Computational Statistics and Data Analysis 71 (2014) 417–433

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Dependent mixture models: Clustering and borrowing information

Antonio Lijoi [a,c], Bernardo Nipoti [b,c], Igor Prünster [b,c,*]

[a] Department of Economics and Management, University of Pavia, Via S. Felice 5, 27100 Pavia, Italy

[b] Department of Economics and Statistics, University of Torino, C.so Unione Sovietica 218/bis, 10134 Torino, Italy

[c] Collegio Carlo Alberto, via Real Collegio 30, 10024 Moncalieri, Italy

ARTICLE INFO

Article history:
Received 19 November 2012
Received in revised form 16 June 2013
Accepted 18 June 2013
Available online 25 June 2013

Keywords:
Bayesian nonparametrics
Dependent process
Dirichlet process
Generalized Pólya urn scheme
Mixture models
Normalized σ-stable process
Partially exchangeable random partition

ABSTRACT

Most of the Bayesian nonparametric models for non-exchangeable data that are used in applications are based on some extension to the multivariate setting of the Dirichlet process, the best known being MacEachern's dependent Dirichlet process. A comparison of two recently introduced classes of vectors of dependent nonparametric priors, based on the Dirichlet and the normalized σ-stable processes respectively, is provided. These priors are used to define dependent hierarchical mixture models whose distributional properties are investigated. Furthermore, their inferential performance is examined through an extensive simulation study. The models exhibit different features, especially in terms of the clustering behavior and the borrowing of information across studies. Compared to popular Dirichlet process based models, mixtures of dependent normalized σ-stable processes turn out to be a valid choice being capable of more effectively detecting the clustering structure featured by the data.

1. Introduction

Bayesian inference, either in parametric or nonparametric form, is commonly based on the assumption that the observations $X_1, \ldots, X_n$ are drawn from an exchangeable sequence of random elements $(X_i)_{i\geq 1}$. This means that, for any n, the distribution of the vector $(X_1, \ldots, X_n)$ is invariant with respect to permutations of its components. Such an assumption reflects an idea of analogy or homogeneity of the data and forms the basis for predictive inference. Furthermore, it is nicely translated into a property of conditional independence and identity in distribution by virtue of the de Finetti representation theorem, namely

$$\begin{aligned} &X_i \mid \tilde{p} \overset{\text{i.i.d.}}{\sim} \tilde{p}, \quad i = 1, \ldots, n, \\ &\tilde{p} \sim Q, \end{aligned} \tag{1}$$

where $\tilde{p}$ is some random probability measure whose distribution Q plays the role of a prior for Bayesian inference.

It is apparent that exchangeability of observations is a strong assumption that fails in many problems of practical interest. This is the case, for instance, when the data originate from different studies or refer to experiments performed under different conditions: in such a context it is reasonable to preserve the homogeneity condition within data that are generated from the same study or experimental condition, while, at the same time, dropping the conditional identity in distribution for

* Corresponding author at: Department of Economics and Statistics, University of Torino, C.so Unione Sovietica 218/bis, 10134 Torino, Italy. Tel.: +39 011 6705281.

E-mail addresses: lijoi@unipv.it (A. Lijoi), bernardo.nipoti@unito.it (B. Nipoti), igor.pruenster@unito.it (I. Prünster).

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.csda.2013.06.015

data emerging from different studies/experiments. Recent literature in Bayesian nonparametric inference has addressed this issue by proposing models that can accommodate for more general forms of dependence than exchangeability. Most of the proposals rely on the notion of partial exchangeability, as set forth by de Finetti (1938), that formalizes the above idea: although not valid across the whole set of observations, exchangeability can hold true within k separate subgroups of observations. Here, for ease of exposition and with no loss of generality, we confine ourselves to considering the case where $k = 2$. More formally, let $\mathbb{X}$ be a complete and separable metric space whose Borel σ-algebra is henceforth denoted as $\mathscr{X}$ and let $P_{\mathbb{X}}$ denote the space of all probability measures on $(\mathbb{X}, \mathscr{X})$. Introduce two (ideally) infinite sequences $X^{(\infty)} = (X_n)_{n\geq 1}$ and $Y^{(\infty)} = (Y_n)_{n\geq 1}$ of $\mathbb{X}$-valued random elements defined on the probability space $(\Omega, \mathscr{F}, P)$. The sequence $(X, Y)^{(\infty)} = (X_1, X_2, \ldots, Y_1, Y_2, \ldots)$ is termed *partially exchangeable* if, for any $n_1, n_2 \geq 1$ and for all permutations λ_1 and λ_2 of $\{1, \ldots, n_1\}$ and $\{1, \ldots, n_2\}$, respectively, the distributions of $(X_1, \ldots, X_{n_1})$ and $(Y_1, \ldots, Y_{n_2})$ coincide, respectively, with the distributions of $(X_{\lambda_1(1)}, \ldots, X_{\lambda_1(n_1)})$ and $(Y_{\lambda_2(1)}, \ldots, Y_{\lambda_2(n_2)})$. This notion is equivalently formulated as

$$\mathbb{P}[X^{(\infty)} \in A^{(n_1)},\ Y^{(\infty)} \in B^{(n_2)}] = \int_{P_{\mathbb{X}}\times P_{\mathbb{X}}} \prod_{i=1}^{n_1} p_1(A_i) \prod_{j=1}^{n_2} p_2(B_j) Q(\mathrm{d}p_1, \mathrm{d}p_2), \tag{2}$$

for any $n_1 \geq 1$ and $n_2 \geq 1$, where $A^{(n_1)} = A_1 \times \cdots \times A_{n_1} \times \mathbb{X}^{\infty}$, $B^{(n_2)} = B_1 \times \cdots \times B_{n_2} \times \mathbb{X}^{\infty}$ with A_i and B_j in $\mathscr{X}$ for all i and j. Furthermore, Q, the de Finetti measure of $(X, Y)^{(\infty)}$, is a distribution of some vector $(\tilde{p}_1, \tilde{p}_2)$ of random probability measures (RPMs) on $\mathbb{X}$. Like in the exchangeable case (1), from a Bayesian perspective Q represents a prior distribution. In this framework, proposing a model for partially exchangeable observations is equivalent to specifying a distribution Q. A convenient definition of such a distribution should display a large topological support in $P_{\mathbb{X}} \times P_{\mathbb{X}}$ and a suitable degree of flexibility in describing a whole variety of dependence structures that range from independence of $\tilde{p}_1$ and $\tilde{p}_2$ to their almost sure identity, the latter corresponding to a Q degenerate on $P_{\mathbb{X}}$.

The first proposal of Q in (2) dates back to 1978 and appears in Cifarelli and Regazzini (1978), where a nonparametric prior for partially exchangeable arrays, defined as mixture of Dirichlet processes (DP), is defined. More recently, MacEachern proposed a general class of dependent processes (MacEachern, 1999) and defined a related dependent Dirichlet process (DDP) (MacEachern, 2000), which represented the seminal contribution for a large and highly influential body of literature. Reviews and key references can be found in Hjort et al. (2010). The use of these new classes of models has been made accessible also to practitioners by virtue of the development of suitable MCMC sampling techniques that allow to draw approximate posterior inferences. Furthermore, it should be mentioned that an R package, named *DPpackage*, allows straightforward applications to a variety of dependent models. See Jara et al. (2011) for details. The present paper inserts itself in this line of research and its focus will be on a particular class of dependent RPMs that arise as mixtures of independent RPMs, where one component is common to all mixtures. This structure of dependence first appeared in Müller et al. (2004), where vectors of RPMs were defined as mixtures of two DPs, one idiosyncratic and the other in common. More recently and still in the Dirichlet setting, in Hatjispyros et al. (2011) a multivariate Dirichlet process with a similar dependence structure has been considered and applied to the estimation of vectors $(f_1, \ldots, f_m)$ of densities, by resorting to a slice sampler. In Lijoi et al. (2013) a similar approach has been followed in a general setup: dependent RPMs are defined as normalization of dependent completely random measures, obtained as mixtures of one common and one idiosyncratic component. This approach leads to the definition of a whole class of dependent RPMs that turn out to be analytically tractable and amenable of use in applications.

As a matter of fact, most of the dependent RPMs used in applications can be thought of as extensions to the multivariate setting of the DP. This is a natural choice, the univariate DP being a widely studied object with well known properties. Nonetheless, as shown e.g. in Ishwaran and James (2001, 2003), Lijoi et al. (2005, 2007a) for the exchangeable case, other choices for the nonparametric component are indeed possible and allow to overcome some of the drawbacks of the DP such as, for instance, its sensitivity to the total mass parameter and its simplistic predictive structure. See Lijoi et al. (2007a,b) for a discussion. Carrying out a comparative analysis of structural features of such models also in the multivariate setting is an important task, which, to the best of our knowledge, has not yet been addressed. Such an analysis, in addition to its practical implications, allows also to gain a deeper understanding of the inferential implications of the various modeling choices. This paper aims at giving a contribution in this direction by comparing a bivariate DP with a bivariate normalized σ-stable process. The analysis that is going to be developed relies on the construction proposed in Lijoi et al. (2013). Moreover, dependent DPs and normalized σ-stable processes are the natural candidates to compare since many quantities of interest can be obtained in closed form. The nature of our comparison will therefore be two-fold: on the one hand, we will analyze different properties of the two vectors of RPMs by investigating their predictive distributions, while on the other hand we will resort to a simulation study in order to appreciate the difference between the two models when applied to problems of density estimation and clustering. Importantly, the results of the simulation study find intuitive explanations by means of the insights gained on the predictive structures of the models.

The outline of the paper is as follows. In Section 2 we concisely summarize the vector of bivariate RPMs introduced in Lijoi et al. (2013). A description of the dependent mixtures and a sketch of the MCMC algorithm that is implemented for drawing posterior inferences is provided in Section 3. In Section 4 we compare the properties of bivariate Dirichlet and normalized σ-stable processes by investigating the structure of their predictive distributions and the distribution of the total number of clusters that both models induce on two vectors of observations. Finally, Section 5 is devoted to an extensive simulation study. The inferential impact of the two models choices and of their characterizing parameters is analyzed by focusing on