Advances in Complex Systems

BAYESIAN IDENTIFICATION OF MULTIPLE CHANGE POINTS IN POISSON DATA

R. H. Loschi & F. R. B. Cruz*

Statistics Department, Federal University of Minas Gerais,
31270-901 - Belo Horizonte - MG, Brazil
{*loschi,fcruz*}*@est.ufmg.br*

Received (received date)
Revised (revised date)

The identification of multiple change point is a problem shared by many subject areas, including disease and criminality mapping, medical diagnosis, industrial control, and finance. An algorithm based on the Product Partition Model (PPM) is developed to solve the multiple change point identification problem in Poisson data sequences. In order to attack the PPM a simple and easy to implement Gibbs sampling scheme is derived. A sensitivity analysis is performed, for different prior specifications. The algorithm is then applied to the analysis of a real data sequence. The results show that the method is quite effective and provides useful inferences.

Keywords: Beta prior distribution, Student-t distribution, Yao's cohesions, Gibbs sampling.

1. Introduction

The identification of multiple change points is a problem encountered in many subject areas, ranging from criminality and disease mapping to finance and industrial control. Given a time series (or a data sequence), as the one seen in Figure 1, the problem is to know whether or not change points occurred in its level (or variance). Certainly, multiple change point identification is not a brand new problem. Indeed, many tools were already considered to tackle it, including Bayesian [1, 12, 14] and non-Bayesian approaches [9, 11, 21]. In particular, this paper is concerned about a Bayesian approach to the multiple change point identification problem in Poisson data sequences, based on the Product Partition Model (PPM).

The PPM was introduced originally by Hartigan [8] and it may be seen as a generalization of several previously developed models such as the model by Smith [20], Menzefricke [16], or Hsu [10], for instance. Advantages of the PPM over other popular methodologies, such as the threshold model of Chen & Lee [3], include

*Corresponding author. Prof. Frederico Cruz. E-mail: fcruz@est.ufmg.br. Phone: (+55 31) 3499 5929. Fax: (+55 31) 3499 5924.

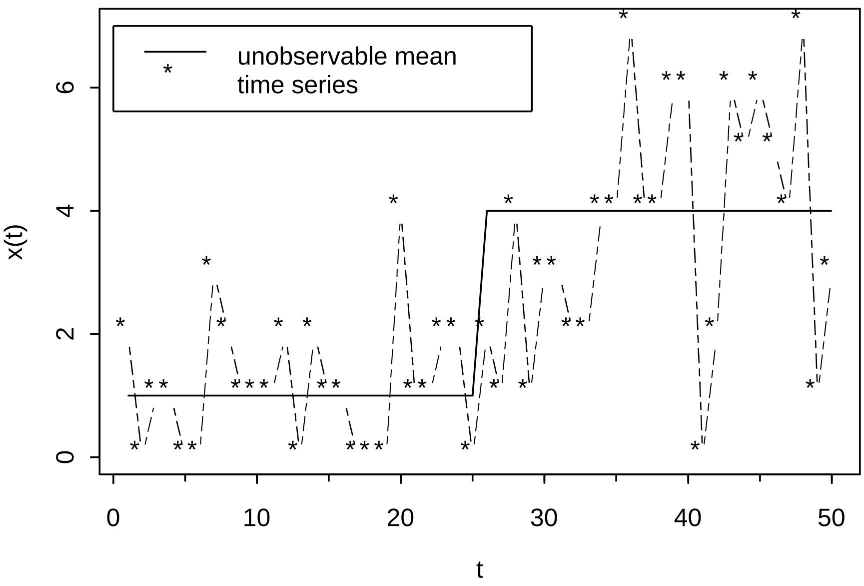


Fig. 1. Poisson data sequence with one change point.

the flexible number of change points in the sequence, which is random instead of a predefined number. Later, the PPM was applied to the identification of multiple change points in normal means by Barry & Hartigan [1] and Crowley [4]. Afterwards Loschi & Cruz [12] extended the PPM and applied the method to identify multiple change points both in the means and variances of normal data, developing a Gibbs sampling scheme to compute new important measures besides the product estimates, including the posterior distributions for the number of change points and for the instants when changes occurred [14], and admitting a prior specification for the probability of having a change, p [13].

The aim of this paper is twofold. First, we derive an original version of the PPM, suitable for identifying multiple change points in the means of Poisson data sets, θ. A gamma prior distribution was assumed for the parameter θ and a beta prior distribution was assumed for the probability of having a change, p. The algorithm developed provides (i) the product estimates for θ, (ii) the posterior distributions for the number of change points, $(B-1)$, and for p, and (iii) the posterior probabilities of each instant to be a change point. Second, a sensitivity analysis for the above estimates is presented, for different prior specifications for p.

The paper is organized as follows. Section 2 reviews the parametric approach for the PPM, presents inferential solutions to identify change points for Poisson random variables, and details a Gibbs sampling scheme to implement the PPM. In Section 3, some computational results are presented and discussed. Section 4 presents the case study. Section 5 concludes the paper with final remarks and topics for further investigation in the area.

2. Product Partition Model

The PPM is a Bayesian model. For the interest reader, details on Bayesian statistics can be found easily in the literature (see, for instance, the book by Migon & Gamerman [17]).

In the parametric approach of the PPM it is considered that the sequence of random variables $X_1, \ldots, X_n$ has marginal densities $f_1(X_1|\theta_1), \ldots, f_n(X_n|\theta_n)$, conditional on $\theta_1, \ldots, \theta_n$. It is assumed that given a partition $\rho = \{i_0, \ldots, i_b\}$ of the set $I \cup \{0\}$, for $I = \{1, \ldots, n\}$ and $b \in I$, such that $0 = i_0 < i_1 < \cdots < i_b = n$, one has that $\theta_i = \theta_{[i_{(r-1)}i_r]}$ for every $i_{r-1} < i \leq i_r$, for $r = 1, \ldots, b$, and that $\theta_{[i_0 i_1]}, \ldots, \theta_{[i_{(b-1)}i_b]}$ are independent, with $\theta_{[ij]}$ having (block) prior density $\pi_{[ij]}(\theta)$, in which $\theta \in \Theta_{[ij]}$, and $\Theta_{[ij]}$ is the parameter space that corresponds to the common parameter, say, $\theta_{[ij]} = \theta_{i+1} = \cdots = \theta_j$, which indexes the conditional density of $\mathbf{X}_{[ij]} = (X_{i+1}, \cdots, X_j)'$. Denote by $c_{[ij]}$, $i, j \in I \cup \{0\}$, $i < j$, the prior cohesion associated with the block $[ij]$, which can be seen as the transition probabilities in the Markov chain defined by the endpoints of the blocks in ρ. That is, the prior cohesion $c_{[ij]}$ denotes the probability of having a change in j, given that a change took place in i. Thus, $(X_1, \ldots, X_n, \rho)$ follows the PPM if

(i) the prior distribution of ρ is the following product distribution

$$P(\rho = \{i_0, \ldots, i_b\}) = \frac{\prod_{j=1}^{b} c_{[i_{(j-1)}i_j]}}{\sum_{\mathcal{C}} \prod_{j=1}^{l} c_{[i_{(j-1)}i_j]}}, \tag{1}$$

in which $\mathcal{C}$ is the set of all possible partitions of the set I into l contiguous blocks with endpoints $i_1, \ldots, i_l$, satisfying the condition $0 = i_0 < i_1 < \cdots < i_l = n$, for all $l \in I$;

(ii) conditional on $\rho = \{i_0, \ldots, i_b\}$, the sequence $X_1, \ldots, X_n$ has the joint density

$$f(X_1, \ldots, X_n|\rho = \{i_0, \ldots, i_b\}) = \prod_{j=1}^{b} \int_{\Theta_{[i_{j-1}i_j]}} f(\mathbf{X}_{[i_{j-1}i_j]}|\theta)\pi_{[i_{j-1}i_j]}(\theta)d\theta,$$

in which $\Theta_{[ij]}$ denotes the parametric space of $\theta_{[ij]}$.

In the PPM, the posterior expectations of θ_k (also called the product estimates) are given by

$$E(\theta_k|X_1, \ldots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} \left(r^*_{[ij]} E(\theta_k|\mathbf{X}_{[ij]}) \right), \ k = 1, \ldots, n, \tag{2}$$

in which $r^*_{[ij]}$ is the posterior relevance for the block $[ij]$, given by

$$r^*_{[ij]} = \frac{\lambda_{[0i]} c^*_{[ij]} \lambda_{[jn]}}{\lambda_{[0n]}}, \tag{3}$$

with $c^*_{[ij]} = c_{[ij]} f_{[ij]}(\mathbf{X}_{[ij]})$ and $\lambda_{[ij]} = \sum \Pi_{k=1}^{l} c^*_{[i_{(k-1)} i_k]}$, and the summation is over all partitions of $\{i+1, \ldots, j\}$ into l blocks with endpoints $i_0, i_1, \ldots, i_l$, satisfying the condition $i = i_0 < i_1 < \cdots < i_l = j$.

Another parameter of interest is the number of blocks B in ρ (or the number of change points, $B-1$). Let $\mathbf{X}_{[0n]}$ be the vector $(X_1, \ldots, X_n)$. In the PPM, the posterior distribution of B is given by

$$P(B = b|\mathbf{X}_{[0n]}) \propto \sum_{\mathcal{C}_b} \Pi_{j=1}^{b} c^*_{[i_{(j-1)} i_j]}.$$

in which $\mathcal{C}_b \subseteq \mathcal{C}$ is the set of all partitions of I into b contiguous blocks.

The posterior distribution of ρ has the same form as its prior distribution and it is obtained by using the posterior cohesions $c^*_{[ij]}$ in Eq. (1). However, since each value of ρ usually receives low mass, the posterior distribution of ρ does not provide a good idea about when changes occurred, as much as the posterior probability for each instant to be a change point would do. Thus, to obtain the posterior probability of each observed data time point to be a change point, consider $\mathcal{C}_k$ a subset of $\mathcal{C}$, which contains all partitions that include the kth instant as a change point. That is, each partition in $\mathcal{C}_k$ assumes the form $\{i_0, \ldots, i_{l-1}, i_l = k, i_{l+1}, \ldots, i_b\}$, for any $l \in I$. Let us denote by A_k the event that the kth instant is a change point, for $k = 2, \ldots, n$. Thus [13]

$$\begin{aligned} P(A_k|\mathbf{X}_{[0n]}) &= \sum_{\mathcal{C}_k} P(\rho = \{i_0, \ldots, i_{l-1}, i_l = k-1, i_{l+1}, \ldots, i_b\}|\mathbf{X}_{[0n]}) \\ &\propto \sum_{\mathcal{C}_k} \Pi_{j=1}^{l-1} c^*_{[i_{(j-1)} i_j]} c^*_{[i_{(l-1)}(k-1)]} c^*_{[(k-1) i_{(l+1)}]} \Pi_{j=l+1}^{b} c^*_{[i_{(j-1)} i_j]}. \end{aligned}$$

2.1. *Poisson Case*

For the Poisson case, given $\theta_1, \ldots, \theta_n$, it is assumed that $X_1, \ldots, X_n$ are independent and that $X_k|\theta_k, \sim \mathcal{P}(\theta_k)$, for $k = 1, \ldots, n$. It is also assumed that the common parameter $\theta_{[ij]}$, related to the block $[ij]$, has the conjugate gamma prior distribution denoted by $\theta_{[ij]} \sim \mathcal{G}(\tau_{1[ij]} + 1, \tau_{0[ij]})$, with density function given by

$$f(\theta_{[ij]}|\tau_{0[ij]}, \tau_{1[ij]}) = \frac{\Big(\tau_{0[ij]}\Big)^{\tau_{1[ij]}+1}}{\Gamma(\tau_{1[ij]} + 1)} \Big(\theta_{[ij]}\Big)^{\tau_{1[ij]}} \exp(-\tau_{0[ij]}\theta_{[ij]}).$$

in which $\tau_{0[ij]} > 0$ and $\tau_{1[ij]} > -1$.

Consequently, the random vector $\mathbf{X}_{[ij]}$ follows a distribution with density function given by

$$f(\mathbf{X}_{[ij]}) = \prod_{k=i+1}^{j} \frac{1}{X_k!} \frac{\Gamma(\tau^*_{1[ij]})}{\Gamma(\tau_{1[ij]}+1)} \left(\frac{\tau_{0[ij]}}{\tau^*_{0[ij]}}\right)^{\tau_{1[ij]}+1} \left(\frac{1}{\tau^*_{0[ij]}}\right)^{\sum_{k=i+1}^{j} X_k}, \qquad (4)$$

in which

$$\begin{cases} \tau^*_{0[ij]} = \tau_{0[ij]} + j - i, \\ \tau^*_{1[ij]} = \tau_{1[ij]} + \sum_{k=i+1}^{j} X_k + 1, \end{cases}$$

for all $i = 0, \ldots, n-1$, and $j = i+1, \ldots, n$.

Given $X_{[ij]}$, the conditional distribution of $\theta_{[ij]}$ is a gamma distribution with parameters $\tau^*_{0[ij]}$ and $\tau^*_{1[ij]}$, that is

$$\theta_{[ij]} | X_{[ij]} \sim \mathcal{G}\Big(\tau^*_{1[ij]}, \tau^*_{0[ij]}\Big).$$

Consequently, the blocks estimates are given by

$$\hat{\theta}_{[ij]} = E(\theta_{[ij]} | X_{[ij]}) = \frac{\tau^*_{1[ij]}}{\tau^*_{0[ij]}}, \qquad (7)$$

and, from Eq. (2) and Eq. (7), it follows that the product estimates for θ_k are

$$\hat{\theta}_k = E(\theta_k | X_1, \ldots, X_n) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} r^*_{[ij]} \hat{\theta}_{[ij]}, \; k = 1, \ldots, n.$$

The posterior relevancies $r^*_{[ij]}$ can be obtained from Eq. (3), taking into consideration the density given in Eq. (4).

Remark: *Notice that in this model we only admit simultaneous changes in the means and variances. Suppose, now, that only a shift in the mean is presented in the data. A possible way to treat the one-change-point problem is to assume, for example, that*

$$\begin{aligned} Y_i &\sim \mathcal{P}(\lambda), \; \textit{for } i = 1, \ldots k; \\ Y_i &= Z + \mu, \; \textit{for } i = k+1, \ldots n; \end{aligned}$$

with $Z \sim \mathcal{P}(\lambda)$. Notice that, in this case, the variance does not change at the instant k but there is a change in the mean, from λ to $\lambda + \mu$. Moreover, after the instant k the distribution of the Y_i, given λ and μ, is

$$f(y) = \frac{\exp\{-\lambda\} \lambda^{y-\mu}}{(y-\mu)!},$$

which clearly is not the regular Poisson distribution. However, the PPM still can be used to identify k because the method also admits changes in the distribution.

2.2. *A Gibbs Sampling Scheme Applied to the PPM*

An extraordinary array of problems in Bayesian inference has been solved by Markov chain Monte Carlo (MCMC) methods since the seminal paper by Gelfand & Smith [6] illustrated how easily a variety of intractable problems could be approximately solved. Such ease of use led to an explosion of research on complex Bayesian models without analytical solution, which could be now treated by the MCMC methods. Recent research results and overviews of the research in this area includes the papers by Besag et al. [2] and Robert [19], to cite just a few. The purpose here is to use the Gibbs sampling [7] as a posterior distribution generation scheme.

In this paper, Yao's [22] prior cohesions will be considered. Thus, let p be the probability of a change to occur at any instant in the sequence. Therefore, the prior cohesion for the block $[ij]$ is given by

$$c_{[ij]} = \begin{cases} p(1-p)^{j-i-1}, \text{ if } j < n, \\ (1-p)^{j-i-1}, \ \text{ if } j = n, \end{cases} \tag{8}$$

for all $i, j \in I$, and $i < j$. Notice that $c_{[ij]}$ corresponds to the probability of a new change to take place after $j-i$ instants, given that a change took place at the instant i.

Remark: *Notice that the prior cohesions are subjective choices and should disclose the similarity among the observations into the same block. For instance, considering $c_{[ij]} = 1$, for all $i, j \in I$, $i < j$, we are admitting a discrete uniform distribution for ρ. As another possibility, Quintana & Iglesias [18] had elicted a prior cohesion depending on the number of observation in the block. By their side, the Yao's cohesions shown in (8) are appropriate whenever it is reasonable to assume that the past change points are noninformative about the future change points and that each instant has the probability p of being a change point. Large values for p should be assigned if it is believed that many change points will take place in the sequence. Finally, if it is believed that different instants have different probabilities of being a change point, the Yao's cohesions become $c_{[ij]} = p_j \Pi_{l=i+1}^{j-1} p_l$, where p_l denotes the probability of the instant l being a change point. Notice that in the latter the calculations involved in the PPM become considerably more complex.*

Supposing that p has the prior distribution $\pi(p)$ and assuming that, given ρ, $\theta_k \in [ij]$, for $k = 1, \ldots, n$, for $i, j \in I$, and $i < j$, we have that the full conditional distributions of p, ρ, and θ_k are given, respectively, by

$$\begin{aligned} \pi(p|\rho, \theta, \mathbf{X}_{[0n]}) &\propto p^{b-1}(1-p)^{n-b}\pi(p), \\ \pi(\rho|p, \theta, \mathbf{X}_{[0n]}) &\propto \left(\Pi_{j=1}^{b} f_{[i_{j-1}i_j]}(X_{[i_{j-1}i_j]})\right) p^{b-1}(1-p)^{n-b}, \\ \pi(\theta_k|\rho, p, \theta_{-k}, \mathbf{X}_{[0n]}) &\propto (\theta_{[ij]})^{\tau_{1[ij]^*}} \exp(-\tau^*_{0[ij]}\theta_{[ij]}), \ k = 1, \ldots, n, \end{aligned}$$

in which $\mathbf{X}_{[0n]} = (X_1, \ldots, X_n)$, $\theta = (\theta_1, \ldots, \theta_n)$, θ_{-k} denotes the vector $(\theta_1, \ldots, \theta_{k-1}, \theta_{k+1}, \ldots, \theta_n)$ and $f(X_{[ij]})$ is given in Eq. (4).

Notice that it is not easy to sample directly from the full conditional distribution of ρ in the Poisson case. In order to estimate the posterior relevance of each block

$[ij]$, the posterior distribution of B, and the posterior distribution of ρ, we will use the auxiliary random quantity U_i, which reflects whether or not a change point occurred at the time i, that is

$$U_i = \begin{cases} 0, \text{ if } \theta_i \neq \theta_{i+1}, \\ 1, \text{ if } \theta_i = \theta_{i+1}, \end{cases}$$

for $i = 1, \ldots, n-1$. Notice that given any particular vector $(U_1, \ldots, U_{n-1})$, the corresponding ρ is immediately identified.

In order to generate the vectors $\mathbf{U}^k$'s considering a beta prior distribution for the probability p of change, denoted by $p \sim \mathcal{B}(\alpha, \beta)$, it is sufficient to consider the ratio given by the expression

$$R_r = \frac{f_{[xy]}(X_{[xy]})}{f_{[xr]}(X_{[xr]})f_{[ry]}(X_{[ry]})} \frac{\Gamma(n+\beta-b+1)\Gamma(b+\alpha-2)}{\Gamma(b+\alpha-1)\Gamma(n+\beta-b)},$$

in which x denotes the last change point before r and y denotes the next change point following r. The rth element at the kth step, U_r^k, is generated from the conditional distribution

$$U_r^k | U_1^k, \ldots, U_{r-1}^k, U_{r+1}^{k-1}, \ldots, U_{n-1}^{k-1};\ X_1, \ldots, X_n; p, \theta,$$

for $r = 1, \ldots, n-1$, starting from an initial vector $\mathbf{U}^0 = (U_1^0, \ldots, U_{n-1}^0)$. Notice that in the Poisson case $f_{[ij]}(X_{[ij]})$ is the distribution given in Eq. (4).

Consequently, the criterion for choosing the values U_r^k becomes

$$U_r^k = \begin{cases} 1, \text{ if } R_r \geq (1-u)/u, \\ 0, \text{ otherwise}, \end{cases}$$

in which u is a random variable uniformly distributed, $u \sim \mathcal{U}(0, 1)$.

Notice that the posterior relevance of the blocks used in Eq. (2) to estimate θ_k can be obtained by

$$\hat{r}^*_{[ij]} = \frac{M}{T},$$

in which M is the number of vectors $\mathbf{U}^k$'s for which it is observed that $U_i^k = 0$, $U_{i+1}^k = \ldots = U_{j-1}^k = 1$, and $U_j^k = 0$, and T is the total number of vectors generated in the Gibbs sampling scheme.

As mentioned earlier, the corresponding random quantity ρ is immediately identified from the vector $\mathbf{U}^k$. Consequently, one can estimate the posterior probability for each particular partition $\rho = \{i_0, i_1, \ldots, i_b\}$ into b contiguous blocks. Also notice that it is possible to estimate the number of blocks in ρ by

$$B = 1 + \sum_{i=1}^{n-1}(1 - U_i),$$

algorithm
  read $X_1, \ldots, X_n$
  read *all prior specifications* $\tau_{0[ij]}, \tau_{1[ij]}, \alpha, \beta$
  for $k = 1$ **to** `SAMPLES` **do**
    generate $\mathbf{U}^k$
    $B_k = 1 + \sum_{i=1}^{n-1}(1 - U_i^k)$
    $p_k \sim \mathcal{B}(\alpha + B_k - 1, n + \beta - B_k)$
  end for
  for all $i, j \in \{0, \ldots, n\}$ **such that** $i < j$ **do**
    $r^*_{[ij]} \leftarrow$ proportion of samples such that
      $U_i^k = 0, U_{i+1}^k = \cdots = U_{j-1}^k = 1, U_j^k = 0$
    $\tau^*_{0[ij]} \leftarrow \tau_{0[ij]} + j - i$
    $$\tau^*_{1[ij]} \leftarrow \tau_{1[ij]} + \sum_{k=i+1}^{j} X_k + 1$$
    $\hat{\theta}_{[ij]} \leftarrow \tau_{1[ij]}/\tau_{0[ij]}$
  end for
  for $k = 1$ **to** n **do**
    $$E(\theta_k|\ X_1, \ldots, X_n) \leftarrow \sum_{i=0}^{k-1} \sum_{j=k}^{n} r^*_{[ij]} \hat{\theta}_{[ij]}$$
    compute $P(A_k)$
  end for
  write $B_k, p_k, E(\theta_k), P(A_k)$
end algorithm

Fig. 2. PPM Gibbs sampling algorithm.

and to estimate the posterior distribution of B (or the posterior distribution of the number of change points, $B - 1$) by

$$P(B = b|\mathbf{X}_{[0n]}) = \frac{\sum_{s=1}^{T} \mathbf{1}\{B_s = b\}}{T},$$

in which $\mathbf{1}\{D\}$ is the indicator function of the event D. Additionally, for a beta prior distribution for the probability p of change, each sample of the posterior distribution of p may be generated from the following beta distribution

$$p_k|\mathbf{X}_{[0n]} \sim \mathcal{B}(\alpha + B_k - 1, n + \beta - B_k),$$

in which B_k is the number of blocks in the kth vector $\mathbf{U}^k$. Similarly, estimates of the posterior probability of each instant k to be a change point are

$$P(A_i|\mathbf{X}_{[0n]}) = \frac{N}{T},\ i = 2, \ldots, n,$$

in which N is the number of vectors for which it is observed that $U_{i-1}^{k} = 0$. Figure 2 shows the complete algorithm in pseudo-code.

3. Computational Experiments

Because of its computationally intensive nature, the algorithm presented in Figure 2 was coded in C++. The code is available upon request. All tests were performed in a PC, Pentium processor 400 MHz, 256 MB RAM, taking less than one minute of CPU time. In order to estimate the posterior relevancies $r^{*}_{[ij]}$, the posterior distribution of B (or the number of change points, $B-1$), and the posterior distribution of p, 4,600 samples of 0–1 values were generated with the dimension of the time series, starting from a sequence of zeros. The initial 100 iterations were discarded for *burn-in* and a lag of one was selected to get stable results independent on the starting point (discussion about the number of iterations to be discarded and the lag to be taken, can be found easily in the literature [5]).

3.1. *Prior Specifications and the Data set*

In order to verify the accuracy of the approach, computational experiments were conducted with the simulated data sequence shown in Figure 1. The observations were assumed to be conditionally independent and distributed according to the Poisson distribution with rate θ. Additionally, one change point occurred at the 26th observation, such that

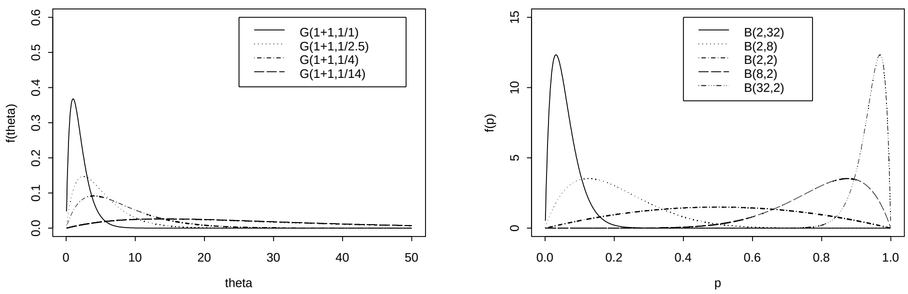
$$X_i|\theta_i \sim \mathcal{P}(1.0),\ i = 1,\dots,25,$$
$$X_i|\theta_i \sim \mathcal{P}(4.0),\ i = 26,\dots,50.$$

For the analysis, the natural conjugate prior distribution was considered for the parameters $\theta_{[ij]}$, which is in this case a gamma distribution. This assumption is not too restrictive, since the gamma distribution is rich enough to describe the uncertainty about the parameters under many practical circumstances, as seen in Figure 3-a. Three different prior distributions for $\theta_{[ij]}$, shown in Table 1 (see also Figure 3-a), were considered. Notice that if $\theta_{[ij]} \sim \mathcal{G}(2,1)$, the prior estimates for the rate is 2.0, for the squared error loss penalty function.

Table 1. Parameters of the prior distributions for $\theta_{[ij]}$.

Prior	τ_1	τ_0	Mode	Mean	Variance
$\theta_{[ij]} \sim \mathcal{G}(2,1)$	**1.0**	**1.0**	1.0	**2.0**	2.0
$\theta_{[ij]} \sim \mathcal{G}(2,1/2.5)$	1.0	1/2.5	2.5	5.0	12.5
$\theta_{[ij]} \sim \mathcal{G}(2,1/4)$	1.0	1/4.0	4.0	8.0	32.0

The truncated geometric distribution with parameter p was considered as prior cohesions because it was assumed that the past change points were non-informative



a) $\mathcal{G}(\tau_1 + 1, \tau_0)$ b) $\mathcal{B}(\alpha, \beta)$

Fig. 3. Probability densities.

about the future change points. One last decision that had to be made concerned the probability p of having a changing point. Thus, it was assumed that $p \sim \mathcal{B}(2, 8)$, plotted in Figure 3-b. That is, a small number of changes was expected in the data sequence. Notice that for the squared error loss penalty function, the prior estimate for the probability p of a change is 0.2 and the variance is 0.0145. Other similar settings for $\mathcal{B}(\alpha, \beta)$ were considered but the results (not shown) did not differ significantly.

3.2. *Numerical Results*

For the sake of conciseness, only results for the simulated data sequence presented in Figure 1 are shown. Additional simulations were carried out with similar simulated series but the results (not shown) did not differ significantly. The main advantage of this analysis is that we could can control for errors in the method since the actual (unobservable) means were known.

Figure 4 presents the posterior estimates (the product estimates) for the rate, θ, for the three prior specifications shown in Table 1. The estimates are also contrasted with the real values of θ at each instant. In spite of the fact that the prior estimates for θ were very different among themselves, the posterior estimates were very similar and close to the real values of θ, mainly before the change and after the 35th observation. It is also noticeable that the product estimates do capture the change in θ. However, the PPM was not able to do it immediately right after the 26th observation but 10 observations ahead. On this matter, we could notice that the length of the time series plays a key role. Longer were the time series, sooner the changes were identified.

Additional information available through the method include Figure 5, which presents the posterior distribution for the number of blocks in ρ, B (or for the number of change points, $B - 1$). As expected, the posterior distribution of the

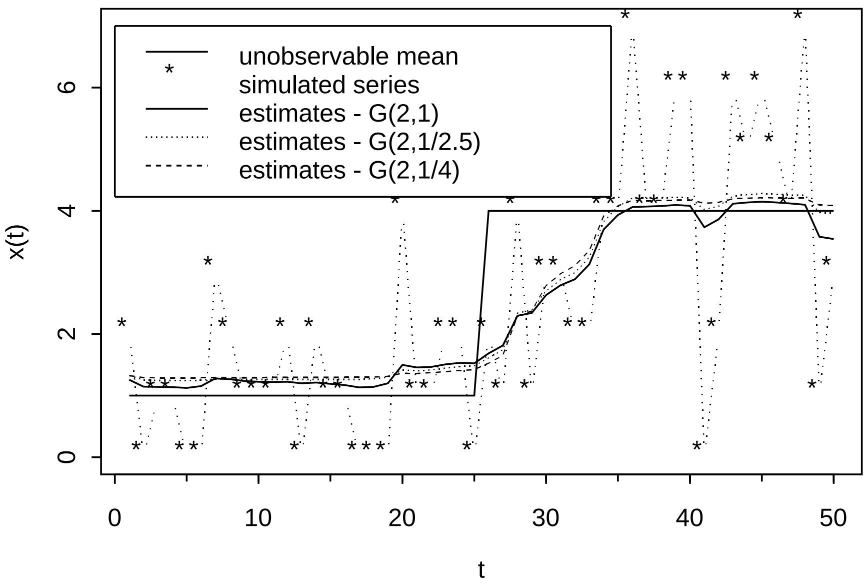


Fig. 4. Product estimates for θ.

number of blocks concentrates most of its mass in small values. For $p \sim \mathcal{B}(2, 8)$, it can be shown that the prior expected number of change points is 8.8 and variance is 49.9. From Tables 1 and 2, it can be noticed that all posterior estimates were more precise, that is, for all prior specifications, the posterior variances were reduced. It can also be observed that surprisingly the best estimate for $B - 1$ was obtained for $\mathcal{G}(2, 1/4)$, that is, when the least informative prior distribution was considered. Finally, it is noticeable from Figure 5 that, for $\mathcal{G}(2, 1/4)$ and $\mathcal{G}(2, 1/2.5)$, the most probable number of blocks was 2, actually the real value, with probability of $\approx 32\%$.

Table 2. Descriptive statistics for the posterior distributions of B.

Prior	Mean	StDev	Q1	Median	Q3	Min	Max
$\theta_{[ij]} \sim \mathcal{G}(2, 1)$	5.28	3.28	3	4	6.25	2	23
$\theta_{[ij]} \sim \mathcal{G}(2, 1/2.5)$	3.49	1.70	2	3	4	2	18
$\theta_{[ij]} \sim \mathcal{G}(2, 1/4)$	2.69	1.01	2	2	3	2	12

Figure 6 shows the prior and posterior distribution for the probability p of having a change at any instant. In the prior evaluation, the value of p was estimated in 0.2

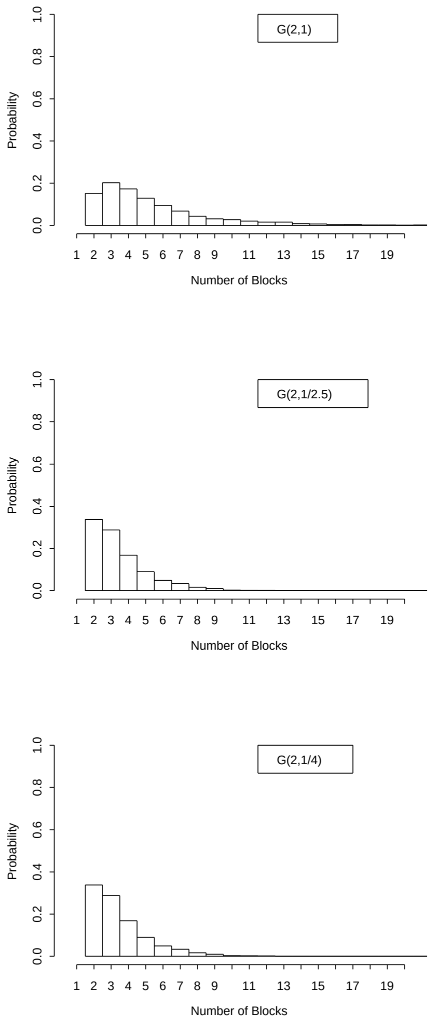


Fig. 5. Posterior distribution of B.

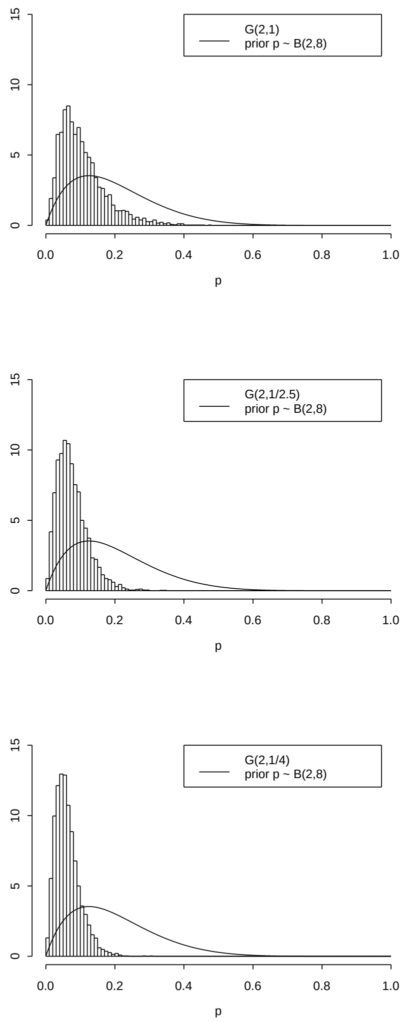


Fig. 6. Prior and posterior distributions for p.

and the standard deviation for the prior distribution of p was 0.12. From Figure 6 and Table 3, it can be noticed that the posterior estimates for p were smaller than 0.2 for all prior specification for $\theta_{[ij]}$. For $\mathcal{G}(2, 1/4)$, for example, the posterior estimate for the probability of a change is more precise and is only 0.0625.

Yet another important observation concerns the most probable partition and the posterior probability for each point (or instant) to be a change point. Figure 7 shows the probability of each instant to be a change point for all prior specification

Table 3. Descriptive statistics for the posterior distributions of p.

Prior	Mean	StDev	Q1	Median	Q3	Min	Max
$\theta_{[ij]} \sim \mathcal{G}(2,1)$	0.105	0.0663	0.0576	0.0908	0.136	0.00336	0.474
$\theta_{[ij]} \sim \mathcal{G}(2,1/2.5)$	0.0758	0.0439	0.0437	0.0676	0.0986	0.00192	0.348
$\theta_{[ij]} \sim \mathcal{G}(2,1/4)$	**0.0625**	0.0354	0.0368	0.0556	0.0808	0.00198	0.308

for $\theta_{[ij]}$. It is noticeable that no point had probability above 27% of being a change point. For all models, the 28th observation was identified as the most probable point to experience a change (the probability of the 28th observation to be a change point was ≈ 0.26). Also, for all models, the partition $\rho = \{0, 27, 50\}$ was identified as the most probable partition. This partition indicates that the 28th observation is the only change point in the sequence (just two observation away from the real change point).

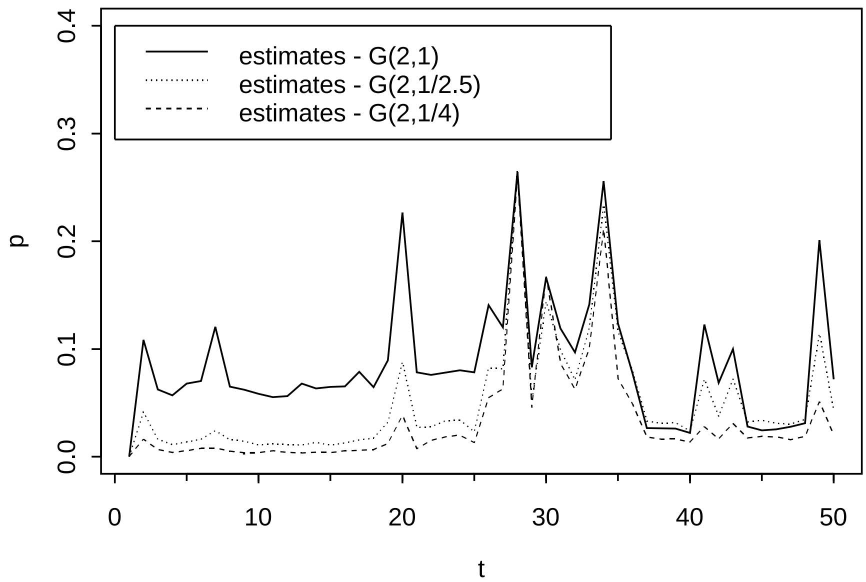


Fig. 7. Posterior probability of a change point.

From Table 4, we notice that the posterior probability of occurrence of the real partition was very small for all prior specifications. We also observe that the posterior probability of the 26th observation to be a change point is much smaller than 26%.

Table 4. Posterior probabilities.

Prior	most probable partition	real partition	real change point
$\theta_{[ij]} \sim \mathcal{G}(2,1)$	0.045	0.0087	0.141
$\theta_{[ij]} \sim \mathcal{G}(2,1/2.5)$	0.097	0.0129	0.082
$\theta_{[ij]} \sim \mathcal{G}(2,1/4)$	0.147	0.0182	0.055

4. A Case Study

The case study will focus on the data sequence "Hyde Park purse snatchings in Chicago", 28 day periods, from Jan., 69, to Sep. 73, from McCleary & Hay, Jr [15]. The data can be seen in Figure 8, along with the product estimates for the rate θ of purse snatchings. The goal here is to verify if this rate changes along the time.

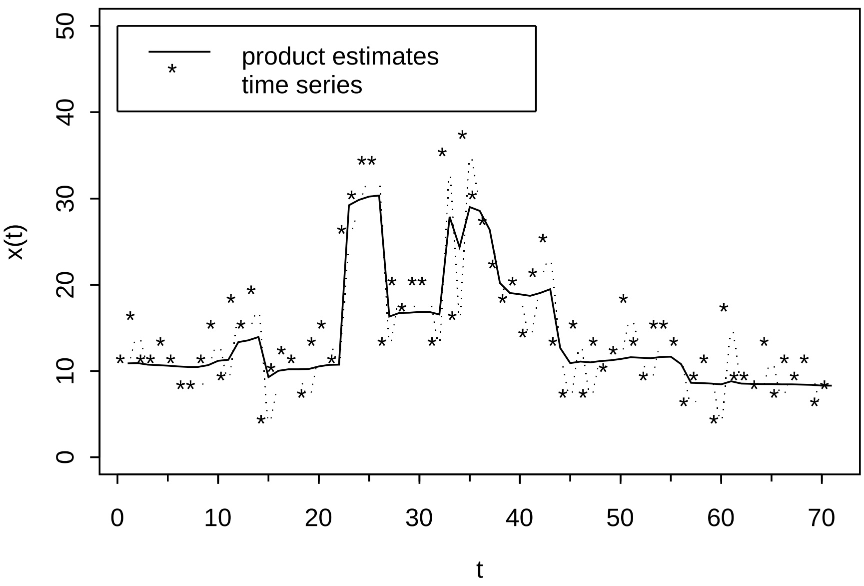


Fig. 8. Data and product estimates.

For the analysis of the data set, it was assumed that $\theta_{[ij]} \sim \mathcal{G}(1+1, 1/14)$ and

$p \sim \mathcal{B}(2, 8)$ shown in Figure 3. It should be noticed that the prior distribution for θ has mean equal to 28 and variance 393, which means that little information is available for θ.

From Figure 8, it can be noticed that between the 23th and 45th observations, Hyde Park experienced with a period with high rate of purse snatchings. Immediately after the 45th observation, the estimate for the rate of purse snatchings in the park was similar to the estimates obtained before the 23th observation. After the 46th observation the rate of purse snatchings reached its smallest level.

Figure 9 shows the most probable partition and the probability of each instant to be a change point. It is noticeable that only the observations 15 (with probability 0.569), 23 (with probability 0.992), 27 (with probability 0.988), 33 (with probability 0.852), 37 (with probability 0.651), 44 (with probability 0.768), and 57 (with probability 0.569) had probability above 50% of being a change point. However, the posterior probability of the partition formed by these most probable change points was only 0.00267. The most probable partition occured with probability 0.004, which is in agreement with the main changes observed in Figure 8. Besides the points mentioned above, the most probable partition also included the 12th observation, which had just the probability of 0.386 of being a change point.

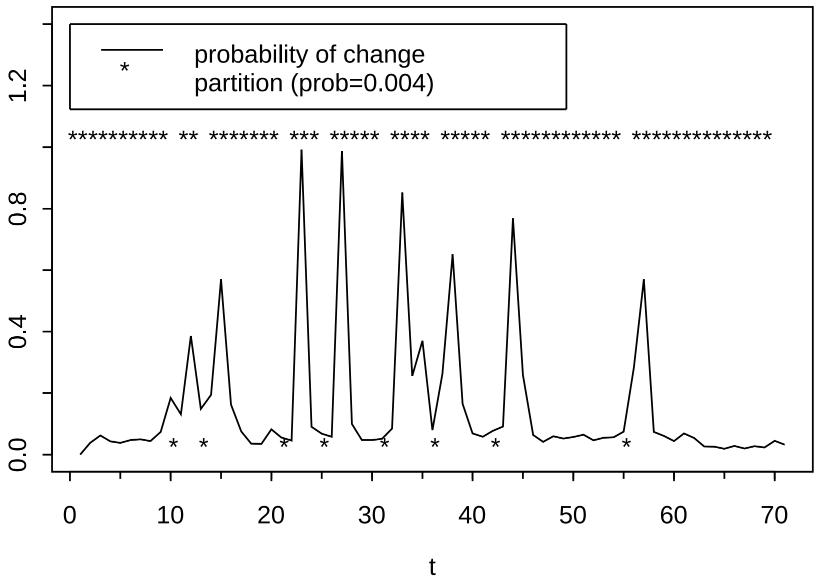
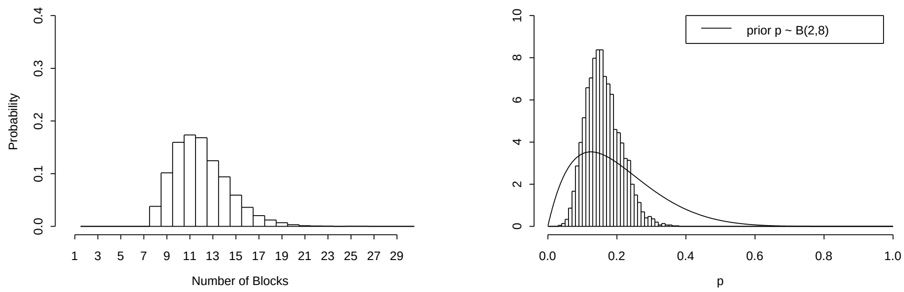


Fig. 9. Posterior probability of a change point and the most probable partition.

For the squared error loss penalty function, the prior estimate for the number of blocks in the partition is 14, which means that, in the prior evaluation, 13 change points were expected in the rate of purse snatchings in the Hyde Park. The variance for B, in this case, is 81.58. Notice from Figure 10-a and from Table 5 that the posterior estimates for B are more precise (the standard deviation is 2.39) and decreases to 11.96.

Figure 10-b shows the prior and the posterior distributions for the probability p of a change to take place in any instant in the purse snatchings rate. From Table 5, it is noticeable that the posterior estimate for p is 0.161 (which is smaller than 0.2, the prior estimate) and the standard deviation for the posterior distribution of p is smaller than the prior standard deviation.



a) B b) p

Fig. 10. Posterior distributions.

Table 5. Descriptive statistics for the posterior distributions of B and p.

	Mean	StDev	Q1	Median	Q3	Min	Max
B	**11.96**	**2.39**	10	12	13	8	26
p	**0.161**	0.0506	0.125	0.156	0.193	0.0362	0.378

5. Conclusions and Final Remarks

The problem of identifying multiple change point in Poisson data sequences was treated by an original version of the Product Partition Model (PPM). The PPM was described and its importance to change point identification problems was stressed, particularly in analyzing data sequences. A Gibbs sampling scheme was derived to implement the PPM, overcoming its inherent computational difficulties. The algorithm proposed proved to be an efficient and useful tool in analyzing change point problems in Poisson data sequences.

In the simulated data sequence to which it was applied, the method performed satisfactory. It could be noticed that, despite of the prior estimates for the rate in each instant have not held on the change exactly at the 26th observation, the PPM successfully identified the change. It is also noticeable that the PPM could identify the change with some delay. Another important fact to be pointed out is that the product estimates for the rate are not strongly influenced by different prior specifications to the rate, which concentrate most of their mass in small values. The number of blocks (or, equivalently, the number of change points in the sequence), was correctly identified (the mode of the posterior distribution of B is two) when the prior specification for the rate had large variance.

Some open research questions remain. How long would the treatable series be? How well would the methodology be for other subject areas? These and other similar questions are interesting and relevant topics for future research in this area.

Acknowledgments

Rosangela H. Loschi acknowledges CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*) of the Ministry for Science and Technology of Brazil (grants 300325/2003-7, 472066/2004-8) and PRPq-UFMG (grants 3893-UFMG/ RTR/ PRPq/ FUNDO/ 98 and 4801-UFMG/ RTR/ FUNDO/ PRPq/ RECEMD-OUTORES/ 00) for a partial allowance to her research. The research of Frederico R. B. Cruz has been partially funded by the CNPq, grants 301809/96-8, 201046/94-6, 300609/2002-7, 307702/2004-9, and 472066/2004-8, the FAPEMIG (*Fundação de Amparo à Pesquisa do Estado de Minas Gerais*), grants CEX-289/98 and CEX-855/98, and PRPq-UFMG, grant 4081-UFMG/ RTR/ FUNDO/ PRPq/ 99.

References

[1] D. Barry & J. A. Hartigan. A Bayesian analysis for change point problem. *Journal of the American Statistical Association* 88(421):309–319 (1993).

[2] J. Besag, D. Higdon P. Green & K. Mengersen. Bayesian computation and stochastic systems. *Statistical Science* 10:3–66 (1995).

[3] C. W. S. Chen & J. C. Lee. Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* 16(5):483–492 (1995).

[4] E. M. Crowley. Product partition models for normal means. *Journal of the American Statistical Association* 92(437):192–198 (1997).

[5] D. Gamerman, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. (Chapman & Hall, London, UK, 1997).

[6] A. E. Gelfand & A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409 (1990).

[7] S. Geman & D. Geman. Stochastic relaxation, Gibbs distribution and Bayesian restoration of images. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 6:721–741 (1984).

[8] J. A. Hartigan. Partition models. *Communication in Statistics - Theory & Methods*, 19(8):2745–2756 (1990).

[9] D. M. Hawkins. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37(3):323–341 (2001).

[10] D. A. Hsu. A Bayesian robust detection of shift in the risk structure of stock market returns. *Journal of the American Statistical Association*, 77(2):29–39 (1982).

[11] V. K. Jandhyala, S. B. Fotopoulos & D. M. Hawkins. Detection and estimation of abrupt changes in the variability of a process. *Computational Statistics & Data Analysis*, 40(1):1–19 (2002).

[12] R. H. Loschi & F. R. B. Cruz. Applying the product partition model to the identification of multiple change points. *Advances in Complex Systems*, 5(4): 371–387 (2002).

[13] R. H. Loschi & F. R. B. Cruz. Extension to the product partition model: Computing the probability of a change. *Computational Statistics & Data Analysis*, 48(2):255–268 (2005).

[14] R. H. Loschi, F. R. B. Cruz, P. L. Iglesias & R. B. Arellano-Valle. A Gibbs sampling scheme to the product partition model: An application to change point problems. *Computers & Operations Research*, 30(3):463–482 (2003).

[15] R. McCleary & R. A. Hay, Jr. *Applied Time Series Analysis for the Social Sciences*. (Sage Publications, Beverly Hills, CA, USA, 1980).

[16] U. Menzefricke. A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. *Applied Statistics*, 30(2):141–146 (1981).

[17] H. S. Migon & D. Gamerman. *Statistical Inference: An Integrated Approach*. (Arnold, London, Great Britain, 1999).

[18] F. A. Quintana & P. L. Iglesias. Bayesian clustering and product partition models. *Journal of the Royal Statistical Society B*, 65(2):557–574, 2003.

[19] C. Robert. Convergence control methods for Markov chain Monte Carlo algorithms. *Statistical Science*, 10:231–253 (1995).

[20] A. F. M. Smith. A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62(2):407–416 (1975).

[21] D. Stauffer. Percolation models of financial market dynamics. *Advances in Complex Systems*, 4(1):19–27 (2001).

[22] Y. Yao. Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. *The Annals of Statistics*, 12(4):1434–1447 (1984).