Change-point Detection in Astronomical Data by using a Hierarchical Model and a Bayesian Sampling Approach

Nicolas Dobigeon[†] Jean-Yves Tourneret[†]
Jeffrey. D. Scargle[◊]

[†] IRIT/ENSEEIHT/TéSA
Toulouse, FRANCE

[◊] Space Science Division, NASA
Moffett Field, CA, USA

IEEE Workshop on Statistical Signal Processing, Bordeaux, 2005

Signal and Image Segmentation using Bayesian Inference

- **Biomedical**
  - M. Lavielle, "Optimal segmentation of random processes," *IEEE Trans. on Signal Processing*, 1998.
  - A. Gacek et al, "A genetic segmentation of ECG signals," IEEE Trans. on Biomedical Engineering, 2003
- **Image Segmentation**
  - R. Fjortoft et al, "An Optimum multiedge Detector for SAR image segmentation," *IEEE Trans. on Geosci. Remote Sensing*, 1998.
  - J-Y Tourneret et al, "Bayesian off-line detection of multiple change-points corrupted by multiplicative noise: application to SAR image edge detection," *Signal Processing*, 2003.
- **Astronomy**
  - 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*, 1998.
  - B. Jackson et al, "An algorithm for optimal partitioning of data on an interval," *IEEE Sig. Proc. Letters*, 2005.

Astronomical Data

1D Data

- The number of photons counted in successive equally spaced intervals (bins) is distributed according to a Poisson distribution.
- The Poisson rate parameter varies 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 consists of a series of pulses.

☞ *Determine rise and decay times of the pulses*

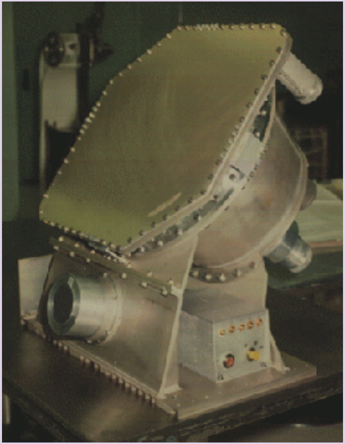
Multi-dimensional Data

- The energies are recorded in four energy channels: $25 - 60$keV, $60 - 110$keV, $110 - 325$keV and > 325keV by BATSE.

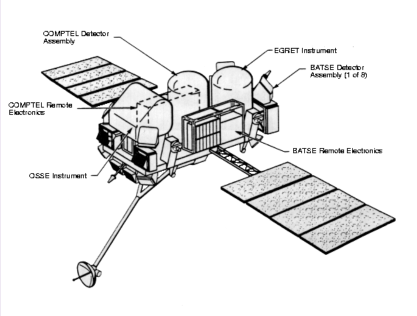
☞ *How GRB variability depends on the energy?*

Introduction

BATSE module
Burst And Transient Source Experiment



The Compton γ-Ray Observatory



Problem formulation

Modeling

Arrival time of photons are modeled by a discrete time Poisson counting process:

$$y_i \sim \mathcal{P}\left(\lambda_k\right)$$

where $k = 1, ..., K$, $i \in \mathcal{I}_k = [l_{k-1} + 1, l_k]$, and :

- $\mathcal{P}(\lambda)$ denotes a Poisson distribution with parameter λ,
- K is the number of segments in the observed signal,
- l_k is the sample point after which the kth change occurs in the signal (by convention $l_0 = 0$ and $l_K = n$ where n is the number of observed samples).

Problem

Estimation of (l_k, λ_k) from data $y = (y_i)_{i=1,...,n}$

Hierarchical Bayesian Model

A standard reparametrization

Indicators :

$$\begin{cases} r_i = 1 \text{ if there is a changepoint at lag } i, \\ r_i = 0 \text{ otherwise,} \end{cases}$$

The unknown parameter vector

$\theta = (r, \lambda) \in \Theta = \{0, 1\}^n \times \mathbb{R}^K$

- $r = (r_1, \ldots, r_n)$ with $r_n = 1$,
- $\lambda = (\lambda_1, \ldots, \lambda_K)$ with $K(r) = \sum_{i=1}^n r_i$,

Hierarchical Bayesian inference

Bayes' theorem :

$$f(\theta|y) \propto \int f(y|\theta) f(\theta|\phi) f(\phi) d\phi$$

Likelihood function

Likelihood function

$$f(y|\theta) = \prod_{k=1}^{K} \prod_{i \in I_k} \frac{\lambda_k^{y_i} \exp(-\lambda_k)}{y_i!}$$

$$= \frac{1}{\prod_{i=1}^{n} y_i!} \prod_{k=1}^{K} \lambda_k^{s_k(r)} \exp(-\lambda_k n_k(r))$$

$$\propto \prod_{k=1}^{K} \lambda_k^{s_k(r)} \exp(-\lambda_k n_k(r)),$$

- $s_k(r) = \sum_{i \in I_k} y_i$
- $n_k(r) = l_k - l_{k-1}$ (number of samples in the kth interval I_k)

Parameter priors

Indicator vector

- $\mathbf{P}(r_i = 0) = 1 - P$ and $\mathbf{P}(r_i = 1) = P$ (do not depend on i),
- the variables r_i (for $i = 1, ..., n$) are *a priori* independent,

The indicator prior distribution :

$$f(r|P) = \prod_{i=1}^{n-1} P^{r_i}(1-P)^{1-r_i},$$
$$= P^{\sum_{i=1}^{n-1} r_i}(1-P)^{n-1-\sum_{i=1}^{n-1} r_i},$$

Poisson parameters

Gamma distribution:

$$f(\lambda_k|\nu,\gamma) \sim \mathcal{G}(\nu,\gamma),$$

where $\nu = 1$ and γ is an adjustable hyperparameter

Hyperparameter priors

Hyperparameter γ

Noninformative Jeffreys' prior:

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

Hyperparameter P

Uniform distribution on $[0, 1]$:

$$f(P) = \mathbb{I}_{[0,1]}(P).$$

Hyperparameter vector $\phi = (\gamma, P)$

Assuming the independence of different hyperparameters:

$$f(\Phi) = \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)\mathbb{I}_{[0,1]}(P).$$

Hierarchical Bayesian Model

Posterior changepoint distribution

After integration respect to the nuisance parameters P and λ_k:

$$f(r,\gamma|y) \propto C(r|y)\frac{1}{\gamma}\left(\frac{\gamma^{\nu}}{\Gamma(\nu)}\right)^{K}\prod_{k=1}^{K}\frac{\Gamma\left(n_k\left(r\right)+\nu\right)}{\left(s_k\left(r\right)+\gamma\right)^{n_k(r)+\nu}},$$

with

$$C(r|y) = \Gamma\left(\sum_{i=1}^{n-1} r_i + 1\right)\Gamma\left(n - \sum_{i=1}^{n-1} r_i\right),$$

where $\Gamma(t)$ is the Gamma function.

A too complex posterior distribution...

☞ Simulation of samples distributed according to the $f(r,\gamma|y)$ by using MCMC methods.

Gibbs sampler for change-point detection

Generation of samples distributed according to $f(r|\gamma, y)$

- n-1 Bernoulli draws: $P(r_i = \epsilon | r_{-i}, \gamma, y) \propto f(r_i(\epsilon), \gamma | y)$.

Generation of samples distributed according to $f(\gamma | r, y)$

- Draw samples according to $f(\lambda|\gamma, r, y)$

$$\lambda_k | \gamma, r, y \sim \mathcal{G}\left(s_k\left(r\right) + \nu, n_k\left(r\right) + \gamma\right),$$

- Draw samples according to $f(\gamma|\lambda, r, y)$

$$\gamma | \lambda, r, y \sim \mathcal{G}\left(\nu K, \sum_{k=1}^{K} \lambda_k\right).$$

Updating P

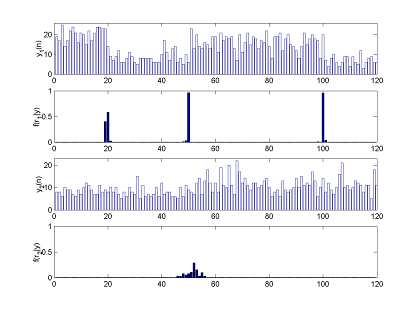
$$f(P|r, y) \propto P^{\sum_{i=1}^{n-1} r_i} (1-P)^{n-1-\sum_{i=1}^{n-1} r_i}.$$

Synthetic data

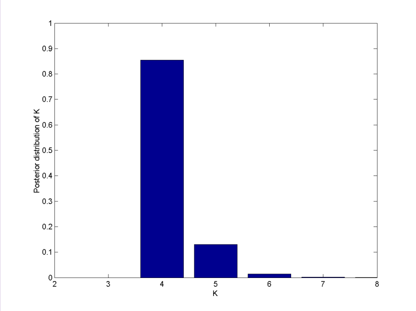
Simulation Parameters

- Signal parameters: $n = 120$, $K = 4$, $l = (20, 50, 100, 120)$, $\lambda = (19, 9, 17, 7)$
- Algorithm: 64 Markov-Chains, 200 burn-in iterations, 800 iterations to compute the estimates

Posterior distribution of r

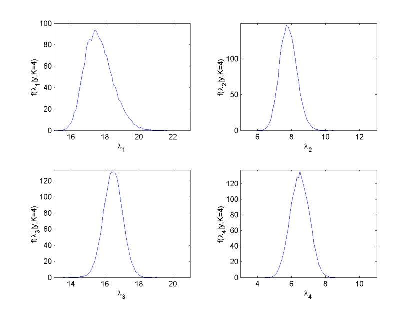


Posterior distribution of K

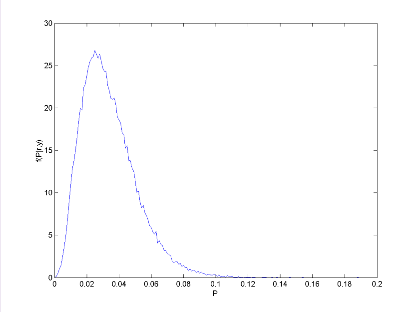


Synthetic data

Posterior distribution of λ



Posterior distribution of P



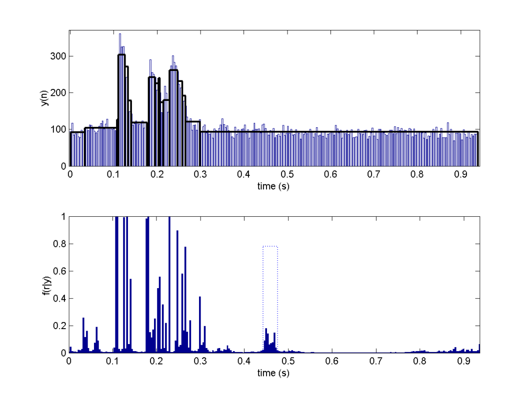
☞ *In good agreement with the theoretical results...*

Real Astronomical Data

Parameters of the simulation

Raw data: 29000 photons, 256 time bins of 3.68ms,
Algorithm: 64 Markov-Chains, 50 burn-in iterations, 1500 computation iterations.

Posterior distribution of changepoint locations



Généralisation multi-capteurs

Modélisation

La statistique des données observées dans les diverses bandes énergétiques peut être décrite comme suit :

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

où $j = 1, ..., J$, $k = 1, ..., K_j$, $i \in \mathcal{I}_{j,k} = [l_{j,k-1} + 1, l_{j,k}[$, et :

- J est le nombre de signaux à segmenter,
- K_j est le nombre de segments du $j^{\text{ième}}$ signal observé,
- $l_{j,k}$ correspond à la $k^{\text{ième}}$ rupture dans le $j^{\text{ième}}$ signal,
- $\mathcal{P}(\lambda)$ désigne une loi de Poisson de paramètre λ.

Problème

Estimation conjointe des $l_{j,k}$ à partir des données :

$$y = \left(y_{j,i}\right)_{j\in\{1,\dots,J\},i\in\{1,\dots,n\}}$$

Modèle Bayésien Hiérarchique

Un reparamètrage classique

Introduction d'indicatrices :

$$\begin{cases} r_{j,i} = 1 \text{ s'il y a une rupture à l'instant } i \text{ du } j^{\text{ième}} \text{ signal}, \\ r_{j,i} = 0 \text{ sinon}, \end{cases}$$

Le vecteur des paramètres inconnus

$\theta = (r, \lambda) \in \Theta = \{0,1\}^{nJ} \times \prod_{j=1}^{J} \mathbb{R}_{+}^{K_j}$, $K(r) = \sum_{i=1}^{n} r_i$ avec :

- $r = (r_1, \ldots, r_n)$ et $r_i = (r_{1,i}, \ldots, r_{J,i})^T$,
- $\lambda = (\lambda_1, \ldots, \lambda_J)$ et $\lambda_j = (\lambda_{j,1}, \ldots, \lambda_{j,K_j})^T$,

Inférence Bayésienne Hiérarchique

Théorème de Bayes :

$$f(\theta|y) \propto \int f(y|\theta) f(\theta|\phi) f(\phi) d\phi$$

Vraisemblance

Hypothèses

Les séquences $y_l = (y_{l,1}, ..., y_{l,n})$ et $y_m = (y_{m,1}, ..., y_{m,n})$ sont indépendantes pour $l \neq m$

Fonction de vraisemblance

Elle s'exprime comme :

$$f(y|\theta) \propto \prod_{j=1}^{J} \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(r)} \exp\left(-\lambda_{j,k} n_{j,k}(r)\right),$$

- $s_{j,k}(r) = \sum_{i \in \mathcal{I}_{j,k}} y_{j,i}$ (somme des points du $k^{\text{ième}}$ intervalle $\mathcal{I}_{j,k}$ du $j^{\text{ième}}$ signal)
- $n_{j,k}(r) = l_{j,k} - l_{j,k-1}$ (longueur du segment $\mathcal{I}_{j,k}$)

Lois a priori : Vecteur des indicatrices

Hypothèses

- Les probabilités $P(r_i = \epsilon)$ ne dépendent pas de i avec :
$$r_i = (r_{1,i}, \ldots, r_{J,i})^T, \qquad \epsilon = (\epsilon_1, \ldots, \epsilon_J)^T \in \{0,1\}^J,$$
- Les variables r_i (pour $i = 1, \ldots, n$) sont *a priori* indépendantes,

Loi a priori des indicatrices

Sous ces hypothèses :
$$f(r|P) = \prod_{\epsilon \in \{0,1\}^J} P_\epsilon^{S_\epsilon(r)},$$

- $P_\epsilon \in \{P_{0\ldots0}, \ldots, P_{1\ldots1}\}$,
- $S_\epsilon(r)$ est le nombre d'instants tels que $r_i = \epsilon$

Introduction d'une corrélation entre les séquences

- Grande valeur de $P_{0\ldots0} \Rightarrow$ absence de ruptures simultanées
- Grande valeur de $P_{1\ldots1} \Rightarrow$ présence de ruptures simultanées

Lois a priori : Paramètres de Poisson

Hypothèses (lois conjuguées)

- $f(\lambda_{j,k}|\nu,\gamma) \sim \mathcal{G}(\nu,\gamma)$, où :
  - $\nu = 1$,
  - γ est un hyperparamètre,
- les paramères $\lambda_{j,k}$ sont *a priori* indépendants,

Loi a priori des paramètres de Poisson

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

où $\mathbb{I}_{\mathbb{R}^+}$ est la fonction indicatrice sur $\mathbb{R}^+$.

Lois a priori des hyperparamètres

Hyperparamètre γ

Loi non-informative de Jeffrey : $f(\gamma) = \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)$.

Hyperparamètre $P = (P_\epsilon)_{\epsilon \in \{0,1\}^J}$

Loi de Dirichlet de vecteur-paramètre $(\alpha_{0\ldots0}, \ldots, \alpha_{1\ldots1})^T$ définie sur le simplexe $\mathfrak{P} = \left\{P; \sum_{\epsilon \in \{0,1\}^J} P_\epsilon = 1, P_\epsilon > 0\right\}$:

$$f(P|\alpha) \sim \mathcal{D}_{2^J}(\alpha).$$

Vecteur d'hyperparamètres $\phi = (\gamma, P)$

En supposant l'indépendance des différents hyperparamètres, :

$$f(\Phi|\alpha) = \left(\prod_{\epsilon \in \{0,1\}^J} P_\epsilon^{\alpha_\epsilon - 1}\right) \frac{1}{\gamma}\mathbb{I}_{\mathbb{R}^+}(\gamma)\mathbb{I}_{\mathfrak{P}}(P).$$

Modèle Bayésien Hiérarchique

Loi a posteriori des instants de ruptures

Après intégration par rapport aux paramètres de nuisance P et λ_k:

$$f(r, \gamma | y) \propto \frac{C(r|y)}{\gamma} \left(\frac{\gamma^{\nu}}{\Gamma(\nu)} \right)^K \prod_{k=1}^{K} \frac{\Gamma(n_k(r) + \nu)}{(s_k(r) + \gamma)^{n_k(r)+\nu}},$$

avec

$$C(r|y) = \Gamma\left(\sum_{i=1}^{n-1} r_i + 1\right) \Gamma\left(n - \sum_{i=1}^{n-1} r_i\right),$$

où $\Gamma(t)$ est la fonction Gamma.

Une loi a posteriori trop complexe...

... pour obtenir une expression simple des estimateurs bayésiens.

☞ Simulation d'échantillons asymptotiquement distribués suivant $f(r, \gamma | y)$ à l'aide d'une méthode MCMC.

Echantillonneur de Gibbs pour la détection de rupture

Génération suivant $f(r|\gamma, y)$

$$P(r_i = \epsilon | r_{-i}, \gamma, y) \propto f(r_i(\epsilon), \gamma | y),$$

Génération suivant $f(\gamma|r, y)$

- Tirage d'échantillons suivant $f(\lambda|\gamma, r, y)$

$$\lambda_{j,k}|\gamma, r, y \sim \mathcal{G}\left(s_{j,k}(r) + \nu, n_{j,k}(r) + \gamma\right),$$

- Tirage d'échantillons suivant $f(\gamma|\lambda, r, y)$

$$\gamma|\lambda, r, y \sim \mathcal{G}\left(\nu \sum_{j=1}^{J} K_j, \sum_{j=1}^{J} \sum_{k=1}^{K_j} \lambda_{j,k}\right).$$

Mise à jour de P

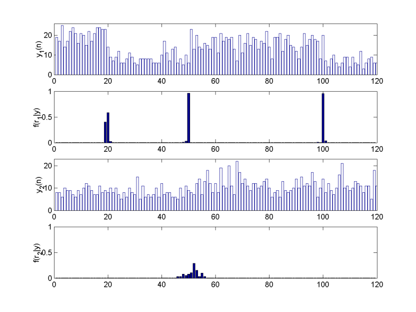
$$f(P|r, y, \alpha) \sim \mathcal{D}_{2^J}(\alpha_\epsilon + S_\epsilon(r)).$$

Simulations : données synthétiques

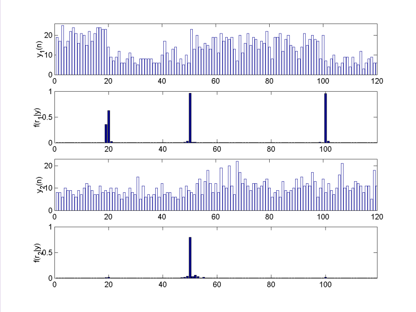
Paramètres

- Paramètres du signal : $J = 2$, $n = 120$, $K = 4$, $l_1 = (20, 50, 100, 120)$, $l_2 = (50, 120)$, $\lambda_1 = (19, 9, 16, 6)$, $\lambda_2 = (8, 11)$,
- Algorithme : 200 itérations de chauffage, 800 itérations utilisées pour les estimations.

Segmentation non conjointe

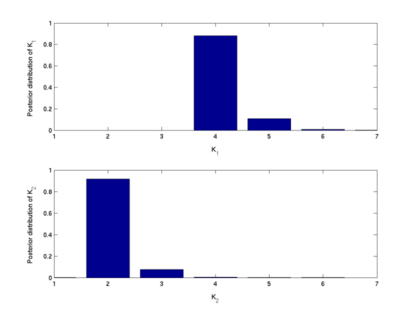


Segmentation conjointe

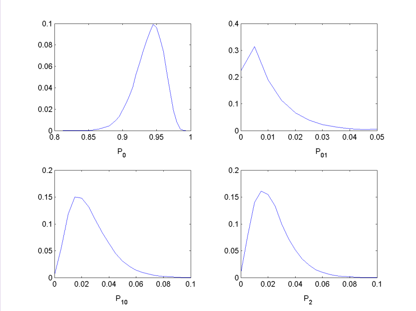


Simulations : données synthétiques

Nombre de changements



Loi a posteriori de P



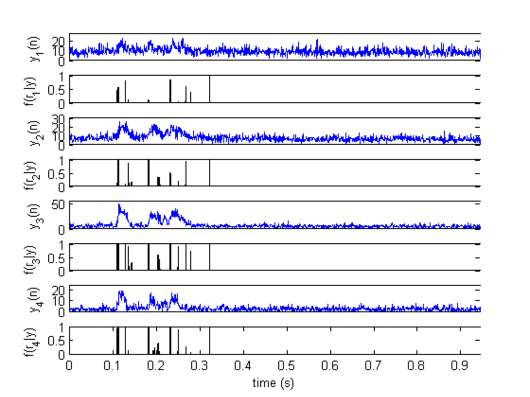
☞ *En accord avec les résultats théoriques...*

Simulations : données réelles

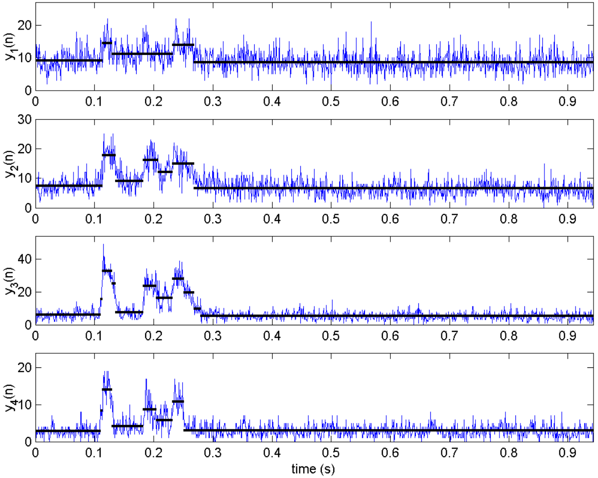
Paramètres de simulation

4 bandes d'énergie : $25 - 60$keV, $60 - 110$keV, $110 - 325$keV et > 325keV.
Algorithme : 200 itérations de chauffage, 3300 itérations d'intérêts.

Loi a posteriori des instants de ruptures



Reconstruction du signal



Conclusions

Travail réalisé

Une méthode de segmentation conjointe de données astrophysiques issues de multiples capteurs basée sur :

- le caractère poissonnien des données,
- l'introduction d'une corrélation *a priori* des instants de ruptures (approche conjointe),
- un modèle Bayésien Hiérarchique,
- l'utilisation d'un échantillonneur de Gibbs.

Perspectives

- Traitement de données astronomiques avec une autre mise en forme,
- Généralisation à d'autres types de signaux (vent, arc-tracking, bio-médical, ...).