Submitted to the Bernoulli

Bayesian inference with dependent normalized completely random measures

ANTONIO LIJOI,[1,3] BERNARDO NIPOTI,[2,3] and IGOR PRÜNSTER[2,3]

[1] *Department of Economics and Management, University of Pavia, Via San Felice 5, 27100 Pavia, Italy, E-mail:* lijoi@unipv.it

[2] *Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy, E-mail:* bernardo.nipoti@unito.it; igor@econ.unito.it

[3] *Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy*

The proposal and study of dependent nonparametric priors has been a major research focus in the recent Bayesian nonparametric literature. In this paper we introduce a flexible class of dependent nonparametric priors, investigate their properties and derive a suitable sampling scheme which allows their concrete implementation. The proposed class is obtained by normalizing dependent completely random measures, where the dependence arises by virtue of a suitable construction of the Poisson random measures underlying the completely random measures. We first provide general distributional results for the whole class of dependent completely random measures and then we specialize them to two specific priors, which represent the natural candidates for concrete implementation due to their analytic tractability: the bivariate Dirichlet and normalized σ–stable processes. Our analytical results, and in particular the partially exchangeable partition probability function, form also the basis for the determination of a Markov Chain Monte Carlo algorithm for drawing posterior inferences, which reduces to the well–known Blackwell MacQueen Pólya urn scheme in the univariate case. Such an algorithm can be used for density estimation and for analyzing the clustering structure of the data and is illustrated through a real two–sample dataset example.

AMS 2000 subject classifications: Primary 60G57, 62G05, 62F15.
Keywords: Completely random measure, Dependent Poisson processes, Dirichlet process, Generalized Polýa urn scheme, Infinitely divisible vector, Normalized σ–stable process, Partially exchangeable random partition.

1. Introduction

The construction of dependent random probability measures for Bayesian inference has attracted considerable attention in the last decade. The seminal contributions of MacEachern [27, 28], who introduced a general class of dependent processes including a popular dependent version of the Dirichlet process, paved the way to a burst in the literature on (covariate) dependent processes and their application in a variety of frameworks such as, e.g., nonparametric regression, inference on time series data, meta-analysis, two–sample

problems. Reviews and key references can be found in, e.g., [29, 7, 39]. Most contributions to this line of research rely on random probability measures defined by means of a stick–breaking procedure, a popular method set forth in its generality for the first time in [17]. Dependence among different *stick–breaking priors* is created by indexing either the stick–breaking weights or the locations or both to relevant covariates. To be more specific, if $\mathcal{Z}$ denotes the covariate space and $\{(\omega_{j,z})_{j\geq 1} : z \in \mathcal{Z}\}$ is a collection of sequences of independent non–negative weights, the stick–breaking procedure consists in defining $p_{1,z} = \omega_{1,z}$ and $p_{j,z} = \omega_{j,z} \prod_{i=1}^{j-1}(1-\omega_{i,z})$. A typical choice is then $\omega_{i,z} \sim \text{Beta}(a_{i,z}, b_{i,z})$ with parameters $(a_{i,z}, b_{i,z})$ such that $\sum_{j\geq 1} p_{j,z} = 1$, almost surely. If one further considers collections of sequences $\{(X_{i,z})_{i\geq 1} : z \in \mathcal{Z}\}$ with the $X_{i,z}$, for $i \geq 1$, taking values in a space $\mathbb{X}$ and i.i.d. from a non–atomic probability measure $P_{0,z}$, a *covariate dependent random probability measure* $\tilde{p}_z = \sum_{j\geq 1} p_{j,z}\, \delta_{X_{j,z}}$ is obtained. The dependence between weights $\omega_{i,z}$ and $\omega_{j,z'}$ and/or between the support points $X_{i,z}$ and $X_{j,z'}$, for $z \neq z'$, induces dependence between $\tilde{p}_z$ and $\tilde{p}_{z'}$. This general framework is then tailored to the specific application at issue. One of the main reasons of the success of stick–breaking constructions is their attractiveness from computational point of view along with their flexibility since, as shown in [3], they have full weak support under mild assumptions. On the other hand, a drawback is represented by the difficulty of studying their distributional properties due to their analytical intractability. In this paper we propose a radically different approach to the construction of dependent nonparametric priors that relies on *completely random measures* (CRMs) introduced by Kingman [21]. For the case of exchangeable setting, in [24] it has been shown that CRMs represent a unifying concept of the Bayesian Nonparametrics given most discrete nonparametric priors can be seen as transformations of CRMs. Our general plan consists in defining a broad class of dependent CRMs thus obtaining a vector of dependent random probability measures via a suitable transformation. A relevant motivation for undertaking such an approach is represented by the consideration that the study of distributional properties of the models are essential for their deep understanding and sound applications. In this respect, even though CRMs are infinite–dimensional objects, they can be summarized by a single measure, i.e. their intensity, which allows to derive key distributional properties.

1.1. Dependent Poisson random measures

A key idea of our approach consists in defining dependent CRMs by creating dependence at the level of the underlying Poisson random measures (PRM). To this end, we resort to a class of bivariate dependent PRMs devised by Griffiths and Milne in [15]. In particular, let $\tilde{N}$ be a PRM on $\mathbb{Y}$ with intensity measure $\bar{\nu}$. The corresponding Laplace functional transform, which completely characterizes the PRM, is then given by

$$\mathbb{E}\left[\mathrm{e}^{-\int f \,\mathrm{d}\tilde{N}}\right] = \mathrm{e}^{-\bar{\nu}(1-\mathrm{e}^{-f})}.$$

for any measurable function $f : \mathbb{Y} \to \mathbb{R}$ such that $\int |f| \,\mathrm{d}\tilde{N} < \infty$ (a.s.). Recall also that a Cox process is a PRM with random intensity. See [6] for an exhaustive account.

Consider now a vector of (possibly dependent) PRMs $(\tilde{N}_1, \tilde{N}_2)$ on $\mathbb{Y}$ with the same marginal intensity measure $\bar{\nu}$. Griffiths and Milne [15] prove that the $\tilde{N}_i$'s admit an additive representation

$$\tilde{N}_i = M_i + M_0 \qquad\qquad i = 1, 2 \tag{1}$$

where M_1, M_2 and M_0 are independent Cox processes with respective random intensities ν, ν and ν_0 such that $\nu_0 \leq \bar{\nu}$ (a.s.) and $\nu = \bar{\nu} - \nu_0$ if and only if the Laplace transform admits the following form

$$\mathbb{E}\left[\mathrm{e}^{-\tilde{N}_1(f_1)-\tilde{N}_2(f_2)}\right] = \mathrm{e}^{-\sum_{i=1}^{2}\bar{\nu}(1-\mathrm{e}^{-f_i})}\,\varphi((1-\mathrm{e}^{-f_1})(1-\mathrm{e}^{-f_2})), \tag{2}$$

for some functional φ. Such a result is appealing for at least two reasons. From an intuition point of view it provides a neat additive representation (1) of the $\tilde{N}_i$'s with a common and idiosyncratic component, M_0 and M_i, for $i = 1, 2$, respectively. From an operational point of view it yields a well identified structure (2) for the Laplace functional, which becomes completely explicit in the cases where one is able to determine the form of φ. In fact, when working with PRM and CRMs, the Laplace functional is the main operational tool for deriving analytical results useful for Bayesian inference and such a relatively simple structure is actually quite surprising for the dependent case.

The pair of PRMs constructed according to (1) is, then, used to define a vector of dependent CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$. Recall that CRMs are random measures giving rise to mutually independent random variables when evaluated on pairwise disjoint measurable sets. Moreover, they can always be represented as functionals of an underlying PRM, which in the particular case of $\mathbb{Y} = \mathbb{R}^+ \times \mathbb{R}^d$ corresponds to the celebrated *Lévy–Ito decomposition*. Therefore, by setting $\mathbb{Y} = \mathbb{R}^+ \times \mathbb{X}$, from $(\tilde{N}_1, \tilde{N}_2)$ one can define the corresponding vector of CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$ with components given by $\tilde{\mu}_i(\mathrm{d}x) = \int_{\mathbb{R}^+} s\,\tilde{N}_i(\mathrm{d}s, \mathrm{d}x)$.

Finally a vector of dependent random probability measures on $\mathbb{X}$ is obtained as $(\tilde{p}_1, \tilde{p}_2) \stackrel{d}{=} (T(\tilde{\mu}_1), T(\tilde{\mu}_2))$ where T is a transformation of the CRM such that $T(\tilde{\mu}_i)(\mathbb{X}) = 1$ a.s. Here we focus on one of the most intuitive transformations, namely "normalization", which corresponds to $T(\tilde{\mu}) = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$. Such a normalization procedure is widely used in the univariate case. Already Ferguson [11] showed that the Dirichlet process can be defined as normalization of a gamma CRM. Such a procedure has then been extended and analyzed for general univariate CRMs in [38, 19, 20]. More recently an interesting construction of a subclass of normalized CRMs has been proposed in [33]. See [24] for a review of other possible transformations T exploited in the literature.

In the literature there are already some proposals, although not in a general framework and analytical depth as set forth here, making use of dependent CRMs for defining dependent random probability measures. For example in [22] and in [37] one can find a model that coincides with a special case we consider in this paper, namely a version of the bivariate Dirichlet process. In these two papers, the authors devise samplers that take advantage of a mixture representation of $\tilde{p}_1$ and of $\tilde{p}_2$ whose weights are, only for their special case, independent from the $\tilde{p}_i$'s. In a similar fashion, [30] proposes dependent convex linear combinations of Dirichlet processes as a tool for examining data originated from different experiments. Vector CRMs, whose dependence is induced by suitable Lévy

copulas, are proposed in [8] for defining a vector of dependent neutral to the right processes and in [23] in order to introduce a bivariate two–parameter Poisson–Dirichlet process. In addition to the great generality of our results, two important features of our proposal are to be highlighted: it preserves computational efficiency since we are able to deduce a generalization of the Blackwell–MacQueen urn scheme for the dependent setting implementable in real–world applications, and it sheds light on theoretical properties of the vector of random probability measures we are proposing, therefore improving the understanding of the model.

1.2. Goals and outline of the paper

As mentioned above, we will investigate vectors of random probabilities $(\tilde{p}_1, \tilde{p}_2)$ obtained by normalizing pairs of dependent CRMs $(\tilde{\mu}_1, \tilde{\mu}_2)$. The distribution of $(\tilde{p}_1, \tilde{p}_2)$ plays the role of mixing measure in the representation of the law of a pair of partially exchangeable sequences or, in other terms, of prior distribution for a partially–exchangeable observation process. We will determine an expression for the probability distribution of the partially exchangeable partition induced by $(\tilde{p}_1, \tilde{p}_2)$. Such a result will also lead us to achieve an extension of the univariate Blackwell–MacQueen Pólya urn scheme. The corresponding Gibbs sampler is then implemented to draw a full Bayesian analysis for density estimation and cluster analysis in two–sample problems. The general results will, then, be specialized to two specific priors where: (i) the $\tilde{\mu}_i$'s are gamma CRMs thus yielding a vector of dependent Dirichlet processes; (ii) the $\tilde{\mu}_i$'s are σ–stable CRMs that give rise to a vector of dependent normalized σ–stable processes.

The outline of the paper is as follows. In Section 2 we introduce some notation and formalize the form of dependence we briefly touched upon before. In Section 3 we consider pairs of partially exchangeable sequences directed by the distribution of $(\tilde{p}_1, \tilde{p}_2)$ and describe some of their distributional properties. Section 4 considers dependent mixtures and introduces the main distributional tools that are needed for their application to the analysis of partially exchangeable data. Section 5 provides a description of the prior specification we adopt and the sampler we resort to. Finally Section 6 contains an illustration with a real dataset which is analyzed through mixture models with both dependent Dirichlet and normalized σ–stable. The proofs are postponed to the Appendix. A key tool for proving our results is represented by an extension to the partial exchangeable case of a technique introduced and subsequently refined in [36, 19, 20]. Such a technique was originally developed for deriving conditional distributions of normalized random measures [38] but, as highlighted in [24], it can be actually applied to any exchangeable model based on completely random measures. Therefore, it is worth remarking that the extension to the partial exchangeable setup is also of independent interest.

2. Dependent completely random measures

Let us start by stating more precisely some of the concepts sketched in the Introduction. Consider a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and denote by $M_{\mathbb{X}}$ the set of boundedly finite

measures on a complete and separable metric space $\mathbb{X}$. Further, the Borel σ–algebras on $M_{\mathbb{X}}$ and $\mathbb{X}$ are denoted by $\mathscr{M}_{\mathbb{X}}$ and $\mathfrak{X}$, respectively. A completely random measure (CRM) μ on $(\mathbb{X}, \mathfrak{X})$ is a measurable function on $(\Omega, \mathscr{F}, \mathbb{P})$ taking values in $(M_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ such that for any pair of disjoint sets A and B in $\mathfrak{X}$ the random variables $\mu(A)$ and $\mu(B)$ are independent. It is well–known that if N is a Poisson random measure on $\mathbb{R}^+ \times \mathbb{X}$, then

$$\mu(B) = \int_{\mathbb{R}^+ \times B} s\, N(\mathrm{d}s, \mathrm{d}x) \qquad \forall B \in \mathfrak{X} \tag{3}$$

is a CRM on $(\mathbb{X}, \mathfrak{X})$. See [21, 6] and, e.g, [18] for uses of representation (3) for Bayesian modeling. If $\bar{\nu}$ is the intensity of N and for brevity $\mu(f) := \int f \mathrm{d}\mu$, the Laplace exponent of $\mu(f)$ is of the form

$$-\log\left(\mathbb{E}\left[\mathrm{e}^{-\mu(f)}\right]\right) = \int_{\mathbb{R}^+ \times \mathbb{X}} [1 - \mathrm{e}^{-sf(x)}]\, \bar{\nu}(\mathrm{d}s, \mathrm{d}x) =: \psi(f) \tag{4}$$

for any measurable function $f : \mathbb{X} \to \mathbb{R}$ such that $\mu(|f|) = \int |f| \,\mathrm{d}\mu < \infty$, almost surely. By virtue of (3), we can construct dependent CRMs as linear functionals of dependent PRMs determined according to (1). To state it more precisely, let P_0 be a non–atomic probability measure on $(\mathbb{X}, \mathfrak{X})$ and $r(\mathrm{d}s) = \rho(s)\,\mathrm{d}s$ a (possibly infinite) measure on $\mathbb{R}^+$. Suppose, further, that $\tilde{N}_1$ and $\tilde{N}_2$ are defined as in (1), where M_1, M_2 and M_0 are three independent Cox processes with respective random intensities ν, ν and ν_0 such that $\nu + \nu_0 = \bar{\nu}$, almost surely. Henceforth, we shall assume $\bar{\nu}(\mathrm{d}s, \mathrm{d}x) = cP_0(\mathrm{d}x)\,\rho(s)\,\mathrm{d}s$.

Definition 1. Let $(\tilde{N}_1, \tilde{N}_2)$ be a vector of Griffiths–Milne (GM) dependent PRMs as in (1) and define the CRMs $\tilde{\mu}_i(\mathrm{d}x) = \int_{\mathbb{R}^+} s\, \tilde{N}_i(\mathrm{d}s, \mathrm{d}x)$, for $i = 1, 2$. Then $(\tilde{\mu}_1, \tilde{\mu}_2)$ is said to be a vector of *GM–dependent* CRMs. The marginal intensity of $\tilde{\mu}_i$ coincides with $\bar{\nu}$.

In the sequel we will focus on a simple class of Cox processes defined through an intensity of the form

$$\nu(\mathrm{d}s, \mathrm{d}x) = cZP_0(\mathrm{d}x)\,\rho(s)\,\mathrm{d}s, \tag{5}$$

for some $[0, 1]$–valued random variable Z. To ease the exposition, and with no loss of generality, we will work conditionally on a fixed value $Z = z$ which makes the Cox processes in (1) coincide with PRMs. According to the definition above, the marginals of a vector of GM–dependent CRMs are equally distributed and

$$\tilde{\mu}_i(\mathrm{d}x) = \int_{\mathbb{R}^+} s\, M_i(\mathrm{d}s, \mathrm{d}x) + \int_{\mathbb{R}^+} s\, M_0(\mathrm{d}s, \mathrm{d}x) = \mu_i(\mathrm{d}x) + \mu_0(\mathrm{d}x) \tag{6}$$

where μ_i, with $i = 1, 2$, and μ_0 are independent CRMs with Laplace functional transforms

$$\mathbb{E}\left[\mathrm{e}^{-\mu_i(f)}\right] = \mathrm{e}^{-cz\psi(f)} \qquad\qquad \mathbb{E}\left[\mathrm{e}^{-\mu_0(f)}\right] = \mathrm{e}^{-c(1-z)\psi(f)}$$

where ψ is defined as in (4). Given the simple form of the intensities specified in (5), one can determine the form of φ in (2) explicitly and straightforwardly obtains a tractable

expression for the joint Laplace functional transform of $(\tilde{\mu}_1, \tilde{\mu}_2)$ given by

$$\mathbb{E}\left[\mathrm{e}^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}\right] = \mathrm{e}^{-cz[\psi(f_1)+\psi(f_2)]-c(1-z)\psi(f_1+f_2)} \tag{7}$$

for any pair of measurable functions $f_i : \mathbb{X} \to \mathbb{R}$, for $i = 1, 2$, such that $\mathbb{P}[\tilde{\mu}_i(|f_i|) < \infty] = 1$. In order to further clarify the above concepts and construction, let us consider two special cases involving well–known CRMs.

Example 1. (*Gamma process*). Set $\rho(s) = \mathrm{e}^{-s}\, s^{-1}$ in (5) which results in μ being a gamma CRM. The corresponding Laplace exponent reduces to $\psi(f) = \int \log(1+f)\,\mathrm{d}P_0$ for any measurable function f such that $\int \log(1+|f|)\,\mathrm{d}P_0 < \infty$. The bivariate Laplace functional transform of the vector of dependent gamma CRMs, then, coincides with

$$\mathbb{E}\left[\mathrm{e}^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}\right] = \mathrm{e}^{-c\int \log(1+f_1+f_2)\mathrm{d}P_0 - cz\int \log\frac{(1+f_1)(1+f_2)}{(1+f_1+f_2)}\,\mathrm{d}P_0}$$

Example 2. (σ–*stable process*). Set $\rho(s) = \sigma\, s^{-1-\sigma}/\Gamma(1-\sigma)$, with $\sigma \in (0,1)$, in (5) which results in μ being a σ–stable CRM. Let $f_i : \mathbb{X} \to \mathbb{R}$ be such that $\int |f_i|^\sigma \,\mathrm{d}P_0 < \infty$, for $i = 1, 2$. Then

$$\mathbb{E}\left[\mathrm{e}^{-\tilde{\mu}_1(f_1)-\tilde{\mu}_2(f_2)}\right] = \mathrm{e}^{-cz\int (f_1^\sigma+f_2^\sigma)\mathrm{d}P_0 - c(1-z)\int (f_1+f_2)^\sigma\,\mathrm{d}P_0}$$

The final step needed for obtaining the desired vector of dependent random probability measures consists in normalizing the previously constructed CRMs, in the same spirit as in [38] for the univariate case. To perform the normalization we need to ensure $\mathbb{P}[\mu_i(\mathbb{X}) \in (0,\infty)] = 1$, for $i = 0, 1, 2$, which is guaranteed by requesting $\int_0^\infty \rho(s)\,\mathrm{d}s = \infty$ (see [38]) and corresponds to considering CRMs which jump infinitely often on any bounded set. By normalizing $\tilde{\mu}_1$ and $\tilde{\mu}_2$, we can then define the vector of dependent random probability measures

$$(\tilde{p}_1,\ \tilde{p}_2) \stackrel{d}{=} (\tilde{\mu}_1/\tilde{\mu}_1(\mathbb{X}),\ \tilde{\mu}_2/\tilde{\mu}_2(\mathbb{X})) \tag{8}$$

to be termed *GM–dependent normalized CRM* in the following.

Having described the main concepts and tools we are resorting to, our next goal is the application of $(\tilde{p}_1, \tilde{p}_2)$ as a nonparametric prior for the statistical analysis of partially exchangeable data.

3. Partially exchangeable sequences

For our purposes, we resort to the notion of partial exchangeability as set forth by de Finetti in [12] and described as follows. Let $X = (X_n)_{n\geq 1}$ and $Y = (Y_n)_{n\geq 1}$ be two sequences of $\mathbb{X}$–valued random elements defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and $P_{\mathbb{X}}$ is the space of probability measures on $(\mathbb{X}, \mathfrak{X})$. If $\boldsymbol{X}^{(n_1)} = (X_1, \ldots, X_{n_1})$ and $\boldsymbol{Y}^{(n_2)} =$

$(Y_1, \ldots, Y_{n_2})$ are the first n_1 and n_2 values of the sequences X and Y, respectively, we have

$$\mathbb{P}\left[\boldsymbol{X}^{(n_1)} \in A_1,\, \boldsymbol{Y}^{(n_2)} \in A_2\right] = \int_{P_{\mathbb{X}}^2} p_1^{n_1}(A_1)\, p_2^{n_2}(A_2)\, \Phi(\mathrm{d}p_1, \mathrm{d}p_2) \tag{9}$$

for any $A_1 \in \mathfrak{X}^{n_1}$, $A_2 \in \mathfrak{X}^{n_2}$, with $p_i^{n_i}$ being the n–fold product measure $p_i \times \cdots \times p_i$ and Φ is a probability distribution on $P_{\mathbb{X}}^2 = P_{\mathbb{X}} \times P_{\mathbb{X}}$ which acts as nonparametric prior for Bayesian inference. We also denote as Φ_i the marginal distribution of $\tilde{p}_i$ on $P_{\mathbb{X}}$. Since $\tilde{p}_i$ is a normalized CRM, then the weak support of Φ_i contains all probability measures on $\mathbb{X}$ whose support is contained in the support of the base measure P_0. Hence, if the support of P_0 coincides with $\mathbb{X}$, a GM–dependent normalized CRM $(\tilde{p}_1, \tilde{p}_2)$ has full weak support with respect to the product topology on $P_{\mathbb{X}}^2$. Having a large support is a minimal requirement a nonparametric prior must comply with in order to ensure some degree of flexibility in statistical analysis.

It should be also noted that the dependence structure displayed in assumption (9) is also the starting point in [4] where the authors propose an example (the first we are aware of in the literature) of nonparametric prior for partially exchangeable arrays which coincides with a mixture of products of Dirichlet processes. Furthermore, (9) defines the framework in which recent proposals of dependent nonparametric priors can be embedded.

3.1. Dependence between $\tilde{p}_1$ and $\tilde{p}_2$

An important preliminary result we state concerns the mixed moment of $(\tilde{p}_1(A), \tilde{p}_2(B))$ for any A and B in $\mathfrak{X}$. To this end, define the following quantity

$$\tau_q(u) := \int_0^\infty s^q \mathrm{e}^{-us}\, \rho(s)\, \mathrm{d}s \tag{10}$$

for any $q \geq 1$. Moreover, to simplify the notation set $\psi(u) = \psi(u\mathbb{1}_{\mathbb{X}})$ for any $u > 0$, where $\mathbb{1}_A$ is the indicator function on set A. One can, then, prove the following

Proposition 1. *Let $(\tilde{p}_1,\, \tilde{p}_2)$ be a vector of GM–dependent normalized CRM defined in (8). For any A and B in $\mathfrak{X}$ one has*

$$\begin{aligned}\mathbb{E}\left[\tilde{p}_1(A)\, \tilde{p}_2(B)\right] &= P_0(A)P_0(B) + \left[P_0(A \cap B) - P_0(A)P_0(B)\right] \\ &\quad \times\; c(1-z) \int_{(0,\infty)^2} \mathrm{e}^{-cz[\psi(u)+\psi(v)]-c(1-z)\psi(u+v)} \tau_2(u+v)\, \mathrm{d}u\mathrm{d}v \end{aligned} \tag{11}$$

Moreover, it follows that

$$\mathrm{Corr}(\tilde{p}_1(A), \tilde{p}_2(B)) = \frac{(1-z)\left[P_0(A \cap B) - P_0(A)P_0(B)\right]}{\sqrt{P_0(A)[1-P_0(A)]}\sqrt{P_0(B)[1-P_0(B)]}}\, \mathcal{I}(c, z) \tag{12}$$

where

$$\mathcal{I}(c,z) := \frac{\int_0^\infty \int_0^\infty \mathrm{e}^{-cz[\psi(u)+\psi(v)]-c(1-z)\psi(u+v)}\,\tau_2(u+v)\,\mathrm{d}u\mathrm{d}v}{\int_0^\infty u\,\mathrm{e}^{-c\psi(u)}\,\tau_2(u)\,\mathrm{d}u}$$

It can be easily seen that if $A = B$, then the correlation in (12) reduces to $(1-z)\mathcal{I}(c,z)$ and does not depend on the specific set where the two random probabilities $\tilde{p}_1$ and $\tilde{p}_2$ are evaluated. This fact is typically used to motivate $(1-z)\mathcal{I}(c,z)$ as a measure of the (overall) dependence between $\tilde{p}_1$ and $\tilde{p}_2$. Coherently with our construction $\tilde{p}_1$ and $\tilde{p}_2$ are uncorrelated if $z = 1$, and the same can be said if A and B are independent with respect to the baseline probability measure P_0. The previous expression is structurally neat and, as will be shown in the following illustrations, in some important special cases the double integral $\mathcal{I}(c,z)$ can be made sufficiently explicit so to allow a straightforward computation.

EXAMPLE 1. (continued). If $\tilde{\mu}_1$, $\tilde{\mu}_2$ are two dependent CRMs, one has $\tau_q(u) = \Gamma(q)\,(1+u)^{-q}$ and the correlation between the corresponding GM–dependent Dirichlet processes coincides with (12) where

$$\mathcal{I}(c,z) = \frac{c}{c+1}\,{}_3F_2(c-cz+2,1,1;c+2,c+2;1) \tag{13}$$

where ${}_3F_2$ is the generalized hypergeometric function

$${}_3F_2(\alpha,\beta,\rho;\gamma,\sigma;x) = \sum_{j\geq 0} \frac{(\alpha)_j(\beta)_j(\rho)_j}{j!(\gamma)_j(\sigma)_j}\,x^j \tag{14}$$

The above series converges if $|x| < 1$ and it does for $x = 1$ provided that $\mathrm{Re}(\gamma+\sigma-\alpha-\beta-\rho) > 0$, with $\mathrm{Re}(z)$ denoting the real part of a complex number z. □

EXAMPLE 2. (continued). If $\tilde{\mu}_1$, $\tilde{\mu}_2$ are σ–stable dependent CRMs, one has $\tau_q(u) = \sigma(1-\sigma)_{q-1}\,u^{\sigma-q}$ and the correlation between the corresponding dependent normalized σ–stable processes is equal to (12) with

$$\mathcal{I}(c,z) = \frac{1}{\sigma}\int_0^1 \frac{w^{1/\sigma-1}}{[1+z(1-w^{1/\sigma})^\sigma - z(1-w)]}\,\mathrm{d}w$$

Even if we are not able to evaluate the above integral analytically, a numerical approximation can be easily determined. □

3.2. Partition probability function

The procedure adopted for determining an expression for the mixed moments of $\tilde{p}_1$ and $\tilde{p}_2$ can be extended to provide a form for the partially exchangeable partition probability function (pEPPF) for the n_1+n_2 random variables (r.v.'s) $\boldsymbol{X}^{(n_1)}$ and $\boldsymbol{Y}^{(n_2)}$. It is worth

recalling that the concept of EPPF plays an important role in modern Probability Theory (see [35] and references therein) and, implicitly, in numerous MCMC algorithms one ends up "sampling from the partition" as well. First, note that if $z < 1$

$$\mathbb{P}[X_i = Y_j] = c \int_0^\infty u \, \mathrm{e}^{-c\psi(u)} \, \tau_2(u) \, \mathrm{d}u > 0$$

for any i and j: hence, with positive probability any of the elements of the first sample $\boldsymbol{X}^{(n_1)}$ can coincide with any element from $\boldsymbol{Y}^{(n_2)}$. This leads us to address the issue of determining the probability that the two samples are partitioned into $K = K_1 + K_2 + K_0$ clusters of distinct values where

(a) K_1 is the number of distinct values in the first sample $\boldsymbol{X}^{(n_1)}$ not coinciding with any of the Y_j's
(b) K_2 is the number of distinct values in the second sample $\boldsymbol{Y}^{(n_2)}$ not coinciding with any of the X_j's
(c) K_0 is the number of distinct values that are shared by both samples $\boldsymbol{X}^{(n_1)}$ and $\boldsymbol{Y}^{(n_2)}$.

Moreover, we denote by $\boldsymbol{N}^{(i)} = (N_{1,i}, \ldots, N_{K_i,i})$ the vector of frequencies for the K_i unshared clusters and with $\boldsymbol{Q}^{(i)} = (Q_{1,i}, \ldots, Q_{K_0,i})$ the vector of frequencies the sample $\boldsymbol{X}^{(n_1)}$, if $i = 1$, or the sample $\boldsymbol{Y}^{(n_2)}$, if $i = 2$, contributes to each of the shared clusters. Correspondingly, we introduce the sets of vectors of positive integers

$$\Delta_{n_i,k_i,k_0} := \left\{ (\boldsymbol{n}^{(i)}, \boldsymbol{q}^{(i)}) : \sum_{l=1}^{k_i} n_{l,i} + \sum_{r=1}^{k_0} q_{r,i} = n_i \right\}$$

where we the more concise notation $\boldsymbol{n}^{(i)} = (n_{1,i}, \ldots, n_{k_i,i})$ and $\boldsymbol{q}^{(i)} = (q_{1,i}, \ldots, q_{k_0,i})$ is used, for $i = 1, 2$. The result we are going to state provides the probability distribution of the random partition of $(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)})$ as encoded by the vector of positive integers $(K_1, K_2, K_0, \boldsymbol{N}^{(1)}, \boldsymbol{N}^{(2)}, \boldsymbol{Q}^{(1)}, \boldsymbol{Q}^{(2)})$. Such a distribution has masses at points $(k_1, k_2, k_0, \boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)})$ that we denote as $\Pi_k^{(n_1+n_2)}(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)})$, where $k = k_1 + k_2 + k_0$.

Proposition 2. *Let $(\tilde{p}_1, \tilde{p}_2)$ be a GM–dependent normalized CRM defined in (8). For any $(\boldsymbol{n}^{(i)}, \boldsymbol{q}^{(i)}) \in \Delta_{n_i,k_i,k_0}$, with $i = 1, 2$, and for any non-negative integers k_1, k_2 and k_0 such that $k_l + k_0 \in \{1, \ldots, n_l\}$, for $l = 1, 2$, one has*

$$\Pi_k^{(n_1+n_2)}(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}) = \frac{c^k}{\Gamma(n_1)\Gamma(n_2)} \sum_{(*)} (1-z)^{k_0+|\boldsymbol{i}|+|\boldsymbol{l}|} z^{k_1+k_2-|\boldsymbol{i}|-|\boldsymbol{l}|}$$

$$\times \int_0^\infty \int_0^\infty u^{n_1-1} v^{n_2-1} \, \mathrm{e}^{-cz[\psi(u)+\psi(v)]-c(1-z)\psi(u+v)} \prod_{j=1}^{k_1} \tau_{n_{j,1}}(u + i_j v)$$

$$\times \prod_{j=1}^{k_2} \tau_{n_{j,2}}(l_j u + v) \prod_{r=1}^{k_0} \tau_{q_{r,1}+q_{r,2}}(u+v)\,\mathrm{d}u\,\mathrm{d}v$$

where the sum runs over the set of all vectors of integers $\boldsymbol{i} = (i_1, \ldots, i_{k_1}) \in \{0,1\}^{k_1}$ *and* $\boldsymbol{l} = (l_1, \ldots, l_{k_2}) \in \{0,1\}^{k_2}$*, whereas* $|\boldsymbol{i}| = \sum_{j=1}^{k_1} i_j$ *and* $|\boldsymbol{l}| = \sum_{j=1}^{k_2} l_j$*.*

The expression, though in closed form and of significant theoretical interest, is quite difficult to evaluate due to the presence of the sum with respect to the integer vectors $\boldsymbol{i}$ and $\boldsymbol{l}$. Nonetheless, Proposition 2 is going to be a fundamental tool for the derivation of the MCMC algorithm we adopt for density estimation and for inferring on the clustering structure of the two samples. We will be able to skip the evaluation of the sum by resorting to suitable auxiliary variables whose full conditionals can be determined and evaluated. To clarify this point, consider the first sample $\boldsymbol{X}^{(n_1)}$, fix $\boldsymbol{i} \in \{0,1\}^{k_1}$ and denote by $\boldsymbol{n}^0_{(\boldsymbol{i})}$ the vector of cluster frequencies that correspond to labels in $\boldsymbol{i}$ equal to 0 whereas $\boldsymbol{n}^1_{(\boldsymbol{i})}$ is the vector of cluster frequencies corresponding to labels in $\boldsymbol{i}$ equal to 1. In a similar fashion, for the second sample $\boldsymbol{X}^{(n_2)}$, for $\boldsymbol{l} \in \{0,1\}^{k_2}$, set $\boldsymbol{n}^0_{(\boldsymbol{l})}$ and $\boldsymbol{n}^1_{(\boldsymbol{l})}$. Finally, let $\boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})} = (\boldsymbol{n}^1_{(\boldsymbol{i})}, \boldsymbol{n}^1_{(\boldsymbol{l})}, q_{1,1} + q_{1,2}, \ldots, q_{k_0,1} + q_{k_0,2})$. From these definitions it is obvious that $\boldsymbol{n}^0_{(\boldsymbol{i})}$, $\boldsymbol{n}^0_{(\boldsymbol{l})}$ and $\boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})}$ are vectors with $k_1 - |\boldsymbol{i}|$, $k_2 - |\boldsymbol{l}|$ and $k_0 + |\boldsymbol{i}| + |\boldsymbol{l}|$ coordinates, respectively. Moreover, let λ_1, λ_2 and λ_0 be permutations of the coordinates of the vectors $\boldsymbol{n}^0_{(\boldsymbol{i})}$, $\boldsymbol{n}^0_{(\boldsymbol{l})}$ and $\boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})}$. We shall further denote

$$\Pi^{(n_1+n_2)}_{k,\boldsymbol{i},\boldsymbol{l}}(\boldsymbol{n}^0_{(\boldsymbol{i})}, \boldsymbol{n}^0_{(\boldsymbol{l})}, \boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})})$$

as the pEPPF conditional on independent random variables $\boldsymbol{i}$ and $\boldsymbol{l}$ whose distribution is Bernoulli with parameter $(1-z)$. Moreover, note that the pEPPF $\Pi^{(n_1+n_2)}_k$ depends on the vectors $\boldsymbol{q}^{(i)}$, for $i = 1, 2$, through their componentwise sum $\boldsymbol{q}^* = (q_{1,1} + q_{1,2}, \ldots, q_{k_0,1} + q_{k_0,2})$. Hence we can also write

$$\Pi^{(n_1+n_2)}_k(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}) = \Pi^{(n_1+n_2)}_k(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*)$$

and shall denote as λ_1', λ_2' and λ_0' permutations of the components in $\boldsymbol{n}^{(1)}$, $\boldsymbol{n}^{(2)}$ and $\boldsymbol{q}^*$, respectively. Therefore, as a straightforward consequence of Proposition 2 we obtain the following invariance property for $\Pi^{(n_1+n_2)}_k$ and for $\Pi^{(n_1+n_2)}_{k,\boldsymbol{i},\boldsymbol{l}}$ whose proof is omitted since it is immediate.

PROPOSITION 3. *Let* $(\tilde{p}_1, \tilde{p}_2)$ *be a GM–dependent normalized CRM defined in* (8)*. Then*

$$\Pi^{(n_1+n_2)}_k(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*) = \Pi^{(n_1+n_2)}_k(\lambda_1'\,\boldsymbol{n}^{(1)}, \lambda_2'\,\boldsymbol{n}^{(2)}, \lambda_0'\,\boldsymbol{q}^*) \tag{15}$$

$$\Pi^{(n_1+n_2)}_{k,\boldsymbol{i},\boldsymbol{l}}(\boldsymbol{n}^0_{(\boldsymbol{i})}, \boldsymbol{n}^0_{(\boldsymbol{l})}, \boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})}) = \Pi^{(n_1+n_2)}_{k,\boldsymbol{i},\boldsymbol{l}}(\lambda_1\boldsymbol{n}^0_{(\boldsymbol{i})}, \lambda_2\boldsymbol{n}^0_{(\boldsymbol{l})}, \lambda_0\boldsymbol{n}_{(\boldsymbol{i},\boldsymbol{l})}) \tag{16}$$

The invariance property in (15) entails that exchangeability holds true within three separate groups of clusters: those with non–shared values and the clusters shared by the two samples. Such a finding is not a surprise since it reflects the partial exchangeability assumption. On the other hand, (16) implies that, conditional on a realization of $\boldsymbol{i}$ and $\boldsymbol{l}$ whose components are i.i.d. Bernoulli random variables with parameter $1-z$, a similar partially exchangeable structure is revealed even if it now involves different groupings of the clusters that are still three: two groups with non–shared values that are labeled either by i_j or l_j equal to 0, and the group containing both observations shared by the two samples and non–shared values labeled by either i_j or l_j equal to 1. Moreover, unlike (15) these three groups of clusters are governed by independent random probability measures. The invariance structure displayed in (16) corresponds to a mixture decomposition for $\tilde{p}_1$ and $\tilde{p}_2$ that is going to be displayed in the next Section and is also relevant in simplifying the MCMC sampling scheme we are going to devise. Note that (15) holds true since the sum appearing in the representation of $\Pi_k^{(n_1+n_2)}$ is over all possible $\{0,1\}$–valued indices i_j and l_j: hence a permutation of the frequency vectors within the three groups simply yields a permutation of the summands in Proposition 2. On the contrary, fixing the indices i_j and l_j as in (16) corresponds to dropping the sum in $\Pi_k^{(n_1+n_2)}$ and, then, the invariance is restricted to those frequencies that correspond to the same index values.

Example 1. (continued). Let $(\tilde{\mu}_1, \tilde{\mu}_2)$ be a vector of GM–dependent gamma CRMs. If $\boldsymbol{i} = (i_1, \ldots, i_{k_1}) \in \{0,1\}^{k_1}$ and $\boldsymbol{l} = (l_1, \ldots, l_{k_2}) \in \{0,1\}^{k_2}$ define $\bar{n}_1 = \sum_{j=1}^{k_1}(1-i_j)n_{j,1}$, $\bar{n}_2 = \sum_{j=1}^{k_2}(1-l_j)n_{j,2}$, $\bar{n}_{1,0} = \sum_{j=1}^{k_1} i_j n_{j,1}$. Moreover, to further simplify notation, set

$$\xi_\sigma(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*) = \prod_{j=1}^{k_1}(1-\sigma)_{n_{j,1}-1} \prod_{i=1}^{k_2}(1-\sigma)_{n_{i,2}-1} \prod_{r=1}^{m}(1-\sigma)_{q_{r,1}+q_{r,2}-1},$$

$\alpha' = c + cz + |\boldsymbol{q}^*|$ and $\beta' = c + \bar{n}_{1,0} + |\boldsymbol{q}^*|$. It can then be shown that the pEPPF of the GM–dependent Dirichlet process is then given by

$$\Pi_k^{(n_1+n_2)}(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*) = c^k\, \xi_0(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*)$$
$$\times \sum_{(*)} \frac{z^{k_1+k_2-|\boldsymbol{i}|-|\boldsymbol{j}|}(1-z)^{k_0+|\boldsymbol{i}|+|\boldsymbol{j}|}}{(\alpha')_{n_1}(\beta')_{n_2}}\, {}_3F_2(cz+\bar{n}_2, \beta', n_1; n_1+\alpha', n_2+\beta'; 1)$$

for any $(\boldsymbol{n}^{(i)}, \boldsymbol{q}^{(i)}) \in \Delta_{n_i,k_i,k_0}$, for $i = 1, 2$, and for any $k_1 \leq n_1$, $k_2 \leq n_2$ and k_0 such that $k = k_1 + k_2 + k_0 \in \{1, \ldots, n_1 + n_2\}$. Note also that if there is only one sample, namely $n_1 n_2 = 0$, the previous pEPPF reduces to the EPPF of the Dirichlet process determined in [10, 1]. □

Example 2. (continued). When $(\tilde{\mu}_1, \tilde{\mu}_2)$ is a vector of GM–dependent σ–stable CRMs, one obtains a pEPPF of the form

$$\Pi_k^{(n_1+n_2)}(\boldsymbol{n}^{(1)},\boldsymbol{n}^{(2)},\boldsymbol{q}^*)=\frac{\Gamma(k)}{\Gamma(n_1)\Gamma(n_2)}\sigma^{k-1}\,\xi_\sigma(\boldsymbol{n}^{(1)},\boldsymbol{n}^{(2)},\boldsymbol{q}^*)$$

$$\times\sum_{(*)} z^{k_1+k_2-|\boldsymbol{i}|-|\boldsymbol{l}|}\,(1-z)^{k_0+|\boldsymbol{i}|+|\boldsymbol{l}|}\int_0^1\frac{w^{n_1-\bar n_1+(k_1-|\boldsymbol{i}|)\sigma-1}(1-w)^{n_2-\bar n_2+(k_2-|\boldsymbol{l}|)\sigma-1}}{[1-z+zw^\sigma+z(1-w)^\sigma]^k}\,\mathrm{d}w$$

where $\bar n_1$, $\bar n_2$, $\bar n_{1,0}$ are defined as in Example 1. Note that the one–dimensional integral above has the same structure as the one appearing in $\mathcal{I}(c,z)$ and can be evaluated numerically. Also in this case, if $n_1n_2=0$ the above expression reduces to the EPPF of the normalized σ–stable process. See, e.g., [35]. □

Remark 1. Following a request of the referees, we also sketch the extension to more than a pair of dependent random probability measures the most natural being $\tilde\mu_i=\mu_i+\mu_0$, for each $i=1,\dots,N$ and $N>2$. If the mutually independent CRMs μ_i are identical in distribution, for $i=1,\dots,N$, and independent from the common source of randomness μ_0, one immediately obtains that the joint Laplace transform of the vector $(\tilde\mu_1,\dots,\tilde\mu_N)$ evaluated at a vector function $(f_1,\dots,f_N)$ is given by

$$\mathbb{E}\left[\mathrm{e}^{-\sum_{i=1}^N\tilde\mu_i(f_i)}\right]=\mathrm{e}^{-c(1-z)\,\psi(|\boldsymbol{f}|)-cz\sum_{i=1}^N\psi(f_i)}$$

where ψ is the Laplace exponent defined in (4) and shared by the μ_i's $(i=0,1,\dots,N)$ and $|\boldsymbol{f}|=\sum_{i=1}^N f_i$. This expression can be used to mimic the proof of Proposition 2 thus yielding a straightforward generalization of the pEPPF associated to the vector $(\tilde p_1,\tilde p_2)$ to the N–dimensional case obtaining an expression of the form

$$\frac{c^k}{\prod_{j=1}^N\Gamma(n_j)}\sum_{(*)}(1-z)^{k_0+\sum_{j=1}^N|\boldsymbol{i}_j|}\,z^{\sum_{j=1}^N(k_j-|\boldsymbol{i}_j|)}\int_{(0,\infty)^N}\prod_{j=1}^N u_j^{n_j-1}$$

$$\times\ \mathrm{e}^{-c(1-z)\,\psi(|\boldsymbol{u}|)-cz\sum_{i=1}^N\psi(u_i)}\prod_{j=1}^N\prod_{l=1}^{k_j}\tau_{i_{l,j}}(u_j)\,\tau_{n_{l,j}-i_{l,j}}(|\boldsymbol{u}|)$$

$$\times\prod_{l=1}^{k_0}\tau_{q_l}(|\boldsymbol{u}|)\,\mathrm{d}u_1,\dots,\mathrm{d}u_N$$

where the $(*)$ is the set of all vectors $\boldsymbol{i}_j=(i_{1,j},\dots,i_{k_j,j})\in\times_{l=1}^{k_j}\{0,n_{l,j}\}$, for $j=1,\dots,N$, $|\boldsymbol{u}|=\sum_{i=1}^N u_i$ and $|\boldsymbol{i}_j|=\sum_{l=1}^{k_j}i_{l,j}$. Moreover, the definition of τ_q in (10) is extended to cover the case with $q=0$ as $\tau_0(u)=1$ for any $u>0$. The previous expression provides the probability of observing an array of N samples, with respective sizes $n_1,\dots,n_N$, with observations partitioned into k_j clusters specific to the j–th sample and k_0 groups shared by two or more samples. The exact evaluation of the above N–dimensional integral poses some additional challenges and its implementation within a sampling scheme is more demanding. A notable exception is given by the GM–dependent Dirichlet process where for computational purposes one can avoid the use of the pEPPF and rely on a mixture representation of $\tilde p_i$ and $\tilde p_0$ that will be detailed at the beginning of the next section. □

4. Dependent mixtures

We now apply the general results for GM–dependent normalized CRMs $(\tilde{p}_1, \tilde{p}_2)$ to mixture models with random dependent densities. In fact, we consider data that are generated from random densities $\tilde{f}_1$ and $\tilde{f}_2$ defined by $\tilde{f}_i(x) = \int_\Theta h_i(x;\theta)\,\tilde{p}_i(\mathrm{d}\theta)$, for $i = 1, 2$, with Θ being a complete and separable metric space equipped with the corresponding Borel σ–algebra. If $\boldsymbol{\theta}^{(i)} = (\theta_{1,i}, \ldots, \theta_{n_i,i})$, for $i = 1, 2$, stand for vectors of latent variables corresponding to the two samples, the mixture model can be represented in hierarchical form as

$$\begin{aligned} X_{i,1} \,|\, (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) &\overset{\text{ind}}{\sim} h_1(\,\cdot\,;\theta_{i,1}) \qquad i = 1, \ldots, n_1 \\ Y_{j,2} \,|\, (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) &\overset{\text{ind}}{\sim} h_2(\,\cdot\,;\theta_{j,2}) \qquad j = 1, \ldots, n_2 \\ (X_{i,1}, Y_{j,2}) \,|\, (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}) &\overset{\text{ind}}{\sim} h_1(\,\cdot\,;\theta_{i,1})\, h_2(\,\cdot\,;\theta_{j,2}) \\ \theta_{j,i} \,|\, (\tilde{p}_1, \tilde{p}_2) &\overset{\text{iid}}{\sim} \tilde{p}_i \qquad\qquad j = 1, \ldots, n_i; \quad i = 1, 2 \\ (\tilde{p}_1, \tilde{p}_2) &\overset{d}{=} \text{ GM–dependent normalized CRM} \end{aligned} \tag{17}$$

Henceforth, we will set $h_1 = h_2 = h$; the case of $h_1 \neq h_2$ can be handled in a similar fashion, with the obvious variants. The investigation of distributional properties of the model is eased by rewriting $\tilde{p}_1$ and $\tilde{p}_2$ in the following mixture form

$$\tilde{p}_i = w_i p_i + (1 - w_i) p_0 \qquad i = 1, 2 \tag{18}$$

where $w_i = \mu_i(\mathbb{X})\{\mu_i(\mathbb{X}) + \mu_0(\mathbb{X})\}^{-1}$, the p_i's and p_0 are independent normalized CRMs with Lévy intensities $czP_0(\mathrm{d}x)\,\rho(s)\mathrm{d}s$ and $c(1-z)P_0\,\rho(s)\mathrm{d}s$, respectively. Obviously w_1 and w_2 are dependent. In general, the weights w_i and the p_i's are dependent, the only exception being the case in Example 1 where the p_i's are independent Dirichlet processes. Details about this special case will be provided later.

Remark 2. An interesting aspect of (18) is that each $\tilde{p}_i$ can be decomposed into two independent sources of randomness: an idiosyncratic one, p_i, and a common one, p_0. This is close in spirit to the model of Müller, Quintana and Rosner [30], which is based on a vector of dependent random probability measures $(\tilde{p}_1, \ldots, \tilde{p}_n)$ defined as

$$\tilde{p}_i = \omega p_i + (1 - \omega) p_0, \tag{19}$$

where p_i and p_0 are independent Dirichlet processes and the distribution of ω is a mixture with point masses $\omega = 0$ and $\omega = 1$ and the remaining mass spread on $(0, 1)$ through a beta density. Despite their similarity, there are however some crucial differences among GM–dependent normalized CRMs and the model in (19) so that it is not possible to interpret one as the generalization of the other, nor viceversa. The first thing to note is that (19) assumes common weights, ω and $1-\omega$, for each $\tilde{p}_i$ whereas in our proposal the weights of the mixtures w_i in (18) do not coincide for different i even if they have the same marginal distributions. More importantly, the random probability measures defined

in [30] via (19) are, in general, marginally not Dirichlet processes. In our framework, preserving the marginal Dirichlet structure or, in general, a normalized CRM structure is relevant: it guarantees the degree of analytical tractability we need for determining distributional results and devising suitable sampling strategies. The latter can then be thought of as alternative to the existing algorithms for dependent random probability measures such as, e.g., the one proposed in [30].

On the basis of the decomposition displayed in (18), one can introduce two collections of auxiliary random variables, $(\zeta_{j,1})_{j\geq 1}$ and $(\zeta_{j,2})_{j\geq 1}$, defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $\{0,1\}^\infty$ and $\{0,2\}^\infty$, and provide an useful alternative representation of the mixing measure in (17) in terms of these auxiliary variables as

$$\begin{aligned} \theta_{i,1} \,|\, \zeta_{i,1}, \mu_1, \mu_2, \mu_0 &\overset{\text{ind}}{\sim} p_{\zeta_{i,1}} && i = 1, \ldots, n_1 \\ \theta_{j,2} \,|\, \zeta_{j,2}, \mu_1, \mu_2, \mu_0 &\overset{\text{ind}}{\sim} p_{\zeta_{j,2}} && j = 1, \ldots, n_2 \\ (\zeta_{i,1}, \zeta_{j,2}) \,|\, \mu_1, \mu_2, \mu_0 &\overset{\text{ind}}{\sim} \text{bern}\,(w_1; \{0,1\}) \times \text{bern}\,(w_2; \{0,2\}) \end{aligned} \tag{20}$$

where $X \sim \text{bern}(w; \{a,b\})$ means that $\mathbb{P}[X = b] = 1 - \mathbb{P}[X = a] = w$ for $w \in [0,1]$ and $a, b \in \mathbb{R}$. The latent variables $\boldsymbol{\theta}^{(i)}$ are, then, governed by GM–dependent normalized CRMs. Therefore, we can resort to results established in Section 3.2 to obtain the full conditional distributions for all the quantities that need to be sampled in order to attain posterior inferences. Given the structure of the model, the latent $\boldsymbol{\theta}^{(i)}$, $i = 1, 2$, might feature ties which generate, according to the notation we have already introduced, $k_1 + k_2 + k_0$ clusters. Our analysis of the partition of the $\boldsymbol{\theta}^{(i)}$'s will further benefit from the following fact that is a straightforward consequence of Proposition 2.

Corollary 1. *Let $(\tilde{p}_1, \tilde{p}_2)$ be a GM–dependent normalized CRM defined in (8). Suppose P_0 is a non–atomic probability measure on $(\mathbb{X}, \mathfrak{X})$. Then*

$$\mathbb{P}[\theta_{i,1} = \theta_{j,2} \,|\, \zeta_{i,1} \neq \zeta_{j,2}] = 0. \tag{21}$$

Hence, (21) entails that ties between the two groups $\boldsymbol{\theta}^{(1)}$ and $\boldsymbol{\theta}^{(2)}$ may arise with positive probability only if any two $\theta_{i,1}$ and $\theta_{j,2}$ share the same label $\zeta_{i,1} = \zeta_{j,2} = 0$. This is a structural property of the model and it intuitively means that there cannot be overlaps between the different sources of randomness involved, which seems desirable.

Suppose $\boldsymbol{\theta}^{(i)}_* = (\theta^*_{1,i}, \ldots, \theta^*_{k_i,i})$, for $i = 1, 2$, and $\boldsymbol{\theta}_* = (\theta^*_1, \ldots, \theta^*_{k_0})$ denote the vectors of unique distinct values associated to the $K = K_1 + K_2 + K_0$ clusters. The corresponding partition is

$$\widetilde{\boldsymbol{\pi}}_{n_1,n_2} = \bigcup_{i=1}^{2} \{C_{j,i} : \, j = 1, \ldots, K_i\} \cup \{C_{j,i,0} : \, j = 1, \ldots, K_0\} \tag{22}$$

where $r, s \in C_{j,i}$ means that $\theta_{r,i} = \theta_{s,i} = \theta^*_{j,i}$, whereas $r_1 \in C_{j,1,0}$ and $r_2 \in C_{j,2,0}$ implies that $\theta_{r_1,1} = \theta_{r_2,2} = \theta^*_j$. It is clear, from the specification of the model (17), that the

conditional density of the data $(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)})$, given the partition $\widetilde{\boldsymbol{\pi}}_{n_1.n_2} = \boldsymbol{\pi}_{n_1,n_2}$ and the distinct latent variables $\boldsymbol{\theta}^* = (\boldsymbol{\theta}^{(1)}_*, \boldsymbol{\theta}^{(2)}_*, \boldsymbol{\theta}_*)$, coincides with

$$f(\boldsymbol{x}, \boldsymbol{y} \,|\, \boldsymbol{\theta}^*, \boldsymbol{\pi}_{n_1,n_2}) = \prod_{j=1}^{k_1} \prod_{i \in C_{j,1}} h(x_i; \theta^*_{j,1}) \prod_{\ell=1}^{k_2} \prod_{i \in C_{\ell,2}} h(y_i; \theta^*_{\ell,2}) \prod_{r=1}^{k_0} \prod_{i \in C_{r,1,0}} h(x_i; \theta^*_r) \prod_{\ell \in C_{r,2,0}} h(y_\ell; \theta^*_r) \quad (23)$$

Finally, set

$$\mathcal{L}^*(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}, \mathrm{d}\pi, \mathrm{d}\boldsymbol{\theta}, \mathrm{d}\boldsymbol{\zeta}) \quad (24)$$

as the distribution of the data $(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)})$, the partition $\widetilde{\boldsymbol{\pi}}_{n_1,n_2}$ in (22), the vector of unique values in $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)})$ and the labels $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)})$. If $n = n_1 + n_2$, then $\mathcal{L}^*$ is a probability distribution on the product space $\mathbb{X}^n \times \mathcal{P}_n \times \Theta^n \times \{0,1\}^{n_1} \times \{0,2\}^{n_2}$, where $\mathcal{P}_n$ is the space of all possible realizations of the random partition $\widetilde{\boldsymbol{\pi}}_{n_1,n_2}$ in (22). The determination of $\mathcal{L}^*$ will be first given for any pair of GM–dependent normalized CRMs. The specific expressions valid for dependent mixtures of the Dirichlet and the normalized σ–stable processes will be established as straightforward corollaries. In the sequel, we also denote as g_0 a density of P_0 with respect to some σ–finite dominating measure H on Θ, namely $g_0 = \mathrm{d}P_0/\mathrm{d}H$.

PROPOSITION 4. *Let $(\tilde{p}_1, \tilde{p}_2)$ be a GM–dependent normalized CRM defined in (8). Moreover, let $\boldsymbol{\zeta}^*_i = (\zeta^*_{1,i}, \ldots, \zeta^*_{k_i,i})$ be the vectors of labels corresponding to the distinct latent variables $\boldsymbol{\theta}^{(i)}_*$, with $i = 1, 2$. For the dependent mixture model in (17), the distribution $\mathcal{L}^*$ in (24) has density given by*

$$g(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{\zeta}^*)\, f(\boldsymbol{x}, \boldsymbol{y} \,|\, \boldsymbol{\theta}^*, \boldsymbol{\pi}_{n_1,n_2}) \prod_{i=1}^{k} g_0(\theta^*_i) \quad (25)$$

where

$$\begin{aligned} g(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{\zeta}^*) &= \frac{c^k\, z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{k_1+k_2-\tilde{k}_1-\tilde{k}_2}}{\Gamma(n_1)\Gamma(n_2)} \\ &\times \int_0^\infty \int_0^\infty u^{n_1-1} v^{n_2-1} \mathrm{e}^{-cz[\psi(u)+\psi(v)]-c(1-z)\psi(u+v)} \\ &\times \prod_{j=1}^{k_1} \tau_{n_{j,1}}(u + (1-\zeta^*_{j,1})v) \prod_{j=1}^{k_2} \tau_{n_{j,2}}((1-\zeta^*_{j,2}/2)u + v) \\ &\times \prod_{r=1}^{k_0} \tau_{q_{r,1}+q_{r,2}}(u+v)\, \mathrm{d}u\, \mathrm{d}v \end{aligned} \quad (26)$$

where $\tilde{k}_1 = |\boldsymbol{\zeta}_*^{(1)}|$ *and* $\tilde{k}_2 = |\boldsymbol{\zeta}_*^{(2)}|/2$ *identify the number of clusters with label* 1 *and* 2*, respectively.*

Before examining the details of the models we will refer to for illustrative purposes, it should be recalled that our approach yields posterior estimates of $\tilde{f}_1$ and $\tilde{f}_2$ and of the number of clusters $K_{\boldsymbol{X}}$ and $K_{\boldsymbol{Y}}$ into which one can group the two sample data. Another interesting issue concerns the estimation of statistical functionals of $\tilde{f}_1$ and of $\tilde{f}_2$, which has been addressed in the exchangeable case by Gelfand and Kottas [13]. Their approach is based on a suitable truncation of the stick-breaking representation of the Dirichlet process. In order to extend their techniques to this setting, a representation of the posterior distribution of a pair of GM–dependent normalized CRMs is still missing.

4.1. Dependent mixtures of Dirichlet processes

If the vector $(\tilde{p}_1, \tilde{p}_2)$ is a GM–dependent Dirichlet process as in Example 1, then one finds out that the weights (w_1, w_2) in (18) and the Dirichlet process components p_i, for $i = 0, 1, 2$, are independent and the density function of the vector (w_1, w_2) is

$$f(w_1, w_2) = \frac{\Gamma(c + cz)}{\Gamma^2(cz)\Gamma(c - cz)} \frac{(w_1 w_2)^{cz-1}[(1 - w_1)(1 - w_2)]^{c-1}}{(1 - w_1 w_2)^{c+cz}} \mathbb{1}_{[0,1]^2}(w_1, w_2) \tag{27}$$

This corresponds to the bivariate beta distribution introduced [32]. This model is analyzed in [22, 37], where independence between (w_1, w_2) and (p_0, p_1, p_2) is used to devise a sampler that includes sampling the weights w_i. Here we marginalize with respect to both the weights (w_1, w_2) and the random independent Dirichlet processes p_i, for $i = 0, 1, 2$. The first marginalization is trickier and is achieved by virtue of the results in Section 3.2.

Corollary 2. *Let* $(\tilde{p}_1, \tilde{p}_2)$ *be a GM–dependent Dirichlet process. A density of the probability distribution* $\mathcal{L}^*$ *defined in* (24) *coincides with*

$$\frac{c^k z^{\tilde{k}_1 + \tilde{k}_2}(1 - z)^{k_1 + k_2 - \tilde{k}_1 - \tilde{k}_2}}{(\alpha)_{n_1}(\beta)_{n_2}} {}_3F_2(\alpha - cz + n_1 - \bar{n}_1, n_1, n_2; \alpha + n_1, \beta + n_2; 1)$$

$$\times\ \xi_0(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*)\, f(\boldsymbol{x}, \boldsymbol{y} \,|\, \boldsymbol{\theta}^*, \boldsymbol{\pi}_{n_1,n_2}) \prod_{i=1}^{k} g_0(\theta_i^*)$$

where $\bar{n}_1 = |\boldsymbol{\zeta}^{(1)}|$, $\bar{n}_2 = |\boldsymbol{\zeta}^{(2)}|/2$, $\alpha = c + n_2 - \bar{n}_2$ *and* $\beta = c + n_1 - \bar{n}_1$.

As for the actual implementation of the model, a Gibbs sampler easily follows from Corollary 2. A key issue is the sampling of the labels. This can be done by first observing the following facts: (i) if $\theta_{i,1} = \theta_{j,2}$ then, by Corollary 1, the corresponding labels are zero, namely $\zeta_{i,1} = \zeta_{j,2} = 0$; (ii) given the partition $\boldsymbol{\pi}$, the dimensions of label vectors can be shrunk so that one basically has k labels corresponding to the $k = k_1 + k_2 + k_0$

clusters of the partition. Remark (i) implies that we do not need to sample the labels associated to $\theta_{i,1}$ values coinciding with any of the $\theta_{j,2}$'s and viceversa. Moreover, Remark (ii) implies that for any $r, s \in C_{j,i}$ one has $\zeta_{r,i} = \zeta_{s,i}$ and, thus, we need to sample only labels $\zeta^*_{j,i}$ corresponding to distinct values $\theta^*_{j,i}$. Finally, there might be $\theta_{j,1}$'s (or $\theta_{j,2}$'s) associated to $\zeta_{j,1} = 0$ (or $\zeta_{j,2} = 0$) that do not coincide with any of the $\theta_{i,2}$'s (or of the $\theta_{i,1}$'s): the corresponding labels are not degenerate and must be sampled from their full conditionals. If $\boldsymbol{\zeta}^{(1)}_{-j,*}$ stands for the vector $\boldsymbol{\zeta}^{(1)}_*$ with the j–th component removed, we use the short notation

$$\pi_{j,1}(x) := \mathbb{P}[\zeta^*_{j,1} = x \,|\, \boldsymbol{\zeta}^{(1)}_{-j,*}, \boldsymbol{\zeta}^{(2)}_*, \boldsymbol{\theta}^*, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}].$$

Hence, if $\theta^*_{j,1}$ does not coincide with any of the distinct values of the latent variables for the second sample, it can be easily deduced that

$$\begin{aligned}\pi_{j,1}(x) \propto \mathbb{1}_{\{0,1\}}(x)\, \frac{z^x\,(1-z)^{1-x}}{(\alpha)_{n_2}(\beta_x)_{n_2}} \\ \times\ {}_3F_2\left(\alpha - cz + n_1 - \bar{n}_{-j,1} - x n_{j,1}, n_1, n_2; \alpha + n_1, \beta_x + n_2; 1\right) \quad (28)\end{aligned}$$

where $\bar{n}_{-j,1} := \sum_{i \neq j} n_{i,1}\zeta^*_{i,1}$ with $n_{i,1}$ denoting the size of the cluster identified by $\theta^*_{i,1}$. Moreover, $\beta_x = c + n_1 - \bar{n}_{-j,1} - x n_{j,1}$. Obviously, the normalizing constant is determined by $\pi_{j,i}(0) + \pi_{j,i}(1) = 1$. The full conditionals for the $\zeta^*_{j,2}$ can be determined analogously.

As for the full conditionals of the $\theta_{j,i}$'s, these reduce to the ones associated to the univariate mixture of the Dirichlet process, since one is conditioning on the labels $\zeta_{j,i}$ as well. Hence, one can sample $\theta_{j,1}$ from

$$w_0 P^*_{j,1}(\mathrm{d}\theta) + \sum_{l \in \mathcal{J}_{-j,\zeta_{j,1}}} w_l \delta_{\tilde{\theta}_{l,\zeta_{j,1}}}(\mathrm{d}\theta) \quad (29)$$

where $\tilde{\theta}_{l,\zeta_{j,1}}$ are the distinct θ values in the urn labeled $\zeta_{j,1}$ and $\mathcal{J}_{-j,\zeta_{j,1}}$ is the set of indices of distinct values from the urn labeled $\zeta_{j,1}$ after excluding $\theta_{j,1}$. Moreover

$$\begin{aligned} w_0 &\propto c\,(1-z)^{1-\zeta_{j,1}}\, z^{\zeta_{j,1}} \int_\Theta h\,(x_j;\theta)\, P_0(\mathrm{d}\theta) \\ w_l &\propto n^{(-j)}_{l,1}\, h(x_j; \tilde{\theta}_{l,\zeta_{j,1}}). \end{aligned} \quad (30)$$

In the weights above, $P^*_{j,1}(\mathrm{d}\theta) = h(x_j;\theta)P_0(\mathrm{d}\theta)/\int_\Theta h(x_j;\theta)P_0(\mathrm{d}\theta)$ and $n^{-j}_{l,1}$ is the size of the cluster containing $\tilde{\theta}_{l,\zeta_{j,1}}$, after deleting $\theta_{j,1}$. With obvious modifications one also obtains the full conditional for generating $\theta_{j,2}$. This last point suggests that, conditional on the labels, one needs to run three independent Blackwell–MacQueen Pólya urn schemes: two are related to the idiosyncratic (and independent) components and one is related to the common component. Given this, the only difficulty in implementing the algorithm is due to the generalized hypergeometric function ${}_3F_2(a, b, c; e, f; x)$. Indeed, when such a function is evaluated at $x = 1$, as in our case, the convergence of the series defining it

can be very slow, depending on the magnitude of $e+f-a-b-c>0$: the lower such a value, the slower the convergence of the series. The efficiency of the algorithm can, thus, be improved by suitably resorting to identities that involve generalized hypergeometric functions in order to obtain equivalent expressions with a larger value of $e+f-a-b-c$. In particular, in the examples considered here we have been able to considerably speed up the implementation of the algorithm by applying an identity that can be found in [2, p.14].

4.2. Dependent mixtures of normalized σ–stable processes

Consider a GM–dependent normalized σ–stable CRM vector $(\tilde{p}_1, \tilde{p}_2)$ as in Example 2. The corresponding model is somehow more complicated to deal with, but at the same time it is more representative of what happens in the general case given the simplifications typical of the Dirichlet process do not occur. Specifically, the weights (w_1, w_2) are no longer independent from the normalized σ–stable processes p_i in (18). Moreover, the density of (w_1, w_2) is not available in closed form for any $\sigma \in (0,1)$, but only for $\sigma = 1/2$. Nonetheless, it is still possible to obtain analytic forms for the full conditionals allowing to estimate the marginal densities $\tilde{f}_i$ and to analyze the clustering structure featured by the two–sample data. Indeed, one can show the following

COROLLARY 3. *Let $(\tilde{p}_1, \tilde{p}_2)$ be a GM–dependent normalized σ–stable CRM. A density of the probability distribution $\mathcal{L}^*$ defined in (24) coincides with*

$$\frac{z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{k_1+k_2-\tilde{k}_1-\tilde{k}_2}}{\Gamma(n_1)\,\Gamma(n_2)}\,\sigma^{k-1}\,\Gamma(k)\,f(\boldsymbol{x},\boldsymbol{y}\,|\,\boldsymbol{\theta}^*,\boldsymbol{\pi}_{n_1,n_2})\prod_{i=1}^{k} g_0(\theta_i^*)$$
$$\times\ \xi_\sigma(\boldsymbol{n}^{(1)},\boldsymbol{n}^{(2)},\boldsymbol{q}^*)\int_0^1 \frac{w^{n-\bar{n}_1+\tilde{k}_1\sigma-1}(1-w)^{n_2-\bar{n}_2+\tilde{k}_2\sigma-1}}{\{1-z+zw^\sigma+z(1-w)^\sigma\}^k}\,\mathrm{d}w$$

where $\bar{n}_1 = |\boldsymbol{\zeta}^{(1)}|$ and $\bar{n}_2 = |\boldsymbol{\zeta}^{(2)}|/2$.

In a similar fashion to the dependent Dirichlet process case, from Corollary 3 one can deduce the full conditionals for both the labels $\boldsymbol{\zeta}_*^{(i)}$ and the $\boldsymbol{\theta}_*^{(i)}$. As for the former, if $\zeta_{j,1}^*$ corresponds to a distinct value $\theta_{j,i}^*$ not coinciding with any value $\theta_{l,2}$ from the second sample, then

$$\pi_{j,1}(x) \propto \mathbb{1}_{\{0,1\}}(x)\,z^x(1-z)^{1-x}$$
$$\times\int_0^1 \frac{w^{n-\bar{n}_{-j,1}-xn_{j,1}+(\tilde{k}_{-j,1}+x)\sigma-1}(1-w)^{n_2-\bar{n}_2+\tilde{k}_2\sigma-1}}{\{1-z+zw^\sigma+z(1-w)^\sigma\}^k}\,\mathrm{d}w \quad (31)$$

where $\bar{n}_{-j,1} = \sum_{i\neq j} n_{i,1}\,\zeta_{i,1}^*$ and $\tilde{k}_{-j,1} = |\boldsymbol{\zeta}_{-j,*}^{(1)}|$.

Interestingly, the full conditionals for the latent random variables are as simple as in the Dirichlet process case. Since we are again conditioning on the labels $\boldsymbol{\zeta}^{(1)}$, it is apparent that one just needs to run three independent Blackwell-MacQueen Pólya urn schemes. For $\theta_{j,1}$ the full conditional coincides with (29) with different weights

$$\begin{aligned} w_0 &\propto k_{-j,\zeta_{j,1}} \sigma\,(1-z)^{1-\zeta_{j,1}}\, z^{\zeta_{j,1}} \int_\Theta h(x_j;\theta)\, P_0(\mathrm{d}\theta), \\ w_l &\propto (n_{l,1}^{(-j)} - \sigma)\, h(x_j;\tilde{\theta}_{l,\zeta_{j,1}}), \end{aligned} \tag{32}$$

where $k_{-j,\zeta_{j,1}}$ above is the number of clusters associated to $p_{\zeta_{j,1}}$ after excluding $\theta_{j,1}$.

5. Full conditional distributions

The results in Sections 3 and 4 form the basis for the concrete implementation of the model (17) to both synthetic and real datasets in the following section. Here we provide a detailed description of the algorithm set forth in Section 4 for specific choices of the kernel $h(\,\cdot\,;\,\cdot\,)$ and of the random probability measures $\tilde{p}_1$ and $\tilde{p}_2$. In particular, we make the standard assumption of $h(\,\cdot\,; M, V)$ being Gaussian with mean M and variance V and consider GM–dependent Dirichlet and normalized σ–stable processes as mixing measures. As for the specification of the base measures P_0 of such mixing measures (see (5)), we propose a natural extension to the partially exchangeable case of the quite standard specification of Escobar and West [9], which greatly contributed to popularizing the mixture of Dirichlet process model. In particular, we take P_0 to be a normal/inverse–Gamma distribution

$$P_0(\mathrm{d}M, \mathrm{d}V) = P_{0,1}(\mathrm{d}V)\, P_{0,2}(\mathrm{d}M \,|\, V)$$

with $P_{0,1}$ being an inverse–Gamma probability distribution with parameters (s, S) and $P_{0,2}$ is Gaussian with mean m and variance τV. Moreover, the corresponding hyperpriors are of the form

$$\begin{aligned} \tau^{-1} &\sim \mathrm{Ga}(w/2, W/2) \\ m &\sim \mathrm{N}(a, A) \\ z &\sim U(0,1) \\ c &\sim \mathrm{Ga}(a_0, b_0) \end{aligned} \tag{33}$$

for some $w > 0$, $W > 0$, $A > 0$, $a_0 > 0$, $b_0 > 0$ and real a. In the following we focus on the two special cases and provide the analytic expressions for the corresponding full conditional distributions. In terms of the notation set in Section 4 the latent variables now become $\theta_{j,i} = (M_{j,i}, V_{j,i}) \in \mathbb{R} \times \mathbb{R}^+$, for any $j = 1, \ldots, n_i$ and $i = 1, 2$. Moreover $\tilde{\theta}_{j,i} = (\tilde{M}_{j,i}, \tilde{V}_{j,i})$, for $i = 0, 1, 2$, represent the j–th distinct value of the latent variables with label i. Also recall that the number of distinct values with label i, for $i = 1, 2$, is equal to $\tilde{k}_i$ and set $\tilde{k}_0 = k_1 + k_2 - \tilde{k}_1 - \tilde{k}_2$.

5.1. GM–dependent Dirichlet processes

Let us first deal with the hierarchical mixture model (17) with $(\tilde{p}_1, \tilde{p}_2)$ a vector of GM–dependent Dirichlet processes with parameters $(c, z; P_0)$, which we will denote by GM–$\mathscr{D}(c, z; P_0)$ in the sequel. With this specification and the auxiliary variable representation of the mixing measure laid out in (20), the weights of the predictive (29) are similar to those described in [9], the only differences being related to the bivariate structure, which results in the dependence on z (see (5)) and on the label $\zeta_{j,i}$. These identify the full conditional for the latent $\theta_{j,i}$.

In order to determine the full conditionals for the other parameters to be sampled, let $\boldsymbol{D}_{-r}$ stand for the set of all (hyper)parameters of the model but r. As for the full conditional for z, one has

$$\kappa_z(z \,|\, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-z}) \propto \kappa_z(z)\, z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{\tilde{k}_0} \times {}_3F_2\left(\alpha - cz + n_1 - \bar{n}_1, n_1, n_2; \alpha + n_1, \beta + n_2; 1\right),$$

where κ_z is the prior distribution of z, which in our specification coincides with the uniform on $(0,1)$. On the other hand, an expression for the full conditional for c is obtained as follows

$$\kappa_c(c \,|\, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-c}) \propto \kappa_c(c)\, \frac{c^k}{(\alpha)_{n_1}(\beta)_{n_2}} \times {}_3F_2\left(\alpha - cz + n_1 - \bar{n}_1, n_1, n_2; \alpha + n_1, \beta + n_2; 1\right),$$

where κ_c is the prior distribution of c that is supposed coincide with $\mathrm{Ga}(a_0, b_0)$. Moreover, note that both the coefficients α and β appearing in the generalized hypergeometric function ${}_3F_2$ above depend on c. See Corollary 2. Finally, τ and m are sampled from the following distributions

$$\tau \,|\, (\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-\tau}) \sim \mathrm{IG}\left(\frac{w+k}{2}, \frac{W+W'}{2}\right), \tag{34}$$

$$m \,|\, (\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-m}) \sim \mathrm{N}(RT, T), \tag{35}$$

where $\mathrm{IG}(a, b)$ denotes the inverse–gamma distribution with density function $g(s) \propto s^{-a-1}\, \mathrm{e}^{-\beta/s}\, \mathbb{1}_{\mathbb{R}^+}(s)$, $W' = \sum_{i=0}^{2} \sum_{l=1}^{\tilde{k}_i} (\tilde{M}_{l,i} - m)^2 / \tilde{V}_{l,i}$ and

$$T = \left[\frac{1}{A} + \frac{1}{\tau}\left(\sum_{i=1}^{\tilde{k}_1} \frac{1}{\tilde{V}_{i,1}} + \sum_{j=1}^{\tilde{k}_2} \frac{1}{\tilde{V}_{j,2}} + \sum_{r=1}^{\tilde{k}_0} \frac{1}{\tilde{V}_{r,0}}\right)\right]^{-1},$$

$$R = \left[\frac{a}{A} + \frac{1}{\tau}\left(\sum_{i=1}^{\tilde{k}_1} \frac{\tilde{M}_{i,1}}{\tilde{V}_{i,1}} + \sum_{j=1}^{\tilde{k}_2} \frac{\tilde{M}_{j,2}}{\tilde{V}_{j,2}} + \sum_{r=1}^{\tilde{k}_0} \frac{\tilde{M}_{r,0}}{\tilde{V}_{r,0}}\right)\right]^{-1}.$$

5.2. GM–dependent σ–stable normalized random measures

When $(\tilde{p}_1, \tilde{p}_2)$ is a vector of GM–dependent normalized σ–stable processes with parameters (z, P_0) we set the short notation GM–st(σ, z, P_0). The full conditionals are then derived from Corollary 3. In particular, explicit expressions for the weights in (32) can be deduced and the full conditional for z which coincides with

$$\kappa_z(z \,|\, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-z}) \propto \kappa_z(z)\, z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{\tilde{k}_0} \times \int_0^1 \frac{w^{n_1-\bar{n}_1+\tilde{k}_1\sigma-1}(1-w)^{n_2-\bar{n}_2+\tilde{k}_2\sigma-1}}{\{1-z+zw^\sigma+z(1-w)^\sigma\}^k}\mathrm{d}w,$$

where κ_z is, as in Section 5.1, uniform on $(0,1)$. Moreover, if a prior on $(0,1)$ is assigned to the parameter σ, the corresponding full conditional is given by

$$\kappa_\sigma(\sigma \,|\, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-\sigma}) \propto \kappa_\sigma(\sigma)\, \sigma^{k-1}\, \xi_\sigma(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*) \times \int_0^1 \frac{w^{n_1-\bar{n}_1+\tilde{k}_1\sigma-1}(1-w)^{n_2-\bar{n}_2+\tilde{k}_2\sigma-1}}{\{1-z+zw^\sigma+z(1-w)^\sigma\}^k}\mathrm{d}w$$

Finally, the full conditionals for τ and m coincide with those displayed in (34) and (35) since they depend only on h and P_0 and not on the specific vector of random probabilities $(\tilde{p}_1, \tilde{p}_2)$ driving the respective dependent mixtures.

5.3. Accelerated algorithm

It is well–known that univariate Pólya urn samplers like the one proposed in [9] tend to mix slowly when the probability of sampling a new value, w_0, is much smaller than the probability to sample an already observed one. When this occurs, the sampler can get stuck at the current set of distinct values and it may take many iterations before any new value is generated. Such a concern clearly extends also to our bivariate Pólya urn sampler and, in particular, to (29) and (32) leading the algorithm to get stuck in some specific $\{\tilde{\theta}_{i,l} : \ l = 0, 1, 2; \ i = 1, \ldots, \tilde{k}_l\}$. To circumvent this problem we resort to the the method suggested in [40] and [26]: it consists in resampling, at the end of every iteration, the distinct values $\tilde{\theta}_{l,i}$ from their conditional distribution. Since this distribution depends on the choice of $\tilde{p}_1$ and $\tilde{p}_2$ only through their base measure P_0, it is the same for the Dirichlet and σ–stable cases. In particular, for every $i = 1, \ldots, \tilde{k}_1$, the required full conditional density of $\tilde{\theta}_{i,1}$ is

$$\mathcal{L}^*(\tilde{\theta}_{i,1} \,|\, \boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)}, \boldsymbol{D}_{-\tilde{\theta}_{i,1}}) \propto g_0(\tilde{\theta}_{i,1}) \prod_{j \in C_{i,1}} h(x_j, \tilde{\theta}_{i,1}), \tag{36}$$

where $\mathcal{L}^*$ is the joint law defined in (24). With our specification, the full conditional distribution of $\tilde{\theta}_{i,1}$ in (36) becomes normal/inverse–Gamma with

$$\tilde{V}_{i,1}^{-1} \sim \mathrm{Ga}\left(s + \frac{n_{i,1}}{2}, S + \frac{\sum x_j^2}{2} + \frac{m^2 n_{i,1} - \sum x_j \,(2m + \tau \sum x_j)}{2(1+\tau n_{i,1})}\right),$$

$$\tilde{M}_{i,1} \,|\, \tilde{V}_{i,1} \sim \mathrm{N}\left(\frac{m + \tau \sum x_j}{1+\tau n_{i,1}}, \tilde{V}_{i,1}\, \frac{\tau}{1+\tau n_{i,1}}\right)$$

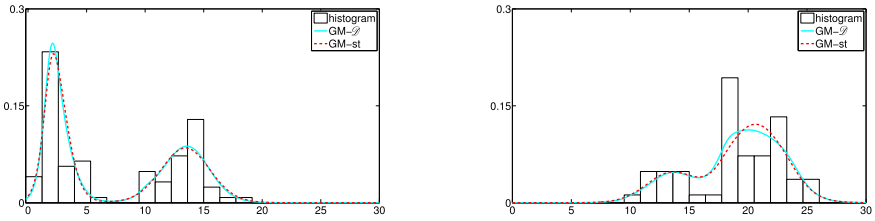
where $\sum x_j$ is a shortened notation for $\sum_{j \in C_{i,1}} x_j$. Analogous expressions, with obvious modifications, hold true for $\tilde{\theta}_{i,2}$ and $\tilde{\theta}_{i,0}$.

6. Illustration

In this section we illustrate the inferential performance of the proposed model on a two–sample dataset and to this end we implement the Gibbs sampling algorithm devised in the previous section for (17). We shall consider $(\tilde{p}_1, \tilde{p}_2)$ being either a GM–$\mathscr{D}(c, z, P_0)$ or a GM–st(σ, z, P_0). In terms of computational efficiency, we note in advance that the algorithm with the GM–st mixture is remarkably faster than the one associated to the GM–$\mathscr{D}$ mixture. As already pointed out in the previous sections this is due to the need of repeated evaluations of generalized hypergeometric function ${}_3F_2$ in the GM–$\mathscr{D}$ case. In contrast, the numerical evaluation of the one–dimensional integral in Corollary 3, for the GM-st mixture, is straightforward.

We shall analyze the well–known Iris dataset, which contains measures of 4 features of 3 different species of *Iris* flowers: *Setosa*, *Versicolor* and *Virginica*. For each of these species 150 records of sepal length, sepal width, petal length and petal width of flowers are available. These data are commonly used in the literature as an illustrative example for discriminant analysis. Indeed, it has been noted that *Setosa* is very well separated from the other two species, which partially overlap. Of the 4 measured features, here we consider the petal width expressed in millimeters. A total number of 50 observations per species have been recorded. The 150 observations are, then, used to form two samples $\boldsymbol{X}^{(n_1)}$ and $\boldsymbol{Y}^{(n_2)}$ as follows. We set $n_1 = 90$ and let the first sample consist of 50 observations of *Setosa* and 40 of *Versicolor*. Correspondingly $n_2 = 60$ and includes 50 observations of *Virginica* and the remaining 10 observations of *Versicolor*. The particular design of the experiment is motivated by the idea that the *Versicolor* species identifies the shared component between the two mixtures, thus making our approach for modeling dependence appropriate. Moreover, on the basis of previous considerations it is expected that the two species in the first dataset are more clearly separated than the two species forming the second sample.

Our statistical analysis has the following two goals: on the one hand we wish to estimate the densities generating the two samples and, on the other, we aim at obtaining an approximation of the posterior distribution of the number of clusters in each sample.



(a) f_1 (b) f_2

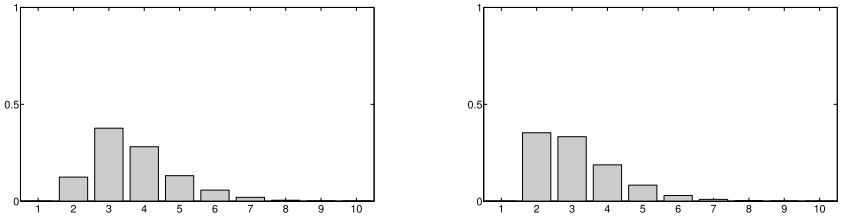
Figure 1. GM–$\mathscr{D}(c, z, P_0)$ (solid line) vs. GM–st(σ, z, P_0) (dashed line) mixture with random c and σ, respectively: posterior estimates of the densities f_1 and f_2. The displayed histograms are generated by the actual two–sample data.

This allows to draw a direct comparison of the inferential outcomes produced by the GM–$\mathscr{D}(c, z, P_0)$ and GM–st(σ, z, P_0) mixtures. As for the specifications of the hyperparameters in (33) we essentially adopted the quite standard specifications of [9]. Hence, we have set $(w, W) = (1, 100)$, $(s, S) = (1, 1)$, $(a, A) = ((n_1\bar{X} + n_2\bar{Y})/(n_1 + n_2), 2)$ and $(a_0, b_0) = (2, 1)$ where $\bar{X}$ and $\bar{Y}$ are the sample means for $\boldsymbol{X}^{(n_1)}$ and $\boldsymbol{Y}^{(n_2)}$, respectively. As for the other parameters involved, we suppose that $c \sim$Ga$(2, 1)$, whereas σ and z are both uniform on $[0, 1]$. Moreover, these three parameters are independent. All estimates will be based on 80000 iterations of the algorithm after 20000 burn–in sweeps.

Table 1. GM–$\mathscr{D}(c, z, P_0)$ vs. GM–st(σ, z, P_0) mixture with random c and σ, respectively: estimated number of clusters (Col. 1 and 2), maximum a posteriori values ($\hat{K}_{\boldsymbol{X}}$ and $\hat{K}_{\boldsymbol{Y}}$) and probability of more than 4 clusters per sample (Col. 5 and 6).

	$\mathbb{E}[K_{\boldsymbol{X}}\|\ldots]$	$\mathbb{E}[K_{\boldsymbol{Y}}\|\ldots]$	$\hat{K}_{\boldsymbol{X}}$	$\hat{K}_{\boldsymbol{Y}}$	$\mathbb{P}(K_{\boldsymbol{X}} \geq 4)$	$\mathbb{P}(K_{\boldsymbol{Y}} \geq 4)$
GM-$\mathscr{D}(c, z, P_0)$	3.72	3.15	3	2	0.50	0.31
GM–st(σ, z, P_0)	2.70	2.30	2	2	0.13	0.05

The estimated densities are displayed in Figure 1 and there seem to be no significant differences. However, regardless the particular mixture model specification, the two species forming each sample are clearly better separated in the first sample. This is not surprising, given that the second sample is formed by two overlapping species. See also the histogram in the background of Figure 1. The results on the clustering structure are reported in Figure 2 and in Table 1. Figure 2 shows that the posterior distributions of the number of clusters corresponding to the GM–st mixture is characterized by a lower variability than in the GM–$\mathscr{D}$ mixture case. Moreover, if one roughly thinks of each species of flowers in a sample as forming a single cluster, then it is apparent that the GM–st mixture better estimates both $K_{\boldsymbol{X}}$ and $K_{\boldsymbol{Y}}$. See also Table 1. These results seems to suggest that the parameter σ, associated to the stable CRM, has a benefical impact on the estimation of the clustering structure. This is in line with the findings of [25] in



(a) f_1, GM–$\mathscr{D}$ (b) f_2, GM–$\mathscr{D}$

(c) f_1, GM–st (d) f_2, GM–st

Figure 2. GM–$\mathscr{D}(c, z, P_0)$ (top row) vs. GM–st(σ, z, P_0) (bottom row) mixture with random c and σ, respectively: posterior distributions of the number of clusters $K_{\boldsymbol{X}}$ and $K_{\boldsymbol{Y}}$.

the exchangeable case, where it is pointed out that σ induces a reinforcement mechanism which improves the capability of learning the clustering structure from the data. We believe this aspect is of great relevance and, hence, deserves further investigation.

Appendix

A.1. Proof of Proposition 1

By combining the definition of GM–dependent normalized CRMs given in (8) with the gamma integral, it is possible to write

$$\mathbb{E}\left[\tilde{p}_1(A)\,\tilde{p}_2(B)\right] = \int_0^\infty \int_0^\infty \mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(\mathbb{X})-v\tilde{\mu}_2(\mathbb{X})}\,\tilde{\mu}_1(A)\,\tilde{\mu}_2(B)\right]\,\mathrm{d}u\mathrm{d}v$$

Since $\tilde{\mu}_i = \mu_i + \mu_0$ for $i = 1, 2$, with μ_0, μ_1 and μ_2 independent, one has

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{e}^{-u\mu_1(\mathbb{X})-(u+v)\mu_0(\mathbb{X})}\,\mu_1(A)\,\mu_0(B)\right] \\
&= \mathbb{E}\left[\mathrm{e}^{-u\mu_1(\mathbb{X})}\,\mu_1(A)\right]\,\mathbb{E}\left[\mathrm{e}^{-(u+v)\mu_0(\mathbb{X})}\,\mu_0(B)\right] \\
&= c^2 z(1-z)P_0(A)\,P_0(B)\,\mathrm{e}^{-cz\psi(u)-c(1-z)\psi(u+v)}\,\tau_1(u)\,\tau_1(u+v)
\end{aligned}$$

Use the symbol A^i to denote A if $i = 1$ and A^c if $i = 0$. Hence $\{A^i \cap B^j : i, j = 0, 1\}$ is the partition of $\mathbb{X}$ generated by $\{A, B\}$. Hence

$$\mathbb{E}\Big[\mathrm{e}^{-(u+v)\mu_0(\mathbb{X})}\,\mu_0(A)\,\mu_0(B)\Big] = \sum_{i,j=0}^{1} \mathbb{E}\left[\mathrm{e}^{-(u+v)\mu_0(\mathbb{X})}\,\mu_0(A\cap B^i)\,\mu_0(A^j\cap B)\right]$$

This implies that

$$\begin{aligned}
\mathbb{E}\left[\mathrm{e}^{-(u+v)\mu_0(\mathbb{X})}\,\mu_0(A)\mu_0(B)\right] &= \mathrm{e}^{-c(1-z)\psi(u+v)}\,c(1-z)\,\{P_0(A\cap B)\,\tau_2(u+v) \\
&\quad + c(1-z)\tau_1^2(u+v)\sum_{i,j=0}^{1} P_0(A\cap B^i)P_0(A^j\cap B)\} \\
&= \mathrm{e}^{-c(1-z)\psi(u+v)}\,c(1-z)\,\{P_0(A\cap B)\,\tau_2(u+v) \\
&\quad + c(1-z)P_0(A)P_0(B)\tau_1^2(u+v)\}.
\end{aligned}$$

Summing up, it follows that

$$\mathbb{E}\left[\tilde{p}_1(A)\,\tilde{p}_2(B)\right] = \int_0^\infty \int_0^\infty \mathrm{e}^{-z(\psi(u)+\psi(v))-c(1-z)\psi(u+v)}$$

$$\times\ c\,\{(1-z)P_0(A\cap B)\,\tau_2(u+v)+c^2P_0(A)P_0(B)[(1-z)^2\tau_1^2(u+v)$$
$$+z(1-z)\tau_1(u+v)(\tau_1(u)+\tau_1(v))+z^2\tau_1(u)\tau_1(v)]\}\,\mathrm{d}u\mathrm{d}v$$

If in the previous expression one sets $A=B=\mathbb{X}$, then the following identity holds true

$$c^2\int_0^\infty\int_0^\infty \mathrm{e}^{-cz(\psi(u)+\psi(v))-c(1-z)\psi(u+v))}\,[(1-z)^2\,\tau_1^2(u+v)$$
$$+z(1-z)\tau_1(u+v)(\tau_1(u)+\tau_1(v))+z^2\tau_1(u)\tau_1(v)]\,\mathrm{d}u\,\mathrm{d}v$$
$$=1-c(1-z)\int_0^\infty\int_0^\infty \mathrm{e}^{-c(1-z)\psi(u+v)}\,\tau_2(u+v)\,\mathrm{d}u\mathrm{d}v$$

The results in (11) and in (12) then follows. □

A.2. Proof of Proposition 2

We first determine the probability distribution of $(\tilde\pi_{n_1,n_2},\boldsymbol{X}^{(n_1)},\boldsymbol{Y}^{(n_2)})$. Here $\tilde\pi_{n_1,n_2}$ denotes a random partition of $\{\boldsymbol{X}^{(n_1)},\boldsymbol{Y}^{(n_2)}\}$ whose generic realization, π_{n_1,n_2}, splits the n_1+n_2 observations into $\sum_{i=0}^2 k_i$ groups of distinct values with respective frequencies $\{n_{j,1}\}_{j=1}^{k_1}$, $\{n_{\ell,1}\}_{\ell=1}^{k_2}$ and $\{q_{r,1}+q_{r,2}\}_{r=1}^{k_0}$. Henceforth, we shall use the shorter notation

$$\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A})=(\tilde\pi_{n_1,n_2},\boldsymbol{X}^{(n_1)},\boldsymbol{Y}^{(n_2)})^{-1}(\pi_{n_1,n_2},\mathcal{A})$$

with $\mathcal{A}$ standing for the collection of pairwise disjoint sets $\{A_{j,1},A_{\ell,2},A_r : j=1,\ldots,k_1;\ \ell=1,\ldots,k_2;\ r=1,\ldots,k_0\}$. Moreover, for any pair of set function m_1 and m_2 on $(\mathbb{X},\mathscr{X})$ we set $\boldsymbol{m}_i^{\boldsymbol{n}^{(i)}}(\mathcal{A}_i)=\prod_{j=1}^{k_i} m_i^{n_{j,i}}(A_{j,i})$ and $(\boldsymbol{m}_1^{\boldsymbol{q}^{(1)}}\times\boldsymbol{m}_2^{\boldsymbol{q}^{(2)}})(\mathcal{A}_0)=\prod_{r=1}^{k_0} m_1^{q_{r,1}}(A_r)m_2^{q_{r,2}}(A_r)$. By virtue of (9) one has

$$\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A})\right]=\int_{P_{\mathbb{X}}^2}\boldsymbol{p}_1^{\boldsymbol{n}^{(1)}}(\mathcal{A}_1)\,\boldsymbol{p}_2^{\boldsymbol{n}^{(2)}}(\mathcal{A}_2)\,(\boldsymbol{p}_1^{\boldsymbol{q}^{(1)}}\times\boldsymbol{p}_2^{\boldsymbol{q}^{(2)}})(\mathcal{A}_0)\,\Phi(\mathrm{d}p_1,\mathrm{d}p_2)\qquad(37)$$

Since each $\tilde p_i$ is equal, in distribution, to the normalized measure $\tilde\mu_i/\tilde\mu_i(\mathbb{X})$ one can proceed in a similar fashion as in the proof of Proposition 1 and write

$$\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A})\right]=\frac{1}{\Gamma(n_1)\Gamma(n_2)}\int_0^\infty\mathrm{d}u\int_0^\infty\mathrm{d}v$$
$$\mathbb{E}\left[\mathrm{e}^{-u\tilde\mu_1(\mathbb{X})-v\tilde\mu_2(\mathbb{X})}\,\tilde{\boldsymbol{\mu}}_1^{\boldsymbol{n}^{(1)}}(\mathcal{A}_1)\,\tilde{\boldsymbol{\mu}}_2^{\boldsymbol{n}^{(2)}}(\mathcal{A}_2)\,(\tilde{\boldsymbol{\mu}}_1^{\boldsymbol{q}^{(1)}}\times\tilde{\boldsymbol{\mu}}_2^{\boldsymbol{q}^{(2)}})(\mathcal{A}_0)\right].$$

Since CRMs give rise to mutually independent random variables when evaluated on disjoint sets, which identifies the so–called independence property of CRMs, the expected value in the integral above is shown to coincide with

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(\mathbb{X}^*)-v\tilde{\mu}_2(\mathbb{X}^*)}\right]\prod_{i=1}^{2}\prod_{j=1}^{k_i}\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(A_{j,i})-v\tilde{\mu}_2(A_{j,i})}\tilde{\mu}_i^{n_{j,i}}(A_{j,i})\right]$$

$$\times\prod_{r=1}^{k_0}\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(A_r)-v\tilde{\mu}_2(A_r)}\tilde{\mu}_1^{q_{r,1}}(A_r)\,\tilde{\mu}_2^{q_{r,2}}(A_r)\right]$$

where $\mathbb{X}^*=\mathbb{X}\setminus\{(\cup_{i=1}^{2}\cup_{j=1}^{k_i}A_{j,i})\cup(\cup_{r=1}^{k_0}A_r)\}$. In the first product let us consider $i=1$. A similar line of reasoning holds for $i=2$ as well. If we set $h_z(u,v)=z(\psi(u)+\psi(v))+(1-z)\psi(u+v)$, by virtue of the Faà di Bruno formula the j–th factor coincides with

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(A_{j,1})-v\tilde{\mu}_2(A_{j,1})}\,\tilde{\mu}_1^{n_{j,1}}(A_{j,1})\right]=(-1)^{n_{j,1}}\,\frac{\partial^{n_{j,1}}}{\partial u^{n_{j,1}}}\mathrm{e}^{-G(A_{j,1})\,h_z(u,v)}$$

$$=\mathrm{e}^{-G(A_{j,1})\,h_z(u,v)}\,\{G(A_{j,1})(z\tau_{n_{j,1}}(u)+(1-z)\tau_{n_{j,1}}(v))+R_j(A_{j,1})\}$$

where $R_j(A_{j,1})$ is a polynomial in $G(A_{j,1})$ of order greater than 1 and $G=cP_0$. Moreover, a multivariate version of the Faà di Bruno formula, see [5], leads to

$$\mathbb{E}\left[\mathrm{e}^{-u\tilde{\mu}_1(A_r)-v\tilde{\mu}_2(A_r)}\,\tilde{\mu}_1^{q_{r,1}}(A_r)\,\tilde{\mu}_2^{q_{r,2}}(A_r)\right]$$

$$=(-1)^{q_{r,1}+q_{r,2}}\,\frac{\partial^{q_{r,1}+q_{r,2}}}{\partial u^{q_{r,1}}\partial v^{q_{r,2}}}\,\mathrm{e}^{-G(A_r)h_z(u,v)}$$

$$=\mathrm{e}^{-G(A_r)h_z(u,v)}\,\{G(A_r)(1-z)\tau_{q_{r,1}+q_{r,2}}(u+v)+R_r^*(A_r)\}$$

with $R_r^*(A_r)$ denoting a polynomial in $G(A_r)$ of degree greater than 1. Combining all these facts together, one obtains

$$\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A})\right]=\frac{\prod_{j,\ell,r}G(A_{j,1})G(A_{\ell,2})G(A_r)}{\Gamma(n_1)\Gamma(n_2)}$$

$$\times\sum_{\boldsymbol{i}\in\{0,1\}^{k_1}}\sum_{\boldsymbol{l}\in\{0,1\}^{k_2}}(1-z)^{k_0+|\boldsymbol{i}|+|\boldsymbol{l}|}\,z^{k_1+k_2-|\boldsymbol{i}|-|\boldsymbol{l}|}\int_0^\infty\int_0^\infty\prod_{j=1}^{k_1}\tau_{n_{j,1}}(u+i_jv)$$

$$\times\prod_{l=1}^{k_2}\tau_{n_{l,2}}(\ell_l u+v)\prod_{r=1}^{k_0}\tau_{q_{r,1}+q_{r,2}}(u+v)\,\mathrm{d}\,\mathrm{d}v+R_r^{**}(\mathcal{A})$$

where $R_r^{**}(\mathcal{A})$ is a polynomial of order greater than $k=k_1+k_2+k_0$ in the variables $G(A_{j,1})$, with $j=1,\ldots,k_1$, $G(A_{\ell,2})$, with $\ell=1,\ldots,k_2$, and $G(A_r)$, with $r=1,\ldots,k_0$. It is apparent that the probability distribution of $(\boldsymbol{X}^{(n_1)},\boldsymbol{Y}^{(n_2)})$, conditional on $\tilde{\pi}_{n_1,n_2}=\pi_{n_1,n_2}$, is absolutely continuous with respect to P_0^k and recall that P_0 is non–atomic. In order to determine a density of $(\boldsymbol{X}^{(n_1)},\boldsymbol{Y}^{(n_2)})$, conditional on $\tilde{\pi}_{n_1,n_2}=\pi_{n_1,n_2}$, define $\mathcal{A}_\varepsilon$ as the collection of sets $\{A_{j,1}^\varepsilon,A_{\ell,2}^\varepsilon,A_r^\varepsilon:\,j=1,\ldots,k_1;\,\ell=1,\ldots,k_2;\,r=1,\ldots,k_0\}$ with

$$A_{j,1}^\varepsilon\downarrow\{x_j\}\qquad A_{\ell,2}^\varepsilon\downarrow\{y_\ell\}\qquad A_r^\varepsilon\downarrow\{z_r\}$$

as $\varepsilon \downarrow 0$. Hence, a version of the conditional density of $(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)})$, conditional on $\tilde{\pi}_{n_1,n_2} = \pi_{n_1,n_2}$, with respect to P_0^k and evaluated at $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is proportional to

$$\lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A}^\varepsilon)\right]}{\prod_{j=1}^{k_1} P_0(A_{j,1}^\varepsilon) \prod_{\ell=1}^{k_2} P_0(A_{\ell,2}^\varepsilon) \prod_{r=1}^{k_0} P_0(A_r^\varepsilon)}$$

and, from previous expansion, it can be easily seen to coincide with 1. And this proves the statement. □

A.3. Proof of Proposition 4

The probability distribution $\mathcal{L}^*$ defined in (24) can be decomposed as follows

$$\mathcal{L}(\boldsymbol{\theta}^*, \pi_{n_1,n_2}, \boldsymbol{\zeta})\, \mathcal{L}(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)} \,|\, \boldsymbol{\theta}^*, \pi_{n_1,n_2}, \boldsymbol{\zeta}).$$

In a similar fashion to the proof of Proposition 2 we use the notation

$$\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A}) = (\tilde{\pi}_{n_1,n_2}, \boldsymbol{\theta}^*)^{-1}(\pi_{n_1,n_2}, \mathcal{A})$$

with $\mathcal{A}$ standing for the collection of pairwise disjoint sets $\{A_{j,1}, A_{\ell,2}, A_r : j = 1, \ldots, k_1;\ \ell = 1, \ldots, k_2;\ r = 1, \ldots, k_0\}$. By virtue of (9) and by definition of $\boldsymbol{\zeta} = (\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)})$ one has

$$\begin{aligned}\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A}) \,|\, \boldsymbol{\zeta}\right] = \int_{P_{\mathbb{X}}^3} & \boldsymbol{p}_1^{\boldsymbol{n}^{(1)}\cdot\boldsymbol{\zeta}_1^*}(\mathcal{A}_1) \boldsymbol{p}_2^{\boldsymbol{n}^{(2)}\cdot\boldsymbol{\zeta}_2^*}(\mathcal{A}_1) \\ & \times \boldsymbol{p}_0^{\boldsymbol{n}^{(1)}\cdot(\mathbf{1}-\boldsymbol{\zeta}_1^*)}(\mathcal{A}_1) \boldsymbol{p}_0^{\boldsymbol{n}^{(2)}\cdot(\mathbf{1}-\boldsymbol{\zeta}_2^*)}(\mathcal{A}_2)\, \boldsymbol{p}_0^{\boldsymbol{q}^*}(\mathcal{A}_0)\, \Phi'(\mathrm{d}p_0, \mathrm{d}p_1, \mathrm{d}p_2), \end{aligned} \tag{38}$$

where Φ' corresponds to the probability distribution of the random vector

$$\left(\frac{\mu_0}{\mu_0(\mathbb{X})}, \frac{\mu_1}{\mu_1(\mathbb{X})}, \frac{\mu_2}{\mu_2(\mathbb{X})}\right)$$

on $P_{\mathbb{X}}^3$ and we have used vector notation to denote the inner products $\boldsymbol{n}^{(i)} \cdot \boldsymbol{\zeta}_i^* = \sum_{j=1}^{k_i} n_{j,i}\zeta_{j,i}^*$ and $\boldsymbol{n}^{(i)} \cdot (\mathbf{1} - \boldsymbol{\zeta}_i^*) = \sum_{j=1}^{k_i} n_{j,i}(1 - \zeta_{j,i})$ for $i = 1, 2$. Moreover, note that

$$\begin{aligned}\mathbb{P}[\boldsymbol{\zeta} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}) \,|\, \mu_0, \mu_1, \mu_2] & \\ = \frac{\mu_1(\mathbb{X})^{|\boldsymbol{a}^{(1)}|}\, \mu_2(\mathbb{X})^{|\boldsymbol{a}^{(2)}|/2}\, \mu_0(\mathbb{X})^{n_1+n_2-|\boldsymbol{a}^{(1)}|-|\boldsymbol{a}^{(2)}|/2}}{(\mu_0(\mathbb{X}) + \mu_1(\mathbb{X}))^{n_1}\, (\mu_0(\mathbb{X}) + \mu_2(\mathbb{X}))^{n_2}} & \end{aligned}$$

for any $\boldsymbol{a} = (\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}) \in \{0,1\}^{n_1} \times \{0,2\}^{n_2}$. Thus, by similar arguments to those employed in the proofs of Propositions 1 and 2, we can write

$$\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A}),\, \boldsymbol{\zeta} = \boldsymbol{a}\right] = \frac{1}{\Gamma(n_1)\Gamma(n_2)} \int_0^\infty \mathrm{d}u \int_0^\infty \mathrm{d}v\, u^{n_1-1}\, v^{n_2-1}$$

$$\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(\mathbb{X})+\mu_1(\mathbb{X}))-v(\mu_0(\mathbb{X})+\mu_2(\mathbb{X}))}\,\boldsymbol{\mu}_1^{\boldsymbol{n}^{(1)}\cdot\boldsymbol{\zeta}_1^*}(\mathcal{A}_1)\boldsymbol{\mu}_0^{\boldsymbol{n}^{(1)}\cdot(\mathbf{1}-\boldsymbol{\zeta}_1^*)}(\mathcal{A}_1)\right.$$
$$\left.\times\,\boldsymbol{\mu}_2^{\boldsymbol{n}^{(2)}\cdot\boldsymbol{\zeta}_2^*}(\mathcal{A}_2)\,\boldsymbol{\mu}_0^{\boldsymbol{n}^{(2)}\cdot(\mathbf{1}-\boldsymbol{\zeta}_2^*)}(\mathcal{A}_2)\,(\boldsymbol{\mu}_0^{\boldsymbol{q}^*})(\mathcal{A}_0)\right]$$

where $\boldsymbol{a}$ is a vector such that $\boldsymbol{a}^{(i)}$ contains k_i labels $\zeta_{j,i}^*$ such that

$$\mathbb{P}[\theta_{j,i}^* \in A \,|\, p_0, p_1, p_2] = p_{\zeta_{j,i}^*}(A).$$

Using the independence property of CRMs and the independence of μ_0, μ_1 and μ_2, the expected value in the integral above can be rewritten as

$$\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(\mathbb{X}^*)+\mu_1(\mathbb{X}^*))-v(\mu_0(\mathbb{X}^*)+\mu_2(\mathbb{X}^*))}\right]$$
$$\times\prod_{i=1}^{2}\prod_{j=1}^{k_i}\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(A_{j,i})+\mu_1(A_{j,i}))-v(\mu_0(A_{j,i})+\mu_2(A_{j,i}))}\mu_i(A_{j,i})^{n_{j,i}\zeta_{j,i}^*}\,\mu_0(A_{j,i})^{n_{j,i}(1-\zeta_{j,i}^*)}\right]$$
$$\times\prod_{r=0}^{k_0}\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(A_r)+\mu_1(A_r))-v(\mu_0(A_r)+\mu_2(A_r))}\mu_0(A_r)^{q_{r,1}+q_{r,2}}\right]$$

where $\mathbb{X}^* = \mathbb{X}\setminus\{(\cup_{i=1}^{2}\cup_{j=1}^{k_i}A_{j,i})\cup(\cup_{r=1}^{k_0}A_r)\}$. In the first product consider $i=1$, a similar line of reasoning holds then for $i=2$. The j–th factor coincides with

$$\mathbb{E}\left[\mathrm{e}^{-v\mu_2(A_{j,1})}\right]\,\mathbb{E}\left[\mathrm{e}^{-u\mu_1(A_{j,1})}\mu_1(A_{j,1})^{n_{j,1}\zeta_{j,1}^*}\right]$$
$$\times\,\mathbb{E}\left[\mathrm{e}^{-(u+v)\mu_0(A_{j,1})}\mu_0(A_{j,1})^{n_{j,1}(1-\zeta_{j,1}^*)}\right], \quad (39)$$

where

$$\mathbb{E}\left[\mathrm{e}^{-v\mu_2(A_{j,1})}\right] = \mathrm{e}^{-c\,P_0(A_{j,1})\,\psi(v)}$$

and, by virtue of the Faà di Bruno formula,

$$\mathbb{E}\left[\mathrm{e}^{-u\mu_1(A_{j,1})}\mu_1(A_{j,1})^{n_{j,1}\zeta_{j,1}^*}\right] = (-1)^{n_{j,1}\zeta_{j,1}^*}\frac{\partial^{n_{j,1}\zeta_{j,1}^*}}{\partial u^{n_{j,1}\zeta_{j,1}^*}}\mathrm{e}^{-cz\,P_0(A_{j,1})\,\psi(u)}$$
$$= \mathrm{e}^{-cz\,P_0(A_{j,1})\,\psi(u)}\left\{cz\left[P_0(A_{j,1})\tau_{n_{j,1}}(u)+R_{j,1}(A_{j,1})\right]\right\}^{\zeta_{j,1}^*}$$

and

$$\mathbb{E}\left[\mathrm{e}^{-(u+v)\mu_0(A_{j,1})}\mu_0(A_{j,1})^{n_{j,1}(1-\zeta_{j,1}^*)}\right]$$
$$= (-1)^{n_{j,1}(1-\zeta_{j,1}^*)}\,\frac{\partial^{n_{j,1}(1-\zeta_{j,1}^*)}}{\partial s^{n_{j,1}(1-\zeta_{j,1}^*)}}\mathrm{e}^{-c(1-z)\,P_0(A_{j,1})\,\psi(s)}\Big|_{s=u+v}$$
$$= \mathrm{e}^{-c(1-z)\,P_0(A_{j,1})\,\psi(u+v)}\left\{c(1-z)\left[P_0(A_{j,1})\tau_{n_{j,1}}(u+v)+R_{j,1}(A_{j,1})\right]\right\}^{1-\zeta_{j,1}^*}.$$

In the previous expressions we have agreed that $\partial^0/\partial s^0$ is the identity operator and that $R_{j,1}(A_{j,1})$ is some polynomial in $P_0(A_{j,1})$ of order greater than 1. Thus, the product in (39) is equal to

$$\mathrm{e}^{-c\,P_0(A_{j,1})h_z(u,v)}c\,z^{\zeta^*_{j,1}}(1-z)^{1-\zeta^*_{j,1}} \times \left\{P_0(A_{j,1})\tau_{n_{j,1}}(u+(1-\zeta^*_{j,1})v)+R_{j,1}(A_{j,1})\right\}. \quad (40)$$

Analogously, one has

$$\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(\mathbb{X}^*)+\mu_1(\mathbb{X}^*))-v(\mu_0(\mathbb{X}^*)+\mu_2(\mathbb{X}^*))}\right]=\mathrm{e}^{-c\,P_0(\mathbb{X}^*)\,h_z(u,v)} \quad (41)$$

and

$$\mathbb{E}\left[\mathrm{e}^{-u(\mu_0(A_r)+\mu_1(A_r))-v(\mu_0(A_r)+\mu_2(A_r))}\mu_0(A_r)^{q_{r,1}+q_{r,2}}\right] = \mathrm{e}^{-c\,P_0(A_r)\,h_z(u,v)}\,c\,(1-z)\left\{P_0(A_r)\tau_{q_{r,1}+q_{r,2}}(u+v)+R_r(A_r)\right\}, \quad (42)$$

where $R_r(A_r)$ is some polynomial in $P_0(A_r)$ of order greater than 1. By combining the expressions (40)–(42) we obtain that $\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A})\,,\,\boldsymbol{\zeta}=\boldsymbol{a}\right]$ coincides with

$$\frac{c^k\,z^{|\boldsymbol{\zeta}^*_1|+|\boldsymbol{\zeta}^*_2|}(1-z)^{k_1+k_2-|\boldsymbol{\zeta}^*_1|-|\boldsymbol{\zeta}^*_2|}}{\Gamma(n_1)\Gamma(n_2)}P_0^k(\mathcal{A}) \int_0^\infty\int_0^\infty u^{n_1-1}\,v^{n_2-1}\mathrm{e}^{-ch_z(u,v)}\prod_{j=1}^{k_1}\tau_{n_{j,1}}(u+(1-\zeta^*_{j,1})v) \prod_{j=1}^{k_2}\tau_{n_{j,2}}((1-\zeta^*_{j,2})u+v))\prod_{r=1}^{k_0}\tau_{q_{r,1}+q_{r,2}}(u+v)\mathrm{d}u\mathrm{d}v+R^*(\mathcal{A}),$$

where $R^*(\mathcal{A})$ is a polynomial in the variables $P_0(A_{j,1})$, with $j=1,\ldots,k_1$, $P_0(A_{\ell,2})$, with $\ell=1,\ldots,k_2$, and $P_0(A_r)$, with $r=1,\ldots,k_0$, of order greater than $k=k_1+k_2+k_0$ and $P_0^k(\mathcal{A})=\prod_{i=1}^{k_1}\prod_{j=1}^{k_2}\prod_{r=1}^{k_0}P_0(A_{i,1})P_0(A_{j,2})P_0(A_r)$. It is apparent that the probability distribution of $(\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)})$, conditional on $\widetilde{\pi}_{n_1,n_2}=\pi_{n_1,n_2}$, is degenerate on Θ^k and the probability distribution of the distinct values $\boldsymbol{\theta}^*=(\boldsymbol{\theta}^{(1)}_*,\boldsymbol{\theta}^{(2)}_*,\boldsymbol{\theta}_*)$ is absolutely continuous with respect to P_0^k. In order to determine a density of $(\boldsymbol{\theta}^*,\boldsymbol{\zeta}^*,\widetilde{\pi}_{n_1,n_2})$, introduce $\mathcal{A}_\varepsilon$ as in the proof of Proposition 2 with

$$A^\varepsilon_{j,1}\downarrow\{\theta^*_{j,1}\}\quad A^\varepsilon_{\ell,2}\downarrow\{\theta^*_{\ell,2}\}\quad A^\varepsilon_r\downarrow\{\theta^*_r\}$$

as $\varepsilon\downarrow 0$ and observe that

$$\lim_{\varepsilon\downarrow 0}\frac{\mathbb{P}\left[\Lambda_{\boldsymbol{n},\boldsymbol{k}}(\mathcal{A}_\varepsilon)\right]}{P_0^k(\mathcal{A}_\varepsilon)}=g(\boldsymbol{n}^{(1)},\boldsymbol{n}^{(2)},\boldsymbol{q}^{(1)},\boldsymbol{q}^{(2)},\boldsymbol{\zeta}^*)$$

and that

$$\mathcal{L}(\boldsymbol{\theta}^*, \boldsymbol{\pi}_{n_1,n_2}, \boldsymbol{\zeta}) = g(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{\zeta}^*) \prod_{i=1}^{k} g_0(\theta_i^*). \tag{43}$$

Since the vector $(\boldsymbol{X}^{(n_1)}, \boldsymbol{Y}^{(n_2)})$, given the partition $\tilde{\pi}_{n_1.n_2} = \pi_{n_1,n_2}$ and the distinct values $(\boldsymbol{\theta}_*^{(1)}, \boldsymbol{\theta}_*^{(2)}, \boldsymbol{\theta}_*)$, is independent from the labels $\boldsymbol{\zeta}$, the result follows from (23). □

A.4. Proof of Corollary 2

If $(\tilde{\mu}_1, \tilde{\mu}_2)$ are GM–dependent gamma CRMs, then one has $\tau_q = \Gamma(q)(1+u)^{-q}$ and $\psi(u) = \log(1+u)$. By plugging these expressions into (26) and resorting to identity 3.197.1 in [14] we obtain that $g(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{\zeta}^*)$ is equal to

$$c^k\, z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{\tilde{k}_0} \frac{\Gamma(c+n_1-\bar{n}_1)}{\Gamma(n_1)\Gamma(c+n_1+n_2-\bar{n}_1)}\, \xi(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^*)$$

$$\times \int_0^\infty u^{n_1-1}(1+u)^{-c-n_1+\bar{n}_2}\, {}_2F_1(\bar{n}_2+cz, n_2; n_1+n_2-\bar{n}_1+c; -u)\mathrm{d}u, \tag{44}$$

where we recall that $\tilde{k}_0 = k_1 + k_2 - \tilde{k}_1 - \tilde{k}_2$. The simple change of variable $t = u/(1+u)$ and the transformation formula for hypergeometric functions

$${}_2F_1(\alpha, \beta; \gamma; z) = (1-z)^{-\alpha}\, {}_2F_1(\alpha, \gamma-\beta; \gamma; z/(z-1))$$

let us rewrite the integral in (44) as

$$\int_0^1 t^{n_1-1}(1-t)^{c+cz-1}\, {}_2F_1(\bar{n}_2+cz, c+n_1-\bar{n}_1; c+n_1+n_2-\bar{n}_1; t)\mathrm{d}t.$$

The proof is then completed by resorting to identity 7.512.5 in [14] □

A.5. Proof of Corollary 3

If $(\tilde{\mu}_1, \tilde{\mu}_2)$ are GM–dependent σ-stable CRMs, then one has $\tau_q = \sigma(1-\sigma)_{q-1}u^{\sigma-q}$ and $\psi(u) = u^\sigma$. By plugging these expressions into (26) we obtain that $g(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)}, \boldsymbol{q}^{(1)}, \boldsymbol{q}^{(2)}, \boldsymbol{\zeta}^*)$ is equal to

$$\frac{c^k\, z^{\tilde{k}_1+\tilde{k}_2}(1-z)^{\tilde{k}_0}\, \sigma^k}{\Gamma(n_1)\Gamma(n_2)}\, \xi_\sigma(\boldsymbol{n}^{(n_1)}, \boldsymbol{n}^{(n_2)}, \boldsymbol{q}^*)$$

$$\times \int_0^\infty \int_0^\infty \frac{u^{n_1-\bar{n}_1+\tilde{k}_1\sigma-1} v^{n_2-\bar{n}_2+\tilde{k}_2\sigma-1}(u+v)^{\tilde{k}_0\sigma-n_1-n_2+\bar{n}_1+\bar{n}_2}}{\exp\{c[z(u^\sigma+v^\sigma)+(1-z)(u+v)^\sigma]\}}\mathrm{d}u\mathrm{d}v.$$

The proof is completed by carefully applying the change of variables $w = u/(u+v)$ and $s = u+v$. □

Acknowledgements.

The authors are grateful to an Associate Editor and three referees for their constructive comments and valuable suggestions. This work was supported by the European Research Council (ERC) through StG "N-BNP" 306406. Part of the material presented here is contained in the PhD thesis [31] defended at the University of Pavia (Italy) in June 2011.

References

[1] Antoniak, C.E. (1974). Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1174.
[2] Bailey, W.N. (1964). *Generalized hypergeometric series.* Cambridge University Press.
[3] Barrientos, A.F., Jara, A. and Quintana, F.A. (2012). On the support of MacEachern's dependent Dirichlet processes and extensions. *Bayesian Analysis* **7**, 277–310.
[4] Cifarelli, D.M. and Regazzini, E. (1978). Problemi statistici non parametrici in condizioni di scambiabilità parziale. *Quaderni Istituto Matematica Finanziaria, Università di Torino* Serie III, **12**. English translation available at: http://www.unibocconi.it/wps/allegatiCTP/CR-Scamb-parz[1].20080528.135739.pdf
[5] Constantines, G.M. and Savits, T.H. (1996). A multivariate version of the Faa di Bruno formula. *Trans. Amer. Math. Soc.* **348**, 503–520.
[6] Daley, D.J. and Vere–Jones, D. (1988). *An introduction to the theory of point processes.* Springer, New York.
[7] Dunson, D.B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. Eds.), pp. 223–273. Cambridge University Press, Cambridge.
[8] Epifani, I. and Lijoi, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica* **20**, 1455–1484.
[9] Escobar, M.D. and West, M. (1995). Bayesian density estimation and inference using mixtures, *J. Amer. Statist. Assoc.* **90**, 577–588.
[10] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Population Biology* **3**, 87–112
[11] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.
[12] de Finetti, B. (1938). Sur la condition d'equivalence partielle. *Actualités scientifiques et industrielles* **739**, 5–18. Herman, Paris.
[13] Gelfand, A.E. and Kottas, A. (2002). A computational approach for full nonparametric Bayesian inference under Dirichlet process mixture models. *J. Comput. Graph. Statist.* **11**, 289–305.
[14] Gradshteyn, I.S. and Ryzhik, I.M. (2007). *Table of Integrals, Series and Products*, 7th Ed. Academic Press, New York.
[15] Griffiths, R.C. and Milne, R.K. (1978). A class of bivariate Poisson processes. *J. Mult. Anal.* **8**, 380–395.

[16] Hjort, N.L., Holmes, C.C. Müller, P. and Walker, S.G. (Eds.) (2010). *Bayesian Nonparametrics*. Cambridge University Press, Cambridge.
[17] Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173.
[18] James, L.F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33**, 1771–1799.
[19] James, L.F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist. 33*, 105–120.
[20] James, L.F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36**, 76–97.
[21] Kingman, J.F.C. (1993). *Poisson processes*. Oxford University Press.
[22] Kolossiatis, M., Griffin, J. and Steel, M.F.J. (2011). On Byesian nonparametric modelling of two correlated distributions. *Statistics and Computing*, to appear.
[23] Leisen, F. and Lijoi, A. (2011). Vectors of two–parameter Poisson–Dirichlet processes. *J. Mult. Anal.* **102**, 482–495.
[24] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. Eds.), pp. 80–136. Cambridge University Press, Cambridge.
[25] Lijoi, A., Mena, R.H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Stat. Soc. Ser. B* **69**, 715–740.
[26] MacEachern, S.N. (1994). Estimating normal means with a conjugate style Dirichlet process prior. *Comm. Statist. Simulation Comput.* **23**, 727–741.
[27] MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association.
[28] MacEachern, S.N. (2000). Dependent Dirichlet processes. *Technical Report*, Ohio State University.
[29] Müller, P. and Quintana, F.A. (2004). Nonparametric Bayesian data analysis. *Statist. Sci.* **19**, 95–110.
[30] Müller, P., Quintana, F. and Rosner, G. (2004). A method for combining inference across related nonparametric Bayesian models. *J. Roy. Statist. Soc. Ser. B* **66**, 735–749.
[31] Nipoti, B. (2011). *Dependent completely random measures and statistical applications*. Ph.D. thesis. Department of Mathematics, University of Pavia.
[32] Olkin, I. and Liu, R. (2003). A bivariate beta distribution, *Statist. Probab. Lett.* **62**, 407–412.
[33] Orbanz, P. (2011). Projective limit random probabilities on Polish spaces. *Electr. J. Statist.* **5**, 1354–1373.
[34] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory Related Fields* **102**, 145–158.
[35] Pitman, J. (2006). *Combinatorial stochastic processes*. Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics N. 1875, Springer, New York.
[36] Prünster, I. (2002). *Random probability measures derived from increasing additive processes and their application to Bayesian statistics*. Ph.d thesis, University of Pavia.

[37] Rao, V.A. and Teh, Y.W. (2009). Spatial normalized Gamma processes. In *Advances in Neural Information Processing Systems* **22**.

[38] Regazzini, E., Lijoi, A. and Prünster, I. (2003). Distributional results for means of random measures with independent increments. *Ann. Statist.* **31**, 560–585.

[39] Teh, Y.W. and Jordan, M.I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. Eds.), pp. 158–207. Cambridge University Press, Cambridge.

[40] West, M, Müller, P. and Escobar, M.D. (1994). Hierarchical priors and mixture models, with application in regression and density estimation. In *Aspects of uncertainty*, 363–386. Wiley, Chichester.