridges. From Lemma 1 we can approximate $\sqrt{Z_0^{(K)}}$ by $\sqrt{Z_{\phi_{(1)}}^{(K)}} = \sqrt{\max_{k\in\{1,...,K-1\}} T_{k,\phi_{(1)}}^{(K)}}$ where for $T_{k,\phi_{(1)}}^{(K)}$ use (12),

$$T_{k,\phi_{(1)}}^{(K)} = \sum_{j=1}^{m} \frac{\left(Y_{kj} - \frac{N_k}{N_K} Y_{Kj}\right)^2}{Y_{Kj} \frac{N_k M_k}{N_K^2}} = (\Psi_k^{(K)})^T \Psi_k^{(K)},$$
$$\Psi_k^{(K)} = \left(\frac{Y_{kj} - \frac{N_k}{N_K} Y_{Kj}}{\sqrt{Y_{Kj} \frac{N_k M_k}{N_K^2}}}\right)_{j=1,...,m} \underset{K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(\boldsymbol{0}_m, \boldsymbol{I}_m - \boldsymbol{p}_K^{-\frac{1}{2}} (\boldsymbol{p}_K^{-\frac{1}{2}})^T), \text{ under } \mathcal{H}_0(K).$$

Let $\boldsymbol{G}$ be a orthogonal matrix such that $\boldsymbol{I}_m - \boldsymbol{p}_K^{-\frac{1}{2}}(\boldsymbol{p}_K^{-\frac{1}{2}})^T = \boldsymbol{G}^T\boldsymbol{G}$. Taking into account that the eigenvalues of $\boldsymbol{G}$ are 1 with multiplicity $m-1$ and 0 with multiplicity 1, we can split this matrix as $\boldsymbol{G} = (\widetilde{\boldsymbol{G}}|\widetilde{\boldsymbol{g}})$, where $\widetilde{\boldsymbol{G}} = (\boldsymbol{g}_1, ..., \boldsymbol{g}_{m-1})$ is a $m \times (m-1)$ matrix composed by the orthonormal eigenvectors associated to eigenvalue 1. It is not difficult to see that

$$\widetilde{\boldsymbol{G}}^T \Psi_k^{(K)} \underset{N_K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(\boldsymbol{0}_{m-1}, \boldsymbol{I}_{m-1}), \text{ under } \mathcal{H}_0(K).$$

Now the $s$-th component of $\widetilde{\boldsymbol{G}}^T \Psi_k^{(K)}$, $\boldsymbol{g}_s^T \Psi_k^{(K)}$, $s = 1, ..., m-1$, is given by

$$\sum_{h=1}^{m} \frac{g_{sh}Y_{ks} - g_{sh}\frac{N_k}{N_K}Y_{Kh}}{\sqrt{Y_{Kh}\frac{N_k M_k}{N_K^2}}} = \frac{\sum_{h=1}^{m}\frac{g_{sh}Y_{kh}}{\sqrt{Y_{Kh}}} - \frac{N_k}{N_K}\sum_{h=1}^{m}\frac{g_{sh}Y_{Kh}}{\sqrt{Y_{Kh}}}}{\sqrt{\frac{N_k M_k}{N_K^2}}}$$
$$= \frac{1}{\sqrt{\frac{N_k M_k}{N_K^2}}}\left(\frac{1}{\sqrt{N_K}}\sum_{h=1}^{m} g_{sh}\frac{Y_{kh} - N_k p_{Kh}}{\sqrt{Y_{Kh}/N_K}} - \frac{N_k}{N_K}\frac{1}{\sqrt{N_K}}\sum_{h=1}^{m} g_{sh}\frac{Y_{Kh} - N_K p_{Kh}}{\sqrt{Y_{Kh}/N_K}}\right),$$

where $\sum_{h=1}^{m} \frac{g_{sh}(Y_{kh} - N_k p_{Kh})}{\sqrt{Y_{Kh}}} \underset{N_K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}(0, \lambda_k^{(K)})$, with $\lambda_k^{(K)} \equiv \lim_{N_k\to\infty} \frac{N_k}{N_K}$ and $\boldsymbol{g}_s^T = (g_{s1}, ..., g_{sm})$. Note that

$$\frac{1}{\sqrt{N_K}}\frac{Y_{kh} - N_k p_{Kh}}{\sqrt{Y_{Kh}/N_K}} = \sqrt{\frac{N_k}{N_K}}\frac{1}{\sqrt{N_k}}\frac{Y_{kh} - N_k p_{Kh}}{\sqrt{Y_{Kh}/N_K}} \underset{N_k\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} W_s(\lambda_k^{(K)}), \text{ under } \mathcal{H}_0(K),$$

where $\{W_s(t)\}_{t\geq 0}$ is a standard Brownian motion process associated to the the $s$-th dimension ($s = 1, ..., m-1$), and

$$\frac{1}{\sqrt{N_K}}\sum_{h=1}^{m} g_{sh}\frac{Y_{kh} - N_k p_{Kh}}{\sqrt{Y_{Kh}/N_K}} = \sqrt{\frac{N_k}{N_K}}\frac{1}{\sqrt{N_k}}\sum_{h=1}^{m} g_{sh}\frac{Y_{kh} - N_k p_{Kh}}{\sqrt{Y_{Kh}/N_K}} \underset{N_k\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} W_s(\lambda_k^{(K)}), \text{ under } \mathcal{H}_0(K),$$

are mutually independent for each dimension. Since $W_s(\lambda_k^{(K)}) - \lambda_k^{(K)} W_s(1)$, $s = 1, ..., m-1$ is a Brownian bridge, we denote it by $W_{0,s}(\lambda_k^{(K)})$ and we obtain

$$\frac{1}{\sqrt{\frac{N_k M_k}{N_K^2}}}\left(\sum_{h=1}^{m} g_{sh}\frac{Y_{kh}}{\sqrt{Y_{Kh}}} - \frac{N_k}{N_K}\sum_{h=1}^{m} g_{sh}\frac{Y_{Kh}}{\sqrt{Y_{Kh}}}\right) \underset{N_K\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \frac{W_{0,s}(\lambda_k^{(K)})}{\sqrt{\lambda_k^{(K)}(1-\lambda_k^{(K)})}}, \text{ under } \mathcal{H}_0(K).$$

From the almost sure continuity of the Brownian bridge, this means that with general values of $n_i$, $i = 1, ..., K$, test statistic

$$\max_{k\in\{1,...,K-1\}} T_{k,\phi_{(1)}}^{(K)} = \max_{k\in\{1,...,K-1\}} \left(\widetilde{\boldsymbol{G}}^T \Psi_k^{(K)}\right)^T \widetilde{\boldsymbol{G}}^T \Psi_k^{(K)}$$
$$= \max_{t\in\left\{\frac{N_k}{N_K}\right\}_{k=1}^{K-1}} \frac{1}{t(1-t)}\sum_{s=1}^{m-1}\left(\sum_{h=1}^{m} g_{sh}\frac{Y_{kh}}{\sqrt{Y_{Kh}}} - \frac{N_k}{N_K}\sum_{h=1}^{m} g_{sh}\frac{Y_{Kh}}{\sqrt{Y_{Kh}}}\right)^2 = \max_{t\in\left\{\frac{N_k}{N_K}\right\}_{k=1}^{K-1}} \frac{1}{t(1-t)}\left\|\boldsymbol{W}_0^{(m-1)}(t)\right\|^2$$
$$= \max_{t\in\left\{\frac{k}{K}\right\}_{k=1}^{K-1}} \frac{1}{g(t)(1-g(t))}\left\|\boldsymbol{W}_0^{(m-1)}(g(t))\right\|^2$$

must have the same asymptotic distribution as test statistic

$$\max_{k\in\{1,...,K-1\}} T_{k,\phi_{(1)}}^{(K)} = \max_{t\in\left\{\frac{k}{K}\right\}_{k=1}^{K-1}} \frac{1}{t(1-t)}\left\|\boldsymbol{W}_0^{(m-1)}(g(t))\right\|^2, \tag{17}$$

with general values of $n_i$, $i = 1, ..., K$. Taking into account that (17) is the basic structure for obtaining the asymptotic distribution of $\max_{k\in\{1,...,K-1\}} T_{k,\phi_{(1)}}^{(K)}$ (see Gombay and Horváth (1996), or Yao and Davis (1986) for $m = 2$), the desired result is obtained. ■

For the following result, when the values of $n_i$, $i = 1, ..., K$ are not equal, it is assumed that there is a monotone, continuous function $g$ such that $g(0) = 0$ and $\lim_{K\to\infty} \max_{k\in\{1,...,K\}} \left| \frac{N_k}{N_K} - g\left(\frac{k}{K}\right) \right| = 0$.

**Lemma 4** *If $\mathcal{H}_0(K)$, given in (1), holds then we have*

$$\alpha(\log(K-1))\sqrt{Z_0^{(K)}} - \beta_{m-1}(\log(K-1)) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{G}.$$

**Proof.** It is an immediate consequence of the previous Lemma if we consider the remark made in the previous proof regarding functions $\alpha$ and $\beta_{m-1}$, if we take $h(K) = K - 1$. ■

**Lemma 5** *f $\mathcal{H}_0(K)$, given in (1), holds and there is a monotone, continuous function $g$ such that $g(0) = 0$ and $\lim_{K\to\infty} \max_{k\in\{1,...,K\}} \left| \frac{N_k}{N_K} - g\left(\frac{k}{K}\right) \right| = 0$, then we have*

$$W_0(K) = \max_{k\in\{1,...,K-1\}} \frac{N_k M_k(K)}{N_K^2} T_{k,0}^{(K)} \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{W}.$$

**Proof.** See Theorem 1.2 in Horváth and Serbinowska (1995). ■

**Theorem 6** *If $\mathcal{H}_0(K)$, given in (1), holds and there is a monotone, continuous function $g$ such that $g(0) = 0$ and $\lim_{K\to\infty} \max_{k\in\{1,...,K\}} \left| \frac{N_k}{N_K} - g\left(\frac{k}{K}\right) \right| = 0$ and (15), then we have*

$$G_\phi(K) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{G} \quad \text{and} \quad G'_\phi(K) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{G},$$
$$W_\phi(K) \underset{K\to\infty}{\xrightarrow{\mathcal{L}}} \mathcal{W}.$$

**Proof.** The result is obtained as consequence of applying Lemma 1 to the previous three lemmas. ■

# 4 Monte Carlo study

In this section a study is performed to compare the approximation of the limit distribution to the null distribution of nine test-statistics inside the the families of test-statistics proposed in Section 3,

$$G_{\phi_\lambda}(K) = \alpha(\log(K-1)) \left( \max_{k\in\{1,...,K-1\}} T_{k,\phi_\lambda}^{(K)} \right)^{\frac{1}{2}} - \beta_{m-1}(\log(K-1)),$$
$$G'_{\phi_\lambda}(K) = \alpha(\log N_K) \left( \max_{k\in\{1,...,K-1\}} T_{k,\phi_\lambda}^{(K)} \right)^{\frac{1}{2}} - \beta_{m-1}(\log N_K),$$
$$W_{\phi_\lambda}(K) = \max_{k\in\{1,...,K-1\}} \frac{N_k M_k(K)}{N_K^2} T_{k,\phi_\lambda}^{(K)},$$

($\alpha$ and $\beta_d$ functions were defined in page 4) with $\lambda \in \{0, 1, 2\}$, i.e.

$$T_{k,\phi_0}^{(K)} = 2\left(\sum_{j=1}^{m} Y_{kj} \log\left(\frac{N_K Y_{kj}}{N_k Y_{Kj}}\right) + \sum_{j=1}^{m} Z_{kj}(K) \log\left(\frac{N_K Z_{kj}(K)}{M_k(K) Y_{Kj}}\right)\right),$$

$$T_{k,\phi_1}^{(K)} = \left(\frac{N_k}{N_K}\right)^2 \sum_{j=1}^{m} \frac{Y_{Kj}^2}{Y_{kj}} + \left(\frac{M_k(K)}{N_K}\right)^2 \sum_{j=1}^{m} \frac{Y_{Kj}^2}{Z_{kj}(K)} - N_K,$$

$$T_{k,\phi_2}^{(K)} = \frac{1}{3}\left(\left(\frac{N_k}{N_K}\right)^3 \sum_{j=1}^{m} \frac{Y_{Kj}^3}{Y_{kj}^2} + \left(\frac{M_k(K)}{N_K}\right)^3 \sum_{j=1}^{m} \frac{Y_{Kj}^3}{Z_{kj}^2(K)} - N_K\right).$$

Their asymptotic distribution, dealt in Section 4, is the same for different values of $\lambda \in \{0, 1, 2\}$. The approximated distribution function of $G_{\phi_\lambda}(K)$ and $G'_{\phi_\lambda}(K)$ can be considered to be (13) for $K$ large enough, and the approximated distribution function of $W_{\phi_\lambda}(K)$ can be found in (3.21) of Kiefer (1959), as well as its tabulation (see tables 1 and 2 in Kiefer (1959)).

For the simulation study we have taken sequences of length $K \in \{64, 300, 500\}$, and it is considered a sequence of multinomial distribution. Dimension $m = 3$ is considered for the figures and $m \in \{2, 3\}$ for results summarized in the tables. Since this paper's results are valid for different values of $n_i$, $i = 1, ..., K$, we have chosen $n_1 = \ldots = n_{[K/2]} = 28$ and $n_{[K/2]+1} = \ldots = n_K = 48$ ($[\bullet]$ is the integer part function). All the results are based on 5000 replications of the experiment that were designed by the authors in FORTRAN. In Figure 1 it is shown that the behaviour of $G_{\phi_\lambda}(K)$ tends to approximate the asymptotic distribution, $\mathcal{G}$, much better than $G'_{\phi_\lambda}(K)$ (this figure is for $\lambda = 2$, $K = 300$, but it happens the same for $\lambda \in \{0, 1\}$ and $K \in \{64, 500\}$). Looking at Figure 2 we can see that the approximation of $G_{\phi_\lambda}(K)$ to the limit distribution seems to be at least so good as $W_{\phi_\lambda}(K)$. In Figure 3 the empirical distribution functions based on the likelihood ratio test-statistic ($\lambda = 0$), chi-square test-statistic ($\lambda = 1$) and $T_{k,\phi_\lambda}^{(K)}$ with $\lambda = 2$, are shown. For other values of $n_i$, $K$, $\lambda$ that are omitted it was concluded that $G_{\phi_\lambda}(K)$, with $\lambda = 2$, has the best approximation to the limit distribution. For $W_{\phi_\lambda}(K)$ the difference between the distribution functions with different values of $\lambda$ is very small, this is why we have omitted its corresponding figure. In figures 4 and 5, how the limit distribution is reached as $K$ is increased is shown for $\lambda = 2$. The behaviour for test statistics $G_{\phi_\lambda}(K)$ seems to be at least as good as for $W_{\phi_\lambda}(K)$. Exclusively with $m = 3$, these good performance is also repeated in the results shown in tables for empirical quantiles and type I error with $\alpha = 0.05$ nominal size when the cutoff is coming from the asymptotic distribution. But with $m = 2$ it seems that the approximation to the limit distribution is a slightly better for $W_{\phi_\lambda}(K)$ than for $G_{\phi_\lambda}(K)$. Comparing the values of $\lambda$, for both dimension sizes it can be seen that the quantiles and type I error of $G_{\phi_\lambda}(K)$ are specially good approximated with $\lambda = 2$, and there is also a small improvement for $W_{\phi_\lambda}(K)$. The approximation of $G'_{\phi_\lambda}(K)$ and $G_{\phi_\lambda}(K)$ with $\lambda = 0$ is quite bad, this fact coincides with the results shown in Horváth in Serbinowska (1995) where it was concludes that $W_{\phi_\lambda}(K)$ was much better than $G'_{\phi_\lambda}(K)$ with $\lambda = 0$. As consequence of this simulation study we recommend for sequences of binomial distributions ($m = 2$) to use $W_{\phi_\lambda}(K)$ based on either the likelihood ratio test-statistic, chi-square or $T_{k,\phi_\lambda}^{(K)}$ with $\lambda = 2$ and $G_{\phi_\lambda}(K)$ for sequences of trinomial distributions ($m = 3$) to use $W_{\phi_\lambda}(K)$ based on $T_{k,\phi_\lambda}^{(K)}$ with $\lambda = 2$.

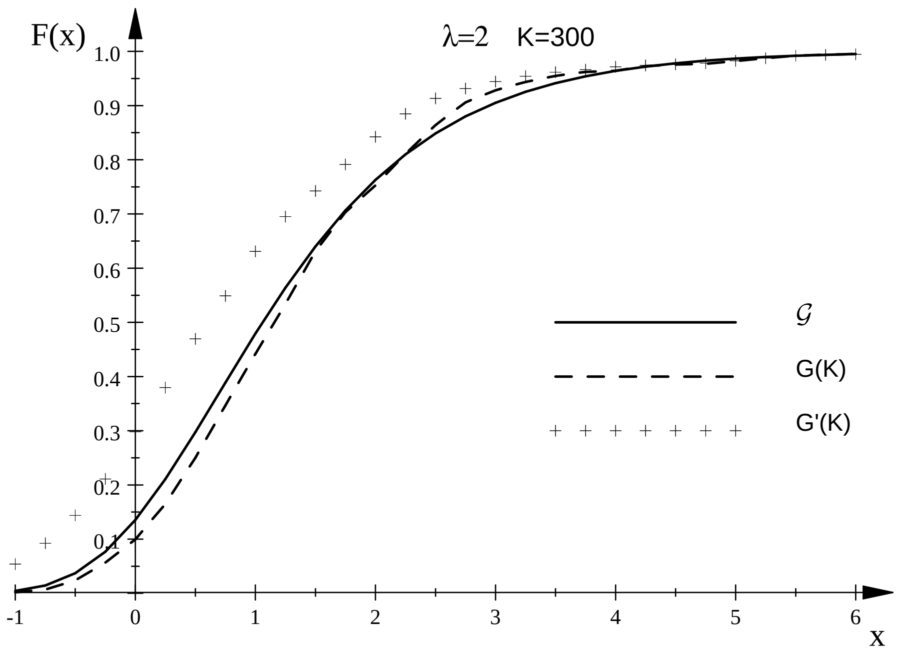



Figure 1: Empirical distribution function of $G_{\phi_\lambda}(K)$, $G'_{\phi_\lambda}(K)$, and limit distribution function ($\mathcal{G}$).

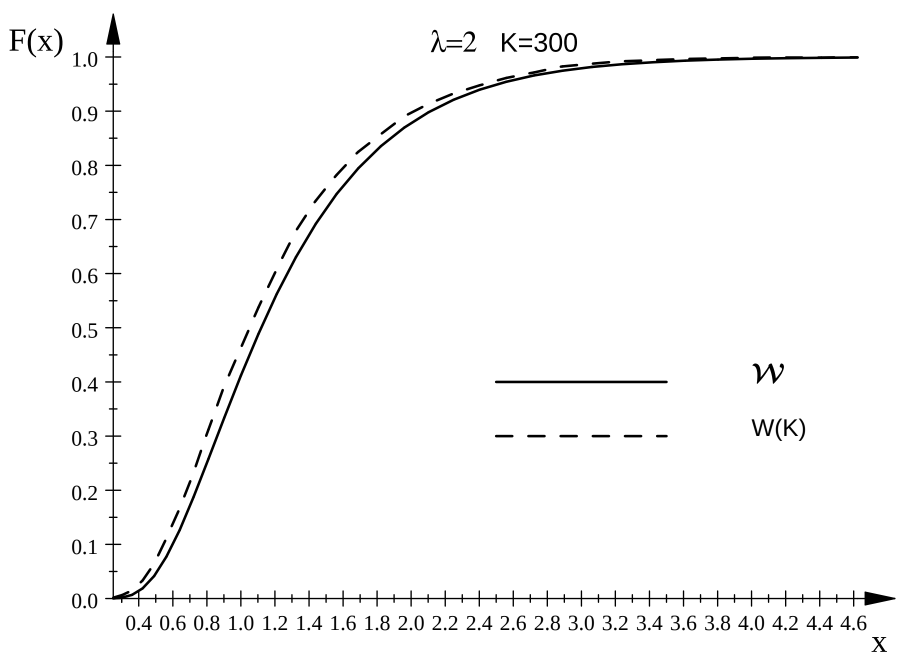



Figure 2: Empirical distribution function of $W_{\phi_\lambda}(K)$ and limit distribution function ($\mathcal{W}$).

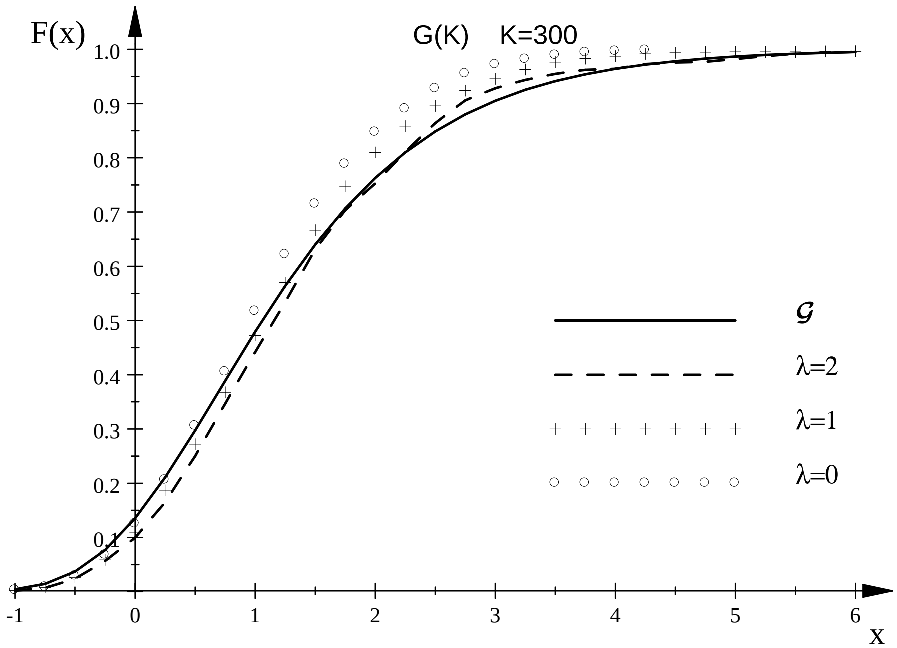



Figure 3: Empirical distribution function of $G_{\phi_\lambda}(K)$ for $\lambda \in \{0, 1, 2\}$, and limit distribution function $(\mathcal{G})$.

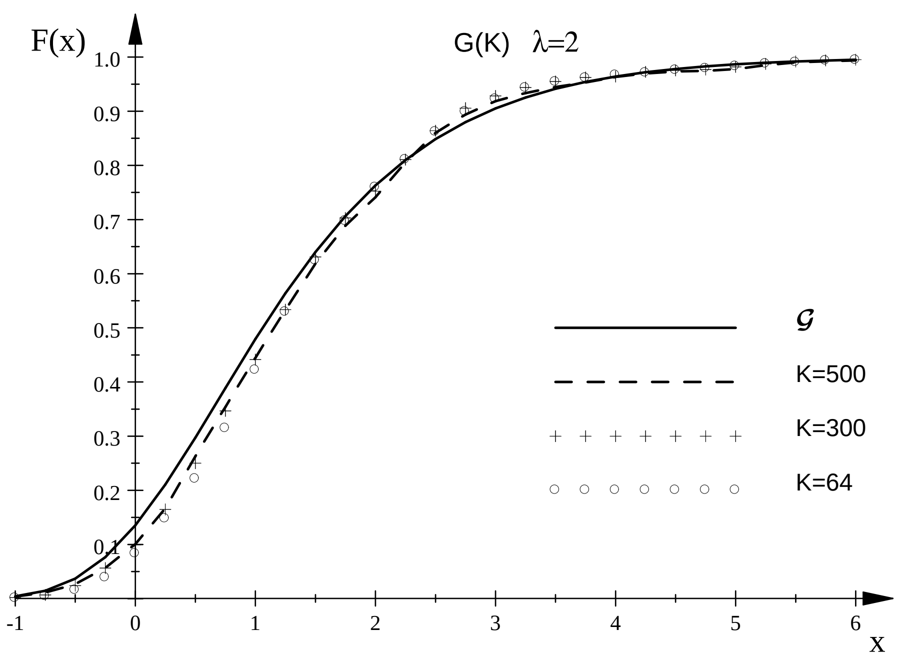



Figure 4: Empirical distribution function of $G_{\phi_\lambda}(K)$ for sequence legths $K \in \{64, 300, 500\}$, and limit distribution function $(\mathcal{G})$.

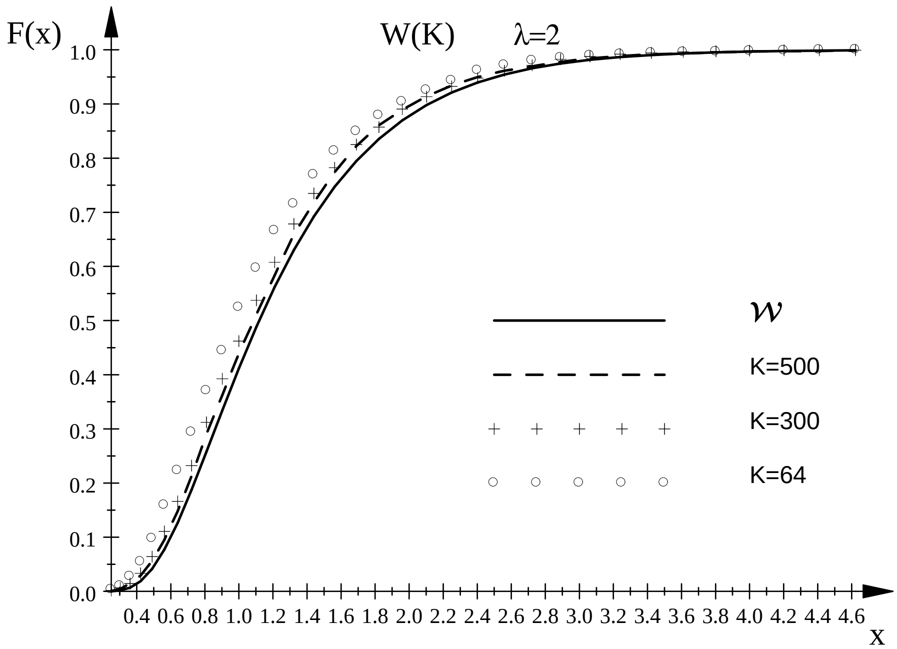



Figure 5: Empirical distribution function of $W_{\phi_\lambda}(K)$ for sequence legths $K \in \{64, 300, 500\}$, and limit distribution function ($\mathcal{W}$).

Table 1: Empirical quantiles and sizes for $G_{\phi_\lambda}(K)$, $G'_{\phi_\lambda}(K)$ and $W_{\phi_\lambda}(K)$ with $\lambda = 0$.

| | | | $K=64$ | | $K=300$ | | $K=500$ | | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|
| | | $1-\alpha$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ |
| $m=2$ | $G_{\phi_0}(K)$ | 0.90 | 2.214 | 0.0330 | 2.320 | 0.0382 | 2.362 | 0.0442 | 2.943 |
| | | 0.95 | 2.689 | 0.0096 | 2.754 | 0.0136 | 2.848 | 0.0140 | 3.663 |
| | | 0.99 | 3.618 | 0.0002 | 3.832 | 0.0002 | 3.957 | 0 | 5.293 |
| | $W_{\phi_0}(K)$ | 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 |
| | $G'_{\phi_0}(K)$ | 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 | 5.293 |
| $m=3$ | $G_{\phi_0}(K)$ | 0.90 | 2.240 | 0.0278 | 2.307 | 0.0312 | 2.343 | 0.0408 | 2.943 |
| | | 0.95 | 2.615 | 0.0074 | 2.701 | 0.0078 | 2.813 | 0.0114 | 3.663 |
| | | 0.99 | 3.464 | 0 | 3.546 | 0 | 3.705 | 0.0002 | 5.293 |
| | $W_{\phi_0}(K)$ | 0.90 | 1.900 | 0.0652 | 2.008 | 0.0840 | 2.010 | 0.0838 | 2.114 |
| | | 0.95 | 2.267 | 0.0292 | 2.419 | 0.0424 | 2.400 | 0.0410 | 2.508 |
| | | 0.99 | 3.109 | 0.0056 | 3.151 | 0.0060 | 3.312 | 0.0090 | 3.396 |
| | $G'_{\phi_0}(K)$ | 0.90 | 1.705 | 0.0154 | 1.913 | 0.0192 | 1.981 | 0.0270 | 2.943 |
| | | 0.95 | 2.156 | 0.0038 | 2.359 | 0.0042 | 2.508 | 0.0078 | 3.663 |
| | | 0.99 | 3.177 | 0 | 3.317 | 0 | 3.506 | 0.0002 | 5.293 |

Table 2: Empirical quantiles and sizes for $G_{\phi_\lambda}(K)$, $G'_{\phi_\lambda}(K)$ and $W_{\phi_\lambda}(K)$ with $\lambda = 1$.

| | | | $K = 64$ | | $K = 300$ | | $K = 500$ | | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|
| | | $1-\alpha$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ |
| $m = 2$ | $G_{\phi_1}(K)$ | 0.90 | 2.312 | 0.0416 | 2.425 | 0.0436 | 2.462 | 0.0506 | 2.943 |
| | | 0.95 | 2.825 | 0.0146 | 2.865 | 0.0182 | 2.952 | 0.0174 | 3.663 |
| | | 0.99 | 3.949 | 0.0008 | 3.968 | 0.0008 | 4.144 | 0.0010 | 5.293 |
| | $W_{\phi_1}(K)$ | 0.90 | 1.305 | 0.0666 | 1.388 | 0.0788 | 1.420 | 0.0860 | 1.498 |
| | | 0.95 | 1.628 | 0.0328 | 1.714 | 0.0372 | 1.741 | 0.0400 | 1.844 |
| | | 0.99 | 2.609 | 0.0094 | 2.485 | 0.0074 | 2.532 | 0.0074 | 2.649 |
| | $G'_{\phi_1}(K)$ | 0.90 | 1.825 | 0.0274 | 2.058 | 0.0318 | 2.123 | 0.0350 | 2.943 |
| | | 0.95 | 2.441 | 0.0116 | 2.558 | 0.0130 | 2.671 | 0.0146 | 3.663 |
| | | 0.99 | 3.793 | 0.0001 | 3.806 | 0.0008 | 4.005 | 0.0010 | 5.293 |
| $m = 3$ | $G_{\phi_1}(K)$ | 0.90 | 2.490 | 0.0582 | 2.523 | 0.0596 | 2.617 | 0.0676 | 2.943 |
| | | 0.95 | 3.081 | 0.0208 | 3.059 | 0.0188 | 3.176 | 0.0254 | 3.663 |
| | | 0.99 | 4.142 | 0.0044 | 4.179 | 0.0046 | 4.257 | 0.0052 | 5.293 |
| | $W_{\phi_1}(K)$ | 0.90 | 1.920 | 0.0688 | 2.010 | 0.0844 | 2.012 | 0.0840 | 2.114 |
| | | 0.95 | 2.281 | 0.0306 | 2.418 | 0.0426 | 2.404 | 0.0412 | 2.508 |
| | | 0.99 | 3.113 | 0.0060 | 3.159 | 0.0062 | 3.311 | 0.0088 | 3.396 |
| | $G'_{\phi_1}(K)$ | 0.90 | 2.006 | 0.0382 | 2.158 | 0.0390 | 2.280 | 0.0486 | 2.943 |
| | | 0.95 | 2.716 | 0.0152 | 2.766 | 0.0148 | 2.914 | 0.0196 | 3.663 |
| | | 0.99 | 3.992 | 0.0048 | 4.034 | 0.0046 | 4.123 | 0.0052 | 5.293 |

Table 3: Empirical quantiles and sizes for $G_{\phi_\lambda}(K)$, $G'_{\phi_\lambda}(K)$ and $W_{\phi_\lambda}(K)$ with $\lambda = 2$.

| | | | $K = 64$ | | $K = 300$ | | $K = 500$ | | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|
| | | $1-\alpha$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ | $\widehat{\alpha}$ | $x_{1-\alpha}$ |
| $m = 2$ | $G_{\phi_2}(K)$ | 0.90 | 2.393 | 0.0542 | 2.482 | 0.0572 | 2.531 | 0.0632 | 2.943 |
| | | 0.95 | 3.018 | 0.0188 | 3.040 | 0.0214 | 3.0708 | 0.0210 | 3.663 |
| | | 0.99 | 4.214 | 0.0022 | 4.422 | 0.0026 | 4.539 | 0.0030 | 5.293 |
| | $W_{\phi_2}(K)$ | 0.90 | 1.308 | 0.0678 | 1.390 | 0.0788 | 1.420 | 0.0860 | 1.498 |
| | | 0.95 | 1.632 | 0.0332 | 1.714 | 0.0376 | 1.741 | 0.0400 | 1.844 |
| | | 0.99 | 2.619 | 0.0096 | 2.486 | 0.0076 | 2.534 | 0.0074 | 2.649 |
| | $G'_{\phi_2}(K)$ | 0.90 | 1.922 | 0.0342 | 2.123 | 0.0362 | 2.201 | 0.0398 | 2.943 |
| | | 0.95 | 2.673 | 0.0160 | 2.756 | 0.0182 | 2.805 | 0.0188 | 3.663 |
| | | 0.99 | 4.110 | 0.0022 | 4.322 | 0.0026 | 4.448 | 0.0030 | 5.293 |
| $m = 3$ | $G_{\phi_2}(K)$ | 0.90 | 2.758 | 0.0828 | 2.711 | 0.0746 | 2.809 | 0.0844 | 2.943 |
| | | 0.95 | 3.390 | 0.0406 | 3.383 | 0.0392 | 3.670 | 0.0506 | 3.663 |
| | | 0.99 | 5.440 | 0.0122 | 5.365 | 0.0114 | 5.490 | 0.0144 | 5.293 |
| | $W_{\phi_2}(K)$ | 0.90 | 1.935 | 0.0732 | 2.011 | 0.0852 | 2.017 | 0.0838 | 2.114 |
| | | 0.95 | 2.303 | 0.0316 | 2.422 | 0.0428 | 2.404 | 0.0412 | 2.508 |
| | | 0.99 | 3.134 | 0.0066 | 3.162 | 0.0062 | 3.312 | 0.0088 | 3.396 |
| | $G'_{\phi_2}(K)$ | 0.90 | 2.328 | 0.0550 | 2.370 | 0.0578 | 2.503 | 0.0688 | 2.943 |
| | | 0.95 | 3.088 | 0.0366 | 3.132 | 0.0356 | 3.467 | 0.0422 | 3.663 |
| | | 0.99 | 5.552 | 0.0132 | 5.377 | 0.0114 | 5.504 | 0.0144 | 5.293 |

# 5 Numerical example: classical Lindisfarne Scribes problem

The Lindisfarne Scribes problem, in the framework of the model that is followed in this paper, considers that the Lindisfarne Gospels are divided into $K = 64$ consecutive sections (see Ross (1950) for more details). It is supposed that each section could have been written by one scribe and the same scribe is associated only with consecutive sections. We consider a triple problem:

**P.1)** It is counted $n_i$ as the total of observed frequencies that the third singular appears in each section $i = 1, ..., 64$.

**P.2)** It is counted $n_i$ as the total of observed frequencies that the second plural appears in each section $i = 1, ..., 64$.

**P.3)** It is counted $n_i$ as the total of observed frequencies that the third singular or second plural appears in each section $i = 1, ..., 64$. These values are obtained as the sum of the frequencies of problems P.1 and P.2 section by section.

In all of them, random variable $(X_{i1}, X_{i2})$ represents how many times endings $-s$ and $-\delta$ appear ($m = 2$). The observations are summarized in Table 4. Our aim is to identify how many scribes took part in writing Lindisfarne Golspels. 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.

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 we have mentioned in Section 1. 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 the the Bonferroni's inequality. Suppose that $T \in \{G_{\phi_\lambda}(K), G'_{\phi_\lambda}(K), W_{\phi_\lambda}(K)\}$ is the test-statistic we are dealing with and $x_{1-\alpha}$ the quantile of order $1 - \alpha$ associated with its asymptotic distribution. Symbol $T_k$ refers to $T^{(*)}_{k,\phi_\lambda}$ if $T \in \{G_{\phi_\lambda}(*), G'_{\phi_\lambda}(*)\}$, and refers to $\frac{N_k M_k(*)}{N_K^2} T^{(*)}_{k,\phi_\lambda}$ if $T \in \{W_{\phi_\lambda}(*)\}$. Symbol $*$ is denoting that depending on the step of the algorithm we are not necessary maximizing on $K$ terms. The algorithm is as follows:

1. Set $\ell = 1$, $K^{(\ell-1)} = K$.
2. For $k = 1, ..., K^{(\ell-1)} - 1$, obtain $T_k$ and then $T$:
   (a) If $T \geq x_{1-\alpha}$ then set $\ell = \ell + 1$ and $K^{(\ell-1)} = \arg\max_{k\in\{1,...,K^{(\ell-1)}-1\}} T_k$ is considered to be a change-point. REPEAT step 2.
   (b) If $T < x_{1-\alpha}$ then it is considered that there are no change point in $[1, K^{(\ell-1)}]$. GO step 3.
3. Are there consecutive change points $K' > 1$ and $K'' > K'$ without making hypothesis testing in segment $[K' + 1, K'']$?
   - If yes, then take the segment with smallest $K'$ and for $k = K' + 1, ..., K'' - 1$ obtain $T_k$ and then $T$ (maximized on $T_k$, $k = K' + 1, ..., K'' - 1$):
     – If $T \geq x_{1-\alpha}$ then set $\ell = \ell + 1$ and $K^{(\ell-1)} = \arg\max_{k\in\{K'+1,...,K''-1\}} T_k$ is considered to be a change-point. REPEAT step 3.
     – If $T < x_{1-\alpha}$ then it is considered that there are no change point in $[K' + 1, K'']$. REPEAT step 3.
   - If no, STOP. There are $\ell - 1$ change points located at $K^{(1)}, ..., K^{(\ell-1)}$.

Focussed on problem P.1 and choosing $\lambda = 2$, in Figure 6 (a) $T_k$, $k = 1, ..., K-1$ are shown for $T = G_{\phi_\lambda}(K)$ in circles and for $T = W_{\phi_\lambda}(K)$ in crosses. Since the maximum value is reached at 18 for both and $G_{\phi_2}(K) = 24.405 \geq x_{0.99} = 5.293$ and $W_{\phi_2}(K) = 54.127 \geq x_{0.99} = 2.649$ approve that $K^{(1)} = 18$, in Figure 6 (b) $T_k$, $k = 1, ..., K^{(1)} - 1$ and $T_k$, $k = K^{(1)} + 1, ..., K - 1$ are shown. At this time test-statistics accept the hypothesis of no change point. Table 5 contains the summary for problem P.1 and also for P.2 and P.3 with $T \in \{G_{\phi_\lambda}(K), G'_{\phi_\lambda}(K), W_{\phi_\lambda}(K)\}$ and $\lambda = 2$. Hence, in the Lindisfarne Scribes problem we can conclude:

- For problem P.1:
  - According to $G_{\phi_2}(*)$ and $G'_{\phi_2}(*)$, two scribes are identified associated to segments $[1, 18]$, $[19, 64]$.
  - According to $W_{\phi_2}(*)$, two scribes are identified associated to segments $[1, 18]$, $[19, 64]$.
- For problem P.2:
  - According to $G_{\phi_2}(*)$ and $G'_{\phi_2}(*)$, two scribes are identified associated to segments $[1, 18]$, $[19, 64]$.
  - According to $W_{\phi_2}(*)$, four scribes are identified associated to segments $[1, 15]$, $[16, 18]$, $[19, 35]$, $[36, 64]$.
- For problem P.3:
  - According to $G_{\phi_2}(*)$ and $G'_{\phi_2}(*)$, five scribes are identified associated to segments $[1, 18]$, $[19, 24]$, $[25, 31]$, $[32, 45]$, $[46, 64]$.
  - According to $W_{\phi_2}(*)$, six scribes are identified associated to segments $[1, 10]$, $[11, 18]$, $[19, 24]$, $[25, 31]$, $[32, 45]$, $[46, 64]$.

All the methods coincide in locating a change point at 18. In overall terms we can say that the Lindisfarne Golpels could have been written by at least two scribes and at most by six scribes. This conclusions tends to be more conservatives compared with the results obtained for the same problem in Horváth and Servinowska (1995). Looking at Figure 6 and comparing it with Figures 2 and 3 in Horváth and Serbinowska, it seems that such a conservative behaviour is related with the trend of having small values of $T_{k,\phi_0}^{(K)}$ with $\lambda = 2$ than with $\lambda = 0$. This idea could be also related with the small improvement in the approximations of the quantiles and type I error that was found in the Monte Carlo study.

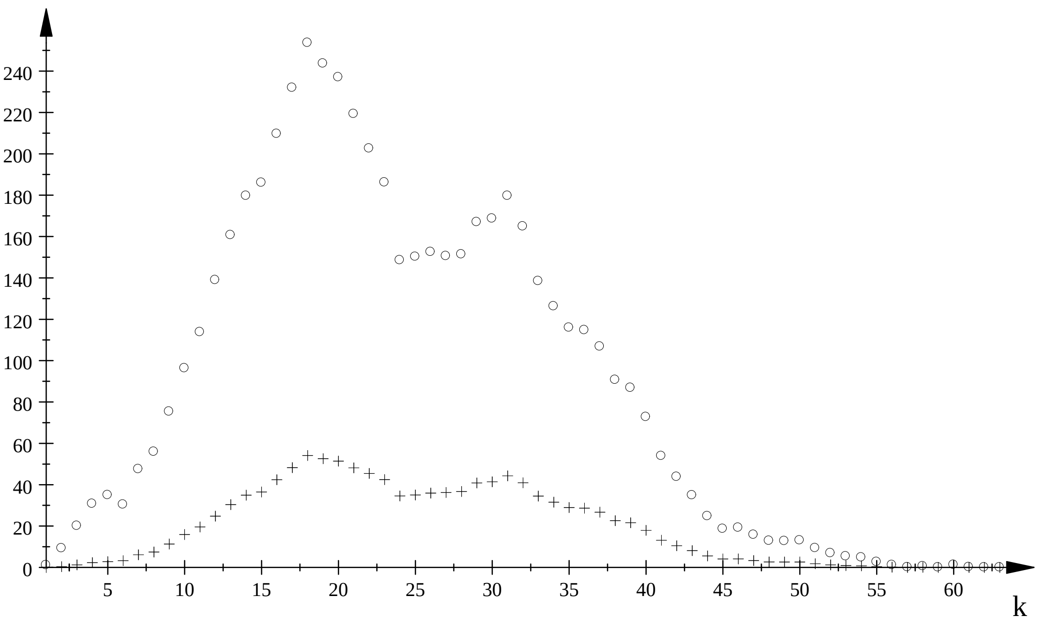
Table 4: Lindisfarne Scribes problem's data

| id. of part | $-s$ of 3rd singular | $-\delta$ of 3rd singular | $-s$ of 2nd plural | $-\delta$ of 2nd plural |
|---|---|---|---|---|
| 1 | 12 | 9 | 0 | 0 |
| 2 | 26 | 10 | 3 | 0 |
| 3 | 31 | 13 | 0 | 0 |
| 4 | 17 | 4 | 4 | 0 |
| 5 | 7 | 2 | 7 | 3 |
| 6 | 28 | 24 | 13 | 1 |
| 7 | 34 | 11 | 15 | 2 |
| 8 | 10 | 1 | 20 | 3 |
| 9 | 29 | 8 | 10 | 0 |
| 10 | 30 | 9 | 5 | 3 |
| 11 | 16 | 2 | 10 | 1 |
| 12 | 17 | 0 | 15 | 1 |
| 13 | 24 | 7 | 6 | 1 |
| 14 | 14 | 2 | 3 | 2 |
| 15 | 5 | 1 | 14 | 1 |
| 16 | 17 | 3 | 16 | 0 |
| 17 | 17 | 4 | 19 | 0 |
| 18 | 16 | 4 | 12 | 1 |
| 19 | 4 | 6 | 6 | 9 |
| 20 | 1 | 3 | 1 | 0 |
| 21 | 3 | 9 | 5 | 6 |
| 22 | 8 | 14 | 4 | 2 |
| 23 | 5 | 13 | 0 | 2 |
| 24 | 0 | 24 | 3 | 1 |
| 25 | 3 | 2 | 11 | 4 |
| 26 | 7 | 6 | 6 | 4 |
| 27 | 15 | 19 | 6 | 1 |
| 28 | 7 | 8 | 12 | 5 |
| 29 | 15 | 7 | 14 | 3 |
| 30 | 4 | 4 | 12 | 8 |
| 31 | 9 | 4 | 7 | 1 |
| 32 | 5 | 19 | 0 | 0 |
| 33 | 2 | 27 | 1 | 0 |
| 34 | 1 | 14 | 0 | 0 |
| 35 | 1 | 13 | 5 | 4 |
| 36 | 1 | 3 | 0 | 1 |
| 37 | 1 | 11 | 9 | 14 |
| 38 | 2 | 24 | 3 | 1 |
| 39 | 1 | 7 | 1 | 5 |
| 40 | 6 | 31 | 4 | 15 |
| 41 | 2 | 36 | 3 | 10 |
| 42 | 7 | 32 | 7 | 16 |
| 43 | 4 | 27 | 4 | 10 |
| 44 | 7 | 38 | 3 | 7 |
| 45 | 5 | 27 | 4 | 6 |
| 46 | 6 | 8 | 7 | 6 |
| 47 | 2 | 15 | 4 | 15 |

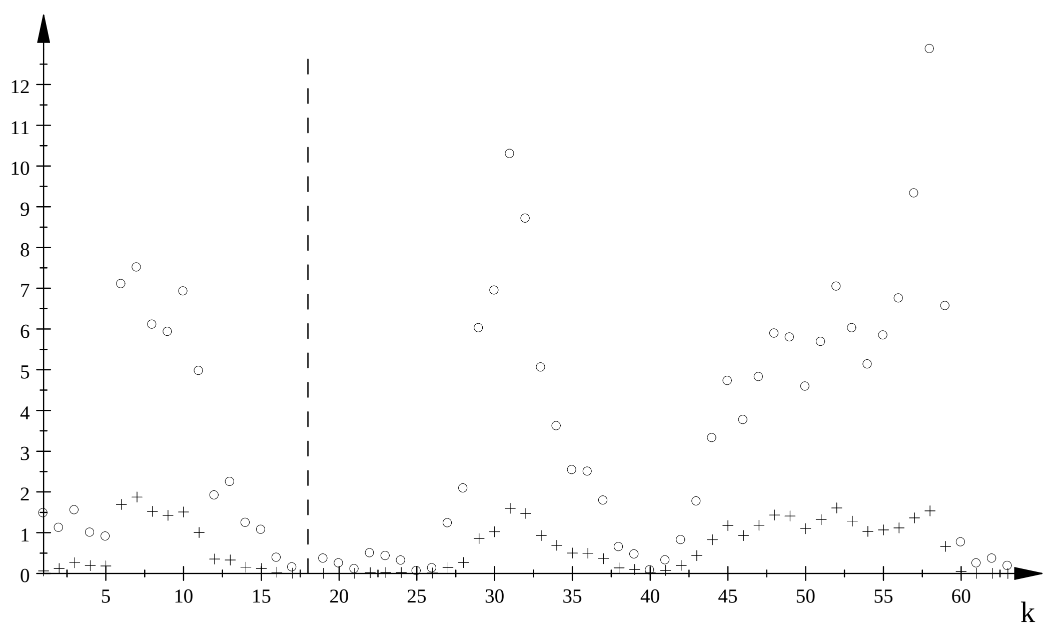
| | | | | |
|---|---|---|---|---|
| 48 | 2 | 14 | 6 | 9 |
| 49 | 1 | 2 | 1 | 5 |
| 50 | 8 | 12 | 3 | 3 |
| 51 | 6 | 25 | 2 | 5 |
| 52 | 2 | 15 | 1 | 11 |
| 53 | 15 | 32 | 4 | 4 |
| 54 | 10 | 19 | 7 | 1 |
| 55 | 10 | 30 | 2 | 3 |
| 56 | 8 | 25 | 7 | 7 |
| 57 | 6 | 27 | 9 | 2 |
| 58 | 9 | 32 | 3 | 1 |
| 59 | 16 | 9 | 5 | 3 |
| 60 | 26 | 21 | 14 | 4 |
| 61 | 19 | 39 | 11 | 16 |
| 62 | 1 | 3 | 3 | 6 |
| 63 | 2 | 3 | 1 | 3 |
| 64 | 2 | 3 | 4 | 7 |

Table 5: Summary of the binary segmentation algorithm for problems P.1, P.2, P.3.

| | segments | $G_{\phi_2}(*)$ ($\geq 5.293?$, $K^{(\ell)}$) | $W_{\phi_2}(*)$ ($\geq 2.649?$, $K^{(\ell)}$) | $G'_{\phi_2}(*)$ ($\geq 5.293?$, $K^{(\ell)}$) |
|---|---|---|---|---|
| | $[1, 64]$ | 24.405 (YES, $K^{(1)} = 18$) | 54.127 (YES, $K^{(1)} = 18$) | 28.020 (YES, $K^{(1)} = 18$) |
| P.1 | $[1, 18]$ | 2.423 (NO) | 1.876 (NO) | 1.865 (NO) |
| | $[19, 64]$ | 3.619 (NO) | 1.607 (NO) | 3.427 (NO) |
| | $[1, 64]$ | 21.227 (YES, $K^{(1)} = 18$) | 40.466 (YES, $K^{(1)} = 18$) | 23.677 (YES, $K^{(1)} = 18$) |
| | $[1, 18]$ | 4.3604(NO) | 3.135 (YES, $K^{(2)}$ =15) | 4.437 (NO) |
| | $[19, 64]$ | 3.877 (NO) | 2.966 (YES, $K^{(3)}$ =35) | 3.779 (NO) |
| P.2 | $[1, 15]$ | – | 0.053 (NO) | – |
| | $[16, 18]$ | – | −0.025 (NO) | – |
| | $[19, 35]$ | – | 0.697 (NO) | – |
| | $[36, 64]$ | – | 2.418 (NO) | – |
| | $[1, 64]$ | 29.922 (YES, $K^{(1)} = 18$) | 77.763 (YES, $K^{(1)} = 18$) | 34.902 (YES, $K^{(1)} = 18$) |
| | $[1, 18]$ | 3.619 (NO) | 3.009 (YES, $K^{(2)} = 10$) | 3.420 (NO) |
| | $[1, 10]$ | – | 0.936 (NO) | – |
| | $[11, 18]$ | – | 0.412 (NO) | – |
| | $[19, 64]$ | 6.227 (YES, $K^{(2)} = 31$) | 4.616 (YES, $K^{(3)} = 31$) | 6.587 (YES, $K^{(2)} = 31$) |
| P.3 | $[19, 31]$ | 4.772 (YES, $K^{(3)} = 24$) | 4.651 (YES, $K^{(4)} = 24$) | 5.076 (YES, $K^{(3)} = 24$) |
| | $[32, 64]$ | 6.085 (YES, $K^{(4)} = 45$) | 6.415 (YES, $K^{(5)} = 45$) | 6.496 (YES, $K^{(4)} = 45$) |
| | $[19, 24]$ | 3.084 (NO) | 1.707 (NO) | 2.818 (NO) |
| | $[25, 31]$ | 1.004 (NO) | 0.371 (NO) | −0.777 (NO) |
| | $[32, 45]$ | 0.109 (NO) | 0.124 (NO) | −1.438 (NO) |
| | $[46, 64]$ | 2.858 (NO) | 2.021 (NO) | 2.406 (NO) |




(a) $T^{(K)}_{k,\phi_2}$ (circles) and $\frac{N_k M_k(K)}{N_K^2} T^{(K)}_{k,\phi_2}$ (crosses).




(b) $T^{(K)}_{k,\phi_2}$ (circles) and $\frac{N_k M_k(K)}{N_K^2} T^{(K)}_{k,\phi_2}$ (crosses) in two segments.

Figure 6: Lindisfarne Scribes problem: Statistics ($\lambda = 2$) to be maximized for the 3rd singular data (P.1).

**Acknowledgement**

This work was supported by Grant MTM 2009-10072.

# References

[1] Chen, J. and Gupta, A.K. (2000). *Parametric Statistical Change Point Analysis.* Birkhäuser.

[2] Csörgo, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* Wiley Series in probability and Statistics.

[3] Darlin, D. and Erdös, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Mathematical Journal,* **23**,143–155.

[4] 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.

[5] Hall, P. (1979). On the rate of convergence of normal extremes. *Journal of Applied Probability*, **16**, 433-439.

[6] Hinkley, D.V. and Hinkley, E.A. (1970). Inference About the Change-point in a Sequence of Binomial Variables. *Biometrika*, **57**, 477–488.

[7] Hawkins, D.M. (2001). Fitting multiple change-point models to data. *Computational Statistics and Data Analysis*, **37**, 323–341

[8] 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.

[9] Horváth, L. and Serbinowska, M. (1995). Testing for Changes in Multinomial Observations: the Lindisfarne Scribes problem. *Scandinavian Journal of Statistics*, **22**, 371–384.

[10] Kiefer (1959). K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-V. Mises Tests. Annals of Mathematical Statististics, **30**, 420–447.

[11] Pardo, L. , Pardo, M.C. and Zografos, K. (1999). Homogeneity for multinomial Populations based on phi-divergences. *Journal Japan Statistical Society*, **29**, 213–228.

[12] Pardo, L. (2006). *Statistical Inference Based on Divergence Measures.* Chapman & Hall/CRC, Boca de Raton.

[13] Read, T. and Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer, NY.

[14] Ross, A.S.C. (1950). Philological probability problems. *Journal of the Royal Statistical Society – Series B*, **12**, 19–59.

[15] Vostrikova, L.J. (1981). Detection of "disorder" in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.

[16] Wolfe, D. A. and Chen, Y. (1990). The changepoint problem in a multinomial sequence. *Communication in Statistics*, **19**, 603–618.

[17] Yao, Y.-C. 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 Series A*, **48** , 339–353.