Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 6151 – 6165
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

# The use of cumulative sums for detection of changepoints in the rate parameter of a Poisson Process

Pedro Galeano*

*Departamento de Estadística e Investigación Operativa, Universidad de Santiago de Compostela, 15782 Santiago de Compostela, Spain*

Received 26 June 2006; received in revised form 21 December 2006; accepted 21 December 2006
Available online 2 January 2007

**Abstract**

The problem of estimation of the rate parameter of a Poisson process when the rate is piecewise constant with unknown number of pieces and changepoint locations is studied. A binary segmentation algorithm in conjunction with a cumulative sum statistic for detection of the changepoints is proposed. The asymptotic distribution of the proposed statistic is derived, its consistency is proved and the limiting distribution of the estimate of the changepoint is obtained. Also, inference of the piecewise constant rate parameter is addressed. A Monte Carlo analysis shows the good performance of the proposed cumulative sum approach, which is illustrated with a real data example.


*Keywords:* Binary segmentation procedure; Changepoints; Cumulative sum statistic; Kolmogorov statistics; Poisson processes

## 1. Introduction

The estimation of the rate parameter of a Poisson process when the rate is piecewise constant with unknown number of pieces and changepoint locations is a widely studied problem in the literature. In the hypothesis testing framework, some early attempts to test for a single changepoint at an unknown location are found in Akman and Raftery (1986) and Siegmund (1988). Worsley (1986) proposed the use of the maximum likelihood ratio test statistic for testing for one changepoint at an unknown location and derived the exact distribution of the maximum likelihood changepoint estimator. Unfortunately, this exact distribution has not a standard form and, in order to obtain critical values for the test, it is necessary to run an iterative algorithm that provides reliable results for sample sizes not larger than 100. Moreover, the asymptotic distribution of the likelihood ratio test statistic has been shown to be non-existing, see, for instance, Yao (1987). Some lower and upper bounds of some moments of the asymptotic distribution of a normalized version of the maximum likelihood changepoint estimator has been given by Jandhyala and Fotopoulos (1999) and Jandhyala et al. (2000).

The case of multiple changepoints at unknown locations has been analyzed by Worsley (1986) and Hawkins (2001). Worsley (1986) proposed to use the likelihood ratio test statistic and the binary segmentation procedure proposed by Vostrikova (1981), which under certain conditions, has been shown to be consistent to estimate the number of changepoints. This approach encounters several problems. First, each time that a time point is selected as a candidate

* Tel.: +34 981 563100x13207; fax: +34 981 597054.
*E-mail address:* pgaleano@usc.es.

0167-9473/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.12.042

to be a changepoint, it is necessary to run an iterative procedure needed to obtain the corresponding critical value which, besides, may be not reliable for large sample sizes. Thus, the procedure may be computationally expensive if the number of changes and/or events of the Poisson process is large. Second, the likelihood ratio test is affected by endpoint effects: some events near the beginning or end of the sample can cause spurious changepoints. Hawkins (2001) proposed a dynamic programming procedure for obtaining the maximum likelihood estimator of a fixed number of changepoints. The complexity of the algorithm is quadratic, which may be computationally burdensome when the number of events is large. Also, this approach does not provide a method to test how many changepoints are in the sample. In fact, Hawkins (2001) showed that his approach leads to a generalized likelihood ratio test statistic for which neither its exact distribution is known, nor its asymptotic distribution exists.

On the other hand, some papers dealing with the problem under the Bayesian point of view are Raftery and Akman (1986), Carlin et al. (1992) and Raftery (1994), who analyzed the one changepoint case, and Green (1995), Chib (1998) and Yang and Kuo (2001), who analyzed the multiple changepoints case by means of different Bayesian procedures relying on Markov Chain Monte Carlo and Bayes factor methods.

The purpose of this paper is to develop a cumulative sum approach for detection of changepoints in the piecewise constant rate of a Poisson process by means of a binary segmentation procedure, which is not affected by the problems of the maximum likelihood approach. The segmentation algorithm proposed is based on splitting the data into pieces after detecting a change. Thus, it is only necessary to test for a change in each step overcoming the complexity of multiple changepoint testing. A Bonferroni type adjustment has been considered in the binary segmentation algorithm in order to avoid spurious change point detection due to sequential testing.

A cumulative sum statistic is used to test for the presence of a single changepoint, for which its asymptotic distribution is derived and its consistency is established. Also, the distribution of the changepoint estimator is obtained, which is useful to obtain confidence intervals. The procedure here proposed is efficient, fast and very easy to implement. Binary segmentation algorithms in conjunction with cumulative sum statistics have been proposed in several papers to locate multiple mean and variance changepoints in univariate and multivariate time series. See, for instance, Inclán and Tiao (1994), Carnero et al. (2003) and Galeano et al. (2006), which also provide several advantages of cusum statistics over likelihood ratio test statistics in other frameworks.

It is important to note that the proposed procedure falls into the category of retrospective detection of changepoints, which is somewhat different than the category of prospective detection procedures. These two types of analyses are based on different purposes. In a retrospective detection, the analysis is carried out for a fixed data set and the aim is to jointly estimate the number of changes and the parameters of the model based on the observed data. In a prospective detection, the analysis is carried out with data accumulating over time and the aim is to detect as soon as possible the presence of a changepoint. Both kind of procedures, although related, are different. See, for instance, Gan (1994) and Kenett and Pollak (1996) for more information on prospective procedures for Poisson processes.

The rest of this article is organized as follows. Section 2 defines a cumulative sum statistic to test for the presence of one changepoint in the rate parameter of a Poisson process, obtains its asymptotic distribution, proves its consistency and derives the limiting distribution of the changepoint estimator. Section 3 proposes a binary segmentation procedure for detection and estimation of multiple change points. Section 4 studies the performance of the procedure in several Monte Carlo experiments for different Poisson processes, number of Poisson events, and number and changepoint locations. It is shown that the proposed procedure in conjunction with the cumulative sum statistic yields satisfactory results for changepoints detection in all the situations considered. Section 5 illustrates the procedure by means of a real data example. Finally, Section 6 concludes.

## 2. Testing for a change in the rate parameter of a Poisson process

Following Karr (1985), two sampling procedures can be used to describe a Poisson Process defined on $\Re^+$. The synchronous sampling is formed with observations until occurrence time $t_n$ of the $n$th event, where $n$ is fixed, and yields synchronous data, $t_1, \ldots, t_n$. In this case, the observation interval $(0, t_n]$ over which the Poisson process is observed is random. The asynchronous sampling is formed with observations $N_t$, the number of events up to and including time $t \in (0, T]$, where $T$ is fixed, and yields asynchronous data, $N_1, \ldots, N_T$. In order to derive asymptotic properties of the process, $n \to \infty$ is used for synchronous data, while $T \to \infty$ is used for asynchronous data.

In this article, the problem of detection of changepoints in the rate parameter of a Poisson process is analyzed assuming synchronous sampling. Thus, it is assumed that a Poisson process on $\Re^+$ is observed until time of the $n$th event. Let $t_1, \ldots, t_n$ be the times of the $n$ observed events, such that $0 < t_1 < t_2 < \cdots < t_n$, and let $d_1, \ldots, d_n$, be the interarrival times between consecutive events, defined by $d_1 = t_1$ and $d_i = t_i - t_{i-1}$, for $i = 2, \ldots, n$. The problem is to test if the rate parameter of the Poisson process is constant for the $n$ events, or, on the contrary, it is piecewise constant. Throughout this paper, one of the following two alternative assumptions is considered:

**Assumption 1.** The rate parameter of the Poisson process is constant for all the $n$ events, that is, $\lambda(i)=\lambda > 0, i=1, \ldots, n$, or in other words, the interarrival times are independent and exponentially distributed with constant rate $\lambda$. The mean of the Poisson process is then given by $\mu(i) = 1/\lambda(i) = 1/\lambda$.

**Assumption 2.** The rate parameter of the Poisson process is piecewise constant and given by

$$\lambda(i) = \begin{cases} \lambda_0 & 0 < i \leqslant n_1, \\ \lambda_1 & n_1 < i \leqslant n_2, \\ \vdots & \vdots \\ \lambda_p & n_p < i \leqslant n, \end{cases} \qquad i = 1, \ldots, n, \tag{1}$$

with $\lambda_j > 0$, $j = 0, \ldots, p$. The mean of the process is also piecewise constant and given by $\mu(i) = 1/\lambda(i)$. The event times $t_1, \ldots, t_{n_1}$ have rate $\lambda_0$, $t_{n_1+1}, \ldots, t_{n_2}$ have rate $\lambda_1$, etc... The following conditions also hold:

(2.A) The number of changepoints, $p$, is fixed. Note that $p$ has to verify $p < n$.

(2.B) The changepoint locations $n_1, \ldots, n_p$ are such that $n_j = [n\tau_j]$, $j = 1, \ldots, p$, where [] is the integer valued function, for some $\tau_j \in (0, 1)$ and $0 < \tau_1 < \cdots < \tau_p < 1$.

(2.C) The values of the rate in consecutive intervals are different, i.e., $\lambda_j \neq \lambda_{j+1}$, $j = 0, \ldots, p-1$.

Given the observed process, the number of changes, $p$, if any, the constant rates, $\lambda_0, \ldots, \lambda_p$, and changepoint locations, $n_1, \ldots, n_p$, are unknown. To test for the presence of changepoints, the statistic $C_i$ defined as follows:

$$C_i = |D_i|, \quad i = 1, \ldots, n \tag{2}$$

is used, where $D_i$ is the centered and normalized cumulative sums of the interarrival times given by

$$D_i = \sqrt{n}\left(\frac{t_i}{t_n} - \frac{i}{n}\right) = \sqrt{n}\left(\frac{\sum_{j=1}^{i} d_j}{\sum_{j=1}^{n} d_j} - \frac{i}{n}\right). \tag{3}$$

The statistic $C_i$ compares the cumulative sum of the interarrival times until the $i$th event with respect to the cumulative sum until the $n$th event. Note that the proposed test is a difference between the theoretical and the empirical distribution functions. Thus, the statistic (2) falls into the class of Kolmogorov statistics. If the rate is constant, the ratio between both cumulative sums should be around $i/n$, but if the rate is piecewise constant, the ratio can be very different from $i/n$. To see this, the behavior of the statistic $C_i$ in (2) is explored under several situations which are illustrated in Fig. 1. The columns in this matrix of plots represent three different generating processes, corresponding to zero, one and two changes, respectively. The first row in Fig. 1 shows the events generated from each process until the 500th event, while the second row shows the values of the statistic $C_i$, computed with the events in the same column. In the first column in Fig. 1, the case of a constant rate with $\lambda = 1$, the statistics plotted in the second row are under a straight line at the height of the 95th percentile of the distribution of the maximum of the statistics $C_i$, computed as explained later. In the second column, where the rate changes from $\lambda(i)=1$ to $\lambda(i)=0.5$ after the 250th event, there is a significant maximum around this event. In the third column, where the rate changes from $\lambda(i) = 1$ to $\lambda(i) = 0.5$ after the 166th event, and then changes from $\lambda(i) = 1$ to $\lambda(i) = 0.5$ after the 333th event, there are two significant maxima around these two events. This behavior suggests to search for a changepoint looking at the maximum of the statistics (2).

In the next section, a binary segmentation algorithm is proposed, which only needs to test for a single changepoint in different pieces of the data. Consequently, the behavior of the statistic (2) is explored under the assumptions of zero and one change, respectively. The rest of this section is devoted to present some theoretical results regarding three different aspects: first, the asymptotic distribution of the statistic (2); second, the asymptotic behavior of the changepoint estimator; and third, the asymptotic distribution of the estimate of the piecewise constant rate. The proofs and technical details of the next three theorems are given in a separate appendix for interested readers.

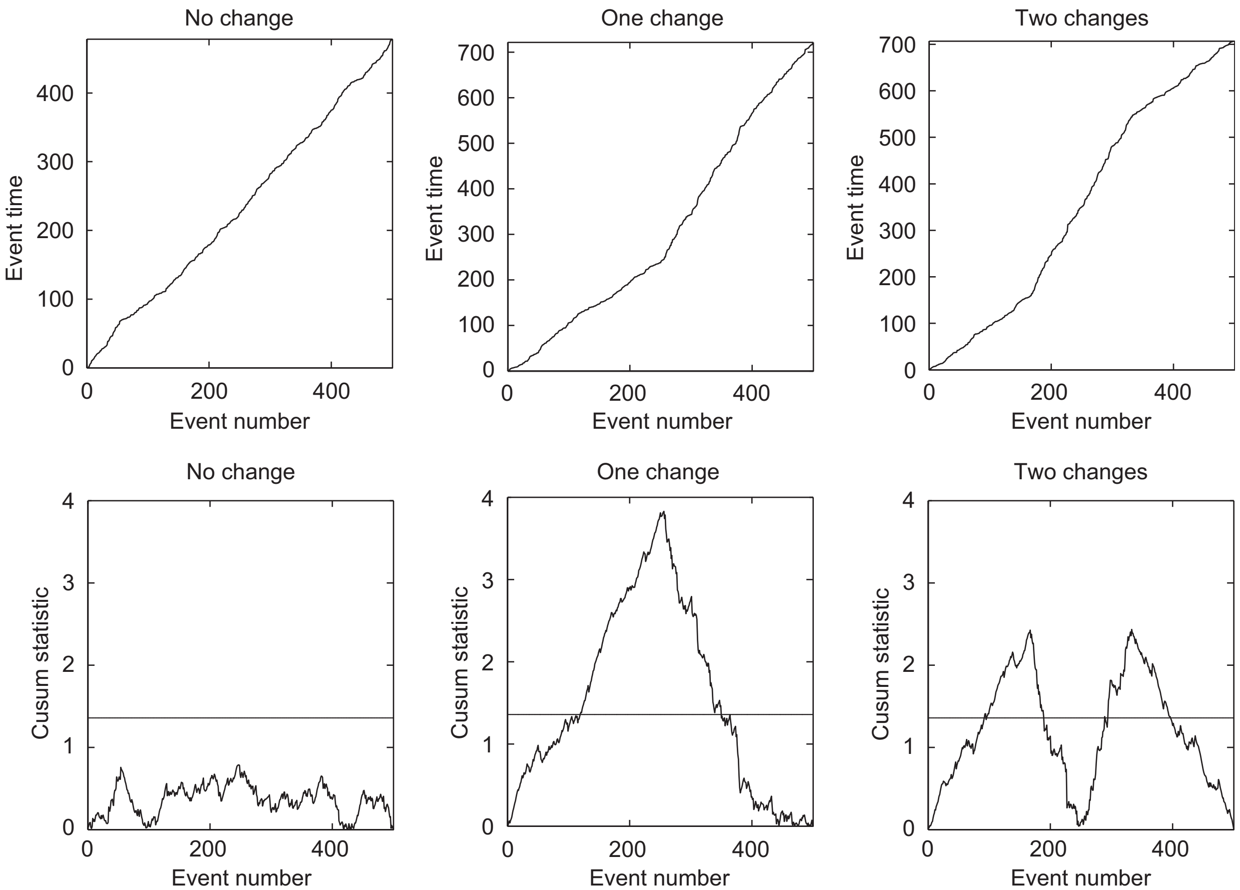


Fig. 1. Three different Poisson processes (top) and the corresponding $C_i$ statistics for changepoint detection (bottom). The first column is the case of no changes, the second column is the case of one change after the 250th event and the third column is the case of two changes after the 166th and 333th events.

First, the centered and normalized cumulative sum statistics under the null hypothesis of no change in other related problems, such as changes in the mean or variance of white noise series, behave asymptotically like a standard Brownian bridge on [0, 1], see Inclán and Tiao (1994) and Carnero et al. (2003). Theorem 1 shows that, in the analyzed problem, this is indeed the case. For that, the statistic (2) can be written as follows:

$$C_{[n\tau]} = |D_{[n\tau]}|, \quad i = 1, \ldots, n \tag{4}$$

where,

$$D_{[n\tau]} = \sqrt{n}\left(\frac{t_{[n\tau]}}{t_n} - \frac{[n\tau]}{n}\right) = \sqrt{n}\left(\frac{\sum_{j=1}^{[n\tau]} d_j}{\sum_{j=1}^{n} d_j} - \frac{[n\tau]}{n}\right),$$

with $\tau \in [0, 1]$, and such that $i = [n\tau]$.

**Theorem 1.** *Under the Assumption* 1, *the statistic* $D_{[n\tau]} \xrightarrow{\mathrm{d}} M^0(\tau)$, *where* $M^0(\tau)$ *denotes a standard Brownian bridge on* [0, 1].

As mentioned previously, the interest is in the maximum of the statistics (4) over $\tau \in [0, 1]$, which is given by

$$\Lambda_{\max} = \max\{C_{[n\tau]}, i = [n\tau], 1 \leqslant i \leqslant n\}. \tag{5}$$

The limiting distribution of $\Lambda_{\max}$ in (5) can be expressed in terms of the distribution of $\sup\{|M^0(\tau)| : 0 \leqslant \tau \leqslant 1\}$, which has the following cumulative distribution function (see, Billingsley, 1968):

$$F(\sup |M^0(\tau)| \leqslant x) = 1 + 2\sum_{j=1}^{\infty} (-1)^j \exp(-2j^2x^2), \tag{6}$$

that provides asymptotic critical values of the statistic (5).

Second, the behavior of the statistic (5) is studied under the assumption of a single changepoint after the $n_1$th event. In this case, the estimate of the true changepoint, $\tau_1 = n_1/n$, provided by the statistic (5) is the point in which the statistic $C_{[n\tau_1]}$ achieves its maximum value, which is denoted by $\widehat{\tau}$. Theorem 2 below establishes the consistency of the changepoint estimator. The rate of convergence (7) is the typical one associated with changepoint estimators. Theorem 2 also provides the distribution of $\widehat{\tau}$, which can be used to construct confidence intervals for $\tau_1$.

**Theorem 2.** *Under the Assumption* 2, *for the case of a single changepoint*, *i.e.* $p = 1$, *the following holds:*

(1) $\widehat{\tau}$ *is a consistent estimator of* $\tau_1$ *with*,

$$|\widehat{\tau} - \tau_1| = \mathrm{O_p}(n^{-1}|\mu_0 - \mu_1|^{-2}). \tag{7}$$

(2) *If additionally*, $|\mu_0 - \mu_1|$ *depends on n in such a way that* $|\mu_0 - \mu_1| \to 0$ *and* $n^{1/2}|\mu_0 - \mu_1|/(\log n)^{1/2} \to \infty$, *if* $n \to \infty$, *then*

$$\frac{n|\widehat{\mu}_0 - \widehat{\mu}_1|^2(\widehat{\tau} - \tau_1)}{\widehat{\sigma}_y^2} \xrightarrow{\mathrm{d}} \arg\max_{w\in(-\infty,\infty)} \left\{ W(w) - \frac{1}{2}|w| \right\}, \tag{8}$$

*where* $W(w)$ *is a two-sided Brownian motion defined in* $(-\infty, \infty)$, $\widehat{\mu}_0$ *and* $\widehat{\mu}_1$ *are the sample means of the sequences* $d_1, \ldots, d_{\widehat{n}_1}$ *and* $d_{\widehat{n}_1+1}, \ldots, d_n$, *respectively*, *where* $\widehat{n}_1 = [n\widehat{\tau}_1]$, *and* $\widehat{\sigma}_y^2$ *is the sample variance of* $\{d_1, \ldots, d_n\}$.

The additional conditions in part 2 of Theorem 2 assume that $|\mu_0 - \mu_1|$ is small enough, which is equivalent to assume that $|\lambda_0 - \lambda_1|$ is small compared with the product $\lambda_0\lambda_1$. The case of small changes is considered because the estimation of a large size changepoint is rather accurate, and the asymptotic distribution of $\widehat{\mu}_0 - \widehat{\mu}_1$ depends on $|\mu_0 - \mu_1|$ in a complex way (see Hinkley, 1970). Therefore, confidence intervals are not easily constructed.

The asymptotic distribution of the changepoint depends on the distribution of $W(w)$, a two-sided Brownian motion process defined in the interval $(-\infty, \infty)$, such that $W(w) = M_1(-w)$ for $w < 0$ and $W(w) = M_2(w)$ for $w \geqslant 0$, where $M_1(w)$ and $M_2(w)$ are two independent Brownian motions. The cumulative distribution function of $\arg\max_w\{W(w) - 1/2|w|\}$ is given by (see, Yao, 1987),

$$F(w) = 1 + \sqrt{\frac{w}{2\pi}} \exp\left(-\frac{w}{8}\right) - \frac{(w+5)}{2}\Phi\left(-\frac{\sqrt{w}}{2}\right) + \frac{3\exp(w)}{2}\Phi\left(-\frac{3\sqrt{w}}{2}\right),$$

for $w > 0$ and $F(w) = 1 - F(-w)$, for $w < 0$, where $\Phi(w)$ is the cumulative distribution function of the standard normal distribution. A confidence interval for the changepoint location at the $100 \times (1 - \alpha)$ confidence level is straighforwardly obtained from (8)

$$\left( \left[ \widehat{\tau}_1 - \frac{w_{1-\alpha/2}\widehat{\sigma}_y^2}{n|\widehat{\mu}_0 - \widehat{\mu}_1|^2} \right], \left[ \widehat{\tau}_1 - \frac{w_{\alpha/2}\widehat{\sigma}_y^2}{n|\widehat{\mu}_0 - \widehat{\mu}_1|^2} \right] \right), \tag{9}$$

where $w_{\alpha/2}$ and $w_{1-\alpha/2}$ are the $100 \times \alpha/2$th and $100 \times (1 - \alpha/2)$th percentiles of the distribution of $\arg\max_w\{W(w) - 1/2|w|\}$.

Finally, given the previous results, it may be expected that the limiting distributions of the estimates of $\lambda_0$ and $\lambda_1$ will be similar to the ones in the case in which the changepoint location is known. This is indeed the case as shown in Theorem 3. If the changepoint location is known (see Karr, 1985),

$$n_1^{1/2}\left(\frac{\widetilde{\mu}_0}{\mu_0} - 1\right) = n_1^{1/2}\left(\frac{\lambda_0}{\widetilde{\lambda}_0} - 1\right), \tag{10}$$

and,

$$(n-n_1)^{1/2}\left(\frac{\widetilde{\mu}_1}{\mu_1}-1\right)=(n-n_1)^{1/2}\left(\frac{\lambda_1}{\widetilde{\lambda}_1}-1\right), \tag{11}$$

have a limiting standard normal distribution, where $\widetilde{\mu}_0$ and $\widetilde{\mu}_1$ are the sample means of the sequences $d_1,\ldots,d_{n_1}$ and $d_{n_1+1},\ldots,d_n$, respectively. The rate values are given by $\widetilde{\lambda}_0=1/\widetilde{\mu}_0$ and $\widetilde{\lambda}_1=1/\widetilde{\mu}_1$, respectively.

**Theorem 3.** *Under the Assumption* 2, *for the case of a single changepoint*, *i.e.* $p=1$, *the following holds:*

(1) $\widehat{n}_1^{1/2}(\lambda_0/\widehat{\lambda}_0-1)$ *has a limiting standard normal distribution.*
(2) $(n-\widehat{n}_1)^{1/2}(\lambda_1/\widehat{\lambda}_1-1)$ *has a limiting standard normal distribution.*

Thus, the piecewise rate parameter values, $\lambda_0$ and $\lambda_1$, can be estimated consistently, and the limiting distributions of their estimates are the same as if $n_1$ is assumed known. Therefore, asymptotic confidence intervals at the $100\times(1-\alpha)$ confidence level for $\lambda_0$ and $\lambda_1$, are given by

$$\lambda_0\in\left(\widehat{\lambda}_0\left(1-\frac{z_{1-\alpha/2}}{\widehat{n}_1^{1/2}}\right),\widehat{\lambda}_0\left(1+\frac{z_{1-\alpha/2}}{\widehat{n}_1^{1/2}}\right)\right), \tag{12}$$

and,

$$\lambda_1\in\left(\widehat{\lambda}_1\left(1-\frac{z_{1-\alpha/2}}{(n-\widehat{n}_1)^{1/2}}\right),\widehat{\lambda}_1\left(1+\frac{z_{1-\alpha/2}}{(n-\widehat{n}_1)^{1/2}}\right)\right), \tag{13}$$

respectively, where $z_{1-\alpha/2}$ denotes the $100\times(1-\alpha/2)$th percentile of the standard normal distribution.

## 3. Detection of multiple changes: binary segmentation algorithm

When $p$ is larger than one, the usefulness of the statistic (5) is questionable due to the possibility of masking effects. Edwards and Cavalli-Sforza (1965) proposed a binary segmentation algorithm as a method for splitting data into clusters, which was afterward popularized by Vostrikova (1981) by establishing its consistency for detecting the true number of changepoints. The algorithm is based on splitting the data into two pieces after detecting a change. Thus, the idea behind the algorithm is to isolate each changepoint. The proposed binary segmentation procedure proceeds as follows:

(1) Let $\ell=1$. Obtain,

$$\Lambda_{\max}=\max\{C_i,1\leqslant i\leqslant n\}.$$

(a) If $\Lambda_{\max}\geqslant\upsilon_\alpha$, where $\upsilon_\alpha$ is the asymptotic critical value for a given significance level $1-\alpha$, define $m_1$ as the point in which the statistic $C_i$ achieves its maximum value. Then, go to step 2.
(b) If $\Lambda_{\max}<\upsilon_\alpha$, it is assumed that the rate is constant and the procedure ends.

(2) Let $\ell=\ell+1$. Repeat this step until,

$$\max\{\Lambda_{\max}^k,k=1,\ldots,\ell+1\}<\upsilon_\alpha,$$

where,

$$\Lambda_{\max}^k=\max\{C_i,m_{k-1}+1\leqslant i\leqslant m_k\},\quad k=1,\ldots,\ell+1$$

such that $m_0=0$, $m_{\ell+1}=n$, and $m_k$, for $k=1,\ldots,\ell$, are the possible changepoints detected in previous iterations sorted in increasing order. Note that the expression of the statistic $C_i$ for $m_{k-1}+1\leqslant i\leqslant m_k$ is given by

$$C_i=\sqrt{m_k-m_{k-1}}\left|\frac{\sum_{j=m_{k-1}+1}^{i}d_j}{\sum_{j=m_{k-1}+1}^{m_k}d_j}-\frac{i-m_{k-1}}{m_k-m_{k-1}}\right|.$$

(3) Let $M = (m_0, m_1, \ldots, m_{\ell+1})$ where $m_0 = 0$, $m_{\ell+1} = n$, and $m_1, \ldots, m_\ell$ are the possible changepoints detected in Steps 1 and 2 sorted in increasing order. Check each possible changepoint by calculating the statistic $C_i$ in the interval $m_j + 1 \leqslant i \leqslant m_{j+2}$. If the maximum of $C_i$ in this interval is significant, keep the point. Otherwise, eliminate it from the list. Repeat Step 3 until the number of possible changepoints does no longer change, and the points found in previous iterations do not differ from those in the last one. The elements of the vector $(m_1, \ldots, m_\ell)$ are the detected changepoints.

(4) Finally, the rate parameter (1) is estimated by means of the inverse of the sample means of the data between changepoints.

Some comments regarding the proposed binary segmentation algorithm are in order. First, the critical values used in the procedure are the percentiles of the distribution of the maximum of the absolute value of a standard Brownian Bridge on [0, 1] with distribution function (6). Several of them are provided in the next section. Second, different critical values are used depending on the number of changepoints detected in each step of the algorithm. The use of the same critical value in steps 1 and 2 may lead to overestimation of the number of changepoints. Note that if the same critical value, $\upsilon_\alpha$, for a given significance level, $1-\alpha$, is used, then, in step 1, the probability of detecting a false changepoint is $1-(1-\alpha)$. After detecting a first changepoint, the probability of detecting a false changepoint in the second iteration is $1-(1-\alpha)^2 > \alpha$. In general, the probability of detecting a false changepoint in the $r$th iteration is $1-(1-\alpha)^r$ which tends to 1 when $r$ increases. Therefore, after detecting a new change, there is an increasing probability of detecting a spurious changepoint after splitting the data again. To avoid this problem, in the $\ell$th iteration, thus after detecting the $(\ell-1)$th changepoint, a critical value, denoted by $\alpha_\ell$, verifying $\alpha_0 = 1-(1-\alpha_\ell)^{\ell+1}$, is taken, where $1-\alpha_0$ is the significance level used in step 1, usually $\alpha_0 = 0.05$. This election ensures that the probability of detecting a false changepoint is always the same in each step. Third, as in Inclán and Tiao (1994), the step 3 in the procedure is included for avoiding false changepoints. In this step, if there are $\ell$ remaining changes, the critical value $\alpha_{\ell-1}$ is used. Fourth, a minimum distance between changes is required larger to be than a positive integer number $d$ in order to estimate the constant rate between changes. In the Monte Carlo experiments and the real data example of the next sections, the choice has been $d=5$ because this election works well in the simulations. Thus, the results given by the procedure are conditioned to this election, which may produce some masking effects for changepoints extremely close.

## 4. Monte Carlo experiments

The Monte Carlo results in this section and the analysis of the real data example in the next one have been carried out by means of various routines written by the author in MATLAB (developed by The MathWorks, Inc) which are available under request. Three aspects of the proposed procedure are analyzed via several Monte Carlo experiments: first, the asymptotic and finite sample critical values of the statistic (5); second, the empirical size of the procedure; and third, the power of the procedure.

First, the finite sample behavior of the quantiles of the statistic (5) under the hypothesis of no change is analyzed and compared with the asymptotic ones. For that, 10 000 realizations from a Poisson process with constant rate parameter $\lambda = 1$, and different number of events, $n = 100$, 200, 500 and 1000 are generated and the statistics (5) are computed.

Table 1 provides some quantiles of the distribution of $\Lambda_{\max}$ for different values of $n$ assuming a constant rate. It can be seen that the finite sample quantiles are always smaller than the asymptotic ones implying that taking the asymptotic

Table 1
Empirical quantiles of the $\Lambda_{\max}$ statistics based on 10 000 realizations for $\lambda = 1$ and $n = 100$, 200, 500 and 1000

| $n$ | Probability | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.95 | 0.975 | 0.983 | 0.987 | 0.990 | 0.991 | 0.992 | 0.993 | 0.994 | 0.995 |
| 100 | 1.271 | 1.402 | 1.462 | 1.508 | 1.538 | 1.565 | 1.584 | 1.607 | 1.617 | 1.633 |
| 200 | 1.305 | 1.427 | 1.495 | 1.545 | 1.570 | 1.597 | 1.614 | 1.624 | 1.633 | 1.653 |
| 500 | 1.332 | 1.453 | 1.518 | 1.561 | 1.591 | 1.624 | 1.644 | 1.663 | 1.683 | 1.698 |
| 1000 | 1.347 | 1.477 | 1.541 | 1.585 | 1.617 | 1.650 | 1.671 | 1.691 | 1.709 | 1.726 |
| $\infty$ | 1.358 | 1.478 | 1.544 | 1.590 | 1.624 | 1.652 | 1.675 | 1.694 | 1.712 | 1.728 |

Table 2
Results for type I errors with $\lambda = 1$ and $n = 100$, 200, 500 and 1000

| $\lambda$ | $n$ | Frequency | | |
|---|---|---|---|---|
| | | 0 | 1 | $\geqslant 2$ |
| 1 | 100 | 96.5 | 3.4 | 0.1 |
| 1 | 200 | 96.0 | 3.9 | 0.1 |
| 1 | 500 | 95.6 | 4.3 | 0.1 |
| 1 | 1000 | 95.4 | 4.5 | 0.1 |

Table 3
Detection frequencies of the proposed procedure for one changepoint

| $\lambda$ | $n$ | $n_1$ | Frequency | | | $\lambda$ | $n$ | $n_1$ | Frequency | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 1 | $\geqslant 2$ | | | | 0 | 1 | $\geqslant 2$ |
| 0.25 | 100 | 25 | 0.1 | 98.0 | 1.9 | 2 | 100 | 25 | 29.6 | 69.2 | 1.2 |
| | | 50 | 0.0 | 97.5 | 2.5 | | | 50 | 14.7 | 83.8 | 1.5 |
| | | 75 | 0.7 | 97.4 | 1.9 | | | 75 | 57.4 | 41.7 | 0.9 |
| 0.25 | 200 | 50 | 0.0 | 96.6 | 3.4 | 2 | 200 | 50 | 4.4 | 93.2 | 2.4 |
| | | 100 | 0.0 | 96.2 | 3.8 | | | 100 | 0.8 | 96.7 | 2.5 |
| | | 150 | 0.0 | 97.3 | 2.7 | | | 150 | 12.8 | 85.0 | 2.2 |
| 0.25 | 500 | 125 | 0.0 | 96.3 | 3.7 | 2 | 500 | 125 | 0.0 | 96.2 | 3.8 |
| | | 250 | 0.0 | 95.6 | 4.4 | | | 250 | 0.0 | 96.2 | 3.8 |
| | | 375 | 0.0 | 95.2 | 4.8 | | | 375 | 0.0 | 96.5 | 3.5 |
| 0.5 | 100 | 25 | 57.4 | 41.9 | 0.7 | 4 | 100 | 25 | 0.9 | 97.1 | 2.0 |
| | | 50 | 14.8 | 83.4 | 1.8 | | | 50 | 0.0 | 97.4 | 2.6 |
| | | 75 | 29.1 | 69.7 | 1.2 | | | 75 | 0.1 | 97.9 | 2.0 |
| 0.5 | 200 | 50 | 12.6 | 85.4 | 2.0 | 4 | 200 | 50 | 0.0 | 97.1 | 2.9 |
| | | 100 | 0.6 | 96.8 | 2.6 | | | 100 | 0.0 | 96.1 | 3.9 |
| | | 150 | 4.6 | 93.1 | 2.3 | | | 150 | 0.0 | 97.2 | 2.8 |
| 0.5 | 500 | 125 | 0.0 | 96.6 | 3.4 | 4 | 500 | 125 | 0.0 | 95.6 | 4.4 |
| | | 250 | 0.0 | 96.6 | 3.4 | | | 250 | 0.0 | 95.1 | 4.9 |
| | | 375 | 0.0 | 96.6 | 3.4 | | | 375 | 0.0 | 96.4 | 3.6 |

quantiles for testing is a conservative decision. Therefore, it is expected that the type I error will be usually smaller than the nominal one. In order to know more precisely the effect of using the asymptotic critical values, the size (type I error) of the proposed procedure is analyzed by generating 10 000 realizations from a Poisson process with constant rate $\lambda = 1$ and number of events $n = 100$, 200 and 500, and applying the proposed procedure with critical values taken from Table 1, starting with $v_{\alpha_0} = 1.358$, which corresponds to the significance level $1 - \alpha_0 = 0.95$. The results are shown in Table 2, where columns 3–5 report the number of changepoints detected by the algorithm. The type I error frequencies are always smaller than 5% for all the sample sizes considered. In particular, note that with $n = 100$ observations, the type I error is 3.5%, which is not very far to the nominal 5%.

Next, a Monte Carlo experiment is carried out in order to study the power of the proposed procedure for one changepoint detection. For that, three changepoint locations are considered: $n_1 = [0.25n]$, $[0.50n]$ and $[0.75n]$, where $n = 100, 200$ and 500. The changes are introduced by transforming the original rate $\lambda = 1$, into one of the rates, 0.25, 0.5, 2 and 4. For each case, 10 000 realizations are generated. Then, the procedure is applied with critical values taken from Table 1, starting with $v_{\alpha_0} = 1.358$ which corresponds to the significance level $1 - \alpha_0 = 0.95$. The results are shown in Tables 3 and 4 . Columns 4–6 and 10–12 of Table 3 report the number of changes detected by the algorithm, which shows that the procedure performs quite well for one changepoint, with most of the cases over the 95% of detection frequency. When the number of events is small, the procedure detects more frequently the changepoints located at

Table 4
Estimation results for one changepoint

| $\lambda$ | $n$ | $n_1$ | $\widehat{n}_1$ | | $\lambda$ | $n$ | $n_1$ | $\widehat{n}_1$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Med. | Mad | | | | Med. | Mad |
| 0.25 | 100 | 25 | 27 | 2 | 2 | 100 | 25 | 25 | 3 |
| | | 50 | 51 | 1 | | | 50 | 48 | 3 |
| | | 75 | 75 | 1 | | | 75 | 63 | 9 |
| 0.25 | 200 | 50 | 51 | 1 | 2 | 200 | 50 | 50 | 3 |
| | | 100 | 101 | 5 | | | 100 | 98 | 3 |
| | | 150 | 150 | 1 | | | 150 | 142 | 8 |
| 0.25 | 500 | 125 | 126 | 2 | 2 | 500 | 125 | 125 | 4 |
| | | 250 | 251 | 1 | | | 250 | 248 | 3 |
| | | 375 | 375 | 1 | | | 375 | 370 | 6 |
| 0.5 | 100 | 25 | 37 | 9 | 4 | 100 | 25 | 25 | 1 |
| | | 50 | 52 | 3 | | | 50 | 49 | 1 |
| | | 75 | 75 | 3 | | | 75 | 73 | 2 |
| 0.5 | 200 | 50 | 58 | 8 | 4 | 200 | 50 | 50 | 1 |
| | | 100 | 102 | 3 | | | 100 | 99 | 1 |
| | | 150 | 150 | 3 | | | 150 | 149 | 1 |
| 0.5 | 500 | 125 | 130 | 6 | 4 | 500 | 125 | 125 | 1 |
| | | 250 | 252 | 3 | | | 250 | 249 | 1 |
| | | 375 | 375 | 4 | | | 375 | 374 | 2 |

Table 5
Detection frequencies of the proposed procedure for two changepoints

| $\lambda_1$ | $\lambda_2$ | $n$ | $n_1$ | $n_2$ | Frequency | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 0 | 1 | 2 | $\geqslant 3$ |
| 2 | 0.5 | 100 | 33 | 66 | 2.1 | 46.7 | 50.7 | 0.5 |
| | | 200 | 66 | 133 | 0.0 | 10.1 | 87.7 | 2.2 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 96.4 | 3.6 |
| 0.5 | 2 | 100 | 33 | 66 | 1.5 | 39.4 | 58.5 | 0.6 |
| | | 200 | 66 | 133 | 0.0 | 7.5 | 90.7 | 1.8 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 96.4 | 3.6 |
| 4 | 0.25 | 100 | 33 | 66 | 0.0 | 1.4 | 96.6 | 2.0 |
| | | 200 | 66 | 133 | 0.0 | 0.0 | 95.9 | 4.1 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 95.3 | 4.7 |
| 0.25 | 4 | 100 | 33 | 66 | 0.0 | 0.2 | 97.7 | 2.1 |
| | | 200 | 66 | 133 | 0.0 | 0.0 | 95.6 | 4.4 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 95.3 | 4.7 |
| 2 | 0.5 | 100 | 40 | 75 | 7.2 | 38.1 | 54.2 | 0.5 |
| | | 200 | 80 | 150 | 0.1 | 6.1 | 91.6 | 2.2 |
| | | 500 | 200 | 375 | 0.0 | 0.0 | 96.7 | 3.3 |
| 0.5 | 2 | 100 | 40 | 75 | 16.9 | 23.8 | 58.9 | 0.4 |
| | | 200 | 80 | 150 | 0.0 | 4.8 | 93.3 | 1.9 |
| | | 500 | 200 | 375 | 0.0 | 0.0 | 96.6 | 3.4 |
| 4 | 0.25 | 100 | 40 | 75 | 0.0 | 0.3 | 97.9 | 1.8 |
| | | 200 | 80 | 150 | 0.0 | 0.0 | 96.2 | 3.8 |
| | | 500 | 200 | 375 | 0.0 | 0.0 | 94.9 | 5.1 |
| 0.25 | 4 | 100 | 40 | 75 | 0.0 | 0.1 | 98.5 | 1.4 |
| | | 200 | 80 | 150 | 0.0 | 0.0 | 96.6 | 3.4 |
| | | 500 | 200 | 375 | 0.0 | 0.0 | 94.9 | 5.1 |

Table 6
Estimation results for two changepoints

| $\lambda_1$ | $\lambda_2$ | $n$ | $n_1$ | $n_2$ | $\widehat{n}_1$ | | $\widehat{n}_2$ | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Med. | Mad | Med. | Mad |
| 0.5 | 2 | 100 | 33 | 66 | 31 | 2 | 67 | 1 |
| | | 200 | 66 | 133 | 64 | 3 | 134 | 1 |
| | | 500 | 166 | 333 | 164 | 3 | 334 | 1 |
| 2 | 0.5 | 100 | 33 | 66 | 35 | 2 | 65 | 1 |
| | | 200 | 66 | 133 | 68 | 3 | 132 | 1 |
| | | 500 | 166 | 333 | 168 | 3 | 332 | 1 |
| 0.25 | 4 | 100 | 33 | 66 | 32 | 1 | 66 | 1 |
| | | 200 | 66 | 133 | 65 | 1 | 134 | 1 |
| | | 500 | 166 | 333 | 165 | 1 | 334 | 1 |
| 4 | 0.25 | 100 | 33 | 66 | 34 | 1 | 66 | 0 |
| | | 200 | 66 | 133 | 67 | 1 | 132 | 1 |
| | | 500 | 166 | 333 | 167 | 1 | 332 | 1 |
| 0.5 | 2 | 100 | 40 | 75 | 38 | 3 | 75 | 1 |
| | | 200 | 80 | 150 | 78 | 3 | 151 | 1 |
| | | 500 | 200 | 375 | 198 | 3 | 376 | 1 |
| 2 | 0.5 | 100 | 40 | 75 | 42 | 2 | 73 | 2 |
| | | 200 | 80 | 150 | 82 | 3 | 148 | 2 |
| | | 500 | 200 | 375 | 202 | 3 | 373 | 2 |
| 0.25 | 4 | 100 | 40 | 75 | 39 | 1 | 75 | 0 |
| | | 200 | 80 | 150 | 79 | 1 | 150 | 0 |
| | | 500 | 200 | 375 | 199 | 1 | 375 | 0 |
| 4 | 0.25 | 100 | 40 | 75 | 41 | 1 | 74 | 1 |
| | | 200 | 80 | 150 | 81 | 1 | 149 | 1 |
| | | 500 | 200 | 375 | 201 | 1 | 374 | 1 |

the middle of the observational period. As the number of events increases and the size of the change gets larger, the procedure works better. Note that the power appears to depend on the relationship between the location and the size of the change. See for instance, the case of $n = 100$, in which, first, the rate 0.5 at the location 25 and the rate 2 at the location 75, and second, the rate 0.5 at the location 75 and the rate 2 at the location 25 give very close powers. Thus, if the rate decreases, a change at the end of the observational period is better detected than a change at the beginning. On the contrary, if the rate increases, a change at the beginning of the observational period is better detected than a change at the end. The reason of this relationship is the symmetry of both experiments. Note that in the first case, the first 25 observations have rate 1, while the second 75 observations have rate 2, while in the second case, the first 75 observations have rate 1, while the second 25 observations have rate 0.5. Thus, there are 75 observations with a rate which is the double than the rate of 25 observations, implying that both situations are perfectly symmetric. On the other hand, columns 4–5 and 9–10 in Table 4 show the median and mean absolute deviation of the changepoint estimators in each case. The median of the estimates are quite close to the true changepoint locations. Note that the larger is the size of the change, the better is estimated its location. Also the relationship between the location and the size of the change is clearly seen here in the same cases.

Finally, for two changepoints, two pairs of changepoint locations at $(n_1, n_2) = ([0.33n], [0.66n])$ and $(n_1, n_2) = ([0.40n], [0.75n])$ are considered, where $n = 100$, 200 and 500. Thus, the rate changes from $\lambda = 1$ to one of the rates 0.25, 0.5, 2 and 4 at the first changepoint, and then changes again to a different rate among the set 0.25, 0.5, 2 and 4 at the second changepoint. Four combinations are considered for both the symmetric and the asymmetric situations. For each case, 10 000 realizations with the corresponding changes are generated, and the procedure with critical values taken from Table 1 is applied. The results are shown in Tables 5 and 6. Columns 6–9 in Table 5 are the number of changes detected by the algorithm. As in the previous case, the proposed procedure works quite well, with most of the detection frequencies of the two changepoints over 95%. The procedure works better if the number of events increases and the sizes of the changes are larger. It is important to note that the asymmetry of the changepoints does not appear to highly affect the efficiency of the procedure. On the other hand, columns 6–9 show the median and mean absolute deviation of the estimates of the changepoint locations. Note that the median of the estimates are quite close to the true ones. Also, it appears that the larger is the size of the change, the better is estimated its location. Finally, the percentage of false changepoints detected in both cases, one and two changepoints, is smaller than the nominal 5% in almost all the cases.

## 5. A real data example

The point process of dates of serious British coal-mining disasters is a frequently used data set for illustrating methods for changepoint analysis in Poisson processes. The data was initially gathered by Maguire et al. (1952), corrected by Jarrett (1979) and finally extended by Raftery and Akman (1986) to include until the 191th accident. Fig. 2 shows the times of occurrence of the 191 disasters. With the final set, Raftery and Akman (1986) assumed a single changepoint and estimated the posterior mode of the changepoint location between the 124th and 125th accidents. Carlin et al. (1992), assuming a single changepoint, estimated the mode of the change around the 127th accident. Green (1995) illustrated the Reversible Jump Markov Chain Monte Carlo (RJMCMC) method with this dataset and concluded that, assuming that the number of changepoints has a priori a Poisson distribution with mean three, the model with two changepoints has the largest posterior probability. The posterior modes of both changepoints were about the 122th and 182th accidents, respectively. Yang and Kuo (2001), using a binary segmentation algorithm in conjunction with Bayes factors, found one changepoint between the 124th and 125th accidents, but using a BIC approximation to the Bayes factor found two changepoints after the 124th and 186th accidents.

Here the proposed procedure is illustrated with the coal-mining disasters data, which allow us to compare it with previous approaches. Fig. 3 summarizes the results given by the procedure. We start by obtaining the value of the statistics in (2) for all the set of events, which are plotted in the first row in Fig. 3. A possible changepoint is found at $\widehat{n}=124$ ($\widehat{\tau}=\frac{124}{191}$), where the value of the statistic (5) is 4.152. Then, the dataset is splitted and the events up to and including the 124th accident are considered first. The statistics (2) in this period are plotted in the second row in Fig. 3. The value of the statistic (5)

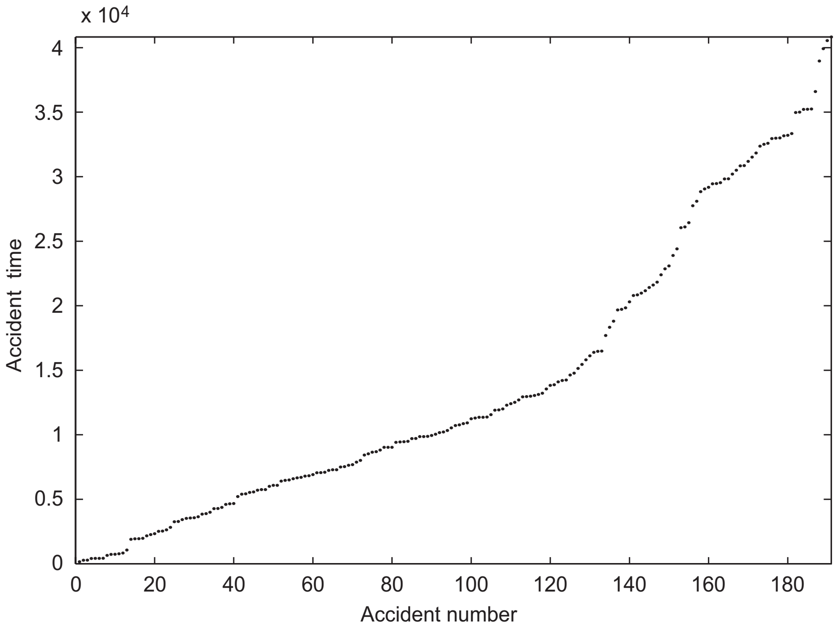


Fig. 2. Coal mining disasters.

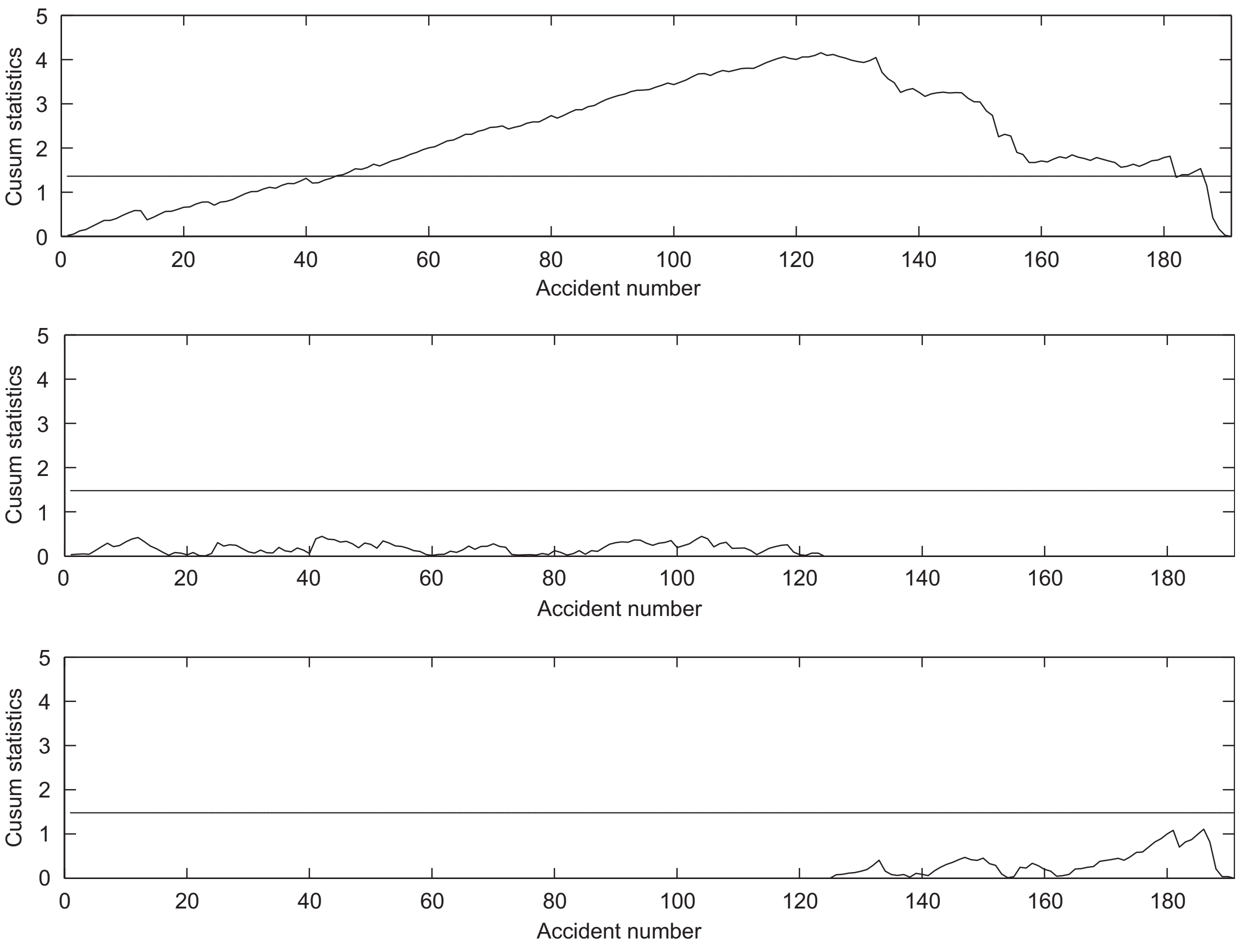


Fig. 3. Summary of the proposed procedure.

is 0.501 at $\widehat{n} = 104$. Next, the statistics (2) from the 125th accident up to and including the 191th accident are obtained and plotted in the last row in Fig. 3. The value of the statistic (5) is 1.125 at $\widehat{n} = 186$. As both statistics are less than 1.478, which is the corresponding critical value taken from Table 1, the proposed procedure supports that there exists only one changepoint between the 124th and 125th accidents, as in Raftery and Akman (1986) and Yang and Kuo (2001). A confidence interval for the changepoint at the 95% confidence level, obtained from (9), is (117,131). This interval is somewhat shorter than the ones given by Akman and Raftery (1986), Raftery and Akman (1986), Green (1995) and Jandhyala et al. (2000). The estimated rate values between consecutive accidents are $\widehat{\lambda}_1 = 0.0087$ and $\widehat{\lambda}_2 = 0.0025$, respectively. Based on the formulas (12) and (13), confidence intervals at the 95% confidence level for $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ are (0.0072,0.0102) and (0.0019,0.0031), respectively.

## 6. Conclusions

This paper has studied the problem of multiple changepoint detection in the rate parameter of a Poisson process. A cusum statistic have been proposed to test for a change at an unknown location. The asymptotic distribution of the test has been derived. Also, it has been shown the consistency of the estimator of the changepoint location and its asymptotic distribution has been obtained. Finally, the asymptotic distribution of the estimate of the rate has been derived. Both asymptotic distributions allow to obtain confidence intervals of the changepoint location and the piecewise rate values. Finally, in order to handle multiple changepoints, a binary segmentation procedure with a Bonferroni type adjustment has been implemented. Several Monte Carlo experiments have shown the efficiency of the proposed procedure, which has been illustrated with a real data example.

## Acknowledgments

The author thanks three anonymous referees for their excellent comments and suggestions which considerably improved the paper. The author gratefully acknowledges the support and hospitality of Ruey S. Tsay during his visit at the Graduate School of Business of the University of Chicago, in which this paper was written. The author also acknowledges financial support by Xunta de Galicia under the Isidro Parga Pondal Program and Ministerio de Educación y Ciencia Grant SEJ2004-03303.

## Appendix

This appendix provides the proofs and technical details of the Theorems 1–3.

**Proof of Theorem 1.** Let $\xi_i = d_i - \mu$, for $i = 1, \ldots, n$, such that $E[\xi_i] = 0$ and $E[\xi_i^2] = \mu^2$. The process $\xi_i$ satisfies the conditions of Theorem 3.4 in Phillips and Solo (1992). Therefore, as $\widehat{\mu} = (1/n)\sum_{i=1}^n d_i$ is a consistent estimator of the mean of the interarrival times $\mu$,

$$\frac{1}{\widehat{\mu}\sqrt{n}} \sum_{i=1}^{[n\tau]} \xi_i - \frac{[n\tau]}{n} \frac{1}{\widehat{\mu}\sqrt{n}} \sum_{i=1}^{n} \xi_i = \sqrt{n} \left( \frac{\sum_{i=1}^{[n\tau]} d_i}{\sum_{i=1}^{n} d_i} - \frac{[n\tau]}{n} \right) \xrightarrow{\text{d}} M^0(\tau),$$

which proves the stated result. $\square$

The proof of Theorem 2 is based on the Hájek-Rényi inequality for a transformation of the interarrival times. Once that the inequality has been established, the proof is almost standard and basically similar to the ones of Proposition 2 and Proposition 3 in Bai (1994) for changes in the mean of linear processes.

**Proof of Theorem 2.** Let, $\xi_i = d_i - \mu(i)$, for $i = 1, \ldots, n$. From the Hájek-Rényi inequality,

$$\Pr \left( \max_{m \leqslant i \leqslant n} c_i \left| \sum_{j=m}^{i} \xi_j \right| > \alpha \right) \leqslant \frac{1}{\alpha^2} \sum_{j=m}^{n} (c_j^2 E[\xi_j^2]) \leqslant \frac{\mu_{\max}^2}{\alpha^2} \sum_{j=m}^{n} c_j^2, \tag{A.1}$$

where $\{c_j : j = m, \ldots, n\}$ is any sequence of decreasing positive constants and $\mu_{\max}^2 = \max\{\mu_0^2, \mu_1^2\}$. Taking $c_j = 1/j$ and because $\sum_{j=m}^{\infty} j^{-2} = \mathrm{O}(m^{-1})$, from (A.1),

$$\Pr \left( \max_{m \leqslant i} \frac{1}{i} \left| \sum_{j=m}^{i} \xi_j \right| > \alpha \right) \leqslant \frac{Q_1}{\alpha^2 m}, \tag{A.2}$$

holds for some constant $Q_1 > 0$. On the other hand, taking $c_j = 1/\sqrt{j}$, and because $\sum_{j=1}^{n} j^{-1} = \mathrm{O}(\log n)$, from (A.1),

$$\Pr \left( \max_{1 \leqslant i \leqslant n} \frac{1}{\sqrt{i}} \left| \sum_{j=1}^{i} \xi_j \right| > \alpha \right) \leqslant \frac{\mu_{\max}^2}{\alpha^2} \sum_{j=1}^{n} j^{-1} \leqslant \frac{Q_2 \log n}{\alpha^2}, \tag{A.3}$$

holds for some constant $Q_2 > 0$.

Now, the statistic (3) can be rewritten as follows:

$$D_i = \sqrt{n} \left( \frac{\sum_{j=1}^{i} d_j}{\sum_{j=1}^{n} d_j} - \frac{i}{n} \right) = \frac{\sqrt{n}}{\overline{d}} \frac{i}{n} \left( 1 - \frac{i}{n} \right) \left( \overline{d}_i - \overline{d}_i^* \right), \quad i = 1, \ldots, n,$$

where $\overline{d}$, $\overline{d}_i$ and $\overline{d}_i^*$ are the sample means of the interarrival time sequences $\{d_1, \ldots, d_n\}$, $\{d_1, \ldots, d_i\}$ and $\{d_{i+1}, \ldots, d_n\}$, respectively. If $V_i = b_i(\overline{d}_i - \overline{d}_i^*)$, where $b_i = i/n(1 - i/n)$, then,

$$i_{\max} = \arg\max_i C_i = \arg\max_i |V_i|. \qquad \square$$

The expression $V_i$ is similar to the one used in Bai (1994) except for the term $b_i$. Now, using the expressions (A.1)–(A.3) and $V_i$, the rest of the proof is similar to the ones of Proposition 2, Proposition 3 and Theorem 1 in Bai (1994).

**Proof of Theorem 3.** Consider the case of $\widehat{\mu}_0 = \overline{d}_{\widehat{n}_1}$. Let

$$\begin{aligned} n^{1/2}\left(\frac{1}{\widehat{n}_1}\sum_{j=1}^{\widehat{n}_1} d_j - \frac{1}{n_1}\sum_{j=1}^{n_1} d_j\right) &= I(\widehat{n}_1 \leqslant n_1)\left(n^{\frac{1}{2}}\frac{n_1-\widehat{n}_1}{\widehat{n}_1 n_1}\sum_{j=1}^{\widehat{n}_1}\xi_j - n^{\frac{1}{2}}\frac{1}{n_1}\sum_{j=\widehat{n}_1+1}^{n_1}\xi_j\right) \\ &+ I(\widehat{n}_1 > n_1)\left(n^{\frac{1}{2}}\frac{n_1-\widehat{n}_1}{\widehat{n}_1 n_1}\sum_{j=1}^{n_1}\xi_j + n^{\frac{1}{2}}\frac{1}{\widehat{n}_1}\sum_{j=n_1+1}^{\widehat{n}_1}\xi_j + n^{\frac{1}{2}}\frac{\widehat{n}_1-n_1}{\widehat{n}_1}(\mu_1-\mu_0)\right), \end{aligned} \tag{A.4}$$

where $I$ is the indicator variable. From Theorem 2, $\widehat{n}_1 = n_1 + \mathrm{O_p}((\mu_1-\mu_0)^{-2})$ and $n(\mu_1-\mu_0)^2 \to \infty$. Therefore, (A.4) is $(n^{1/2}(\mu_1-\mu_0))^{-1}\mathrm{O_p}(1)$, which converges to 0 in probability. Thus, $\widehat{\mu}_0$ and $\widetilde{\mu}_0$ have the same asymptotic distribution, so that the same happens with $\widehat{\lambda}_0$ and $\widetilde{\lambda}_0$. The common asymptotic distribution is the stated in the Theorem in view of (10) and (11). The proof in the case of $\widehat{\mu}_1$ is similar. $\square$

## References


Akman, V.E., Raftery, A.E., 1986. Asymptotic inference for a change-point Poisson process. Ann. Statist. 14, 1583–1590.

Bai, J., 1994. Least squares estimation of a shift in linear processes. J. Time Ser. Anal. 15, 453–472.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Carlin, B.P., Gelfand, A.E., Smith, A.F.M., 1992. Hierarchical Bayesian analysis of changepoint problems. J. Roy. Statist. Soc. Ser. C 41, 389–405.

Carnero, M.A., Peña, D., Ruiz, E., 2003. Detecting level shifts in the presence of conditional heteroskedasticity. Working Paper, 03-63, Universidad Carlos III de Madrid.

Chib, S., 1998. Estimation and comparison of multiple change-point models. J. Econometrics 86, 221–241.

Edwards, A.W.F., Cavalli-Sforza, L.L., 1965. A method for cluster analysis. Biometrics 21, 362–375.

Galeano, P., Peña, D., Tsay, R.S., 2006. Outlier detection in multivariate time series by Projection Pursuit. J. Amer. Statist. Assoc. 101, 654–669.

Gan, F.F., 1994. Design of optimal exponential CUSUM control charts. J. Qual. Tech. 26, 109–124.

Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. Biometrika 82, 711–732.

Hawkins, D.M., 2001. Fitting multiple change-point models to data. Comput. Statist. Data Anal. 37, 323–341.

Hinkley, D., 1970. Inference about the change point in a sequence of random variables. Biometrika 57, 1–17.

Inclán, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. J. Amer. Statist. Assoc. 89, 913–923.

Jandhyala, V.K., Fotopoulos, S.B., 1999. Capturing the distributional behavior of the maximum likelihood estimation of a changepoint. Biometrika 86, 129–140.

Jandhyala, V.K., Fotopoulos, S.B., Evaggelopoulos, N.E., 2000. A comparison of unconditional and conditional solutions to the maximum likelihood estimation of a change-point. Comput. Statist. Data Anal. 34, 315–334.

Jarrett, R.G., 1979. A note on the intervals between coal-minning disasters. Biometrika 66, 191–193.

Karr, A.F., 1985. Point Processes and Their Statistical Inference. Marcel Dekker, New York.

Kenett, R.S., Pollak, M., 1996. Data-analytic aspects of the Shiryayev-Roberts control chart: Surveillance of a non-homogeneous Poisson process. J. Appl. Statist. 23, 125–137.

Maguire, B.A., Pearson, E.S., Wynn, A.H.A., 1952. The time intervals between industrial accidents. Biometrika 39, 168–180.

Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Ann. Statist. 20, 971–1001.

Raftery, A.E., 1994. Change point and change curve modeling in stochastic processes and spatial statistics. J. Appl. Statist. Sci. 1, 403–424.

Raftery, A.E., Akman, V.E., 1986. Bayesian analysis of a Poisson Process with a change-point. Biometrika 73, 85–89.

Siegmund, D., 1988. Confidence sets in change points problems. Int. Stat. Rev. 56, 31–48.

Vostrikova, L.J., 1981. Detecting "disorder" in multidimensional random processes. Sov. Math. Dokl. 24, 55–59.

Worsley, K.J., 1986. Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika 73, 91–104.

Yang, T.Y., Kuo, L., 2001. Bayesian binary segmentation procedure for a Poisson process with multiple changepoints. J. Comput. Graph. Statist. 10, 772–785.

Yao, Y., 1987. Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. Ann. Statist. 15, 1321–1328.