JOINT SEGMENTATION OF MULTIVARIATE POISSONIAN TIME SERIES. APPLICATION TO BURST AND TRANSIENT SOURCE EXPERIMENTS

Nicolas Dobigeon[†], *Jean-Yves Tourneret*[†] *and Jeffrey D. Scargle*[*]

[†]IRIT/ENSEEIHT/TéSA - 2 rue Charles Camichel, BP 7122, 31071 Toulouse cedex 7, France
[*]Space Science Division, NASA, Ames Research Center, Moffett Field, CA, USA
email: Nicolas.Dobigeon@enseeiht.fr, Jean-Yves.Tourneret@enseeiht.fr, Jeffrey.D.Scargle@nasa.gov

ABSTRACT

This paper addresses the problem of detecting significant intensity variations in multiple Poissonian time-series. This detection is achieved by using a constant Poisson rate model and a hierarchical Bayesian approach. An appropriate Gibbs sampling strategy allows joint estimation of the unknown parameters and hyperparameters. An extended model that includes constraints on the segment lengths is also proposed. Simulation results performed on synthetic and real data illustrate the performance of the proposed algorithm.

1. INTRODUCTION

Signal segmentation has received considerable interest in the signal processing and statistical literature (see for instance [1, 2, 3] and references therein). In particular, Bayesian estimators have shown very interesting properties for this problem [4, 5]. The complexity of the posterior distributions for the unknown parameters generally requires to develop appropriate simulation methods such as Markov Chain Monte Carlo (MCMC) methods [4, 5]. These methods can also be used to estimate the unknown hyperparameters by introducing a second level of hierarchy within the Bayesian paradigm. The hyperparameters are then integrated out from the joint posterior distribution or estimated from the observed data [5].

The problem of segmenting burst and transient source experiment (BATSE) data was considered more recently. Bayesian algorithms based on a constant Poisson rate model were studied in [6, 7]. The algorithm studied in [6] decomposed the observed signal into two segments and iterated the process several times until a stopping rule was satisfied. A new algorithm avoiding the use of a stopping rule was proposed in [7]. However, the algorithm required to define an appropriate prior distribution for the number of changepoints. These limitations were removed in [8] where a hierarchical Bayesian model allowed to estimate the number of changepoints, their locations as well as confidence intervals for the estimated parameters. This paper extends the algorithm proposed in [8] to multiple Poissonian time series.

The paper is organized as follows. The joint segmentation problem is formulated in section 2. Section 3 describes the different elements of the hierarchical model which will be used to solve this segmentation problem. Section 4 studies a Gibbs sampler which generates samples distributed according to the posteriors of the unknown parameters. Some simulation results on synthetic data are presented in Section 5. Section 6 introduces a markovian model for the changepoints which allows to reject segmentations involving segments shorter than a given minimal length. Section 7 applies the proposed methodology to real data recorded by the NASA compton Gamma ray observatory's BATSE (Burst and transient source experiment). Conclusions are reported in the last section.

2. PROBLEM FORMULATION

The statistical properties of the time series studied in this paper are defined as follows:

$$y_{j,i} \sim \mathscr{P}\left(\lambda_{j,k}\right),$$

where $j = 1,...,J$, $k = 1,...,K_j$, $i \in I_{j,k} = \{l_{j,k-1}+1,...,l_{j,k}\}$, and the following notations have been used: $\mathscr{P}(\lambda)$ denotes a Poisson distribution with parameter λ, J is the number of signals to be segmented, K_j is the number of segments in the j^{th} observed signal and $l_{j,k}$ is the sample point after which the k^{th} change occurs in the j^{th} signal (by convention $l_{j,0} = 0$ and $l_{j,K_j} = n$ where n is the number of observed samples). Moreover, the sequences $\mathbf{y}_l = [y_{l,1},\ldots,y_{l,n}]$ and $\mathbf{y}_m = [y_{m,1},\ldots,y_{m,n}]$ are assumed to be independent for $l \neq m$. Segmenting the astronomical time series $\mathbf{y}_j = [y_{j,1},...,y_{j,n}]$ jointly consists of estimating the change-points numbers K_j and their positions $l_{j,k}$ (for $j = 1,...,J$ and $k = 1,...,K_j$) from the observations contained in $\mathbf{Y} = [\mathbf{y}_1,\ldots,\mathbf{y}_J]^{\mathsf{T}}$.

3. HIERARCHICAL BAYESIAN MODEL

The unknown parameters for the segmentation problem (introduced in the previous section) are the numbers of segments K_j, the change-point locations $l_{j,k}$ and the Poisson parameters $\lambda_{j,k}$ (with $\lambda_j = [\lambda_{j,1},\ldots,\lambda_{j,K_j}]^{\mathsf{T}}$ and $\Lambda = \{\lambda_1,\ldots,\lambda_J\}$). A standard reparameterization consists of introducing indicators $r_{j,i}$, $j \in \{1,\ldots,J\}$, $i \in \{1,\ldots,n\}$ such that:

$$\begin{cases} r_{j,i} = 1 \text{ if there is a changepoint at time } i \text{ in the signal } j, \\ r_{j,i} = 0 \text{ otherwise,} \end{cases}$$

with $r_{j,n} = 1$ (this condition ensures that the number of changepoints and the number of steps of the j^{th} signal are equal to $K_j = \sum_{i=1}^{n} r_{j,i}$). The unknown parameter vector resulting from this reparameterization is $\theta = \{\theta_1,\ldots,\theta_J\}$ where $\theta_j = \left\{\mathbf{r}_j, \lambda_j\right\}$ and $\mathbf{r}_j = [r_{j,1},\ldots,r_{j,n}]$. Note that the unknown parameter vector θ belongs to a space $\Theta = \{0,1\}^{J\times n} \times \prod_{j=1}^{J} \mathbb{R}_+^{K_j}$ whose dimension depends on the parameters K_j, $j = 1,\ldots,J$. This paper proposes to estimate the unknown parameter vector θ by using Bayesian estimation theory. Bayesian inference on θ is based on the posterior distribution $f(\theta|\mathbf{Y}) \propto f(\mathbf{Y}|\theta) f(\theta)$. This posterior distribution is related to the likelihood of the observations $f(\mathbf{Y}|\theta)$ and the parameter priors summarized in $f(\theta)$. These distributions are detailed in the next section.

3.1 Likelihood

The likelihood of the observed vector $\mathbf{Y}$ can be expressed as:

$$\begin{aligned} f(\mathbf{Y}|\theta) &= \prod_{j=1}^{J}\prod_{k=1}^{K_j}\prod_{i\in I_{j,k}} \frac{\lambda_{j,k}^{y_{j,i}} \exp\left(-\lambda_{j,k}\right)}{y_{j,i}!} \\ &\propto \prod_{j=1}^{J}\prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(\mathbf{r}_j)} \exp\left(-\lambda_{j,k} n_{j,k}\left(\mathbf{r}_j\right)\right), \end{aligned} \tag{1}$$

where $\propto$ means "proportional to", $s_{j,k}\left(\mathbf{r}_j\right) = \sum_{i\in I_{j,k}} y_{j,i}$ and $n_{j,k}\left(\mathbf{r}_j\right) = l_{j,k} - l_{j,k-1}$ (number of samples in the k^{th} interval $I_{j,k}$ of the j^{th} signal).

3.2 Parameter Priors

3.2.1 Indicator Vector

The indicator vectors $\mathbf{R}_i = [r_{1,i}, \ldots, r_{J,i}]^\mathsf{T}$ and $\mathbf{R}_{i'} = [r_{1,i'}, \ldots, r_{J,i'}]^\mathsf{T}$ are assumed to be independent for any $i \neq i'$. Moreover, the possible correlations between the change locations in the different observed signals are adjusted by choosing an appropriate prior distribution $f(\mathbf{R}|\mathbf{P})$, where $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_J]^\mathsf{T}$ and $\mathbf{P}$ is defined below. We assume that the probability of having $[r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}$ does not depend on i (with $\boldsymbol{\varepsilon} \in \mathscr{E} = \{0,1\}^J$) and is denoted $\mathrm{P}_{\boldsymbol{\varepsilon}}$. As a consequence, the indicator prior distribution is:

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\boldsymbol{\varepsilon} \in \mathscr{E}} \mathrm{P}_{\boldsymbol{\varepsilon}}^{S_{\boldsymbol{\varepsilon}}(\mathbf{R})},$$

where $\mathbf{P} = \{\mathrm{P}_{\boldsymbol{\varepsilon}}\}_{\boldsymbol{\varepsilon} \in \mathscr{E}}$, $\mathrm{P}_{\boldsymbol{\varepsilon}} \in \{\mathrm{P}_{0\ldots0}, \ldots, \mathrm{P}_{1\ldots1}\}$ and $S_{\boldsymbol{\varepsilon}}(\mathbf{R})$ is the number of lags such that $[r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}$. The more likely the configuration $[r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}$, the higher the probability $\mathrm{P}_{\boldsymbol{\varepsilon}}$. This choice induces correlation between change-point locations in the different time series.

3.2.2 Poisson Parameters

The parameters $\lambda_{j,k}$ are assumed *a priori* independent and Gamma distributions are assigned to these Poisson parameters:

$$\lambda_{j,k}|\nu,\gamma \sim \mathscr{G}(\nu,\gamma),$$

where $\nu = 1$ (as in [5]), γ is an adjustable hyperparameter and $\mathscr{G}(a,b)$ denotes the Gamma distribution with parameters a and b. The previous assumptions yield the following prior for Λ:

$$f(\Lambda|\gamma) = \prod_{j=1}^{J} \left[\frac{\gamma^{\nu K_j} e^{-\gamma \sum_{k=1}^{K_j} \lambda_{j,k}}}{\Gamma(\nu)^{K_j}} \prod_{k=1}^{K_j} \left(\lambda_{j,k}^{\nu-1} \mathbb{I}_{\mathbb{R}^+}(\lambda_{j,k}) \right) \right],$$

where $\mathbb{I}_{\mathbb{R}^+}(\cdot)$ is the indicator function defined on $\mathbb{R}^+$ (i.e. $\mathbb{I}_{\mathbb{R}^+}(t) = 1$ if $t \geq 0$ and $\mathbb{I}_{\mathbb{R}^+}(t) = 0$ otherwise) and $\Gamma(\cdot)$ is the Gamma function (i.e. $\Gamma(t) = \int_0^{+\infty} u^{t-1} e^{-u} du, \quad t > 0$). The Gamma distribution, which is the conjugate prior for the parameters $\lambda_{j,k}$, allows to integrate out these parameters from the joint posterior (see section 3.4).

The hyperparameter vector associated with the priors defined above is $\boldsymbol{\Phi} = (\mathbf{P}, \gamma)$. Of course, the quality of the Bayesian segmentation depends on the values of the hyperparameters. In particular applications, these hyperparameters can be fixed from available information regarding the observed signals as in [9]. However, in order to increase the accuracy of the algorithm, hyperparameters can be considered as random variables with noninformative priors as in [5]. This strategy, involving different levels in a Bayesian prior hierarchy, results in so-called *hierarchical Bayesian models*. Such models require that one defines hyperparameter priors (sometimes referred to as *hyper-priors*), as detailed in the next section.

3.3 Hyperparameter Priors

3.3.1 Hyperparameter γ

The prior distribution for γ is a noninformative Jeffreys' prior (as in [5]) which reflects the absence of knowledge regarding this hyperparameter:

$$f(\gamma) = \frac{1}{\gamma} \mathbb{I}_{\mathbb{R}^+}(\gamma).$$

3.3.2 Hyperparameter $\mathbf{P}$

This paper assumes that the prior distribution for $\mathbf{P}$ is a Dirichlet distribution defined on the simplex $\mathscr{P} = \{\mathbf{P}; \sum_{\boldsymbol{\varepsilon} \in \mathscr{E}} \mathrm{P}_{\boldsymbol{\varepsilon}} = 1, \mathrm{P}_{\boldsymbol{\varepsilon}} > 0\}$ with parameter vector $\boldsymbol{\alpha} = [\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1}]^\mathsf{T}$ denoted as:

$$\mathbf{P}|\boldsymbol{\alpha} \sim \mathscr{D}_{2^J}(\boldsymbol{\alpha}).$$

The Dirichlet distribution is a very common distribution for parameters summing to 1. Assuming that the different hyperparameters are *a priori* independent, the prior distribution for the hyperparameter vector $\boldsymbol{\Phi}$ can be written as follows:

$$f(\boldsymbol{\Phi}|\boldsymbol{\alpha}) \propto \left(\prod_{\boldsymbol{\varepsilon} \in \mathscr{E}} \mathrm{P}_{\boldsymbol{\varepsilon}}^{\alpha_{\boldsymbol{\varepsilon}}-1} \right) \frac{1}{\gamma} \mathbb{I}_{\mathbb{R}^+}(\gamma) \mathbb{I}_{\mathscr{P}}(\mathbf{P}), \tag{2}$$

where $\alpha_{\boldsymbol{\varepsilon}} \in \{\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1}\}$. This paper has assumed all values of $\alpha_{\boldsymbol{\varepsilon}}$ are equal. In this case, the Dirichlet distribution reduces to the uniform distribution on $\mathscr{P}$ reflecting the lack of information regarding $\mathbf{P}$.

3.4 Posterior Distribution of θ

The posterior distribution of the unknown parameter vector $\theta = \{\Lambda, \mathbf{R}\}$ can be computed from the following hierarchical structure:

$$f(\theta|\mathbf{Y}) = \int f(\theta, \boldsymbol{\Phi}|\mathbf{Y}) d\boldsymbol{\Phi} \propto \int f(\mathbf{Y}|\theta) f(\theta|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) d\boldsymbol{\Phi},$$

where $f(\mathbf{Y}|\theta)$ and $f(\boldsymbol{\Phi})$ have been defined in (1) and (2). The previous priors and hyperpriors allow one to integrate out the nuisance parameters Λ and $\mathbf{P}$ from the joint distribution $f(\theta, \boldsymbol{\Phi}|\mathbf{Y})$, yielding:

$$\frac{f(\mathbf{R}, \gamma|\mathbf{Y})}{C(\mathbf{R}|\mathbf{Y})} \propto \frac{1}{\gamma} \prod_{j=1}^{J} \left[\left(\frac{\gamma^{\nu}}{\Gamma(\nu)} \right)^{K_j} \prod_{k=1}^{K_j} \frac{\Gamma(s_{j,k} + \nu)}{(n_{j,k} + \gamma)^{s_{j,k}+\nu}} \right] \mathbb{I}_{\mathbb{R}^+}(\gamma), \tag{3}$$

with

$$C(\mathbf{R}|\mathbf{Y}) = \frac{\prod_{\boldsymbol{\varepsilon} \in \mathscr{E}} \Gamma(S_{\boldsymbol{\varepsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\varepsilon}})}{\Gamma(\sum_{\boldsymbol{\varepsilon} \in \mathscr{E}} (S_{\boldsymbol{\varepsilon}}(\mathbf{R}) + \alpha_{\boldsymbol{\varepsilon}}))}.$$

The posterior distribution (3) is too complex to obtain closed-form expressions of the Bayesian estimators for the unknown parameters (such as the minimum mean square error (MMSE) estimator or the maximum *a posteriori* (MAP) estimator). In this case, it is quite common to apply MCMC methods to generate samples which are asymptotically distributed according to the posteriors of interest. The samples can then be used to estimate the unknown parameters.

4. GIBBS SAMPLER FOR CHANGE-POINT DETECTION

The Gibbs sampler is an iterative sampling strategy which consists of generating samples distributed according to the full conditional distributions of each parameter. This paper proposes to sample according to the distribution $f(\mathbf{R}, \gamma|\mathbf{Y})$ defined in (3). The main steps of the algorithm are summarized in Sections 4.1 to 4.3. The reader is invited to consult [10] for more details.

4.1 Generation of Samples Distributed According to $f(\mathbf{R}|\gamma, \mathbf{Y})$

This generation is achieved by using the Gibbs Sampler to draw $(n-1)$ samples distributed according to $f(r_{1,i}, \ldots, r_{J,i}|\gamma, \mathbf{Y})$. This random variable is discrete and takes its values in $\mathscr{E} = \{0,1\}^J$. Consequently, its distribution is fully characterized by the probabilities $P([r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}|\gamma, \mathbf{Y}), \boldsymbol{\varepsilon} \in \mathscr{E}$. By using the notations $\mathbf{R}_{-i}$ to denote the matrix $\mathbf{R}$ whose ith column has been removed, the following result can be obtained:

$$P\left([r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}|\mathbf{R}_{-i}, \gamma, \mathbf{Y} \right) \propto f\left(\mathbf{R}_i(\boldsymbol{\varepsilon}), \gamma|\mathbf{Y} \right),$$

where $\mathbf{R}_i(\boldsymbol{\varepsilon})$ is the matrix $\mathbf{R}$ whose ith column has been replaced by the vector $\boldsymbol{\varepsilon}$. This yields a closed-form expression of the probabilities $P\left([r_{1,i}, \ldots, r_{J,i}]^\mathsf{T} = \boldsymbol{\varepsilon}|\mathbf{R}_{-i}, \gamma, \mathbf{Y} \right)$ after appropriate normalization.

4.2 Generation of Samples Distributed According to $f(\gamma|\mathbf{R}, \mathbf{Y})$

To obtain samples distributed according to $f(\gamma|\mathbf{R}, \mathbf{Y})$, it is very convenient to simulate vectors distributed according to the joint distribution distribution $f(\gamma, \Lambda|\mathbf{R}, \mathbf{Y})$ by using Gibbs moves. This step can be decomposed as follows:

- **Draw samples according to** $f(\Lambda|\mathbf{R},\gamma,\mathbf{Y})$
  Looking carefully at the joint distribution $f(\theta,\Phi|\mathbf{Y})$, the following results can be obtained:

$$\lambda_{j,k}|\mathbf{R},\gamma,\mathbf{Y} \sim \mathscr{G}\left(s_{j,k}\left(\mathbf{r}_j\right)+\nu, n_{j,k}\left(\mathbf{r}_j\right)+\gamma\right).$$

- **Draw samples according to** $f(\gamma|\mathbf{R},\Lambda,\mathbf{Y})$
  This is achieved as follows:

$$\gamma|\mathbf{R},\Lambda \sim \mathscr{G}\left(\nu\sum_{j=1}^{J}K_j, \sum_{j=1}^{J}\sum_{k=1}^{K_j}\lambda_{j,k}\right).$$

4.3 Posterior Distribution of hyperparameter $\mathbf{P}$

The hyperparameter $\mathbf{P}$ carries information regarding the probability of having simultaneous changes at a given location. As a consequence, its estimation may be interesting in practical applications. The posterior distribution of this parameter conditioned upon the indicator vector $\mathbf{R}$, the vector of observed samples $\mathbf{Y}$ and the parameters α can be easily derived. This is a 2^J-Dirichlet distribution with parameters $(\alpha_\varepsilon + S_\varepsilon(\mathbf{R}))_{\varepsilon\in\mathscr{E}}$:

$$\mathbf{P}|\mathbf{R},\mathbf{Y},\alpha \sim \mathscr{D}_{2^J}(\alpha_\varepsilon + S_\varepsilon(\mathbf{R})). \tag{4}$$

5. SIMULATIONS

Many simulations have been conducted to validate the previous segmentation algorithm. The simulations presented in this section have been obtained for $J=2$ signals of $n=120$ samples. The change-point locations for these two sequences are $l_1=(30,70,120)$ and $l_2=(30,120)$. The parameters of the Poisson distributions are $\lambda_1=[8,23,12]^\mathsf{T}$ and $\lambda_2=[14,12]^\mathsf{T}$. The hyperparameters have been set to $\nu=2$ and $\alpha_\varepsilon=1,\forall\varepsilon$. All figures have been obtained after averaging the results of 64 Markov chains. The total number of runs for each Markov chain is $N_{MC}=400$, including $N_{bi}=50$ burn-in iterations. Thus, only the last 350 Markov chain output samples are used for the estimations. The results provided by the joint segmentation procedure are compared with those provided by two 1D segmentations (which consists of performing the proposed algorithm on the two sequences independently).

5.1 Posterior Distribution of the Change-point Locations

The MMSE estimates of the change locations are depicted on figures 1 and 2 for 1D and joint approaches. Note that these estimates are the posterior probabilities to have changes at the different locations (since $r_{j,i}$ is a binary random variable). For example, there is a very high posterior probability that a change occurred at lags $i=30$ and $i=70$ in the first sequence (with both 1D and 2D approaches). The advantage of the joint segmentation procedure is illustrated on these two figures: the change at lag $i=30$ in the second sequence is much better estimated by the joint approach (fig. 2) than by the 1D approach (fig. 1). Thus, the joint 2D segmentation procedure provides better results than two independent 1D segmentations on this example.

5.2 Posterior Distribution of (K_1,K_2)

The proposed algorithm generates samples $\left(\mathbf{R}^{(t)},\gamma^{(t)}\right)$ distributed according to the posterior distribution $f(\mathbf{R},\gamma|\mathbf{Y})$, which allows for model selection. Indeed, the change-point number in each sequence can be estimated by $K_j^{(t)}=\sum_{i=1}^{n} r_{j,i}^{(t)}$. Figure 3 shows the estimated posterior of K_j in each sequence (computed from the last 350 Markov chain samples) for the joint segmentation algorithm. The maximum values of these posteriors provide the MAP estimates of the changepoint numbers $\widehat{K}_1=3$ and $\widehat{K}_2=2$, which corresponds to the actual numbers of changes. Note again that there is a change at lag $n=120$ in all signals, by convention.

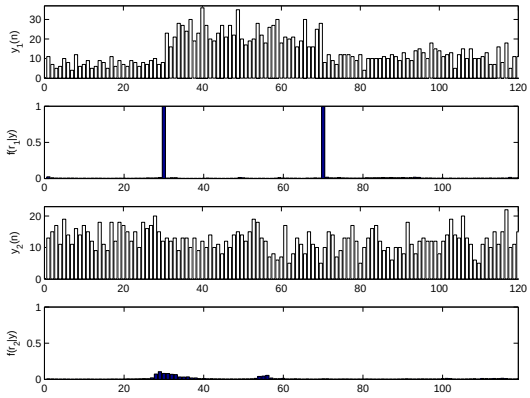

Figure 1: Posterior distribution of the change-point locations (1D segmentations).

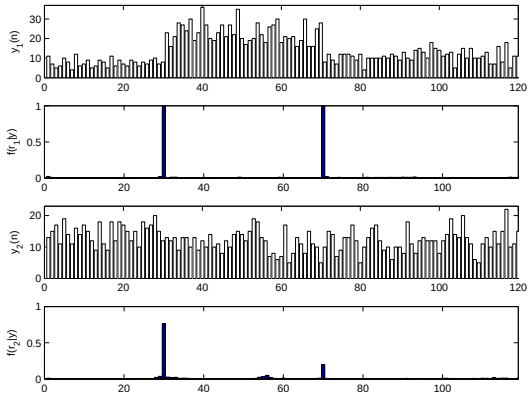

Figure 2: Posterior distribution of the change-point locations (2D segmentation).

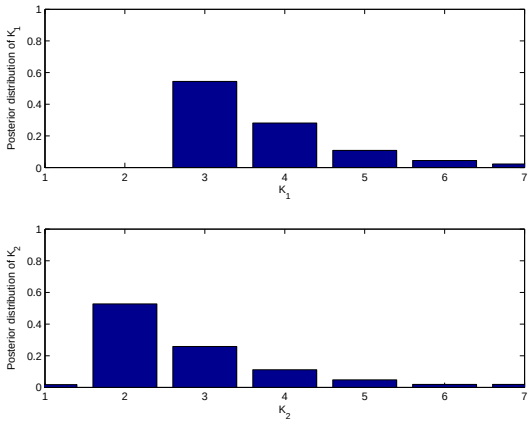

Figure 3: Posterior distribution of the number of change-points (2D segmentation).

5.3 Poisson Parameter Estimation

The estimation of the Poisson parameters is interesting since it allows for signal reconstruction. The posterior distributions of the parameters $\lambda_{2,k}$ conditionally upon $K_2=2$ are depicted on figure 4. They are clearly in good agreement with the actual values of the parameters $\lambda_2=[14,12]^\mathsf{T}$. Similar results could be obtained for Poisson parameters $\lambda_{1,k}$. They are omitted here for brevity.

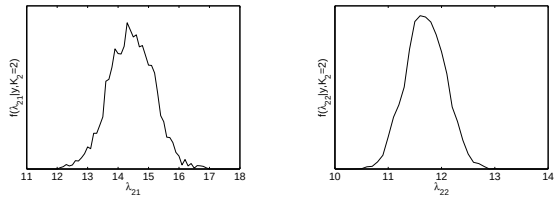

Figure 4: Posterior distribution of the Poisson parameters $\lambda_{2,i}$ (for $i = 1,2$) conditionally on $K_2 = 2$.

5.4 Hyperparameter Estimation

The last simulation results illustrate the performance of the method for the estimation of the hyperparameter vector $\mathbf{P}$. The posterior distributions of P_ε are depicted on figure 5. They are clearly in agreement with the actual posterior distributions given by the Dirichlet distribution (4).

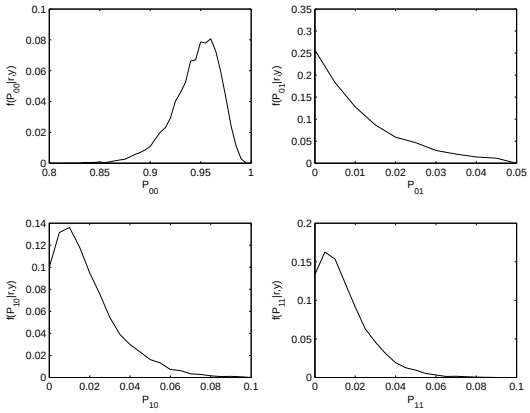

Figure 5: Posterior distribution of the hyperparameters $\mathrm{P}_{00}, \mathrm{P}_{01}, \mathrm{P}_{10}$ and P_{11}.

It is important to mention the following point: the number of burn-in iterations as well as the number of samples used for the estimation have been determined by using appropriate convergence diagnosis. The reader is invited to consult [10] for more details.

6. MARKOVIAN MODEL

The vectors $\mathbf{R}_i = [r_{1,i}, \ldots, r_{J,i}]^\mathsf{T}$ $(i = 1, \ldots, n)$ have been previously assumed to be independent. As a consequence, the algorithm can hesitate between close change-point values. Introducing constraints on the length of the segment could be an efficient means to limit this phenomenon. This section presents a standard discrete-time finite state Markovian model on $\mathbf{R}$ that rejects all segmentation models involving segments shorter than a minimal length L. For the sake of clarity, we have considered the simple case $L = 1$. However this procedure could be generalized to any value of L.

We propose a 2^J-state Markovian model on $\mathbf{R}$ with the following transition matrix:

$$\mathbf{P} = \begin{pmatrix} \mathrm{P}_{0\ldots0} & \mathrm{P}_{0\ldots1} & \cdots & \mathrm{P}_{1\ldots1} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}.$$

By denoting $\Pi(\mathbf{R}_i) = \prod_{\varepsilon \in \mathscr{E}} \mathrm{P}_\varepsilon^{\delta(\mathbf{R}_i - \varepsilon)}$ and $\Xi(\mathbf{R}_i) = \delta(\mathbf{R}_i - 0)$ (with $\delta(\mathbf{R}_i - \varepsilon) = 1$ if $\mathbf{R}_i = \varepsilon$, $\delta(\mathbf{R}_i - \varepsilon) = 0$ otherwise), it can be shown that:

$$f(\mathbf{R}_i | \mathbf{R}_{i-1}, \mathbf{P}) = \begin{cases} \Pi(\mathbf{R}_i) & \text{if } \Xi(\mathbf{R}_{i-1}) = 1, \\ \Xi(\mathbf{R}_i) & \text{otherwise.} \end{cases}$$

As a consequence, the prior for $\mathbf{R}$ can be written:

$$f(\mathbf{R}|\mathbf{P}) = f(\mathbf{R}_1) \prod_{i=2}^{n-1} \big[(\Pi(\mathbf{R}_i) - \Xi(\mathbf{R}_i)) \times \Xi(\mathbf{R}_{i-1}) + \Xi(\mathbf{R}_i) \big] f(\mathbf{R}_i).$$

The posterior distribution of P_ε is still a Dirichlet distribution:

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathscr{D}_{2^J}\left(\widetilde{S}_\varepsilon(\mathbf{R}) + \alpha_\varepsilon\right),$$

where $\widetilde{S}_\varepsilon(\mathbf{R})$ is the number of lags such as $\mathbf{R}_i = \varepsilon$ and $\mathbf{R}_{i-1} = \mathbf{0}$.

We illustrate the performance of the method by processing the synthetic data presented in the section 5. However, we have modified the first sequence by inserting a new segment between lags $i = 80$ and $i = 83$ with Poisson parameter $\lambda = 3$. We perform the proposed algorithm with $L = 4$. Fig.'s 6 and 7 show the posterior distributions of the change-locations obtained for the two time-series with the two approaches. The initial algorithm clearly detects successive model changes between $i = 80$ and $i = 83$ (fig. 6), contrary to the modified algorithm that includes a Markovian model for $\mathbf{R}$ (fig. 7).

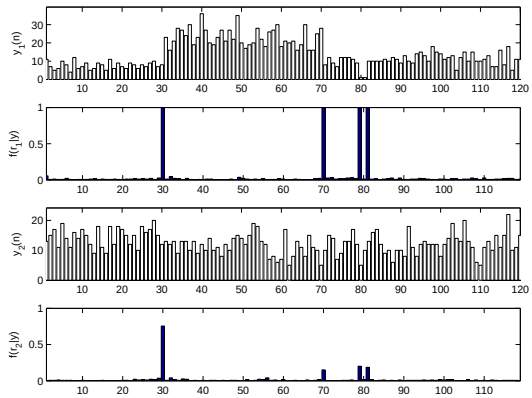

Figure 6: Posterior distribution of the change-point locations (initial model).

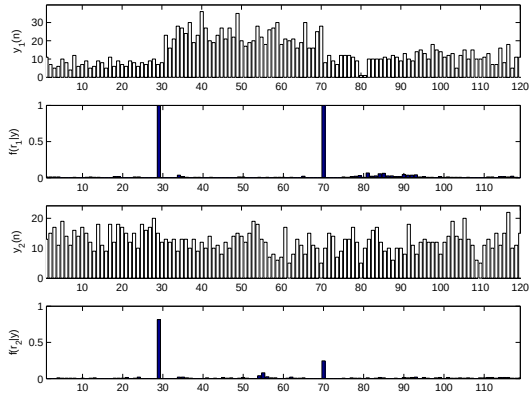

Figure 7: Posterior distribution of the change-point locations (Markovian model).

7. APPLICATION TO BATSE DATA

This section presents the analysis of a small sample of data obtained by the NASA Compton Gamma Ray Observatory's BATSE (Burst and Transient Source Experiment) [11]. By the very nature of this

photon-counting experiment, the time series can be accurately modeled as sequences of Poissonian data whose Poisson rate parameters vary as determined by the actual changes in brightness of the gamma-ray burst (GRB) source. The intensity of the GRB as a function of time often consists of a series of short-time-scale structures, called pulses. Of considerable interest is how the GRB variability depends on the energy of the radiation. In the data mode analyzed here, BATSE recorded the energies of the photons in four energy channels – analogous to four colors in ordinary visible radiation. The unit of energy is *keV* (thousand electron volts), and the nominal energy channels are: $25-60keV$, $60-110keV$, $110-325keV$, and $>325keV$. The variability curves at low and high energies are typically very similar, but there can be a delay or lag between them.

The hierarchical method presented in this paper has been applied to the BATSE data. The raw counting data (which consists of about 29000 photons recorded in four energy channels) have been transformed into binned data by counting the number of photons distributed in 256 time bins of width $3.68ms$. Note that $J=4$ in this example. The results have been averaged from 64 Markov chains with $N_{MC}=3500$ runs and $N_{bi}=200$ burn-in iterations. Figure 8 shows that the MAP estimates of parameters K_j are $\widehat{K_1}=5$, $\widehat{K_2}=7$, $\widehat{K_3}=11$ and $\widehat{K_4}=8$. The estimated posterior distribution of $\mathbf{R}$ can then be used to estimate the change locations in each channel. Here, in each sequence, segments are obtained from the $\widehat{K_j}$ largest values of the posteriors. In the last step of the analysis, the different Poisson intensities are then estimated by averaging the signal on each segment of each sequence (which corresponds to the intensity MMSE estimator conditioned to K_j). This procedure yields Bayesian blocks introduced in [6]. The resulting Bayesian blocks are shown in figure 9. Most results are in good agreement with those presented in [12]. However, the proposed joint approach makes it possible to find out changes that were not initially detected by the iterative method. For example, the second and third changepoints $l_{1,2}$ and $l_{1,3}$ in the first channel (respectively at $0.1294s$ and $0.2316s$) are detected by the joint approach and not by independent 1D segmentations. The presence of changes at the same position in the other channels explains this detection.

8. CONCLUSIONS

This paper studied a joint Bayesian segmentation procedure for Poissonian time series. The proposed Bayesian formalism allowed to define correlations between the change locations of these time series. A constraint ensuring a minimal segment length can be included in the algorithm by using a Markovian prior for the change locations. The segmenter performance was illustrated on several synthetic data sets. The application to real data recorded by the NASA Compton Gamma Ray Observatory's BATSE was finally investigated.

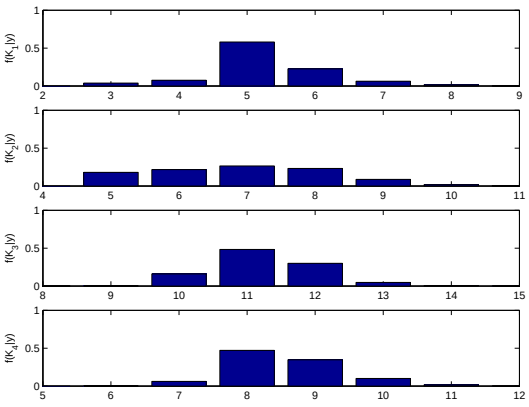

Figure 8: Posterior distribution of the change-point number (4D astronomical data).

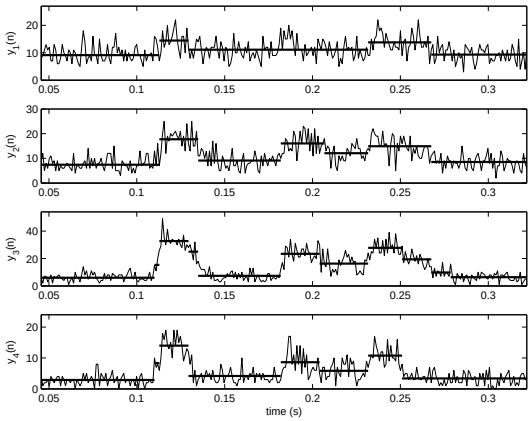

Figure 9: Block representation (4D astronomical data).

REFERENCES

[1] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs NJ: Prentice-Hall, 1993.

[2] B. Brodsky and B. Darkhovsky, *Nonparametric methods in change-point problems*. Boston (MA): Kluwer Academic Publishers, 1993.

[3] F. Gustafsson, *Adaptive filtering and change detection*. New-York: Wiley, 2000.

[4] M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1365–1373, 1998.

[5] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, pp. 747–758, March 2002.

[6] J. D. Scargle, "Studies in astronomical time series analysis: v. Bayesian blocks, a new method to analyze structure in photon counting data," *The Astrophysical Journal*, vol. 504, pp. 405–418, Sept. 1998.

[7] B. Jackson *et al.*, "An algorithm for optimal partitioning of data on an interval," *IEEE Signal Processing Lett.*, vol. 12, pp. 105–108, Feb. 2005.

[8] N. Dobigeon, J.-Y. Tourneret, and J. Scargle, "Change-point detection in astronomical data by using a hierarchical model and a bayesian sampling approach," in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, (Bordeaux, France), July 2005.

[9] R. McCulloch and R. Tsay, "Bayesian inference and prediction for mean and variance shifts in autoregressive time series," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 968–978, 1993.

[10] N. Dobigeon, J.-Y. Tourneret, and J. Scargle, "Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model," *submitted to IEEE Trans. Signal Processing*, 2005.

[11] W. S. Paciesas and *al.*, "The fourth BATSE burst revised catalog," *Astrophysical Journal Supplement Series*, vol. 122, pp. 465–497, June 1999. The data are available at `http://cossc.gsfc.nasa.gov/batse/`.

[12] J. D. Scargle, J. Norris, and B. Jackson, "Studies in astronomical time series analysis: vi. optimal segmentation : blocks, histograms and triggers," to be submitted.