Journal of Multivariate Analysis 102 (2011) 482–495

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Vectors of two-parameter Poisson–Dirichlet processes

Fabrizio Leisen [a], Antonio Lijoi [b,c,d,*]

[a] Universidad Carlos III de Madrid, Departamento de Estadistica, calle de Madrid 126, 28903 Getafe (Madrid), Spain
[b] Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, via San Felice 5, 27100 Pavia, Italy
[c] Collegio Carlo Alberto, via Real Collegio 30, 10024 Moncalieri, Italy
[d] CNR–IMATI, via Bassini 15, 20133 Milano, Italy

ARTICLE INFO

Article history:
Received 8 January 2010
Available online 16 October 2010

AMS 2000 subject classifications:
62F15
62H05
60G57
60G51

Keywords:
Bayesian nonparametric statistics
Bivariate completely random measures
Lévy copula
Partial exchangeability
Poisson–Dirichlet process
Posterior distribution

ABSTRACT

The definition of vectors of dependent random probability measures is a topic of interest in applications to Bayesian statistics. They represent dependent nonparametric prior distributions that are useful for modelling observables for which specific covariate values are known. In this paper we propose a vector of two-parameter Poisson–Dirichlet processes. It is well-known that each component can be obtained by resorting to a change of measure of a σ-stable process. Thus dependence is achieved by applying a Lévy copula to the marginal intensities. In a two-sample problem, we determine the corresponding partition probability function which turns out to be partially exchangeable. Moreover, we evaluate predictive and posterior distributions.

1. Introduction

Random probability measures are a primary tool in the implementation of the Bayesian approach to statistical inference since they can be used to define nonparametric priors. The Dirichlet process introduced in [7] represents the first well-known example. After the appearance of Ferguson's work, a number of generalizations of the Dirichlet process have been proposed. In the present paper, attention is focused on one of such extensions, namely the Poisson–Dirichlet process with parameters (σ, θ), introduced in [16], which hereafter we denote for short as PD(σ, θ). In particular, we confine ourselves to considering values of (σ, θ) such that $\sigma \in (0, 1)$ and $\theta > -\sigma$. It is worth recalling that the PD(σ, θ) process also emerges in various research areas which include, for instance, population genetics, statistical physics, excursions of stochastic processes and combinatorics. See [18] and references therein. Its use within Bayesian nonparametric and semiparametric models has recently become much more frequent. There are various reasons that explain such a growing popularity in statistical practice. Firstly, the PD(σ, θ) process yields a more flexible model for clustering than the one provided by the Dirichlet process. Indeed, if $X_1, \dots, X_n$ are the first n terms of an infinite sequence of exchangeable random variables directed by a PD(σ, θ) process, then the probability that $X_1, \dots, X_n$ cluster into k groups of distinct values with respective positive frequencies

* Corresponding author at: Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, via San Felice 5, 27100 Pavia, Italy.
E-mail addresses: fabrizio.leisen@gmail.com (F. Leisen), lijoi@unipv.it (A. Lijoi).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2010.10.008

$n_1, \ldots, n_k$ coincides with

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

for $k \in \{1, \ldots, n\}$ and for any vector of positive integers $(n_1, \ldots, n_k)$ such that $\sum_{j=1}^k n_j = n$, where $(a)_m = a(a+1)\cdots(a+m-1)$ for any $m \geq 1$ and $(a)_0 \equiv 1$. See [16]. The parameter σ can be used to tune the reinforcement mechanism of larger clusters as highlighted in [13]. Another feature which makes convenient the use of a PD(σ, θ) process for Bayesian inference is its stick-breaking representation. In order to briefly recall the construction, let $(\xi_i)_{i\geq 1}$ be a sequence of independent and identically distributed random variables whose probability distribution α is non-atomic and let $(V_i)_{i\geq 1}$ be a sequence of independent random variables where V_i is beta distributed with parameters $(1-\sigma, \theta + i\sigma)$. If

$$\tilde{p}_1 = V_1 \qquad \tilde{p}_j = V_j \prod_{i=1}^{j-1}(1-V_i) \quad j \geq 2 \tag{2}$$

then the random probability measure $\tilde{p} = \sum_{j\geq 1} \tilde{p}_j\, \delta_{\xi_j}$ coincides in distribution with a PD(σ, θ) process. The simple procedure described in (2) suggests an algorithm for simulating the trajectories of the process. An alternative construction, based on a transformation of the σ-stable completely random measure, will be used in the next sections. Finally, the proposal and implementation of suitable Markov Chain Monte Carlo algorithms has made the application of PD(σ, θ) process quite straightforward even in more complex hierarchical mixture models. A work that has had a remarkable impact in this direction is [9].

Stimulated by the importance of PD(σ, θ) prior in Bayesian nonparametric modelling, our main goal in the present paper is the proposal of a definition of a two-dimensional vector of PD(σ, θ) processes along with an analysis of some of its distributional properties. In this respect our work connects to a very active research area which is focused on the definition of random probability measures suited for applications to nonparametric regression modelling. They are obtained as families of priors $\{\tilde{p}_w : w \in \mathcal{W}\}$ where $\mathcal{W}$ is a covariate space and any two random probabilities $\tilde{p}_{w_1}$ and $\tilde{p}_{w_2}$, for $w_1 \neq w_2$, are dependent. The proposals that have appeared in the literature so far are based on variations of the stick-breaking representation in (2). A typical strategy for introducing covariate-dependence in $\tilde{p}$ consists of letting the distribution of the V_i's or of the ξ_i's, or both, depend on w. Among various recent contributions, we confine ourselves to mentioning [15,4,5,21]. This approach, though fruitful from a computational point of view, has some limitations if one aims to obtain analytical results related to the clustering structure of the observations or the posterior distribution of the underlying dependent random probabilities. Besides these noteworthy applications to Bayesian nonparametric regression, other recent contributions point towards applications to computer science and machine learning. For example, in [24] a hierarchical Dirichlet process is applied to problems in information retrieval and text modelling. The authors in [23] propose a dependent two parameter Poisson–Dirichlet process prior which generalises the hierarchical Dirichlet process of [24] and apply it to segmentation of object categories from image databases. Finally, [22] have proposed a dependent prior which takes on the name of the Mondrian process and is used to model relational data.

In the present paper we resort to a construction of $\tilde{p}$ in terms of a completely random measure $\tilde{\mu}$, a strategy that can be adopted for defining the Dirichlet process itself, as pointed out by [7]. Hence, any two random probability measures $\tilde{p}_{w_1}$ and $\tilde{p}_{w_2}$ are dependent if the completely random measures, say $\tilde{\mu}_1$ and $\tilde{\mu}_2$, that define them are dependent. We will deal with the case where the covariate is binary so that $\mathcal{W}$ consists of two points. This is a typical setting for statistical inference with two-sample data. Dependence between $\tilde{\mu}_1$ and $\tilde{\mu}_2$ is induced by a Lévy copula acting on the respective marginal intensities. A similar approach has been undertaken in [6] with the aim of modelling two-sample survival data, thus yielding a generalization of neutral to the right priors. Assuming within group exchangeability and conditional independence between data from the two groups, we obtain a description of the partition probability function generated by the process we propose as a mixture of products of Gauss' hypergeometric functions. Moreover, we deduce a posterior characterization which allows to evaluate the corresponding family of predictive distributions. The structure of the paper is as follows. In Section 2, the bivariate two parameter PD(σ, θ) random probability measure is defined. In Section 3, the analysis of the induced partition structure is developed for a generic vector of two parameter PD(σ, θ) processes. A specific case is considered in Section 4, where the PD(σ, θ) process vector is generated by a suitable Lévy–Clayton copula. Finally, Section 5 provides a posterior characterization, conditional on a vector of latent non-negative random variables, thus generalizing a well-known result valid for the univariate case.

2. A bivariate PD process

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $(\mathbb{X}, \mathscr{X})$ a measure space, with $\mathbb{X}$ Polish and $\mathscr{X}$ the Borel σ-algebra of subsets of $\mathbb{X}$. Suppose $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are two completely random measures (CRMs) on $(\mathbb{X}, \mathscr{X})$ with respective marginal Lévy measures

$$\bar{\nu}_i(\mathrm{d}x, \mathrm{d}y) = \alpha(\mathrm{d}x)\nu_i(\mathrm{d}y) \quad i = 1, 2.$$

The probability measure α on $\mathbb{X}$ is non-atomic and ν_i is a measure on $\mathbb{R}^+$ such that $\int_{\mathbb{R}^+} \min(y, 1)\nu_i(\mathrm{d}y) < \infty$. For background information on CRMs one can refer to [12]. We further suppose that both $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are σ-stable CRMs, *i.e.*

$$\nu_i(\mathrm{d}y) = \frac{\sigma}{\Gamma(1-\sigma)} y^{-1-\sigma}\mathrm{d}y \quad i = 1, 2 \tag{3}$$

with $\sigma \in (0, 1)$. Moreover, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are dependent and the random vector $(\tilde{\mu}_1, \tilde{\mu}_2)$ has independent increments, in the sense that given A and B in $\mathscr{X}$, with $A \cap B = \varnothing$, then $(\tilde{\mu}_1(A), \tilde{\mu}_2(A))$ and $(\tilde{\mu}_1(B), \tilde{\mu}_2(B))$ are independent. This implies that for any pair of measurable functions $f : \mathbb{X} \to \mathbb{R}$ and $g : \mathbb{X} \to \mathbb{R}$, such that $\int |f|^\sigma \mathrm{d}\alpha < \infty$ and $\int |g|^\sigma \mathrm{d}\alpha < \infty$, one has

$$\mathbb{E}\left[\mathrm{e}^{-\tilde{\mu}_1(f)-\tilde{\mu}_2(g)}\right] = \exp\left\{-\int_{\mathbb{X}}\int_{(0,\infty)^2}\left[1-\mathrm{e}^{-y_1 f(x)-y_2 g(x)}\right]\nu(\mathrm{d}y_1, \mathrm{d}y_2)\alpha(\mathrm{d}x)\right\}. \tag{4}$$

The representation (4) entails that the jump heights of $(\tilde{\mu}_1, \tilde{\mu}_2)$ are independent from the locations where the jumps occur. Moreover, these jump locations are common to both CRMs and are governed by α.

An important issue is the definition of the measure ν in (4): we will determine it in such a way that it satisfies the condition

$$\int_0^\infty \nu(\mathrm{d}x, A) = \int_0^\infty \nu(A, \mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)}\int_A y^{-1-\sigma}\mathrm{d}y \tag{5}$$

for any $A \in \mathscr{B}(\mathbb{R}^+)$. In other words, the marginal Lévy intensities coincide with ν_i in (3). This can be achieved by resorting to the notion of Lévy copula whose description is postponed to Section 4. It is worth pointing out that a similar construction has been provided for bivariate gamma processes in [10]. Indeed, they define a vector of random measures in a similar fashion as we do with $(\tilde{\mu}_1, \tilde{\mu}_2) = \sum_{i\ge 1}(J_{i,1}, J_{i,2})\,\delta_{X_i}$. There are two main differences with the present paper. In [10] the marginal CRMs are gamma and the dependence between jump heights $J_{i,1}$ and $J_{i,2}$ is induced by some dependent scaling random factors. On the other hand, here we consider marginal σ-stable random measures with dependence between the jump heights $J_{i,1}$ and $J_{i,2}$ induced indirectly through a Lévy copula. Of course, both the scale invariance approach by [10] and the Lévy copula approach can be extended to deal with CRMs different from the gamma and the σ-stable ones, respectively.

The model we adopt for the observables is as follows. We let $(X_n, Y_n)_{n\ge 1}$ be a sequence of exchangeable random vectors taking values in $\mathbb{X}^2$ for which the following representation holds true

$$\begin{aligned}\mathbb{P}\left[(X_1, Y_1) \in A_1, \ldots, (X_n, Y_n) \in A_n\right] &= \int_{\boldsymbol{P}_{\mathbb{X}^2}}\left\{\prod_{i=1}^n \int_{A_i} p(\mathrm{d}x, \mathrm{d}y)\right\} Q(\mathrm{d}p)\\ &= \int_{\boldsymbol{P}_{\mathbb{X}}^2}\left\{\prod_{i=1}^n \int_{A_i} p_1(\mathrm{d}x)p_2(\mathrm{d}y)\right\} Q^*(\mathrm{d}p_1, \mathrm{d}p_2)\end{aligned}$$

where $\boldsymbol{P}_{\mathbb{X}^2}$ is the space of probability measures on $(\mathbb{X}^2, \mathscr{X}^2)$, $\boldsymbol{P}_{\mathbb{X}}^2 = \boldsymbol{P}_{\mathbb{X}} \times \boldsymbol{P}_{\mathbb{X}}$ is the space of vectors (p_1, p_2) where both p_1 and p_2 are probability measures on $(\mathbb{X}, \mathscr{X})$ and the above representation is valid for any $n \ge 1$ and any choice of sets $A_1, \ldots, A_n$ in $\mathscr{X}^2$. It then follows that Q is a probability distribution on $(\boldsymbol{P}_{\mathbb{X}^2}, \mathscr{P}_{\mathbb{X}^2})$ which degenerates on $(\boldsymbol{P}_{\mathbb{X}}^2, \mathscr{P}_{\mathbb{X}}^2)$. In order to define Q^* we will make use of the σ-stable CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$. Suppose $\mathbb{P}_{i,\sigma}$ is the probability distribution of $\tilde{\mu}_i$, for $i = 1, 2$. Hence $\mathbb{P}_{i,\sigma}$ is supported by the space of all boundedly finite measures $\boldsymbol{M}_{\mathbb{X}}$ on $\mathbb{X}$ endowed with the Borel σ-algebra $\mathscr{M}_{\mathbb{X}}$ with respect to the $w^\sharp$-topology ("weak-hash" topology). Recall that a sequence of measures $(m_i)_{i\ge 1}$ in $\boldsymbol{M}_{\mathbb{X}}$ converges, in the $w^\sharp$-topology, to a measure m in $\boldsymbol{M}_{\mathbb{X}}$ if and only if $m_i(A) \to m(A)$ for any bounded set $A \in \mathscr{X}$ such that $m(\partial A) = 0$. See [3] for further details. Introduce, now, another probability distribution $\mathbb{P}_{i,\sigma,\theta}$ on $(\boldsymbol{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ such that $\mathbb{P}_{i,\sigma,\theta} \ll \mathbb{P}_{i,\sigma}$ and

$$\frac{\mathrm{d}\mathbb{P}_{i,\sigma,\theta}}{\mathrm{d}\mathbb{P}_{i,\sigma}}(\mu) = \frac{\Gamma(\theta+1)}{\Gamma\left(\frac{\theta}{\sigma}+1\right)}[\mu(\mathbb{X})]^{-\theta}.$$

We denote with $\tilde{\mu}_{i,\sigma,\theta}$ a random element defined on $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $(\boldsymbol{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ whose probability distribution coincides with $\mathbb{P}_{i,\sigma,\theta}$. The random probability measure $\tilde{p}_i = \tilde{\mu}_{i,\sigma,\theta}/\tilde{\mu}_{i,\sigma,\theta}(\mathbb{X})$ is a PD(σ, θ) process. See, e.g., [19,18]. Hence, Q^* is the probability distribution of the vector $(\tilde{p}_1, \tilde{p}_2)$ of Poisson–Dirichlet random probability measure on $(\mathbb{X}, \mathscr{X})$. We are then assuming that the sequence of random variables $(X_n, Y_n)_{n\ge 1}$ is exchangeable, such that

$$\mathbb{P}\left[\boldsymbol{X}_{n_1} \in \times_{i=1}^{n_1} A_i;\ \boldsymbol{Y}_{n_2} \in \times_{j=1}^{n_2} B_j \mid (\tilde{p}_1, \tilde{p}_2)\right] = \prod_{i=1}^{n_1}\tilde{p}_1(A_i)\prod_{j=1}^{n_2}\tilde{p}_2(B_j) \tag{6}$$

with $\boldsymbol{X}_{n_1} = (X_1, \ldots, X_{n_1})$ and $\boldsymbol{Y}_{n_2} = (Y_1, \ldots, Y_{n_2})$. We will particularly focus on the case where the dependence between $\tilde{p}_1$ and $\tilde{p}_2$ is determined by the copula $C_{1/\sigma}$ as described later in (13).

3. Partition structure

The description of the model as provided by (6) implies that we are considering the two samples $(X_1, \ldots, X_{n_1})$ and $(Y_1, \ldots, Y_{n_2})$ as independent, conditional on $(\tilde{p}_1, \tilde{p}_2)$. Each $\tilde{p}_i$ is, almost surely, discrete so that

$$\tilde{p}_1\tilde{p}_2 = \sum_{i\geq 1}\sum_{j\geq 1}\omega_{1,i}\omega_{2,j}\delta_{Z_i}\delta_{Z_j} \tag{7}$$

where δ_x is the usual notation for the unit mass concentrated at x, $\sum_{i\geq 1}\omega_{1,i} = \sum_{i\geq 1}\omega_{2,i} = 1$ ($\mathbb{P}$-almost surely), and the Z_i's are i.i.d. from the non-atomic probability distribution α on $(\mathbb{X}, \mathscr{X})$.

Given the discrete nature of the random probability measure in (7), there might be ties, i.e. common values with certain multiplicities, among X_i's and the Y_i's. It, then, follows that there are $1 \leq K \leq n_1 + n_2$ distinct values, say $Z_1^*, \ldots, Z_K^*$ among the components of $\boldsymbol{X}_{n_1} = (X_1, \ldots, X_{n_1})$ and $\boldsymbol{Y}_{n_2} = (Y_1, \ldots, Y_{n_2})$. Moreover, let

$$N_{i,1} = \sum_{l=1}^{n_1}\mathbb{1}_{X_l=Z_i^*} \qquad N_{j,2} = \sum_{l=1}^{n_2}\mathbb{1}_{Y_l=Z_j^*} \quad i,j = 1,\ldots,K$$

be the frequencies associated to each distinct value from the two samples. It is clear that there might also be values in common between the $\boldsymbol{X}_{n_1}$ and the $\boldsymbol{Y}_{n_2}$ sample so that for any $i \in \{1, \ldots, k\}$ both $N_{i,1}$ and $N_{i,2}$ are positive integers with positive probability. According to this, for our purposes the data can be described as the set

$$\{K, N_{1,1}, \ldots, N_{K,1}, N_{1,2}, \ldots, N_{K,2}, Z_1^*, \ldots, Z_K^*\}.$$

In particular, in the present section we will investigate the probability distribution of the partition of $\boldsymbol{X}_{n_1}$ and $\boldsymbol{Y}_{n_2}$ expressed in terms of $K, \boldsymbol{N}_1 = (N_{1,1}, \ldots, N_{K,1})$ and $\boldsymbol{N}_2 = (N_{1,2}, \ldots, N_{K,2})$. This takes on the name of *partition probability function* according to the terminology adopted in [16] and we shall denote it as

$$\Pi_k^{(n_1,n_2)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{P}[K = k, \boldsymbol{N}_1 = \boldsymbol{n}_1, \boldsymbol{N}_2 = \boldsymbol{n}_2]$$

for $1 \leq k \leq n$ and for vectors of non-negative integers $\boldsymbol{n}_i = (n_{1,i}, \ldots, n_{k,i})$ such that $\sum_{j=1}^k n_{j,i} = n_i$, for $i = 1, 2$, and $n_{j,1} + n_{j,2} \geq 1$ for $j = 1, \ldots, k$. As a consequence of (6) one has

$$\Pi_k^{(n_1,n_2)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{E}\left[\int_{\mathbb{X}^k}\pi_k^{(n_1,n_2)}(\mathrm{d}\boldsymbol{z})\right] \tag{8}$$

where

$$\pi_k^{(n_1,n_2)}(\mathrm{d}\boldsymbol{z}) = \prod_{j=1}^k\left(\frac{\tilde{\mu}_{1,\sigma,\theta}(\mathrm{d}z_j)}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,1}}\left(\frac{\tilde{\mu}_{2,\sigma,\theta}(\mathrm{d}z_j)}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,2}}.$$

As we shall shortly see, an important lemma for obtaining an expression for $\Pi_{\boldsymbol{k}}^{(n_1,n_2)}$ in (8) is the following

Lemma 1. *Let $(\tilde{\mu}_1, \tilde{\mu}_2)$ be a vector of CRMs with Laplace exponent $\psi(\cdot,\cdot)$. If $C_\epsilon \in \mathscr{X}$ is such that* $\mathrm{diam}(C_\epsilon) \downarrow 0$ *as $\epsilon \downarrow 0$, then*

$$\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}_1(C_\epsilon)-t\tilde{\mu}_2(C_\epsilon)}\prod_{i=1}^2\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right] = (-1)^{q_1+q_2-1}\alpha(C_\epsilon)\mathrm{e}^{-\alpha(C_\epsilon)\psi(s,t)} \times \frac{\partial^{q_1+q_2}}{\partial s^{q_1}\partial t^{q_2}}\psi(s,t) + o(\alpha(C_\epsilon)) \tag{9}$$

as $\epsilon \downarrow 0$.

Proof. The proof follows from a simple application of a multivariate version of the Faá di Bruno formula as given in [1]. For notational simplicity, let $|\boldsymbol{w}| := \sum_{i=1}^d w_i$ for any vector $\boldsymbol{w} = (w_1, \ldots, w_d)$ in $\mathbb{R}^d$. We then recall a linear order on the set $\mathbb{N}_0^d$ of d-dimensional vectors of non-negative integers adopted in [1]. Given two vectors $\boldsymbol{x} = (x_1, \ldots, x_d)$ and $\boldsymbol{y} = (y_1, \ldots, y_d)$ in $\mathbb{N}_0^d$, then $\boldsymbol{x} \prec \boldsymbol{y}$ if either $|\boldsymbol{x}| < |\boldsymbol{y}|$ or $|\boldsymbol{x}| = |\boldsymbol{y}|$ and $x_1 < y_1$ or if $|\boldsymbol{x}| = |\boldsymbol{y}|$ with $x_i = y_i$ for $i = 1, \ldots, j$ and $x_{j+1} < y_{j+1}$ for some j in $\{1, \ldots, d\}$. Hence note that

$$\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}(C_\epsilon)-t\tilde{\mu}(C_\epsilon)}\prod_{i=1}^2\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right] = (-1)^{q_1+q_2}\frac{\partial^{q_1+q_2}}{\partial s^{q_1}\partial t^{q_2}}\,\mathrm{e}^{-\alpha(C_\epsilon)\psi(s,t)}$$

and by virtue of Theorem 2.1 in [1] one has that the right-hand side above coincides with

$$\mathrm{e}^{-\alpha(C_\epsilon)\psi(s,t)}q_1!q_2!\sum_{k=1}^{q_1+q_2}(-1)^k[\alpha(C_\epsilon)]^k \times \sum_{j=1}^{q_1+q_2}\sum_{p_j(q_1,q_2,k)}\prod_{i=1}^j\frac{1}{\lambda_i!(s_{1,i}!s_{2,i}!)^{\lambda_i}}\left(\frac{\partial^{s_{1,i}+s_{2,i}}}{\partial s^{s_{1,i}}\partial t^{s_{2,i}}}\psi(s,t)\right)^{\lambda_i}$$

where $p_j(q_1, q_2, k)$ is the set of vectors $(\boldsymbol{\lambda}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_j)$ with $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_j)$ a vector whose positive coordinates are such that $\sum_{i=1}^j\lambda_i = k$ and the $\boldsymbol{s}_i = (s_{1,i}, s_{2,i})$ are vectors such that $\boldsymbol{0} \prec \boldsymbol{s}_1 \prec \cdots \prec \boldsymbol{s}_j$. Obviously, in the previous sum, all terms with $k \geq 2$ are $o(\alpha(C_\epsilon))$ as $\epsilon \downarrow 0$. Hence, by discarding these summands one has the result stated in (9). □

If we further suppose that the bivariate Lévy measure is of finite variation, i.e. $\int_{\|\boldsymbol{y}\|\le 1}\|\boldsymbol{y}\|\nu(y_1,y_2)\mathrm{d}y_1\mathrm{d}y_2<\infty$ where $\|\boldsymbol{y}\|$ stands for the Euclidean norm of the vector $\boldsymbol{y}=(y_1,y_2)$, then one also has $\int_{\|\boldsymbol{y}\|\le 1} y_1^{n_1}y_2^{n_2}\nu(y_1,y_2)\mathrm{d}y_1\mathrm{d}y_2<\infty$ for any n_1 and n_2 positive integers. Consequently, one can interchange derivative and integral signs to obtain from (9) the following expression

$$\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}(C_\epsilon)-t\tilde{\mu}(C_\epsilon)}\prod_{i=1}^{2}\{\tilde{\mu}_i(C_\epsilon)\}^{q_i}\right]=\alpha(C_\epsilon)\mathrm{e}^{-\alpha(C_\epsilon)\psi(s,t)}g_\nu(q_1,q_2;s,t)+o(\alpha(C_\epsilon)) \tag{10}$$

as $\epsilon\downarrow 0$, for any $s>0$ and $t>0$, where

$$g_\nu(q_1,q_2;s,t):=\int_{(0,\infty)^2} y_1^{q_1}y_2^{q_2}\mathrm{e}^{-sy_1-ty_2}\nu(y_1,y_2)\mathrm{d}y_1\mathrm{d}y_2.$$

One can now state the main result which provides a probabilistic characterization of the partition structure induced by the random probability distribution structure (7).

Theorem 1. *For any positive integers n_1,n_2 and k and vectors $\boldsymbol{n}_1=(n_{1,1},\dots,n_{k,1})$ and $\boldsymbol{n}_2=(n_{1,2},\dots,n_{k,2})$ such that $\sum_{j=1}^k n_{j,i}=n_i$ and $n_{i,1}+n_{i,2}\ge 1$, for $i=1,2$, one has*

$$\Pi_k^{(n_1,n_2)}(\boldsymbol{n}_1,\boldsymbol{n}_2)=\frac{\sigma^2}{\Gamma^2\left(\frac{\theta}{\sigma}\right)}\frac{1}{\prod_{i=1}^{2}(\theta)_{n_i}}\int_{(0,\infty)^2}s^{\theta+n_1-1}t^{\theta+n_2-1}\mathrm{e}^{-\psi(s,t)}\times\prod_{j=1}^{k}g_\nu(n_{j,1},n_{j,2};s,t)\mathrm{d}s\mathrm{d}t. \tag{11}$$

Proof. For simplicity, we let $\tilde{\mu}_i$ denote the i-th σ-stable completely random measure $\tilde{\mu}_{i,\sigma,0}$, for $i=1,2$. By virtue of the definition of the two-parameter Poisson–Dirichlet process one can then evaluate $\Pi_k^{(n)}$ in (8) by replacing $\pi_k^{(n_1,n_2)}$ with

$$\tilde{\pi}_k^{(n_1,n_2)}(\boldsymbol{n}_1,\boldsymbol{n}_2,\mathrm{d}\boldsymbol{z})=\frac{\sigma^2\Gamma^2(\theta)}{\Gamma^2\left(\frac{\theta}{\sigma}\right)\prod_{i=1}^{2}\left[\tilde{\mu}_i(\mathbb{X})\right]^{\theta+n_i}}\prod_{j=1}^{k}\left[\tilde{\mu}_1(\mathrm{d}z_j)\right]^{n_{j,1}}\left[\tilde{\mu}_2(\mathrm{d}z_j)\right]^{n_{j,2}}$$

for any $k\ge 1$ and $\boldsymbol{n}_i=(n_{1,i},\dots,n_{k,i})$ such that $\sum_{j=1}^k n_{j,i}=n_i$ for $i=1,2$. We will now show that the probability distribution $\mathbb{E}\left[\tilde{\pi}_k^{(n_1,n_2)}\right]$ admits a density on $\mathbb{N}^{2k}\times\mathbb{X}^k$ with respect to the product measure $\gamma^{2k}\times\alpha^k$, where γ is the counting measure on the positive integers, and will determine its form. To this end, suppose $C_{\epsilon,x}$ denotes a neighbourhood of $x\in\mathbb{X}$ of radius $\epsilon>0$ and $B_\epsilon=\times_{j=1}^k C_{\epsilon,z_j}$. Then

$$\begin{aligned}\int_{B_\epsilon}\mathbb{E}\left[\tilde{\pi}_k^{(n_1,n_2)}(\boldsymbol{n}_1,\boldsymbol{n}_2,\mathrm{d}\boldsymbol{z})\right]&=\frac{\sigma^2}{\Gamma^2\left(\frac{\theta}{\sigma}\right)\prod_{i=1}^{2}(\theta)_{n_i}}\int_0^\infty\int_0^\infty s^{\theta+n_1-1}t^{\theta+n_2-1}\\&\quad\times\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}_1(\mathbb{X})-t\tilde{\mu}_2(\mathbb{X})}\prod_{j=1}^{k}\left[\tilde{\mu}_1(C_{\epsilon,z_j})\right]^{n_{j,1}}\left[\tilde{\mu}_2(C_{\epsilon,z_j})\right]^{n_{j,2}}\right]\mathrm{d}s\mathrm{d}t.\end{aligned}$$

Define $\mathbb{X}_\epsilon$ to be the whole space $\mathbb{X}$ with the neighbourhoods C_{ϵ,z_r} deleted for all $j=1,\dots,k$. By virtue of the independence of the increments of the CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$, the expression above reduces to

$$\frac{\sigma^2}{\Gamma^2\left(\frac{\theta}{\sigma}\right)\prod_{i=1}^{2}(\theta)_{n_i}}\int_0^\infty\int_0^\infty s^{\theta+n_1-1}t^{\theta+n_2-1}\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}_1(\mathbb{X}_\epsilon)-t\tilde{\mu}_2(\mathbb{X}_\epsilon)}\right]\times\prod_{j=1}^{k}M_{j,\epsilon}(s,t)\mathrm{d}s\mathrm{d}t$$

where, by virtue of Lemma 1,

$$\begin{aligned}M_{j,\epsilon}(s,t)&:=\mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}_1(C_{\epsilon,z_j})-t\tilde{\mu}_2(C_{\epsilon,z_j})}\left[\tilde{\mu}_1(C_{\epsilon,z_j})\right]^{n_{j,1}}\left[\tilde{\mu}_2(C_{\epsilon,z_j})\right]^{n_{j,2}}\right]\\&=\alpha(C_{\epsilon,z_j})\mathrm{e}^{-\alpha(C_{\epsilon,z_j})\psi(s,t)}g_\nu(n_{j,1},n_{j,2};s,t)+o(\alpha(C_{\epsilon,z_j})).\end{aligned}$$

This shows that $\mathbb{E}[\tilde{\pi}^k]$ admits a density with respect to $\gamma^{2k}\times\alpha^k$ and it is given by

$$\frac{\sigma^2}{\Gamma^2\left(\frac{\theta}{\sigma}\right)\prod_{i=1}^{2}(\theta)_{n_i}}\int_0^\infty\int_0^\infty s^{\theta+n_1-1}t^{\theta+n_2-1}\mathrm{e}^{-\psi(s,t)}\prod_{j=1}^{k}g_\nu(n_{j,1},n_{j,2};s,t)\mathrm{d}s\mathrm{d}t.$$

And this completes the proof. □

It is worth noting that the results displayed in the previous Theorem 1 can be adapted to obtain an evaluation of the mixed moment of the vector $(\tilde{p}_1(A), \tilde{p}_2(B))$ for any A and B in $\mathscr{X}$. Indeed, one has

Theorem 2. *Let A and B be any two sets in $\mathscr{X}$. Then*

$$\mathbb{E}\left[\tilde{p}_1(A)\tilde{p}_2(B)\right] = \alpha(A)\alpha(B) + \frac{\alpha(A\cap B) - \alpha(A)\alpha(B)}{\left[\Gamma(\frac{\theta}{\sigma}+1)\right]^2}\int_{(\mathbb{R}^+)^2} (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t)\mathrm{d}s\mathrm{d}t. \tag{12}$$

Proof. Proceeding in a similar fashion as in the proof of the previous Theorem 1, one has

$$\mathbb{E}\left[\tilde{p}_1(A)\tilde{p}_2(B)\right] = \frac{\sigma^2}{\theta^2\Gamma^2(\theta/\sigma)}\int_{(\mathbb{R}^+)^2} (st)^\theta \mathbb{E}\left[\mathrm{e}^{-s\tilde{\mu}_1(\mathbb{X})-t\tilde{\mu}_2(\mathbb{X})}\tilde{\mu}_1(A)\tilde{\mu}_2(B)\right]\mathrm{d}s\mathrm{d}t.$$

It now suffices to consider the partition of $\mathbb{X}$ induced by $\{A, B\}$ which allows to exploit the independence of the increments of $(\tilde{\mu}_1, \tilde{\mu}_2)$ and resort to the following identity

$$\int_{(\mathbb{R}^+)^2} (st)^\theta \mathrm{e}^{-\psi(s,t)}\left\{g_\nu(1, 0; s, t)g_\nu(0, 1; s, t) + g_\nu(1, 1; s, t)\right\}\mathrm{d}s\mathrm{d}t = \Gamma^2\left(\frac{\theta}{\sigma}+1\right),$$

for any $\theta > -\sigma, \sigma \in (0, 1)$ and ν. Then the application of the multivariate Faá di Bruno formula yields the claimed result. □

The expression in (12) can be used to determine the correlation between $\tilde{p}_1(A)$ and $\tilde{p}_2(B)$, a quantity which is of great interest for prior specification in Bayesian nonparametric inference. Recalling that $\mathbb{E}[\tilde{p}_i(C)] = \alpha(C)$ for any $C \in \mathscr{X}$ and for any $i = 1, 2$, then

$$\mathrm{cov}(\tilde{p}_1(A), \tilde{p}_2(B)) = \frac{\alpha(A\cap B) - \alpha(A)\alpha(B)}{\left[\Gamma\left(\frac{\theta}{\sigma}+1\right)\right]^2}\int_{(\mathbb{R}^+)^2} (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t)\mathrm{d}s\mathrm{d}t.$$

As expected, if the two events A and B are independent with respect to the probability measure α, then the corresponding random probability masses $\tilde{p}_1(A)$ and $\tilde{p}_2(B)$ are uncorrelated. Moreover, if one recalls that for a Poisson–Dirichlet process $\tilde{p}$ with parameters (σ, θ) and baseline measure α one has $\mathrm{var}(\tilde{p}(A)) = \alpha(A)[1-\alpha(A)](1-\sigma)/(\theta+1)$, one can straightforwardly note that

$$\mathrm{corr}(\tilde{p}_1(B), \tilde{p}_2(B)) = \frac{\theta+1}{(1-\sigma)\Gamma^2\left(\frac{\theta}{\sigma}+1\right)}\int_{(\mathbb{R}^+)^2} (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t)\mathrm{d}s\mathrm{d}t$$

for any B in $\mathscr{X}$. The fact that the previous correlation does not depend on the specific set B is usually seen as a desired property in applications to Bayesian inference, since it can be considered as an overall measure of dependence between random probability measures $\tilde{p}_1$ and $\tilde{p}_2$.

4. Lévy–Clayton copula

Let us now focus on the case where the $\tilde{\mu}_i$'s are both σ-stable CRMs whose dependence is determined by a Lèvy copula. See [2,11]. A well-known example is the so-called Lévy–Clayton copula defined as

$$C_\lambda(x_1, x_2) = (x_1^{-\lambda} + x_2^{-\lambda})^{-\frac{1}{\lambda}} \tag{13}$$

with $\lambda > 0$ and its name is due to fact it is reminiscent of the Clayton copula for probability distributions. In this construction λ is a parameter that tunes the degree of dependence between $\tilde{\mu}_1$ and $\tilde{\mu}_2$. See [2]. As a consequence of Theorem 5.4 in [2], the Lévy intensity of the random vector $(\tilde{\mu}_1, \tilde{\mu}_2)$ is

$$\nu(y_1, y_2) = \left.\frac{\partial^2 C_\lambda(x_1, x_2)}{\partial x_1 \partial x_2}\right|_{x_1=U_1(y_1),\ x_2=U_2(y_2)} \nu_1(y_1)\nu(y_2)$$

where $U_i(y) = \nu_i(y, +\infty)$, for $i = 1, 2$, are the marginal tail integrals. It can be easily checked that in this case one would have

$$\nu(y_1, y_2) = \frac{(\lambda+1)\sigma^2}{\Gamma(1-\sigma)}\frac{y_1^{\lambda\sigma-1}y_2^{\lambda\sigma-1}}{(y_1^{\lambda\sigma}+y_2^{\lambda\sigma})^{\frac{1}{\lambda}+2}}\mathbb{1}_{(0,+\infty)^2}(y_1, y_2). \tag{14}$$

A direct use of this bivariate Lévy intensity in Theorem 1 makes it difficult to provide an exact analytic evaluation of the function $g_\nu(n_{j,1}, n_{j,2}; s, t)$. On the other hand, if we confine ourselves to considering the case where $\lambda = 1/\sigma$ one has

$$\nu(y_1, y_2) = \frac{\sigma(1+\sigma)}{\Gamma(1-\sigma)}(y_1+y_2)^{-\sigma-2}\mathbb{1}_{(0,+\infty)^2}(y_1, y_2) \tag{15}$$

and the function g_ν can be exactly evaluated. Besides this analytical advantage, it should also be noted that setting $\lambda = 1/\sigma$ links the parameter governing the dependence between $\tilde{\mu}_1$ and $\tilde{\mu}_2$ and the parameter that influences the clustering structure induced by the bivariate PD processes $(\tilde{p}_1, \tilde{p}_2)$. The effect of this assumption is a lower bound on λ since it implies that $\lambda \in (1, \infty)$. In other terms, λ cannot approach values yielding the independent copula from (13), see [2]. Nonetheless, if one is willing to preserve the general form of the intensity in (14), with the additional parameter λ governing the correlation structure, it is possible to proceed with a numerical evaluation of the integral defining $g_\nu(n_{j,1}, n_{j,2}; s, t)$. Alternatively, a full Bayesian analysis based on this vector prior can be developed by relying on a simulation algorithm as devised in [2]. Here we do not pursue this issue which will be the object of future research.

If we take (15) as the bivariate Lévy intensity, for any $s, t > 0$, with $s \neq t$, the Laplace exponent of $(\tilde{\mu}_1, \tilde{\mu}_2)$ is

$$\psi(s,t) := \int_{(0,+\infty)^2} \left[1 - \mathrm{e}^{-sy_1 - ty_2}\right] \nu(y_1, y_2)\mathrm{d}y_1\mathrm{d}y_2 = \frac{t^{\sigma+1} - s^{\sigma+1}}{t-s}. \tag{16}$$

Moreover, $\psi(t,t) = (\sigma+1)t^\sigma$ for any $t > 0$. Interestingly note that ψ is symmetric, i.e. $\psi(s,t) = \psi(t,s)$ for any $s > 0$ and $t > 0$. Given this, we now proceed to determine the partially exchangeable partition probability function corresponding to the bivariate PD process. Define the function

$$\zeta_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z) := \prod_{j=1}^k \frac{n_{j,1}! n_{j,2}!}{(\bar{n}_j+1)!} \frac{{}_2F_1(n_{j,2}+1, \bar{n}_j - \sigma; \bar{n}_j + 2; 1-z)}{{}_2F_1(1, -\sigma; 2; 1-z)}$$

$$\xi_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z) = \prod_{j=1}^k \frac{n_{j,1}! n_{j,2}!}{(\bar{n}_j+1)!} \frac{{}_2F_1(n_{j,1}+1, \bar{n}_j - \sigma; \bar{n}_j + 2; 1-z)}{{}_2F_1(1, -\sigma; 2; 1-z)}$$

where $\bar{n}_j := n_{j,1} + n_{j,2} \geq 1$, for any j, and ${}_2F_1$ denotes the Gauss hypergeometric function. Hence, one can deduce the following result

Theorem 3. *For any integer $n \geq 1$ and vector $(\boldsymbol{k}, \boldsymbol{n}, \boldsymbol{l})$ in A_{n_1,n_2}, one has*

$$\begin{aligned} \Pi_k^{(n_1,n_2)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = {} & \frac{\sigma^{1+k}\Gamma\left(\frac{2\theta}{\sigma}+k\right)}{\Gamma^2\left(\frac{\theta}{\sigma}\right)\prod_{i=1}^2(\theta)_{n_i}} \prod_{j=1}^k (1-\sigma)_{\bar{n}_j-1} \int_0^1 \left(\frac{1-z}{1-z^{\sigma+1}}\right)^{\frac{2\theta}{\sigma}} \\ & \times \left[z^{\theta+n_2-1}\zeta_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z) + z^{\theta+n_1-1}\xi_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z)\right] \mathrm{d}z. \end{aligned} \tag{17}$$

Proof. Set $\bar{q} := q_1 + q_2$, for any integers q_1 and q_2, and suppose $\bar{q} \geq 1$. Since $\psi(s,t)$ is evaluated as in (16) one obtains

$$\begin{aligned} g_\nu(q_1, q_2; s, t) &= (-1)^{\bar{q}-1} \frac{\partial^{\bar{q}}}{\partial s^{q_1} \partial t^{q_2}} \psi(s,t) \\ &= \sum_{j=0}^{q_2} [\sigma+1]_j (-1)^{q_1-j+1} (\bar{q}-j)! \binom{q_2}{j} t^{\sigma+1-j} (t-s)^{-\bar{q}-1+j} \\ &\quad + \sum_{i=0}^{q_1} [\sigma+1]_i (-1)^{q_2-i+1} (\bar{q}-i)! \binom{q_1}{i} s^{\sigma+1-i} (s-t)^{-\bar{q}-1+i} \end{aligned}$$

where $[a]_j = \prod_{i=1}^j (a-i+1)$ is the j-th descending factorial coefficient of a, with $[a]_0 \equiv 1$. First split the area of integration in (11) into the two disjoint regions $A^+ = \{(s,t) : 0 < t \leq s < \infty\}$ and $A^- = \{(s,t) : 0 < s \leq t < \infty\}$. For $(s,t) \in A^+$, one can resort to Proposition 7 in the Appendix A to obtain

$$g_\nu(q_1, q_2; s, t) = \frac{q_1! q_2! \sigma(\sigma+1)(1-\sigma)_{\bar{q}-1}}{(\bar{q}+1)!} s^{\sigma-\bar{q}} {}_2F_1\left(q_2+1, \bar{q}-\sigma; \bar{q}+2; 1-\frac{t}{s}\right)$$

and the change of variable $(t/s, s) = (z, w)$ leads to

$$\begin{aligned} & \int_{A^+} s^{\theta+n_1-1} t^{\theta+n_2-1} \mathrm{e}^{-\psi(s,t)} \prod_{j=1}^k g_\nu(n_{j,1}, n_{j,2}; s, t)\mathrm{d}s\mathrm{d}t \\ & = \sigma^k(\sigma+1)^k \prod_{j=1}^k \frac{n_{j,1}! n_{j,2}! (1-\sigma)_{\bar{n}_j-1}}{(\bar{n}_j+1)!} \int_0^\infty w^{2\theta+k\sigma-1} \int_0^1 \mathrm{e}^{-w^\sigma \frac{1-z^{\sigma+1}}{1-z}} z^{\theta+n_2-1} \\ & \quad \times \prod_{j=1}^k {}_2F_1(n_{j,2}+1, \bar{n}_j-\sigma; \bar{n}_j+2; 1-z)\mathrm{d}z\mathrm{d}w \\ & = \sigma^{k-1}\Gamma\left(\frac{2\theta}{\sigma}+k\right) \prod_{j=1}^k (1-\sigma)_{\bar{n}_j-1} \int_0^1 z^{\theta+n_2-1} \left(\frac{1-z}{1-z^{\sigma+1}}\right)^{\frac{2\theta}{\sigma}} \zeta_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z)\mathrm{d}z \end{aligned}$$

where the last equality follows from $1 - z^{\sigma+1} = (\sigma+1)(1-z)_2F_1(1, -\sigma; 2; 1-z)$. One works in a similar fashion for $(s, t) \in A^-$ since, in this case, Proposition 7 yields

$$g_\nu(q_1, q_2; s, t) = \frac{q_1! q_2! \sigma(\sigma+1)(1-\sigma)_{\bar{q}-1}}{(\bar{q}+1)!} t^{\sigma-\bar{q}} {}_2F_1\left(q_1+1, \bar{q}-\sigma; \bar{q}+2; 1-\frac{s}{t}\right)$$

so that

$$\int_{A^-} s^{\theta+n_1-1} t^{\theta+n_2-1} \mathrm{e}^{-\psi(s,t)} \prod_{j=1}^k g_\nu(n_{j,1}, n_{j,2}; s, t) \mathrm{d}s\mathrm{d}t = \sigma^{k-1} \Gamma\left(\frac{2\theta}{\sigma}+k\right) \prod_{j=1}^k (1-\sigma)_{\bar{n}_j-1} \times \int_0^1 z^{\theta+n_1-1} \left(\frac{1-z}{1-z^{\sigma+1}}\right)^{\frac{2\theta}{\sigma}} \xi_k(\boldsymbol{n}_1, \boldsymbol{n}_2; z) \mathrm{d}z.$$

The proof of (17) is then completed. □

The representation obtained in Theorem 3 suggests a few considerations that are of great interest if compared to the well-known results for the univariate two parameter PD process. A nice feature about the exchangeable partition probability function $\Pi_k^{(n)}$ in (1) is its symmetry: for any permutation τ of the integers $(1, \ldots, k)$ one has $\Pi_k^{(n)}(n_1, \ldots, n_k) = \Pi_k^{(n)}(n_{\tau(1)}, \ldots, n_{\tau(k)})$. The exchangeability property can be extended to the partition probability function of bivariate PD process in the following terms. Let $\boldsymbol{r}_j = (n_{j,1}, n_{j,2})$ and note that there might be $\boldsymbol{r}_j$ vectors whose first or second coordinate is zero. To take this into account, we introduce disjoint sets of indices I_1 and I_2 identifying those $\boldsymbol{r}_j$ with the first or the second zero coordinate, respectively. Hence, if k_i is the cardinality of I_i one has $0 \le k_1 + k_2 \le k$. An interesting configuration arises when $I_1 \cup I_2 = \{1, \ldots, k\}$ which implies this corresponds to the case where $X_i \neq Y_j$ for any i and j and $k_1 + k_2 = k$. When this happens, we set $I_1 = \{i_1, \ldots, i_{k_1}\}$, $I_2 = \{j_1, \ldots, j_{k_2}\}$ and

$$\Pi_k^{(n_1,n_2)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \Pi_k^{(n_1,n_2)}(n_{i_1,1}, \ldots, n_{i_{k_1},1}, n_{j_1,2}, \ldots, n_{j_{k_2},2}).$$

One can now immediately deduce the following

Corollary 4. *The partition probability function in (17), seen as a function of* $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_k)$ *is symmetric in the sense that for any permutation* τ *of* $(1, \ldots, k)$ *one has*

$$\Pi_k^{(n_1,n_2)}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_k) = \Pi_k^{(n_1,n_2)}(\boldsymbol{r}_{\tau(1)}, \ldots, \boldsymbol{r}_{\tau(k)}).$$

Moreover, if $k_1 + k_2 = k$, *then for any permutations* τ_1 *and* τ_2 *of integers in* $I_1 = \{i_1, \ldots, i_{k_1}\}$ *and* $I_2 = \{j_1, \ldots, j_{k_2}\}$, *respectively, one has*

$$\Pi_k^{(n_1,n_2)}(n_{i_1,1}, \ldots, n_{i_{k_1},1}, n_{j_1,2}, \ldots, n_{j_{k_2},2}) = \Pi_k^{(n_1,n_2)}(n_{\tau_1(i_1),1}, \ldots, n_{\tau_1(i_{k_1}),1}, n_{\tau_2(j_1),2}, \ldots, n_{\tau_2(j_{k_2}),2}).$$

Hence one observes that, seen as a function of the pairs of integers $(n_{j,1}, n_{j,2})$, $\Pi_k^{(n_1,n_2)}$ is symmetric. On the other hand, if $\Pi_k^{(n_1,n_2)}$ is restricted to those partitions of $n_1 + n_2$ data such that $X_i \neq Y_j$ for any i and j, then $\Pi_k^{(n_1,n_2)}$ is *partially exchangeable* with respect to the single frequencies $n_{i,1}$ and $n_{j,2}$.

As for the correlation between $\tilde{p}_1(A)$ and $\tilde{p}_2(B)$, one has

$$g_\nu(1, 1; s, t) = \frac{s^{\sigma-2}(-\sigma-1)_3}{3!} {}_2F_1\left(2-\sigma, 2; 4; 1-\frac{t}{s}\right) \quad 0 \le t \le s < \infty$$

and, by virtue of symmetry of the function $(s, t) \mapsto (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t)$,

$$\begin{aligned}\int_{(\mathbb{R}^+)^2} (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t) \mathrm{d}s\mathrm{d}t &= 2\int_0^\infty \mathrm{d}s \int_0^s \mathrm{d}t (st)^\theta \mathrm{e}^{-\psi(s,t)} g_\nu(1, 1; s, t) \\ &= 2\int_0^\infty \mathrm{d}w \int_0^1 \mathrm{d}z w^{2\theta+\sigma-1} \mathrm{e}^{-w^\sigma \frac{1-z^{\sigma+1}}{1-z}} z^\theta (1-z)^{-3} \\ &\quad \times \left\{-\sum_{j=0}^1 [\sigma+1]_j (2-j)! z^{\sigma+1-j}(1-z)^j + \sum_0^1 [\sigma+1]_i (2-i)! (1-z)^i\right\} \mathrm{d}z \\ &= \frac{\Gamma\left(\frac{2\theta}{\sigma}+1\right)(1-\sigma^2)}{3} \int_0^1 z^\theta \left(\frac{1-z}{1-z^{\sigma+1}}\right)^{\frac{2\theta}{\sigma}+1} {}_2F_1(2-\sigma, 2; 4; 1-z) \mathrm{d}z.\end{aligned}$$

This implies that for any A and B in $\mathscr{X}$

$$\operatorname{corr}(\tilde{p}_1(A), \tilde{p}_2(B)) = \frac{\alpha(A\cap B)-\alpha(A)\alpha(B)}{\sqrt{\alpha(A)\alpha(B)(1-\alpha(A))(1-\alpha(B))}} \frac{(\theta+1)(\sigma+1)\Gamma\left(\frac{2\theta}{\sigma}+1\right)}{3\Gamma^2\left(\frac{\theta}{\sigma}+1\right)} \times \int_0^1 z^\theta \left(\frac{1-z}{1-z^{\sigma+1}}\right)^{\frac{2\theta}{\sigma}+1} {}_2F_1(2-\sigma, 2; 4; 1-z)\mathrm{d}z$$

from which one easily deduces $\operatorname{corr}(\tilde{p}_1(B), \tilde{p}_2(B))$. A simplification occurs if $\theta = 0$, in which case $\tilde{p}_i$ is a normalised σ-stable completely random measure and

$$\operatorname{corr}(\tilde{p}_1(B), \tilde{p}_2(B)) = \frac{1}{3}\int_0^1 \frac{{}_2F_1(2-\sigma, 2; 4; 1-z)}{{}_2F_1(-\sigma, 1; 2; 1-z)}\mathrm{d}z \tag{18}$$

for any B in $\mathscr{X}$. The correlation between $\tilde{p}_1(B)$ and $\tilde{p}_2(B)$ does not depend on the specific set B: this is an argument commonly used in Bayesian statistics to interpret the expression in (18) as a prior guess on the correlation between $\tilde{p}_1$ and $\tilde{p}_2$.

5. A posterior characterization

The determination of the posterior distribution of the vector $(\tilde{p}_1, \tilde{p}_2)$, given $\boldsymbol{D} = \{X_i, Y_j : i = 1, \ldots, n_1, j = 1, \ldots, n_2\}$ represents an important issue in Bayesian statistical inference. In the one-sample case, in [17], it is shown that if $(X_i)_{i\geq 1}$ is a sequence of exchangeable random elements directed by a $\mathrm{PD}(\sigma, \theta)$ process, then, conditional on a sample $X_1, \ldots, X_n$ featuring k distinct values $X_1^*, \ldots, X_k^*$ with respective frequencies $n_1, \ldots, n_k$, the random probability measure $\tilde{p}_{\sigma,\theta}$ is identical in distribution to

$$\sum_{j=1}^k w_j \delta_{X_j^*} + (1 - w_1 - \cdots - w_k)\,\tilde{p}^*$$

with the vector $(w_1, \ldots, w_k)$ being distributed according to a k-variate Dirichlet distribution with parameters $(n_1 - \sigma, \ldots, n_k - \sigma, \theta + k\sigma)$ and $\tilde{p}^*$ coinciding in distribution with a $\mathrm{PD}(\sigma, \theta + k\sigma)$ process. Here we aim at extending this result to a vector $(\tilde{p}_1, \tilde{p}_2)$ of dependent $\mathrm{PD}(\sigma, \theta)$ random probability measures.

To this end, we suppose the $n_1 + n_2$ data in $\boldsymbol{D}$ are represented by k distinct values, $Z_1^*, \ldots, Z_k^*$, with corresponding frequencies $(m_1, \ldots, m_k)$ where $m_i = n_{i,1} + n_{i,2}$ for any $i = 1, \ldots, k$. The vector of completely random measures $(\tilde{\mu}_1, \tilde{\mu}_2)$ has intensity ν on $(\mathbb{R}^+)^2$ and we let (S, T) be a vector of non-negative random variables whose distribution, conditional on $\boldsymbol{D}$, admits a density

$$f(s, t|\boldsymbol{D}) \propto s^{\theta+n_1-1} t^{\theta+n_2-1} \mathrm{e}^{-\psi(s,t)} \prod_{j=1}^k g_\nu(n_{j,1}, n_{j,2}; s, t) \tag{19}$$

with $\propto$ meaning that equality holds up to a proportionality constant. As in the previous section, $(\tilde{\mu}_1, \tilde{\mu}_2)$ has the intensity ν specified in (15) so that the marginal components are both σ-stable. Moreover, conditional on (S, T) and on $\boldsymbol{D}$, $(\tilde{\mu}_1^*, \tilde{\mu}_2^*)$ is a vector of completely random measures with Lévy intensity $\nu^*(y_1, y_2) = \mathrm{e}^{-Sy_1-Ty_2}\nu(y_1, y_2)$. From this one can deduce that the conditional univariate Lévy intensities of $\tilde{\mu}_1^*$ and of $\tilde{\mu}_2^*$ are

$$\nu_1^*(y) = \frac{\sigma(\sigma+1)t^{\sigma+1}}{\Gamma(1-\sigma)} \mathrm{e}^{-(s-t)y}\Gamma(-\sigma-1; ty)$$

$$\nu_2^*(y) = \frac{\sigma(\sigma+1)s^{\sigma+1}}{\Gamma(1-\sigma)} \mathrm{e}^{-(t-s)y}\Gamma(-\sigma-1; sy)$$

respectively. The Laplace exponent of $(\tilde{\mu}_1^*, \tilde{\mu}_2^*)$ is given by $\psi^*(\lambda_1, \lambda_2) = \psi(s + \lambda_1, t + \lambda_2) - \psi(s, t)$ for any $s, t, \lambda_1, \lambda_2$, where ψ is as in (16). Note that this implies, for example, that the marginal Laplace exponent for $\tilde{\mu}_1^*$ coincides with

$$\psi_1^*(\lambda_1) = \frac{\sigma(\sigma+1)t^{\sigma+1}}{\Gamma(1-\sigma)} \int_0^\infty \left[1 - \mathrm{e}^{-\lambda_1 y}\right] \mathrm{e}^{-(s-t)y}\Gamma(-\sigma-1; ty)\mathrm{d}y$$

for any s, t and λ_1 such that $s \neq t$ and $s + \lambda_1 \neq t$. Finally, introduce a collection of k mutually independent random pairs $\{(M_{j,1}, M_{j,2}) : j = 1, \ldots, k\}$ and the density function of $(M_{j,1}, M_{j,2})$, conditional on (S, T) and on $\boldsymbol{D}$, is

$$\frac{x^{n_{j,1}} y^{n_{j,2}} \mathrm{e}^{-Sx-Ty}\nu(x, y)}{g_\nu(n_{j,1}, n_{j,2}; S, T)} \mathbb{1}_{(0,+\infty)^2}(x, y). \tag{20}$$

At this point, we are able to describe a characterization of the posterior distribution of $(\tilde{\mu}_{1,\sigma,\theta}, \tilde{\mu}_{2,\sigma,\theta})$ given the data $\boldsymbol{D}$ and the random vector (S, T).

Theorem 5. *Let $(\tilde{p}_1, \tilde{p}_2)$ be a vector of dependent two-parameter Poisson–Dirichlet processes and suppose the data $\boldsymbol{D} = (\boldsymbol{X}_{n_1}, \boldsymbol{Y}_{n_2})$ satisfy (6). Moreover, let $\boldsymbol{D}$ contain k distinct values $Z_1^*, \ldots, Z_k^*$ with respective frequencies $n_{1,1}+n_{1,2}, \ldots, n_{k,1}+n_{k,2}$. The posterior distribution of $(\tilde{\mu}_{1,\sigma,\theta}, \tilde{\mu}_{2,\sigma,\theta})$ given D and (S, T) coincides with the distribution of the CRM*

$$(\tilde{\mu}_1^*, \tilde{\mu}_2^*) + \sum_{j=1}^{k} (M_{j,1}\delta_{Z_j^*}, M_{j,2}\delta_{Z_j^*}) \tag{21}$$

where $(\tilde{\mu}_1^*, \tilde{\mu}_2^*)$ and the k vectors of jump heights $(M_{j,1}, M_{j,2})$, $j = 1, \ldots, k$, are independent.

An analogous description for the univariate PD(σ, θ) process is provided by [14], where the mixture is explicitly evaluated to reproduce Pitman's result. It is worth noting that from (20) one can deduce an expression for the marginal distributions of the jumps, conditional on (S, T) and on $\boldsymbol{D}$. Resorting to (3.383.4) and (7.621.3) in [8] one finds out that the density function of $M_{j,1}$, conditional on (S, T) and on $\boldsymbol{D}$, is

$$f_{M_{j,1}}(x) = \frac{(\bar{n}_j+1)!}{n_{j,1}!\,\Gamma(\bar{n}_j-\sigma)} \frac{t^{1-\frac{n_{j,2}-\sigma}{2}} s^{-1-n_{j,1}}}{{}_2F_1\left(n_{j,1}+1, \sigma+2; \bar{n}_j+2; 1-\frac{t}{s}\right)} \times x^{n_{j,1}+\frac{n_{j,2}-\sigma}{2}-1} \mathrm{e}^{-\left(s-\frac{t}{2}\right)x} W_{-\frac{n_{j,2}+\sigma+2}{2}, -\frac{n_{j,2}-1-\sigma}{2}}(tx)$$

where $\bar{n}_j := n_{j,1} + n_{j,2}$, $W_{\lambda,\mu}$ is the Whittaker function and ${}_2F_1(a, b; c; x)$ is to be interpreted as the analytic continuation of the series representation of the hypergeometric function for x in the complex plane cut along $[1, \infty)$.

Example 1. A possible use of previous result is the determination of the predictive distribution with a bivariate two parameter Poisson–Dirichlet process. For ease of exposition, here we confine ourselves to considering a sample of size $n_1 = 1$ and $n_2 = 0$ and will determine $\mathbb{P}[Y_1 \in A|X_1, S, T]$ and $\mathbb{P}[X_2 \in A|X_1, S, T]$. In this case, conditional on (X_1, S, T), by virtue of Theorem 5 one has $(\tilde{\mu}_{1,\sigma,\theta}, \tilde{\mu}_{2,\sigma,\theta}) \overset{d}{=} (\tilde{\mu}_1^*, \tilde{\mu}_2^*) + (M_1\delta_{X_1}, M_2\delta_{X_1})$. This leads to

$$\mathbb{P}[Y_1 \in A|X_1, S, T] = \mathbb{E}\left[\frac{\mu_2^*(A) + M_2\delta_{X_1}(A)}{\mu_2^*(\mathbb{X}) + M_2}\right].$$

Since M_2 is a non-negative random variable whose density, conditional on (X_1, S, T), coincides with

$$f_{M_2}(y) = \frac{2}{\Gamma(1-\sigma)} \frac{t^{-1} s^{\frac{\sigma+1}{2}}}{{}_2F_1\left(1, \sigma+2; 3; 1-\frac{s}{t}\right)} y^{\frac{1-\sigma}{2}-1} \mathrm{e}^{-\left(t-\frac{s}{2}\right)y} W_{-\frac{\sigma+3}{2}, \frac{\sigma}{2}}(sy).$$

Using this fact, one obtains

$$\mathbb{P}[Y_1 \in A|X_1, S, T] = \alpha(A)\omega_0 + \delta_{X_1}(A)\omega_1 \tag{22}$$

where

$$\begin{aligned}
\omega_0 &= \frac{\sigma(\sigma+1)}{2} \frac{s^{\sigma+1} t^{\sigma-2} \mathrm{e}^{\psi(s,t)}}{{}_2F_1\left(1, 2+\sigma; 3; 1-\frac{s}{t}\right)} \int_0^\infty (t+u)^{-1} \mathrm{e}^{-\psi(s,t+u)} \\
&\quad \times {}_2F_1\left(1, 2+\sigma; 3; 1-\frac{s}{t+u}\right) {}_2F_1\left(1, 1-\sigma; 3; 1-\frac{s}{t+u}\right) \mathrm{d}u \\
\omega_1 &= \frac{1-\sigma}{3} \frac{s^{\frac{\sigma+1}{2}} t^{-1} \mathrm{e}^{\psi(s,t)}}{{}_2F_1\left(1, 2+\sigma; 3; 1-\frac{s}{t}\right)} \times \int_0^\infty (t+u)^{-2} \mathrm{e}^{-\psi(s,t+u)} {}_2F_1\left(2, 2+\sigma; 4; 1-\frac{s}{t+u}\right) \mathrm{d}u.
\end{aligned}$$

Similarly, one can show that

$$\mathbb{P}[X_2 \in A|X_1, S, T] = \alpha(A)\omega_0' + \delta_{X_1}(A)\omega_1' \tag{23}$$

with

$$\begin{aligned}
\omega_0' &= \frac{\sigma(\sigma+1)}{(2+\sigma)\Gamma(1-\sigma)} \frac{\mathrm{e}^{\psi(s,t)} t^{\sigma+2} s^{-2}}{{}_2F_1\left(2, 2+\sigma; 3; 1-\frac{t}{s}\right)} \int_0^\infty (s+u)^{\sigma-2} \mathrm{e}^{-\psi(s+u,t)} \\
&\quad \times {}_2F_1\left(2+\sigma, 2+\sigma; 3+\sigma; 1-\frac{t}{s+u}\right) {}_2F_1\left(1, 1-\sigma; 3; 1-\frac{t}{s+u}\right) \mathrm{d}u \\
\omega_1' &= \frac{2}{(3+\sigma)\Gamma(1-\sigma)} \frac{t^{\sigma+2} s^{-2} \mathrm{e}^{\psi(s,t)}}{{}_2F_1\left(2, 2+\sigma; 3; 1-\frac{t}{s}\right)} \int_0^\infty (s+u)^{\sigma-3} \mathrm{e}^{-\psi(s+u,t)} \\
&\quad \times {}_2F_1\left(3+\sigma, 2+\sigma; 4+\sigma; 1-\frac{t}{s+u}\right) \mathrm{d}u.
\end{aligned}$$

It should be recalled that an analogous expression could have been attained by resorting to the partition probability function described in Section 4.

Acknowledgments

Antonio Lijoi is partially supported by MIUR grant no. 2008MK3AFZ and by the Collegio Carlo Alberto project "New approaches to stochastic modeling in Economics via Bayesian Nonparametrics".

Appendix A. Technical results

We first prove a combinatorial result that serves as a tool for proving the Proposition 7.

Lemma 6. *Suppose $h \in \{0, 1, \ldots, m\}$ and let $n \geq m$. Then*

$$\sum_{j=0}^{h} (-1)^j \binom{h}{j} \binom{n-j}{m} = \frac{h!(n-h)!}{m!(n-m)!} \binom{m}{h}. \tag{A.1}$$

If $h \in \{m+1, \ldots, n\}$, then

$$\sum_{j=0}^{h} (-1)^j \binom{h}{j} \binom{n-j}{m} = 0. \tag{A.2}$$

Proof. Indeed

$$\sum_{j=0}^{h} (-1)^j \binom{h}{j} \binom{n-j}{m} = \frac{1}{m!} \frac{\partial^m}{\partial s^m} \sum_{j=0}^{h} (-1)^j \binom{h}{j} s^{n-j} \Big|_{s=1} = \frac{1}{m!} \frac{\partial^m}{\partial s^m} s^{n-h} (s-1)^h \Big|_{s=1}$$
$$= \frac{1}{m!} \sum_{i=0}^{m} \binom{m}{i} \frac{\partial^i}{\partial s^i} s^{n-h} \Big|_{s=1} \frac{\partial^{m-i}}{\partial s^{m-i}} (s-1)^h \Big|_{s=1} = \frac{1}{m!} \binom{m}{h} [n-h]_{m-h} h!$$

and (A.1) follows. The proof of (A.2) can be deduced in a similar fashion. □

Let us now state the main proposition which is involved in the proof of the results stated in Theorem 2.

Proposition 7. *Let q_1 and q_2 be two non-negative integers and $\bar{q} = q_1 + q_2$. For any $z \in (0, 1)$ one has*

$$\sum_{j=0}^{q_1} [\sigma+1]_j (\bar{q}-j)! \binom{q_1}{j} z^{\sigma+1-j} (1-z)^j - \sum_{i=0}^{q_2} (-1)^i [\sigma+1]_i (\bar{q}-i)! \binom{q_2}{i} (1-z)^i$$
$$= (1-z)^{\bar{q}+1} \frac{(-1)^{q_1} q_1! q_2! (-\sigma-1)_{\bar{q}+1}}{\Gamma(\bar{q}+2)} {}_2F_1(q_1+1, \bar{q}-\sigma; \bar{q}+2; 1-z).$$

Proof. Suppose that $q_2 > q_1$. Since $z^{\sigma+1-j} = \sum_{v \geq 0} [\sigma+1-j]_v (-1)^v (1-z)^v / v!$, then

$$\sum_{j=0}^{q_1} [\sigma+1]_j (\bar{q}-j)! \binom{q_1}{j} z^{\sigma+1-j} (1-z)^j = \sum_{v=0}^{\infty} \frac{(-1)^v}{v!} \sum_{j=0}^{q_1} [\sigma+1]_{j+v} (\bar{q}-j)! \binom{q_1}{j} (1-z)^{j+v}$$
$$= \sum_{h=0}^{\infty} [\sigma+1]_h (1-z)^h \sum_{v=0 \vee (h-q_1)}^{h} \frac{(-1)^v}{v!} (\bar{q}-h+v)! \binom{q_1}{h-v}$$
$$= \left(\sum_{h=0}^{q_1} + \sum_{h=q_1+1}^{q_2} + \sum_{h=q_2+1}^{\bar{q}} + \sum_{h=\bar{q}+1}^{\infty} \right) [\sigma+1]_h (1-z)^h \times \sum_{v=0 \vee (h-q_1)}^{h} \frac{(-1)^v}{v!} (\bar{q}-h+v)! \binom{q_1}{h-v}$$

and $a \vee b := \max\{a, b\}$. As for the first sum above, note that for any $h \in \{0, 1, \ldots, q_1\}$

$$\sum_{v=0}^{h} \frac{(-1)^v}{v!} (\bar{q}-h+v)! \binom{q_1}{h-v} = \frac{q_1! q_2!}{h!} \sum_{v=0}^{h} (-1)^v \binom{h}{v} \binom{\bar{q}-h+v}{q_2}$$
$$= \frac{q_1! q_2!}{h!} (-1)^h \frac{h!(\bar{q}-h)!}{q_1! q_2!} \binom{q_2}{h}$$

where the last equality follows from (A.1). When $h \in \{q_1+1, \ldots, q_2\}$ one has

$$\sum_{v=h-q_1}^{h} \frac{(-1)^v}{v!}(\bar{q}-h+v)!\binom{q_1}{h-v} = (-1)^h(\bar{q}-h)!\sum_{j=0}^{q_1}(-1)^j\binom{q_1}{j}\binom{\bar{q}-j}{\bar{q}-h} = (-1)^h(\bar{q}-h)!\binom{q_2}{h}$$

by virtue of (A.2) to show that the second sum is zero. On the other hand, for any $h \in \{q_2+1, \ldots, \bar{q}\}$ it can be seen that

$$\sum_{v=h-q_1}^{h} \frac{(-1)^v}{v!}(\bar{q}-h+v)!\binom{q_1}{h-v} = (-1)^h\sum_{j=0}^{q_1}(-1)^j\binom{q_1}{j}[\bar{q}-j]_{\bar{q}-h} = 0.$$

Hence, one is left just with the last sum where $h \geq \bar{q}+1$. In this case, from Eq. 0.160.2 in [8] one has

$$\sum_{v=h-q_1}^{h} \frac{(-1)^v}{v!}(\bar{q}-h+v)!\binom{q_1}{h-v} = (-1)^{h+q_1}\frac{\Gamma(h-q_2)q_2!}{h!\Gamma(h-\bar{q})}.$$

Consequently, one has

$$\begin{aligned}
&\sum_{j=0}^{q_1}[\sigma+1]_j(\bar{q}-j)!\binom{q_1}{j}z^{\sigma+1-j}(1-z)^j - \sum_{i=0}^{q_2}(-1)^i[\sigma+1]_i(\bar{q}-i)!\binom{q_2}{i}(1-z)^i \\
&\quad = \sum_{h=\bar{q}+1}^{\infty}[\sigma+1]_h(-1)^{h+q_1}\frac{q_2!\Gamma(h-q_2)}{h!\Gamma(h-\bar{q})}(1-z)^h \\
&\quad = \sum_{j=0}^{\infty}[\sigma+1]_{j+\bar{q}+1}(-1)^{j+q_2+1}\frac{\Gamma(j+q_1+1)q_2!}{\Gamma(j+\bar{q}+2)j!}(1-z)^{j+\bar{q}+1} \\
&\quad = \frac{(-1)^{q_1}q_1!q_2!}{\Gamma(\bar{q}+2)}(1-z)^{\bar{q}+1}\sum_{j=0}^{\infty}\frac{\Gamma(-\sigma+j+\bar{q})}{\Gamma(-\sigma-1)}\frac{(q_1+1)_j}{j!(\bar{q}+2)_j}(1-z)^j
\end{aligned}$$

which yields the stated result. For the case $q_2 \leq q_1$ one works in a similar fashion. □

Appendix B. Proof of Theorem 5

The result will be proved by evaluating the posterior Laplace transform of the vector $(\tilde{\mu}_{1,\sigma,\theta}(A), \tilde{\mu}_{2,\sigma,\theta}(A))$, given D and (S, T). To this end, we resort to a technique introduced in [20]. The idea is to evaluate an approximation of the posterior which is simpler to handle and, then, obtain the posterior via a limiting procedure. This is better illustrated as follows. First note that since $\mathbb{X}$ is separable there exists a sequence $(\Pi_m)_{m\geq1}$ of measurable partitions, with $\Pi_m = \{A_{m,i} : i = 1, \ldots, k_m\}$, such that: (a) Π_{m+1} is a refinement of Π_m; (b) if $\mathscr{G}_m = \sigma(\Pi_m)$, then $\mathscr{X} = \sigma(\cup_{m\geq1}\mathscr{G}_m)$; (c) $\max_{1\leq i\leq k_m+1}\operatorname{diam}(A_{m,i}) \to 0$ as $m \to \infty$. Accordingly, define sequences $(X'_{m,i})_{i\geq1}$ and $(Y'_{m,i})_{i\geq1}$ of $\mathbb{X}$-valued random elements with $X'_{m,l} = \sum_{i=1}^{k_m+1} x_{m,i}\delta_{A_{m,i}}(X_l)$ and $Y'_{m,l} = \sum_{i=1}^{k_m+1} y_{m,i}\delta_{A_{m,i}}(Y_l)$, for any $l \geq 1$, where $x_{m,i}$ and $y_{m,i}$ are points in $A_{m,i}$. It follows that

$$\mathbb{P}[X'_{m,r} \in A, Y_{m,s} \in B \mid (\tilde{\mu}_{1,\sigma,\theta}, \tilde{\mu}_{2,\sigma,\theta})] = \sum_{i,j=1}^{k_m+1}\frac{\tilde{\mu}_{1,\sigma,\theta}(A_{m,i})}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\frac{\tilde{\mu}_{2,\sigma,\theta}(A_{m,j})}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\delta_{x_{m,i}}(A)\delta_{y_{m,j}}(B).$$

It is apparent that if $\mathscr{F}^{(m)}_{n_1,n_2} = \sigma(\boldsymbol{X}'_{m,n_1}, \boldsymbol{Y}'_{m,n_2})$ is the σ-algebra generated by $\boldsymbol{X}'_{m,n_1} = (X'_{m,1}, \ldots, X'_{m,n_1})$ and $\boldsymbol{Y}'_{m,n_2} = (Y'_{m,1}, \ldots, Y'_{m,n_2})$, then

$$\mathscr{F}^{(m)}_{n_1,n_2} \subset \mathscr{F}_{n_1,n_2} := \sigma(X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}).$$

Moreover, set $\boldsymbol{j} = (j_1, \ldots, j_{n_1+n_2}) \in \{1, \ldots, k_m+1\}$ and $R_{m,\boldsymbol{j}} = \times_{i=1}^{n_1+n_2} A_{m,j_i}$, and note that

$$\mathbb{E}[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)} \mid \mathscr{F}^{(m)}_{n_1,n_2}] = \sum_{\boldsymbol{j}}\mathbb{1}_{R_{m,\boldsymbol{j}}}(\boldsymbol{X}'_{m,n_1}, \boldsymbol{Y}'_{m,n_2}) \times \frac{\mathbb{E}\left[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)}\prod_{i=1}^{n_1}\frac{\tilde{\mu}_{1,\sigma,\theta}(A_{m,j_i})}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\prod_{l=n_1+1}^{n_1+n_2}\frac{\tilde{\mu}_{2,\sigma,\theta}(A_{m,j_l})}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\right]}{\mathbb{E}\left[\prod_{i=1}^{n_1}\frac{\tilde{\mu}_{1,\sigma,\theta}(A_{m,j_i})}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\prod_{l=n_1+1}^{n_1+n_2}\frac{\tilde{\mu}_{2,\sigma,\theta}(A_{m,j_l})}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\right]}$$

for any positive λ_1 and λ_2. An application of Proposition 2 in [20] implies that

$$\mathbb{E}[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)} \mid \mathscr{F}^{(m)}_{n_1,n_2}] \to \mathbb{E}[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)} \mid \mathscr{F}_{n_1,n_2}] \tag{B.1}$$

almost surely, as $m \to \infty$. Our main goal will then be the evaluation of the left hand side of (B.1), so that the stated equivalence in distribution with (21) can be achieved by taking the limit as $m \to \infty$. Let us suppose that, for m large enough, the data are gathered into $1 \le k \le n_1+n_2$ sets $A_{m,i_1}, \ldots, A_{m,i_k}$ and set the frequencies $n_{j,1} = \sum_{r=1}^{n_1} \mathbb{1}_{A_{m,j}}(X_r)$, $n_{j,2} = \sum_{r=1}^{n_2} \mathbb{1}_{A_{m,j}}(Y_r)$. The left hand side of (B.1) reduces to

$$\frac{\mathbb{E}\left[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)} \prod_{j=1}^{k} \left(\frac{\tilde{\mu}_{1,\sigma,\theta}(A_{m,j})}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,1}} \left(\frac{\tilde{\mu}_{2,\sigma,\theta}(A_{m,j})}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,2}}\right]}{\mathbb{E}\left[\prod_{j=1}^{k} \left(\frac{\tilde{\mu}_{1,\sigma,\theta}(A_{m,j})}{\tilde{\mu}_{1,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,1}} \left(\frac{\tilde{\mu}_{2,\sigma,\theta}(A_{m,j})}{\tilde{\mu}_{2,\sigma,\theta}(\mathbb{X})}\right)^{n_{j,2}}\right]}. \tag{B.2}$$

By virtue of Theorem 1, the denominator coincides with

$$\frac{\sigma^2 \prod_{j=1}^{k} \alpha(A_{m,j})}{\Gamma^2\left(\frac{\theta}{\sigma}\right) \prod_{i=1}^{2} (\theta)_{n_i}} \int_0^\infty \int_0^\infty s^{\theta+n_1-1} t^{\theta+n_2-1} \mathrm{e}^{-\psi(s,t)} \prod_{j=1}^{k} g_\nu(n_{j,1}, n_{j,2}; s, t)\mathrm{d}s\mathrm{d}t + a_m$$

as $m \to \infty$, where a_m is such that $\lim_{m\to\infty} = a_m/(\prod_{j=1}^k \alpha(A_{m,j})) = 0$, and we are taking into account that α is a non-atomic probability measure on $(\mathbb{X}, \mathscr{X})$. On the other hand, one can check that the numerator of (B.2) is

$$\frac{\sigma^2 \prod_{j=1}^{k} \alpha(A_{m,j})}{\Gamma^2\left(\frac{\theta}{\sigma}\right) \prod_{i=1}^{2} (\theta)_{n_i}} \int_0^\infty \int_0^\infty s^{\theta+n_1-1} t^{\theta+n_2-1} \mathrm{e}^{-\alpha(A)\psi(s+\lambda_1,t+\lambda_2)-\alpha(A^c)\psi(s,t)}$$
$$\times \prod_{j: A\cap A_{m,j}\neq\varnothing} \int_0^\infty \int_0^\infty x^{n_{j,1}} y^{n_{j,2}} \mathrm{e}^{-(\lambda_1+s)x-(\lambda_2+t)y} \nu(x,y)\mathrm{d}x\mathrm{d}y\mathrm{d}s\mathrm{d}t + a'_m$$

as $m \to \infty$, where a'_m is such that $\lim_{m\to\infty} = a'_m/\left(\prod_{j=1}^k \alpha(A_{m,j})\right) = 0$. When taking the limit as $m \to \infty$ one finds out that for any $A \in \mathscr{X}$

$$\mathbb{E}\left[\mathrm{e}^{-\lambda_1\tilde{\mu}_{1,\sigma,\theta}(A)-\lambda_2\tilde{\mu}_{2,\sigma,\theta}(A)} | \mathrm{D}\right] = \int_{(\mathbb{R}^+)^2} f(s,t|\boldsymbol{D}) \mathrm{e}^{-\alpha(A)[\psi(s+\lambda_1,t+\lambda_2)-\psi(s,t)]}$$
$$\times \prod_{i: Z_i^*\in A} \frac{\int_{(\mathbb{R}^+)^2} \mathrm{e}^{-\lambda_1 x-\lambda_2 y} x^{n_{j,1}} y^{n_{j,2}} \mathrm{e}^{-sx-ty} \nu(x,y)\mathrm{d}x\mathrm{d}y}{g_\nu(n_{j,1}, n_{j,2}; s, t)} \mathrm{d}s\mathrm{d}t.$$

And this yields the representation in (21). □

References

[1] G.M. Constantines, T.H. Savits, A multivariate version of the Faa di Bruno formula, Trans. Amer. Math. Soc. 348 (1996) 503–520.
[2] R. Cont, P. Tankov, Financial Modelling with Jump Processes, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[3] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, vol. I, Springer, New York, 2003.
[4] M. De Iorio, P. Müller, G.L. Rosner, S.N. MacEachern, An ANOVA model for dependent random measures, J. Amer. Statist. Assoc. 99 (2004) 205–215.
[5] D.B. Dunson, Y. Xue, L. Carin, The matrix stick-breaking process: flexible Bayes meta-analysis, J. Amer. Statist. Assoc. 103 (2008) 317–327.
[6] I. Epifani, A. Lijoi, Nonparametric priors for vectors of survival functions, Statist. Sinica (2010) 1455–1484.
[7] T.S. Ferguson, A Bayesian analysis of some nonparametric problems, Ann. Statist. 1 (1973) 209–230.
[8] I.S. Gradshteyn, J.M. Ryzhik, Table of Integrals, Series, and Products, 7th ed., Academic Press, New York, 2007.
[9] H. Ishwaran, L.F. James, Gibbs sampling methods for stick-breaking priors, J. Amer. Statist. Assoc. 96 (2001) 161–173.
[10] H. Ishwaran, M. Zarepour, Series representations for multivariate generalized gamma processes via a scale invariance principle, Statist. Sinica 19 (2009) 1665–1682.
[11] J. Kallsen, P. Tankov, Characterization of dependence of multidimensional Lévy processes using Lévy copulas, J. Multivariate Anal. 97 (2006) 1551–1572.
[12] J.F.C. Kingman, Poisson Processes, Oxford University Press, Oxford, 1993.
[13] A. Lijoi, R.H. Mena, I. Prünster, Controlling the reinforcement in Bayesian nonparametric mixture models, J. R. Stat. Soc. Ser. B 69 (2007) 715–740.
[14] A. Lijoi, I. Prünster, Beyond the Dirichlet process, in: C.C. Holmes, N.L. Hjort, P. Müller, S.G. Walker (Eds.), Bayesian Nonparametrics, Cambridge University Press, Cambridge, 2010, pp. 80–136.
[15] S.N. MacEachern, Dependent nonparametric processes, in: ASA Proceedings of the Section on Bayesian Statistical Science, American Statistical Association, Alexandria, VA, 1999.

[16] J. Pitman, Exchangeable and partially exchangeable random partitions, Probab. Theory Related Fields 102 (1995) 145–158.
[17] J. Pitman, Some developments of the Blackwell–MacQueen URN scheme, in: T.S. Ferguson, L.S. Shapley, J.B. MacQueen (Eds.), Statistics, Probability and Game Theory, in: IMS Lecture Notes Monogr. Ser., vol. 30, Inst. Math. Statist., Hayward, 1996, pp. 245–267.
[18] J. Pitman, Combinatorial Stochastic Processes (Ecole d'Eté de Probabilités de Saint-Flour XXXII 2002), in: Lecture Notes in Mathematics, vol. 1875, Springer, Berlin, 2006.
[19] J. Pitman, M. Yor, The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator, Ann. Probab. 25 (1997) 855–900.
[20] E. Regazzini, V.V. Sazonov, Approximation of distributions of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian statistical inferences, Theory Probab. Appl. 45 (2000) 93–110.
[21] A. Rodríguez, D.B. Dunson, A. Gelfand, The nested Dirichlet process, J. Amer. Statist. Assoc. 103 (2008) 1131–1144.
[22] D.M. Roy, Y.W. Teh, The Mondrian process, in: D. Koller, Y. Bengio, D. Schuurmans, L. Bottou (Eds.), Advances in Neural Information Processing Systems (NIPS), vol. 21, 2009.
[23] E. Sudderth, M.I. Jordan, Shared segmentation of natural scenes using dependent Pitman–Yor processes, in: D. Koller, Y. Bengio, D. Schuurmans, L. Bottou (Eds.), Advances in Neural Information Processing Systems (NIPS), vol. 21, 2009.
[24] Y.W. Teh, M.I. Jordan, M.J. Beal, D.M. Blei, Hierarchical Dirichlet processes, J. Amer. Statist. Assoc. 101 (2006) 1566–1581.