

A Bayesian nonparametric model for combining data from different experiments

Un modello bayesiano nonparametrico per l'analisi di dati provenienti da più campioni

Antonio Lijoi, Bernardo Nipoti and Igor Prünster

Abstract A number of models have been recently proposed in the Bayesian nonparametric literature for dealing with data arising from different related studies. In this paper we consider a modeling approach that relies on canonically correlated Poisson random measures. These lead to define vectors of dependent random probability measures, which are useful in the contexts of density estimation and survival analysis. With reference to the former we point out results useful for devising a Gibbs sampling algorithm. This is then used to emphasize some remarkable features, especially in terms of the clustering behavior and the borrowing information across datasets, of a class of dependent nonparametric priors based on the normalized sigma-stable process.

Abstract *In questo lavoro viene presentato un modello recentemente proposto per l'analisi dati provenienti da più campioni in ambito bayesiano. Nello specifico, verrà considerata una classe di misure di probabilità aleatorie ottenute a partire da un vettore di misure aleatorie di Poisson aventi correlazioni canoniche. Verranno evidenziati alcuni risultati utili per lo sviluppo di un algoritmo di Gibbs sampling per la stima di densità. Le illustrazioni, basate su dati simulati, consentono di evidenziare le caratteristiche di modelli basati sul processo stabile normalizzato in relazione a problemi di clustering.*

Key words: Clustering, Completely random measures, Density estimation, Dependent priors, Partial exchangeability, Poisson random measures, σ -stable process.

Antonio Lijoi

Department of Economics and Management, University of Pavia, via San Felice 5, 27100 Pavia, e-mail: lijoi@unipv.it

Bernardo Nipoti

Department of Economics and Statistics, University of Torino, c.so Unione Sovietica 218/bis, 10134 Torino, e-mail: bernardo.nipoti@unito.it

Igor Prünster

Department of Economics and Statistics, University of Torino, c.so Unione Sovietica 218/bis, 10134 Torino, e-mail: igor.pruenster@unito.it

1 Introduction

The present paper provides an illustration of the properties of a flexible class of non-parametric priors suitable for inference with non-exchangeable data. These arise in a number of relevant applied problems such as, e.g, regression, time series analysis, spatial statistics, meta-analysis, two-sample problems. Here, attention is focused on the case where data originate from different studies or refer to experiments performed under different conditions. In such a context it is reasonable to preserve the homogeneity (or exchangeability) condition within data that are generated from the same study or under the same experimental condition while, at the same time, dropping the conditional identity in distribution for 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. The first seminal contribution in this direction is [2], but it is only with the key contributions [15, 16] that the subject has gained greater popularity and interest. The modeling framework can be described as follows. Suppose there are k different studies generating \mathbb{X} -valued observations $\mathbf{X}_i = (X_{i,1}, \dots, X_{i,n_i})$, with $i = 1, \dots, k$, and let $\tilde{\mathbf{P}} = (\tilde{P}_1, \dots, \tilde{P}_k)$ be a vector of random probability measures on \mathbb{X} . Then it is assumed that

$$\begin{aligned} X_{\ell,i}, X_{\ell',j} | \tilde{\mathbf{P}} &\stackrel{\text{ind}}{\sim} \tilde{P}_\ell \times \tilde{P}_{\ell'} & \ell \neq \ell' \\ \tilde{\mathbf{P}} &\sim Q \end{aligned} \tag{1}$$

where Q is a probability distribution on the space of vectors of k -dimensional probability measure vectors on \mathbb{X} . The main goal in this setting is the definition of Q and most contributions to this line of research rely on the popular stick-breaking construction. Dependence among different *stick-breaking priors* is created by indexing either the stick-breaking weights or the locations (or both) to relevant covariates. To make things more concrete let $\{(w_{\ell,i})_{i \geq 1} : \ell = 1, \dots, k\}$ be a collection of sequences of non-negative “stick-breaking” weights i.e. $p_{\ell,1} = w_{\ell,1}$ and $p_{\ell,j} = w_{\ell,j} \prod_{i=1}^{j-1} (1 - w_{\ell,i})$ such that $\sum_{i \geq 1} p_{\ell,i} = 1$, with probability 1, for each ℓ . If one further considers collections of sequences $\{(\xi_{\ell,i})_{i \geq 1} : \ell = 1, \dots, k\}$, with the $\xi_{\ell,i}$, for $i \geq 1$, taking values in a space \mathbb{X} and i.i.d. from a non-atomic probability measure $P_{0,\ell}$, a collection of *dependent random probability measures* $\tilde{P}_\ell = \sum_{j \geq 1} p_{\ell,j} \delta_{\xi_{\ell,j}}$ is obtained. The dependence between weights $w_{\ell,i}$ and $w_{\ell',j}$ and/or between the support points $\xi_{\ell,i}$ and $\xi_{\ell',j}$, for $\ell \neq \ell'$, induces dependence between \tilde{P}_ℓ and $\tilde{P}_{\ell'}$. The use of these new classes of models has been made accessible also to practitioners due to the development of suitable MCMC sampling techniques that allow to draw approximate posterior inferences. Furthermore an R package, named *DP-package*, has been recently made available and it allows straightforward applications of a variety of dependent models to real data. See [7] for details. Stimulating and nice reviews of the topic can be found in [18, 20].

Here we rely on a different approach that makes use of a class of canonically correlated Poisson random measures introduced and investigated in [5] by R.C. Grif-

fiths and R.K. Milne. The main goal is to define a vector of completely random measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_k)$ on \mathbb{X} as linear functionals of canonically correlated Poisson random measures and, then, $\tilde{\boldsymbol{P}}$ in (1) is obtained through a suitable transformation of $\tilde{\boldsymbol{\mu}}$. To this end, it is worth recalling that if N is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{X}$ with intensity ν , then for any measurable set A in $\mathbb{R}^+ \times \mathbb{X}$ such that $\nu(A) < \infty$ one has $N(A) \sim \text{Pois}(\nu(A))$. Moreover, if

$$\tilde{\mu}(B) = \int_{\mathbb{R}^+ \times B} s N(ds, dx) = \sum_{i \geq 1} J_i \delta_{\xi_i}(B) \quad (2)$$

then $\tilde{\mu}$ is a *completely random measure* (CRM). This means that for any collection of pairwise disjoint sets A_1, \dots, A_d , the random variables $\tilde{\mu}(A_1), \dots, \tilde{\mu}(A_d)$ are independent and for any measurable function $f : \mathbb{X} \rightarrow \mathbb{R}^+$ such that $\int_{\mathbb{X}} f(x) \tilde{\mu}(dx) < \infty$, with probability 1,

$$-\log \left(\mathbb{E} \left[e^{-\int f d\tilde{\mu}} \right] \right) = \int_{\mathbb{R}^+ \times \mathbb{X}} \left[1 - e^{-sf(x)} \right] \nu(ds, dx) =: \psi(f) \quad (3)$$

is the *Laplace exponent* of $\tilde{\mu}$ evaluated at f . Henceforth we assume that the Lévy intensity has the form

$$\nu(ds, dx) = \rho(s) ds \alpha(dx) \quad (4)$$

for some σ -finite and non-null measure α on \mathbb{X} . In (2), the pairs (J_i, ξ_i) are the points of the Poisson random measure N : the J_i s identify the jump heights and the ξ_i s are the corresponding jump locations. See [8] for some background material on Poisson random measures and CRMs. When moving to the multidimensional case, the dependence among the $\tilde{\mu}_i$'s is induced by the dependence among the canonically correlated Poisson random measures. The approach we undertake has the advantage of preserving some analytical tractability which leads us to obtain some results that are also useful for devising a Gibbs sampling strategy used for approximating posterior inferences on quantities of interest.

The outline of the paper is as follows. In Section 2 we introduce the main construction and hint at applications to density estimation and survival analysis. Section 3 focuses on the former and displays some distributional properties of the proposed model whose relevance becomes, then, apparent in Section 4, where a few illustrations based on simulated data are provided.

2 Correlated random probability measures

If Q in (1) is such that $Q(\tilde{P}_1 = \dots = \tilde{P}_k) = 1$, then the observations are exchangeable and many popular nonparametric priors in this setting can be seen as suitable transformations of CRMs. See [13] for a survey of various classes of discrete nonparametric priors using CRMs as unifying concept. Two such classes are of particular relevance to our treatment: normalized random measures with independent

increments, introduced in [19] (see [1] for a review), and mixture hazard models originated from [3, 14] and laid out in great generality in [6]. Here we also consider models with CRMs as basic building blocks and rely on the idea of [11] to define a vector $\tilde{\mathbf{P}}$ of dependent random probability measures by transformation of the underlying dependent CRMs. In particular, our focus is on dependent normalized random measures and dependent mixture hazards to be used for density estimation or survival analysis, respectively. In order to ease the exposition we consider the case $k = 2$.

2.1 Dependent random measures

Suppose N_1 and N_2 are two Poisson random measures on some space \mathbb{S} with the same intensity measure ν . In [5] a characterization of a class of vectors (N_1, N_2) such that N_1 and N_2 are *canonically correlated* is provided. This connects to a body of literature that has been very popular around the end of the '50s and the 60's and that has drawn inspiration from the seminal work of H.O. Lancaster. See, e.g., [9]. The main advantage from Griffiths and Milne's proposal [5] is the fact that, under suitable assumptions we will adopt here, the joint Laplace transform of (N_1, N_2) is readily available and has a simple form. This is very useful since the evaluation of posterior inferences is achieved by relying on such a transform. In [5] it is, thus, shown that N_1 and N_2 are canonically correlated if and only if there exist three independent Cox processes N'_1, N'_2 and N'_0 such that $N_\ell = N'_\ell + N'_0$, for $\ell = 1, 2$. Moreover, the random intensities of N'_1, N'_2 and N'_0 denoted as $\bar{\nu}, \bar{\nu}$ and ν_0 , respectively, are such that $\bar{\nu} \leq \nu$ and $\nu_0 = \nu - \bar{\nu}$. Hence, given a pair of canonically correlated Poisson random measure (N_1, N_2) on $\mathbb{S} = \mathbb{R}^+ \times \mathbb{X}$, one can define a vector of dependent CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$ such that $\tilde{\mu}_\ell(dx) = \int_{\mathbb{R}^+} s N_\ell(ds, dx)$, for $\ell = 1, 2$. The additive structure, with a random measure in common, that characterizes (N_1, N_2) clearly carries over to the CRM case, so that $\tilde{\mu}_\ell = \mu'_\ell + \mu'_0$ with $\tilde{\mu}'_1, \tilde{\mu}'_2$ and $\tilde{\mu}'_0$ being independent CRMs defined as $\mu'_j(dx) = \int_{\mathbb{R}^+} s N'_j(ds, dx)$.

In the sequel we consider a suitable specification of the underlying Cox processes such that $\bar{\nu} = Z\nu$ and $\nu_0 = (1 - Z)\nu$ for some $[0, 1]$ -valued random variable Z . It can be easily seen that, conditional on $Z = z$, one has

$$\mathbb{E}_z \left[e^{-\tilde{\mu}_1(f_1) - \tilde{\mu}_2(f_2)} \right] = \exp \{ -z(\psi(f_1) + \psi(f_2)) - (1 - z)\psi(f_1 + f_2) \} \quad (5)$$

If $\nu(ds, dx) = \sigma s^{-1-\sigma} ds \alpha(dx) / \Gamma(1 - \sigma)$ for some $\sigma \in (0, 1)$ and f_ℓ is such that $\int_{\mathbb{X}} f_\ell^\sigma(x) P_0(dx) < \infty$, the Laplace transform in (5) equals

$$e^{-z \int_{\mathbb{X}} (f_1^\sigma(x) + f_2^\sigma(x)) \alpha(dx) - (1-z) \int_{\mathbb{X}} (f_1(x) + f_2(x))^\sigma \alpha(dx)}.$$

and we shall refer to $(\tilde{\mu}_1, \tilde{\mu}_2)$ as a vector of *GM-dependent σ -stable processes*. Though the illustration we provide focuses on random probability measures that are generated by GM-dependent σ -stable processes, it should be noted that the

approach applies to all CRMs for which the jump location component of the Lévy intensity (4), namely α , is a non-atomic measure on \mathbb{X} . For instance, one could alternatively work with gamma processes which correspond to $\nu(ds, dx) = e^{-s} s^{-1} ds \alpha(dx)$. In this case, if $f_\ell : \mathbb{X} \rightarrow \mathbb{R}^+$ (with $\ell = 1, 2$) are such that $\int_{\mathbb{X}} \log(1 + f_\ell(x)) \alpha(dx) < \infty$, then (5) reduces to

$$e^{-z \int_{\mathbb{X}} \log(1+f_1(x))(1+f_2(x)) \alpha(dx) - (1-z) \int_{\mathbb{X}} \log(1+f_1(x)+f_2(x)) \alpha(dx)}$$

and $(\tilde{\mu}_1, \tilde{\mu}_2)$ is a vector of *GM-dependent gamma processes*. Two possible uses of suitably transformed $(\tilde{\mu}_1, \tilde{\mu}_2)$ in Bayesian nonparametric inference are density estimation and survival analysis.

2.2 Density estimation

If ν in (4) is such that $\int_{\mathbb{R}^+} \rho(s) ds = \infty$ and $\alpha(\mathbb{X}) = c < \infty$, then $\tilde{P}_\ell = \tilde{\mu}_\ell / \tilde{\mu}_\ell(\mathbb{X})$ defines a random probability measure on \mathbb{X} , for $\ell = 1, 2$. Due to the specific construction used we refer to $\tilde{\mathbf{P}} = (\tilde{P}_1, \tilde{P}_2)$ as a vector of *GM-dependent normalized random measures*. Note that starting from a vector of σ -stable or gamma GM-dependent processes, one obtains normalized σ -stable and Dirichlet GM-dependent processes, respectively. Moreover, due to the additive structure of $\tilde{\mu}_\ell$ one has the following representation

$$\tilde{P}_\ell = \pi_\ell P'_\ell + (1 - \pi_\ell) P'_0 \quad \ell = 1, 2 \quad (6)$$

where $\pi_\ell = \mu'_\ell(\mathbb{X}) / [\mu'_\ell(\mathbb{X}) + \mu'_0(\mathbb{X})]$ and $P'_j = \mu'_j / \mu'_j(\mathbb{X})$ for $j = 0, 1, 2$. Note that the P'_j s are independent though they are not necessarily independent from the mixing weight (π_1, π_2) . The representation in (6) can be useful in order to gain further understanding on the dependence structure between \tilde{P}_1 and \tilde{P}_2 . It is worth noting that when $\tilde{\mu}_\ell$ are gamma processes, each \tilde{P}_ℓ is marginally a Dirichlet process and the \tilde{P}'_j and the π_j are independent. Using this representation, we consider two random samples $\mathbf{Y}_1 = (Y_{1,1}, \dots, Y_{1,n_1})$ and $\mathbf{Y}_2 = (Y_{2,1}, \dots, Y_{2,n_2})$ such that

$$\begin{aligned} (Y_{1,i}, Y_{2,j}) &\stackrel{\text{iid}}{\sim} h(\cdot; X_{1,i}) h(\cdot; X_{2,j}) \\ (X_{1,i}, X_{2,j}) | (\tilde{P}_1, \tilde{P}_2) &\stackrel{\text{iid}}{\sim} \tilde{P}_1 \times \tilde{P}_2 \end{aligned} \quad (7)$$

and $h(\cdot; \cdot)$ is some kernel such that $\int_{\mathbb{R}} h(y, x) dy = 1$ for any x in \mathbb{X} . The model in (7) implies that, conditional on $(\tilde{P}_1, \tilde{P}_2)$, the two-sample data \mathbf{Y}_1 and \mathbf{Y}_2 are independent from densities $\tilde{f}_1(y) = \int_{\mathbb{X}} h(y, x) \tilde{P}_1(dx)$ and $\tilde{f}_2(y) = \int_{\mathbb{X}} h(y, x) \tilde{P}_2(dx)$, respectively. The model we propose is reminiscent of the one developed in [17] where the authors consider $\pi_1 = \pi_2$ and confine their attention to the Dirichlet process case.

2.3 Survival analysis

When $\mathbf{T}_1 = (T_{1,1}, \dots, T_{1,n_1})$ and $\mathbf{T}_2 = (T_{2,1}, \dots, T_{2,n_2})$ are two samples of survival data, subject to some censoring mechanism, a possible strategy for defining a bivariate vector of priors would lead to introduce mixtures $\tilde{r}_\ell(t) = \int_{\mathbb{X}} r(t, x) \tilde{\mu}_\ell(dx)$, for $\ell = 1, 2$. Under suitable conditions, $(\tilde{r}_1, \tilde{r}_2)$ defines a vector of *GM-dependent random hazard rate functions* for \mathbf{T}_1 and \mathbf{T}_2 . One accordingly has

$$\mathbb{P}[T_{1,i} > t_1, T_{2,j} > t_2 | (\tilde{\mu}_1, \tilde{\mu}_2)] = \exp \left\{ - \int_0^{t_1} \tilde{r}_1(s) ds - \int_0^{t_2} \tilde{r}_2(s) ds \right\}$$

and, due to the nice structure featured by $(\tilde{\mu}_1, \tilde{\mu}_2)$, one can still determine expressions of quantities of interest for posterior inference. Due to space constraints we do not pursue this model further here and defer the interested reader to [10] for a comprehensive development.

3 Clustering and borrowing information

With reference to the model in (7) it is apparent that, due to the discreteness of the underlying $\tilde{\mu}_\ell$, the GM-dependent normalized random measures \tilde{P}_1 and \tilde{P}_2 select almost surely discrete distributions. Hence, $\mathbb{P}[X_{\ell,i} = X_{\ell,j}] > 0$ for any $i \neq j$ and $\ell = 1, 2$. Moreover, if $\mathbb{P}[Z < 1] > 0$ then $\mathbb{P}[X_{1,i} = X_{2,j}] > 0$ for any i and j so that there is a positive probability of detecting ties also between \mathbf{X}_1 and \mathbf{X}_2 . Such properties naturally lead to consider the random partition induced by the latent variables and, then, determine the probability of observing a sample $(\mathbf{X}_1, \mathbf{X}_2)$ having a specific clustering structure into

- k_1 distinct values $X_{1,1}^*, \dots, X_{1,k_1}^*$ specific to \mathbf{X}_1 with respective frequencies $\mathbf{n}_1 = (n_{1,1}, \dots, n_{1,k_1})$
- k_2 distinct values $X_{2,1}^*, \dots, X_{2,k_2}^*$ specific to \mathbf{X}_2 with respective frequencies $\mathbf{n}_2 = (n_{2,1}, \dots, n_{2,k_2})$
- k_0 distinct values $X_1^*, \dots, X_{k_0}^*$ shared by \mathbf{X}_1 and \mathbf{X}_2 with frequencies $\mathbf{q}_1 + \mathbf{q}_2 = (q_{1,1} + q_{2,1}, \dots, q_{1,k_0} + q_{2,k_0})$.

This provides a natural extension to a partially exchangeable framework of the exchangeable partition probability function (EPPF). In [11, Proposition 2] a closed form expression for the probability distribution of $(\mathbf{X}_1, \mathbf{X}_2)$ encoded by the positive integers $(k_1, k_2, k_0, \mathbf{n}_1, \mathbf{n}_2, \mathbf{q}_1, \mathbf{q}_2)$ is obtained. Such an expression is uniquely identified by the marginal Lévy intensities of $\tilde{\mu}_1$ and $\tilde{\mu}_2$ and by Z and, after a suitable augmentation, can be used to devise a Gibbs sampling algorithm. This, in turn, allows to estimate the densities of the two populations and to infer on the number of clusters featured by each of the two samples. Indeed, we shall consider sequences of random elements $(\zeta_{1,j})_{j \geq 1}$ and $(\zeta_{2,j})_{j \geq 1}$ that act as labels and take values in $\{0, 1\}^\infty$ and $\{0, 2\}^\infty$, respectively. They are such that

$$\mathbb{P}[X_{\ell,i} \in \cdot \mid \zeta_{\ell,i}, \mu'_1, \mu'_2, \mu'_0] = \tilde{P}'_{\zeta_{\ell,i}}(\cdot) \quad \forall i = 1, \dots, n_\ell.$$

Hence, the $\zeta_{\ell,i}$ s identify the specific urn/distribution $X_{\ell,i}$ is sampled from. This implies that $\mathbb{P}[X_{1,i} = X_{2,j} \mid \zeta_{1,i} \neq \zeta_{2,j}] = 0$ and, then, the labels associated to shared distinct values are degenerate at 0. In summary, corresponding to the $k_\ell + k_0$ distinct values in \mathbf{X}_ℓ one has distinct labels $(\zeta_{\ell,1}^*, \dots, \zeta_{\ell,k_\ell}^*, \zeta_0^*, \dots, \zeta_{k_0}^*)$ where it is obvious that $\zeta_0^* = \dots = \zeta_{k_0}^* = 0$ and do need not to be sampled. By relying on [11, Proposition 2] one can deduce full conditional distributions for sampling the $\zeta_{\ell,i}^*$. Specific closed forms for the GM-dependent Dirichlet and the normalized σ -stable cases are detailed in [11, Corollaries 2 & 3]. If $P_0 = \alpha/c$, the full conditionals for the latent variables \mathbf{X}_1 and \mathbf{X}_2 can be easily determined and for $X_{\ell,j}$ one has

$$w_0 P_\ell^* + \sum_{(*)} w_i \delta_{\tilde{X}_{\zeta_{\ell,j},i}}$$

where $P_\ell^*(dx) = h(Y_{\ell,j};x) P_0(dx) / \int_{\mathbb{X}} h(Y_{\ell,j};x) P_0(dx)$, the $\tilde{X}_{\zeta_{\ell,j},i}$ are the distinct X -values in the urn labeled $\zeta_{\ell,j}$ and $(*)$ is the set of indices of distinct values from the urn labeled $\zeta_{\ell,j}$ after excluding $X_{\ell,j}$. The weights w_i in the above convex linear combination are determined through the expression of the partition probability function given in [11, Proposition 2]. In particular, when $(\tilde{P}_1, \tilde{P}_2)$ is a GM-dependent normalized σ -stable process one has

$$w_0 \propto k_{\zeta_{\ell,j},-j} \sigma (1-z)^{1-\zeta_{\ell,j} z^{\zeta_{\ell,j}}} \int_{\mathbb{X}} h(Y_{\ell,j};x) P_0(dx), \quad w_i \propto (n_{\ell,i}^{(-j)} - \sigma) h(Y_{\ell,j}; \tilde{X}_{\zeta_{\ell,j},i})$$

where $n_{\ell,i}^{(-j)}$ is the size of the cluster corresponding to $\tilde{X}_{\zeta_{\ell,j},i}$ and $k_{\zeta_{\ell,j},-j}$ is the number of clusters associated to $P'_{\zeta_{\ell,j}}$, namely to the urn labeled by $\zeta_{\ell,j}$, after deleting $X_{\ell,j}$.

4 Illustration

We now perform a simulation study on mixture models (7) by implementing a Gibbs sampling algorithm that extends the ubiquitous Blackwell-MacQueen Pólya urn scheme. We specifically focus on the posterior estimation of a pair of dependent densities and of the marginal clustering structures. Due to the lack of space, here we provide only a partial description of the properties of the proposed model. A more extensive analysis is provided in [12]. With the shorthand notation GM-st($\sigma, Z; P_0$) we denote the GM-dependent normalized σ -stable process with base measure P_0 and dependence parameter Z .

The model specification is an extension to the partially exchangeable case of the quite standard specification of [4]. In particular, we shall assume that the latent variables $X = (M, V) \in \mathbb{R} \times \mathbb{R}^+$ identify mean M and variance V of a Gaussian kernel $h(\cdot; M, V)$. We also take P_0 to be a normal/inverse-gamma distribution $P_0(dM, dV) = P_{0,1}(dV) P_{0,2}(dM \mid V)$, with $P_{0,1}$ being an inverse-gamma probability

distribution with parameters $(1, 1)$ and $P_{0,2}$ is Gaussian with mean m and variance τV . Moreover, the corresponding hyperpriors are of the form

$$\begin{aligned}\tau^{-1} &\sim \text{Ga}(1/2, 50), \\ m &\sim \text{N}(\bar{D}, 2), \\ Z &\sim \text{U}(0, 1), \\ \sigma &\sim \text{U}(0, 1)\end{aligned}\tag{8}$$

where $\bar{D} = (\sum_{i=1}^{n_1} Y_{1,i} + \sum_{j=1}^{n_2} Y_{2,j}) / (n_1 + n_2)$ is the overall sample mean. In the above specification, $\text{Ga}(a, b)$ stands for the gamma distribution with expected value a/b . The experiment is completed by the simulation of $n_1 = 200$ and $n_2 = 100$ data from the following densities $f_1 \sim \sum_{i=1}^7 a_i \text{N}(\mu_i, \sigma_i)$ and $f_2 \sim \sum_{i=1}^7 b_i \text{N}(\mu_i, \sigma_i)$, respectively, where the vectors of means and standard deviations are respectively equal to $\boldsymbol{\mu} = (-7, -5, -3, -1, 1, 3, 5, 7)$ and $\boldsymbol{\sigma} = (3, 2, 4, 2, 4, 2, 3, 2)/8$. On the other hand, the weights are $\boldsymbol{a} = (1, 1, 1, 0, 1, 1, 1, 1)/7$ and $\boldsymbol{b} = (0, 1, 0, 1, 0, 1, 0, 0)/3$. This implies that one can rewrite $f_j = g_j + g_0$, for $j = 1, 2$, where g_j is the idiosyncratic component and g_0 is the common component consisting of a mixture of densities centered on $\mu = -5$ and $\mu = 3$. See Figure 1.

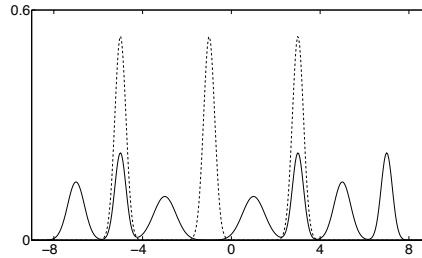


Fig. 1: True densities f_1 and f_2 generating the simulated dataset.

We shall now resort to GM-st(σ, Z, P_0) mixtures to estimate both f_1 and f_2 and to infer on the number of clusters, $K_{\mathbf{Y}_1}$ and $K_{\mathbf{Y}_2}$, of each sample. As for the latter, we shall consider $K_{\mathbf{Y}_\ell} = K_{\mathbf{X}_\ell}$ thus relying on the partition structure induced at the level of the latent random samples. Estimates of $K_{\mathbf{Y}_\ell}$ are reported in Table 1.

It is worth noting that the numerical results we achieve allow for a straightforward comparison between GM-dependent mixtures with $\mathbb{P}[Z < 1] > 0$ and mixtures with $\mathbb{P}[Z = 1] = 1$ which corresponds to considering the two samples independent. In the latter case, no borrowing strength takes place and this affects posterior estimates. When considering $K_{\mathbf{Y}_1}$, it can be seen that GM-dependent mixtures yield more accurate results since estimates are closer to the correct value, namely 7, if compared to the outcome with independent mixtures for which $Z = 1$ with probability 1. This

Table 1: GM–st mixture vs. independent univariate normalized σ –stable process mixtures: posterior expected number of clusters (Cols. 1 and 2), maximum a posteriori values (\hat{K}_{Y_1} , \hat{K}_{Y_2}), and posterior probability of exact number of components, variance of the posterior distribution of K_{Y_1} , K_{Y_2} .

	$\mathbb{E}[K_{Y_1} \cdot]$	$\mathbb{E}[K_{Y_2} \cdot]$	\hat{K}_{Y_1}	\hat{K}_{Y_2}	$\mathbb{P}[K_{Y_1}=7 \cdot]$	$\mathbb{P}[K_{Y_2}=3 \cdot]$	$\text{Var}(K_{Y_1})$	$\text{Var}(K_{Y_2})$
GM–st(σ, Z, P_0)	7.62	3.04	7	3	0.56	0.96	0.73	0.04
GM–st($\sigma, 1, P_0$)	9.64	3.12	8	3	0.12	0.89	10.02	0.13

is due to a borrowing strength phenomenon that helps identifying the actual number of clusters also in the most difficult scenario as identified by f_1 . The variability appears significantly different in the two models. This is also apparent from the approximate distributions of K_{Y_1} that arise from the MCMC output and that are depicted in Figure 2

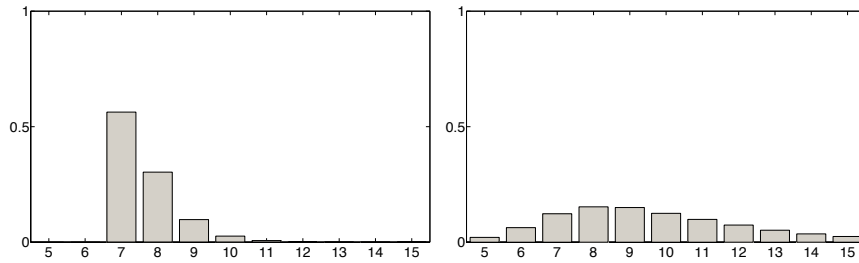


Fig. 2: Posterior distributions of the number of clusters K_{Y_1} of the first sample ($n_1 = 200$) with a GM–st(σ, Z, P_0) mixture (left) and with a GM–st($\sigma, 1, P_0$) mixture (right).

For the mixture with 3 components, f_2 , the effect of the borrowing strength is not so apparent since given the simple structure of the mixture the correct number of clusters can be easily detected even without introducing dependence between the two samples. The estimated densities are depicted in Figure 3.

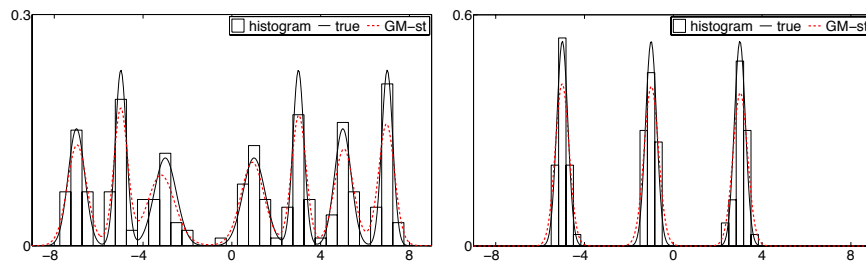


Fig. 3: True data generating densities with histograms of simulated data. Corresponding estimates are obtained with the GM–st(σ, Z, P_0) mixture model.

Acknowledgements The authors are also affiliated to the Collegio Carlo Alberto in Moncalieri, whose support is acknowledged. The first and the third author are also supported by the European Research Council (ERC) through StG “N-BNP” 306406.

References

1. Barrios, E., Lijoi, A., Nieto-Barajas, L.E., Prünster, I.: Modeling with normalized random measure mixture models. *Statistical Science* **28**, 313-334 (2013).
2. Cifarelli, D.M., Regazzini, E.: Problemi statistici non parametrici in condizioni di scambiabilità parziale. *Quaderni Istituto Matematica Finanziaria, Università di Torino Serie III*, **12** (1978).
3. Dykstra, R. L., Laud, P.: A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367 (1981).
4. Escobar, M.D., West, M.: Bayesian density estimation and inference using mixtures. *J. Amer. Statist. Assoc.* **90**, 577–588 (1995).
5. Griffiths, R.C., Milne, R.K.: A class of bivariate Poisson processes. *J. Mult. Anal.* **8**, 380–395 (1978).
6. James, L.F.: Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33**, 1771–1799 (2005).
7. Jara, A., Hanson, T., Quintana, F.A., Müller, P., Rosner, G.: DPpackage: Bayesian non- and semi-parametric modelling in R. *J. Statist. Soft.* **40**, 1–30 (2011).
8. Kingman, J.F.C.: *Poisson processes*. Oxford University Press, Oxford (1993).
9. Lancaster, H.O.: The structure of bivariate distributions. *Ann. Math. Statist.* **29**, 719–736 (1958).
10. Lijoi, A., Nipoti, B.: A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc.* (2014), doi: 10.1080/01621459.2013.869499.
11. Lijoi, A., Nipoti, B., Prünster, I.: Bayesian inference with dependent normalized completely random measures. *Bernoulli* (2014), doi: 10.3150/13-BEJ521.
12. Lijoi, A., Nipoti, B., Prünster, I.: Dependent mixture models: Clustering and borrowing information. *Comput. Statist. Data Anal.* **71**, 417–433 (2014).
13. Lijoi, A., Prünster, I.: Models beyond the Dirichlet process. In: Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (eds.) *Bayesian Nonparametrics*, pp. 80–136. Cambridge University Press, Cambridge (2010).
14. Lo, A.Y., Weng, C.-S.: On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41**, 227–245 (1989).
15. MacEachern, S.N.: Dependent nonparametric processes. In: *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA, (1999).
16. MacEachern, S.N. Dependent Dirichlet processes.: Technical Report, Ohio State University (2000).
17. Müller, P., Quintana, F.A., Rosner, G.: A method for combining inference across related non-parametric Bayesian models. *J. Roy. Statist. Soc. Ser. B* **66**, 735–749 (2004).
18. Müller, P., Mitra, R.: Bayes nonparametric inference - Why and how. *Bayesian Anal.* **8**, 269–302 (2013).
19. Regazzini, E., Lijoi, A., Prünster, I.: Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585 (2003).
20. Teh, Y.W., Jordan, M.I.: Hierarchical Bayesian nonparametric models with applications. In: Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (eds.) *Bayesian Nonparametrics*, pp. 158–207. Cambridge University Press, Cambridge (2010).