UNIVERSIDAD CARLOS III DE MADRID working papers

Working Paper 04-68
Statistics and Econometrics Series 16
December 2004

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

# USE OF CUMULATIVE SUMS FOR DETECTION OF CHANGEPOINTS IN THE RATE PARAMETER OF A POISSON PROCESS

Pedro Galeano*

**Abstract**

This paper studies the problem of multiple changepoints in rate parameter of a Poisson process. We propose a binary segmentation algorithm in conjunction with a cumulative sums statistic for detection of changepoints such that in each step we need only to test the presence of a simple changepoint. We derive the asymptotic distribution of the proposed statistic, prove its consistency and obtain the limiting distribution of the estimate of the changepoint. A Monte Carlo analysis shows the good performance of the proposed procedure, which is illustrated with a real data example.

**Keywords:** Binary segmentation procedure; Cumulative sums statistics; Changepoints; Poisson Processes.

* Galeano, Departamento de Estadística, Universidad Carlos III de Madrid, c/Madrid, 126, 28903 Getafe (Madrid), e-mail: pedro.galeano@uc3m.es. This paper was written while the author was visiting the Graduate School of Business at the University of Chicago. The author gratefully acknowledges the support and hospitality shown him by Ruey S. Tsay during his visit at this University. The author is very grateful for helpful comments from María Concepción Ausín.

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

Pedro Galeano

Departamento de Estadística, Universidad Carlos III de Madrid, Spain

**Abstract**

This paper studies the problem of multiple changepoints in rate parameter of a Poisson process. We propose a binary segmentation algorithm in conjunction with a cumulative sums statistic for detection of changepoints such that in each step we need only to test the presence of a simple changepoint. We derive the asymptotic distribution of the proposed statistic, prove its consistency and obtain the limiting distribution of the estimate of the changepoint. A Monte Carlo analysis shows the good performance of the proposed procedure, which is illustrated with a real data example.

Keywords: Binary segmentation procedure; Cumulative sums statistics; Changepoints; Poisson Processes.

## 1 Introduction

The problem of estimation of the rate parameter of a Poisson process when the rate is piecewise linear with unknown number of pieces and locations is a widely studied problem in different disciplines. Maguire et al (1952), Akman and Raftery (1986), Worsley (1986) and Siegmund (1988) gave frequentist analysis by proposing several statistics to test for one changepoint at a given location. Lately, most of the works dealing with changepoints in Poisson processes are from a Bayesian perspective. Raftery and Akman (1986), Carlin et al (1992) and Raftery (1994) study the single changepoint case. Multiple changepoints are analyzed by Green (1995), who proposed a reversible jump Markov chain Monte Carlo method for locating the multiple changepoints and estimating the corresponding rate parameters. Chib (1998) proposed a hidden Markov model an a MCMC algorithm for its estimation. Finally, Yang and Kuo (2001) proposed the use of a Bayesian binary segmentation algorithm based on a sequence of nested hypothesis tests using Bayes factors and a BIC approximation to the Bayes factors.

The propose of this paper is to develop a cumulative sum approach for detection of several changepoints at unknown locations in Poisson processes using a binary segmentation procedure. This type of algorithm has been extensively used in changepoint detection and Vostrikova (1981) proves its consistency for detecting changepoints in multidimensional random processes under mild conditions. This algorithm is based on consecutive divisions of the whole sample into pieces after detecting one change, such that it is only required testing for zero changes against one change in each step. This overcomes the complexity of testing for several changes at the same time. To test for the presence of a changepoint, we use a cumulative sums statistic for which we have derived its asymptotic distribution and proved the consistency of the estimate of the changepoint. We also have obtained the distribution of the estimate of the changepoint, that can be used for obtain confidence intervals of the locations of the changepoints. The procedure here proposed is efficient, fast and very easy to implement. The binary segmentation algorithm in conjunction with cumulative sums statistics have been used previously by Inclán and Tiao (1994), Carnero et al (2003) and Galeano et al (2004), to locate multiple mean and variance changepoints in univariate and multivariate time series. Other successfully implementations of the algorithm can be found in Chen and Gupta (1997), Yang and Kuo (2001) and Yang (2004).

The rest of this article is organized as follows. In section 2, we present a cumulative sum statistic that can be used for testing for changes in the rate parameter of a Poisson process, obtain its asymptotic distribution, prove its consistency and obtain the limiting distribution of the estimate of the location of the changepoint. In section 3, a binary segmentation procedure for detection and estimation of these changes is proposed. In section 4, we study the performance of the procedure in several Monte Carlo experiments for different models, sample sizes, number and location of the changepoints. We conclude that the proposed procedure yields satisfactory results for changepoints detection in all the situations considered. Finally, section 5 illustrates the procedure by means of a real data example.

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

Suppose that we observe $n$ independent events occurring in the interval $(0, T]$, $x = (x_1, \ldots, x_n)'$, such that $0 < x_1 \leq \ldots \leq x_n < T$, taken from a Poisson process with rate parameter $\lambda(t)$, given by:

$$\lambda(t) = \begin{cases} \lambda_0 & h_0 < t \leq h_1 \\ \vdots & \vdots \\ \lambda_p & h_p < t \leq h_{p+1} \end{cases} \tag{1}$$

where $h_0 = 0 < h_1 < \cdots < h_p < h_{p+1} = T$, $\lambda_j \neq \lambda_{j+1}$, $j = 0, \ldots, p-1$ and $p$ are unknown parameters. Here $p$ is the number of changepoints and $h_i$ is the location of the $j-$th changepoint. The problem is to estimate the true number and locations of the changepoints based either on the arrival times, $x = (x_1, \ldots, x_n)'$, or the interarrival times between consecutive events, $y = (y_1, \ldots, y_n)'$, given by $y_1 = x_1$, $y_i = x_i - x_{i-1}$, for $i = 2, \ldots, n$. These interarrival times are independent, identically distributed exponential random variables with rate parameter $\lambda(t)$. Let,

$$D_i = \sqrt{n}\left(\frac{x_i}{x_n} - \frac{i}{n}\right) = \sqrt{n}\left(\frac{\sum_{j=1}^{i} y_j}{\sum_{j=1}^{n} y_j} - \frac{i}{n}\right) \tag{2}$$

be the centered and normalized cumulative sums of the interarrival times. The statistic $D_i$ compares the cumulative sum of the interarrival times until time $x_i$ with respect to the cumulative sum of all the interarrival times. In the case of a constant rate, the ratio between both cumulative sums should be around $i/n$, but if the rate is piecewise linear, the ratio can be very different from $i/n$. To see this, we explore the behavior of the statistic $D_i$ in (2) under several situations which are illustrated in Figure 1. The three columns in this matrix of plots represents three different generating processes. The first column corresponds to the case of constant rate, $\lambda(t) = 1$. The second column corresponds to the case of a single change in the rate at $h_1 = x_{[n/2]}$, where the rate changes from $\lambda(t) = 1$ to $\lambda(t) = 2$. The third column corresponds to the case of two changes, the first at $h_1 = x_{[n/3]}$, where the rate changes as in the previous case, and the second at $h_2 = x_{[2n/3]}$, where the rate goes back to be $\lambda(t) = 1$. The rows represent the cumulative sums of the time events from the Poisson process and the corresponding statistics $D_i$. The first row in Figure 1 shows $n = 500$ time events generated from the first process, and the second row shows the $D_i$ statistic computed with the event times in the same column. In the first column in Figure 1, the constant rate case, the statistics plotted in the second row are under the two straight lines computed as explain next as for the 95% critical value of the distribution of the maximum of the statistics $D_i$ in absolute value. In the second column, a single change at $h_1 = x_{250}$, the maximum of the statistics in absolute value is around the event $i = 250$, and is larger than the critical value, so the hypothesis of no change is rejected. In the third column, two rate changes at $h_1 = x_{166}$ and $h_2 = x_{333}$, which appear as two significant extremes around the times of the changes. This behavior leads to search for a changepoint as follows. Let $i_{\max}$ be the value of $i$ at which $\max_i \{|D_i| : i = 1, \ldots, n\}$ is achieved. If this maximum exceeds a given boundary, we may conclude that there exists a changepoint and that $x_{i_{\max}}$ is the estimate of its location.

In the next section, we propose a binary segmentation algorithm such that in each step we need only to test the null hypothesis of no change against the alternative of one change at an unknown point in different pieces of the data. Therefore, we only need to explore the asymptotic behavior of the statistic (2) under the null hypothesis of no change and the alternative hypothesis of one change, respectively. First, we show that under the null hypothesis, the statistic (2) behaves like a Brownian bridge asymptotically. The proofs of Lemma 1 and Theorem 2 are in the appendix.

**Lemma 1** *Let $x = (x_1, \ldots, x_n)'$ be $n$ events occurring in the interval $(0, T]$ such that $0 < x_1 \leq \ldots \leq x_n < T$, taken from a Poisson process with constant rate parameter $\lambda(t) = \lambda_0$ and let $y = (y_1, \ldots, y_n)'$ be the corresponding interarrival times. Then, for every $i = 1, \ldots, n$,*

$$E[D_i] = o\left(n^{-\frac{1}{2}}\right).$$

Consequently, the mean of the statistic $D_i$ is asymptotically 0 for every $i = 1, \ldots, n$. Let $M$ be a Brownian motion process verifying $E[M(r)] = 0$, and $E[M(r) M(s)] = s$, where $0 \leq s < r \leq 1$. Let $M^0$ denote a Brownian bridge given by $M^0(r) = M(r) - rM(1)$, verifying $E\left[M^0(1)\right] = 0$,

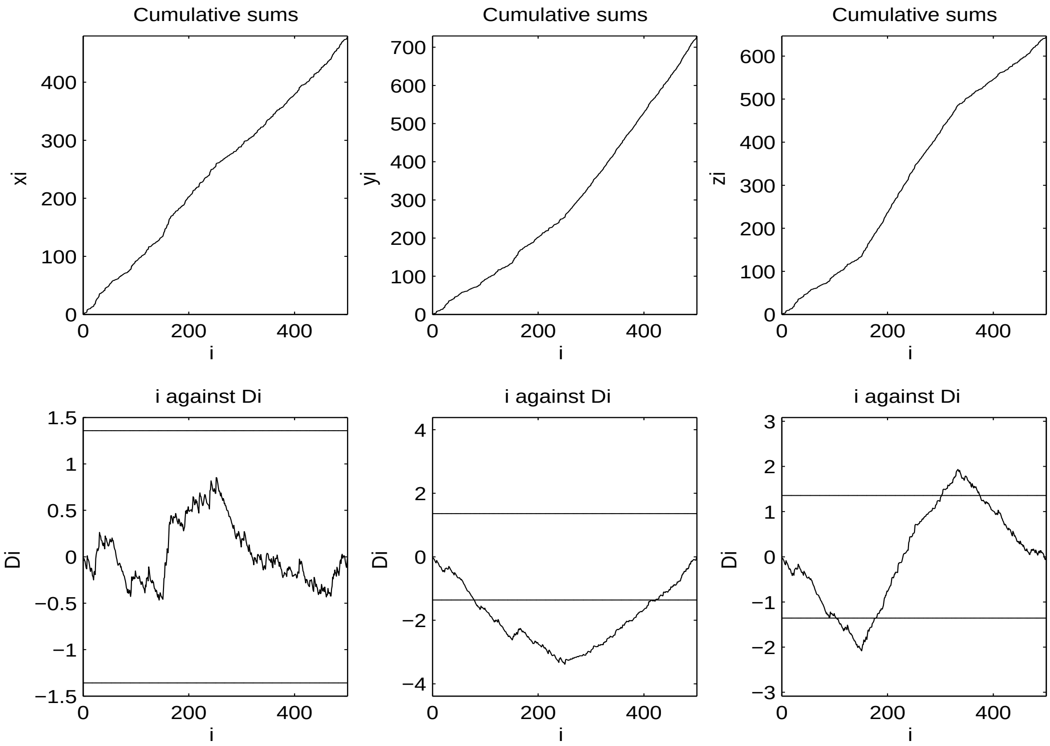


Figure 1: Three different Poisson processes (up) and the corresponding $D_h$ statistics for change-points detection (down). The first column is the no changes case, the second column a change at $h = 250$ and the third to two changes at $h = 166$ and $h = 333$.

$E\left[M^0(r)\, M^0(s)\right] = s(1-r)$, $0 \leq s < r \leq 1$, and $M^0(0) = M^0(1) = 0$, with probability 1. The asymptotic distribution of the statistic $D_i$ is obtained in the following theorem.

**Theorem 2** *Under the conditions of Lemma 1, the statistic* $D_i \xrightarrow{d} M^0$.

As we mention previously, we are interested in the maximum of the absolute value of statistics (2) which is given by,

$$\Lambda_{\max}(i_{\max}) = \max\{|D_i|, 1 \leq i \leq n\}, \tag{3}$$

where $i_{\max}$ is the point in which the statistic $|D_i|$ achieves its maximum, and $x_{i_{\max}}$ is the estimate of the location of the changepoint. The distribution of $\Lambda_{\max}$ in (3) is asymptotically the distribution of sup $\left\{\left|M^{0}(r)\right| : 0 \leq r \leq 1\right\}$ which is given by (see, Billingsley, pg. 85, 1968),

$$P\left\{\sup\left|M^{0}(r)\right| \leq a : 0 \leq r \leq 1\right\} = 1 + 2\sum_{j=1}^{\infty}(-1)^{j}\exp\left(-2j^{2}a^{2}\right),$$

and critical values can be obtained from this distribution.

Next, we study the behavior of the statistic (3) under the alternative hypothesis of one change at $t = h$. Let $W$ be a two-sided Brownian motion process defined in the interval $[-w, w]$ for every $w > 0$, such that $W(w) = M_1(-w)$ for $w < 0$ and $W(w) = M_2(w)$ for $w \geq 0$, where $M_1$ and $M_2$ are two independent Brownian motions. Let $i_h = \arg\min\{|x_i - h| : i = 1, \ldots, n\}$ and assume that $i_h \to \infty$ and $n \to \infty$, such that $i_h/n = \tau_h$, a constant. The following theorem, which proof is in the appendix, states that $i_{\max}$ estimated according to (3) is consistent to the true changepoint, $h$, with probability approaching 1 as $n \to \infty$, and obtain the limiting distribution of the changepoint estimator $i_{\max}$.

**Theorem 3** *Let $x = (x_1, \ldots, x_n)'$ be $n$ events occurring in the interval $(0, T]$ such that $0 < x_1 \leq \ldots \leq x_n < T$, taken from a Poisson process with rate parameter $\lambda(t)$ given by,*

$$\lambda(t) = \begin{cases} \lambda_0 & 0 < t \leq h \\ \lambda_1 & h < t \leq T \end{cases} \tag{4}$$

*and let $y = (y_1, \ldots, y_n)'$ be the corresponding interarrival times. Let $h$ be the true position of the changepoint under (4). Let $x_{i_{\max}}$ be the estimate of $h$ given by (3). Assume that $i_{\max}/n = \tau_{\max}$ and $i_h/n = \tau_h$. Then,*

1. *$\tau_{\max}$ is a consistent estimator of $\tau$ with,*

$$|\tau_{\max} - \tau_h| = O_p\left(n^{-1}|\lambda_0 - \lambda_1|^{-1}\right). \tag{5}$$

2. *If in addition, we assume that,*

   (a) *$|\lambda_0 - \lambda_1|$ depends on $n$ such that $|\lambda_0 - \lambda_1| \to 0$, if $n \to \infty$, and,* *(b)* $n^{\frac{1}{2}} |\lambda_0 - \lambda_1| / (\log n)^{\frac{1}{2}} \to \infty$,

*therefore,*

$$\frac{n \left|\widehat{\lambda}_0 - \widehat{\lambda}_1\right|^2 (\tau_{\max} - \tau_h)}{\widehat{\sigma}_y^2} \xrightarrow{d} \arg\max_v \left\{ W(v) - \frac{1}{2}|v| \right\}, \tag{6}$$

*where* $W(v)$ *is a two-sided Brownian motion,* $\widehat{\lambda}_1 = \overline{y}_{i_{\max}}$, $\widehat{\lambda}_2 = \overline{y}^*_{i_{\max}}$, *where* $\overline{y}_{i_{\max}}$ *and* $\overline{y}^*_{i_{\max}}$ *are the means of the sequences* $\{y_1, \ldots, y_{i_{\max}}\}$ *and* $\{y_{i_{\max}+1}, \ldots, y_n\}$, *respectively, and* $\widehat{\sigma}_y^2$ *is the estimated variance of* $\{y_i : i = 1, \ldots, n\}$.

Therefore, accordingly with Theorem 3, under the alternative hypothesis of one change, the statistic (3) will detect the presence and location of the changepoint asymptotically. As $n \to \infty$ implies $T \to \infty$, $(x_{i_h} - h)/T \to 0$. The rate of convergence (5) is a typical rate associated with changepoint estimators, while the limiting distribution (6) can be used to obtain confidence intervals for the location of the changepoint. For that, an expression of the cumulative distribution function of $\arg\max_v \left\{W(v) - \frac{1}{2}|v|\right\}$ can be found in Bai (1997). Finally, the limiting distributions of $\widehat{\lambda}_1 = \overline{y}_{i_{\max}}$ and $\widehat{\lambda}_2 = \overline{y}^*_{i_{\max}}$ are given in Theorem 4.

**Theorem 4** *Under the conditions of Theorem 3,*

1. $n^{\frac{1}{2}}\widehat{\lambda}_1/\lambda_1$ *has a limiting Gamma distribution with mean* $n^{\frac{1}{2}}$ *and variance* $1/\tau_h$.

2. $n^{\frac{1}{2}}\widehat{\lambda}_2/\lambda_2$ *has a limiting Gamma distribution with mean* $n^{\frac{1}{2}}$ *and variance* $1/(1 - \tau_h)$.

Thus, we can estimate consistently the rate parameters $\lambda_1$ and $\lambda_2$. Moreover, the limiting distributions are the same as if $h$ is assumed known.

## 3 Detection of multiple changes: Binary segmentation algorithm

If several changes have occurred in the data, the usefulness of the statistic (3) 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, while Vostrikova (1981) used it for detection of changepoints in multidimensional random processes. It is based on splitting the data into two pieces when a change is detected, thus the algorithm tries to isolate each changepoint. The proposed algorithm proceeds as follows:

1. Let $i_1 = 1$. Obtain $\Lambda_{\max}(i_{\max})$ in (3) for $i = 1, \ldots, n$. If $\Lambda_{\max}(i_{\max}) > C_\alpha$, where $C_\alpha$ is the asymptotic critical value for a critical level, $\alpha$, go to step 2. If $\Lambda_{\max}(i_{\max}) < C_\alpha$, it is assumed that there is not a change in the sequence and the procedure ends.

2. Step 2 has three substeps:

   (a) Obtain $\Lambda_{\max}(i_{\max})$ for $i = 1, \ldots, i_2$, where $i_2 = i_{\max}$. If $\Lambda_{\max}(i_{\max}) > C_\alpha$, redefine $i_2 = i_{\max}$ and repeat Step 2(a) until $\Lambda_{\max}(i_{\max}) < C_\alpha$. When this happens, define $i_{first} = i_2$ where $i_2$ is the last value such that $\Lambda_{\max}(i_{\max}) > C_\alpha$.

   (b) Repeat a similar search in the interval $i_2 \leq i \leq n$, where $i_2$ is the point $i_{\max}$ obtained in Step 1. For that, define $i_1 = i_{\max} + 1$, where $i_{\max} = \arg\max\{|D_i| : i = i_1, \ldots, n\}$ and repeat it until $\Lambda_{\max}(i_{\max}) < C_\alpha$. Define $i_{last} = i_1 - 1$, where $i_1$ is the last value such that $\Lambda_{\max}(i_{\max}) > C_\alpha$.

   (c) If $|i_{last} - i_{first}| < d$, there is just one change point and the algorithm ends here. Otherwise, keep both values as possible changepoints and repeat Steps 1 and 2 for $i_1 = i_{first}$ and $n = i_{last}$, until no more possible change points are detected. Then, go to step 3.

3. Define a vector $\ell = (\ell_1, \ldots, \ell_s)$ where $\ell_1 = 1$, $\ell_s = n$ and $\ell_2, \ldots, \ell_{s-1}$ are the points detected in Steps 1 and 2 in increasing order. Obtain the statistic $D_i$ in each one of the intervals $(\ell_i, \ell_{i+2})$ and check if its maximum in absolute value is still significant. If it is not, eliminate the corresponding point. Repeat Step 3 until the number of possible changepoints does not change, and the points found in previous iterations do not differ from those in the last one. The vector $(\ell_2, \ldots, \ell_{s-1})$ are the final changepoints.

4. Finally, we estimate the rate parameter (1) by means of the sample mean in each of the intervals between changes.

Some comments regarding the proposed binary segmentation algorithm are in order. First, the critical values used in the procedure are the asymptotic critical values of the maximum absolute value of a Brownian Bridge. In the next section we provide some of them. Second, we use different critical values depending on the number of changepoints detected in each step of the algorithm. If we use the same critical value in steps 1 and 2, this can leads to overdetect the number of changepoints. To see this point, assume that we use the same critical value, $C_\alpha$, for a given critival level, $\alpha$. In step 1, the probability of detecting one changepoint, unless it does not exist, is $1-(1-\alpha)$. Assume now that we detect a changepoint. Therefore, the probability of detecting at least a second changepoint, unless it does not exist, is $1-(1-\alpha)^2 > \alpha$, and this can be repeated iteratively. Therefore, each time that we detect a new changepoint, there is an increasing probability of detecting a spurious changepoint in after splitting the data again. To avoid this problem, we proceed by taking a critical value, denoted by $\alpha_m$, after detecting the $m$ changepoint, verifying $\alpha_0 = 1-(1-\alpha_m)^{m+1}$, where $\alpha_0$ is the critical level used in step 1, usually $\alpha_0 = 0.05$. This ensures that the probability of detecting a false changepoint is always the same in each step. Third, as in Inclán and Tiao (1994), we include the step 3 in the procedure for avoiding any remaining false change. In this step, we use the critical level $\alpha_0$. Finally, we require a minimum distance between changes larger than an positive integer number $d$. In the Monte Carlo experiments and the real data example of the next sections we have taken $d = n/10$ because this election works well in the simulations.

# 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). Although the asymptotic distribution of the statistic $\Lambda_{\max}$ in (3) is known, we study the finite sample behavior of the quantiles of this statistic under the hypothesis of no change. For that, we generate 10000 realizations from an exponential distribution with rate parameter $\lambda(t) = 1$, which are considered as the interarrival times of the Poisson Process, for each of the sample sizes, $n = 100$, 200, 500 and 1000 and compute the statistics (3). Table 1 provides some quantiles of the distribution of $\Lambda_{\max}$ for different sample sizes under the null hypothesis of no change in the rate parameter. As we can see, the finite sample quantiles are always smaller than

Table 1: Empirical quantiles of the $\Lambda_{\max}$ statistics based on 10000 realizations for $n$ = 100, 200, 500 and 1000.

| Probability | 0.05 | 0.50 | 0.95 | 0.975 | 0.983 | 0.987 | 0.990 | 0.991 | 0.992 | 0.993 | 0.994 | 0.995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| n=100 | 0.453 | 0.754 | 1.280 | 1.395 | 1.461 | 1.502 | 1.538 | 1.568 | 1.598 | 1.613 | 1.632 | 1.648 |
| n=200 | 0.471 | 0.784 | 1.316 | 1.430 | 1.504 | 1.541 | 1.577 | 1.597 | 1.628 | 1.634 | 1.651 | 1.670 |
| n=500 | 0.491 | 0.797 | 1.324 | 1.449 | 1.515 | 1.567 | 1.594 | 1.623 | 1.640 | 1.652 | 1.668 | 1.680 |
| n=1000 | 0.501 | 0.804 | 1.336 | 1.458 | 1.519 | 1.570 | 1.598 | 1.632 | 1.652 | 1.657 | 1.672 | 1.694 |
| n=$\infty$ | 0.520 | 0.828 | 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.

| $\lambda$ | $n$ | $h$ | **frequency** | | |
|---|---|---|---|---|---|
| | | | 0 | 1 | $\geq 2$ |
| 1 | 100 | — | 96.3 | 3.6 | 0.1 |
| 1 | 200 | — | 96.1 | 3.8 | 0.1 |
| 1 | 500 | — | 95.8 | 4.0 | 0.2 |
| 1 | 1000 | — | 95.4 | 4.2 | 0.4 |

the asymptotic ones implying that the use of the asymptotic quantile is a conservative decision and therefore, the type I error will not increase.

First, we analyze the type I error of the proposed procedure by generating 10000 realizations from a Poisson process with constant rate parameter $\lambda(t) = 1$ for each of the sample sizes $n$ = 100, 200 and 500 and apply the proposed procedure with the critical values from Table 1, starting with $C_{\alpha_0} = 1.358$ which corresponds to the critical level $\alpha_0 = 0.05$. The results are shown in Table 2, where columns 4 to 6 report the number of changepoints detected by the algorithm. The type I errors are around 5 % in all the sample sizes considered.

Next, we consider the case of one changepoint and make a Monte Carlo experiment in order to study the size and power of the proposed procedure. For that, for each $n$ = 100, 200 and 500, we consider three locations of the change point, $h = x_{[0.25n]}$, $x_{[0.50n]}$ and $x_{[0.75n]}$. The changes are introduced by transforming the original rate parameter $\lambda(t) = 1$, into one of the rates, $\lambda_1 = 0.5$, $\lambda_2 = 0.25$, $\lambda_3 = 2$ and $\lambda_4 = 4$. For each case, we generate 10000 realizations. Then, we apply the proposed procedure with the critical values from Table 1, starting with $C_{\alpha_0} = 1.358$ which

Table 3: Results for one changepoint.

| $\lambda$ | $n$ | $h$ | frequency | | | $\widehat{h}$ | | | $\lambda$ | $n$ | $h$ | frequency | | | $\widehat{h}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | 1 | $\geq 2$ | Mod. | Med. | Mad | | | | 0 | 1 | $\geq 2$ | Mod. | Med. | Mad |
| | | 25 | 29.3 | 70.0 | 0.7 | 25 | 25 | 3 | | | 25 | 57.5 | 41.9 | 0.6 | 25 | 38 | 9 |
| $\lambda_1$ | 100 | 50 | 14.2 | 84.5 | 1.4 | 50 | 48 | 3 | $\lambda_3$ | 100 | 50 | 14.7 | 84.2 | 1.1 | 50 | 52 | 3 |
| | | 75 | 57.5 | 41.7 | 0.8 | 75 | 63 | 9 | | | 75 | 29.6 | 69.5 | 0.9 | 75 | 75 | 3 |
| | | 50 | 0.0 | 97.8 | 2.2 | 50 | 50 | 1 | | | 50 | 13.6 | 84.0 | 2.4 | 50 | 60 | 9 |
| $\lambda_1$ | 200 | 100 | 0.0 | 97.1 | 2.9 | 100 | 99 | 1 | $\lambda_3$ | 200 | 100 | 0.7 | 97.2 | 2.1 | 100 | 102 | 3 |
| | | 150 | 0.0 | 97.0 | 3.0 | 150 | 145 | 5 | | | 150 | 4.6 | 93.9 | 1.5 | 150 | 150 | 3 |
| | | 125 | 0.0 | 97.1 | 2.9 | 125 | 125 | 4 | | | 125 | 0.0 | 96.1 | 3.9 | 125 | 135 | 10 |
| $\lambda_1$ | 500 | 250 | 0.0 | 96.6 | 3.4 | 250 | 248 | 3 | $\lambda_3$ | 500 | 250 | 0.0 | 96.5 | 3.5 | 250 | 252 | 3 |
| | | 375 | 0.0 | 96.7 | 3.3 | 375 | 365 | 10 | | | 375 | 0.0 | 97.1 | 2.9 | 375 | 375 | 4 |
| | | 25 | 0.1 | 98.8 | 1.2 | 25 | 25 | 1 | | | 25 | 0.8 | 96.8 | 2.4 | 25 | 29 | 4 |
| $\lambda_2$ | 100 | 50 | 0.0 | 98.0 | 2.0 | 50 | 49 | 1 | $\lambda_4$ | 100 | 50 | 0.0 | 97.5 | 2.5 | 50 | 51 | 1 |
| | | 75 | 0.8 | 97.2 | 2.0 | 75 | 71 | 4 | | | 75 | 0.1 | 98.7 | 1.3 | 75 | 75 | 1 |
| | | 50 | 0.0 | 97.7 | 2.3 | 50 | 50 | 1 | | | 50 | 0.0 | 97.0 | 3.0 | 50 | 54 | 4 |
| $\lambda_2$ | 200 | 100 | 0.0 | 96.8 | 3.2 | 100 | 99 | 1 | $\lambda_4$ | 200 | 100 | 0.0 | 97.0 | 3.0 | 100 | 101 | 1 |
| | | 150 | 0.0 | 97.1 | 2.9 | 150 | 145 | 5 | | | 150 | 0.0 | 97.9 | 2.1 | 150 | 150 | 1 |
| | | 125 | 0.0 | 96.6 | 3.4 | 125 | 125 | 1 | | | 125 | 0.0 | 96.6 | 3.4 | 125 | 130 | 5 |
| $\lambda_2$ | 500 | 250 | 0.0 | 96.1 | 3.9 | 250 | 249 | 1 | $\lambda_4$ | 500 | 250 | 0.0 | 96.0 | 4.0 | 250 | 251 | 1 |
| | | 375 | 0.0 | 96.9 | 3.1 | 375 | 370 | 5 | | | 375 | 0.0 | 96.8 | 3.2 | 375 | 375 | 1 |

correspond to the critical level $\alpha_0 = 0.05$. The results are shown in Table 3, where columns 4 to 6 and 13 to 15 report the number of rate changes detected by the algorithm and columns 7 to 9 and 16 to 18 show the mode, median and mean absolute deviation of the estimates of the locations of the changepoints for each case. From Table 3, we conclude that the procedure performs quite well with most of the cases over the 90% of detection frequency. As the sample size increases and the change is larger the procedure works better. The results also suggest that the procedure detects more frequently the changes located at the middle of the sample when the sample size and the size of the change are small. We note that, in all the experiments considered, the estimated modes of the locations of the changepoints coincide with the true locations.

For two changepoints, we consider the same sample sizes and the situation of the changepoints at $(h_1, h_2) = \left(x_{[0.33n]}, x_{[0.66n]}\right)$. Each changepoint is associated with two rates $\lambda_i$, $i = 1, 2, 3, 4$, which give the rate parameter of the process after each change. Four combinations are considered.

Table 4: Results for two changepoints.

| $\lambda_1$ | $\lambda_2$ | $n$ | $h_1$ | $h_2$ | frequency | | | | $\widehat{h}_1$ | | | $\widehat{h}_2$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 0 | 1 | 2 | $\geq 3$ | Mod. | Med. | Mad | Mod. | Med. | Mad |
| | | 100 | 33 | 66 | 2.1 | 59.9 | 37.2 | 0.8 | 33 | 31 | 2 | 66 | 67 | 1 |
| 0.5 | 2 | 200 | 66 | 133 | 0.2 | 15.4 | 81.9 | 2.5 | 66 | 64 | 3 | 133 | 134 | 1 |
| | | 500 | 166 | 333 | 0.1 | 0.0 | 95.4 | 4.5 | 166 | 164 | 3 | 333 | 334 | 1 |
| | | 100 | 33 | 66 | 0.3 | 25.4 | 72.8 | 1.4 | 33 | 32 | 1 | 66 | 67 | 1 |
| 0.25 | 4 | 200 | 66 | 133 | 0.2 | 3.0 | 94.2 | 2.6 | 66 | 65 | 1 | 133 | 134 | 1 |
| | | 500 | 166 | 333 | 0.1 | 0.0 | 95.3 | 4.6 | 166 | 165 | 1 | 333 | 334 | 1 |
| | | 100 | 33 | 66 | 1.6 | 40.8 | 57.2 | 0.4 | 33 | 35 | 2 | 66 | 65 | 1 |
| 2 | 0.5 | 200 | 66 | 133 | 0.0 | 7.7 | 90.6 | 1.7 | 66 | 68 | 3 | 133 | 132 | 1 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 96.8 | 3.2 | 166 | 168 | 3 | 333 | 332 | 1 |
| | | 100 | 33 | 66 | 0.0 | 0.3 | 98.7 | 1.0 | 33 | 34 | 1 | 66 | 65 | 1 |
| 4 | 0.25 | 200 | 66 | 133 | 0.0 | 0.0 | 97.9 | 2.1 | 66 | 67 | 1 | 133 | 132 | 1 |
| | | 500 | 166 | 333 | 0.0 | 0.0 | 97.0 | 3.0 | 166 | 167 | 1 | 333 | 332 | 1 |

For each case, we generate 10000 realizations with the corresponding changes. Then, we apply the proposed procedure with the same critical values that in the previous case. The results are shown in Table 4. Columns 6 to 9 in the table are the number of rate changes detected by the algorithm, and columns 10 to 15 show the mode, median and mean absolute deviation of the estimates of the change points. For two changepoints, the proposed procedure works quite well, with several cases over the 90 % of detection frequency. As in the previous case, as the sample size increases and the size of the change is larger, the procedure works better. It also appears that the estimate of the second changepoint has smallest mad, suggesting that the procedure detects more precisely the change at the end of the series. The median of the estimates are quite closed to the real locations of the changepoints, except with the smallest sample size and the smallest changes. As in the case of one changepoint, the modes of the estimated locations coincide with the true ones. Finally, the percentage of false changepoints detected in both cases, one and two changepoints, is always smaller than the nominal 5%.

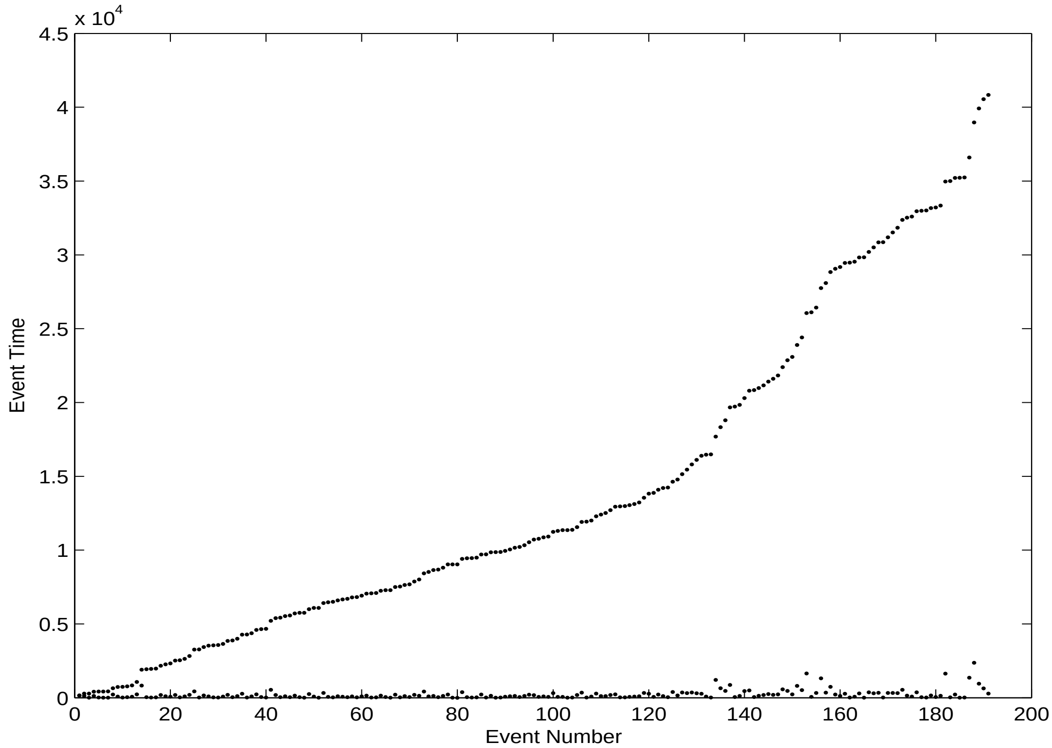


Figure 2: Coal Mining disasters.

# 5 A real data example

The point processes of dates of serious British coal-mining disasters is a data set frequently used in illustrating methods for changepoint analysis. The data was initially gathered by Maguire et al (1952), corrected by Jarrett (1979) and finally extended by Raftery and Akman (1986) to the period between January 1, 1851, and December 31, 1962. Figure 2 shows the dates of the 192 disasters together with the cumulative counting process. With the final set, Raftery and Akman (1986) assumed a single changepoint and estimated the posterior mode of the location of the change 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 method with this dataset and concluded that a model with two changepoints has the largest posterior mode. The posterior modes of both changepoints are about the 122th and 182th accidents, respectively. Yang and Kuo (2001), using the Binary segmentation

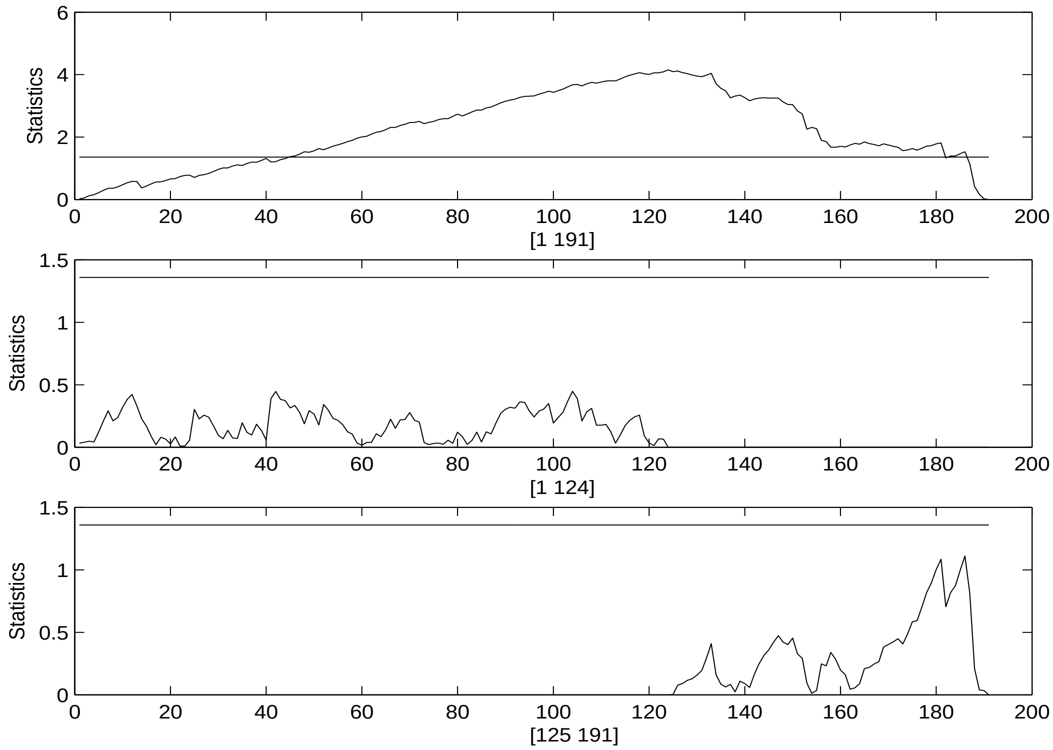


Figure 3: Summary of the proposed procedure.

algorithm with the Bayes factors, found one changepoint between the 124th and 125th accidents, but using a BIC approximation to the Bayes factors found two changepoints around the 124th and 186th accidents.

Here we illustrate the proposed procedure applied to the coal-mining disasters data. Figure 3 summarize the procedure. We start obtaining the value of the statistics in (2) for all the data and are plotted in absolute value at the first row in Figure 3. We find a possible changepoint at $i = 124$ or $h = x_{124}$, where the value of the statistic (3) is 4.152. Then, we cut the data and consider the accidents in the interval $[0, x_{124}]$. The statistics (2) in this period are plotted in second row in Figure 3. The value of the statistic (3) is 0.447 at $i = 104$, so that $i_{first} = 124$. Now we consider the second part, where the statistics (2) are plotted in the last row of Figure 3. The value of the statistic (3) is 1.110 at $i = 186$, so that $i_{last} = 125$. We conclude that 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 can be found based on the limiting distribution given in Theorem 3. In this case, $(117, 131)$ and $(x_{117}, x_{131})$ are the confidence intervals at the 95 % significance level for the index event and time event of the changepoint, respectively. This interval is somewhat shorter that the ones given by Akman and Raftery (1986), Raftery and Akman (1986) and Green (1995). The estimated expected value between consecutive accidents are $\widehat{\lambda}_1 = 114.83$ and $\widehat{\lambda}_2 = 391.54$ in both periods, respectively. Based on the limiting distributions given in Theorem 4, confidence intervals at the 95 % significance level for $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ are $(97.03, 138.07)$ and $(330.83, 470.75)$, respectively.

**Acknowledgements**

This paper was written while the author was visiting the Graduate School of Business at the University of Chicago. The author gratefully acknowledges the support and hospitality shown him by Ruey S. Tsay during his visit at this University. The author is very grateful for helpful comments from María Concepción Ausín.

# 6 Appendix

**Proof of Lemma 1.** The second order Taylor expansion of the ratio $x_i/x_n$ about the value $(E[x_i], E[x_n])$ is:

$$E\left[\frac{x_i}{x_n}\right] = \frac{E[x_i]}{E[x_n]} - \frac{E[x_i x_n]}{E[x_n]^2} + \frac{E[x_i]\, E[x_n^2]}{E[x_n]^3} + o\left(n^{-1}\right).$$

Taking into account that $E[x_i] = i\lambda$, and

$$E[x_i x_n] = E\left[x_i\left(\sum_{j=1}^{i} y_j + \sum_{j=i+1}^{n} y_j\right)\right] = E\left[x_i^2\right] + E\left[\sum_{j=1}^{i} y_j\right] E\left[\sum_{j=i+1}^{n} y_j\right] =$$
$$= i\lambda^2 + i^2\lambda^2 + i(n-i)\lambda^2 = (n+1)\, i\lambda^2,$$

the ratio $E[x_i/x_n]$ can be written as:

$$E\left[\frac{x_i}{x_n}\right] = \frac{i}{n} - \frac{(n+1)\, i}{n^2} + \frac{(n+1)\, i}{n^2} + o\left(n^{-1}\right) = \frac{i}{n} + o\left(n^{-1}\right),$$

and therefore,

$$E\left[D_i\right] = E\left[\sqrt{n}\left(\frac{x_i}{x_n} - \frac{i}{n}\right)\right] = o\left(n^{-\frac{1}{2}}\right).$$

■

**Proof of Theorem 2.** Let $\xi_i = y_i - \lambda$, such that $E\left[\xi_i\right] = 0$ and $E\left[\xi_i^2\right] = \lambda^2$. Let $X_n\left(r\right) = \frac{1}{\lambda\sqrt{n}}S_{[nr]} + \left(nr - [nr]\right)\frac{1}{\lambda\sqrt{n}}\xi_{[nr]+1}$, where $S_n = \sum_{i=1}^{n}\xi_i$. By Donsker's Theorem, $X_n \xrightarrow{d} M$, so $\left\{X_n\left(r\right) - rX_n\left(1\right)\right\} \xrightarrow{D} M_0$, see Billingsley (1968, Th. 10.1 and Th. 5.1). Let $nr = i$, $i = 1, \ldots, n$. Then,

$$\begin{aligned} X_n\left(r\right) - rX_n\left(1\right) &= \frac{1}{\lambda\sqrt{n}}S_{[nr]} + \left(nr - [nr]\right)\frac{1}{\lambda\sqrt{n}}\xi_{[nr]+1} - r\frac{1}{\lambda\sqrt{n}}S_{[n]} = \\ &= \frac{1}{\lambda\sqrt{n}}\left(S_i - \frac{i}{n}S_n\right) = \frac{i}{\lambda\sqrt{n}}\left(\overline{y}_i - \overline{y}\right) \end{aligned}$$

Finally, as $\frac{1}{n}\sum_{j=1}^{n} y_j \to \lambda$,

$$\frac{i}{\lambda\sqrt{n}}\left(\overline{y}_i - \overline{y}\right) = \frac{i}{\sqrt{n}}\frac{\overline{y}}{\lambda}\left(\frac{\overline{y}_i}{\overline{y}} - 1\right) \xrightarrow{d} M^0,$$

that proves the stated result. ■

**Proof of Theorem 3.** The proof of this Theorem is similar to the proofs of Proposition 2, Proposition 3 and Theorem 1 in Bai (1994) after some preliminar results, so we only show here these preliminar results to save space and refer to Bai (1994) for a more detailed proof. First, the statistic (2) can be written as follows:

$$D_i = \frac{i}{\sqrt{n}\overline{y}}\left(\overline{y}_i - \overline{y}\right) = \frac{\sqrt{n}}{\overline{y}}\frac{i}{n}\left(1 - \frac{i}{n}\right)\left(\overline{y}_i - \overline{y}_i^*\right),$$

where $\overline{y}_i^*$ is the sample mean of $\{y_i, \ldots, y_n\}$. Defining $V_i = b_i\left(\overline{y}_i - \overline{y}_i^*\right)$, where $b_i = \frac{i}{n}\left(1 - \frac{i}{n}\right)$, we get,

$$i_{\max} = \arg\max_i \left|D_i\right| = \arg\max_i \left|V_i\right|.$$

The expression $V_i$ is similar to the one used in Bai (1994) except for the term $b_i$. Second, define, $z_i = y_i - \lambda_i$, $i = 1, \ldots, n$. Due to the Kolmogorov-Hájek-Rényi inequality (see, Whittle (1976), for instance),

$$\Pr\left(\max_{m\leq i\leq n} c_i\left|\sum_{j=m}^{i} z_j\right|>\alpha\right)\leq\frac{1}{\alpha^2}\sum_{j=m}^{n}\left(c_j^2E\left[z_j^2\right]\right)\leq\frac{\lambda_{\max}^2}{\alpha^2}\sum_{j=m}^{n}c_j^2 \tag{7}$$

where $\{c_j : j = m,\dots,n\}$ is a sequence of decreasing positive constants and $\lambda_{\max}^2 = \max\left\{\lambda_1^2,\lambda_2^2\right\}$. Due to (7), by taking $c_j = 1/j$, and because $\sum_{j=m}^{\infty} j^{-2} = O\left(m^{-1}\right)$, we have,

$$\Pr\left(\max_{m\leq i}\frac{1}{i}\left|\sum_{j=m}^{i} z_j\right|>\alpha\right)\leq\frac{C_1}{\alpha^2 m}, \tag{8}$$

for some $C_1 > 0$. By taking $c_j = 1/\sqrt{j}$, and because $\sum_{j=1}^{n} j = O(\log n)$, we have,

$$\Pr\left(\max_{1\leq i\leq n}\frac{1}{\sqrt{i}}\left|\sum_{j=1}^{i} z_j\right|>\alpha\right)\leq\frac{\lambda_{\max}^2}{\alpha^2}\sum_{j=1}^{n} j\leq\frac{C_2\log n}{\alpha^2}, \tag{9}$$

for some $C_2 > 0$. Now, using the expressions (7), (8) and (9), the rest of the proof is similar to the proofs of Proposition 2, Proposition 3 and Theorem 1 in Bai (1994). ■

**Proof of Theorem 4.** Consider the case of $\widehat{\lambda}_1 = \overline{y}_i$. Let:

$$\begin{aligned} n^{\frac{1}{2}}\left(\frac{1}{i_{\max}}\sum_{j=1}^{i_{\max}} y_j-\frac{1}{i_h}\sum_{j=1}^{i_h} y_j\right) &= I\left(i_{\max}\leq i_h\right)\left(n^{\frac{1}{2}}\frac{i_h-i_{\max}}{i_{\max}i_h}\sum_{j=1}^{i_{\max}} z_j-n^{\frac{1}{2}}\frac{1}{i_{\max}}\sum_{j=i_{\max}+1}^{i_h} z_j\right)+ \\ &+I\left(i_{\max}>i_h\right)\left(n^{\frac{1}{2}}\frac{i_h-i_{\max}}{i_{\max}i_h}\sum_{j=1}^{i_{\max}} z_j+n^{\frac{1}{2}}\frac{1}{i_{\max}}\sum_{j=i_{\max}+1}^{i_h} z_j+n^{\frac{1}{2}}\frac{i_h-i_{\max}}{i_{\max}i_h}\left(\lambda_2-\lambda_1\right)\right), \end{aligned} \tag{10}$$

and as $i_{\max} = i_h + O_p\left(\left(\lambda_2-\lambda_1\right)^{-2}\right)$ and $n\left(\lambda_2-\lambda_1\right)^2\to\infty$, then (10) is $\left(n^{\frac{1}{2}}\left(\lambda_2-\lambda_1\right)\right)^{-1}O_p(1)$, which converges to 0 in probability. This implies that $\overline{y}_{i_{\max}}$ and $\overline{y}_{i_h}$ have the same asymptotic distributions. As $y_j$, $j = 1,\dots,i_h$ are exponential distributed with mean $\lambda_1$, $\overline{y}_{i_h}$ has a Gamma distribution with mean $i_h$ and variance $\lambda_1^2/i_h$, and thus, the distribution of $n^{\frac{1}{2}}\widehat{\lambda}_1/\lambda_1$ is the stated in the Theorem. The proof in the case of $\widehat{\lambda}_2$ is similar. ■

## References

Akman, V. E. and Raftery, A. E. (1986) Asymptotic inference for a Change-Point Poisson Process.

Annals of Statistics, 14, 1583-1590.

Bai, J. (1994) Least squares estimation of a shift in linear processes. Journal of Time Series Analysis, 15, 453-472.

Bai, J. (1997) Estimation of a change point in multiple regression models. The Review of Economics and Statistics, 79, 551-563.

Billingsley, P. (1968) Convergence of probability measures. John Wiley and Sons, New York.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. M. (1992) Hierarchical Bayesian analysis of changepoint problems. Applied Statistics, 41, 389-405.

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

Chen, J. and Gupta, A. K. (1997) Testing and locating variance changepoints with application to stock prices. Journal of the American Statistical Association, 92, 739-747.

Chib, S. (1998) Estimation and comparison of multiple change-point models. Journal of Econometrics, 86, 221-241.

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

Galeano, P., Peña, D. and Tsay, R. S. (2004) Outlier detection in multivariate time series via Projection Pursuit. Working Paper, 04-42, Universidad Carlos III de Madrid.

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

Inclán, C. and Tiao, G. C. (1994) Use of cumulative sums of squares for retrospective detection of changes of variance. Journal of the American Statistical Association, 89, 913-923.

Jarrett, R. G. (1979) A note on the intervals between Coal-Minning disasters. Biometrika, 66, 191-193.

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

Raftery, A. E. (1994) Change point and change curve modeling in stochastic processes and spatial statistics. Journal of Applied Statistical Science, 1, 403-424.

Raftery, A. E. and 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. International Statistical Review, 56, 31-48.

Vostrikova, L. J. (1981) Detecting "disorder" in multidimensional random processes. Soviet Mathematics Doklady, 24, 55-59.

Whittle, P. (1976) Probability. Wiley-Interscience, New York.

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. (2004) Bayesian binary segmentation procedure for detecting streakiness in sports. Journal of the Royal Statistical Society, Series A, 167, 627-637.

Yang, T. Y. and Kuo, L. (2001) Bayesian binary segmentation procedure for a Poisson Process with multiple changepoints. Journal of Computational and Graphical Statistics, 10, 772-785.