Multiple change-point detection for Poisson processes

C. Dion-Blanc, E. Lebarbier, Stephane S. Robin

Multiple change-point detection for Poisson processes

C. Dion-Blanc

Sorbonne Université, CNRS - LPSM, 75013 Paris, France.

E. Lebarbier

Université Paris Nanterre, CNRS, Modal'X, UPL, 92000 Nanterre, France

E-mail: emilie.lebarbier@parisnanterre.fr

S. Robin

Sorbonne Université, CNRS - LPSM, 75013 Paris, France.

Abstract. Change-point detection aims at discovering behavior changes lying behind time sequences data. In this paper, we investigate the case where the data come from an inhomogenous Poisson process or a marked Poisson process. We present an offline multiple change-point detection methodology based on minimum contrast estimator. In particular we explain how to deal with the continuous nature of the process together with the discrete available observations. Besides, we select the appropriate number of regimes through a cross-validation procedure which is really convenient here due to the nature of the Poisson process. Through experiments on simulated and realworld datasets, we show the interest of the proposed method. The proposed method has been implemented in the `CptPointProcess` R package.

Keywords: Point process; Change-point detection; Exact optimization; Dynamic programming; Cross-Validation

1. Introduction

Multiple change-point detection is one of the most typical task in statistics, data analysis and signal processing. The problem can be stated in the following way: considering a process observed along time, identify the times, called change-points, before and after which the intensity of the process is different. We consider here the detection of change-points in the distribution of a point process. More specifically, we consider an heterogeneous Poisson process, assuming that its intensity function is piecewise constant and we aim at finding the times at which the intensity value changes.

1.1. State of the art

Change-point detection. Change-point detection has been widely studied in the literature (see e.g. [21] or [27] for surveys). Existing methods can be organized according to few main alternatives: (a) single vs multiple change-point detection, (b) on-line vs off-line setting, (c) frequentist vs Bayesian statistical inference, (d) discrete time vs continuous time; all combinations have been considered in one way or another. The single change-point detection focuses on the existence of a unique change, whereas multiple

detection considers a series of changes. The on-line setting aims at detecting changes while the process is being observed, whereas the off-line setting deals with the detection problem once the process has been observed over a given period of time. We consider here the multiple change-point detection problem in the off-line setting, for which we adopt a frequentist approach. One important specificity of our work is that we deal with a continuous point process.

Poisson process. We consider here Poisson processes (see [3, 11] for a general introduction), that have been used as a reference model in a huge range of domains, including vulcanology [7], epidemiology [26, 28] or even cyber-attack modeling [8, 12]. The reader may refer to [20] for on-line detection of changes in a Poisson process. For the same process, Bayesian approaches have been proposed by [23] or [14] for the detection of one single change-point and by [30] or [24] for the detection of multiple change-points. As said before, we cast our work in the multiple detection, off-line, frequentist framework, likewise [13], who uses multiple testing techniques, whereas we adopt a segmentation viewpoint.

Statistical issues. From a statistical perspective, change-point detection raises three main problems: (*i*) the location of the change-points, (*ii*) the estimation of the parameters ruling the process between change-points and (*iii*) the estimation of the number of change-points. Problem (*ii*) is usually not difficult to deal with using, e.g. maximum likelihood, once the change-points have been identified (see e.g. [30]). The determination of the number of change-points (*iii*) is a typical model selection problem, which is discussed later. The determination of the position of the changes (*i*) raises a more complex problem mostly because is resort the minimization of contrast (say, the negative log-likelihood), which is not a continuous function of the change-points.

Minimizing a non-continuous contrast. The off-line detection of multiple change-points in discrete times is known to raise a specific optimization problem for the determination of the optimal change-point locations (that is: the set of locations that minimizes a given contrast), because classical contrasts, such as least squares or negative log-likelihood, are not continuous with respect to (wrt) the change-points because the change-point locations are themselves discrete (see e.g. [6, 19]). Hence, the so-called segmentation space (that is the – huge, although finite– set of possible change-point locations) needs to be explored in some way. Hopefully, for a given number of segments and provided that the contrast is additive wrt the segments, this can be achieved using dynamic programming [5], with quadratic complexity.

The optimization problem could seem simpler for point processes with change-point in continuous times, but it turns out to be even more complex because as we will show, most classical contrasts are still not continuous functions of the change-points and the segmentation space is now infinite. The latter problem can obviously be circumvented by discretizing the time (see [1] for medicine, [2] for pollution surveillance, and [25] for genomics), at the price of computational efficiency. The exploration of the continuous segmentation space has been considered by [29] or [30] to detect a single change-point in a Poisson process. Importantly, [30] observe that the negative log-likelihood is actually

concave between two successive event times, allowing the problem to be solved in a simple and exact manner. This observation is actually essential and paves the way to the strategy we propose here for the detection of multiple change-points.

1.2. Our contribution

In this paper, we propose an efficient method for the segmentation of an heterogeneous Poisson process. One important feature of our work is to show that, under reasonable assumptions for the contrast, the optimal segmentation belongs to a known and finite grid, so that it can be recovered using dynamic programming. One important assumption is that the contrast is additive wrt the segment, which is the consequence of the general properties of Poisson processes. We also discuss which contrasts are admissible, in the sense that they do not yield undesirable solutions.

Regarding the selection of the number of segments, we adopt a cross-validation procedure inspired from [4]. An important feature is that the validity of this procedure results from the thinning property of Poisson processes. Interestingly, the whole methodology we propose can be extended to the segmentation of marked Poisson processes. The proposed method has been implemented in the `CptPointProcess` R package, which is available on `github.com/Elebarbier/CptPointProcess`.

Outline. We introduce the model right after the present paragraph. Then, the optimization procedure for a given number of segments is described in Section 2 and admissible contrasts are discussed in Section 3. The extension of the methodology to marked Poisson processes is explained in Section 4. Section 5 presents the proposed procedure for the selection of the number of segments. Numerical experiments on synthetic data are presented in Section 6 and the use of the whole methodology is illustrated on earth sciences datasets in Section 7.

1.3. Model

Consider a general Poisson process $\{N_t\}_{0\leq t\leq 1}$ with intensity $t \mapsto \lambda(t)$ and we denote $N_t = N((0,t])$ the number of events that occur on the interval $[0,t]$. Let us denote $T_1, \ldots, T_{N_1}$ the event times observed on the observation time interval $[0,1]$, such that $N_{T_i} - N_{T_i^-} = 1$ with the convention $T_0 = 0$ and $T_{N_1+1} = 1$. We assume that $T_j - T_j^- = 0$ for all j and denote for a sake of simplicity $N_1 = n$. In the following, we work conditionally on this event.

We assume that the intensity of the process is piecewise constant. Let us denote m a partition of $[0,1]$ composed of K segments denoted $I_k := (\tau_{k-1}, \tau_k]$ for $k = 1, \ldots, K$ where $0 = \tau_0 < \tau_1 < \ldots < \tau_K = 1$ are called the change-points, and $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_{K-1})$ the vector of the $K-1$ unknown change-points. Thus, the partition m is defined as $m := \{I_k\}_{1\leq k\leq K}$ and the associated intensity of the process is

$$\lambda(t) := \sum_{k=1}^{K} \lambda_k \mathbb{1}_{I_k}(t) \tag{1}$$

where λ_k is the intensity within segment I_k. The vector of intensities is denotes $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_K)$. This model corresponds to a piecewise homogeneous Poisson process: we name it the Poisson change-point model.

We denote $\Delta\tau_k := \tau_k - \tau_{k-1}$ the length of interval I_k and $\Delta N_k := N((\tau_{k-1}, \tau_k]) = N(\tau_k) - N(\tau_{k-1})$ the number of events that occur in interval I_k. The likelihood of a given path $\{N_t\}_{0 \leq t \leq 1}$ (which we denote by N) for the Poisson change-point model (identified with a subscript P) is then

$$p_P(N; \boldsymbol{\tau}, \boldsymbol{\lambda}) = \prod_{k=1}^{K} e^{-\lambda_k \Delta\tau_k} \lambda_k^{\Delta N_k}. \tag{2}$$

It can be easily seen that, for a given $\boldsymbol{\tau}$, $p_P(N; \boldsymbol{\tau}, \boldsymbol{\lambda})$ is maximal for $\widehat{\lambda}_k(\boldsymbol{\tau}) = \Delta N_k / \Delta\tau_k$, $1 \leq k \leq K$. As a consequence, the purpose is to find the set of maximum-likelihood change-point locations amounts at minimizing the so-called Poisson-contrast

$$\gamma_P(\boldsymbol{\tau}; N) := -\log p_P(N; \boldsymbol{\tau}, \widehat{\boldsymbol{\lambda}}(\boldsymbol{\tau})) = \sum_{k=1}^{K} \Delta N_k \left(1 - \log\left(\frac{\Delta N_k}{\Delta\tau_k}\right)\right). \tag{3}$$

Observe that, because of the independence between event times of disjoint segments, γ_P is additive wrt the segments: this property plays a critical role in the sequel.

2. Optimal change-points

We now consider the determination of the optimal set of change-point locations $\widehat{\boldsymbol{\tau}}$, defined as the minimizer of a given data-dependent contrast $\gamma(\boldsymbol{\tau}; N)$, such as the Poisson contrast γ_P defined in (3). Important feature of the resulting optimization problem over $(0,1)^{K-1}$ is that the function γ may not be convex, nor even continuous wrt the τ_k's (the Poisson contrast γ_P is neither one nor the other). As a consequence, classical optimization strategies such as gradient descent do not apply.

Still, in the single change-point detection context, [30] make the critical observation that the Poisson contrast γ_P is actually concave on each inter-event interval $[T_i, T_{i+1}[$, so that the optimal change-point is necessarily localized at an event time T_i or just before, in T_i^-. Extending [30], we show that, in the context of multiple change-point detection with a given number of segments K, the additivity and concavity assumptions on the contrast function are sufficient to ensure that the optimal change-points $\widehat{\tau}_k$ $(1 \leq k \leq K-1)$ are each located on an event time, or just before. This observation brings back the original continuous-space optimization problem to a discrete-space optimization problem, for which efficient and exact algorithms exist.

2.1. Segmentation space and partitioning

For a fixed number of segments K, we define the segmentation space as the set of all possible partitions of $(0,1)$ into K segments:

$$\mathcal{M}^K := \left\{\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{K-1}) \in (0,1)^{K-1}; 0 = \tau_0 < \tau_1 < \ldots < \tau_K = 1\right\}.$$

An element of $\mathcal{M}^K$ (i.e. a 'segmentation') is therefore a sequence of change-points. Moreover, we define the set of all possible K-uplets of, possibly null, integers, summing to n:

$$\Upsilon^{K,n} := \left\{ \boldsymbol{\nu} := (\nu_1, \ldots, \nu_K) \in \{0, 1, \ldots, n\}^K, \ \sum_{k=1}^{K} \nu_k = n \right\}. \tag{4}$$

It is a finite set with cardinal $\left|\Upsilon^{K,n}\right| = \binom{n+K-1}{K-1}$. Then, for a fixed $\boldsymbol{\nu} \in \Upsilon^{K,n}$ and a given path $N = \{N_t\}_{0 \le t \le 1}$ with n events (that is $N_1 = n$), we define the subset of segmentations from $\mathcal{M}^K$ with a prescribed number of events in each segment given by $\boldsymbol{\nu}$, as

$$\mathcal{M}^K_{\boldsymbol{\nu}}(N) := \left\{ \boldsymbol{\tau} \in \mathcal{M}^K, \forall\ 1 \le k \le K : \ \Delta N_k = \nu_k \right\}. \tag{5}$$

It is important to notice that this constraint is equivalent to impose that, for each $1 \le k \le K-1$,

$$\tau_k \in \left[T_{\sum_{j=1}^{k} \nu_j}, T_{\sum_{j=1}^{k} \nu_j + 1} \right)$$

setting $T_0 = 0$ and $T_{n+1} = 1$. Obviously, the count vectors $\boldsymbol{\nu} \in \Upsilon^{K,n}$ induce a partition of the segmentation space $\mathcal{M}^K$:

$$\mathcal{M}^K = \bigcup_{\boldsymbol{\nu} \in \Upsilon^{K,n}} \mathcal{M}^K_{\boldsymbol{\nu}}(N), \qquad \{\boldsymbol{\nu} \neq \boldsymbol{\nu}'\} \Rightarrow \{\mathcal{M}^K_{\boldsymbol{\nu}}(N) \cap \mathcal{M}^K_{\boldsymbol{\nu}'}(N) = \varnothing\}.$$

Figure 1 displays the segmentation space $\mathcal{M}^K$ for $K = 3$ and its partition into all subsets $\mathcal{M}^K_{\boldsymbol{\nu}}(N)$ for $\boldsymbol{\nu} \in \Upsilon^{K,n}$ with $n = 4$ events. The gray region labeled by $\boldsymbol{\nu} = (2, 1, 1)$ is $\mathcal{M}^K_{\boldsymbol{\nu}}(N)$, that is the set of all segmentations $\boldsymbol{\tau} \in (0, 1)^2$ such that $\Delta N_1 = 2, \Delta N_2 = 1, \Delta N_3 = 1$ or, equivalently $T_2 \le \tau_1 < T_3$ and $T_3 \le \tau_2 < T_4$.

For obvious reasons, we do not consider the configurations $\boldsymbol{\nu}$ that can contain more than two successive zeros, i.e. one of the interval $[T_{\sum_{j=1}^{k} \nu_j}, T_{\sum_{j=1}^{k} \nu_j + 1})$ contains more than two changes. This may occur only when $K > 2$. We therefore restrict the segmentation space to

$$\mathcal{M}^K_\star := \bigcup_{\boldsymbol{\nu} \in \Upsilon^{K,n}_\star} \mathcal{M}^K_{\boldsymbol{\nu}}(N), \tag{6}$$

where

$$\Upsilon^{K,n}_\star = \left\{ \boldsymbol{\nu} \in \Upsilon^{K,n}, \text{ for } 2 \le k \le K-1\text{: if } \nu_k = 0 \text{ then } \nu_{k-1} \neq 0 \text{ and } \nu_{k+1} \neq 0 \right\} \tag{7}$$

has cardinal

$$\left|\Upsilon^{K,n}_\star\right| = \sum_{h=1}^{K} \binom{n-1}{h-1} \binom{h+1}{K-h}$$

(with the convention $\binom{p}{q} = 0$ if $q > p$). In Figure 1, the forbidden segmentation subsets (i.e. the elements of $\Upsilon^{K,n} \setminus \Upsilon^{K,n}_\star$) are the white upper triangular regions labeled with $\boldsymbol{\nu} = (0, 0, 4)$ and $\boldsymbol{\nu} = (4, 0, 0)$, respectively.

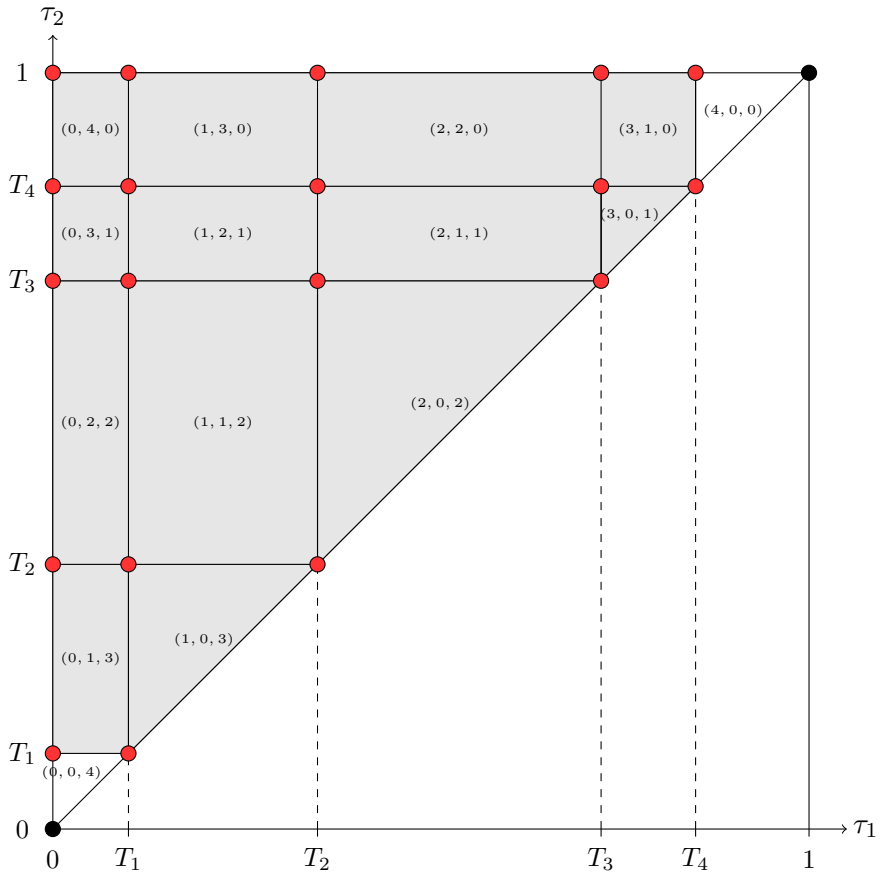

Figure 1. Segmentation space $\mathcal{M}_\star^K$ for $K = 3$ and $N_1 = n = 4$ in gray. Each gray block in the upper triangle corresponds to an element of $\Upsilon_\star^{K,n}$. Each white block in the upper triangle corresponds to an element of $\Upsilon^{K,n} \setminus \Upsilon_\star^{K,n}$.

2.2. Minimum contrast estimation

We now search for the optimal segmentation (i.e. sequence of change-points) $\widehat{\boldsymbol{\tau}}$ with K segments that minimizes a given contrast γ within $\mathcal{M}_\star^K$:

$$\widehat{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau}\in\mathcal{M}_\star^K}{\operatorname{argmin}}\ \gamma(\boldsymbol{\tau}). \tag{8}$$

The number of segments K is constant in the rest of the section so the superscript K may be dropped for the sake of clarity.

The methodology we propose relies on two main assumptions regarding the contrast function γ.

ASSUMPTION 2.1 (SEGMENT-ADDITIVITY ASSUMPTION). *For each $\boldsymbol{\nu} \in \Upsilon_\star^n$ and for all $\boldsymbol{\tau} \in \mathcal{M}_{\boldsymbol{\nu}}(N)$, the contrast γ writes as sum of segment-specific cost functions:*

$$\gamma(\boldsymbol{\tau}) = \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k).$$

As an example, for the Poisson contrast γ_P defined in (3), the cost function is $C(\nu_k, \Delta\tau_k) = \nu_k\,(1 - \log(\nu_k/\Delta\tau_k))$.

ASSUMPTION 2.2 (CONCAVITY ASSUMPTION). *For each $\boldsymbol{\nu} \in \Upsilon_\star^n$ and each $1 \le k \le K$, the cost function $C(\nu_k, \Delta\tau_k)$ is a concave function of $\Delta\tau_k$.*

Under Assumption 2.1, partitioning the segmentation space, the optimization problem (8) can be rewritten as

$$\widehat{\boldsymbol{\tau}} \;=\; \underset{\boldsymbol{\nu}\in\Upsilon_\star^{K,n}}{\operatorname{argmin}}\ \min_{\boldsymbol{\tau}\in\mathcal{M}_{\boldsymbol{\nu}}^K(N)} \gamma(\boldsymbol{\tau}) = \underset{\boldsymbol{\nu}\in\Upsilon_\star^{K,n}}{\operatorname{argmin}}\ \min_{\boldsymbol{\tau}\in\mathcal{M}_{\boldsymbol{\nu}}^K(N)} \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k). \tag{9}$$

We now show that the concavity Assumption 2.2 for each cost function C implies the piecewise concavity of the contrast γ wrt $\boldsymbol{\tau}$.

PROPOSITION 2.3. *Under Assumption 2.2, for each $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$, the contrast function $\boldsymbol{\tau} \to \gamma(\boldsymbol{\tau})$ is concave wrt the segmentation $\boldsymbol{\tau}$ within $\mathcal{M}_{\boldsymbol{\nu}}^K(N)$.*

PROOF. Let fix $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ and take $\boldsymbol{\tau} \in \mathcal{M}_\nu^K(N)$. Denoting by C' and C'' the first and second derivative of C with respect to its second argument, we have, for all $1 \le k < K$,

$$\frac{\partial\gamma(\boldsymbol{\tau})}{\partial\tau_k} = C'(\nu_k, \Delta\tau_k) - C'(\nu_{k+1}, \Delta\tau_{k+1}), \qquad \frac{\partial^2\gamma(\boldsymbol{\tau})}{\partial\tau_k\partial\tau_{k+1}} = -C''(\nu_{k+1}, \Delta\tau_{k+1}) \quad (\text{if } k < K-1)$$

$$\frac{\partial^2\gamma(\boldsymbol{\tau})}{\partial\tau_k^2} = C''(\nu_k, \Delta\tau_k) + C''(\nu_{k+1}, \Delta\tau_{k+1}).$$

The Hessian matrix of the contrast γ is the following $(K-1)\times(K-1)$ matrix

$$H_K = \begin{pmatrix} -(A_1+A_2) & A_2 & 0 & \dots \\ A_2 & -(A_2+A_3) & A_3 & \dots \\ 0 & \ddots & \ddots & \ddots \\ \vdots & 0 & A_{K-1} & -(A_{K-1}+A_K) \end{pmatrix}$$

where $A_k := -C''(\nu_k, \Delta\tau_k)$. For all vector $u \in \mathbb{R}^{K-1} \setminus \{0\}$, we have that

$$\begin{aligned} {}^t u H_K u &= -\sum_{k=1}^{K-1} u_k^2 (A_k + A_{k+1}) + 2 \sum_{k=1}^{K-2} u_k A_{k+1} u_{k+1} \\ &= -\sum_{k=1}^{K-2} (u_k - u_{k+1})^2 A_{k+1} - u_1^2 A_1 - u_{K-1}^2 A_K. \end{aligned}$$

Using the concavity Assumption 2.2 of the contrast function on each segment we have that $A_k \geq 0$ for all k, thus ${}^t u H_K u \leq 0$, which concludes the proof.

2.3. Exact optimization

We now show that the optimal segmentation necessarily belongs to a finite subset of $\mathcal{M}_\star^K$. More specifically, it necessarily belongs to a finite and known grid, so the optimal solution can be thus obtained in an exact and fast manner.

THEOREM 2.4. *Under Assumption 2.1 and 2.2, for a fixed $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ we get*

$$\widehat{\boldsymbol{\tau}} = \underset{\boldsymbol{\tau} \in \mathcal{M}_{\boldsymbol{\nu}}^K(N)}{\operatorname{argmin}} \gamma(\boldsymbol{\tau}) \in \{T_{\nu_1}, T^-_{\nu_1+1}\} \times \{T_{\nu_1+\nu_2}, T^-_{\nu_1+\nu_2+1}\} \times \ldots \times \{T_{\nu_1+\ldots\nu_K}, T^-_{\nu_1+\ldots\nu_K+1}\}.$$

PROOF. The proof results directly from the concavity of the contrast function $\gamma(\boldsymbol{\tau})$ wrt $\boldsymbol{\tau} \in \mathcal{M}_{\boldsymbol{\nu}}^K(N)$ (see Proposition 2.3).

Theorem (2.4) ensures that the $K-1$ optimal change-points in each subset of $\mathcal{M}_{\boldsymbol{\nu}}^K(N)$ is necessarily one of its boundary partitions reducing thus the search in a finite set possible solutions. These solutions are illustrated in Figure 1 by the red circle points for the simple case of $K = 3$ segments (or 2 change-points). As an example, for $\boldsymbol{\nu} = (2, 1, 1)$, $T_2 \leq \tau_1 < T_3$ and $T_3 \leq \tau_2 < T_4$ leading to four possible choices for the optimal solution: $(\widehat{\tau}_1, \widehat{\tau}_2) = (T_2, T_3), (T_2, T_4^-), (T_3^-, T_3)$ or (T_3^-, T_4^-).

Consequently, the optimal change-points in $\mathcal{M}_\star^K$ are necessary localized at an event time T_i or just before, in T_i^-. Moreover, for obvious reasons, for $K > 2$, the configuration $(\nu_k, \Delta\tau_k) = (0, 0)$ are not considered. Indeed, this configuration corresponds to cases where τ_{k-1} and τ_k belong to the same inter-event interval $\left[T_{\sum_{j=1}^{k-1}\nu_j}, T_{\sum_{j=1}^{k-1}\nu_j+1}\right)$ and $\tau_{k-1} = \tau_k$. The global optimization problem is thus reduced to a discrete optimization problem on the finite grid

$$\mathbf{tp} = \{T_1^-, T_1, , T_2^-, T_2, \ldots, T_n^-, T_n\},$$

with cardinal $2n$. This problem is well known in discrete change-point detection setting and can be solved using the usual the classical dynamic programming (DP) algorithm (presented in Section A.1) with a complexity $\mathcal{O}((2n+1)^2 K)$. This algorithm can be used here thanks to the segment-additivity Assumption 2.1 of the contrast function γ.

3. Admissible contrasts

We present some admissible contrasts in the sense that they satisfy the two Assumptions 2.1 and 2.2 and discuss the relevance of their optimal solutions.

Let first consider the classical Poisson contrasts, the negative log-likelihood and the least-squares, for a given path N of the Poisson change-point model. The contrast based on the likelihood criterion has been yet given in the Introduction, Equation (3), and used as example in the previous section. Let us recall it here for a fixed $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ and for all $\boldsymbol{\tau} \in \mathcal{P}_{\boldsymbol{\nu}}^K(N)$, the contrast writes

$$\gamma_P(\boldsymbol{\tau}) \quad = \quad -\log p_P(N; \boldsymbol{\tau}, \widehat{\boldsymbol{\lambda}}) = \sum_{k=1}^{K} \nu_k \left(1 - \log\left(\frac{\nu_k}{\Delta\tau_k}\right)\right) = \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k). \quad (10)$$

Similarly, the least squares criterion of the given path N is

$$LS_P(N; \boldsymbol{\tau}, \boldsymbol{\lambda}) = \sum_{k=1}^{K} \left(\Delta\tau_k \ \lambda_k^2 - \Delta N_k \lambda_k\right), \quad (11)$$

that is minimum for the same intensities $\widehat{\lambda}_k(\boldsymbol{\tau}) = \frac{\Delta N_k}{\Delta\tau_k}$, $1 \leq k \leq K$. The associated contrast function of $\boldsymbol{\tau}$ is thus

$$\tilde{\gamma}_P(\boldsymbol{\tau}) \quad = \quad LS_P(N; \boldsymbol{\tau}, \widehat{\boldsymbol{\lambda}}) = -\sum_{k=1}^{K} \frac{\nu_k}{\Delta\tau_k} = \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k). \quad (12)$$

For both contrasts, the independence between disjoint time-intervals property of the Poisson process ensures that the segment-additivity Assumption 2.1 is true. Then, it is straightforward to see that the cost function C is concave wrt $\Delta\tau_k$, and that Assumption 2.2 is validated.

However, these two contrasts share the same undesirable property: for $K > 2$, the optimal segmentation inevitably contains segments with null length. The reason is easy to see: for both, the cost function $C(1, 0) = -\infty$. For obvious reasons, such solution are not desirable.

We thus propose another admissible based-likelihood contrast which does not have the same drawback. To do so, we adopt a Bayesian approach and assume that the intensities λ_k's are independent random variables and follow a gamma distribution with parameters $a > 0$ and $b > 0$, denoted $\mathcal{G}\text{amma}(a, b)$. The distribution of $\boldsymbol{\lambda}$ is thus the following

$$p_G(\boldsymbol{\lambda}; a, b) = \prod_{k=1}^{K} \frac{b^a}{\Gamma(a)} \ \lambda_k^{(a-1)} \ e^{-b\lambda_k}$$

where Γ stands for the gamma function: $\Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} \mathrm{d}t$. For a given $\boldsymbol{\tau}$, the joint distribution of $(N, \boldsymbol{\lambda})$ is thus

$$p_{PG}(N, \boldsymbol{\lambda}; \boldsymbol{\tau}, a, b) = p_P(N; \boldsymbol{\lambda}, \tau) \ p_G(\boldsymbol{\lambda}; a, b) = \prod_{k=1}^{K} \frac{b^a}{\Gamma(a)} \lambda_k^{(\Delta N_k + a - 1)} \ e^{-(\Delta\tau_k + b)\lambda_k},$$

and the marginal distribution of N is

$$p_{PG}(N;\boldsymbol{\tau},a,b) = \int p_{PG}(N,\boldsymbol{\lambda};\boldsymbol{\tau},a,b)\mathrm{d}\boldsymbol{\lambda} = \prod_{k=1}^{K} \frac{b^a\ \Gamma(\Delta N_k + a)}{\Gamma(a)\ (\Delta\tau_k + b)^{\Delta N_k + a}}.$$

For a fixed $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ and all $\boldsymbol{\tau} \in \mathcal{P}_{\boldsymbol{\nu}}^K(N)$, the proposed contrast, so-called the Poisson-Gamma contrast, is

$$\begin{aligned}\gamma_{PG}(\boldsymbol{\tau}) &= -\log p_{PG}(N \mid \boldsymbol{\tau}; a, b) \\ &= \sum_{k=1}^{K} \Big(-a\log b + \log\Gamma(a) + \widetilde{a}_k \log \widetilde{b}_k - \log\Gamma(\widetilde{a}_k)\Big) = \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k) \quad (13)\end{aligned}$$

where $\widetilde{a}_k = \nu_k + a$, $\widetilde{b}_k = \Delta\tau_k + b$. This contrast also satisfies Assumptions 2.1 and 2.2 and is this admissible. Moreover, $C(1, \Delta\tau_k)$ is now lower-bounded allowing the preference of an other segmentation compared to the previous undesirable ones.

Let us notice that a and b need to be chosen in practice. Since the average of the gamma distribution with parameters (a, b) is a/b, a simple rule can be to choose a and b such that $a/b = n$.

Moreover, the conditional distribution of $(\boldsymbol{\lambda} \mid N; \boldsymbol{\tau})$ is a gamma distribution with parameters $\widetilde{a}_k$ and $\widetilde{b}_k$, thus its posterior mean is

$$\mathbb{E}(\lambda_k \mid N, \tau) = \widetilde{a}_k / \widetilde{b}_k. \quad (14)$$

This provides us with the estimator $\widehat{\lambda}_k = \widetilde{a}_k/\widetilde{b}_k$ and enables us to build the estimator of the intensity process $(\lambda(t))_{t\in[0,1]}$. We will use this estimator in the final algorithm, proposed in Section 5.

4. Extension to marked Poisson process

We extend the proposed methodology to a marked Poisson process for which both the intensity function of the underlying Poisson process and the parameter of the distribution of the marks are affected by the same changes.

4.1. Model

Let us consider a marked Poisson process. More specifically, we consider a Poisson process N with a piecewise constant intensity function λ given in Equation (1) and suppose that a mark X_i is associated with each event time T_i $(1 \le i \le n = N_1)$.
We assume that the marks $\{X_i\}_{i=1,\dots,n}$ are independent exponential random variables with parameter $\rho(T_i)$: $X_i|T_i \sim \mathcal{E}(\rho(T_i))$. The function ρ is assumed to be piecewise constant with the same change-points as λ, for $t \in [0, 1]$:

$$\rho(t) = \sum_{k=1}^{K} \rho_k \mathbb{1}_{I_k}(t), \quad \boldsymbol{\rho} = (\rho_1, \dots, \rho_K). \quad (15)$$

We take the exponential distribution as an example: this setting work can easily be extended to other distributions for the marks.

4.2. Admissible contrasts

The log-likelihood of a given marked Poisson path (N, X) with piecewise intensity and piecewise marks distribution, is given by

$$\log p_{MP}(N, X; \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{K} \left(\Delta N_k \log(\lambda_k) - \lambda_k \Delta\tau_k + \sum_{i, T_i \in I_k} \log(p(X_i | T_i; \rho_k)) \right).$$

It can be easily seen that, for a given $\boldsymbol{\tau}$, $\log p_{MP}(N, X; \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\rho})$ is maximal for $\widehat{\lambda}_k(\boldsymbol{\tau}) = \Delta N_k / \Delta\tau_k$ and

$$\widehat{\rho}_k(\boldsymbol{\tau}) = \frac{\Delta N_k}{S_k}, \qquad \text{with} \quad S_k = \sum_{i, T_i \in I_k} X_i,$$

for $1 \leq k \leq K$. The resulting contrast function for the estimation of the change-points is thus writes for a fixed $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ and for all $\boldsymbol{\tau} \in \mathcal{P}_{\boldsymbol{\nu}}^K(N)$ as

$$\begin{aligned} \gamma_{MP}(\boldsymbol{\tau}) &= -\log p_{MP}(N, X; \boldsymbol{\tau}, \widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\rho}}) \\ &= \sum_{k=1}^{K} \nu_k \left(-\log\left(\frac{\nu_k}{\Delta\tau_k}\right) - \log\left(\frac{\nu_k}{S_k}\right) + 2 \right) = \sum_{k=1}^{K} C(\nu_k, \Delta\tau_k). \end{aligned} \tag{16}$$

As for the Poisson contrast, this based-likelihood contrast is admissible but suffers of the limitation pointed out in the previous Section.

Following the same idea, we define a new Marked Poisson-Gamma-Exponential-Gamma (MPGEG) based-likelihood contrast. Precisely, we assume that the λ_k's and ρ_k's are independent random variables and follow a Gamma distribution with parameters $a_\lambda > 0$ and $b_\lambda > 0$ and a Gamma distribution with parameters $a_\rho > 0$ and $b_\rho > 0$, respectively. For a given $\boldsymbol{\tau}$, the joint distribution of $(N, X, \boldsymbol{\lambda}, \boldsymbol{\rho})$ is thus

$$\begin{aligned} p_{MPGEG}(N, X, \boldsymbol{\lambda}, \boldsymbol{\rho}; \boldsymbol{\tau}, a_\lambda, b_\lambda, a_\rho, b_\rho) &= p_P(N, X; \boldsymbol{\tau}, \boldsymbol{\lambda}, \boldsymbol{\rho})\ p_G(\boldsymbol{\lambda}; a_\lambda, b_\lambda)\ p_G(\boldsymbol{\rho}; a_\rho, b_\rho) \\ &= \prod_{k=1}^{K} \frac{{b_\lambda}^{a_\lambda}}{\Gamma(a_\lambda)} \lambda_k^{(\Delta N_k + a_\lambda - 1)} e^{-(\Delta\tau_k + b_\lambda)\lambda_k} \\ &\quad \times \prod_{k=1}^{K} \frac{{b_\rho}^{a_\rho}}{\Gamma(a_\rho)} \rho_k^{(\Delta N_k + a_\rho - 1)}\ e^{-(S_k + b_\rho)\rho_k}, \end{aligned}$$

and the marginal distribution of (N, X) is

$$\begin{aligned} p_{MPGEG}(N, X; \boldsymbol{\tau}, a_\lambda, b_\lambda, a_\rho, b_\rho) &= \iint p_{PGEG}(N, \boldsymbol{\lambda}, \boldsymbol{\rho}; \boldsymbol{\tau}, a_\lambda, b_\lambda, a_\rho, b_\rho)\ \mathrm{d}\boldsymbol{\lambda}\ \mathrm{d}\boldsymbol{\rho} \\ &= \prod_{k=1}^{K} \frac{{b_\lambda}^{a_\lambda}\ \Gamma(\Delta N_k + a_\lambda)}{\Gamma(a_\lambda)\ (\Delta\tau_k + b_\lambda)^{\Delta N_k + a_\lambda}}\ \frac{{b_\rho}^{a_\rho}\ \Gamma(\Delta N_k + a_\rho)}{\Gamma(a_\rho)\ (S_k + b_\rho)^{\Delta N_k + a_\rho}}. \end{aligned}$$

For a fixed $\boldsymbol{\nu} \in \Upsilon_\star^{K,n}$ and for any $\boldsymbol{\tau} \in \mathcal{P}_{\boldsymbol{\nu}}^K(N)$, the MPGEG contrast is

$$\begin{aligned}\gamma_{MPGEG}(\boldsymbol{\tau}) &= -\log p_{MPGEG}(N; \boldsymbol{\tau}, a_\lambda, b_\lambda, a_\rho, b_\rho) \\ &= \sum_{k=1}^{K} \left(\widetilde{a}_{k,\lambda} \log \widetilde{b}_{k,\lambda} - \log\Gamma(\widetilde{a}_{k,\lambda})\right) + \left(\widetilde{a}_{k,\rho} \log \widetilde{b}_{k,\rho} - \log\Gamma(\widetilde{a}_{k,\rho})\right) \\ &\quad + K\left(-a_\lambda \log b_\lambda + \log\Gamma(a_\lambda) - a_\rho \log b_\rho + \log\Gamma(a_\rho)\right),\end{aligned} \tag{17}$$

where $\widetilde{a}_{k,\lambda} = \nu_k + a_\lambda$, $\widetilde{b}_{k,\lambda} = \Delta\tau_k + b_\lambda$, $\widetilde{a}_{k,\rho} = \nu_k + a_\rho$, and $\widetilde{b}_{k,\rho} = S_k + b_\rho$. This contrast function is also admissible and can avoid segmentations with zero-length segments.

The posterior means of λ_k and ρ_k are

$$\widehat{\lambda}_k = \widetilde{a}_{k,\lambda}/\widetilde{b}_{k,\lambda}, \qquad\qquad \widehat{\rho}_k = \widetilde{a}_{k,\rho}/\widetilde{b}_{k,\rho}.$$

We use these posterior means as estimators of the intensity of the process and of the density parameter of the marks, respectively.

Similarly to the PG contrast, the parameters $a_\lambda, b_\lambda, a_\rho, b_\rho$ need to be chosen. For the two first parameters, we choose $a_\lambda = 1$ and $a_\lambda/b_\lambda = n$. Then, because the conditional distribution of $X \mid N$ a Pareto distribution with parameter (b_ρ, a_ρ) so $\mathbb{E}[X \mid N] = b_\rho/(a_\rho - 1)$ with the condition $a_\rho > 2$ (to ensure the existence of the variance). Thus, one may choose $a_\rho = 2.01$ and $b_\rho = \overline{X}(a_\rho - 1)$.

5. Choosing the number of segments

We now consider the selection of the number of segments K. To this aim, we propose to use a cross-validation strategy which takes advantage of a specific property of the Poisson process, which we remind to the reader in the following.

PROPERTY 5.1 (THINNING). *Consider an heterogeneous Poisson process N with intensity function $\lambda(t)$: $N = \{N_t\}_{0\leq t\leq 1} \sim PP(\lambda)$. Sampling each event time of N with probability f yields in an heterogeneous Poisson process N^A with intensity function $\lambda^A(t) = f\lambda(t)$. Furthermore, the remaining fraction of event times forms a second heterogeneous Poisson process N^B with intensity function $\lambda^B(t) = (1-f)\lambda(t)$, and the processes N^A and N^B are independent.*

This thinning property has two important consequences. First, if the intensity function λ is piecewise constant, then the intensity functions λ^A and λ^B are also piecewise constant, *with same change-points as* λ. Second, whatever the form of the intensity function of N, the ratio between the intensity functions of N^B and N^A is constant and equal to $\lambda^B(t)/\lambda^A(t) \equiv (1-f)/f$. This suggests the following cross-validation procedure: (i) sample the events of the observed process N with probability f to form a *learning process* N^L and form an independent *test process* N^T with the remaining events; (ii) for a series of values of K, get estimates $(\widehat{\boldsymbol{\tau}}^{K,L}, \widehat{\boldsymbol{\lambda}}^{K,L})$ of the change-points and intensities, respectively, using the learning process N^L, (iii) evaluate the contrast for the test process N^T with parameters $(\widehat{\boldsymbol{\tau}}^{K,L}, \frac{1-f}{f}\widehat{\boldsymbol{\lambda}}^{K,L})$.

Let now present the algorithm.

ALGORITHM 5.2 (CROSS-VALIDATION PROCEDURE).

Input: *a realization of the process* N.

Cross-validation: *for* $m = 1$ *to* M

(a) *sample a learning process* $N^{m,L}$ *from* N *with probability* f*, and form the test process* $N^{m,T}$ *with the remaining events;*

(b) *for* $K = 1$ *to* $K_{\max}$,

- *segment the learning process* $N^{m,L}$ *using the Poisson-Gamma contrast* (13) *to get*

$$\widehat{\boldsymbol{\tau}}^{m,K,L} = \underset{\boldsymbol{\nu}\in\Upsilon_\star^{K,n(m,L)}}{\operatorname{argmin}} \min_{\boldsymbol{\tau}\in\mathcal{M}_{\boldsymbol{\nu}}^K(N^{m,L})} \gamma_{PG}(\boldsymbol{\tau}),$$

  where $n(m, L)$ *is the number of events in the learning process,*

- *deduce the estimate* $\widehat{\boldsymbol{\lambda}}^{m,K,L}$ *as the set of posterior means* (14)

$$\widehat{\lambda}_k^{m,K,L} = \mathbb{E}(\lambda_k | N^{m,L}, \widehat{\boldsymbol{\tau}}^{m,K,L}) = \frac{a^{m,L} + \Delta N_k^{m,L}}{b^{m,L} + \Delta \hat{\tau}_k^{m,K,L}} \quad \text{for all } k = 1, \dots K$$

- *and compute the Poisson contrast for the test process*

$$\gamma^{m,K,T} = -\log p_P\left(N^{m,K,T}; \widehat{\boldsymbol{\tau}}^{m,K,L}, \frac{1-f}{f}\widehat{\boldsymbol{\lambda}}^{m,K,L}\right).$$

Averaging: *for* $K = 1$ *to* $K_{\max}$*, compute the average contrast*

$$\overline{\gamma}^{K,T} = \frac{1}{M}\sum_{m=1}^{M} \gamma^{m,K,T}.$$

Selection: *select* K *as*

$$\widehat{K} = \underset{K}{\operatorname{argmin}}\ \overline{\gamma}^{K,T}.$$

Some comments. Few remarks can be made about this procedure. First, Algorithm 5.2 involves three tuning parameters: the number of cross-validation sample M, the maximum number of segments $K_{\max}$ and the sampling probability f. The first two are only limited by the computational burden and can be chosen as large as wanted. The sampling probability f was set to $4/5$ in all the study.

Second, we use the Poisson-Gamma contrast for the estimation of the parameters $\boldsymbol{\tau}$ and $\boldsymbol{\lambda}$ as it is admissible, as explained in Section 3. Still, because the undesirable properties of the Poisson likelihood have no effect when used to measure the fit of parameters $\widehat{\boldsymbol{\tau}}$ and $\widehat{\boldsymbol{\lambda}}$ to an independent process, we use the regular Poisson contrast to evaluate this fit.

Lastly, but most importantly, cross-validation is generally not applicable when dealing with discrete-time segmentation problems because it then amounts at removing observations times, making the location of estimated change-point unclear with respect to the complete dataset. The situation is different in the continuous time setting we consider, thanks to the thinning Property 5.1.

Practical implementation. In practice, to perform the segmentation of an observed process N, we propose to first determine $\widehat{K}$ using Algorithm 5.2 and, then, to estimate $\boldsymbol{\tau}$ using the Poisson-gamma contrast and $\boldsymbol{\lambda}$ as the posterior mean on the whole process N:

$$\widehat{\boldsymbol{\tau}} = \underset{\boldsymbol{\nu} \in \Upsilon_\star^{\widehat{K},n}}{\operatorname{argmin}} \min_{\boldsymbol{\tau} \in \mathcal{M}_{\boldsymbol{\nu}}^{\widehat{K}}(N)} \gamma_{PG}(\boldsymbol{\tau}),$$
$$\widehat{\lambda}_k = \mathbb{E}(\lambda_k | N, \widehat{\boldsymbol{\tau}}) \quad \text{for all } k = 1, \ldots, \widehat{K}.$$

The adaptation of Algorithm 5.2 to the segmentation of a marked Poisson process is straightforward, replacing the Poisson-Gamma contrast with the Marked Poisson-Gamma-Expo-Gamma contrast and the Poisson contrast with the marked Poisson contrast, respectively.

6. Simulation study

We present a simulation study to assess the performances of the proposed methodology for both the Poisson process and the marked Poisson process.

6.1. Simulation design and quality criteria

Simulation design for the Poisson process. We use a simulation design for the change-points inspired from [9]. We fix the number of segments to $K = 6$, with change-point locations $\boldsymbol{\tau} = [0, 7, 8, 14, 16, 20, 24]/24$ (that is with lengths $\Delta\boldsymbol{\tau} = [7, 1, 6, 2, 4, 4]/24$). The total length of odd segments ($k = 1, 3, 5$) is therefore $\Delta\tau_- = 17/24$ and this of even segments is $\Delta\tau_+ = 7/24$. The intensity is set to λ_- in odd segments and to λ_+ in even segments.

The simulation design for these two intensities are based on the two following parameters:

$\overline{\lambda}$: mean intensity: $\overline{\lambda} = \int_0^1 \lambda(t)\mathrm{d}t = (\lambda_-\Delta\tau_- + \lambda_+\Delta\tau_+)$. It controls the expected number of events. A small value for $\overline{\lambda}$ yields a very scarce signal, that is a very poor available information. We considered the values $\overline{\lambda} = 32$, 56, 100, 178, 316, 562 and 1000.

λ_R : ratio between the even and odd intensities: $\lambda_R := \lambda_+/\lambda_- \geq 1$. It controls the contrasts between successive segments. Observe that $\lambda_R = 1$ actually yields one unique segment with intensity $\overline{\lambda}$. We considered the values $\lambda_R = 1$, 2, 3, 4, 6, 8, 11 and 16.

The values of the intensities λ_- and λ_+ are deduced from the both fixed values of $\overline{\lambda}$ and λ_R: $\lambda_- = \overline{\lambda}/\left(\Delta\tau_- + \lambda_R\Delta\tau_+\right)$ and $\lambda_+ = \lambda_T\lambda_-$. Examples of piecewise intensity function λ are given in Figure 2 for $\overline{\lambda} = 100$ and different values of λ_R.

The maximum number of segments was set to $K_{\max} = 12$. We used $M = 500$ samples for the cross-validation procedure and the sampling probability was set to $f = 4/5$, and $B = 100$ processes N were sampled with each parameter configuration. The hyperparameters a and b for the Poisson-Gamma contrast where set to $a = 1$ and $b = 1/n$, where n is the number of events in the process to be segmented.

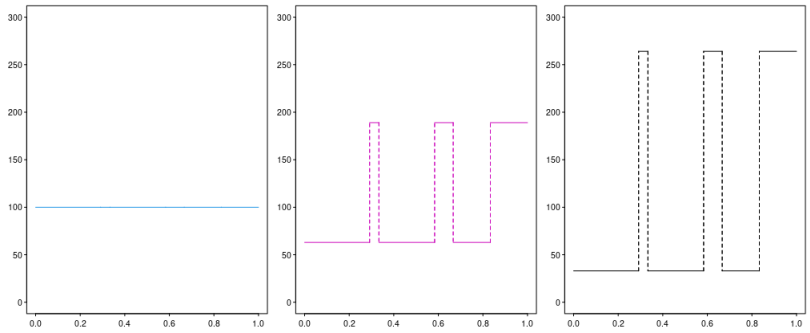

Figure 2. Piecewise intensity function $\lambda(\cdot)$ on $[0,1]$ for $\bar{\lambda} = 100$ and $\lambda_R \in \{1, 3, 8\}$ from left to right.

Simulation design for the marked Poisson process. We aim at studying if the additional change-point information carried by the mark process helps the detection and to know which of the two processes (the Poisson process and the mark one) allows better detection of the change-points. We use the same simulation design as previously for the change-point locations and we fix $\overline{\lambda} = 100$. We consider two values for λ_R: $\lambda_R = 1$ where the ground intensity of the Poisson process does not change from a segment to another, referred to a 'no signal' case, and $\lambda_R = 8$ where the change in the intensity is marked, referred to 'signal case' in λ respectively. For each value of λ_R, the parameter of the mark distribution $\rho(t)$ is either constant and equals to 0.1, referred to a 'no signal' case in ρ or alternates between 0.1 or 0.005 for the odd and the even segments respectively, referred to a 'signal' case in ρ. This leads to four scenarios in terms of detection according to the changes or not in (λ, ρ).

Quality criteria. The performances were assessed according to the following criteria.

- The selected number of segments $\hat{K}$ (obtained with Algorithm 5.2).
- The Hausdorff distance $d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$ between the true change-point location $\boldsymbol{\tau}$ and the estimated $\widehat{\boldsymbol{\tau}}$ (with possibly different number of change-points, as in [9]). More specifically, defining

$$d_1(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}}) = \max_k \min_\ell |\tau_k - \widehat{\tau}_\ell|, \qquad d_2(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}}) = d_1(\widehat{\boldsymbol{\tau}}, \boldsymbol{\tau}),$$
$$d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}}) = \max(d_1(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}}), d_2(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})),$$

  d_1 indicates if each true change-points τ_k is close to an estimated one $\widehat{\tau}_\ell$ (d_1 will typically be small when $\widehat{K} \gg K$), whereas d_2 indicates if each estimated change-point $\widehat{\tau}_\ell$ is close to a true one τ_k. A perfect segmentation results in both null d_1 and d_2 (and $d = 0$).

- The relative L^2-norm between the estimated and the true cumulated intensity functions. More specifically, denoting $\lambda(t)$ and $\widehat{\lambda}(t)$ the true and estimated intensity

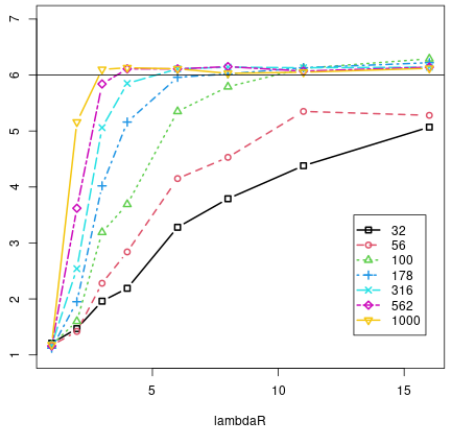

Figure 3. Mean selected number of segments $\widehat{K}$ (averaged over $B = 100$ replicates) as a function of the intensity ratio λ_R. Legend panel = value of the mean intensity $\overline{\lambda}$.

functions and $\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u$ and $\widehat{\Lambda}(t) = \int_0^t \widehat{\lambda}(u)\mathrm{d}u$ the corresponding cumulated intensity functions, we compute:

$$\ell^2(\Lambda, \widehat{\Lambda}) = \left(\int_0^1 (\widehat{\Lambda}(t) - \Lambda(t))^2 \mathrm{d}t \right) \Big/ \overline{\lambda}.$$

6.2. Results

Model selection. Figure 3 shows the mean selected number of segments $\widehat{K}$ as a function of λ_R (a measure of the difficulty of the task). The correct number of segments $K = 6$ is better recovered when either the mean intensity $\overline{\lambda}$ or the ratio λ_R increases, as expected. More precisely for a high mean intensity ($\overline{\lambda} = 1000$), the correct number of segments $K = 6$ is recovered as soon as the ratio λ_R reaches 3. In the typical case where $\overline{\lambda} = 100$, the correct number of segments is obtained when λ_R is about 10. When $\lambda_R = 1$, for all values of $\overline{\lambda}$, the mean number of selected segments $\widehat{K}$ is found to be 1, which is actually the correct number (as the intensity is actually constant in this case). When λ_R is slightly greater than one, in particular for small mean intensities $\overline{\lambda}$, the model selection procedure tends to underestimate the number of segments. This behavior is classical and desired to avoid false detection [see e.g. 10].

Accuracy of the change-points. The left panel of Figure 4 represents the mean Hausdorff distance as a function of the ratio λ_R. It shows the expected behavior in terms of accuracy of the estimated change-points. The Hausdorff distance $d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$ decreases as either the mean intensity $\overline{\lambda}$ or the ratio λ_R increases. The distance is almost zero with a mean intensity $\overline{\lambda} = 1000$ and a ratio $\lambda_R = 3$. In the typical case where $\overline{\lambda} = 100$, the distance decreases very fast as λ_R increases, but remains above .05, meaning that that some

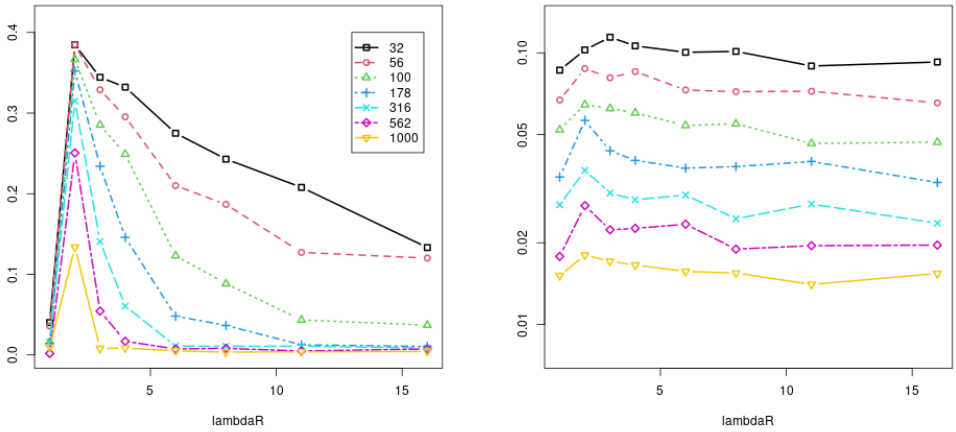

Figure 4. Left: Mean Hausdorff distance $d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$ as a fonction of the ratio λ_R. Legend panel = value of the mean intensity $\overline{\lambda}$. Right: Mean relative distance $\ell_2(\Lambda, \widehat{\Lambda})$ between the true and estimated cumulated intensity function $\Lambda(t)$ as a fonction of the ratio λ_R (same legend as left panel).

uncertainty remains about the precise location of the change point. Observe that when $\lambda_R = 1$, the Hausdorff distance is almost 0 for all values of $\overline{\lambda}$, simply because the selected number of segments is almost always 1, yielding $\widehat{\boldsymbol{\tau}} = \{0, 1\}$, which equals the true $\boldsymbol{\tau}$.

Estimation of the intensity function. The right panel of Figure 4 represents the mean relative ℓ_2 distance between the true and estimated cumulated intensity function, as a function of the ratio λ_R. It shows that the estimation of the cumulated intensity function Λ improves as the mean intensity $\overline{\lambda}$ increases. More interestingly, the ratio λ_R does not seem to strongly affect the precision of $\widehat{\Lambda}$. One possible explanation is that, although a strong contrast between the intensity of neighbor segments yields a better location of the change-points, even a small error in the change-point location induces a high error in terms of $\lambda(t)$ or $\Lambda(t)$.

Results for the marked Poisson process Table 6.2 gives the mean selected number of segments $\widehat{K}$ and the mean Hausdorff distance $d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$ for the four scenarios described in Section 6.1. When both the intensity and the mark parameters are constant, the mean number of selected segments is close to 1, that is the true number, and thus the mean the Hausdorff distance is close to 0. When both processes are significantly affected by the changes, the correct number of segments ($K = 6$) is recovered and the estimated segments is almost always close to the true one (the Hausdorff distance is close to 0). Two interesting results are the following: whether the change-points appear either in intensity only or in mark only do not change the power of detection, and adding the change-point information of the marks increases slightly this power.

Table 1. Mean selected number of segments $\widehat{K}$ and mean Hausdorff distance $d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$.

	$\widehat{K}$		$d(\boldsymbol{\tau}, \widehat{\boldsymbol{\tau}})$	
$\lambda \backslash \rho$	no signal	signal	no signal	signal
no signal	1.132	5.79	0.41	0.12
signal	5.41	5.99	0.11	0.05

7. Illustrations in vulcanology

In this section, as an illustration, we analyze two datasets describing the activity of volcanoes. In Section 7.1, only the eruption dates are considered and we look for homogeneous segments in a Poisson process. In Section 7.2, the duration of each eruption is also considered and we look for change-points in a marked Poisson process.

7.1. Poisson process

We consider here the eruptions of the Kilauea and Mauna Loa volcanoes in Hawaii, presented by [15]. The two datasets consists of the dates of eruptions recorded from year 1750 to year 1983 for the Kilauea volcano and to year 1984 for the Mauna Loa volcano. Over these periods, $n = 63$ eruptions were observed for the former and $n = 40$ for the latter. The original data only reports the number of eruptions per year (ranging from 0 to 4). The continuous times were restored by associated to each eruption a uniformly distributed date, within the corresponding year.

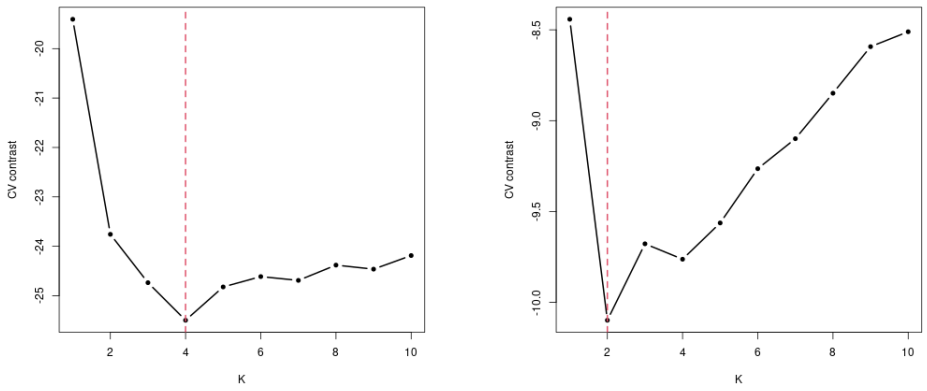

Figure 5. Eruptions of the Kilauea (left) et Mauna Loa (right) volcanos. Selection of the number of segments K: black solid-bullet line = criterion $\overline{\gamma}^{K,T}$, vertical red dotted line = optimal number of segments $\widehat{K}$.

Figure 5 provides the values of the cross-validation criterion $\overline{\gamma}^{K,T}$ as a function of K defined in Section 5 for each series of eruption dates. The criterion indicates the existence of four segments for the Kilauea and two segments (i.e. one change-point) for the Mauna Loa volcano. Observe that the latter criterion admits a local minima at $K = 4$.

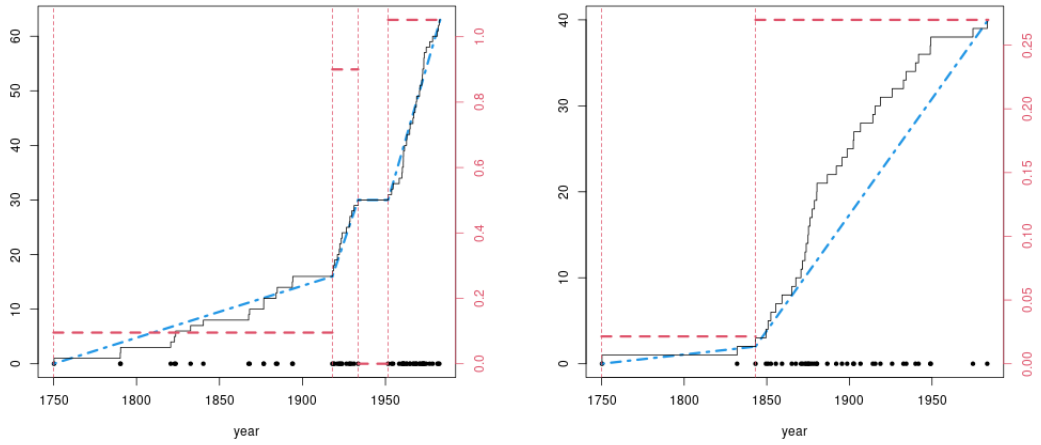

Figure 6. Eruptions of the Kilauea (left) et Mauna Loa (right) volcanos. Final segmentation: black bullets = eruptions times T_i, black solid line = observed count process $N(t)$, red vertical dotted lines = estimated change points $\widehat{\tau}_k$, blue dotted-dashed line = estimated cumulated intensity $\widehat{\Lambda}(t)$, red dashed horizontal lines = estimated piecewise constant intensity $\widehat{\lambda}(t)$ (referred to the right axis). The first vertical red dotted lines is $\tau_0 = 0$.

Figure 6 gives the optimal segmentation of the Kilauea series in $K = 4$ segments (three change-points located at the years 1918, 1934, 1952) and of the Mauna Loa series $K = 2$ segments (one change-point located at the year 1843). The four period segmentation of Kilauea series is consistent with observed process $N(t)$: the estimated (piecewise linear) cumulated intensity $\widehat{\Lambda}(t)$ remains close to $N(t)$ over the whole period of time. The two segment solution obtained for the Mauna Loa volcano displays a poorer fit. The distribution of the eruptions during the second period of time (from 1843 to 1984) displays a remaining heterogeneity, which may indicate a conservative behavior of the model selection procedure in this case, probably due to the small number of events observed over the period ($n = 40$).
The segmentations of the Mauna Loa series with $K = 3, 4, 5$ and 6 segments, given in Appendix A.2, Figure 10, show how the fit improves when K increases.

7.2. Marked Poisson process

The data comes from the technical report [22]. It consists in mount Etna volcano flunk eruption data. Indeed, flunk eruption constitutes one of the most important threat for a volcanic hazard assessment according to the authors. The first variable is the date of eruption of the volcano located in the North Eastern part of the Sicily Island, and the second one is the volume of lava spread. There are 63 events. We investigate these data with the inhomogeneous Poisson model and then with the marked Poisson model that we presented earlier. We search the best number of segments K in the collection $\{1, \ldots, 10\}$. The chosen dimension is $\widehat{K} = 2$ with both implementations. Both contrast function as represented in Figure 7. The obtained segmentations are given in Figure 8. We can

see that the segmentations are different and we believe that we can see the influence of the mark on the second graph. Furthermore, on Figure 9 where we have represented the segmentation obtained with the marked Poisson model, imposing $K = 3$. We can see that the method gives a segmentation with two change-points which are close to be the single change-point found by the Poisson process model and the single change-point found by the marked Poisson process model.

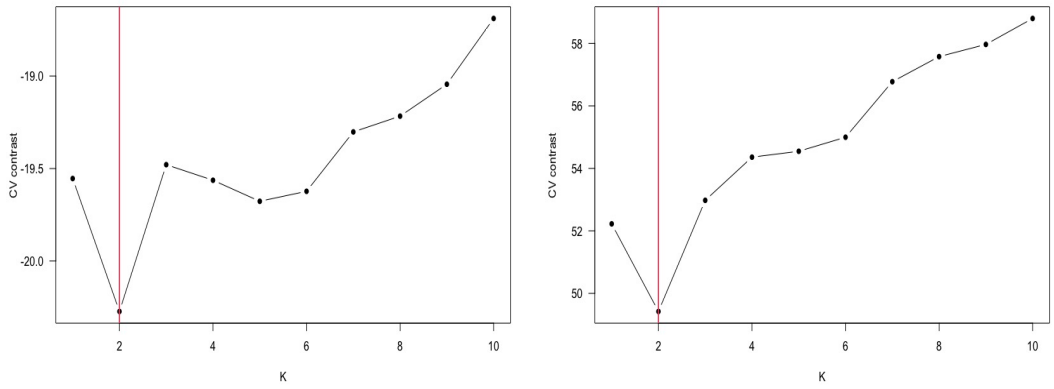

Figure 7. Eruptions of Etna. Left graph: Poisson model, right graph: marked Poisson model. Contrast function as a function of K, and in red the selected dimension $\widehat{K} = 2$ in both cases.

8. Discussion

Segmentation procedure. We have proposed a general frequentist framework for the detection of multiple change-points in a Poisson process. As usual in change-point detection problems, the inference procedure consists in two steps: (1) segment the observed path N into a fixed number of segments, and (2) choose this number. For step (1), we show that, for any contrast function satisfying segment-additive and concavity assumptions, the optimal segmentation in K segments can be obtained in an exact manner and reasonably fast using the dynamic programming algorithm. Note that the use of DP basically relies on the property of segment-additive and this general work includes maximum likelihood or least square inference. For the model selection step (2), we propose to use a cross-validation strategy.

Penalized contrasts. In many cases, the selection of the number of segments K (step (2)) is based on a penalized contrast, where the penalty can depend only of the number of segments (think of the classical BIC and AIC or see [19, 18, 16, 17]) or on both the number of segments and their lengths as in [31]. As proposed in [16], when using a penalty proportional to K, the two separate tasks (1) and (2) can be embedded in a single one using DP because the penalized contrast remains segment-additive.

Our general framework can also include more general penalized contrasts, provided that the penalty also satisfies the same condition for admissibility. For example, consider

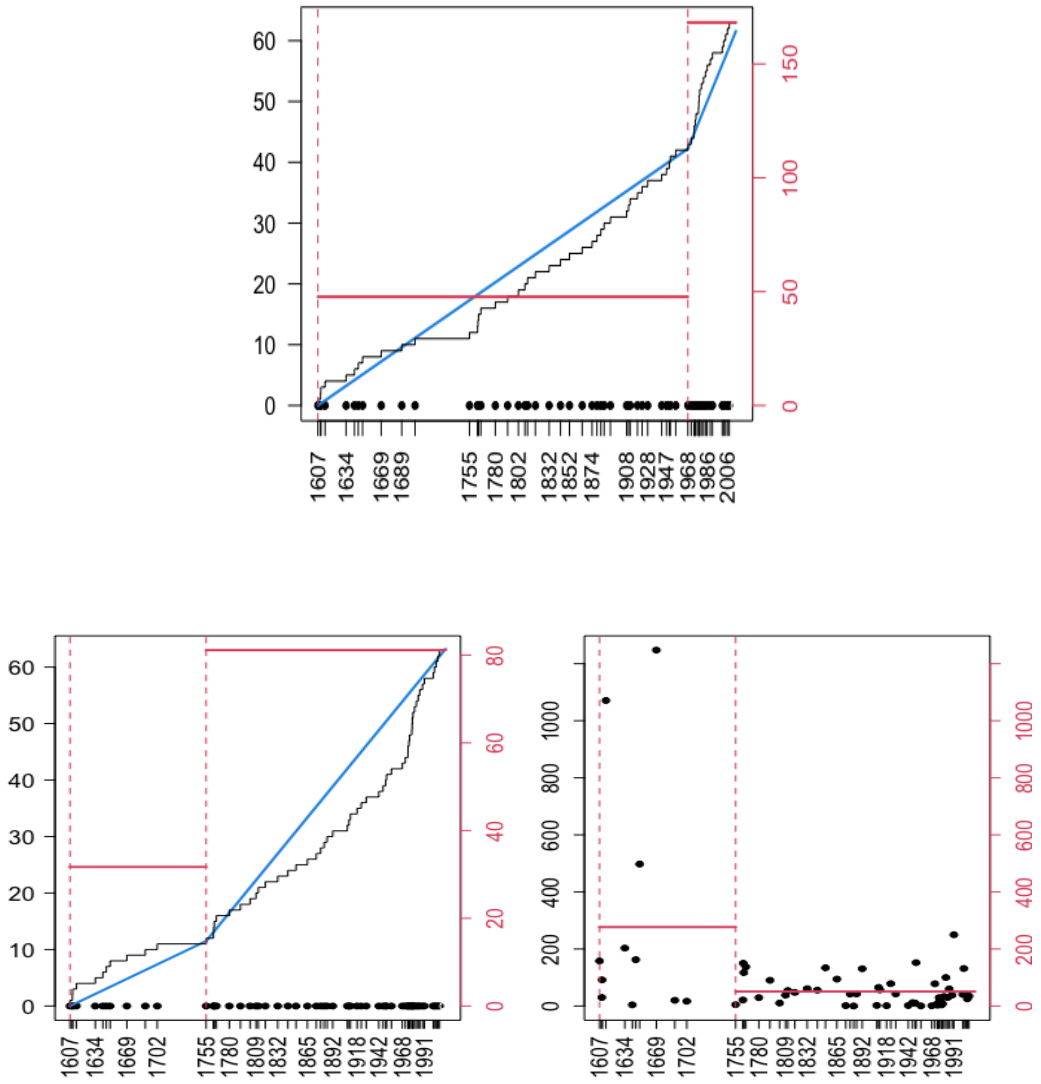

Figure 8. Eruptions of Etna. Top line: Poisson model, bottom line: marked Poisson model. (Left graphs) Final segmentations: black bullets = eruptions times T_i, black solid line = observed count process (N_t), red vertical dotted lines = estimated change points $\widehat{\tau}_k$, blue dotted-dashed line = estimated cumulated intensity $\widehat{\Lambda}(t)$, red dashed horizontal lines = estimated piecewise constant intensity $\widehat{\lambda}(t)$ (referred to the right axis). Bottom right graph: black bullets = marks X_i, red vertical dotted lines = estimated change points, red dashed horizontal lines = estimated piecewise constant parameter of the exponential distribution of the marks $\widehat{\rho}$.

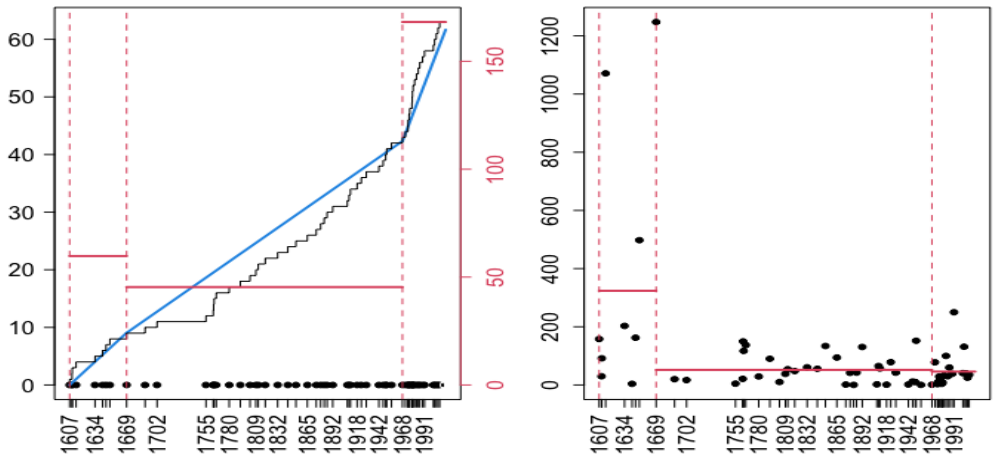

Figure 9. Eruptions of Etna. Marked Poisson model. Same legend as top line of Figure 8 with $K = 3$.

an admissible contrast, to which a penalty is added

$$\sum_{k=1}^{K} C(\nu_k, \Delta\tau_k) + \text{pen}(\boldsymbol{\tau}, K).$$

Suppose that the penalty function has a form similar to this of the modified BIC proposed by [31], that is

$$\text{pen}(\boldsymbol{\tau}, K) = \sum_{k=1}^{K} f(\Delta\tau_k) + \beta K,$$

β being a constant. The penalized cost

$$\widetilde{C}(\Delta N_k, \Delta\tau_k) = C(\nu_k, \Delta\tau_k) + f(\Delta\tau_k) + \beta,$$

is still concave in $\Delta\tau_k$, so the procedure proposed for step (1), applied to the penalized contrast $\gamma(\boldsymbol{\tau}) = \sum_{k=1}^{K} \widetilde{C}(\Delta N_k, \Delta\tau_k)$ achieves steps (1) and (2) at once. This combination obviously improves the computational time of the procedure.

Computational time. The computational cost of DP is quadratic in the number of events n (that is $\mathcal{O}((2n+1)^2 K)$). When dealing with trajectories containing a very high number of events, DP can be time demanding. The time-efficiency could probably be improved using the pruned version of DP proposed by [17] reducing drastically the computational cost (almost linear in n).

Segmentation of multiple trajectories. In the change-point detection literature, simultaneous segmentation refers to the detection of multiple change-points in several series

observed over the same period of time. In this setting, the change points are supposed to be the same for all series, so the contrast for the S series typically writes

$$\gamma(\boldsymbol{\tau}) = \sum_{s=1}^{S} \sum_{k=1}^{K} C(\Delta N_k^s, \Delta\tau_k).$$

The cost function $\sum_{s=1}^{S} C(\Delta N_k^s, \Delta\tau_k)$ is still concave wrt $\Delta\tau_k$ and the problem can be dealt with, with the same complexity as in the univariate case.

Acknowledgments

This work has been conducted within the FP2M federation (CNRS FR 2036). This work is also part of the 2022 DAE 103 EMERGENCE(S) - PROCECO project supported by Ville de Paris.

References

[1] J. Achcar, E. Martinez, A. Ruffino-Netto, C. Paulino, and P. Soares. A statistical model investigating the prevalence of tuberculosis in new york city using counting processes with two change-points. *Epidemiology & Infection*, 136(12):1599–1605, 2008.

[2] J. A. Achcar, E. R. Rodrigues, and G. Tzintzun. Using non-homogeneous poisson models with multiple change-points to estimate the number of ozone exceedances in mexico city. *Environmetrics*, 22(1):1–12, 2011.

[3] P. K. Andersen, Ø. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York, 1993.

[4] S. Arlot and A. Celisse. Segmentation of the mean of heteroscedastic data via cross-validation. *Statistics and Computing*, 21(4):613–632, 2011.

[5] I. E. Auger and C. E. Lawrence. Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.*, 51(1):39–54, 1989.

[6] J. Bai and P. Perron. Computation and analysis of multiple structural change models. *J. Appl. Econ.*, 18:1–22, 2003.

[7] M. S. Bebbington. Identifying volcanic regimes using hidden markov models. *Geophysical Journal International*, 171(2):921–942, 2007.

[8] Y. Bessy-Roland, A. Boumezoued, and C. Hillairet. Multivariate hawkes process for cyber insurance. *Annals of Actuarial Science*, 15(1):14–39, 2021.

[9] S. Chakar, E. Lebarbier, C. Lévy-Leduc, S. Robin, et al. A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli*, 23(2):1408–1447, 2017.

[10] A. Cleynen and É. Lebarbier. Model selection for the segmentation of multiparameter exponential family distributions. 2017.

[11] D. Daley and D. Vere-Jones. *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media, 2007.

[12] N. El Karoui, S. Loisel, and Y. Salhi. Minimax optimality in robust detection of a disorder time in doubly-stochastic poisson processes. *The Annals of Applied Probability*, 27(4):2515–2538, 2017.

[13] M. Fromont, F. Grela, and R. Le Guével. Minimax and adaptive tests for detecting abrupt and possibly transitory changes in a Poisson process. working paper or preprint, Sept. 2021.

[14] A. Gupta and J. Baker. A bayesian change point model to detect changes in event occurrence rates, with application to induced seismicity. In *12th International Conference on Applications of Statistics and Probability in Civil Engineering*, pages 12–15. The Univ. of British Columbia Vancouver, Canada, 2015.

[15] C.-H. Ho and M. Bhaduri. A quantitative insight into the dependence dynamics of the Kilauea and Mauna Loa volcanoes, Hawaii. *Mathematical Geosciences*, 49(7):893–911, 2017.

[16] B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai. An algorithm for optimal partitioning of data on an interval. *Signal Processing Letters, IEEE*, 12(2):105–108, 2005.

[17] R. Killick, P. Fearnhead, and I. A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of The American Statistical Association*, 107(500):1590–1598, 2012.

[18] M. Lavielle. Using penalized contrasts for the change-point problem. *Signal Processing*, 85(8):1501–1510, 2005.

[19] E. Lebarbier. Detecting multiple change-points in the mean of gaussian process by model selection. *Signal processing*, 85(4):717–736, 2005.

[20] S. Li, Y. Xie, M. Farajtabar, A. Verma, and L. Song. Detecting changes in dynamic events over networks. *IEEE Transactions on Signal and Information Processing over Networks*, 3(2):346–359, 2017.

[21] Y. S. Niu, N. Hao, and H. Zhang. Multiple change-point detection: a selective overview. *Statistical Science*, pages 611–623, 2016.

[22] L. Passarelli, B. Sansò, L. Sandri, and W. Marzocchi. Testing forecasts of a new bayesian time-predictable model of eruption occurrence. *Journal of volcanology and geothermal research*, 198(1-2):57–75, 2010.

[23] A. Raftery and V. Akman. Bayesian analysis of a poisson process with a change-point. *Biometrika*, pages 85–89, 1986.

[24] L. Shaochuan. Bayesian multiple changepoint detection for stochastic models in continuous time. *Bayesian Analysis*, 16(2):521–544, 2021.

[25] J. J. Shen and N. R. Zhang. Change-point model on nonhomogeneous poisson processes with application in copy number profiling by next-generation dna sequencing. *The Annals of Applied Statistics*, 6(2):476–496, 2012.

[26] G. Shmueli and H. Burkom. Statistical challenges facing early outbreak detection in biosurveillance. *Technometrics*, 52(1):39–51, 2010.

[27] C. Truong, L. Oudre, and N. Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2020.

[28] K.-L. Tsui, W. Chiu, P. Gierlich, D. Goldsman, X. Liu, and T. Maschek. A review of healthcare, public health, and syndromic surveillance. *Quality Engineering*, 20(4):435–450, 2008.

[29] W. R. West and T. R. Ogden. Continuous-time estimation of a change-point in a poisson process. *Journal of Statistical Computation and Simulation*, 56(4):293–302, 1997.

[30] T. Young Yang and L. Kuo. Bayesian binary segmentation procedure for a poisson process with multiple changepoints. *Journal of Computational and Graphical Statistics*, 10(4):772–785, 2001.

[31] N. R. Zhang and D. O. Siegmund. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63(1):22–32, 2007.

A. Appendix

A.1. Dynamic Programming (DP) algorithm

In this section, we describe the principle of the standard Dynamic Programming algorithm for a discrete segmentation problem and specify how to apply it on a finite and given grid as in our case.

DP for discrete change-points problem. If we observed an ordered sequence of A observations, the goal is to find a partition of the discrete grid $[\![1, A]\!]$ into K segments delimited by $K-1$ change-points denoted a_k for $k = 1, \ldots, K-1$ with the convention $a_0 = 0$ and $a_K = A$. The k-th segment is $[\![a_{k-1}+1, a_k]\!]$ and we define the set of all possible segmentations with $K-1$ change-points:

$$\mathcal{A}_A^K = \{\mathbf{a} = (a_1, a_2, \ldots, a_{K-1}) \in \mathbb{N}^{K-1} : a_0 = 0 < a_1 < \ldots < a_{K-1} < a_K = A\},$$

with cardinality $\binom{A-1}{K-1}$. Even if this space is finite, it is extremely large and a naive search is prohibitive in a computational point of view. The well known solution is to use the DP algorithm that can be applied if and only if the quantity to be optimized is segment-additive:

$$\widehat{\mathbf{a}} = (\widehat{a}_1, \widehat{a}_2, \ldots, \widehat{a}_{K-1}) = \underset{\mathbf{a}\in\mathcal{A}_A^K}{\operatorname{argmin}}\ R_{\mathbf{a}} = \underset{\mathbf{a}\in\mathcal{A}_A^K}{\operatorname{argmin}}\ \sum_{k=1}^{K} C(a_{k-1}+1 : a_k),$$

where $R_{\mathbf{a}}$ is called the cost of the segmentation $\mathbf{a}$ and $C(a_{k-1}+1 : a_k) = C([\![a_{k-1}+1, a_k]\!])$ the cost of the segment $[\![a_{k-1}+1, a_k]\!]$. We define

$$C_{K,A} = \min_{\mathbf{a}\in\mathcal{A}_A^K} \sum_{k=1}^{K} C(a_{k-1}+1 : a_k),$$

the cost of the best segmentation in K segments. If we note $C(i : j) = C([\![i, j]\!])$ the cost of the segment $[\![i, j]\!]$, thanks to the segment-additivity property of $R_{\mathbf{a}}$, DP solved the optimization problem using the following update rule:

$$C_{K,A} = \min_{0<a_1<\ldots<a_{K-1}<A} \sum_{k=1}^{K} C(a_{k-1}+1 : a_k) = \min_{K-1\le h<A} \left\{C_{K-1,h} + C(h+1 : A)\right\}.$$

This algortihm requires the calculation of the cost of each possible segment that is

$$\begin{aligned} C(i : j) &= C([\![i, j]\!]) \quad \text{if } 1 \le i \le j \le A \\ &= +\infty \quad \text{otherwise.} \end{aligned}$$

Search in a fixed and finite grid. Let

$$\mathbf{tp} = \{\mathrm{tp}_1, \mathrm{tp}_2, \ldots, \mathrm{tp}_A\},$$

be this grid where $0 < \mathrm{tp}_i < 1$. If a segment is defined as $(\mathrm{tp}_\ell, \mathrm{tp}_h]$ with $\ell < h$, we can apply DP with

$$\begin{aligned} C(i:j) &= C((\mathrm{tp}_{i-1}, \mathrm{tp}_j]) \quad \text{if } 1 \leq i \leq j \leq A+1 \\ &= +\infty \quad \text{otherwise.} \end{aligned}$$

with the convention $\mathrm{tp}_0 = 0$ and $\mathrm{tp}_{A+1} = 1$. The complexity is thus $\mathcal{O}((A+1)^2 K)$ and the optimal change-points are given by

$$\widehat{\tau}_k = \mathrm{tp}_{\widehat{a}_k}$$

A.2. Illustrations

Figure 10 displays the segmentations of the Mauna Loa series obtained when considering $K = 3, 4, 5$ and 6 segments. The fit gets obviously better when K increases and the cross-validation procedure proposed in Section 5 aims at preventing from over-fitting. The exploration of higher values of K may still be useful to discover tiny heterogeneity, that may deserve further investigations (see, e.g., the 1870-1880 segment exhibited with $K = 6$).

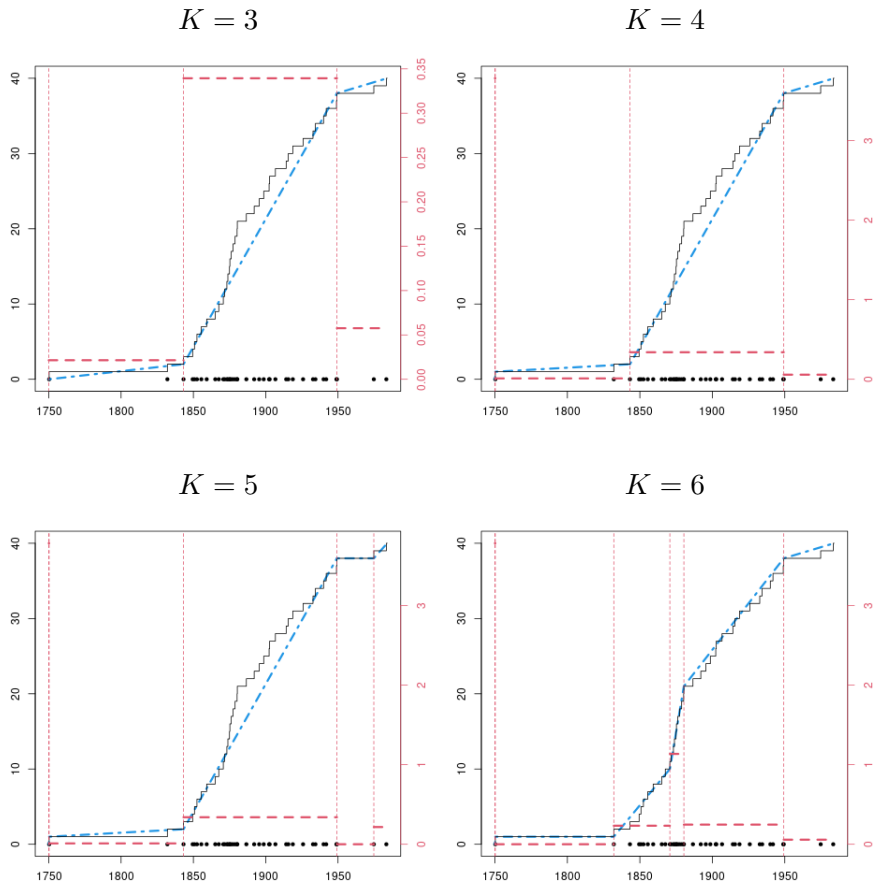

Figure 10. Eruptions of the Mauna Loa volcano. Segmentation with $K = 3, 4, 5, 6$ segments. Same legend as Figure 6. Note that the scale of the right y-axis differs between plots. The first segment for $K = 4, 5$ and 6 is included in the first observation year (1750) and contains only one event.