Bayesian Analysis (2014) **9**, Number 4, pp. 823–858

On a Nonparametric Change Point Detection Model in Markovian Regimes

Asael Fabian Martínez* and Ramsés H. Mena†

Abstract. Change point detection models aim to determine the most probable grouping for a given sample indexed on an ordered set. For this purpose, we propose a methodology based on exchangeable partition probability functions, specifically on Pitman's sampling formula. Emphasis will be given to the Markovian case, in particular for discretely observed Ornstein-Uhlenbeck diffusion processes. Some properties of the resulting model are explained and posterior results are obtained via a novel Markov chain Monte Carlo algorithm.

Keywords: Bayesian nonparametric, Change point detection, Ornstein-Uhlenbeck process, Two-parameter Poisson-Dirichlet process

1 Introduction

We present a model which deals with change point detection problems. Assume dependent data $y = (y_{t_1}, \ldots, y_{t_n})$ are observed at times $t_1, \ldots, t_n$, with $0 < t_i < t_j$ for all $i < j$. Our purpose is to identify changes in their underlying structure by focusing on the posterior distribution

$$\mathbb{P}(\rho_n|y) \propto \mathbb{P}(\rho_n)\mathbb{P}(y|\rho_n), \tag{1}$$

where ρ_n is a random variable modeling all possible groupings for the data preserving the time ordering. In particular, the study of $\mathbb{P}(y|\rho_n)$ can be simplified by splitting it into *regimes*, that is, parts of the likelihood characterized by structurally different parts of y, under the given modeling assumptions. This kind of model is appealing in many fields and applications such as genetics (for instance in DNA segmentation, Braun et al. 2000, and phylogenetic recombination detection, Minin et al. 2007), signal processing (in EEG analysis, Kaplan and Shishkin 2000, and signal segmentation, Punskaya et al. 2002), environmental time series (for detecting changes in wind speed and direction, Dobigeon and Toumeret 2007, and in hydrometeorological data, Perreault et al. 2000), econometrics (detecting changes in variance for stock prices, Chen and Gupta 1997, and analyzing inflation dynamics, Jochmann 2010) among others.

The literature on the subject is vast and encompasses both frequentist and Bayesian approaches. Some up to date reviews, mainly gathering classical nonparametric methods, can be found in Brodsky and Darkhovsky (1993) and Chen and Gupta (2011). On the Bayesian counterpart, early contributions focused on change points induced by a change of parameter within a parametric family modeling sequential observations, e.g. Chernoff and Zacks (1964) and Kander and Zacks (1966). A full Bayesian estimation

*Universidad Nacional Autónoma de México fabian@sigma.iimas.unam.mx
†Universidad Nacional Autónoma de México ramses@sigma.iimas.unam.mx

 DOI:10.1214/14-BA878

approach was proposed by Smith (1975). He framed it under the assumption of exchangeable intra-regime observations, inter-regime independence, and considered only a single change point. Smith's approach was later extended to a nonparametric regime distribution by Muliere and Scarsini (1985), where two independent Dirichlet processes are assumed as models for two different regimes separated by an independent change point random variable. Such independence assumptions were subsequently relaxed by Mira and Petrone (1996) who proposed a mixture of products of Dirichlet processes to model the dependence among the above two regimes.

Whereas the idea behind a generalization to multiple change points follows a similar line of reasoning, its mathematical and computational treatment is more demanding. For instance, within the parametric framework, Green (1995) makes use of reversible jump Markov chain Monte Carlo (MCMC) to draw posterior inferences on the number of change points. Concretely, he focuses on change points among data modeled through a non-homogeneous Poisson process with a step function as intensity and where the number of steps, namely the number of regimes, is determined via a birth and death mechanism. This approach requires the construction of transdimensional samplers which, depending on the choice of the model, does not always lead to a simple way to disentangle the prior and posterior mass assigned to a given change point structure.

Exhausting the full set of possibilities for change points requires considering distributions on all ordered partitions of the data. In this direction, Barry and Hartigan (1992, 1993) proposed a product partition model (PPM). Concretely, they assumed the number of change points as random and assigned them a product distribution, where the probability of each partition is given by Yao's cohesion function (Yao 1984). See Loschi et al. (2003) and Loschi and Cruz (2005) for further developments in this direction.

Also using PPMs, Quintana and Iglesias (2003) propose an algorithm to select a single partition according to some specific decision problem of interest, such as outliers detection. They also underline the relationship between PPMs' cohesion functions and the underlying clustering structure generated by the Dirichlet process. In a similar fashion, Park and Dunson (2010) generalized the Dirichlet process clustering by making it predictor-dependent, see also Müller et al. (2011). It is evident, then, that various change point models available in the literature are based on random partition distributions.

Hence, an important component for the study of change point detection problems is the distribution on partitions that is assumed and its properties. Under this viewpoint, the Bayesian nonparametric literature offers a wide variety of models, i.e. exchangeable partition probability functions (EPPFs), which after appropriate modifications can be considered as an alternative to the PPMs-based approach. Notice, however, that the support of such distributions is not immediately adequate for change point problems as data-partitions which are not subsequently observed should not receive positive probability.

In order to overcome this latter issue, an alternative approach can be devised from a distribution presented in Fuentes-García et al. (2010). With a different purpose in mind, i.e. classification, they introduce a distribution that preserves an ordering con-

straint by restricting the support of a Dirichlet process mixture model. Our proposal elaborates on this idea to construct more suitable distributions for ρ_n. Indeed, it can be seen to resemble some PPMs based approaches, where cohesion functions are given by a suitable modification of EPPFs derived from Bayesian nonparametric mixture models. In particular, we will focus our discussion on a distribution derived from Pitman's sampling formula, also known as the two-parameter Poisson-Dirichlet process (Perman et al. 1992). Within the Bayesian nonparametric literature, this choice of prior constitutes a good balance as it contains the Dirichlet process (Ferguson 1973) and the normalized stable process (Kingman 1975) as particular cases, the most informative and the least informative within the class of Gibbs-type priors (Lijoi et al. 2007b). In particular, one of the parameters controls the reinforcement of group sizes in such a class of priors. This, within the context at issue, will be interpreted as a change point sensitiveness parameter.

As mentioned before, one of the areas where change point models are of great interest is financial econometrics. In particular, detecting structural periods of distributional homogeneity is of interest in several scenarios of volatility and asset prices. Robust change-point-detection strategies can be used to establish financial markers for global financial crisis (Allen et al. 2013), to detect the *de facto* exchange rate regime in operation of a particular central bank (Reinhart and Rogoff 2004; Bubula and Ötker-Robe 2002), to measure business cycles (Harding and Pagan 2008), etc. In order to illustrate our proposal, we analyze the exchange rate between US Dollars and Mexican Pesos during the period of January 2007–December 2012. As it can be seen in Benavides and Capistrán (2009), understanding structural changes in the exchange rate helps also to understand the mechanisms through which monetary policies affect the Mexican economy.

Regarding the underlying model for the data, and in the light of the exchange-rate analysis, there are various aspects to consider. The first and most important one goes along the lines that change point models can be used to study (potentially) non-stationary series by partitioning them in intervals, each one having certain distributional homogeneity (Aggarwal et al. 1999; Mikosch and St**u**aric**u**a 2004; Lavielle and Teyssière 2007). Second, most financial assets, in particular exchange rates, evolve naturally in continuous time, and, thus, a model encompassing this characteristic would be preferred. In addition, to avoid no-identifiability of certain change points intrinsic to the model, we rule out features such as jump discontinuities. Lastly, we would like to be able to compare our proposal with other approaches, e.g. those based on independent distributions.

Although higher-order models could be considered, long-range dependent processes and non-stationary processes tend to be confused when used within change point contexts (Bhattacharya et al. 1983), hence, we confine the model for y to Markovian dependence. Therefore, an excellent candidate and a good compromise between the above requirements, generality and computability, is the Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein 1930). This model is the only stationary, Markovian diffusion process which is Gaussian. In particular, it includes the case of independent Gaussian observations.

It is worth saying that, in different but related directions, Markovian and other sorts of dependence structures have been considered. For instance, in Fearnhead and Liu (2007), the actual change points are modeled through a Markov chain, whereas, in Monteiro et al. (2011), dependence is induced at the inter-regime parameter levels.

Notice that, within a probabilistic approach, change point problems are bound to be model-based. Clearly, undertaking a Bayesian approach, e.g. setting prior distributions for the parameters characterizing different regimes, modifies the marginal probabilistic structure assumed for the data. Hence, in order to highlight the strength of our proposal based on the two-parameter Poisson-Dirichlet, we concentrate on the main problem of detecting change points and their location. In order to do so, we consider the integrated regime likelihood, while keeping the parameter controlling the dependence.

The layout of the paper is as follows. In Sections 2 and 3, we introduce and explain our proposal. First, we focus on the distribution of ρ_n and its properties, and later, we derive the integrated regime likelihood obtained from the Ornstein-Uhlenbeck process. Due to the large support induced by the distribution of ρ_n, the use of MCMC techniques becomes essential, therefore an algorithm is also depicted in Section 4 with details deferred to the appendix. A sensitivity analysis is presented in Section 5, in particular to disentangle the role of the reinforcement parameter underlying to the two-parameter Poisson-Dirichlet process as a change-point-detection vulnerability parameter. Section 6 shows the performance of our approach to detect change points in the exchange rate between US Dollars and Mexican Pesos. Finally, Section 7 contains some concluding remarks.

2 Probability distribution of ρ_n

Let $\mathcal{X} = \{x_1, \ldots, x_n\}$ be an arbitrary collection of n items. In classification problems, every possible grouping is formed by subsets $\{A_1, \ldots, A_k\}$ such that

1. $A_i \subset \mathcal{X}$ and $A_i \neq \emptyset$, for $i = 1, \ldots, k$,
2. $A_1 \cup \cdots \cup A_k = \mathcal{X}$,
3. $A_i \cap A_j = \emptyset$, for each $i \neq j$.

All possible groupings, called hereafter *partitions*, form a combinatorial class known as *set partitions* (of $\mathcal{X}$); its cardinality is given by the Bell number B_n, with n the size of $\mathcal{X}$. Let us denote this class by $\mathcal{P}_\mathcal{X}$.

Example 1. *Let $\mathcal{X} = \{x_1, x_2, x_3\}$. Then there are $B_3 = 5$ partitions in $\mathcal{P}_\mathcal{X}$ and they are:*

$$\{\{x_1, x_2, x_3\}\},\ \{\{x_1\}, \{x_2, x_3\}\},\ \{\{x_1, x_2\}, \{x_3\}\},\ \{\{x_1, x_3\}, \{x_2\}\},\ \{\{x_1\}, \{x_2\}, \{x_3\}\}.$$

Thus, for example, $\{\{x_1, x_2, x_3\}\}$ means that all elements belong to the same single group and $\{\{x_1\}, \{x_2\}, \{x_3\}\}$ that each element belongs to its own group.

Let $[n] := \{1, 2, \ldots, n\}$. If we assume that $\mathcal{X}$ is the observed data, notice that $\mathcal{P}_{\mathcal{X}}$ and $\mathcal{P}_{[n]}$ are isomorphic combinatorial classes, then classifying $\mathcal{X}$ or $[n]$ is an equivalent problem. We use the latter to agree with the common notation.

Perhaps the easiest and more convenient way to define distributions on the set of partitions is via the concept of EPPFs.

Definition 1. *A* random partition Π_n *is a random variable with support* $\mathcal{P}_{[n]}$. *In addition,* Π_n *is exchangeable if for any partition* $\{A_1, \ldots, A_k\}$ *of* $[n]$

$$\mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}) = p(n_1, \ldots, n_k), \tag{2}$$

for some symmetric function p *of integer compositions* $n_j := |A_j|$, $j = 1, \ldots, k$. *Here* $|A|$ *denotes the cardinality of* A. *The function* p *is called an* exchangeable partition probability function *(EPPF).*

Remark 1. *An* integer composition *of* $n \in \mathbb{N}$ *is a sequence* $(n_1, \ldots, n_k) \subset \mathbb{N}^k$ *such that* $n_1 + \cdots + n_k = n$ *for some* $k > 0$.

These classes of functions arise naturally within the Bayesian platform for exchangeable observations, which, for the sake of completeness, we briefly state. Let $\mathsf{P}_{\mathbb{X}}$ denote the set of all probability measures on $\mathbb{X}$. Due to de Finetti's theorem (de Finetti 1931), assuming the property of exchangeability for a set of $\mathbb{X}$-valued observations $(X_i)_{i\geq 1}$ is equivalent to assuming the existence of probability distribution Q on $\mathsf{P}_{\mathbb{X}}$ such that for any $i \geq 1$,

$$\begin{aligned} X_i|\tilde{p} &\overset{\text{iid}}{\sim} \tilde{p} \\ \tilde{p} &\sim Q. \end{aligned}$$

Whenever Q is infinite-dimensional, one speaks of a Bayesian nonparametric inferential problem and Q is interpreted as a nonparametric prior for the random probability measure (RPM) $\tilde{p}$. Of particular relevance are those nonparametric priors selecting discrete distributions almost surely (a.s.), in other words, those associated to RPMs with the representation

$$\tilde{p} = \sum_{j=1}^{\infty} w_j \, \delta_{\xi_j},$$

where $w_j > 0$, $j = 1, 2, \ldots$, and $\sum_{j\geq 1} w_j = 1$ a.s., and $(\xi_j)_{j\geq 1}$ is a sequence of iid $\mathbb{X}$-valued random variables, independent of $(w_j)_{j\geq 1}$. Under such a discrete nature for $\tilde{p}$, the observables $(X_i)_{i=1}^n$ are bound to have ties with positive probability and, thus, to form $K_n = k$ groups with distinct representative values, e.g. $X_1^*, \ldots, X_k^*$, with frequencies $N_1, \ldots, N_k$, where clearly $k \leq n$. Therefore, selecting a model for Q induces an EPPF defined through

$$p(n_1, \ldots, n_k) := \mathbb{P}(K_n = k, N_1 = n_1, \ldots, N_k = n_k) = \int_{\mathbb{X}^k} \mathbb{E}_Q\left(\prod_{j=1}^{k} \tilde{p}^{\,n_j}(\mathrm{d}x_j) \right), \tag{3}$$

which reads as the probability of partitioning a sample of size n into k groups with frequencies $n_1, \ldots, n_k$, provided $\sum_{i=1}^{k} n_i = n$. Moving away from the Dirichlet Process model, a relatively general approach to define classes of Q can be found in Regazzini et al. (2003). Furthermore, an up-to-date review and several of such classes with their resulting EPPFs are discussed in Lijoi and Prünster (2010).

Hence, EPPFs have been used in several contexts, e.g. species sampling models, mixture models, density estimation, classification problems and, in general, form the basis of Bayesian nonparametric inference. However, for change point detection, an EPPF is not directly useful due to the induced order that the times $(t_j)_{j\geq 1}$ impose.

Suppose we observe x_1, x_2, x_3 sequentially and consider Example 1. Since we are interested in the location of change points, partitions like $\{\{x_1, x_3\}, \{x_2\}\}$, namely partitions lacking of ordering in their time indices, are meaningless and should receive probability zero. While it is true that for certain data sets such kind of groupings should receive positive probability, we believe their study would be better framed under more general classification strategies. Therefore, a different support other than $\mathcal{P}_{\mathcal{X}}$ is required. One choice is to restrict $\mathcal{P}_{\mathcal{X}}$, considering only those partitions which preserve the order in the (time) indices of $\mathcal{X}$.

Such a subclass can be built as follows. For a given integer composition $(n_1, \ldots, n_k)$ of n, consider the partition $\{A_1, \ldots, A_k\} \in \mathcal{P}_{\mathcal{X}}$ such that $A_j = \{x_{s_{j-1}+l} : l = 1, \ldots, n_j\}$, $j = 1, \ldots, k$, where $s_j = n_1 + \cdots + n_j$ with $s_0 = 0$. We call this partition a *set composition*. Moreover, we term the resulting combinatorial class formed by all these set compositions *the class of set compositions (of $\mathcal{X}$)*, and it will be denoted by $\mathcal{C}_{\mathcal{X}}$. Clearly, $|\mathcal{C}_{\mathcal{X}}| = 2^{n-1}$.

Example 2. *Let $\mathcal{X} = \{x_1, x_2, x_3\}$ be as in the previous example. Then, there are $2^{3-1} = 4$ compositions in $\mathcal{C}_{\mathcal{X}}$ and they are*

$$\{\{x_1, x_2, x_3\}\},\ \{\{x_1\}, \{x_2, x_3\}\},\ \{\{x_1, x_2\}, \{x_3\}\},\ \{\{x_1\}, \{x_2\}, \{x_3\}\}.$$

This combinatorial class is more adequate for change point detection problems because every single observation y_{t_i}, $i = 1, \ldots, n$, could start a new group. Within the change point context, this means that every first observation's time index within each group will be a change point, except t_1. That is, we could have at most $n - 1$ change points arranged in 2^{n-1} different ways.

Since $\mathcal{C}_{\mathcal{X}} \subseteq \mathcal{P}_{\mathcal{X}}$ for any set $\mathcal{X}$, a distribution of the random variable of interest for grouping, ρ_n, could be obtained by restricting a particular class of EPPF. One way to do this, that keeps the essence of EPPFs, but supported on $\mathcal{C}_{[n]}$ instead of $\mathcal{P}_{[n]}$, can be done through the following definition devised from Pitman (2006).

Definition 2. *A $\mathcal{C}_{[n]}$-valued random variable ρ_n is said to have an* exchangeable random order distribution *if it is given by*

$$p'(n_1, \ldots, n_k) = \binom{n}{n_1, \ldots, n_k} \frac{1}{k!}\, p(n_1, \ldots, n_k), \tag{4}$$

for p the EPPF of some Π_n.

The idea is that the mass for all partitions with $(n_1, \ldots, n_k)$ degenerates on those partitions with the same $(n_1, \ldots, n_k)$ but where the order of the data is preserved.

An attractive statistic within the study and applications of EPPFs is the marginal distribution of K_n derived from (3), i.e.

$$\mathbb{P}(K_n = k) = \sum_{\{A_1,\ldots,A_k\}\in\mathcal{P}_{\mathcal{X}}} \mathbb{P}(\Pi_n = \{A_1, \ldots, A_k\}),$$

which indicates the number of partitions in a collection $\mathcal{X}$ of size n with exactly k groups. Clearly, such quantity tackles important problems like assessing the number of groups in a fully probabilistic way, e.g. in mixture modeling (Lijoi et al. 2007b) or in discovery probability problems (Lijoi et al. 2007a). In our context, the random variable $C_n := K_n - 1$ will be used to model the number of change points.

An appealing feature of the distribution in Definition 2 is that the distribution of K_n is preserved, and so are its asymptotic results (see Lijoi et al. 2007b).

Lemma 1. *Let ρ_n be distributed as (4). Hence*

$$\mathbb{P}_{\mathcal{C}_{\mathcal{X}}}(K_n = k) = \mathbb{P}_{\mathcal{P}_{\mathcal{X}}}(K_n = k).$$

Proof. The proof follows easily after noticing that

$$\begin{aligned}\mathbb{P}_{\mathcal{C}_{\mathcal{X}}}(K_n = k) &= \sum_{\{A_1,\ldots,A_k\}\in\mathcal{C}_{\mathcal{X}}} \mathbb{P}(\rho_n = \{A_1, \ldots, A_k\}) \\ &= \sum_{\{A_1,\ldots,A_k\}\in\mathcal{C}_{\mathcal{X}}} p'(|A_1|, \ldots, |A_k|) \\ &= \sum_{\{B_1,\ldots,B_k\}\in\mathcal{P}_{\mathcal{X}}} p(|B_1|, \ldots, |B_k|),\end{aligned}$$

leading to the stated result. □

This construction of $\mathcal{C}_{[n]}$-valued random variables is an attractive alternative to the PPM approach based on cohesion functions.

Here, we will focus on the EPPF induced by the two-parameter Poisson-Dirichlet process, that is the EPPF characterized by the probability distribution

$$\mathbb{P}(\Pi_n = (n_1, \ldots, n_k)) = \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{(\theta+1)_{n-1\uparrow}} \prod_{j=1}^{k} (1-\sigma)_{n_j-1\uparrow},$$

where $(x)_{n\uparrow} = x(x+1)\cdots(x+n-1)$ denotes the Pochhammer symbol, with $(x)_{0\uparrow} = 1$, and $\sigma \in [0,1)$ with $\theta > -\sigma$ or $\sigma < 0$ with $\theta = m|\sigma|$ for some positive integer m. In particular, we will work with the case $\sigma \in [0,1)$ which, within the context at issue, corresponds to the case of always having new change points as the sample n increases (see De Blasi et al. 2013 for more asymptotic results of this class of distributions).

Therefore, for this process, the restriction to compositions defined as in (4) simplifies to

$$\mathbb{P}(\rho_n = (n_1, \ldots, n_k)) = \frac{n!}{k!} \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{(\theta+1)_{n-1\uparrow}} \prod_{j=1}^{k} \frac{(1-\sigma)_{n_j-1\uparrow}}{n_j!}. \tag{5}$$

In both cases, the marginal distribution for K_n is given by

$$\mathbb{P}(K_n = k) = \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{\sigma^k(\theta+1)_{n-1}} \frac{1}{k!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (-j\sigma)_{n\uparrow}, \quad k = 1, \ldots, n,$$

with mean value given by

$$\mathbb{E}[K_n] = \frac{(\theta+\sigma)_{n\uparrow}}{\sigma(\theta+1)_{n-1\uparrow}} - \frac{\theta}{\sigma}. \tag{6}$$

This model, also known as the Pitman-Yor process, has been widely used in species sampling problems, density estimation, classification and machine learning. As mentioned in the introduction, this EPPF includes the Dirichlet and the normalized stable processes as particular cases when $\sigma = 0$ and $\theta = 0$, respectively. In particular, it stands out within the class of Gibbs-type priors as it places a good tradeoff between being informative or not about the prior information on K_n (Lijoi et al. 2007b).

The above comments about the flexibility of the two-parameter Poisson-Dirichlet process can be seen through the marginal distribution of K_n, as σ varies. This parameter plays an important role in the underlying clustering behavior. Lijoi et al. (2007b) study this effect for the same parameter in the Generalized Gamma process; since both processes belong to the Gibbs-type priors, a similar behavior is observed in this case. Figure 1 shows some examples of the density function of K_n with $\mathbb{E}(K_n)$ fixed. This feature will also play an important role when detecting change points.

Clearly, there are other ways to construct distributions supported on $\mathcal{C}_{[n]}$. Indeed, Fuentes-García et al. (2010) propose to assign probability zero to those partitions not preserving the order. Although similar to our approach, the way the remaining mass is distributed is different. Specifically, within the context of sequentially observed data, they considered

$$p^*(n_1, \ldots, n_k) \propto p(n_1, \ldots, n_k)\, \mathbb{I}(t_1 < t_2 < \cdots < t_n),$$

which maintains the symmetry of the EPPF, just as in (4), but changes the marginal distribution of K_n.

Another approach can be deduced by simply focusing on the Markov chain, $(K_n)_{n=1}^{\infty}$, induced by the generalized Pólya urn characterizing the two-parameter Poisson-Dirichlet process, that is, a Markov chain driven by one-step transition probabilities

$$\mathbb{P}(K_{n+1} = \xi | K_n = k) = \frac{\theta + k\sigma}{\theta + n}\, \mathbb{I}(\xi = k+1) + \frac{n - k\sigma}{\theta + n}\, \mathbb{I}(\xi = k),$$

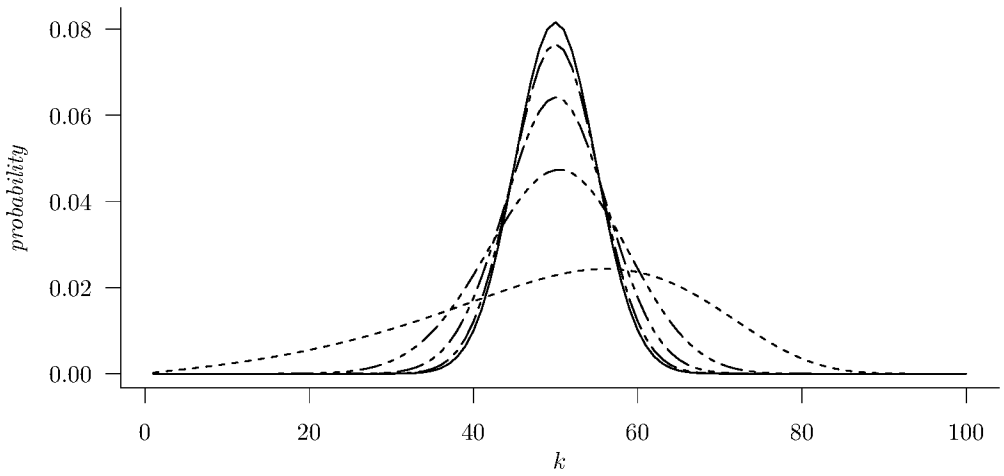


Figure 1: Probability density function of K_{100} for different values of (σ, θ) such that $\mathbb{E}(K_{100}) = 50$; σ was fixed to: —— 0.1, -·— 0.3, --— 0.5, ---- 0.7 and ---- 0.9. Points are connected by straight lines for visual simplification.

and $K_1 = 1$. This approach keeps the marginal characteristics of K_n induced by the two-parameter Poisson-Dirichlet distribution, but breaks the symmetry of the resulting distribution on $\mathcal{C}_{[n]}$, say $p^\star$. In particular, this would mean that change points occurring early would receive higher probabilities. Notice that this approach could be considered as tantamount to the one obtained via infinite hidden Markov models (iHMMs), with the state Markov process restricted to have upper triangular transition matrices.

Despite the above drawback, such an approach to study the marginal statistic K_n could be of interest in other problems. In particular, it clearly leads as an alternative, perhaps easier, approach to solve some species sampling problems (Lijoi et al. 2007a).

In order to make these observations clearer, consider the results in Table 1 where the above probabilities are shown for the case of $n = 4$ and $(\sigma, \theta) = (0.35, 2.7)$. Note how both, symmetry in the compositions and distributional properties for K_4 are preserved by distribution (4), unlike in the other two approaches.

3 Integrated regime likelihood

According to Barry and Hartigan (1992), PPMs are such that, conditioned to the partition ρ_n, the observations in different groups are independent. There also exists a sequence of parameters $(x_j)_{j=1}^n$ which, given the partition ρ_n, are independent and follow certain distribution. It is worth mentioning that some recent approaches, based on PPMs, introduce dependence in the model at the level of parameters (see, for example Monteiro et al. 2011). In both cases, however, the observations within the same group are assumed also independent given the group's parameter x_l^*.

k	*composition*	p'	p^*	$p^\star$	p
1	$\{\{x_1, x_2, x_3, x_4\}\}$	0.029	0.041	0.029	
		0.029	*0.041*	*0.029*	*0.029*
2	$\{\{x_1\}, \{x_2, x_3, x_4\}\}$	0.066	0.047	0.092	
	$\{\{x_1, x_2\}, \{x_3, x_4\}\}$	0.039	0.019	0.046	
	$\{\{x_1, x_2, x_3\}, \{x_4\}\}$	0.066	0.047	0.033	
		0.171	*0.112*	*0.171*	*0.171*
3	$\{\{x_1\}, \{x_2\}, \{x_3, x_4\}\}$	0.136	0.097	0.204	
	$\{\{x_1\}, \{x_2, x_3\}, \{x_4\}\}$	0.136	0.097	0.136	
	$\{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$	0.136	0.097	0.068	
		0.408	*0.290*	*0.408*	*0.408*
4	$\{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}$	0.392	0.557	0.392	
		0.392	*0.557*	*0.392*	*0.392*

Table 1: Distribution of ρ_4 and K_4 under the different approaches of constructing distributions on $\mathcal{C}_{[n]}$: the exchangeable random order p', the approach described by Fuentes-García et al. (2010) p^*, and the Markov chain induced by the two-parameter Poisson-Dirichlet process $p^\star$. The corresponding distributions of K_4 are shown in cursive numbers, as well as for the two-parameter Poisson-Dirichlet EPPF p.

Due to the examples we have in mind, we impose dependency at the level of observations by assuming a continuous time Markovian process modulating each regime. Afterwards, we integrate out the process's driven parameters but we preserve the dependence induced through the correlation parameter. By doing this, we can focus on change points' inferences.

In particular, we assume data $y = (y_{t_1}, \ldots, y_{t_n})$ are modeled by a strictly stationary Markovian process with invariant distribution $\pi(\cdot; x)$ and transition density $p(y_0, y_t; x)$, denoting generically the driven parameter by x. Then, the integrated regime likelihood is given by

$$\mathbb{P}(y|\rho_n = \{A_1, \ldots, A_k\}) = \prod_{j=1}^{k} L(y|A_j), \tag{7}$$

where $L(y|A_j)$ is the marginal likelihood given only the observations in group A_j, integrating out the parameter x, i.e.

$$L(y|A_j) = \int_{\mathbb{X}} \pi(y_{j,1}; x) \prod_{l=1}^{n_j - 1} p(y_{j,l}, y_{j,l+1}; x) P(\mathrm{d}x),$$

where $y_{j,i}$ is the ith observation in group A_j, $n_j = |A_j|$ and P is the prior distribution

of x. So, the integrated regime likelihood (7) is given by

$$\mathbb{P}(y|\rho_n = \{A_1, \ldots, A_k\}) = \prod_{j=1}^{k} \int_{\mathbb{X}} \pi(y_{j,1}; x) \prod_{l=1}^{n_j - 1} p(y_{j,l}, y_{j,l+1}; x) P(\mathrm{d}x). \tag{8}$$

Assuming the nature of the phenomena producing y evolves in continuous time, and following the discussion in Section 1, we complete the above specification by considering an Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein 1930). In particular, this process can be also seen as the solution to

$$\mathrm{d}Y_t = -\alpha(Y_t - \mu)\mathrm{d}t + \sqrt{\frac{2\alpha}{\lambda}}\,\mathrm{d}W_t,$$

for $\mu \in \mathbb{R}$, $\alpha, \lambda > 0$ and $(W_t)_{t \geq 0}$ denotes the standard Brownian motion. This process, apart from being stationary, reversible and Markovian is also Gaussian, allowing us, then, to study directly the independent case through its autocorrelation. Although other models could be also considered here, the model-based nature of change point problems suggests that these do not necessarily need to be very complicated. That is, sudden changes, jumps or other kinds of distributional variations could simply be part of a different regime under the above model.

Following Karatzas and Shreve (1988), it can be easily seen that the invariant and transition densities for this process, setting $\phi := e^{-\alpha}$, are given by

$$\begin{aligned} \pi(y; \mu, \lambda) &= \mathrm{N}(y; \mu, 1/\lambda) \\ p(y_0, y_t; \mu, \lambda, \phi) &= \mathrm{N}(y_t; y_0\phi^t + \mu(1 - \phi^t), (1 - \phi^{2t})/\lambda), \end{aligned}$$

where $\mathrm{N}(y; \mu, \sigma^2)$ denotes the normal density with mean μ and variance σ^2. From this, we have that $\mathbb{E}[Y_t] = \mu$ and $\mathrm{Var}[Y_t] = 1/\lambda$ for all t. In addition, its covariance and correlation functions are $\mathrm{Cov}(Y_s, Y_t) = \phi^{t-s}/\lambda$ and $\mathrm{Cor}(Y_s, Y_t) = \phi^{t-s}$ for all $0 \leq s \leq t$, respectively, with $0 < \phi < 1$. Therefore, parameter ϕ controls the correlation, and the case where observations are independent and normally distributed is obtained when $\phi \to 0$. To refer to the law of such a Gaussian process we will use $\mathrm{OU}(\mu, \lambda, \phi)$.

We computed the integrated regime likelihood under the assumption that observations are recorded at equally spaced times, thus using the simpler notation $t_i = i$, for $i = 1, \ldots, n$. Parameters μ and λ are integrated out using a Normal-Gamma prior distribution $\mathrm{N}(\mu; 0, (c\lambda)^{-1})\mathrm{Ga}(\lambda; a, b)$. Then, given ϕ, the integrated regime likelihood (8) is given by

$$\begin{aligned} \mathbb{P}(y|\rho_n) = \prod_{j=1}^{k} \frac{(2b(1-\phi^2))^a \Gamma(n_j/2 + a)}{\pi^{n_j/2}\Gamma(a)} \left(\frac{c(1+\phi)(1-\phi^2)}{c + n_j - \phi(n_j - c - 2)}\right)^{1/2} \times \\ \left(\mathbf{y}_j' \mathbf{S}_j \mathbf{y}_j - \frac{(1-\phi)\left(\sum_{i=1}^{n_j} y_{j,i} - \phi \sum_{i=2}^{n_j - 1} y_{j,i}\right)^2}{c + n_j - \phi(n_j - c - 2)} + 2b(1-\phi^2)\right)^{-(n_j/2 + a)}, \end{aligned} \tag{9}$$

where $\mathbf{y}_j = (y_{j,i} : i = 1 \ldots, n_j)$ and $\mathbf{S}_j \in \mathbb{R}^{n_j \times n_j}$ is given by

$$\mathbf{S}_j = \begin{pmatrix} 1 & -\phi & 0 & \ldots & 0 \\ -\phi & 1+\phi^2 & -\phi & \ldots & 0 \\ 0 & -\phi & 1+\phi^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}.$$

Notice that parameter ϕ models the correlation of the complete series y. Although, this parameter could also be integrated out, at least numerically, we prefer to keep it as an overall feature across regimes, since otherwise we could have regimes with independent components and others with Markovian components.

4 MCMC simulation algorithm

Even when the number of set compositions is much smaller than the number of partitions, simulating from the posterior distribution of ρ_n

$$\mathbb{P}(\rho_n|y) \propto \left(\frac{n!}{k!} \frac{\prod_{i=1}^{k-1}(\theta + i\sigma)}{(\theta+1)_{n-1\uparrow}} \prod_{j=1}^{k} \frac{(1-\sigma)_{n_j-1\uparrow}}{n_j!} \right) \mathbb{P}(y|\rho_n),$$

where $\mathbb{P}(y|\rho_n)$ is given by (9) becomes unfeasible for bigger datasets. The need of an MCMC algorithm is then evident. Although, we are using a different approach than the one used in Fuentes-García et al. (2010), their first algorithm can be adapted to our framework. Concretely, a split-merge algorithm which updates the number of groups k and the group sizes $(n_1, \ldots, n_k)$ using Metropolis-Hastings steps will be used. Note that unlike other split-merge MCMC algorithms on the space of partitions, the "split" and "merge" steps make complete sense here as the ordering restriction leaves us no other sensible moves.

Hence, two possible choices are available: a split, which creates a new group, or a merge, which combines two consecutive existing groups into a single one. After that, a random pair of adjacent groups is updated proposing new values in order to speed the sampler up. All these steps are performed introducing latent variables, avoiding, then, the need to change dimensions among iterations. The sampler is depicted in the algorithm shown in Figure 2, and further details are given in the appendix.

Additionally, it is possible to assign prior distributions to parameters σ, θ and ϕ so that we are also able to make inferences about them. We include this option in the algorithm as follows. For variables (σ, θ), the likelihood function is taken from (5), the prior for ρ_n. A Beta prior, with parameters (a, b), is assigned to σ and a shifted Gamma prior, with parameters $(\mathsf{c}, \mathsf{d}, -\sigma)$, to θ. We say that a random variable Z is shifted Gamma distributed with parameters (α, β, μ), for $\alpha, \beta > 0$ and $-\infty < \mu < \infty$, if and only if $Z - \mu$ is Gamma distributed with parameters (α, β).

We have decided to work with the case $\sigma \in (0, 1)$ in order to highlight the implied

read $y_1, \ldots, y_n$ and hyper-parameters
initiate k and $(n_1, \ldots, n_k)$
repeat
  with probability $q\,\mathbb{I}(1 < k < n) + \mathbb{I}(k = 1)$ ▷ split
    choose j uniformly from $\{j : 1 \leq j \leq k, n_j > 1\}$
    choose l uniformly from $\{1 \ldots, n_j - 1\}$
    with probability α in (15)
      $(n_1, \ldots, n_{k+1}) \leftarrow (n_1, \ldots, n_{j-1}, l, n_j - l, n_{j+1}, \ldots, n_k)$
      $k \leftarrow k + 1$
    end with
  otherwise ▷ merge
    choose j uniformly from $\{1, \ldots, k-1\}$
    with probability α in (16)
      $(n_1, \ldots, n_{k-1}) \leftarrow (n_1, \ldots, n_{j-1}, n_j + n_{j+1}, n_{j+2}, \ldots, n_k)$
      $k \leftarrow k - 1$
    end with
  end with
  if $k > 1$ **then** ▷ shuffle
    choose i uniformly from $\{1, \ldots, k-1\}$
    choose j uniformly from $\{1, \ldots, n_i + n_{i+1} - 1\}$
    with probability α in (17)
      $n_{i+1} \leftarrow n_i + n_{i+1} - j$
      $n_i \leftarrow j$
    end with
  end if
  [simulate values for additional parameters]
until converge and have the required number of samples
write the sampled values of k and $(n_1, \ldots, n_k)$ [and additional parameters]

Figure 2: Split-merge MCMC algorithm used for change point detection.

reinforcement mechanism. Thus, the posterior distribution of σ is given by

$$p(\sigma|\ldots) \propto \sigma^{\mathsf{a}-1}(1-\sigma)^{\mathsf{b}-1}(\theta+\sigma)^{\mathsf{c}-1}e^{-\mathsf{d}\sigma}\prod_{i=1}^{k-1}(\theta+i\sigma)\prod_{j=1}^{k}(1-\sigma)_{n_j-1\uparrow}$$

with $\sigma \in (\max\{-\theta, 0\}, 1)$. Simulation for this posterior distribution can be easily done via the adaptive Metropolis rejection sampling (ARMS) method (Gilks et al. 1995). For the posterior distribution of θ, we use the following result; the proof is given in the appendix.

Proposition 1. *The augmented full conditional distribution of θ, with likelihood function (5) and prior distribution shifted Gamma with parameters $(\mathsf{c}, \mathsf{d}, -\sigma)$, is given by*

$$p(\theta|y, z, \ldots) \propto \sum_{j=0}^{k+1} w_j \mathrm{Ga}(\theta; \mathsf{c}+j, \mathsf{d}+y-\log(z), -\sigma),$$

where the weights $(w_j)_{j=0}^{k+1}$ *are given by*

$$w_j \propto \frac{\big((n-\sigma)(n+1-\sigma)c_{k-1,j} + (2n+1-2\sigma)\sigma c_{k-1,j-1} + \sigma^2 c_{k-1,j-2}\big)\Gamma(\mathsf{c}+j)}{\big(\sigma(\mathsf{d}+y-\log(z))\big)^j},$$

where k *and* n *are as in* (5), $z \sim \text{Be}(\theta+2, n)$, $y \sim \text{Exp}(\theta+1)$ *and* $c_{k,j}$, $j = 1\ldots, k$, *are the absolute value of the Stirling number of the first kind with* $c_{k,r} = 0$ *for* $r \leq 0$ *and for* $r > k$ *and* $c_{0,0} = 1$.

To the best of our knowledge, previous implementations to simulate from the posterior distribution of θ were limited to Metropolis-Hastings steps (see, for example, Jara et al. 2010 and Nieto-Barajas and Contreras-Cristan 2014). Hence, the above proposition could also be useful in other Bayesian nonparametric implementations based on the two-parameter Poisson-Dirichlet process.

Together with the prior distribution of (σ, θ), we further assume an a priori independent $(0, 1)$-Uniform distribution of the dependency parameter ϕ. Therefore, its posterior distribution is proportional to Equation (9) and, as for parameter σ, the ARMS method can be used to simulate from it.

Regarding the inferences, we focus on the posterior distributions of K_n and ρ_n. Due to the support of ρ_n, estimates like mean, variance or quantiles are not straightforward to obtain. Therefore, we worked with two point estimates: the posterior mode and the least-squares clustering (Dahl 2006). The second estimate corresponds to the grouping g^* that minimizes the sum of squared deviations of the association matrix $\delta(g)$ from $\hat{\pi}$, i.e.

$$g^* = \arg\min_{g \in \mathcal{S}} \sum_{i,j} (\delta_{i,j}(g) - \hat{\pi}_{i,j})^2,$$

where $\mathcal{S}$ denotes the set of all the sampled compositions, $\delta(g)$ is such that its (i, j) element, $\delta_{i,j}(g)$, is the indicator of whether observations y_{t_i} and y_{t_j}, $i, j = 1, \ldots, n$, belong to the same group in g and $\hat{\pi}$ is the element-wise mean of all these association matrices. However, since both estimates coincided for almost all the simulations performed in the next sections, we will only report the mode which, in addition, has a neater interpretation.

Another method used by Loschi and Cruz (2005) to obtain a posterior estimate of ρ_n is based on the posterior probability that each time t_i, $i = 2, \ldots, n$, is a change point. Given a sequence of frequencies $(n_1, \ldots, n_k)$, their corresponding change points are given by $\langle j_1, \ldots, j_{k-1}\rangle$, where $j_i = n_1 + \cdots + n_i + 1$. Define the event T_i as the event where the ith instant t_i is a change point, $i = 2, \ldots, n$, then

$$\tau_i := \mathbb{P}(T_i|y) = \sum_{\langle j_1,\ldots,j_{k-1}\rangle \in \mathcal{S}} \mathbb{P}(\rho_n = \langle j_1, \ldots, j_l = t_i, \ldots, j_{k-1}\rangle|y), \tag{10}$$

with $\mathcal{S}$ as before. Using this, Loschi and Cruz's estimate $\rho_n^\star$ is given by all the times $(t_{j_i})_{i=1}^r$ whose probabilities τ_{j_i}, $i = 1, \ldots, r$, are greater than a certain threshold; hence $\rho_n^\star = \langle t_{j_1}, \ldots, t_{j_r}\rangle$. While this approach might be appealing in monitoring contexts, it

is not clear how to define a threshold for change point determination. Having said this, within the context at issue, it is useful to illustrate the role of parameter σ, so we will use probabilities $(\tau_i)_{i=2}^n$ in the next sections.

5 σ as a change point sensitivity parameter

In this section, we illustrate the role of the σ-parameter when detecting change points as well as the performance of the split-merge algorithm described in Figure 2. First, a small dataset is used in order to study the role of the σ-parameter. Afterwards, three bigger simulated datasets are used to depict the performance of the algorithm under various scenarios, e.g. assuming dependent ($\phi > 0$) or independent ($\phi = 0$) observations and under different mean and variance schemes.

For the first dataset, 15 observations are used. Working with a small dataset allows us to obtain exact results without the need of numeric algorithms and, therefore, we can draw conclusions without the error introduced by MCMC samplers. Then, data are simulated as follows:

$$y_i \overset{\text{iid}}{\sim} \text{N}(0, 0.5) \quad i = 1, \ldots, 6$$
$$y_i \overset{\text{iid}}{\sim} \text{N}(2, 0.5) \quad i = 7, \ldots, 15$$

(see Figure 3). Since dependence is hard to reproduce and detect in small datasets, we use independent observations. Furthermore, we use the prior specification $(a, b, c, \phi) = (1, 1, 0.1, 0)$ for the integrated regime likelihood and various other specifications for the prior parameters of ρ_n. In particular, the σ-parameter was set to 0.0, 0.1, 0.3, 0.6 and 0.9 and then the corresponding values for θ were found such that the prior expected value for C_n matches 1, 5 and 11, namely true, far and too far from the real number of change points. The results are shown in Table 2.

The most illustrative results in this example are those when the prior guess at the number of change points is far from the number that generated the data. As σ increases, the probability assigned to the posterior mode of ρ_n increases, and, at the same time, the posterior mode of C_n shifts towards smaller values. Also observe that the posterior mode of ρ_n contains fewer groups as σ increases. In a similar way, the posterior probabilities $(\tau_i)_{i=2}^{15}$ (Equation 10), shown in Figure 3 for the case $\mathbb{E}[C_n] = 11$, reveal a reinforcement assigning lower probabilities to all the times except the correct one as σ increases. This is related to the reinforcement mechanism explained in Lijoi et al. (2007b).

Hence, this verifies the observations in Figure 1, namely the value for σ controls the prior and posterior knowledge of ρ_n and C_n. Meaning that if $\mathbb{E}[C_n]$ is far from the true value, a higher value for σ is preferred. In the above example this reinforcement forces the posterior mode of C_n to shift towards small values, as σ increases, since the prior was chosen such that $E[C_n] = 11$. However, the direction would be the opposite if the true one would be $C_n = 25$ instead of $C_n = 2$.

Now, we show the performance of the MCMC algorithm using three bigger simulated datasets. Concretely, the first dataset has changes in mean, the second has changes in

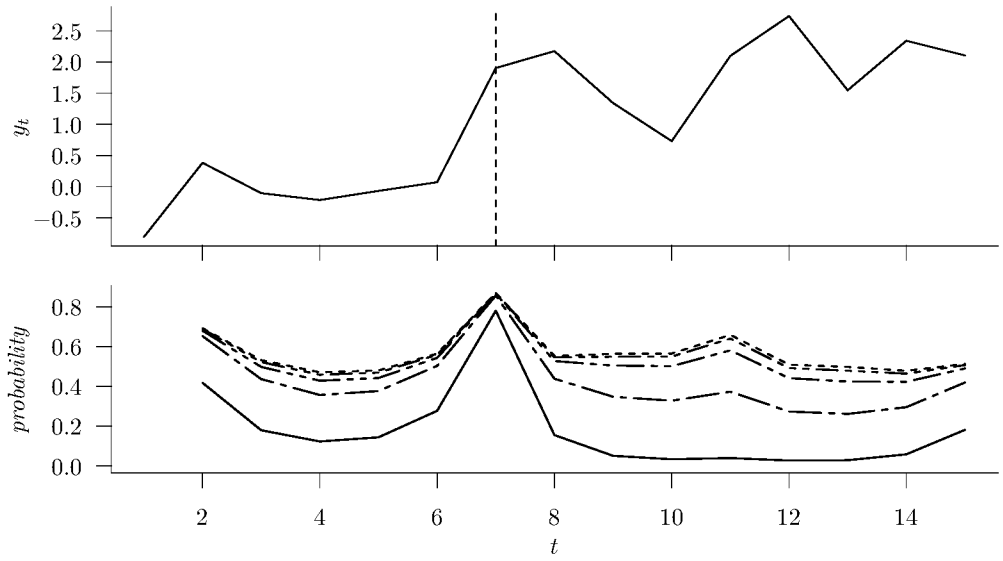


Figure 3: (top) Small simulated dataset with change point at $\langle 7 \rangle$, indicated with a dashed line. (bottom) Posterior probabilities that each time is a change point assuming $\mathbb{E}[C_n] = 11$ *a priori* and taking different values of σ: ---- 0.0, ---- 0.1, ---- 0.3, ---- 0.6 and ——— 0.9. Points are connected by straight lines for visual simplification.

variance and the last one has changes in mean, variance and correlation. Each of them consist of 150 observations and they are described as follows.

1. The first dataset has two changes in mean. This is sequentially simulated from independent Ornstein-Uhlenbeck processes as follows:

$$\begin{aligned} y_i &\sim \mathrm{OU}(0, 2, 0.4) \quad i = 1, \ldots, 50 \\ y_i &\sim \mathrm{OU}(5, 2, 0.4) \quad i = 51, \ldots, 85 \\ y_i &\sim \mathrm{OU}(2, 2, 0.4) \quad i = 86, \ldots, 150. \end{aligned}$$

   So its change points are $\langle 51, 86 \rangle$ (Figure 4a).

2. The second dataset has two changes in variance. Like the previous dataset, it is simulated as follows:

$$\begin{aligned} y_i &\sim \mathrm{OU}(0, 1.5, 0.7) \quad i = 1, \ldots, 50 \\ y_i &\sim \mathrm{OU}(0, 0.2, 0.7) \quad i = 51, \ldots, 120 \\ y_i &\sim \mathrm{OU}(0, 0.8, 0.7) \quad i = 121, \ldots, 150. \end{aligned}$$

   Its change points are $\langle 51, 121 \rangle$ (Figure 4b).

$\mathbb{E}[C_n]$	σ	θ	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
1	0.0	0.356	$\langle 7 \rangle$	0.6985	1	0.7986
	0.1	0.194	$\langle 7 \rangle$	0.6788	1	0.7776
	0.3	−0.114	$\langle 7 \rangle$	0.6301	1	0.7254
	0.6	−0.531	$\langle 7 \rangle$	0.5026	1	0.5868
	0.9	−0.890	$\langle\rangle$	0.8229	0	0.8229
5	0.0	3.201	$\langle 7 \rangle$	0.1457	2	0.3293
	0.1	2.626	$\langle 7 \rangle$	0.1749	2	0.3413
	0.3	1.527	$\langle 7 \rangle$	0.2507	2	0.3508
	0.6	0.097	$\langle 7 \rangle$	0.3923	1	0.4581
	0.9	−0.822	$\langle\rangle$	0.3564	1	0.3630
11	0.0	25.683	$\langle 2, 3, 4, 5, 6, 7 \rangle$	0.0017	8	0.2168
	0.1	22.670	$\langle 2, 3, 4, 5, 6, 7 \rangle$	0.0025	8	0.2103
	0.3	16.672	$\langle 2, 3, 4, 5, 6, 7 \rangle$	0.0052	7	0.1951
	0.6	7.832	$\langle 2, 3, 4, 5, 6, 7 \rangle$	0.0176	6	0.1799
	0.9	0.087	$\langle 7 \rangle$	0.2512	1	0.3179

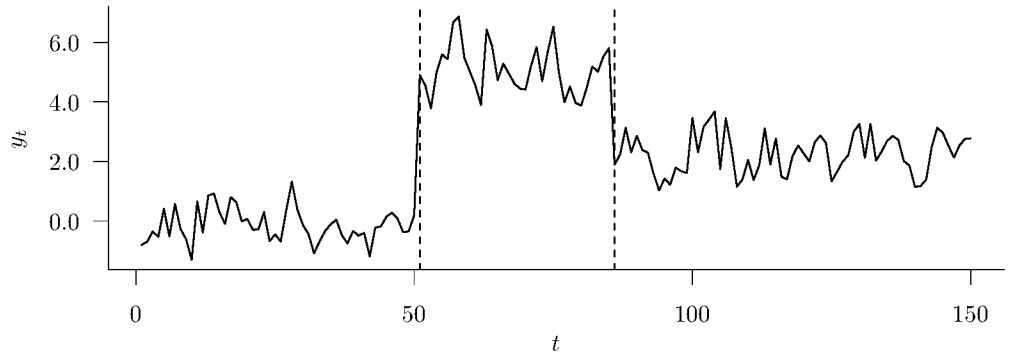
Table 2: Posterior results for the small dataset. The first three columns show the prior parameters for ρ_n. Columns 4 and 5 show the posterior mode, $\tilde{\rho}_n$, together with its probability. The last two columns show the mode for the number of change points, $\tilde{C}_n$, and its corresponding probability.

3. The third dataset has two change points and it is formed by the following independent samples:

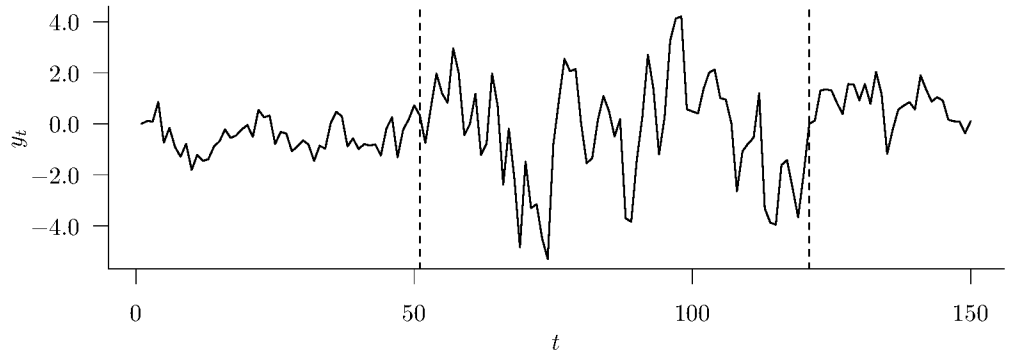
$$\begin{aligned} y_i &\sim \mathrm{OU}(0, 0.5, 0.1) \quad i = 1, \ldots, 50 \\ y_i &\sim \mathrm{OU}(-1, 2, 0.7) \quad i = 51, \ldots, 100 \\ y_i &\sim \mathrm{OU}(0, 1, 0.2) \quad i = 101, \ldots, 150. \end{aligned}$$

Then, its change points are $\langle 51, 101 \rangle$ (Figure 4c). This is a more complex dataset since it has changes in mean, variance and also in correlation. Even when our model uses a single parameter, ϕ, to control the correlation in the data, we want to test how well it works. In particular, the number of change points is expected to be different from the above, as, in principle, more of these might be needed to represent the above data under similar intra-regime dependency scenarios.

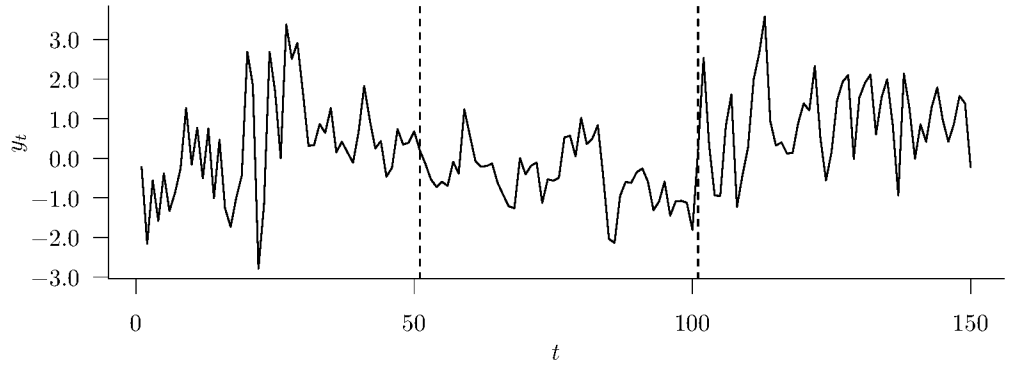
Various simulations were done using different specifications, each of them consisted of $20,000$ iterations after $10,000$ of burn-in with the integrated regime likelihood's prior parameters set as $a = b = 1$ and $c = 0.1$. The independent and dependent cases were considered for ϕ, letting $\phi = 0$ and $\phi \sim \mathrm{U}(0, 1)$, respectively. Regarding the prior parameters for ρ_n, like in the small dataset, parameter σ was set to 0.1, 0.3, 0.6 and 0.9 and their corresponding values for θ were found such that the prior expected value for C_n matches 2, 49 and 99. Additionally, we considered specifications where $\theta|\sigma \sim \mathrm{Ga}(1, 1, -\sigma)$ with $\sigma \sim \mathrm{Be}(1, 1)$ or fixed taking the aforementioned values.



(a) Changes in mean at times $\langle 51, 86 \rangle$.



(b) Changes in variance at times $\langle 51, 121 \rangle$.



(c) Changes in mean, variance and correlation at times $\langle 51, 101 \rangle$.

Figure 4: Simulated datasets; change points are indicated with dashed lines.

For the sake of comparison, we also show the results based on the model developed by Loschi and Cruz (2005), namely a PPM-based approach with Yao's cohesion function and independent inter- and intra-regime Gaussian likelihood. Within this approach, the posterior distribution of ρ_n is given by Equation (1), where $\mathbb{P}(\rho_n) \propto p^{k-1}(1-p)^{n-k}$ and $\mathbb{P}(y|\rho_n)$ is given by (9) with $\phi = 0$. Sampling from the corresponding posterior distribution is done using an extension of the Gibbs sampling scheme proposed by Barry and Hartigan (1993), see also Loschi et al. (2003). In our view, this is one of the most competitive models for change point problems available in the literature, apart from having certain features similar to our approach.

The simulations for this model were also done using different scenarios, taking $20,000$ iterations after a burn-in of $10,000$. Regime parameters and cohesion function's hyper-parameters, (α, β), were taken as described in their paper. Furthermore, hyper-parameters (α, β) were fixed such that the expected number of change points matches the same values as in our model.

The results are shown in the appendix, Tables 4–7. From these results, changes in mean are correctly detected for almost all configurations (Table 4). Both the locations and the number of the change points were correctly identified. The most notable difference related to the performance of ϕ is that the independent case assigns lower probabilities than those corresponding to the dependent case. Regarding the parameter σ, we notice that under a complete misspecification of the prior, i.e. $\mathbb{E}[C_n] = 49$ and $\sigma = 0.1$, neither the locations nor the number of change points were correctly identified. However, as σ increases the results become more favorable. Indeed, by inspecting more extreme cases, e.g. $\mathbb{E}[C_n] = 99$, only for $\sigma = 0.9$, the correct number of change points is recovered.

The second example, reported in Table 5, shows the improvement under a more robust model, i.e. allowing for dependency by setting $\phi \sim \mathrm{U}(0, 1)$. In such a scenario, most change points were correctly detected. Indeed, the small variations might be due to the randomness inherent to the simulation. In the independent case, however, even when the prior guess at the number of change points was close enough to the correct one, the model was unable to correctly identify them. This clearly favors the use of a model able to capture serial dependence. On the other hand, parameter σ shows a similar performance as in the previous dataset. However, in this case the cost of placing a misspecified prior, e.g. $\mathbb{E}[C_n] = 49$, tends to be higher.

For the third dataset, which has change points in mean, variance and correlation, the results are favored under the dependent case, having similar sensitivity conclusions for the parameter σ. See Table 6.

The results of Loschi and Cruz's method are shown in Table 7. Their model performs well for the dataset with changes in mean; hyper-parameters do not affect the results. For the dataset with changes in variances, an extra change point at $t = 4$ is detected in all scenarios; additionally, there are some scenarios including more change points. However, for the third dataset, their method could not detect any change point. This behavior might be due to the fact that their prior distribution $\mathbb{P}(\rho_n)$ is a function of n and k only, that is, giving the same weight to all compositions $(n_1, \ldots, n_k)$ for a given k.

So, we could say that our approach has an advantage over theirs. Our distribution of ρ_n provides more flexibility to detect change points. In particular, the parameter σ plays a key role since it enhances the detection under more uncertain scenarios. Moreover, we can also say that our distribution may be also used as an alternative cohesion function within PPMs restricted to $\mathcal{C}_n$.

Throughout this analysis, we used the modal partition as indicative of change points, however, as mentioned above, the least-squares clustering could also be employed. Indeed, we evaluated these two statistics, but, due to the ordering constraint present when working with set compositions, the resulting change points coincided in almost all scenarios.

6 Application to real data

In this section, we test our proposal with daily records of Mexican Peso/US Dollar exchange rate from January 2007 to December 2012 (Figure 5), available at the website `www.federalreserve.gov`. Like in the previous section, we compare our proposal with the PPM of Loschi and Cruz (2005), which has been widely used in change point detection in financial data; see, for example Zantedeschi et al. (2011).

Estimations are based on $20,000$ iterations after a burn-in of $10,000$. Integrated regime likelihood's parameters were set as before, i.e. $a = b = 1$ and $c = 0.1$, and, for ϕ, the dependent ($\phi \sim \mathrm{U}(0,1)$) and independent ($\phi = 0$) cases were considered. From the analysis made in the previous section, we fixed parameter $\sigma = 0.9$ to allow a strong reinforcement and let $\theta \sim \mathrm{Ga}(1, 1, -0.9)$. Regarding Loschi and Cruz's proposal, we test it under a similar scenario. Regime likelihood's parameters were set as the ones used in Loschi and Cruz (2005) and the cohesion function's hyper-parameters (α, β) were fixed to $(1, 1)$. For simplicity during the comparison, our dependent and independent cases are called model A and model B, respectively, whereas Loschi and Cruz's method is called model C. These simulations took 238, 66 and 361 seconds, respectively, using a PC with an Intel® Core™ i5-3570 CPU at 3.4GHz with 8Gb of RAM. In all cases, we performed the Gelman and Rubin (1992) visual convergence test and the Raftery and Lewis (1992) diagnostic convergence test; both indicated satisfactory results.

For model A, the most probable change points, occurring with probability 0.0245, are detected at: *September 10, 2008*, *May 4, 2009*, *July 16, 2010*, *August 4, 2011* and *August 3, 2012*. For model B, the modal change points, occurring with probability 0.116, are at: *April 1, 2008*, *October 7, 2008*, *January 14, 2009*, *April 2, 2009*, *November 24, 2009*, *January 3, 2011* and *September 8, 2011*. Finally, for model C, with probability 0.08465, the mode is detected at: *September 9, 2008*, *March 11, 2009*, *March 18, 2011* and *August 25, 2011*. Figure 5 indicates all these change points for each model and Figure 6 shows posterior probabilities $(\tau_i)_{i=2}^{1566}$ (Equation 10). Figure 7 exhibits the posterior distributions of θ, for Models A and B, and ϕ, for Model A. Additionally, Table 3 displays the posterior distribution of the number of change points.

These results share change points around relevant events in Mexico and USA: 1) the

2008 financial crisis with a sharp spike in September, 2) the flu pandemic suffered by Mexico in the period of March–May 2009, where the pharmaceutical industry tried to reactivate the global economy, and 3) the US debt-ceiling crisis in 2011. The rest of the change points detected by Model A are also close to other relevant events: 4) the so-called *currency war* in 2010 as a consequence of the 2008 financial crisis, and 5) the Mexican presidential election.

Both Loschi and Cruz's proposal and ours can be subject to interpretability. In particular, in our view, Model A, i.e. with a high value for σ and with positive correlation, provides a more sensible detection. Clearly, both approaches are model-based, one working with independent observations and the other with a stationary Markovian model.

7 Discussion

We have proposed a change point detection model which constitutes an alternative to other approaches available in the literature, in particular to the PPM-based approach. As mentioned in Section 2, these kinds of problems are essentially classification problems but with a restriction imposed by the sequential nature of the data. Indeed, the PPM approach shares some similarities with the posterior inference obtained under a Bayesian non parametric approach, but with the distinctive difference of having a restricted support on $\mathcal{C}_{[n]}$ instead of $\mathcal{P}_{[n]}$. One of the advantages of such a connection, and thus of our proposal, is that it allows to build up other classes of $\mathcal{C}_{[n]}$-valued distributions. In particular, take the EPPF of your preference and apply Definition 2, hence providing a wide set of alternatives to the cohesion function PPM-based approach. Notice that PPMs are, in general, not applicable to change point problems as these are usually also supported in $\mathcal{P}_{[n]}$.

Concerning the distribution of ρ_n, clearly, as we already said, there are many other ways to define it. In particular, the aforementioned alternative construction devised through the Markov chain $(K_n)_{n=1}^{\infty}$, which also resembles the approach via iHMMs, does not necessarily lead to an analytic form of the implied distribution on partitions and therefore the need of more elaborate MCMC schemes becomes evident.

Therefore, our approach could be considered as a coherent one and that goes along with the EPPF approach. In particular, it inherits some appealing marginal features such as the distribution of the number of groups, used here to model the number of change points.

Furthermore, the role of parameter σ in the two-parameter Poisson-Dirichlet process also resulted in an appealing feature of the EPPF used through this paper. Its ability to control the clustering, not achieved by other models, such as the Dirichlet process, led us to a change points sensitivity parameter.

An important point to emphasize is the model-based nature of change point detection models. In this work, we deal with stationary Markovian observations, however other modeling schemes could be also considered, while keeping the same idea.

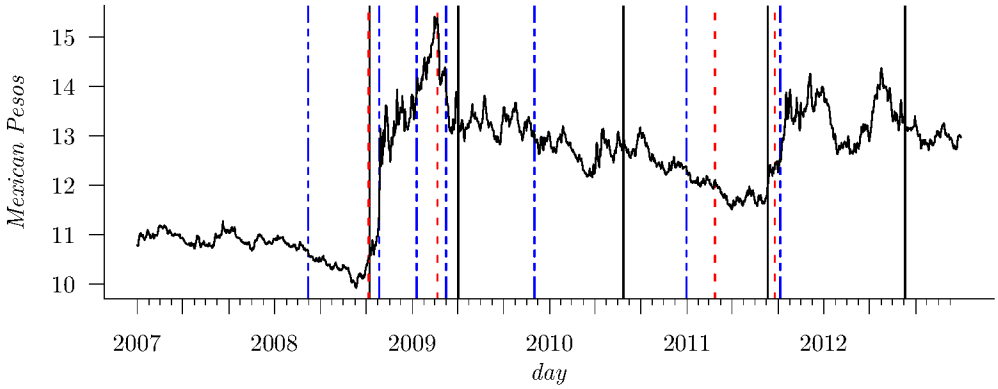


Figure 5: Price of one US Dollar in Pesos. Change points are indicated with vertical lines for the different models: A) —— our approach with $\phi \sim U(0,1)$, B) - - - our approach with $\phi = 0$, and C) - - - Loschi and Cruz's approach.

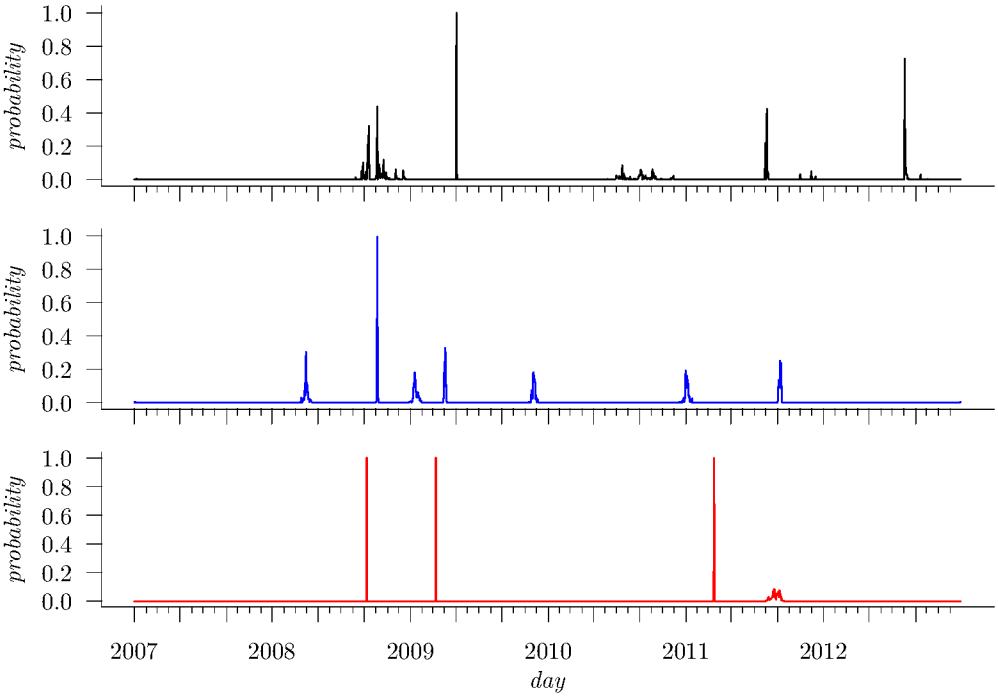
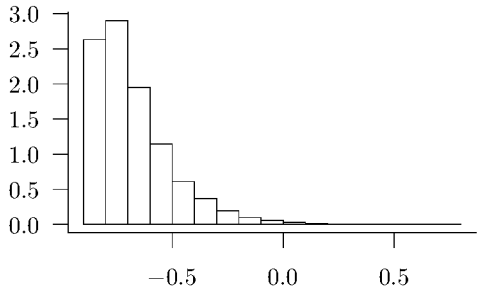


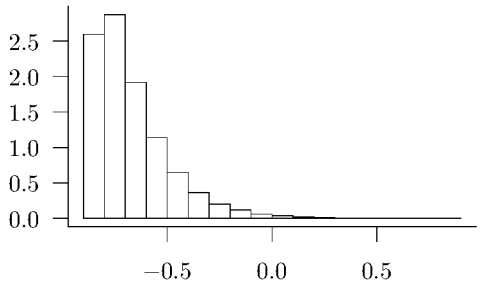
Figure 6: Posterior probabilities that each time is a change point for the different models: (top) our approach with $\phi \sim U(0,1)$, (middle) our approach with $\phi = 0$, and (bottom) Loschi and Cruz's approach.

Model	*Number of change points*					
	4	5	6	7	8	9
A: *ours with* $\phi \sim U(0,1)$	—	0.23265	0.23015	0.49245	0.04365	0.00110
B: *ours with* $\phi = 0$	—	—	—	0.83290	0.15695	0.01015
C: *Loschi and Cruz's*	0.00025	0.99975	—	—	—	—

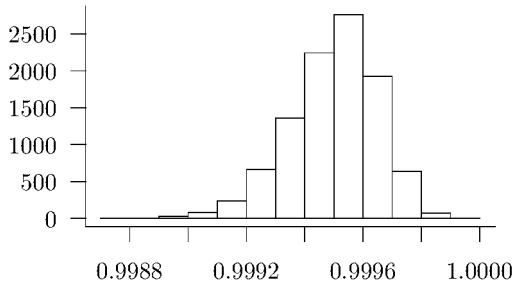
Table 3: Posterior probabilities for the number of change points, C_n, for each model.



(a) Parameter θ, model A



(b) Parameter θ, model B



(c) Parameter ϕ, model A

Figure 7: Posterior distributions for our approach with $\phi \sim U(0,1)$ (model A) and $\phi = 0$ (model B).

Various other extensions could be also evaluated. For example, one could restrict the partition distribution resulting from dependent nonparametric approaches, which would then include cases with inter-regime dependence. Also, by assuming a continuous time-dependent nonparametric Markovian regime, a generalization in the same direction of the approach by Muliere and Scarsini (1985) and Mira and Petrone (1996), could be achieved.

The above mentioned features of our approach seem to work well together for their main purpose: the detection of change points. This was shown in the exchange rate application where change points were indeed related to important events in the countries which affected them. Finally, an appealing aspect still to be explored, within the class of models here studied, is the prediction of change points which is part of our on-going research.

Finally, another appealing aspect devised from the ideas here presented is the proposed MCMC algorithm. Indeed, this can be easily extended to other $\mathcal{C}_{[n]}$-valued distributions, such as those mentioned in Section 2 and those implied from PPMs-based approaches.

References

Aggarwal, R., Inclán, C., and Leal, R. (1999). "Volatility in emerging stock markets." *Journal of Financial and Quantitative Analysis*, 34: 1–17. 825

Allen, D. E., McAleer, M., Powell, R. J., and Kumar-Singh, A. (2013). "Nonparametric Multiple Change Point Analysis of the Global Financial Crisis." URL http://dx.doi.org/10.2139/ssrn.2270029 825

Barry, D. and Hartigan, J. (1992). "Product partition models for change point problems." *The Annals of Statistics*, 20(1): 260–279. 824, 831

Barry, D. and Hartigan, J. A. (1993). "A Bayesian Analysis for Change Point Problems." *Journal of the American Statistical Association*, 88(421): 309–319. 824, 841

Benavides, G. and Capistrán, C. (2009). "Una Nota sobre la Volatilidad de la Tasa de Interés y del Tipo de Cambio bajo Diferentes Instrumentos de Política Monetaria: México 1998–2008." Working Paper 2009-10, Banco de México, México. 825

Bhattacharya, R. N., Gupta, V. K., and Waymire, E. (1983). "The Hurst effect under trends." *Journal of Applied Probability*, 20: 649–662. 825

Braun, J. V., Braun, R. K., and Müller, H. G. (2000). "Multiple Changepoint Fitting via Quasilikelihood, with Application to DNA Sequence Segmentation." *Biometrika*, 87(2): 301–314. 823

Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Change–Point Problems*. Springer. 823

Bubula, A. and Ötker-Robe, I. (2002). "The Evolution of Exchange Rate Regimes Since 1990: Evidence from De Facto Policies." Working Paper 02/155, International Monetary Fund. 825

Chen, J. and Gupta, A. K. (1997). "Testing and locating variance changepoints with application to stock prices." *Journal of the American Statistical Association*, 92(438): 739–747. 823

— (2011). *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhauser, 2nd edition. 823

Chernoff, H. and Zacks, S. (1964). "Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time." *Annals of Mathematical Statistics*, 35(3): 999–1018. 823

Dahl, D. B. (2006). "Model-Based Clustering for Expression Data via a Dirichlet Process Mixture Model." In Do, K.-A., Müller, P., and Vannucci, M. (eds.), *Bayesian Inference for Gene Expression and Proteomics*, 201–218. Cambridge University Press. 836

De Blasi, P., Favaro, S., Lijoi, A., Mena, R., Prünster, I., and Ruggiero, M. (2013). "Are Gibbs-Type Priors the Most Natural Generalization of the Dirichlet Process?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
URL http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.217 829

de Finetti, B. (1931). *Funzione Caratteristica Di un Fenomeno Aleatorio*, 251–299. 6. Memorie. Academia Nazionale del Linceo. 827

Dobigeon, N. and Toumeret, J. Y. (2007). "Joint segmentation of wind speed and direction using a hierarchical model." *Computational Statistics and Data Analysis*, 51: 5603–5621. 823

Fearnhead, P. and Liu, Z. (2007). "Online Inference for Multiple Changepoint Problems." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 69: 589–605. 826

Ferguson, T. (1973). "A Bayesian analysis of some nonparametric problems." *The Annals of Statistics*, 1(2): 209–230. 825

Fuentes-García, R., Mena, R. H., and Walker, S. G. (2010). "A Probability for Classification Based on the Dirichlet Process Mixture Model ." *Journal of Classification*, 27: 389–403. 824, 830, 832, 834

Gelman, A. and Rubin, D. B. (1992). "Inference from Iterative Simulation Using Multiple Sequences." *Statistical Science*, 7(4). 842

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). "Adaptive rejection Metropolis sampling." *Applied Statistics*, 44: 455–472. 835

Godsill, S. (2001). "On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty." *Journal of Computational and Graphical Statistics*, 10(2): 230–248. 851

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82(4): 711–732. 824

Harding, D. and Pagan, A. R. (2008). "Business cycle measurement." In Durlauf, S. N. and Blume, L. E. (eds.), *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2nd edition. 825

Jara, A., Lesaffre, E., De Iorio, M., and Quintana, F. (2010). "Bayesian semiparametric inference for multivariate doubly–interval–censored data." *Annals of Applied Statistics*, 4(4): 2126–2149. 836

Jochmann, M. (2010). "Modeling U.S. Inflation Dynamics: A Bayesian Nonparametric Approach." Technical Report 2010–06, Scottish Institute for Research in Economics (SIRE). 823

Kander, Z. and Zacks, S. (1966). "Test Procedures for Possible Changes in Parameters of Statistical Distributions Occurring at Unknown Time Points." *Annals of Mathematical Statistics*, 37(5): 1196–1210. 823

Kaplan, A. Y. and Shishkin, S. L. (2000). "Application of the change-point analysis to the investigation of the brain's electrical activity." In Brodsky, B. E. and Darkhovsky, B. S. (eds.), *Non-Parametric Statistical Diagnosis: Problems and Methods*, 333–388. Springer. 823

Karatzas, I. and Shreve, S. E. (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag. 833

Kingman, J. (1975). "Random discrete distributions." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 37(1): 1–22. 825

Lavielle, M. and Teyssière, G. (2007). "Adaptive Detection of Multiple Change-Points in Asset Price Volatility." In *Long Memory in Economics*, 129–156. Springer Berlin Heidelberg. 825

Lijoi, A., Mena, R. H., and Prünster, I. (2007a). "Bayesian nonparametric estimation of the probability of discovering a new species." *Biometrika*, 94: 769–786. 829, 831

— (2007b). "Controlling the reinforcement in Bayesian non-parametric mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4): 715–740. 825, 829, 830, 837

Lijoi, A. and Prünster, I. (2010). "Models beyond the Dirichlet process." In Hjort, N. L., Holmes, C. C., Müller, P., and Walker, S. G. (eds.), *Bayesian Nonparametrics*, 80–136. Cambridge University Press. 828

Loschi, R. H. and Cruz, F. R. B. (2005). "Extension to the product partition model: computing the probability of a change." *Computational Statistics & Data Analysis*, 48(2): 255–268. 824, 836, 841, 842

Loschi, R. H., Cruz, F. R. B., Iglesias, P. L., and Arellano-Valle, R. B. (2003). "A Gibbs sampling scheme to the product partition model: an application to change point problems." *Computers & Operations Research*, 30(3): 463–482. 824, 841

Mikosch, T. and St**u**aric**u**a, C. (2004). "Changes of structure in financial time series and the GARCH model." *REVSTAT - Statistical Journal*, 2(1): 41–73. 825

Minin, V. N., Dorman, K. S., Fang, F., and Suchard, M. A. (2007). "Phylogenetic Mapping of Recombination Hotspots in Human Immunodeficiency Virus via Spatially Smoothed Change-Point Processes." *Genetics*, 175(4): 1773–1785. 823

Mira, A. and Petrone, S. (1996). "Bayesian hierarchical nonparametric inference for change–point problems." In Bernardo, J., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 5*, 693–703. Oxford University Press. 824, 846

Monteiro, J. V. D., Assunção, R., and Loschi, R. H. (2011). "Product partition models with correlated parameters." *Bayesian Analysis*, 6(4): 691–726. 826, 831

Muliere, P. and Scarsini, M. (1985). "Change–point problems: A Bayesian nonparametric approach." *Aplikace matematiky*, 30(6): 397–402. 824, 846

Müller, P., Quintana, F., and Rosner, G. L. (2011). "A Product Partition Model With Regression on Covariates." *Journal of Computational and Graphical Statistics*, 20(1): 260–278. 824

Nieto-Barajas, L. E. and Contreras-Cristan, A. (2014). "A Bayesian Nonparametric Approach for Time Series Clustering." *Bayesian Analysis*, 9(1): 147–170. 836

Park, J.-H. and Dunson, D. B. (2010). "Bayesian generalized product partition model." *Statistica Sinica*, 20(20): 1203–1226. 824

Perman, M., Pitman, J., and Yor, M. (1992). "Size–biased sampling of Poisson point processes and excursions." *Probability Theory and Related Fields*, 92: 21–39. 825

Perreault, L., Bernier, J., Bobée, B., and Parent, E. (2000). "Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited." *Journal of Hydrology*, 235(3–4): 221–241. 823

Pitman, J. (2006). *Combinatorial stochastic processes*. Ecole d'eté de probabilités de Saint-Flour XXXII - 2002. Springer. 828

Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. J. (2002). "Bayesian curve fitting using MCMC with applications to signal segmentation." *IEEE Transactions on Signal Processing*, 50: 747–758. 823

Quintana, F. A. and Iglesias, P. L. (2003). "Bayesian Clustering and Product Partition Models." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 65(2): 557–574. 824

Raftery, A. E. and Lewis, S. M. (1992). "How many iterations in the Gibbs sampler?" In Bernardo, J., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, 763–773. Oxford University Press. 842

Regazzini, E., Lijoi, A., and Prünster, I. (2003). "Distributional results for means of normalized random measures with independent increments." *Annals of Statistics*, 31: 560–585. 828

Reinhart, C. M. and Rogoff, K. S. (2004). "The Modern History of Exchange Rate Arrangements: A Reinterpretation." *The Quarterly Journal of Economics*, 119(1): 1–48. 825

Smith, A. F. M. (1975). "A Bayesian approach to inference about a change–point in a sequence of random variables." *Biometrika*, 62(2): 407–416. 824

Uhlenbeck, G. E. and Ornstein, L. (1930). "On the theory of Brownian motion." *Physical Review*, 36(36): 823–841. 825, 833

West, M. (1992). "Hyperparameter estimation in Dirichlet process mixture models." Technical report, Institute of Statistics and Decision Sciences, Duke University. 853

Yao, Y.-C. (1984). "Estimation of a Noisy Discrete-Time Step Function: Bayes and Empirical Bayes Approaches." *Annals of Statistics*, 12(4): 1434–1447. 824

Zantedeschi, D., Damien, P., and Polson, N. G. (2011). "Predictive Macro-Finance With Dynamic Partition Models." *Journal of the American Statistical Association*, 106(494): 427–439. 842

Acknowledgments

The authors are grateful to the Editor, an Associate Editor, two referees and the Production Editor, Kary Myers, for their constructive comments and suggestions that have led to a substantial improvement of the manuscript. The first author was supported by PAPIIT project IN100411 and the second author by CONACyT project 131179.

Appendix

Simulation algorithm

Let $p(k, \gamma_k) = p(n_1, \ldots, n_k|y)$ where γ_k is a grouping for the data y of size k with each group having size n_j. Then, at each iteration, the MCMC algorithm will update values for k and γ_k.

The support for k is $\{1, \ldots, n\}$. When updating this variable, the chain will change the grouping from one with $k^{(i)}$ groups to another with $k^{(i+1)}$, involving perhaps a change in dimension. To avoid this issue, a series of latent variables, γ_j, $j = 1, \ldots, n$ and $j \neq k$, is introduced (Godsill 2001). The resulting density is written as

$$p(k, \gamma) = p(k, \gamma_k) \prod_{j=1}^{k-1} p(\gamma_j|\gamma_{j+1}) \prod_{j=k+1}^{n} p(\gamma_j|\gamma_{j-1}). \tag{11}$$

Updating k is done via Metropolis-Hastings with target distribution $p(k|\gamma)$. Let $0 < q < 1$ and let

$$p(k|r) = q\,\mathbb{I}(k = r + 1) + (1 - q)\,\mathbb{I}(k = r - 1) \tag{12a}$$

be the proposal distribution of k, whenever $1 < r < n$, and

$$p(2|1) = p(n - 1|n) = 1, \tag{12b}$$

otherwise. Thus, the acceptance probability to update k is given by

$$\alpha = \min\left\{\frac{p(k|k')}{p(k'|k)}\frac{p(k'|\gamma)}{p(k|\gamma)}, 1\right\},$$

with k' simulated from the proposal (12).

For a given k, there are only two possible values for updating it: $k + 1$ and $k - 1$, namely split and merge moves, respectively. These moves are related to the conditional distributions in (11), since $p(\gamma_{k+1}|\gamma_k)$ represents a split move and $p(\gamma_{k-1}|\gamma_k)$ a merge move. So it is necessary to be able to simulate from them.

When a split is performed, one group with size greater than one is selected and, then, split into two. These two choices are made uniformly, so

$$p(\gamma_{k+1}|\gamma_k) = \frac{1}{n_{g,k}(n_s - 1)}, \tag{13}$$

where $n_{g,k}$ is the number of groups of size greater than one and n_s the size of the selected group. On the other hand, when a merge is performed, two adjacent groups are selected to be merged. Again, they are chosen uniformly, so

$$p(\gamma_{k-1}|\gamma_k) = \frac{1}{k - 1}. \tag{14}$$

Therefore, updating k is done as follows. When a split move is proposed, i.e. $k' = k+1$, Equations (11–14) simplify α to

$$\alpha = \min\left\{\frac{1-q}{q}\frac{p(k+1,\gamma_{k+1})}{p(k,\gamma_k)}\frac{n_{g,k}(n_s-1)}{k}, 1\right\}, \tag{15a}$$

whenever $1 < k < n$, and

$$\alpha = \min\left\{(1-q)\,(n-1)\frac{p(2,\gamma_2)}{p(1,\gamma_1)}, 1\right\}, \tag{15b}$$

if $k = 1$. Otherwise, a merge move is proposed, i.e. $k' = k-1$. Then two adjacent groups are selected, say n_s and n_{s+1}, so Equations (11–14) simplify α to

$$\alpha = \min\left\{\frac{q}{1-q}\frac{p(k-1,\gamma_{k-1})}{p(k,\gamma_k)}\frac{k-1}{n_{g,k-1}(n_s+n_{s+1}-1)}, 1\right\}, \tag{16a}$$

whenever $1 < k < n$, and

$$\alpha = \min\left\{q\,(n-1)\frac{p(n-1,\gamma_{n-1})}{p(n,\gamma_n)}, 1\right\}, \tag{16b}$$

when $k = n$.

A second step, called shuffle, is included in the MCMC algorithm in order to improve it. This step takes two adjacent groups, say n_s and n_{s+1}, and updates their sizes uniformly by splitting the combined group into two, of sizes n_s^* and n_{s+1}^*, with $n_s^*, n_{s+1}^* \geq 1$. In this case, the acceptance probability is given by

$$\alpha = \min\left\{\frac{p(k,\gamma_k^*)}{p(k,\gamma_k)}\frac{(n_s^*+n_{s+1}^*-1)}{(n_s+n_{s+1}-1)}, 1\right\} = \min\left\{\frac{p(k,\gamma_k^*)}{p(k,\gamma_k)}, 1\right\}. \tag{17}$$

Proof of Proposition 1

The posterior distribution of θ is given by

$$p(\theta|\dots) \propto \frac{(\theta+\sigma)^{c-1}e^{-d\theta}}{(\theta+1)_{n-1\uparrow}}\prod_{i=1}^{k-1}(\theta+i\sigma)\mathbb{I}(\theta > -\sigma). \tag{18}$$

To derive the augmented full conditional distribution, notice that

$$\prod_{i=1}^{k-1}(\theta+i\sigma) = \sigma^{k-1}\sum_{j=0}^{k-1}c_{k-1,j}\Big(\frac{\theta+\sigma}{\sigma}\Big)^j,$$

where $c_{k-1,j}$ is the absolute value of the Stirling numbers of the first kind, and that

$$\frac{1}{(\theta+1)_{n-1\uparrow}} = \frac{\Gamma(\theta+1)}{\Gamma(\theta+n)} = \frac{(\theta+n)(\theta+n+1)}{(\theta+1)\Gamma(n)}\beta(\theta+2,n),$$

where β is the Beta function. Therefore, the posterior distribution (18) can be rewritten as

$$\sum_{j=0}^{k-1} \frac{c_{k-1,j}}{\sigma^j} \frac{(\theta+n)(\theta+n+1)}{(\theta+1)} \beta(\theta+2,n)(\theta+\sigma)^{c+j-1} e^{-d\theta} \mathbb{I}(\theta > -\sigma).$$

Using data augmentation, let $y \sim \text{Exp}(\theta+1)$ and, following the approach of West (1992) to update the Dirichlet process's parameter, let $z \sim \text{Be}(\theta+2, n)$. Thus,

$$p(\theta|y,z,\dots) \propto \sum_{j=0}^{k-1} \frac{c_{k-1,j}}{\sigma^j} (\theta+n)(\theta+n+1)(\theta+\sigma)^{c+j-1} e^{-(d+y-\log(z))\theta} \mathbb{I}(\theta > -\sigma).$$

Expanding the product $(\theta+n)(\theta+n+1)$ as function of $(\theta+\sigma)$ and rearranging the terms of the sum with respect to $(\theta+\sigma)^{c+j-1}$, $j = 0, 1, \dots, k+1$, we have

$$p(\theta|y,z,\dots) \propto \sum_{j=0}^{k+1} w'_j (\theta+\sigma)^{c+j-1} e^{-(d+y-\log(z))\theta} \mathbb{I}(\theta > -\sigma),$$

where

$$w'_j = \frac{(n-\sigma)(n+1-\sigma)c_{k-1,j} + (2n+1-2\sigma)\sigma c_{k-1,j-1} + \sigma^2 c_{k-1,j-2}}{\sigma^j},$$

with $c_{k,r} = 0$ for $r \leq 0$ and for $r > k$. Finally, completing the density in each term leads us to the stated result.

Simulation results for the synthetic datasets

The results of the simulations performed for each dataset described in Section 5 are shown in Tables 4, 5 and 6.

ϕ	$\mathbb{E}[C_n]$	σ	θ	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
0.0	2	0.1	0.1897	$\langle 51, 86\rangle$	0.7880	2	0.7966
		0.3	−0.1634	$\langle 51, 86\rangle$	0.7379	2	0.7454
		0.6	−0.5709	$\langle 51, 86\rangle$	0.6757	2	0.6757
		0.9	−0.8979	$\langle 51, 86\rangle$	0.6130	2	0.6142
	49	0.1	21.4127	$\langle 2, 50, 51, 86, 87, 149, 150\rangle$	0.0008	15	0.1195
		0.3	13.0593	$\langle 51, 86, 87\rangle$	0.0035	9	0.1267
		0.6	3.0021	$\langle 51, 86\rangle$	0.1969	3	0.2842
		0.9	−0.7974	$\langle 51, 86\rangle$	0.6069	2	0.6069
	99	0.1	113.4390	$\langle \dots \rangle_{40}$	0.0003	43	0.0700
		0.3	80.8346	$\langle \dots \rangle_{37}$	0.0004	35	0.0891
		0.6	34.2148	$\langle \dots \rangle_{10}$	0.0008	18	0.0866
		0.9	0.4399	$\langle 51, 86\rangle$	0.4056	2	0.4127
		0.1	*r.v.*	$\langle 51, 86\rangle$	0.7234	2	0.7235
		0.3	*r.v.*	$\langle 51, 86\rangle$	0.6558	2	0.6603
		0.6	*r.v.*	$\langle 51, 86\rangle$	0.5905	2	0.5927
		0.9	*r.v.*	$\langle 51, 86\rangle$	0.5354	2	0.5359
		r.v.	*r.v.*	$\langle 51, 86\rangle$	0.7030	2	0.7030
r.v.	2	0.1	0.1897	$\langle 51, 86\rangle$	0.8134	2	0.8134
		0.3	−0.1634	$\langle 51, 86\rangle$	0.7908	2	0.7908
		0.6	−0.5709	$\langle 51, 86\rangle$	0.7579	2	0.7579
		0.9	−0.8979	$\langle 51, 86\rangle$	0.7631	2	0.7634
	49	0.1	21.4127	$\langle \dots \rangle_{11}$	0.0004	13	0.1416
		0.3	13.0593	$\langle 51, 86\rangle$	0.0071	8	0.1436
		0.6	3.0021	$\langle 51, 86\rangle$	0.3486	2	0.3486
		0.9	−0.7974	$\langle 51, 86\rangle$	0.7531	2	0.7534
	99	0.1	113.4390	$\langle \dots \rangle_{32}$	0.0004	36	0.0946
		0.3	80.8346	$\langle \dots \rangle_{24}$	0.0005	28	0.0893
		0.6	34.2148	$\langle \dots \rangle_{18}$	0.0008	14	0.1205
		0.9	0.4399	$\langle 51, 86\rangle$	0.6336	2	0.6339
		0.1	*r.v.*	$\langle 51, 86\rangle$	0.7414	2	0.7414
		0.3	*r.v.*	$\langle 51, 86\rangle$	0.7046	2	0.7046
		0.6	*r.v.*	$\langle 51, 86\rangle$	0.7139	2	0.7139
		0.9	*r.v.*	$\langle 51, 86\rangle$	0.7632	2	0.7632
		r.v.	*r.v.*	$\langle 51, 86\rangle$	0.7162	2	0.7162

Table 4: Posterior results for the simulated dataset with changes in mean (Figure 4a). The first four columns show the prior specifications; the label *r.v.* means that a prior distribution was assigned. The last four columns show the modal change points, $\tilde{\rho}_n$, and the modal number of change points, $\tilde{C}_n$, together with their corresponding probabilities. The modal change points indicated as $\langle \dots \rangle_k$ denote that there are k of them.

ϕ	$\mathbb{E}[C_n]$	σ	θ	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
0.0	2	0.1	0.1897	$\langle 50, 66, 75, 113, 121\rangle$	0.0094	6	0.2164
		0.3	−0.1634	$\langle 53, 66, 76, 113, 121\rangle$	0.0022	9	0.0927
		0.6	−0.5709	$\langle 54, 55, 111, 112, 113, 121\rangle$	0.0014	19	0.0602
		0.9	−0.8979	$\langle \dots \rangle_{10}$	0.0018	23	0.0376
	49	0.1	21.4127	$\langle \dots \rangle_{21}$	0.0005	33	0.1058
		0.3	13.0593	$\langle \dots \rangle_{29}$	0.0006	31	0.1034
		0.6	3.0021	$\langle \dots \rangle_{26}$	0.0007	26	0.0887
		0.9	−0.7974	$\langle 53, 54, 112, 113, 121\rangle$	0.0034	10	0.0418
	99	0.1	113.4390	$\langle \dots \rangle_{67}$	0.0004	65	0.0818
		0.3	80.8346	$\langle \dots \rangle_{58}$	0.0004	59	0.0827
		0.6	34.2148	$\langle \dots \rangle_{55}$	0.0005	51	0.0767
		0.9	0.4399	$\langle \dots \rangle_{15}$	0.0019	37	0.0583
		0.1	*r.v.*	$\langle 49, 66, 75, 97, 99, 108, 121\rangle$	0.0013	11	0.1011
		0.3	*r.v.*	$\langle \dots \rangle_{13}$	0.0008	15	0.1054
		0.6	*r.v.*	$\langle \dots \rangle_{24}$	0.0008	24	0.0814
		0.9	*r.v.*	$\langle \dots \rangle_{36}$	0.0014	33	0.0486
		r.v.	*r.v.*	$\langle \dots \rangle_{11}$	0.0010	17	0.0746
r.v.	2	0.1	0.1897	$\langle 53, 121\rangle$	0.0559	2	0.4422
		0.3	−0.1634	$\langle 53, 121\rangle$	0.0452	2	0.4258
		0.6	−0.5709	$\langle 53, 121\rangle$	0.0434	2	0.4341
		0.9	−0.8979	$\langle 53, 121\rangle$	0.0486	2	0.5111
	49	0.1	21.4127	$\langle \dots \rangle_{23}$	0.0006	27	0.1013
		0.3	13.0593	$\langle \dots \rangle_{21}$	0.0006	21	0.0891
		0.6	3.0021	$\langle 53, 123\rangle$	0.0086	5	0.1640
		0.9	−0.7974	$\langle 53, 121\rangle$	0.0413	2	0.4763
	99	0.1	113.4390	$\langle \dots \rangle_{67}$	0.0005	66	0.0784
		0.3	80.8346	$\langle \dots \rangle_{71}$	0.0005	60	0.0721
		0.6	34.2148	$\langle \dots \rangle_{39}$	0.0008	49	0.0578
		0.9	0.4399	$\langle 53, 121\rangle$	0.0290	2	0.3403
		0.1	*r.v.*	$\langle 53, 121\rangle$	0.0437	2	0.3218
		0.3	*r.v.*	$\langle 53, 121\rangle$	0.0340	2	0.2949
		0.6	*r.v.*	$\langle 53, 121\rangle$	0.0263	2	0.3730
		0.9	*r.v.*	$\langle 53, 121\rangle$	0.0379	2	0.4616
		r.v.	*r.v.*	$\langle 53, 121\rangle$	0.0405	2	0.3407

Table 5: Posterior results for the simulated dataset with changes in variance (Figure 4b). The first four columns show the prior specifications; the label *r.v.* means that a prior distribution was assigned. The last four columns show the modal change points, $\tilde{\rho}_n$, and the modal number of change points, $\tilde{C}_n$, together with their corresponding probabilities. The modal change points indicated as $\langle \dots \rangle_k$ denote that there are k of them.

ϕ	$\mathbb{E}[C_n]$	σ	θ	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
0.0	2	0.1	0.1897	$\langle 2, 51, 103\rangle$	0.0168	5	0.3322
		0.3	−0.1634	$\langle 2, 51, 82, 91, 101\rangle$	0.0107	5	0.1684
		0.6	−0.5709	$\langle 2, 51, 103\rangle$	0.0291	3	0.1588
		0.9	−0.8979	$\langle\rangle$	0.9071	0	0.9071
	49	0.1	21.4127	$\langle\ldots\rangle_{32}$	0.0004	29	0.0929
		0.3	13.0593	$\langle\ldots\rangle_{19}$	0.0006	24	0.0817
		0.6	3.0021	$\langle 2, 3, 50, 51, 103, 150\rangle$	0.0021	14	0.1111
		0.9	−0.7974	$\langle 2\rangle$	0.5550	1	0.5682
	99	0.1	113.4390	$\langle\ldots\rangle_{71}$	0.0004	64	0.0756
		0.3	80.8346	$\langle\ldots\rangle_{60}$	0.0004	62	0.0731
		0.6	34.2148	$\langle\ldots\rangle_{42}$	0.0005	44	0.0798
		0.9	0.4399	$\langle 2\rangle$	0.0943	7	0.1109
		0.1	*r.v.*	$\langle 2, 51, 103\rangle$	0.0075	6	0.1779
		0.3	*r.v.*	$\langle 2, 51, 82, 91, 101\rangle$	0.0088	6	0.1837
		0.6	*r.v.*	$\langle 2\rangle$	0.0107	8	0.1267
		0.9	*r.v.*	$\langle 2\rangle$	0.1454	1	0.1477
		r.v.	*r.v.*	$\langle 2, 51, 103\rangle$	0.0129	6	0.1807
r.v.	2	0.1	0.1897	$\langle 51, 108\rangle$	0.0846	2	0.4178
		0.3	−0.1634	$\langle 51, 108\rangle$	0.0450	2	0.3212
		0.6	−0.5709	$\langle\rangle$	0.3802	0	0.3802
		0.9	−0.8979	$\langle\rangle$	0.9946	0	0.9946
	49	0.1	21.4127	$\langle\ldots\rangle_{20}$	0.0008	24	0.0909
		0.3	13.0593	$\langle\ldots\rangle_{10}$	0.0006	18	0.0998
		0.6	3.0021	$\langle 51, 108, 118, 120\rangle$	0.0056	7	0.1532
		0.9	−0.7974	$\langle\rangle$	0.5752	0	0.5752
	99	0.1	113.4390	$\langle\ldots\rangle_{67}$	0.0004	64	0.0811
		0.3	80.8346	$\langle\ldots\rangle_{55}$	0.0004	62	0.0649
		0.6	34.2148	$\langle\ldots\rangle_{23}$	0.0004	36	0.0612
		0.9	0.4399	$\langle 2\rangle$	0.2874	1	0.3665
		0.1	*r.v.*	$\langle 51, 108\rangle$	0.0464	2	0.2493
		0.3	*r.v.*	$\langle 51, 108\rangle$	0.0357	2	0.2212
		0.6	*r.v.*	$\langle 51, 108\rangle$	0.0345	3	0.2274
		0.9	*r.v.*	$\langle\rangle$	0.5786	0	0.5786
		r.v.	*r.v.*	$\langle 51, 103\rangle$	0.0254	2	0.2599

Table 6: Posterior results for the simulated dataset with changes in mean, variance and correlation (Figure 4c). The first four columns show the prior specifications; the label *r.v.* means that a prior distribution was assigned. The last four columns show the modal change points, $\tilde{\rho}_n$, and the modal number of change points, $\tilde{C}_n$, together with their corresponding probabilities. The modal change points indicated as $\langle\ldots\rangle_k$ denote that there are k of them.

$\mathbb{E}[C_n]$	α	β	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
2	1.00	73.50	$\langle 51, 86\rangle$	0.9900	2	0.9901
	0.01	1.00	$\langle 51, 86\rangle$	0.9924	2	0.9925
49	1.00	2.04	$\langle 51, 86\rangle$	0.9848	2	0.9849
	0.49	1.00	$\langle 51, 86\rangle$	0.9891	2	0.9892
99	1.00	0.51	$\langle 51, 86\rangle$	0.9847	2	0.9848
	1.98	1.00	$\langle 51, 86\rangle$	0.9773	2	0.9774
75	1	1	$\langle 51, 86\rangle$	0.9847	2	0.9848
3	1	50	$\langle 51, 86\rangle$	0.9900	2	0.9891
14	5	50	$\langle 51, 86\rangle$	0.9660	2	0.9662
146	50	1	$\langle 51, 86\rangle$	0.6242	2	0.6244
136	50	5	$\langle 51, 86\rangle$	0.6347	2	0.6349

(a) Results for the dataset with changes in mean (Figure 4a).

$\mathbb{E}[C_n]$	α	β	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
2	1.00	73.50	$\langle 4, 53, 121\rangle$	0.0456	3	0.3674
	0.01	1.00	$\langle 4, 53, 121\rangle$	0.0463	4	0.3379
49	1.00	2.04	$\langle 4, 53, 121\rangle$	0.0280	4	0.3494
	0.49	1.00	$\langle 4, 53, 121\rangle$	0.0365	4	0.3519
99	1.00	0.51	$\langle 4, 53, 121\rangle$	0.0275	4	0.3471
	1.98	1.00	$\langle 4, 53, 121, 123, 146\rangle$	0.0335	4	0.3332
75	1	1	$\langle 4, 53, 121, 123, 146\rangle$	0.0275	4	0.3474
3	1	50	$\langle 4, 53, 121\rangle$	0.0362	4	0.3595
14	5	50	$\langle 4, 50, 121, 123, 146\rangle$	0.0257	5	0.3485
146	50	1	$\langle 4, 53, 121, 123, 146, 149, 150\rangle$	0.0143	7	0.2914
136	50	5	$\langle 4, 53, 121, 123, 146, 149, 150\rangle$	0.0152	7	0.2970

(b) Results for the dataset with changes in variance (Figure 4b).

$\mathbb{E}[C_n]$	α	β	$\tilde{\rho}_n$	*prob.*	$\tilde{C}_n$	*prob.*
2	1.00	73.50	$\langle\rangle$	1.0000	0	1.0000
	0.01	1.00	$\langle\rangle$	1.0000	0	1.0000
49	1.00	2.04	$\langle\rangle$	1.0000	0	1.0000
	0.49	1.00	$\langle\rangle$	1.0000	0	1.0000
99	1.00	0.51	$\langle\rangle$	1.0000	0	1.0000
	1.98	1.00	$\langle\rangle$	0.9993	0	0.9993
75	1	1	$\langle\rangle$	1.0000	0	1.0000
3	1	50	$\langle\rangle$	1.0000	0	1.0000
14	5	50	$\langle\rangle$	0.9971	0	0.9971
146	50	1	$\langle\rangle$	0.6214	0	0.6214
136	50	5	$\langle\rangle$	0.6050	0	0.6050

(c) Results for the dataset with changes in mean, variance and correlation (Figure 4c).

Table 7: Posterior results for the simulated datasets using Loschi and Cruz's method. For each table, the first three columns show the prior specifications and the last four columns show the modal change points, $\tilde{\rho}_n$, and the modal number of change points, $\tilde{C}_n$, together with their corresponding probabilities.