ISSN 2279-9362

Collegio Carlo Alberto

Distribution theory for hierarchical processes

Federico Camerlenghi
Antonio Lijoi
Peter Orbanz
Igor Pruenster

No. 517
December 2017

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

Distribution theory for hierarchical processes*

Federico Camerlenghi[1†], Antonio Lijoi[2†], Peter Orbanz[3] and Igor Prünster[2]

[1] Department of Economics, Management and Statistics, University of Milano–Bicocca, Piazza dell'Ateneo Nuovo 1, 20126 Milano, Italy.

[2] Department of Decision Sciences, BIDSA and IGIER, Bocconi University, via Röntgen 1, 20136 Milano, Italy.

[3] Department of Statistics, Columbia University, 1255 Amsterdam Avenue, 10027 New York, USA.

Abstract

Hierarchies of discrete probability measures are remarkably popular as nonparametric priors in applications, arguably due to two key properties: (i) they naturally represent multiple heterogeneous populations; (ii) they produce ties across populations, resulting in a shrinkage property often described as "sharing of information". In this paper we establish a distribution theory for hierarchical random measures that are generated via normalization, thus encompassing both the hierarchical Dirichlet and hierarchical Pitman–Yor processes. These results provide a probabilistic characterization of the induced (partially exchangeable) partition structure, including the distribution and the asymptotics of the number of partition sets, and a complete posterior characterization. They are obtained by representing hierarchical processes in terms of completely random measures, and by applying a novel technique for deriving the associated distributions. Moreover, they also serve as building blocks for new simulation algorithms, and we derive marginal and conditional algorithms for Bayesian inference.

Keywords: Bayesian Nonparametrics, Distribution theory, Hierarchical processes, Partition structure, Posterior distribution, Prediction, Random measures, Species sampling models.

AMS Classification 2010: 60G57; 62G05; 62F15.

1 Introduction

The random partition structure induced by discrete nonparametric priors plays a pivotal role in a number of inferential problems related to clustering, density estimation, and prediction. It appears in applications such as species sampling, computational linguistics and topic modeling, genomics, and networks. The theory for the exchangeable case is now well understood and extensively studied. See, e.g., [26, 38, 39, 17] for probabilistic investigations and, e.g., [20, 21, 23, 9] for statistical contributions. However, in most applications data are intrinsically heterogeneous and consistent with a dependence assumption more general than exchangeability. Starting from the seminal contributions of MacEachern [34, 35], an extensive literature has been developed to address inferential issues arising with non–exchangeable observations in a Bayesian nonparametric setting. See [10, 43] for reviews. In document analysis, for example, the overall population consists

*A. Lijoi and I. Prünster are supported by the European Research Council (ERC) through StG "N-BNP" 306406.

†Also affiliated to Collegio Carlo Alberto, Piazza Vincenzo Arbarello 8, 10122 Torino, Italy.

of all words in a collection of documents, but each document constitutes a sub-population with its own distribution. Latent Dirichlet Allocation (LDA) was developed in [2] as a simple and effective solution; its enormous popularity is testament to the importance of the problem. The hierarchical Dirichlet process [42] is a natural nonparametric extension. Further contributions in this direction include [16, 43, 45, 37]. In these models, the induced partition structure determines the inferential outcomes but, due to the analytical complexity, its investigation and that of the associated prediction rules have been quite limited; first contributions in this direction, under different dependence assumptions, can be found in [30, 36, 46]. As far as posterior characterizations are concerned, no results are known beyond the hierarchical Dirichlet case [43]. Such characterizations are of theoretical interest, but also a prerequisite for inference algorithms, which simulate draws from (unobserved) random measures conditionally on data. See [5, 18, 31, 46] for examples, and [15] for a comprehensive list of references.

The present paper deals with a general class of hierarchical processes obtained by normalizing random measures, which encompass hierarchical Dirichlet and Pitman-Yor processes. We establish a distribution theory for this class of processes and determine the two distributional quantities essential for Bayesian inference, namely the induced partition structure and a posterior characterization. These allow to perform prediction density estimation, clustering and the assessment of distributional homogeneity across different samples. The focus on a general class of priors rather than on special cases has a two–fold motivation. On the one hand, it helps to clarify the underlying, probabilistic structure of hierarchical models and its statistical implications. On the other hand, the Dirichlet process has well-known limitations in the plain exchangeable framework, and that is similarly true in the non–exchangeable case. In the former, various extensions of the Dirichlet process have been introduced to provide more flexibility; our results provide counterparts in the latter more general framework.

1.1 Partial exchangeability

A random infinite sequence is exchangeable if its distribution is invariant under the group of all finitary permutations (those which permute an arbitrary but finite number of indices of the sequence). It is *partially exchangeable* if invariance holds under a subgroup of such permutations; see [24] for an extensive bibliography. In the problems considered in the following, partial exchangeability arises naturally: if a population decomposes into (conditionally independent) multiple sub-populations that are each exchangeable in their own right, the overall population is partially exchangeable.

More formally, suppose $\mathbb{X}$ is a complete and separable metric space endowed with the Borel σ–field $\mathscr{X}$. Consider d partially exchangeable sequences $\{(X_{i,j})_{j\geq 1} : i = 1,\dots,d\}$ defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $(\mathbb{X}, \mathscr{X})$. By de Finetti's representation theorem this is equivalent to assuming

$$\mathbb{P}\Big[\Big\{\boldsymbol{X}^{(N_i)} \in A_i : \ i = 1,\dots,d\Big\}\Big] = \int_{\mathsf{P}_{\mathbb{X}}^d} \prod_{i=1}^d p_i^{(N_i)}(A_i)\, Q_d(\mathrm{d}p_1,\dots,\mathrm{d}p_d) \tag{1}$$

for any integer $N_i \geq 1$ and $A_i \in \mathscr{X}^{N_i}$, where $\boldsymbol{X}^{(N_i)} = (X_{i,1},\dots,X_{i,N_i})$ and $p^{(q)} = p \times \cdots \times p$ is the q–fold product measure on $\mathbb{X}^q$, for any $q \geq 1$. Moreover, $\mathsf{P}_{\mathbb{X}}$ is the space of all probability measures on $\mathbb{X}$, which we suppose is endowed with the topology of weak convergence and denote as $\mathscr{P}_{\mathbb{X}}$ the corresponding Borel σ–algebra. The mixing or de Finetti measure Q_d is a probability measure on $(\mathsf{P}_{\mathbb{X}}^d, \mathscr{P}_{\mathbb{X}}^d)$ that plays the role of a prior distribution. Hence, (1) amounts to assuming that, given a vector of random probability measures $(\tilde{p}_1, \dots, \tilde{p}_d) \sim Q_d$, the d samples are independent and the observations $\boldsymbol{X}^{(N_i)}$ of the i–th sample are independent and identically distributed from $\tilde{p}_i$.

As in most of the current literature, here we focus on choices of Q_d that select, with probability 1, vectors of discrete probability measures. This implies that there will be ties, with positive probability, within each sample and typically also across different samples. From a modeling perspective this is a desirable feature since it allows clustering both within and across samples or, in other terms, to have models accounting for heterogeneity in a flexible way. The appearance of ties then naturally leads to look at the induced partition structure. In the exchangeable framework, the partition structure is uniquely characterized by the exchangeable partition probability function (EPPF) (see [39]), which is a key tool for studying clustering properties, deriving prediction rules and sampling schemes.

In the partially exchangeable context one can define an analogous object, which we term partially exchangeable partition probability function (pEPPF), and plays exactly the same role of the EPPF in this more general setup. In order to provide a probabilistic description of the pEPPF, let k be the number of distinct values recorded among the $N = N_1 + \cdots + N_d$ observations in $\{\boldsymbol{X}^{(N_i)} : i = 1, \dots, d\}$. Each distinct value identifies a specific cluster of the partition. Accordingly, $\boldsymbol{n}_i = (n_{i,1}, \dots, n_{i,k})$ denotes the vector of frequency counts and $n_{i,j}$ is the number of elements of the i–th sample that coincide with the j–th distinct value. Clearly $n_{i,j} \geq 0$ for any $i = 1, \dots, d$ and $j = 1, \dots, k$, and $\sum_{i=1}^{d} n_{i,j} \geq 1$ for any $j = 1, \dots, k$. Note that $n_{i,j} = 0$ means that the j–th distinct value does not appear in the i–th sample. The j–th distinct value is shared by any two samples i and κ if and only if $n_{i,j} n_{\kappa,j} \geq 1$. To sum up, the pEPPF is defined as

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \cdots, \boldsymbol{n}_d) = \mathbb{E} \int_{\mathbb{X}^k} \prod_{j=1}^{k} \tilde{p}_1^{n_{1,j}}(\mathrm{d}x_j) \, \dots \, \tilde{p}_d^{n_{d,j}}(\mathrm{d}x_j) \tag{2}$$

with the obvious constraint $\sum_{j=1}^{k} n_{i,j} = N_i$, for each $i = 1, \dots, d$.

1.2 Outline

The main goal of the paper is to establish a distribution theory for prior distributions Q_d displaying a hierarchical structure and selecting discrete random probabilities. We focus on two key aspects. On the one hand, we investigate the random partitions induced by an array of partially exchangeable sequences as in (1), including the distribution of the number of partition sets and its asymptotics when the sample size increases. On the other hand, we provide a posterior characterization for a vector of hierarchical random probability measures $(\tilde{p}_1, \dots, \tilde{p}_d)$, conditional on the data. The former allows one to address two relevant issues in Bayesian nonparametric inference, namely inference on the clustering structure of the data and prediction. The latter is crucial for accurate uncertainty quantification and for devising simulation algorithms that generate trajectories of hierarchical random probability measure, from their posterior distribution.

In Section 2 we introduce some basic elements on completely random measures and provide a description of hierarchical normalized random measures. A probabilistic characterization of the induced partially exchangeable random partition is detailed in Section 3 and this forms the basis for investigating the distributional properties of the number of distinct values in d partially exchangeable samples in Section 4. The main results for establishing a posterior representation of $(\tilde{p}_1, \dots, \tilde{p}_d)$ are, then, stated in Section 5. Finally, the computational algorithms that can be obtained from our theoretical results are described in Section 6. Proofs are deferred to the appendix.

2 Hierarchical normalized random measures

In the present work we rely on random measures as the basic building blocks for the construction of discrete nonparametric priors having a hierarchical structure. Let $\mathsf{M}_{\mathbb{X}}$ be the space of boundedly finite measures on $(\mathbb{X}, \mathscr{X})$, i.e. $m(A) < \infty$ for any $m \in \mathsf{M}_{\mathbb{X}}$ and for any bounded set $A \in \mathscr{X}$, equipped with the corresponding Borel σ–algebra $\mathscr{M}_{\mathbb{X}}$. See [8] for details. We consider random elements $\tilde{\mu}$ defined on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and taking values in $(\mathsf{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$. Furthermore, $\tilde{\mu}$ is assumed to be almost surely discrete and without fixed points of discontinuity. Hence, they can be represented as $\tilde{\mu} = \sum_{i\geq 1} J_i \, \delta_{Y_i}$. We shall henceforth focus on random probabilities obtained as suitable transformations of $\tilde{\mu}$. In particular, we will focus on normalization. Indeed, if $0 < \tilde{\mu}(\mathbb{X}) < \infty$ a.s., we define

$$\tilde{p} = \frac{\tilde{\mu}}{\tilde{\mu}(\mathbb{X})} = \sum_{i\geq 1} \frac{J_i}{\bar{J}} \, \delta_{Y_i} \sim \mathrm{NRM}(P) \tag{3}$$

where $\bar{J} := \sum_{i\geq 1} J_i = \tilde{\mu}(\mathbb{X})$ and $P = \mathbb{E}\tilde{p}$ is a probability distribution on $(\mathbb{X}, \mathscr{X})$. In order to obtain a hierarchical structure, one then assumes that $(Y_i)_{i\geq 1}$ in (3) is exchangeable with $Y_i \,|\, \tilde{p}_0 \overset{\text{iid}}{\sim} \tilde{p}_0$. Moreover, $\tilde{p}_0 = \tilde{\mu}_0/\tilde{\mu}_0(\mathbb{X})$ is obtained by normalizing a random measure $\tilde{\mu}_0 = \sum_{i\geq 1} J_{i,0} \, \delta_{Y_{i,0}}$, where $(Y_{i,0})_{i\geq 1}$ is an i.i.d. sequence taking values in $\mathbb{X}$ and whose common probability distribution P_0 is non–atomic. Therefore, we deal with d sequences $\{(X_{i,j})_{j\geq 1} : i = 1, \ldots, d\}$ that are partially exchangeable according to (1) and the mixing measure Q_d is characterized by

$$\begin{aligned} \tilde{p}_i \,|\, \tilde{p}_0 &\overset{\text{iid}}{\sim} \mathrm{NRM}(\tilde{p}_0) \qquad i = 1, \ldots, d \\ \tilde{p}_0 &\sim \mathrm{NRM}(P_0). \end{aligned} \tag{4}$$

The almost sure discreteness of $\tilde{\mu}$ is clearly inherited by the $\tilde{p}_i$'s and hence, as desired, we have nonparametric priors Q_d selecting discrete distributions and inducing ties within and across the samples $\boldsymbol{X}^{(N_1)}, \ldots, \boldsymbol{X}^{(N_d)}$.

The following subsections focus on two specifications of $(\tilde{\mu}, \tilde{\mu}_0)$, and hence of (4), that will be thoroughly investigated in the paper.

2.1 Hierarchical NRMIs

A first natural choice is to set $\tilde{\mu}$ as a *completely random measure* (CRM), i.e. a random element taking values in $\mathsf{M}_{\mathbb{X}}$ such that for any collection of pairwise disjoint sets $A_1, \ldots, A_k$ in $\mathscr{X}$, and for any $k \geq 1$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_k)$ are mutually independent. See [25]. An appealing feature of CRMs is the availability of their Laplace functional. Indeed, if it is further assumed that $\tilde{\mu}$ does not have fixed points of discontinuity, for any measurable function $f : \mathbb{X} \to \mathbb{R}^+$ one has

$$\mathbb{E}\, \mathrm{e}^{-\int_{\mathbb{X}} f(x)\, \tilde{\mu}(\mathrm{d}x)} = \mathrm{e}^{-\int_{\mathbb{R}^+\times\mathbb{X}} \left[1-\mathrm{e}^{-sf(x)}\right] \nu(\mathrm{d}s, \mathrm{d}x)} \tag{5}$$

where ν is the Lévy intensity uniquely characterizing the CRM $\tilde{\mu}$. See [25, 27] for an exhaustive account. Though the treatment can be developed for any CRM, for the ease of illustration henceforth we consider the case where the jumps J_i's and the locations Y_i's are independent and specifically that

$$\nu(\mathrm{d}s, \mathrm{d}x) = \rho(s)\, \mathrm{d}s\, c\, P_0(\mathrm{d}x) \tag{6}$$

for some measurable function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$, constant $c > 0$ and probability measure P_0 on $(\mathbb{X}, \mathscr{X})$. Noteworthy examples are the gamma process and the σ–stable process, which correspond to CRMs having $\rho(s) = s^{-1}\,\mathrm{e}^{-s}$ and $\rho(s) = \sigma s^{-1-\sigma}/\Gamma(1-\sigma)$, for some $\sigma \in (0,1)$. If $\tilde{p} = \tilde{\mu}/\tilde{\mu}(\mathbb{X})$ we use the notation

$$\tilde{p} \sim \mathrm{NRMI}(\rho, c, P_0),$$

which recalls the acronym of [41], where normalized random measures have first been introduced and studied in the exchangeable framework. The corresponding hierarchical model in (4) is thus termed *hierarchical NRMI.*

For hierarchical NRMIs one can evaluate the correlation between $\tilde{p}_i(A)$ and $\tilde{p}_j(A)$, for any $i \neq j$ and measurable subset A of $\mathbb{X}$, in terms of the underlying parameters (c, ρ, c_0, ρ_0). In order to ease the statement of the result, set $\psi(u) = \int_0^\infty [1 - \mathrm{e}^{-us}]\rho(s)\mathrm{d}s$ and $\psi_0(u) = \int_0^\infty [1 - \mathrm{e}^{-us}]\rho_0(s)\mathrm{d}s$ as the Laplace exponents corresponding to $\tilde{p}$ and $\tilde{p}_0$, respectively.

Theorem 1. *Suppose that* $\tilde{p}_i \,|\, \tilde{p}_0 \overset{\text{iid}}{\sim} \mathrm{NRMI}(\rho, c, \tilde{p}_0)$*, for* $i = 1, \ldots, d$*, and* $\tilde{p}_0 \sim \mathrm{NRMI}(\rho_0, c_0, P_0)$*. Then, for any* $A \in \mathscr{X}$ *and* $i \neq j$

$$\mathrm{corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = \Big\{1 + c_0\, c\, \frac{\int_0^\infty u\,\mathrm{e}^{-c\psi(u)}\,\tau_2(u)\,\mathrm{d}u\, \int_0^\infty u\,\mathrm{e}^{-c_0\psi_0(u)}\,\tau_{1,0}^2(u)\,\mathrm{d}u}{\int_0^\infty u\,\mathrm{e}^{-c_0\psi_0(u)}\,\tau_{2,0}(u)\,\mathrm{d}u}\Big\}^{-1}, \tag{7}$$

where $\tau_q(u) = \int_0^\infty s^q\,\mathrm{e}^{-us}\,\rho(s)\,\mathrm{d}s$ *and* $\tau_{q,0}(u) = \int_0^\infty s^q\,\mathrm{e}^{-us}\,\rho_0(s)\,\mathrm{d}s$.

It is worth stressing two important facts. The correlation coefficient between $\tilde{p}_i(A)$ and $\tilde{p}_j(A)$ is always positive. It does not depend on the specific set A. Moreover, by specifying (c, ρ, c_0, ρ_0) the correlation coefficient (7) becomes readily available as shown in the following examples.

Example 1. If $\rho(s) = \rho_0(s) = s^{-1}\,\mathrm{e}^{-s}$, then $\tilde{p}_0$ is a Dirichlet process and the $\tilde{p}_i$'s are, conditionally on $\tilde{p}_0$, independent and identically distributed Dirichlet processes. Hence, $(\tilde{p}_1, \ldots, \tilde{p}_d)$ is a vector of hierarchical Dirichlet processes as in [42]. A straightforward application of Theorem 1 yields

$$\mathrm{corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = \frac{c+1}{c+1+c_0}.$$

Note that the correlation is increasing in c and decreasing in c_0. As $c_0 \uparrow \infty$ the distribution of $\tilde{p}_0$ degenerates on P_0 and the $\tilde{p}_i$'s are independent, which is consistent with $\mathrm{corr}(\tilde{p}_i(A), \tilde{p}_j(A))$ converging to 0. On the other hand, if $c \uparrow \infty$, then the distribution of each $\tilde{p}_i$, conditional on $\tilde{p}_0$, degenerates on $\tilde{p}_0$ and it is, thus, not surprising that the correlation coefficient between any pair of $\tilde{p}_i(A)$'s converges to 1, for any A in $\mathscr{X}$.

Example 2. The hierarchical stable NRMI arises by setting $\rho(s) = \sigma\, s^{-1-\sigma}/\Gamma(1-\sigma)$ and $\rho_0(s) = \sigma_0\, s^{-1-\sigma_0}/\Gamma(1-\sigma_0)$, for some σ and σ_0 in $(0,1)$. This implies that $\tilde{p}_0$ is a σ_0–stable NRMI and, conditionally on $\tilde{p}_0$, the $\tilde{p}_i$'s are independent and identically distributed σ–stable NRMIs. We will say that $(\tilde{p}_1, \ldots, \tilde{p}_d)$ is a vector of hierarchical stable NRMIs. A plain application of Theorem 1 leads to

$$\mathrm{corr}(\tilde{p}_i(A), \tilde{p}_j(A)) = \frac{1-\sigma_0}{1-\sigma\sigma_0},$$

which is increasing in σ and decreasing in σ_0. Due to the properties of the stable CRM, unsurprisingly the correlation coefficient does not depend on the total masses c_0 and c.

2.2 Hierarchical Pitman–Yor processes

The second relevant construction arises when $\tilde{\mu}$ has a distribution obtained by a suitable transformation of the distribution of a CRM. In particular, let $\mathbb{P}_\sigma$ be the probability distribution on

$(\mathsf{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ of a σ–stable CRM, with $\sigma \in (0,1)$. For $\theta > 0$ define $\mathbb{P}_{\sigma,\theta}$ on $(\mathsf{M}_{\mathbb{X}}, \mathscr{M}_{\mathbb{X}})$ as absolutely continuous w.r.t. $\mathbb{P}_\sigma$ and such that its Radon–Nikodym derivative is

$$\frac{\mathrm{d}\mathbb{P}_{\sigma,\theta}}{\mathrm{d}\mathbb{P}_\sigma}(m) = \frac{\sigma\,\Gamma(\theta)}{\Gamma(\theta/\sigma)}\, m^{-\theta}(\mathbb{X}) \tag{8}$$

The resulting random measure $\tilde\mu_{\sigma,\theta}$ with distribution $\mathbb{P}_{\sigma,\theta}$ is not completely random. Nonetheless, via normalization

$$\tilde p = \frac{\tilde\mu_{\sigma,\theta}}{\tilde\mu_{\sigma,\theta}(\mathbb{X})} \sim \mathrm{PY}(\sigma,\theta;P) \tag{9}$$

one obtains a fundamental process, the *Pitman–Yor process* or *two–parameter Poisson–Dirichlet process*. A different equivalent construction, simpler but less convenient for our purposes, starts from a specific $\mathrm{NRMI}(\rho, c, P_0)$ and puts a gamma prior on the parameter c. See [40] for details on both derivations.

The following results provides the correlation structure for the hierarchical Pitman–Yor process and nicely describes the role of the parameters $(\sigma, \sigma_0, \theta, \theta_0)$.

Theorem 2. *Suppose that* $\tilde p_i \,|\, \tilde p_0 \stackrel{\text{iid}}{\sim} \mathrm{PY}(\sigma,\theta,\tilde p_0)$, *for* $i = 1,\dots,d$, *and* $\tilde p_0 \sim \mathrm{PY}(\sigma_0,\theta_0,P_0)$. *Then, for any* $A \in \mathscr{X}$ *and* $i \neq j$

$$\mathrm{corr}(\tilde p_i(A), \tilde p_j(A)) = \left\{1 + \frac{1-\sigma}{1-\sigma_0}\,\frac{\theta_0+\sigma_0}{\theta+1}\right\}^{-1} \tag{10}$$

Unsurprisingly, also for hierarchical Pitman–Yor processes the correlation between $\tilde p_i(A)$ and $\tilde p_j(A)$, for any $i \neq j$, is positive and does not depend on $A \in \mathscr{X}$. Moreover, from (10) the impact of $(\theta_0, \sigma_0, \theta, \sigma)$ on $\mathrm{corr}(\tilde p_i(A), \tilde p_j(A))$ can be easily deduced.

3 Random partitions induced by hierarchical NRMs

Consider an array of d partially exchangeable sequences with de Finetti measure Q_d given by hierarchies of normalized measures as in (4). As already mentioned, the discreteness of the $\tilde p_i$'s and $\tilde p_0$ entails that $\mathbb{P}[X_{\ell,i} = X_{\kappa,j}] > 0$ for any ℓ and κ, i.e. there is a positive probability of ties both within each sample and across the different samples $\boldsymbol{X}^{(N_i)} = (X_{i,1},\dots,X_{i,N_i})$. A random partition of the samples is, thus, induced, whereby any two elements $X_{\ell,i}$ and $X_{\kappa,j}$ are in the same partition group (or cluster) if and only if they take on the same value. Its probability distribution is identified by the pEPPF $\Pi_k^{(N)}$ in (2). Here we determine a closed form expression for hierarchical NRMIs and the hierarchical Pitman–Yor process.

We first focus on hierarchical NRMIs. In order to gain some intuition on the structure of $\Pi_k^{(N)}$, it is worth recalling the so–called *Chinese restaurant franchise* metaphor described in [42] for the hierarchical Dirichlet process. According to this scheme, a franchise of d restaurants shares the same menu, which includes an infinite number of dishes and is generated by the top level base measure P_0. Each restaurant has infinitely many tables. The first customer sitting at each table of restaurant i chooses the dish and this dish is shared by all other customers who afterwards join the same table. In contrast to the well–known Chinese restaurant process, the same dish can be served at different tables within the same restaurant and across different restaurants. According to this scheme, $X_{i,j}$ represents the dish served in the i–th restaurant to the j–th customer for $j = 1,\dots,N_i$ and $i = 1,\dots,d$. Furthermore, the frequency $n_{i,j}$ in (2) is the number of customers in restaurant i eating the j–th dish and we further let $\ell_{i,j} \in \{1,\dots,n_{i,j}\}$ be the number of tables in restaurant i at which the j–th dish is served, if $n_{i,j} \geq 1$. When $n_{i,j} = 0$ it is obvious that $\ell_{i,j} = 0$ as well. Hence

$$\bar{\ell}_{\bullet j} = \sum_{i=1}^{d} \ell_{i,j}, \quad \bar{\ell}_{i\bullet} = \sum_{j=1}^{k} \ell_{i,j}.$$

denote, respectively, the number of tables serving dish j (across restaurants) and the number of tables in restaurant i (regardless of the served dishes). Moreover, if we further label the tables, with $q_{i,j,t}$ we can identify the number of customers in restaurant i eating dish j at table t so that $\sum_{t=1}^{\ell_{i,j}} q_{i,j,t} = n_{i,j}$. This additional notation suggests we are going to consider a combinatorial structure arising from the composition of random partitions acting at different levels of the hierarchy: one yields a partition where the $N = N_1 + \cdots + N_d$ customers are allocated to $|\boldsymbol{\ell}| = \sum_{i=1}^{d} \sum_{j=1}^{k} \ell_{i,j} \geq k$ tables and these tables are, then, clustered into k groups, with each group being identified by a different distinct dish.

Before providing the pEPPF, we introduce the notation that identifies the composing random partitions. If $\tilde{p}_0 \sim \mathrm{NRMI}(\rho_0, c_0, P_0)$ and P_0 is a diffuse probability measure on $\mathbb{X}$, for any $k \in \{1, \dots, n\}$ and any vector of positive integers $(r_1, \dots, r_k)$ such that $\sum_{i=1}^{k} r_i = n$, we set

$$\Phi_{k,0}^{(n)}(r_1, \dots, r_k) = \frac{c_0^k}{\Gamma(n)} \int_0^\infty u^{n-1} \, \mathrm{e}^{-c_0 \psi_0(u)} \prod_{j=1}^{k} \tau_{r_j,0}(u) \, \mathrm{d}u. \tag{11}$$

Note that according to [23, Proposition 3], $\Phi_{k,0}^{(n)}$ is the EPPF induced by an exchangeable sequence drawn from a NRMI with parameter (c_0, ρ_0).

Theorem 3. *Suppose the sequences $\{(X_{i,j})_{j\geq 1} : i = 1, \dots, d\}$ are partially exchangeable according to (1), with Q_d characterized by*

$$\tilde{p}_i \,|\, \tilde{p}_0 \overset{\text{iid}}{\sim} \mathrm{NRMI}(\rho, c, \tilde{p}_0) \quad (i = 1, \dots, d), \qquad \tilde{p}_0 \sim \mathrm{NRMI}(\rho_0, c_0, P_0).$$

Then

$$\Pi_k^{(N)}(\boldsymbol{n}_1, \cdots, \boldsymbol{n}_d) = \sum_{\boldsymbol{\ell}} \sum_{\boldsymbol{q}} \Phi_{k,0}^{(|\boldsymbol{\ell}|)}(\bar{\ell}_{\bullet 1}, \cdots, \bar{\ell}_{\bullet k}) \times \prod_{i=1}^{d} \prod_{j=1}^{k} \frac{1}{\ell_{i,j}!} \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,\ell_{i,j}}} \Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k}) \tag{12}$$

where, if $n_{i,j} \geq 1$, $\boldsymbol{q}_{i,j} = (q_{i,j,1}, \dots, q_{i,j,\ell_{i,j}})$ is a vector of positive integers such that $|\boldsymbol{q}_{i,j}| = n_{i,j}$, for any $i = 1, \dots, d$ and $j = 1, \dots, k$, and

$$\Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k}) = \frac{c^{\bar{\ell}_{i\bullet}}}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} e^{-c\psi(u)} \prod_{j=1}^{k} \prod_{t=1}^{\ell_{i,j}} \tau_{q_{i,j,t}}(u) \mathrm{d}u. \tag{13}$$

Note that, if $n_{i,j} = 0$, then $\boldsymbol{q}_{i,j} = (0, \dots, 0)$ and

$$\Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k}) = \Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,j-1}, \boldsymbol{q}_{i,j+1}, \dots \boldsymbol{q}_{i,k})$$

The backbone of (12) is

$$\Phi_{k,0}^{(|\boldsymbol{\ell}|)}(\bar{\ell}_{\bullet 1}, \cdots, \bar{\ell}_{\bullet k}) \prod_{i=1}^{d} \Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \dots, \boldsymbol{q}_{i,k}) \tag{14}$$

which displays the random partitions' composition acting at the two levels of the hierarchy: the single samples (or restaurants) and the whole collection of samples (or the franchise). The former is captured by $\prod_{i=1}^{d}\Phi^{(N_i)}_{\bar{\ell}_{i\bullet},i}$ while the latter is identified by $\Phi^{(|\boldsymbol{\ell}|)}_{k,0}$. The resulting expression of $\Pi^{(N)}_k$ then follows from plain marginalization.

We now illustrate the result by considering again the hierarchical Dirichlet process and the hierarchical stable NRMI.

Example 3. Let $(\tilde{p}_1,\dots,\tilde{p}_d)$ be a vector of hierarchical Dirichlet processes as in Example 1. Let $(a)_n=\Gamma(a+n)/\Gamma(a)$ be the ascending factorial and $|\mathfrak{s}(n,k)|$ the signless Stirling number of the first kind. It is then straightforward to show that

$$\begin{aligned}\Pi^{(N)}_k(\boldsymbol{n}_1,\dots,\boldsymbol{n}_d)&=\frac{c_0^k}{\prod_{i=1}^d (c)_{N_i}}\sum_{\boldsymbol{\ell}}\frac{c^{|\boldsymbol{\ell}|}}{(c_0)_{|\boldsymbol{\ell}|}}\prod_{j=1}^{k}(\bar{\ell}_{\bullet j}-1)!\prod_{i=1}^{d}|\mathfrak{s}(n_{i,j},\ell_{i,j})|\\&=c_0^k\left(\prod_{i=1}^{d}\frac{\prod_{j=1}^k (c)_{n_{i,j}}}{(c)_{N_i}}\right)\sum_{\boldsymbol{\ell}}\frac{1}{(c_0)_{|\boldsymbol{\ell}|}}\prod_{j=1}^{k}(\bar{\ell}_{\bullet j}-1)!\prod_{i=1}^{d}\mathbb{P}[K_{n_{i,j}}=\ell_{i,j}]\end{aligned}$$

where $K_{n_{i,j}}$ is a random variable denoting the number of distinct observations, out of $n_{i,j}$ drawn from an exchangeable sequence whose de Finetti measure is a Dirichlet process with concentration parameter c. Alternatively, one can rely on properties of $|\mathfrak{s}(n,k)|$ and deduce the following integral representation

$$\Pi^{(N)}_k(\boldsymbol{n}_1,\dots,\boldsymbol{n}_d)=\frac{c_0^k\,c^{\xi}}{\prod_{i=1}^d (c)_{N_i}}\int_{\Delta_k}D_k(\mathrm{d}\boldsymbol{p}\,;\xi_1,\dots,\xi_k,c_0)\prod_{i=1}^{d}\prod_{\{j:\,n_{i,j}\ge 1\}}(c\,p_j+1)_{n_{i,j}-1}$$

where $\xi_j=\sum_{i=1}^d \mathbb{1}_{\{1,2,\dots\}}(n_{i,j})$ is the number of restaurants sharing the j–th dish, $\xi=\sum_{j=1}^k\xi_j$ and $D_k(\,\cdot\,;a_1,\dots,a_{k+1})$ is the multivariate Dirichlet distribution on the k–dimensional simplex Δ_k, with parameters $(a_1,\dots,a_{k+1})$.

Example 4. Let $(\tilde{p}_1,\dots,\tilde{p}_d)$ be a vector of hierarchical stable NRMIs defined as in Example 2 and $\mathscr{C}(n,k;\sigma)$ be the generalized factorial coefficients defined by

$$(\sigma t)_n=\sum_{k=1}^{n}\mathscr{C}(n,k;\sigma)\,(t)_k.$$

As for the pEPPF, Theorem 3 and some algebra lead to

$$\begin{aligned}\Pi^{(N)}_k(\boldsymbol{n}_1,\dots,\boldsymbol{n}_d)&=\frac{\sigma_0^{k-1}\,\Gamma(k)}{\prod_{i=1}^d\Gamma(N_i)}\sum_{\boldsymbol{\ell}}\frac{\sigma^{|\boldsymbol{\ell}|-d}\prod_{i=1}^d\Gamma(\bar{\ell}_{i\bullet})}{\Gamma(|\boldsymbol{\ell}|)}\prod_{j=1}^{k}(1-\sigma_0)_{\bar{\ell}_{\bullet j}-1}\prod_{i=1}^{d}\prod_{j=1}^{k}\frac{\mathscr{C}(n_{i,j},\ell_{i,j};\sigma)}{\sigma^{\ell_{i,j}}}\\&=\sigma_0^{k-1}\,\sigma^{\xi-d}\prod_{i=1}^{d}\left(\frac{\prod_{\{j:\,n_{i,j}\ge1\}}\Gamma(n_{i,j})}{\Gamma(N_i)}\right)\sum_{\boldsymbol{\ell}}\frac{\prod_{j=1}^k(1-\sigma_0)_{\bar{\ell}_{\bullet j}-1}}{\Gamma(|\boldsymbol{\ell}|)}\\&\quad\times\prod_{i=1}^{d}\left(\frac{\Gamma(\bar{\ell}_{i\bullet})}{\prod_{\{j:\,n_{i,j}\ge1\}}\Gamma(\ell_{i,j})}\right)\prod_{i=1}^{d}\prod_{j=1}^{k}\mathbb{P}[K_{n_{i,j}}=\ell_{i,j}]\end{aligned}$$

with $K_{n_{i,j}}$ denoting the number of distinct observations generated by $n_{i,j}$ observations from an exchangeable sequence whose de Finetti measure is a normalized σ–stable process.

The combinatorial structure yielding the pEPPF in (12) is not specific to hierarchical NRMIs. Indeed, it can be established also for the Pitman–Yor process, which arises as the normalization of a measure that is not completely random.

Theorem 4. *Let $\{(X_{i,j})_{j\geq 1}:\, i=1,\ldots,d\}$ be partially exchangeable as in (1), with Q_d characterized by*

$$\tilde p_i \,|\, \tilde p_0 \overset{\text{iid}}{\sim} \mathrm{PY}(\sigma,\theta;\tilde p_0) \quad (i=1,\ldots,d), \qquad \tilde p_0 \sim \mathrm{PY}(\sigma_0,\theta_0;P_0)$$

Then

$$\Pi_k^{(N)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d)=\sum_{\boldsymbol{\ell}} \frac{\prod_{r=1}^{k-1}(\theta_0+r\sigma_0)}{(\theta_0+1)_{|\boldsymbol{\ell}|-1}} \prod_{j=1}^{k}(1-\sigma_0)_{\bar\ell_{\bullet j}-1} \\ \times \prod_{i=1}^{d}\frac{\prod_{r=1}^{\bar\ell_{i\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_i-1}} \prod_{j=1}^{k}\frac{\mathscr{C}(n_{i,j},\ell_{i,j};\sigma)}{\sigma^{\ell_{i,j}}} \tag{15}$$

This result is related to the findings in [16], whose construction leads to a tree structure used as a language model. In models of this type, termed *sequence memoizer*, the observations take values in the space Σ^* of finite sequences of elements from a countable (typically finite) set symbols Σ. Each random probability measure involved in the hierarchies of the model is supported by Σ and it is, then, apparent that the base measure at the root of the hierarchy is atomic. Our treatment is different, in the sense that the state space coincides with any separable and complete metric space $\mathbb{X}$ and the probability distribution at the root of the hierarchy is diffuse. The latter is crucial for obtaining the expressions of the pEPPF displayed in this paper.

4 Distribution of the number of clusters K_N

Having determined the pEPPF of hierarchical NRMIs and hierarchical Pitman–Yor processes, a natural issue to address is the determination of the probability distribution of the number K_N of distinct values out of $N=N_1+\cdots+N_d$ partially exchangeable observations. This can be achieved by relying on the composition of random partitions in the pEPPF representations in Theorems 3 and 4 and highlighted in (14). For the derivation of the result, it is useful to introduce a collection of sequences of latent random variables $\{(T_{i,j})_{j\geq 1}:\, i=1,\ldots,d\}$. They are such that $T_{i,j}|\tilde q_i \overset{\text{iid}}{\sim} \tilde q_i$, with $\tilde q_i \overset{\text{iid}}{\sim} \mathrm{NRMI}(c,\rho,G)$ for hierarchical NRMIs and $\tilde q_i \overset{\text{iid}}{\sim} \mathrm{PY}(\sigma,\theta,G)$ for hierarchical Pitman–Yor processes, while G is some diffuse probability measure. In terms of the Chinese restaurant franchise metaphor, $T_{i,j}$ is the label of the table where the j–th customer of the i–th restaurant is seated. In view of this, the probability distribution of K_N arises by considering:

(i) independent random variables K'_{i,N_i} that equal, for each $i=1,\ldots,d$, the number of distinct values in $\boldsymbol{T}^{(N_i)}=(T_{i,1},\ldots,T_{i,N_i})$;

(ii) $K_{0,t}$, which represents the number of distinct values out of t exchangeable random elements generated from $\tilde p_0$.

According to the Chinese restaurant metaphor, K'_{i,N_i} is the number of tables where the N_i customers of restaurant i are seated, while $K_{0,t}$ is the number of distinct dishes allocated to the t tables where the N customers of the whole franchise are seated.

Theorem 5. *Suppose K_N is the number of distinct values in the d partially exchangeable samples $\{\boldsymbol{X}^{(N_i)} : i = 1, \dots, d\}$ governed by a vector of hierarchical NRMIs, i.e. $\tilde{p}_i | \tilde{p}_0 \stackrel{\text{iid}}{\sim} \mathrm{NRMI}(c, \rho, \tilde{p}_0)$ and $\tilde{p}_0 \sim \mathrm{NRMI}(c_0, \rho_0, P_0)$, with P_0 being non–atomic. Then, for any $k = 1 \dots, N$ one has*

$$\mathbb{P}[K_N = k] = \sum_{t=k}^{N} \mathbb{P}[K_{0,t} = k]\, \mathbb{P}\Big[\sum_{i=1}^{d} K'_{i,N_i} = t\Big] \tag{16}$$

The probability distributions of $K_{0,t}$ and of K'_{i,N_i} are readily derived from their EPPFs and coincide with

$$\mathbb{P}[K_{0,t} = k] = \frac{1}{k!} \sum_{(r_1,\dots,r_k)\in\Delta_{k,t}} \binom{t}{r_1 \cdots r_k} \Phi_{k,0}^{(t)}(r_1, \dots, r_k) \tag{17}$$

for any $k \in \{1, \dots, t\}$, where $\Delta_{j,n} = \{(r_1, \dots, r_j) : r_i \geq 1, \ \sum_{i=1}^{j} r_i = n\}$, and

$$\mathbb{P}[K'_{i,N_i} = \zeta] = \frac{1}{\zeta!} \sum_{(r_1,\dots,r_\zeta)\in\Delta_{\zeta,N_i}} \binom{N_i}{r_1 \cdots r_\zeta} \Phi_{\zeta,i}^{(N_i)}(r_1, \dots, r_\zeta) \tag{18}$$

for any $\zeta \in \{1, \dots, N_i\}$.

A similar result holds for the hierarchical Pitman–Yor process.

Theorem 6. *Suppose K_N is the number of distinct values in the d partially exchangeable samples $\{\boldsymbol{X}^{(N_i)} : i = 1, \dots, d\}$ governed by a vector of hierarchical Pitman–Yor processes, i.e. $\tilde{p}_i | \tilde{p}_0 \stackrel{\text{iid}}{\sim} \mathrm{PY}(\sigma, \theta; \tilde{p}_0)$ and $\tilde{p}_0 \sim \mathrm{PY}(\sigma_0, \theta_0; P_0)$. Then*

$$\begin{aligned}\mathbb{P}[K_N = k] = \sum_{t=k}^{N} &\frac{\prod_{r=1}^{k-1}(\theta_0 + r\sigma_0)}{(\theta_0+1)_{t-1}} \frac{\mathscr{C}(t, k; \sigma_0)}{\sigma_0^k} \\ &\times \sum_{\{(\zeta_1,\dots,\zeta_d)\in\Delta_{d,t}\}} \prod_{i=1}^{d} \frac{\prod_{r=1}^{\zeta_i-1}(\theta + r\sigma)}{(\theta+1)_{N_i-1}} \frac{\mathscr{C}(N_i, \zeta_i; \sigma)}{\sigma^{\zeta_i}}\end{aligned} \tag{19}$$

Remark 1. In the proofs of Theorems 5–6, based on the expressions of the pEPPFs, we give an alternative equivalent representation of K_N: if $\xi(\boldsymbol{N}) = K'_{1,N_1} + \cdots + K'_{d,N_d}$, from (16) and (19) one deduces for both hierarchical NRMIs and Pitman–Yor processes

$$K_N \stackrel{\text{d}}{=} K_{0,\xi(\boldsymbol{N})}.$$

The equality between K_N and $K_{0,\xi(\boldsymbol{N})}$ can be strengthened, and actually holds almost surely. This fact is useful for the determination of the asymptotic behaviour of K_N.

Before establishing the asymptotic behavior of K_N, as $N \to \infty$, introduce two positive sequences $(\lambda_0(n))_{n\geq1}$ and $(\lambda(n))_{n\geq1}$ such that $\lim_n \lambda_0(n) = \lim_n \lambda(n) = \infty$ and assume λ_0 satisfies the following condition:

(H1) for any pair of positive sequences $(b_1(n))_{n\geq1}$ and $(b_2(n))_{n\geq1}$ such that $\lim_n b_1(n) = \lim_n b_2(n) = \infty$ and $\lim_n(b_1(n)/b_2(n)) = 1$

$$\lim_{n\to\infty} \frac{\lambda_0(b_1(n))}{\lambda_0(b_2(n))} = 1$$

We would like to stress that assumption (H1) is satisfied when λ_0 is a *regularly varying function.*

In the sequel we agree that $Y_n \simeq \lambda(n)$, for $n \to \infty$, means that $\lim_n Y_n/\lambda(n)$ almost surely exists and equals a finite and positive random variable, then one can state the following.

Theorem 7. *Suppose K_N is the number of distinct values in the d partially exchangeable samples $\{\boldsymbol{X}^{(N_i)}:\ i=1,\dots,d\}$ governed by a vector of hierarchical NRMIs such that $K_{0,N} \simeq \lambda_0(n)$ and $K'_{i,N} \simeq \lambda(N)$ as $N \to \infty$, where $(\lambda_0(n))_{n\geq 1}$ satisfies* (H1)*. Moreover, let $N_1 = \cdots = N_d = N^* = N/d$. Then*

$$K_N \simeq \lambda_0(\eta\,\lambda(N/d)) \qquad \text{as } N \to \infty,$$

for some positive and finite random variable η.
In particular, if $(\tilde{p}_1,\dots,\tilde{p}_d)$ is a vector of hierarchical Dirichlet processes, then

$$K_N \simeq \log\log N \qquad \text{as } N \to \infty.$$

Note that the rate of increase of K_N for the hierarchical Dirichlet process has been also displayed in [43] based on a more informal argument. The corresponding result for hierarchical Pitman–Yor process is as follows.

Theorem 8. *Suppose K_N is the number of distinct values in the d partially exchangeable samples $\{\boldsymbol{X}^{(N_i)}:\ i=1,\dots,d\}$ governed by a vector of hierarchical Pitman–Yor processes. Furthermore, let $N_1 = \cdots = N_d = N^* = N/d$. Then*

$$K_N \simeq N^{\sigma\,\sigma_0} \qquad \text{as } N \to \infty$$

Remark 2. These results can be extended to the case where only a subset of the N_i's diverge and the others stay finite. Indeed, if for some $m \leq d$ one has $N_{j_1} = \cdots = N_{j_m} = N^*$, where $N^* \to \infty$, and $N_i < L < \infty$ for any other $i \notin \{j_1,\dots,j_m\}$, then it is possible to conclude that

$$K_N \simeq \lambda_0(\eta\lambda(N/m))$$

as $N^* \to \infty$, which entails $N \to \infty$. This leaves the rates of increase for K_N displayed in Theorems 7–8 unchanged .

Remark 3. With some care the results can be generalized to cover the case of the N_i's diverging at different rates. Indeed, considering the asymptotics as $\max_{1\leq i\leq d} N_i \to \infty$, K_N increases at rates similar to those displayed Theorems 7–8.

5 Posterior characterizations

In order to complete the description of distributional properties of hierarchical processes, it is essential to determine a posterior characterization. To the best of our knowledge, no posterior characterization is available for dependent processes in a partially exchangeable framework, whether constructed in terms of hierarchies or by different means. Hence, our following results are the very first. Despite the theoretical interest, note that, while for prediction the partition probability functions of Theorems 3–4 suffice, inference on non–linear functionals of $(\tilde{p}_1,\dots,\tilde{p}_d)$ requires the posterior distribution of the vector of hierarchical random probabilities.

5.1 Hierarchical NRMI posterior

In the following let $X_1^*,\dots,X_k^*$ denote the distinct observations featured by the whole collection of samples $\boldsymbol{X} = \{\boldsymbol{X}^{(N_i)}:\ i=1,\dots d\}$ and assume U_0 is a positive random variable whose density function, conditional on $\boldsymbol{X}$ and on the latent tables' labels $\boldsymbol{T} = \{\boldsymbol{T}^{(N_i)} : \ i = 1, \dots, d\}$ introduced in Section 4, equals

$$f_0(u|\boldsymbol{X}, \boldsymbol{T}) \propto u^{|\boldsymbol{\ell}|-1} e^{-c_0\psi_0(u)} \prod_{j=1}^{k} \tau_{\bar{\ell}_{\bullet j},0}(u). \tag{20}$$

The posterior characterization is then composed of two blocks, the first concerning the root of the hierarchy in terms of $\tilde{\mu}_0$ and the second concerning the vector of random probabilities.

Theorem 9. *Suppose the data $\boldsymbol{X}$ are partially exchangeable and are modeled as in (4). Then*

$$\tilde{\mu}_0 \,|\, (\boldsymbol{X}, \boldsymbol{T}, U_0) \stackrel{d}{=} \eta_0^* + \sum_{j=1}^{k} I_j \delta_{X_j^*} \tag{21}$$

where the two summands on the right–hand side of the distributional identity are independent and

(i) *η_0^* is a CRM with intensity*

$$\nu_0(\mathrm{d}s, \mathrm{d}x) = \mathrm{e}^{-U_0 s} \rho_0(s) \mathrm{d}s \, c_0 \, P_0(\mathrm{d}x).$$

(ii) *the I_j's are independent and non–negative jumps with density*

$$f_j(s|\boldsymbol{X}, \boldsymbol{T}) \propto s^{\bar{\ell}_{\bullet j}} e^{-sU_0} \rho_0(s)$$

It is worth noting that the posterior of $\tilde{\mu}_0$ depends on sample information across the populations rather than population–specific, most notably the number of different dishes served across restaurants. This clearly serves the purpose of directing the dependence across populations. Theorem 9 allows us then to establish the posterior distribution of a vector $(\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ of hierarchical CRMs, conditional a vector $\boldsymbol{U} = (U_1, \dots, U_d)$ whose components are conditionally independent, given $(\boldsymbol{X}, \boldsymbol{T})$, and with respective densities

$$f_i(u|\boldsymbol{X}, \boldsymbol{T}) \propto u^{N_i - 1} e^{-c\psi(u)} \prod_{j=1}^{k} \prod_{t=1}^{\ell_{i,j}} \tau_{q_{i,j,t}}(u) \qquad i = 1, \dots, d. \tag{22}$$

The fundamental posterior characterization, where population–specific characteristics come into play, can then be stated as follows.

Theorem 10. *Suppose the data $\boldsymbol{X}$ are partially exchangeable and are modeled as in (4). Then*

$$(\tilde{\mu}_1, \dots, \tilde{\mu}_d)|(\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U}, \tilde{\mu}_0) \stackrel{d}{=} (\tilde{\mu}_1^*, \dots, \tilde{\mu}_d^*) + \Big(\sum_{j=1}^{k} \sum_{t=1}^{\ell_{1,j}} J_{1,j,t} \delta_{X_j^*}, \dots, \sum_{j=1}^{k} \sum_{t=1}^{\ell_{d,j}} J_{d,j,t} \delta_{X_j^*} \Big), \tag{23}$$

where the two summands on the right–hand–side are independent, $\sum_{t=1}^{\ell_{i,j}} J_{i,j,t} \equiv 0$ if $n_{i,j} = 0$ and

(i) *$(\tilde{\mu}_1^*, \dots, \tilde{\mu}_d^*)$ is a vector of hierarchical CRMs and, conditional on $\tilde{\mu}_0^* = \eta_0^* + \sum_{j=1}^{k} I_j \delta_{X_j^*}$ in (21), each $\tilde{\mu}_i^*$ has intensity*

$$\nu_i(\mathrm{d}s, \mathrm{d}x) = e^{-U_i s} \rho(s) \mathrm{d}s \, c \, \tilde{p}_0^*(\mathrm{d}x),$$

with $\tilde{p}_0^* = \tilde{\mu}_0^* / \tilde{\mu}_0^*(\mathbb{X})$;

(ii) *the jumps* $J_{i,j,t}$ *are independent and non–negative random variables whose density equals*

$$f_{i,j,t}(s) \propto e^{-U_i s} s^{q_{i,j,t}} \rho(s).$$

when $n_{i,j} \geq 1$*, whereas* $J_{i,j,t} = 0$*, almost surely, if* $n_{i,j} = 0$.

The expressions involved in the posterior characterization of Theorem 10 are somehow reminiscent of the ones provided in [23] for the exchangeable case. This is due to the fact that, once accounted for the dependence structure inherited from the hierarchical construction, one has exchangeability within each population.

We now illustrate the general results by means of two examples, related to the hierarchical Dirichlet process and the hierarchical stable NRMI.

Example 5. Assume that $\rho(s) = \rho_0(s) = e^{-s}/s$, so we are considering a vector of hierarchical Dirichlet processes. Recall that $\psi(u) = \psi_0(u) = \log(1+u)$ and $\tau_q(u) = \tau_{q,0}(u) = \Gamma(q)/(1+u)^q$. In this case

$$f_0(u) = \frac{\Gamma(|\boldsymbol{\ell}| + c_0)}{\Gamma(|\boldsymbol{\ell}|)\Gamma(c_0)} \frac{u^{|\boldsymbol{\ell}|-1}}{(1+u)^{c_0+|\boldsymbol{\ell}|}} \mathbb{1}_{(0,\infty)}(u)$$

implying that $U_0/(1+U_0) \sim \text{Beta}(|\boldsymbol{\ell}|, c_0)$. In the posterior representation of $\tilde{\mu}_0$ as stated in Theorem 9, one has

(a) η_0^* is a gamma CRM with intensity $\mathrm{e}^{-(1+U_0)s} s^{-1} \, \mathrm{d}s \, c_0 \, P_0(\mathrm{d}x)$,

(b) $I_j \overset{\text{ind}}{\sim} \text{Ga}(\bar{\ell}_{\bullet j}, 1+U_0)$, meaning that its density function is

$$\frac{(1+U_0)^{\bar{\ell}_{\bullet j}}}{\Gamma(\bar{\ell}_{\bullet j})} x^{\bar{\ell}_{\bullet j}-1} e^{-(1+U_0)x} \mathbb{1}_{(0,\infty)}(x)$$

Now, since the normalized distributions of (a) and (b) do not depend on the scale U_0, it follows that

$$\tilde{p}_0^* = \tilde{p}_0 | (\boldsymbol{X}, \boldsymbol{T}) \sim \mathscr{D}(c_0 P_0 + \sum_{j=1}^k \bar{\ell}_{\bullet j} \delta_{X_j^*}).$$

with $\mathscr{D}$ indicating a Dirichlet process. As far as the vector of random probabilities $(\tilde{p}_1, \ldots, \tilde{p}_d)$ is concerned, by Theorem 10 one has that, conditional on $\tilde{p}_0^*$ and on $(\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U})$, the CRMs $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are independent, and the distribution of each $\tilde{\mu}_i$ equals the probability distribution of the random measure $\tilde{\mu}_i^* + \sum_{j=1}^k H_{i,j} \delta_{X_{i,j}^*}$ where

(a') $\tilde{\mu}_i^*$ a gamma CRM having intensity $\mathrm{e}^{-(1+U_i)s} s^{-1} \, \mathrm{d}s \, c \, \tilde{p}_0^*(\mathrm{d}x)$

(b') $H_{i,j} = \sum_{t=1}^{\ell_{i,j}} J_{i,j,t}$, where $J_{i,j,t} \overset{\text{ind}}{\sim} \text{Ga}(q_{i,j,t}, U_i + 1)$, for $t = 1, \ldots, \ell_{i,j}$, thus implying that $H_{i,j} \sim \text{Ga}(n_{i,j}, U_i + 1)$ if $n_{i,j} \geq 1$ and $H_{i,j} = 0$ almost surely if $n_{i,j} = 0$, by virtue of Theorem 10(ii).

Moreover, note that $U_i/(1+U_i) \overset{\text{ind}}{\sim} \text{Beta}(c, N_i)$. Hence, by the same arguments as before, one has

$$\tilde{p}_i | (\boldsymbol{X}, \boldsymbol{T}, \tilde{p}_0^*) \sim \mathscr{D}\Big(c\tilde{p}_0^* + \sum_{j=1}^k n_{i,j} \delta_{X_{i,j}^*}\Big)$$

for $i = 1, \ldots, d$. Note that the dependence on the table configuration $\boldsymbol{T}$ is induced solely by $\tilde{p}_0^*$, arguably a quite restrictive feature. $\square$

Example 6. For a hierarchical stable NRMI one has $\rho(s) = \sigma s^{-1-\sigma}\,\mathrm{d}s/\Gamma(1-\sigma)$, for some $\sigma \in (0,1)$, $\psi(u) = u^\sigma$ and $\tau_q(u) = \sigma(1-\sigma)_{q-1}\,u^{\sigma-q}$. Similar expressions hold true for $\rho_0, \tau_{q,0}$ and ψ_0, with $\sigma_0 \in (0,1)$ replacing σ. It is easily seen that U_0 is such that $U_0^{\sigma_0} \sim \mathrm{Ga}(k, c_0)$ and note that the distribution of U_0 depends on the observations only through k. Moreover

(a) η_0^* is a CRM with intensity

$$\frac{\sigma_0}{\Gamma(1-\sigma_0)}\,\frac{\mathrm{e}^{-U_0 s}}{s^{\sigma_0+1}}\mathrm{d}s\; c_0\, P_0(\mathrm{d}x),$$

which is known as generalized gamma CRM (see, e.g, [29]).

(b) $I_j \overset{\text{ind}}{\sim} \mathrm{Ga}(\bar{\ell}_{\bullet j} - \sigma_0, U_0)$.

Hence $\tilde{p}_0^* = (\eta_0^* + \sum_{j=1}^k I_j\,\delta_{X_j^*})/(\eta_0^*(\mathbb{X}) + \sum_{j=1}^k I_j)$. Conditional on $\tilde{p}_0^*$, and on $(\boldsymbol{X}, \boldsymbol{T}, \boldsymbol{U})$, the CRMs $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ are independent and each $\tilde{\mu}_i$ equals, in distribution, $\tilde{\mu}_i^* + \sum_{j=1}^{k_i} H_{i,j}\delta_{X_{i,j}^*}$, where

(a') $\tilde{\mu}_i^*$ is a generalized gamma CRM whose intensity is

$$\frac{\sigma}{\Gamma(1-\sigma)}\,\frac{\mathrm{e}^{-U_i s}}{s^{\sigma+1}}\,\mathrm{d}s\, c\,\tilde{p}_0^*(\mathrm{d}x);$$

(b') $H_{i,j} := \sum_{t=1}^{\ell_{i,j}} J_{i,j,t}$, where $J_{i,j,t} \overset{\text{ind}}{\sim} \mathrm{Ga}(q_{i,j,t} - \sigma, U_i)$, for $t = 1, \ldots, \ell_{i,j}$, thus implying that $H_{i,j} \overset{\text{ind}}{\sim} \mathrm{Ga}(n_{i,j} - \ell_{i,j}\sigma, U_i)$ if $n_{i,j} \geq 1$, while $H_{i,j} = 0$ almost surely if $n_{i,j} = 0$.

Finally, U_i is such that $U_i^\sigma \sim \mathrm{Ga}(k, c)$. This implies that the posterior distribution of $(\tilde{p}_1, \ldots, \tilde{p}_d)$, conditional on the data and a suitable latent structure, is a vector of normalized generalized gamma CRMs with fixed points of discontinuity at the data points. □

5.2 Hierarchical PY posterior

Even if not obtained through the normalization of a CRM, the techniques used in Theorems 9–10 apply, with suitable modifications, to the determination of a posterior characterization of the Pitman–Yor process. Hence, assume that data $\boldsymbol{X}$ are partially exchangeable as in (1) and the prior Q_d is characterized by

$$\tilde{p}_i\,|\,\tilde{p}_0 \overset{\text{iid}}{\sim} \mathrm{PY}(\sigma, \theta; \tilde{p}_0) \quad (i = 1, \ldots, d), \qquad \tilde{p}_0 \sim \mathrm{PY}(\sigma_0, \theta_0; P_0)$$

where $\tilde{p}_0 = \tilde{\mu}_0/\tilde{\mu}_0(\mathbb{X})$ and $\tilde{p}_i = \tilde{\mu}_i/\tilde{\mu}_i(\mathbb{X})$, for $i = 1, \ldots, d$ and, recall that, in view of (8), here the random measures $\tilde{\mu}_0$ and $\tilde{\mu}_i$ are not completely random. The first step is again the posterior characterization of the root of the hierarchy in terms of $\tilde{\mu}_0$.

Theorem 11. *Let V_0 be such that $V_0^{\sigma_0} \sim Ga(k+\theta_0/\sigma_0, 1)$. Then $\tilde{\mu}_0|(\boldsymbol{X}, \boldsymbol{T}, V_0)$ equals, in distribution, the random measure $\eta_0^* + \sum_{j=1}^k I_j\delta_{X_j^*}$, where η_0^* is a generalized gamma CRM whose intensity is*

$$\frac{\sigma_0}{\Gamma(1-\sigma_0)}\,\frac{\mathrm{e}^{-V_0 s}}{s^{1+\sigma_0}}\,\mathrm{d}s P_0(\mathrm{d}x),$$

the jumps $\{I_j : j = 1, \ldots, k\}$ and η_0^* are independent and $I_j \overset{\text{ind}}{\sim} \mathrm{Ga}(\bar{\ell}_{\bullet j} - \sigma_0, V_0)$, for $j = 1, \ldots, k$.

Given this result, one can establish the following posterior characterization of the vector of random measures $(\tilde\mu_1,\dots,\tilde\mu_d)$ whose normalization yields a vector of hierarchical PY processes.

Theorem 12. *Let V_i be such that $V_i^\sigma \stackrel{\text{ind}}{\sim} Ga(\bar\ell_{i\bullet}+\theta/\sigma,\,1)$, for $i=1,\dots,d$. Then $(\tilde\mu_1,\dots,\tilde\mu_d)|(\boldsymbol{X},\boldsymbol{T},\boldsymbol{V},\tilde p_0^*)$ equals, in distribution, the random measure*

$$(\tilde\mu_1^*,\dots,\tilde\mu_d^*)+\Big(\sum_{j=1}^k H_{1,j}\delta_{X_j^*},\dots,\sum_{j=1}^k H_{d,j}\delta_{X_j^*}\Big)$$

where the two summands in the above expression are independent, $\tilde p_0^*=(\eta_0^*+\sum_{j=1}^k I_j\delta_{X_j^*})/(\eta_0^*(\mathbb{X})+\sum_{j=1}^k I_j)$ and

(i) *$\tilde\mu_1^*,\dots,\tilde\mu_d^*$ are independent and each $\tilde\mu_i^*$ is a generalized gamma CRM with intensity*

$$\frac{\sigma}{\Gamma(1-\sigma)}\,\frac{\mathrm{e}^{-V_i s}}{s^{1+\sigma}}\,\mathrm{d}s\,\tilde p_0^*(\mathrm{d}x)$$

(ii) *$H_{i,j}\stackrel{\text{ind}}{\sim}\mathrm{Ga}(n_{i,j}-\ell_{i,j}\sigma,V_i)$ if $n_{i,j}\ge 1$ and $H_{i,j}=0$, almost surely, if $n_{i,j}=0$.*

From Theorems 11– 12 the posterior distribution of $\tilde p_0$ and of the $\tilde p_i$'s, conditional on $\tilde p_0$, immediately follow. However, given the special features of the PY process, one can further simplify such a representation and discard the dependence on the latent random elements V_0 and $\boldsymbol{V}=(V_1,\dots,V_d)$ leading to a simple posterior representation, which completes the picture of the posterior behaviour of hierarchical PY process. In stating the result, we set $k_i=\mathrm{card}\{j:\,n_{i,j}\ge 1\}$ and agree that the Dirichlet distribution with parameters $(n_{i,1}-\ell_{i,1}\sigma,\dots,n_{i,k}-\ell_{i,k}\sigma,\theta+\bar\ell_{i\bullet}\sigma)$ is on the k_i–dimensional simplex, after removing the parameters having $n_{i,j}=0$.

Theorem 13. *The posterior distribution of $\tilde p_0$, conditional on $(\boldsymbol{X},\boldsymbol{T})$, equals the distribution of the random probability measure*

$$\sum_{j=1}^k W_j\delta_{X_j^*}+W_{k+1}\,\tilde p_{0,k} \tag{24}$$

where $(W_1,\dots,W_k)$ is a k–variate Dirichlet random vector with parameters $(\bar\ell_{\bullet 1}-\sigma_0,\dots,\bar\ell_{\bullet k}-\sigma_0,\theta_0+k\sigma_0)$, $W_{k+1}=1-\sum_{i=1}^k W_i$ and $\tilde p_{0,k}\sim\mathrm{PY}(\sigma_0,\theta_0+k\sigma_0;P_0)$. Moreover, conditional on $(\tilde p_0,\boldsymbol{X},\boldsymbol{T})$, the posterior distribution of each $\tilde p_i^*=(\tilde\mu_i^*+\sum_{j=1}^k H_{i,j}\delta_{X_j^*})/(\tilde\mu_i^*(\mathbb{X})+\sum_{j=1}^k H_{i,j})$ equals the distribution of the random measure

$$\sum_{j=1}^k W_{i,j}\,\delta_{X_j^*}+W_{i,k+1}\,\tilde p_{i,k} \tag{25}$$

where $(W_{i,1},\dots,W_{i,k})$ is a k–variate Dirichlet random vector with parameters $(n_{i,1}-\ell_{i,1}\sigma,\dots,n_{i,k}-\ell_{i,k}\sigma,\theta+\bar\ell_{i\bullet}\sigma)$, $W_{i,k+1}=1-\sum_{j=1}^k W_{i,j}$ and $\tilde p_{i,k}\,|\,\tilde p_0\stackrel{\text{ind}}{\sim}\mathrm{PY}(\sigma,\theta+\bar\ell_{i\bullet}\sigma;\tilde p_0)$.

As previously mentioned, in (25) one has $\mathbb{P}[W_{i,j}=0]=1$ whenever $n_{i,j}=0$ and the distribution of $(W_{i,1},\dots,W_{i,k})$ degenerates on a lower–dimensional simplex. Both representations (24) and (25) are reminiscent of the one given in the exchangeable case by [38]. The common thread is the so–called *quasi-conjugacy* property characteristic of the PY process. See [33].

6 Algorithms

The theoretical findings in Sections 3 and 5 are essential for deriving, respectively, marginal and conditional sampling schemes. Note that, based on the pEPPFs provided in Theorems 3–4, one can derive the predictive distributions associated to hierarchical normalized random measures. However, the analytical complexity inherent to the hierarchical construction does not allow to deduce closed form expressions. Therefore, the best route for a concrete implementation is represented by the derivation of suitable sampling schemes. In Section 6.1 we state the marginal sampler arising from the pEPPF in the context of prediction problems, when $\tilde{p}_1, \ldots, \tilde{p}_d$ model directly the data and one is interested in specific features of additional samples $(X_{i,N_i+1}, \ldots, X_{i,N_i+m})$, conditional on $\boldsymbol{X}^{(N_i)} = (X_{i,1}, \ldots, X_{i,N_i})$, for $i = 1, \ldots, d$. The algorithm can be adapted in a straightforward way to mixture models with $\tilde{p}_1, \ldots, \tilde{p}_d$ modeling latent random variables in dependent mixtures. Finally, in Section 6.2 we devise a conditional algorithm, which allows to simulate the trajectories of $(\tilde{p}_1, \ldots, \tilde{p}_d)$ from its posterior distribution. These posterior trajectories can then be immediately used for prediction and mixture modeling.

6.1 Blackwell–MacQueen urn scheme

The pEPPFs established in Theorems 3–4 arise upon marginalizing out the hierarchical random probability measures and naturally lend themselves to be used for addressing predictive inferential issues. To be more specific, conditional on observed data $\boldsymbol{X}^{(N_i)}$, we aim at determining the probability distribution of the m_i additional outcomes for each population $i = 1, \ldots, d$

$$\begin{aligned}\mathbb{P}[\cap_{i=1}^d \{\boldsymbol{X}^{(m_i|N_i)} \in A_i\} | \boldsymbol{X}^{(N_1)}, \ldots, \boldsymbol{X}^{(N_d)}]& \\ = \int_{\mathsf{P}_{\mathbb{X}}} \prod_{i=1}^d p_i^{(m_i)}(A_i)\, Q_d(\mathrm{d}p_1, \ldots, \mathrm{d}p_d | \boldsymbol{X}^{(N_1)}, \ldots, \boldsymbol{X}^{(N_d)})& \end{aligned} \tag{26}$$

where $\boldsymbol{X}^{(m_i|N_i)} = (X_{i,N_i+1}, \ldots, X_{i,N_i+m_i})$ and $A_i \in \mathscr{X}^{m_i}$. Based on (26), one can predict specific features of $\boldsymbol{X}^{(m_i|N_i)}$, for $i = 1, \ldots, d$, such as, e.g., the number of new distinct values in the additional m_i sample data or the number of distinct values that have appeared r times in the observed sample $\boldsymbol{X}^{(N_i)}$ that will be recorded in $\boldsymbol{X}^{(m_i|N_i)}$. These, and a number of related problems, have been extensively studied in the exchangeable case in view of species sampling applications where such quantities can be seen as measures of species diversity. See, e.g., [12, 28]. The results of this paper allow to cover also the more realistic partially exchangeable case for the first time.

The direct evaluation of (26) is unfeasible and one needs to resort to some simulation scheme. To this end, one may rely on the pEPPF in (12)–(15) to devise a Blackwell–MacQueen urn scheme, for any $d \geq 2$, that generates $\boldsymbol{X}^{(m_i|N_i)}$ for any hierarchical NRMI. In order to simplify the notation and the description of the algorithm, we consider the case $d = 2$. The goal is to generate samples $X_{1,N_1+1}, \ldots, X_{1,N_1+m_1}$ and $X_{2,N_2+1}, \ldots, X_{2,N_2+m_2}$, conditional on $\boldsymbol{X}^{(N_1)}$ and $\boldsymbol{X}^{(N_2)}$, for any two positive integers m_1 and m_2. One needs to introduce $N_1 + m_1 + N_2 + m_2$ latent variables $T_{1,1}, \ldots, T_{1,N_1+m_1}$, $T_{2,1}, \ldots, T_{2,N_2+m_2}$, which are the labels identifying the tables at which the different costumers are seated in the restaurants. The determination of the full conditionals follows immediately from Theorems 3–4 and, more specifically, (14). The sampler allows one to generate $(T_{i,1}, \ldots, T_{i,N_i})$ and $(X_{i,N_i+r}, T_{i,N_i+r})$, for $r = 1, \ldots, m_i$ and $i = 1, 2$. In order to provide details on this, the label $-r$ is used to identify a quantity determined after removing r–th element. Hence, for each $i = 1, 2$, one has

(1) At $t=0$, start from an initial configuration $X^{(0)}_{l,N_l+1},\dots,X^{(0)}_{l,N_l+m_l}$ and $T^{(0)}_{l,1},\dots,T^{(0)}_{l,N_l+m_i}$, for $l=1,2$.

(2) At iteration $t\ge 1$

(2.a) With $X_{i,r}=X^*_h$ generate latent variables $T^{(t)}_{i,r}$, for $r=1,\dots,N_i$, from

$$\mathbb{P}(T_{i,r}=\text{“new”}\,|\cdots)=w_{h,r}\,\frac{\Phi^{(N_i)}_{\bar\ell_{i\bullet}^{-r}+1,i}(\boldsymbol{q}^{-r}_{i,1},\dots,(\boldsymbol{q}^{-r}_{i,h},1),\dots,\boldsymbol{q}^{-r}_{i,k})}{\Phi^{(N_i-1)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{1,1},\dots,\boldsymbol{q}^{-r}_{1,h},\dots,\boldsymbol{q}^{-r}_{i,k})}$$

and, for $\kappa=1,\dots,\ell^{-r}_{i,h}$,

$$\mathbb{P}(T_{i,r}=T^{*,-r}_{i,h,\kappa}\,|\cdots)=\frac{\Phi^{(N_i)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,h}+\boldsymbol{1}_\kappa,\dots,\boldsymbol{q}^{-r}_{i,k})}{\Phi^{(N_i-1)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,h},\dots,\boldsymbol{q}^{-r}_{i,k})}$$

where

$$w_{h,r}=\frac{\Phi^{(|\boldsymbol{\ell}^{-r}|+1)}_{k,0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{\bullet h}+1,\dots,\bar\ell^{-r}_{\bullet k})}{\Phi^{(|\boldsymbol{\ell}^{-r}|)}_{k,0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{\bullet h},\dots,\bar\ell^{-r}_{\bullet k})}\,\mathbb{1}_{\{0\}^c}(\bar\ell^{-r}_{i,h})+\mathbb{1}_{\{0\}}(\bar\ell^{-r}_{i,h})$$

and $\boldsymbol{1}_\kappa$ is a vector of dimension $\ell^{-r}_{i,h}$ with all components being zero but the κ–th which equals 1. Moreover, $T^{*,-r}_{i,h,1},\dots,T^{*,-r}_{i,h,\ell^{-r}_{i,h}}$ are the tables at the first restaurant where the h-th dish is served, after the removal of $T_{i,r}$.

(2.b) For $r=1,\dots,m_i$, generate $(X^{(t)}_{i,N_i+r},T^{(t)}_{i,N_i+r})$ from the following predictive distributions

$$\mathbb{P}(X_{i,N_i+r}=\text{“new”},T_{i,N_i+r}=\text{“new”}\,|\cdots)=$$

$$\frac{\Phi^{(|\boldsymbol{\ell}^{-r}|+1)}_{k+j^{-r}+1,0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{\bullet k+j^{-r}},1)}{\Phi^{(|\boldsymbol{\ell}^{-r}|)}_{k+j^{-r},0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{\bullet k+j^{-r}})}\;\frac{\Phi^{(N_i+m_i)}_{\bar\ell_{i\bullet}^{-r}+1,i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,k},1)}{\Phi^{(N_i+m_i-1)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,k})}$$

while, for any $h=1,\cdots,k+j^{-r}$ and $\kappa=1,\cdots,\ell^{-r}_{i,h}$,

$$\mathbb{P}(X_{i,N_i+r}=X^{*,-r}_h,T_{i,N_i+r}=\text{“new”}\,|\cdots)=$$

$$\frac{\Phi^{(|\boldsymbol{\ell}^{-r}|+1)}_{k+j^{-r},0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{1,h}+1,\dots,\bar\ell^{-r}_{\bullet k+j^{-r}})}{\Phi^{(|\boldsymbol{\ell}^{-r}|)}_{k+j^{-r},0}(\bar\ell^{-r}_{\bullet 1},\dots,\bar\ell^{-r}_{\bullet k+j^{-r}})}\;\frac{\Phi^{(N_i+m_i)}_{\bar\ell_{i\bullet}^{-r}+1,i}(\boldsymbol{q}^{-r}_{i,1},\dots,(\boldsymbol{q}^{-r}_{i,h},1),\dots,\boldsymbol{q}^{-r}_{i,k})}{\Phi^{(N_i+m_i-1)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,k})}$$

$$\mathbb{P}(X_{i,N_i+r}=X^{*,-r}_h,T_{i,N_i+r}=T^{*,-r}_{i,h,\kappa}\,|\cdots)=$$

$$\frac{\Phi^{(N_i+m_i)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,h}+\boldsymbol{1}_\kappa,\dots,\boldsymbol{q}^{-r}_{i,k},1)}{\Phi^{(N_i+m_i-1)}_{\bar\ell_{i\bullet}^{-r},i}(\boldsymbol{q}^{-r}_{i,1},\dots,\boldsymbol{q}^{-r}_{i,h},\dots,\boldsymbol{q}^{-r}_{i,k})}\,\mathbb{1}_{\{n^{-r}_{i,h}>0\}}$$

where $X^{*,-r}_h$, for $h=1,\dots,k+j^{-r}$ denote the distinct dishes in the whole franchise after the removal of the r–th observation, while the condition $n^{-r}_{i,h}>0$ entails that the h–th dish is served in the i–th restaurant.

The above algorithm holds for any hierarchical NRMI and only requires insertion of the specific ρ, ρ_0 and P_0 to specialize to a particular instance of hierarchical NRMI. The sampling schemes outlined above can also be tailored, in a quite straightforward way, to the hierarchical Pitman–Yor case (see the appendix for details and [3] for applications). Finally note that the proposed algorithm can also be adapted to yield a marginal sampling schemes for mixture models with dependent hierarchical mixing measures.

6.2 Simulation of $(\tilde{p}_1, \dots \tilde{p}_d)$ from its posterior distribution

The posterior representations derived in Theorems 10 and 13 are of great importance also from a computational standpoint as they allow to establish algorithms that generate the trajectories of $\tilde{p}_1, \dots, \tilde{p}_d$ from their posterior distributions, conditional on $\boldsymbol{T}$. The resulting sampling scheme can be viewed as an extension of a Ferguson & Klass–type algorithm (see [14, 44] for additional details) to a partially exchangeable setting. With respect to the generalized Blackwell–MacQueen urn scheme described in Section 6.1, the possibility of generating posterior samples of hierarchical processes is a significant addition. Just to give an example, it allows to obtain estimates of non–linear functionals, such as credible intervals, of the vector $(\tilde{p}_1, \cdots, \tilde{p}_d)$ that cannot be otherwise achieved.

For the sake of simplicity assume that $\mathbb{X} = \mathbb{R}^+$. Using a representation of X_t given in [14] and the notation of Theorems 9–10, one has

$$\eta_0^*((0,t]) = \sum_{h=1}^{\infty} J_h^{(0)} \mathbb{1}\{V_h \le P_0((0,t])\}, \tag{27}$$

with $V_1, V_2, \dots \overset{\text{iid}}{\sim} \mathrm{U}(0,1)$. The jumps $J_h^{(0)}$ are in decreasing order and can be obtained by solving the following

$$S_{h,0} = c_0 \int_{J_h^{(0)}}^{\infty} e^{-U_0 s} \rho_0(s)\, \mathrm{d}s. \tag{28}$$

where $S_{1,0}, S_{2,0}, \dots$ are the points of a standard Poisson process on $\mathbb{R}^+$, that is to say $S_{h,0} - S_{h-1,0}$ are i.i.d. exponential random variables having unit mean. Similarly, one has

$$\tilde{\mu}_i^*((0,t]) = \sum_{h=1}^{\infty} J_h^{(i)} \mathbb{1}\{V_h \le \tilde{p}_0^*((0,t])\}, \tag{29}$$

where the ordered jumps $J_h^{(i)}$ are now the solution of

$$S_{h,i} = c \int_{J_h^{(i)}}^{\infty} e^{-U_i s}\, \rho(s)\, \mathrm{d}s, \tag{30}$$

where $S_{1,i}, S_{2,i} - S_{1,i}, \dots$ are i.i.d. exponential random variables having unit mean. In view of these representations, once one has sampled the latent variables $\boldsymbol{T}$ through the algorithm described in Section 6.1, one can proceed as follows:

(1) Generate $\tilde{p}_0$ from its posterior distribution, described in Theorem 9, namely:

(1.a) Generate U_0 from $f_0(\,\cdot \mid \boldsymbol{X}, \boldsymbol{T})$ in (20);

(1.b) Generate I_j from $f_j(\,\cdot \mid \boldsymbol{X}, \boldsymbol{T})$ in Theorem 9(ii), for any $j = 1, \dots, k$;

(1.c) Fix $\varepsilon > 0$ and for any $h \geq 1$

- Generate unit mean exponential random variables $S_{h,0} - S_{h-1,0}$
- Determine jumps $J_h^{(0)}$ according to (28)
- Stop at $\bar{h} = \min\{h \geq 1 : J_h^{(0)} \leq \varepsilon\}$
- Generate i.i.d. $V_1, \ldots, V_{\bar{h}}$ from a $\mathrm{U}(0,1)$

and evaluate an approximate draw of η_0^* on $(0,t]$ as

$$\eta_0^*((0,t]) \approx \sum_{h=1}^{\bar{h}} J_h^{(0)} \mathbb{1}\{V_h \leq P_0((0,t])\},$$

(1.d) Evaluate an approximate draw of a posterior sample of $\tilde{p}_0$ as

$$\tilde{p}_0^*((0,t]) \approx \frac{\sum_{h=1}^{\bar{h}} J_h^{(0)} \mathbb{1}\{V_h \leq P_0((0,t])\} + \sum_{j=1}^{k} I_j \delta_{X_j^*}((0,t])}{\sum_{h=1}^{\bar{h}} J_h^{(0)} + \sum_{j=1}^{k} I_j}.$$

Having drawn $\tilde{p}_0^*$, one can now rely on Theorem 10 in order to approximately sample $(\tilde{p}_1, \ldots, \tilde{p}_d)$ from its posterior distribution. This can be easily deduced and described as follows.

(2) For any $i = 1, \ldots, d$, generate $\tilde{p}_i | (\boldsymbol{X}, \boldsymbol{T}, \tilde{p}_0^*)$ as follows

(2.a) Generate U_i from $f_i(\,\cdot\,|\boldsymbol{X}, \boldsymbol{T})$ in (22);

(2.b) Generate $J_{i,j,t}$ from $f_{i,j,t}(\,\cdot\,|\boldsymbol{X}, \boldsymbol{T})$ in Theorem 10(ii)

(2.c) Fix $\varepsilon > 0$ and for any $h \geq 1$

- Generate unit mean exponential random variables $S_{h,i} - S_{h-1,i}$
- Determine jumps $J_h^{(i)}$ according to (30)
- Stop at $\bar{h}_i = \min\{h \geq 1 : J_h^{(i)} \leq \varepsilon\}$

and evaluate an approximate sample of the posterior trajectory of $\tilde{p}_i$ as follows

$$\tilde{p}_i((0,t]) \approx \frac{\sum_{h=1}^{\bar{h}_i} J_h^{(i)} \mathbb{1}\{V_h \leq \tilde{p}_0^*((0,t])\} + \sum_{j=1}^{k} \sum_{t=1}^{\ell_{i,j}} J_{i,j,t} \delta_{X_j^*}((0,t])}{\sum_{h=1}^{\bar{h}_i} J_h^{(i)} + \sum_{j=1}^{k_i} \sum_{t=1}^{\ell_{i,j}} J_{i,j,t}}.$$

An important, and well–known, advantage of the procedure is the fact that it generates jumps $J_h^{(0)}$ and $J_h^{(i)}$, for $i = 1, \ldots, d$, in decreasing order. This entails that the truncation at $\bar{h}$ or $\bar{h}_i$ is such that the most relevant jumps are taken into account and one is discarding a negligible random mass of the actual trajectory. Future work, of more computational nature, will aim at: (i) investigating the implementation of the algorithm to applied problems, such as density estimation with accurate uncertainty quantification, allowed by the conditional structure of the algorithm and (ii) comparing the performance of our proposal with the so–called direct assignment algorithm, which is widely used within estimation problems involving the hierarchical Dirichlet process.

Acknowledgments

F. Camerlenghi has completed the paper while being a postdoctoral fellow at the University of Bologna. He is deeply grateful to the Department of Statistical Sciences for the support.

A Proofs

A.1 Proof of Theorem 1

For the sake of notational simplicity set

$$\mathcal{I}_q = \int_0^\infty u \mathrm{e}^{-c\psi(u)} \tau_q(u)\, \mathrm{d}u, \qquad \mathcal{I}_{q,0} = \int_0^\infty u \mathrm{e}^{-c_0\psi_0(u)} \tau_{q,0}(u)\, \mathrm{d}u.$$

From [22, Proposition 1] one has $\mathrm{Var}(\tilde{p}_0(A)) = c_0 P_0(A)(1-P_0(A))\,\mathcal{I}_{2,0}$ and $\mathrm{Var}(\tilde{p}_i(A)\,|\,\tilde{p}_0) = c\tilde{p}_0(A)(1-\tilde{p}_0(A))\,\mathcal{I}_2$. Hence, for any $i = 1, \dots, d$,

$$\begin{aligned}\mathrm{Var}(\tilde{p}_i(A)) &= \mathbb{E}\,\mathrm{Var}(\tilde{p}_i(A)\,|\,\tilde{p}_0) + \mathrm{Var}(\tilde{p}_0(A)) \\ &= c_0 P_0(A)(1-P_0(A))\{cc_0\,\mathcal{I}_2\mathcal{I}_{1,0} + \mathcal{I}_{2,0}\}.\end{aligned}$$

Moreover, for any $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(\tilde{p}_i(A), \tilde{p}_j(A)) &= \mathbb{E}\,\mathbb{E}[\tilde{p}_i(A)\tilde{p}_j(A)|\tilde{p}_0] - (\mathbb{E}\,\mathbb{E}[\tilde{p}_i(A)|\tilde{p}_0])\,(\mathbb{E}\,\mathbb{E}[\tilde{p}_j(A)|\tilde{p}_0]) \\ &= \mathbb{E}[\mathbb{E}[\tilde{p}_i(A)|\tilde{p}_0]\mathbb{E}[\tilde{p}_j(A)|\tilde{p}_0]] - (\mathbb{E}[\tilde{p}_0(A)])^2 \\ &= \mathbb{E}[\tilde{p}_0^2(A)] - (\mathbb{E}[\tilde{p}_0(A)])^2 = \mathrm{Var}(\tilde{p}_0(A)),\end{aligned}$$

and the result follows. □

A.2 Proof of Theorem 2

The same line of reasoning in the proof of Theorem 1 may be used to establish that $\mathrm{Cov}(\tilde{p}_i(A), \tilde{p}_j(A)) = \mathrm{Var}(\tilde{p}_0(A))$. Moreover, by the definition of Pitman–Yor process in Section 2, one has

$$\begin{aligned}\mathrm{Var}(\tilde{p}_0(A)) &= P_0(A)(1-P_0(A))\,\frac{\sigma_0^2(1-\sigma_0)}{\theta_0(\theta_0+1)\Gamma(\theta_0/\sigma_0)} \int_0^\infty u^{\theta_0+\sigma_0-1}\mathrm{e}^{-u^{\sigma_0}}\,\mathrm{d}u \\ &= P_0(A)(1-P_0(A))\,\frac{1-\sigma_0}{\theta_0+1}\end{aligned}$$

Furthermore, $\mathrm{Var}(\tilde{p}_i(A)) = \mathbb{E}\,\mathrm{Var}(\tilde{p}_i(A)\,|\,\tilde{p}_0) + \mathrm{Var}(\tilde{p}_0(A))$ holds and one obtains

$$\begin{aligned}\mathrm{Var}(\tilde{p}_i(A)) &= \frac{1-\sigma}{\theta+1}\,\mathbb{E}\,\tilde{p}_0(A)(1-\tilde{p}_0(A)) + \frac{1-\sigma_0}{\theta_0+1}P_0(A)(1-P_0(A)) \\ &= \frac{P_0(A)(1-P_0(A))}{\theta_0+1}\left\{(1-\sigma_0) + (\theta_0+\sigma_0)\frac{1-\sigma}{\theta+1}\right\},\end{aligned}$$

from which (10) easily follows. □

A.3 Proof of Theorem 3

In order to prove Theorem 3, we first need to display a useful technical lemma, which follows from a formula by Faà di Bruno. See, e.g., [19]. According to it, the n-th order derivative of $\mathrm{e}^{-m\psi(u)}$, with $\psi(u) = \int_0^\infty [1 - \mathrm{e}^{-uv}]\, \rho(v)\, \mathrm{d}v$, is given by

$$\frac{\mathrm{d}^n}{\mathrm{d}u^n}\mathrm{e}^{-m\psi(u)} = \sum_{\pi} \frac{\mathrm{d}^{|\pi|}}{\mathrm{d}x^{|\pi|}}(\mathrm{e}^{-mx})\Big|_{x=\psi(u)} \prod_{B\in\pi} \frac{\mathrm{d}^{|B|}}{\mathrm{d}u^{|B|}}\psi(u) \tag{31}$$

where $m \in \mathbb{R}$, and the sum is extended over all partitions π of $[n] = \{1, \ldots, n\}$. Clearly (31) can be rewritten as

$$\frac{\mathrm{d}^n}{\mathrm{d}u^n}\mathrm{e}^{-m\psi(u)} = \sum_{i=1}^{n}(-m)^i \mathrm{e}^{-m\psi(u)} \sum_{\pi:|\pi|=i} \prod_{B\in\pi} \frac{\mathrm{d}^{|B|}}{\mathrm{d}u^{|B|}}\psi(u).$$

Since to each unordered partition π of size i there correspond $i!$ ordered partitions of the set $[n]$ into i components, which are obtained by permuting the elements of π in all the possible ways, one has

$$\begin{aligned}
\frac{\mathrm{d}^n}{\mathrm{d}u^n}\mathrm{e}^{-m\psi(u)} &= \sum_{i=1}^{n}(-m)^i \mathrm{e}^{-m\psi(u)} \frac{1}{i!} \sum_{\pi:|\pi|=i}^{\mathrm{o}} \prod_{B\in\pi} \frac{\mathrm{d}^{|B|}}{\mathrm{d}u^{|B|}}\psi(u) \\
&= \sum_{i=1}^{n}(-m)^i \mathrm{e}^{-m\psi(u)} \frac{1}{i!} \sum_{(*)} \binom{n}{q_1, \cdots, q_i} \frac{\mathrm{d}^{q_1}}{\mathrm{d}u^{q_1}}\psi(u) \cdots \frac{\mathrm{d}^{q_i}}{\mathrm{d}u^{q_i}}\psi(u)
\end{aligned}$$

where $\sum^{\mathrm{o}}$ is the sum over the ordered partitions of the set $[n]$ while the sum $(*)$ runs over all vectors $(q_1, \cdots, q_i)$ of positive integers such that $\sum_{j=1}^{i} q_j = n$. The second equality follows upon noting that the derivative of ψ depends only on the number of elements within each component of the partition π and that the number of partitions π of the set $[n]$ containing i elements $(B_1, \cdots, B_i)$, with $(|B_1|, \cdots, |B_i|) = (q_1, \cdots, q_i)$, equals the multinomial coefficient above. In view of these remarks, one has

Lemma 1. *If* $\tau_q(u) = \int_0^\infty v^q\, \mathrm{e}^{-uv}\, \rho(v)\, \mathrm{d}v$ *and*

$$\xi_{n,i}(u) = \sum_{(*)} \frac{1}{i!} \binom{n}{q_1, \cdots, q_i} \tau_{q_1}(u) \cdots \tau_{q_i}(u), \tag{32}$$

the following relation holds

$$(-1)^n \frac{\mathrm{d}^n}{\mathrm{d}u^n} e^{-m\psi(u)} = e^{-m\psi(u)} \sum_{i=1}^{n} m^i \xi_{n,i}(u) \tag{33}$$

In view of this, for any $x_1 \neq \cdots \neq x_k$ in $\mathbb{X}$ and $k \geq 1$, one can evaluate

$$M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(\mathrm{d}x_1, \ldots, \mathrm{d}x_k) = \mathbb{E} \prod_{j=1}^{k} \prod_{i=1}^{d} \tilde{p}_i^{n_{i,j}}(\mathrm{d}x_j)$$

where $\boldsymbol{n}_i = (n_{i,1}, \ldots, n_{i,k})$, for $i = 1, \ldots, d$, and it is apparent that there may be $n_{i,j}$'s equal to zero, though $\sum_{i=1}^{d} n_{i,j} \geq 1$. Indeed, if one sets $A_{j,\varepsilon} = B(x_j; \varepsilon)$ a ball of radius ε around x_j, with $\varepsilon > 0$ small enough so that $A_{i,\varepsilon} \cap A_{j,\varepsilon} = \varnothing$ for any $i \neq j$, then

$$M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(A_{1,\varepsilon} \times \cdots \times A_{k,\varepsilon}) = \mathbb{E} \prod_{i=1}^{d} \mathbb{E}\Big[\prod_{j=1}^{k} \tilde{p}_i^{n_{i,j}}(A_{j,\varepsilon}) \Big| \tilde{p}_0 \Big]$$

$$
\begin{aligned}
&= \mathbb{E} \prod_{i=1}^{d} \frac{1}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} \mathbb{E}[\mathrm{e}^{-u\tilde{\mu}_i(\mathbb{X}_\varepsilon^*)}|\tilde{p}_0] \\
&\qquad \times \prod_{j=1}^{k} \mathbb{E}[\mathrm{e}^{-u\tilde{\mu}_i(A_{j,\varepsilon})} \tilde{\mu}_i^{n_{i,j}}(A_{j,\varepsilon})|\tilde{p}_0] \, \mathrm{d}u \\
&= \mathbb{E} \prod_{i=1}^{d} \frac{1}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} \mathrm{e}^{-c\psi(u)\tilde{p}_0(\mathbb{X}_\varepsilon^*)} \\
&\qquad \times \prod_{j=1}^{k} \Big((-1)^{n_{i,j}} \frac{\mathrm{d}^{n_{i,j}}}{\mathrm{d}u^{n_{i,j}}} e^{-c\psi(u)\tilde{p}_0(A_{j,\varepsilon})} \Big) \, \mathrm{d}u
\end{aligned}
$$

where $\mathbb{X}_\varepsilon^* = \mathbb{X} \setminus (\cup_{j=1}^k A_{j,\varepsilon})$. If $\boldsymbol{\ell} = (\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_d)$ where each $\boldsymbol{\ell}_i = (\ell_{i,1}, \ldots, \ell_{i,k}) \in \times_{j=1}^k \{1, \ldots, n_{i,j}\}$, by virtue of (33) one obtains

$$
\begin{aligned}
&M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(A_{1,\varepsilon} \times \cdots \times A_{k,\varepsilon}) \\
&= \mathbb{E} \prod_{i=1}^{d} \frac{1}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} \, \mathrm{e}^{-c\,\psi(u)} \prod_{j=1}^{k} \sum_{\ell_{i,j}=1}^{n_{i,j}} c^{\ell_{i,j}} \tilde{p}_0^{\ell_{i,j}}(A_{j\varepsilon}) \, \xi_{n_{i,j},\ell_{i,j}}(u) \, \mathrm{d}u \\
&= \sum_{\boldsymbol{\ell}} \mathbb{E} \prod_{j=1}^{k} \tilde{p}_0^{\bar{\ell}_{\bullet j}}(A_{j,\varepsilon}) \prod_{i=1}^{d} \Big(\frac{c^{\bar{\ell}_{i\bullet}}}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} e^{-c\psi(u)} \prod_{j=1}^{k} \xi_{n_{i,j},\ell_{i,j}}(u) \mathrm{d}u \Big)
\end{aligned}
$$

and we further agree that, whenever $n_{i,j} = 0$, we have $\ell_{i,j} = 0$ and $\sum_{\ell_{i,j}=1}^{0} \equiv 1$. If we set

$$
M^0_{\bar{\ell}_{\bullet 1},\ldots,\bar{\ell}_{\bullet k}}(A_{1,\varepsilon}, \ldots, A_{k,\varepsilon}) = \mathbb{E} \prod_{j=1}^{k} \tilde{p}_0^{\bar{\ell}_{\bullet j}}(A_{j,\varepsilon})
$$

as $\varepsilon \to 0$ the non–atomicity of P_0 implies

$$
\begin{aligned}
M^0_{\bar{\ell}_{\bullet 1},\ldots,\bar{\ell}_{\bullet k}}(\mathrm{d}x_1, \ldots, \mathrm{d}x_k) &= \mathbb{E} \prod_{j=1}^{k} \tilde{p}_0^{\bar{\ell}_j}(\mathrm{d}x_j) = \prod_{j=1}^{k} P_0(\mathrm{d}x_j) \\
&\quad \times \frac{c_0^k}{\Gamma(|\boldsymbol{\ell}|)} \int_0^\infty u^{|\boldsymbol{\ell}| - 1} \, \mathrm{e}^{-c_0\psi_0(u)} \prod_{j=1}^{k} \tau_{\bar{\ell}_{\bullet j},0}(u) \, \mathrm{d}u \\
&= \Big(\prod_{j=1}^{k} P_0(\mathrm{d}x_j) \Big) \, \Phi^{(|\boldsymbol{\ell}|)}_{k,0}(\bar{\ell}_{\bullet 1}, \ldots, \bar{\ell}_{\bullet k}).
\end{aligned}
$$

Then

$$
\begin{aligned}
M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(\mathrm{d}x_1, \ldots, \mathrm{d}x_k) &= \Big(\prod_{j=1}^{k} P_0(\mathrm{d}x_j) \Big) \sum_{\boldsymbol{\ell}} \Phi^{(|\boldsymbol{\ell}|)}_{k,0}(\bar{\ell}_{\bullet 1}, \ldots, \bar{\ell}_{\bullet k}) \\
&\qquad \times \prod_{i=1}^{d} \Big(\frac{c^{\bar{\ell}_{i\bullet}}}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} e^{-c\psi(u)} \prod_{j=1}^{k} \xi_{n_{i,j},\ell_{i,j}}(u) \mathrm{d}u \Big)
\end{aligned}
$$

and the result follows by virtue of (32) and by noting that

$$\Pi_k^{(n)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d) = \int_{\mathbb{X}^k} M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(\mathrm{d}x_1,\ldots,\mathrm{d}x_k).$$

□

A.4 Proof of Theorem 4

This can be deduced by working along the same lines as for the Proof of Theorem 3. One only needs to take into account the change of measure (8) in order to work directly with CRMs. Indeed, using the same notation, one has

$$\begin{aligned}
M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(A_{1,\varepsilon}\times\cdots\times A_{k,\varepsilon}) &= \frac{\Gamma^d(\theta+1)}{\Gamma^d\Big(\frac{\theta}{\sigma}+1\Big)}\,\mathbb{E}\prod_{i=1}^d \frac{1}{\Gamma(\theta+N_i)}\int_0^\infty u^{\theta+N_i-1}\,\mathrm{e}^{-\psi(u)}\\
&\quad\times\prod_{j=1}^k\sum_{\ell_{i,j}=1}^{n_{i,j}}\tilde p_0^{\ell_{i,j}}(A_{j\varepsilon})\,\xi_{n_{i,j},\ell_{i,j}}(u)\,\mathrm{d}u\\
&= \frac{\Gamma^d(\theta+1)}{\Gamma^d\Big(\frac{\theta}{\sigma}+1\Big)}\sum_{\boldsymbol{\ell}}\Big(\mathbb{E}\prod_{j=1}^k \tilde p_0^{\bar\ell_{\bullet j}}(A_{j,\varepsilon})\Big)\\
&\quad\times\prod_{i=1}^d\Big(\frac{1}{\Gamma(N_i)}\int_0^\infty u^{N_i-1}e^{-\psi(u)}\prod_{j=1}^k\xi_{n_{i,j},\ell_{i,j}}(u)\mathrm{d}u\Big)
\end{aligned}$$

where $\psi(u)=u^\sigma$ and

$$\begin{aligned}
\xi_{n_{i,j},\ell_{i,j}}(u) &= \frac{\sigma^{\ell_{i,j}}}{u^{n_{i,j}-\ell_{i,j}\sigma}\,\ell_{i,j}!}\sum_{\boldsymbol{q}}\binom{n_{i,j}}{q_{i,1}\ \cdots\ q_{i,\ell_{i,j}}}\prod_{r=1}^{\ell_{i,j}}(1-\sigma)_{q_{i,r}-1}\\
&= \frac{1}{u^{n_{i,j}-\ell_{i,j}\sigma}}\,\mathscr{C}(n_{i,j},\ell_{i,j};\sigma).
\end{aligned}$$

The result now easily follows upon noting that, as $\varepsilon\downarrow 0$,

$$\mathbb{E}\prod_{j=1}^k\tilde p_0^{\bar\ell_{\bullet j}}(A_{j,\varepsilon}) = \prod_{j=1}^k P_0(A_{j\varepsilon})\,\frac{\prod_{i=1}^{k-1}(\theta_0+i\sigma_0)}{(\theta_0+1)_{|\boldsymbol{\ell}|-1}}\,\prod_{j=1}^k(1-\sigma_0)_{\bar\ell_{\bullet j}-1}+\lambda_{k,\varepsilon}$$

where $\lambda_{k,\varepsilon}=\mathrm{o}\Big(\prod_{j=1}^k P_0(A_{j,\varepsilon})\Big)$. □

A.5 Proof of Theorem 5

Let $\bar N_0\equiv 0$ and, for each $i=1,\ldots,d$, $\bar N_i=\sum_{j=1}^i N_j$. Further suppose $\tilde\pi_1,\ldots,\tilde\pi_d$ denote independent random partitions of $\mathbb{N}$ such that the restriction $\tilde\pi_{i,N_i}$ of $\tilde\pi_i$ to $[N_i]=\{\bar N_{i-1}+1,\ldots,\bar N_i\}$ has probability distribution $\Phi_{\zeta_i,i}^{(N_i)}$ as defined in (13). Furthermore, $\tilde\pi_0$ is a random partition of $\mathbb{N}$ such that, conditional on $(\tilde\pi_{1,N_1},\ldots,\tilde\pi_{d,N_d})$, its restriction $\tilde\pi_{0,h}$ to $[h]$ has probability distribution $\Phi_{k,0}^{(h)}$ in (11), where $h=\sum_{i=1}^d|\tilde\pi_{i,N_i}|$ with $|\tilde\pi_{i,N_i}|$ the number of blocks in the partition $\tilde\pi_{i,N_i}$. Relying on Theorem 3 we have

$$\Pi_k^{(N)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d)=\sum_{h=k}^{N}\prod_{i=1}^{d}\sum_{(*)}\mathbb{P}[\tilde\pi_{i,N_i}=\{B_{i,1},\ldots,B_{i,\zeta_i}\}]$$
$$\times\sum_{(**)}\mathbb{P}\Big[\tilde\pi_{0,h}=\{C_1,\ldots,C_k\}\,\Big|\,\bigcap_{i=1}^{d}\Big\{\tilde\pi_{i,N_i}=\{B_{i,1},\ldots,B_{i,\zeta_i}\}\Big\}\Big]$$

where the sums are taken over all partitions such that $\sum_{i=1}^d \zeta_i = h$, for any $h\in\{k\ldots,N\}$, and

$$\sum_{\{t:\,\bar\zeta_{i-1}+t\in C_j\}\cap\{1,\ldots,\zeta_i\}}\mathrm{card}(B_{i,t})=n_{i,j}$$

where $\bar\zeta_0\equiv 0$ and $\bar\zeta_i=\sum_{r=1}^i\zeta_r$, for each $i=1,\ldots,d$. Note that one may have $\{t:\ \bar\zeta_{i-1}+t\in C_j\}\cap\{1,\ldots,\zeta_i\}=\varnothing$, in which case $n_{i,j}=0$, and in the Chinese restaurant franchise terminology it would mean that the j–th dish is not served in restaurant i. According to this notation, one has $K'_{i,N_i}=|\tilde\pi_{i,N_i}|$ and $K_{0,h}=|\tilde\pi_{0,h}|$ and

$$\mathbb{P}[K_N=k]=\mathbb{P}\Big[\bigcup_{t=k}^{N}\ \bigcup_{(t_1,\ldots,t_d)\in\Delta_{d,t}}\{K'_{1,N_1}=t_1,\,\ldots\,,K'_{d,N_d}=t_d,\,K_{0,t}=k\}\Big].$$

The result, then, follows from independence of $K'_{1,N_1},\ldots,K'_{d,N_d}$ and from the fact that

$$\mathbb{P}[K_{0,t}=k\,|\,K'_{1,N_1}=t_1,\ldots,K'_{d,N_d}=t_d]=\mathbb{P}[K_{0,t}=k]\,\mathbb{1}_{\Delta_{d,t}}(t_1,\ldots,t_d).$$

□

A.6 Proof of Theorem 6

The proof follows exactly the same lines as the proof of Theorem 5, the only difference being that K'_{i,N_i} are the number of blocks corresponding to independent random partitions from $\mathrm{PY}(\sigma,\theta;H_0)$ processes, for some diffuse probability measure H_0, and $K_{0,t}$ is the number of blocks of a random partition from a $\mathrm{PY}(\sigma_0,\theta_0;P_0)$. □

A.7 Proof of Theorem 7

First note that the combination of partial exchangeability, which entails conditional independence across samples, and the hierarchical structure of $(\tilde p_1,\ldots,\tilde p_d)$, which implies the sharing of atoms across samples, lead to $K_n=K_{0,\xi(\boldsymbol{N})}$ almost surely. Moreover, in view of the growth assumptions, let

$$\frac{K_{0,n}}{\lambda_0(n)}\xrightarrow{\text{a.s.}}M_0,\qquad \frac{K'_{i,n}}{\lambda(n)}\xrightarrow{\text{a.s.}}M'_i\quad(i=1,\ldots,d)$$

as $n\to\infty$, where M_0 and the M'_is are positive and finite random variables. Moreover, one has $\xi(\boldsymbol{N})=\sum_{i=1}^d K'_{i,N^*}$ and

$$\frac{\xi(\boldsymbol{N})}{\lambda(N^*)}\xrightarrow{\text{a.s.}}\sum_{i=1}^{d}M'_i=\eta$$

as $N^*\to\infty$. Exploiting (H1) and recalling that $N=dN^*$, one then obtains

$$\frac{K_{0,\xi(\boldsymbol{N})}}{K_{0,\eta\lambda(N^*)}}=\frac{\lambda_0(\xi(\boldsymbol{N}))}{\lambda_0(\eta\lambda(N^*))}\,\frac{K_{0,\xi(\boldsymbol{N})}/\lambda_0(\xi(\boldsymbol{N}))}{K_{0,\eta\lambda(N^*)}/\lambda_0(\eta\lambda(N^*))}\xrightarrow{\text{a.s.}}1,$$

which entails, as $N^* \to \infty$,

$$\frac{K_N}{\lambda_0(\eta\lambda(N^*))} = \frac{K_{0,\xi(\boldsymbol{N})}}{K_{0,\eta\lambda(N^*)}} \frac{K_{0,\eta\lambda(N^*)}}{\lambda_0(\eta\lambda(N^*))} \xrightarrow{\text{a.s.}} M_0$$

If $(\tilde{p}_1, \ldots, \tilde{p}_d)$ is a vector of hierarchical Dirichlet processes, then $\lambda_0(n) = \lambda(n) = \log n$ and (H1) holds true. Moreover, $\eta = dc$ and

$$\frac{\lambda_0(\eta\lambda(N))}{\lambda_0(\lambda(N))} \to 1$$

as $N \to \infty$, and the result follows. □

A.8 Proof of Theorem 8

In this case, $\lambda_0(n) = n^{\sigma_0}$ and $\lambda(n) = n^{\sigma}$. Hence (H1) holds and one may work out the proof along the same lines as for Theorem 7. □

A.9 Proof of Theorem 9

The technique introduced for proving Theorem 3 can be suitably adapted and extended to establish also the posterior characterization. Our goal is the determination of the posterior Laplace functional of $\tilde{\mu}_0$

$$\mathbb{E}\Big[\mathrm{e}^{-\tilde{\mu}_0(f)} \,\Big|\, \boldsymbol{X}\Big] = \lim_{\varepsilon \downarrow 0} \frac{\mathbb{E}\,\mathrm{e}^{-\tilde{\mu}_0(f)} \prod_{j=1}^k \prod_{i=1}^d \tilde{p}_i^{n_{i,j}}(A_{j,\varepsilon})}{\mathbb{E} \prod_{j=1}^k \prod_{i=1}^d \tilde{p}_i^{n_{i,j}}(A_{j,\varepsilon})} \tag{34}$$

for any measurable $f : \mathbb{X} \to \mathbb{R}^+$, where the $A_{j,\varepsilon}$ notation is the same used in the Proofs of Theorems 3–4. The denominator is $M_{\boldsymbol{n}_1,\ldots,\boldsymbol{n}_d}(A_{1,\varepsilon} \times \cdots \times A_{k,\varepsilon})$ that has been defined in the proof of Theorem 3 and, as $\varepsilon \downarrow 0$, equals

$$\prod_{j=1}^k P_0(A_{j\varepsilon}) \sum_{\boldsymbol{\ell}} \sum_{\boldsymbol{q}} \Phi_{k,0}^{(|\boldsymbol{\ell}|)}(\bar{\ell}_{\bullet 1}, \cdots, \bar{\ell}_{\bullet k})$$

$$\times \prod_{i=1}^d \prod_{j=1}^k \frac{1}{\ell_{i,j}!} \binom{n_{i,j}}{q_{i,j,1}, \cdots, q_{i,j,\ell_{i,j}}} \Phi_{\bar{\ell}_{i\bullet},i}^{(N_i)}(\boldsymbol{q}_{i,1}, \ldots, \boldsymbol{q}_{i,k}) + \lambda_{k,\varepsilon} \tag{35}$$

where $\lambda_{k,\varepsilon} = \mathrm{o}\Big(\prod_{j=1}^k P_0(A_{j,\varepsilon})\Big)$. In a similar fashion one can proceed to determine the numerator of (34) and

$$\mathbb{E}\,\mathrm{e}^{-\tilde{\mu}_0(f)} \prod_{j=1}^k \prod_{i=1}^d \tilde{p}_i^{n_{i,j}}(A_{j,\varepsilon}) = \mathbb{E}\,\mathrm{e}^{-\tilde{\mu}_0(f)} \prod_{i=1}^d \mathbb{E}\Big[\tilde{p}_i^{n_{i,j}}(A_{j,\varepsilon}) \,\Big|\, \tilde{\mu}_0\Big]$$

$$= \sum_{\boldsymbol{\ell}} \prod_{i=1}^d \frac{c^{\bar{\ell}_{i\bullet}}}{\Gamma(N_i)} \int_0^\infty u^{N_i - 1} \mathrm{e}^{-c\,\psi(u)} \prod_{j=1}^k \xi_{n_{i,j},\ell_{i,j}}(u)\,\mathrm{d}u \Big(\mathbb{E}\mathrm{e}^{-\tilde{\mu}_0(f)} \prod_{j=1}^k \tilde{p}_0^{\bar{\ell}_{\bullet j}}(A_{j,\varepsilon})\Big)$$

It can now be seen that, with $\mathbb{X}_\varepsilon^* = \mathbb{X} \setminus (\cup_{j=1}^k A_{j,\varepsilon})$ and as $\varepsilon \downarrow 0$,

$$\mathbb{E}\mathrm{e}^{-\tilde{\mu}_0(f)} \prod_{j=1}^k \tilde{p}_0^{\bar{\ell}_{\bullet j}}(A_{j,\varepsilon}) = \frac{1}{\Gamma(|\boldsymbol{\ell}|)} \int_0^\infty u^{|\boldsymbol{\ell}|-1} \Big(\mathbb{E}\mathrm{e}^{-\tilde{\mu}_0((f+u)\mathbb{1}_{\mathbb{X}_\varepsilon^*})}\Big)$$

$$\times \Big(\prod_{j=1}^{k} \mathbb{E}\mathrm{e}^{-\tilde\mu_0((f+u)\mathbb{1}_{A_{j,\varepsilon}})}\,\tilde\mu_0^{\bar\ell_{\bullet j}}(A_{j,\varepsilon})\Big)\,\mathrm{d}u$$

$$= \frac{\prod_{j=1}^k c_0\, P_0(A_{j,\varepsilon})}{\Gamma(|\boldsymbol{\ell}|)} \int_0^\infty u^{|\boldsymbol{\ell}|-1}\,\mathrm{e}^{-c_0\,\psi_0(f+u)}$$

$$\times \prod_{j=1}^{k} \tau_{\bar\ell_{\bullet j}}(u+f(X_j^*))\,\mathrm{d}u + \lambda_{k,\varepsilon}$$

Hence, one has

$$\mathbb{E}\,\mathrm{e}^{-\tilde\mu_0(f)} \prod_{j=1}^k\prod_{i=1}^d \tilde p_i^{n_{i,j}}(A_{j,\varepsilon}) = \prod_{j=1}^k P_0(A_{j,\varepsilon})$$

$$\times \sum_{\boldsymbol{\ell},\boldsymbol{q}} \prod_{i=1}^d\prod_{j=1}^k \frac{1}{\ell_{i,j}!}\binom{n_{i,j}}{q_{i,j,1}\ \cdots\ q_{i,j,\ell_{i,j}}}\,\Phi_{\bar\ell_{i\bullet},i}(\boldsymbol{q}_{i,1},\dots,\boldsymbol{q}_{i,k})$$

$$\times \frac{c_0^{|\boldsymbol{\ell}|}}{\Gamma(|\boldsymbol{\ell}|)}\int_0^\infty u^{|\boldsymbol{\ell}|-1}\,\mathrm{e}^{-c_0\,\psi_0(f+u)}\prod_{j=1}^k \tau_{\bar\ell_{\bullet j}}(u+f(X_j^*))\,\mathrm{d}u + \lambda_{k,\varepsilon}.$$

To sum up, if we further condition on $\boldsymbol{T}$, one has

$$\mathbb{E}\Big[\mathrm{e}^{-\tilde\mu_0(f)}\,\Big|\,\boldsymbol{X},\boldsymbol{T}\Big] = \frac{\int_0^\infty u^{|\boldsymbol{\ell}|-1}\,\mathrm{e}^{-c_0\,\psi_0(f+u)}\prod_{j=1}^k \tau_{\bar\ell_{\bullet j}}(u+f(X_j^*))\,\mathrm{d}u}{\Phi_{k,0}^{(|\boldsymbol{\ell}|)}(\bar\ell_{\bullet 1},\dots,\bar\ell_{\bullet k})}$$

and the result follows, as it is easy to check that the normalizing constant of the density $f_0(\,\cdot\,|\boldsymbol{X},\boldsymbol{T})$ in (20) is $\Phi_{k,0}^{(|\boldsymbol{\ell}|)}(\bar\ell_{\bullet 1},\dots,\bar\ell_{\bullet k})$. □

A.10 Proof of Theorem 10

From elementary properties of conditional expectations, one has

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d\tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T}\Big] = \mathbb{E}\Big[\,\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d\tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T},\tilde\mu_0\Big]\,\Big|\,\boldsymbol{X},\boldsymbol{T}\Big]$$

for any choice of measurable functions $f_i : \mathbb{X}\to\mathbb{R}^+$, for $i=1,\dots,d$. The proof, then, boils down to determining posterior Laplace functional

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d\tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T},\tilde\mu_0\Big] = \lim_{\varepsilon\downarrow 0}\frac{\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d\tilde\mu_i(f_i)}\prod_{i=1}^d\prod_{j=1}^k\tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\Big]}{\mathbb{E}\Big[\prod_{i=1}^d\prod_{j=1}^k\tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\Big]} \tag{36}$$

As far as the denominator is concerned, it plainly equals

$$\Big(\prod_{i=1}^d\prod_{j=1}^k\tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\Big)\prod_{i=1}^d\frac{c^{\bar\ell_{i\bullet}}}{\Gamma(N_i)}\int_0^\infty u^{N_i-1}\,\mathrm{e}^{-c\,\psi(u)}\prod_{j=1}^k\prod_{t=1}^{\ell_{i,j}}\tau_{q_{i,j,t}}(u)\,\mathrm{d}u$$

As for the numerator, set

$$\tilde\psi(f) = \int_{\mathbb{X}}\int_0^\infty [1-\mathrm{e}^{-sf(x)}]\,\rho(s)\,\mathrm{d}s\,\tilde p_0(\mathrm{d}x)$$

and note that

$$\mathbb{E}\left[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)}\prod_{i=1}^d\prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\right]=\prod_{i=1}^d \mathbb{E}\Big[\mathrm{e}^{-\tilde\mu_i(f_i)}\prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\Big]$$

$$=\Big(\prod_{i=1}^d\prod_{j=1}^k \tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\Big)\prod_{i=1}^d \frac{c^{\bar\ell_{i\bullet}}}{\Gamma(N_i)}\int_0^\infty u^{N_i-1}\,\mathrm{e}^{-c\,\tilde\psi(f+u)}\prod_{j=1}^k\prod_{t=1}^{\ell_{i,j}}\tau_{q_{i,j,t}}(u+f_i(X_j^*))\,\mathrm{d}u$$

To sum up, the right–hand–side of (36) equals

$$\prod_{i=1}^d \frac{\int_0^\infty u^{N_i-1}\mathrm{e}^{-c\tilde\psi(f_i+u)}\prod_{j=1}^k\prod_{t=1}^{\ell_{i,j}}\tau_{q_{i,j,t}}(u+f_i(X_j^*))\mathrm{d}u}{\int_0^\infty u^{N_i-1}\mathrm{e}^{-c\psi(u)}\prod_{j=1}^k\prod_{t=1}^{\ell_{i,j}}\tau_{q_{i,j,t}}(u)\mathrm{d}u}.$$

which entails

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T},\boldsymbol{U},\tilde\mu_0\Big]=\prod_{i=1}^d\prod_{j=1}^k\prod_{t=1}^{\ell_{i,j}}\frac{\tau_{q_{i,j,t}}(U_i+f_i(X_j^*))}{\tau_{q_{i,j,t}}(U_i)}$$
$$\times\prod_{i=1}^d \exp\left\{-c\int_{\mathbb{X}\times\mathbb{R}^+}(1-e^{-sf_i(x)})\mathrm{e}^{-sU_i}\,\rho(s)\,\mathrm{d}s\,\tilde p_0(\mathrm{d}x)\right\}$$

and the conclusion follows. □

A.11 Proof of Theorem 11

The proof follows the similar lines as that of Theorem 9, the main difference being the polynomial tilting that defines the Pitman–Yor process in (8). Let $\tilde\mu_0'$ and μ_i' denote stable CRMs with respective parameters σ_0 and σ in $(0,1)$. Moreover, the base measure $\tilde\mu_0'$ is P_0 whereas the $\tilde\mu_i'$ are, conditional on $\tilde\mu_0'$, independent and identically distributed CRMs with base measure $\tilde p_0'=\tilde\mu_0'/\tilde\mu_0'(\mathbb{X})$. When determining the posterior Laplace functional transform as in (34), one has

$$\mathbb{E}\,\mathrm{e}^{-\tilde\mu_0(f)}\prod_{j=1}^k\prod_{i=1}^d \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})=\mathbb{E}\Big[\mathrm{e}^{-\tilde\mu_0(f)}\prod_{i=1}^d\mathbb{E}\Big[(\tilde\mu_i(\mathbb{X}))^{-N_i}\prod_{j=1}^k(\tilde\mu_i(A_{j,\varepsilon}))^{n_{i,j}}\,\Big|\,\tilde\mu_0\Big]\Big]. \tag{37}$$

It can now be seen that, for any $i=1,\dots,d$,

$$\mathbb{E}\Big[(\tilde\mu_i(\mathbb{X}))^{-N_i}\prod_{j=1}^k(\tilde\mu_i(A_{j,\varepsilon}))^{n_{i,j}}\,\Big|\,\tilde\mu_0\Big]$$
$$=\frac{\Gamma(\theta+1)}{\Gamma\Big(\frac{\theta}{\sigma}+1\Big)}\,\mathbb{E}\Big[(\tilde\mu_i'(\mathbb{X}))^{-\theta-N_i}\prod_{j=1}^k(\tilde\mu_i'(A_{j,\varepsilon}))^{n_{i,j}}\,\Big|\,\tilde\mu_0\Big]$$
$$=\frac{1}{(\theta+1)_{N_i-1}}\,\frac{1}{\Gamma\Big(\frac{\theta}{\sigma}+1\Big)}\int_0^\infty v^{\theta+N_i-1}\mathbb{E}\Big[\mathrm{e}^{-v\tilde\mu_i'(\mathbb{X})}\prod_{j=1}^k(\tilde\mu_i'(A_{j,\varepsilon}))^{n_{i,j}}\,\Big|\,\tilde\mu_0\Big]\,\mathrm{d}v$$
$$=\frac{1}{(\theta+1)_{N_i-1}}\,\frac{1}{\Gamma\Big(\frac{\theta}{\sigma}+1\Big)}\int_0^\infty v^{\theta+N_i-1}\,\mathrm{e}^{-v^\sigma}\prod_{j=1}^k\sum_{\ell_{i,j}=1}^{n_{i,j}}\tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\,\xi_{n_{i,j},\ell_{i,j}}(v)\,\mathrm{d}v.$$

It is clear that, in order to evaluate (37), one may note that as $\varepsilon \downarrow 0$

$$\mathbb{E}\mathrm{e}^{-\tilde\mu_0(f)} \prod_{j=1}^k \tilde p_0^{\bar\ell_{\bullet j}}(A_{j,\varepsilon}) = \frac{\Gamma(\theta_0+1)}{\Gamma\Big(\frac{\theta_0}{\sigma_0}+1\Big)}\, \mathbb{E}(\tilde\mu_0'(\mathbb{X}))^{-\theta_0-|\boldsymbol{\ell}|}\, \mathrm{e}^{-\tilde\mu_0'(f)} \prod_{j=1}^k (\tilde\mu_0'(A_{j,\varepsilon}))^{\bar\ell_{\bullet j}}$$

$$= \frac{1}{(\theta_0+1)_{|\boldsymbol{\ell}|-1}}\, \frac{\sigma_0^k}{\Gamma\Big(\frac{\theta_0}{\sigma_0}+1\Big)} \Big(\prod_{j=1}^k P_0(A_{j,\varepsilon})\,(1-\sigma_0)_{\bar\ell_{\bullet j}-1}\Big) \int_0^\infty v_0^{\theta_0+|\boldsymbol{\ell}|-1}$$

$$\times\ \mathrm{e}^{-\int_0^\infty (v_0+f(x))^{\sigma_0}\,P_0(\mathrm{d}x)} \prod_{j=1}^k \frac{1}{(v_0+f(X_j^*))^{\bar\ell_{\bullet j}-\sigma_0}}\,\mathrm{d}v_0 + \lambda_{k,\varepsilon}$$

where $\lambda_{k,\varepsilon} = \mathrm{o}(\prod_{j=1}^k P_0(A_{j,\varepsilon}))$. If one further conditions on $\boldsymbol{T}$, it can be seen that

$$\mathbb{E}\Big[\mathrm{e}^{-\tilde\mu_0(f)} \prod_{j=1}^k \prod_{i=1}^d \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T}\Big] = \frac{1}{(\theta_0+1)_{|\boldsymbol{\ell}|-1}}\, \frac{\sigma_0^k}{\Gamma\Big(\frac{\theta_0}{\sigma_0}+1\Big)} \prod_{i=1}^d \frac{\prod_{r=1}^{\bar\ell_{i\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_i-1}}$$

$$\times \prod_{j=1}^k P_0(A_{j,\varepsilon})(1-\sigma_0)_{\bar\ell_{\bullet j}-1} \prod_{i=1}^d \frac{\mathscr{C}(n_{i,j},\ell_{i,j};\sigma)}{\sigma^{\ell_{i,j}}}$$

$$\times \int_0^\infty v_0^{\theta_0+k\sigma_0-1}\,\mathrm{e}^{-v_0^{\sigma_0}}\,\mathrm{e}^{-\int_0^\infty [(v_0+f(x))^{\sigma_0}-v_0^{\sigma_0})\,P_0(\mathrm{d}x)}$$

$$\times \Big(\prod_{j=1}^k \mathbb{E}\,\mathrm{e}^{-I_j f(X_j^*)}\Big)\,\mathrm{d}v_0$$

The proof is, thus, completed by using Theorem 4, which identifies

$$\mathbb{E}\Big[\prod_{i=1}^d \prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T}\Big],$$

as $\varepsilon \downarrow 0$. $\square$

A.12 Proof of Theorem 12

As in the proof of Theorem 10, we aim at determining

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T},\tilde\mu_0\Big]\,\Big|\,\boldsymbol{X},\boldsymbol{T}\Big]$$

for any collection of non–negative and measurable functions $f_1,\ldots,f_d$ defined on $\mathbb{X}$. See (36). This is achieved through the following

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)}\,\Big|\,\boldsymbol{X},\boldsymbol{T},\tilde\mu_0\Big] = \prod_{i=1}^d \mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)} \prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\Big].$$

For any $i=1,\ldots,d$, one has

$$\mathbb{E}\Big[\mathrm{e}^{-\sum_{i=1}^d \tilde\mu_i(f_i)} \prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon})\,\Big|\,\boldsymbol{T},\tilde\mu_0\Big]$$

$$= \frac{1}{(\theta+1)_{N_i-1}} \frac{1}{\Gamma\left(\frac{\theta}{\sigma}+1\right)} \int_0^\infty v^{\theta+N_i-1}\, \mathbb{E}\Big[\mathrm{e}^{-\tilde\mu_i'(v+f_i)} \prod_{j=1}^k (\tilde\mu_i'(A_{j,\varepsilon}))^{n_{i,j}} \,\Big|\, \boldsymbol{T}, \tilde\mu_0\Big]\, \mathrm{d}v$$

$$= \frac{1}{(\theta+1)_{N_i-1}} \frac{1}{\Gamma\left(\frac{\theta}{\sigma}+1\right)} \prod_{j=1}^k \tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\, \mathscr{C}(n_{i,j},\ell_{i,j};\sigma)$$

$$\times \int_0^\infty v^{\theta+N_i-1}\, \mathrm{e}^{-\int_0^\infty [v+f_i(x)]^\sigma \tilde p_0(\mathrm{d}x)} \prod_{j=1}^k \frac{1}{(v+f_i(X_j^*))^{n_{i,j}-\ell_{i,j}\sigma}}\, \mathrm{d}v$$

$$= \frac{1}{(\theta+1)_{N_i-1}} \frac{1}{\Gamma\left(\frac{\theta}{\sigma}+1\right)} \prod_{j=1}^k \tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\, \mathscr{C}(n_{i,j},\ell_{i,j};\sigma)$$

$$\times \int_0^\infty v^{\theta+\bar\ell_{i\bullet}\sigma-1}\, \mathrm{e}^{-v^\sigma}\, \mathrm{e}^{-\int_{\mathbb{X}}[(v+f_i(x))^\sigma - v^\sigma]\tilde p_0(\mathrm{d}x)} \prod_{j=1}^k \frac{1}{\left(1+\frac{f_i(X_j^*)}{v}\right)^{n_{i,j}-\ell_{i,j}\sigma}}\, \mathrm{d}v$$

$$= \frac{\prod_{r=1}^{\bar\ell_{i\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_i-1}} \prod_{j=1}^k \tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\, \frac{\mathscr{C}(n_{i,j},\ell_{i,j};\sigma)}{\sigma^{\ell_{i,j}}}$$

$$\times \int_0^\infty h_i(v) \left(\mathbb{E}_v\, \mathrm{e}^{-\tilde\mu_i^*(f_i)}\right) \left(\prod_{j=1}^k \mathbb{E}_v \mathrm{e}^{-H_{i,j} f_i(X_j^*)}\right) \mathrm{d}v$$

where h_i is the density function of a random variable V_i such that $V_i^\sigma \sim \mathrm{Ga}(\bar\ell_{i\bullet}+\theta/\sigma;1)$ and $\tilde\mu_i^*$ is, conditionally on V_i and on $\tilde\mu_0$, a generalized gamma CRM with parameters (V_i,σ) and base measure $\tilde p_0 = \tilde\mu_0/\tilde\mu_0(\mathbb{X})$. Moreover, the $H_{i,j}$'s are independent gamma random variables whose parameters are $n_{i,j}-\ell_{i,j}\sigma$ and V_i. If one, now, notes that

$$\mathbb{E}\Big[\prod_{i=1}^d \prod_{j=1}^k \tilde p_i^{n_{i,j}}(A_{j,\varepsilon}) \,\Big|\, \boldsymbol{T}, \tilde\mu_0\Big] = \frac{\prod_{r=1}^{\bar\ell_{i\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_i-1}} \prod_{j=1}^k \tilde p_0^{\ell_{i,j}}(A_{j,\varepsilon})\, \frac{\mathscr{C}(n_{i,j},\ell_{i,j};\sigma)}{\sigma^{\ell_{i,j}}}$$

which, combined with the result proved in Theorem 11, yields the result. □

A.13 Proof of Theorem 13

Let $\tilde\mu_\sigma$ denote a σ–stable CRM and, as in (9), $\tilde\mu_{\sigma,\theta}$ indicates a random measure whose normalization yields a Pitman–Yor process with parameters (σ,θ). From Theorem 11, one has that

$$\mathbb{E}\Big[\mathrm{e}^{-\eta_0^*(f)}\Big] = \frac{\sigma_0}{\Gamma\left(\frac{\theta_0}{\sigma_0}+k\right)} \int_0^\infty v^{\theta_0+k\sigma_0-1}\, \mathrm{e}^{-v^{\sigma_0}}\, \mathrm{e}^{-\int_{\mathbb{X}}[(v+f(x))^{\sigma_0}-v^{\sigma_0}]\,P_0(\mathrm{d}x)}\, \mathrm{d}v$$

$$= \frac{\sigma_0}{\Gamma\left(\frac{\theta_0}{\sigma_0}+k\right)} \int_0^\infty v^{\theta_0+k\sigma_0-1} \left(\mathbb{E}\, \mathrm{e}^{-v\tilde\mu_{\sigma_0}(\mathbb{X})-\tilde\mu_{\sigma_0}(f)}\right) \mathrm{d}v$$

$$= \frac{\sigma_0\,\Gamma(\theta_0+k\sigma_0)}{\Gamma\left(\frac{\theta_0}{\sigma_0}+k\right)} \left(\mathbb{E}\, \mathrm{e}^{-\tilde\mu_{\sigma_0}(f)} \{\tilde\mu_{\sigma_0}(\mathbb{X})\}^{-\theta_0-k\sigma_0}\right)$$

$$= \mathbb{E}\, \mathrm{e}^{-\tilde\mu_{\sigma_0,\theta_0+k\sigma_0}(f)}$$

Hence, one can conclude that $\eta_0^* \overset{\mathrm{d}}{=} \tilde\mu_{\sigma_0,\theta_0+k\sigma_0}$ and

$$\tilde p_0 \,|\, (\boldsymbol{X}, \boldsymbol{T}, V_0) \;\overset{\mathrm{d}}{=}\; \sum_{j=1}^k W_{j,V_0}\, \delta_{X_j^*} + W_{k+1,V_0}\, \tilde p_{\sigma_0,\theta_0+k\sigma_0}$$

where $\tilde p_{\sigma_0,\theta_0+k\sigma_0} \sim \mathrm{PY}(\sigma_0, \theta_0 + k\sigma_0; P_0)$. Moreover $W_{j,V_0} = I_j/(\eta_0^*(\mathbb{X}) + \sum_{i=1}^k I_i)$, for any $j = 1, \ldots, k$, and $W_{k+1,V_0} = \eta_0^*(\mathbb{X})/(\eta_0^*(\mathbb{X})+\sum_{i=1}^k I_i)$. If f_v is the density of the vector $(W_{1,V_0}, \ldots, W_{k,V_0})$ on the k–dimensional simplex Δ_k, we would like to determine

$$f(w_1, \ldots, w_k) = \int_0^\infty f_v(w_1, \ldots, w_k)\, \frac{\sigma_0}{\Gamma\Big(\frac{\theta_0}{\sigma_0} + k\Big)}\, v^{\theta_0+k\sigma_0-1}\, \mathrm{e}^{-v^{\sigma_0}}\, \mathrm{d}v.$$

To this end, we denote by h_v the density function of $\eta_0^*(\mathbb{X})$ and, using independence, the vector $(I_1, \ldots, I_k, \eta_0^*(\mathbb{X}))$ has density given by

$$f_v^*(x_1, \ldots, x_k, t) = h_v(t)\, \frac{v^{|\boldsymbol{\ell}|-k\sigma_0}}{\prod_{j=1}^k \Gamma(\bar\ell_{\bullet j} - \sigma_0)}\, \mathrm{e}^{-v\sum_{i=1}^k x_i} \prod_{j=1}^k x_j^{\bar\ell_{\bullet j}-\sigma_0-1}.$$

If $W_{V_0} = \sum_{i=1}^k I_j + \eta_0^*(\mathbb{X})$, a simple vector transformation yields a density function for the vector $(W_{1,V_0}, \ldots, W_{k,V_0}, W_{V_0})$

$$f_v(w_1, \ldots, w_k, w) = \frac{v^{|\boldsymbol{\ell}|-k\sigma_0} \prod_{j=1}^k w_j^{\bar\ell_{\bullet j}-\sigma_0-1}}{\prod_{j=1}^k \Gamma(\bar\ell_{\bullet j} - \sigma_0)} w^{|\boldsymbol{\ell}|-k\sigma_0}\, \mathrm{e}^{-vw|\boldsymbol{w}|}\, h_v(w(1-|\boldsymbol{w}|))$$

where $|\boldsymbol{w}| = \sum_{i=1}^k w_i$. From this, an expression for the density of $(W_{1,V_0}, \ldots, W_{k,V_0})$ easily follows and it turns out to be equal to

$$\begin{aligned} f_v(w_1, \ldots, w_k) &= \frac{v^{|\boldsymbol{\ell}|-k\sigma_0} \prod_{j=1}^k w_j^{\bar\ell_{\bullet j}-\sigma_0-1}}{\prod_{j=1}^k \Gamma(\bar\ell_{\bullet j} - \sigma_0)}\, \frac{1}{(1-|\boldsymbol{w}|)^{|\boldsymbol{\ell}|-k\sigma_0+1}} \int_0^\infty s^{|\boldsymbol{\ell}|-k\sigma_0}\, \mathrm{e}^{-\frac{vs|\boldsymbol{w}|}{1-|\boldsymbol{w}|}}\, h_v(s)\, \mathrm{d}s \\ &= \frac{v^{|\boldsymbol{\ell}|-k\sigma_0} \prod_{j=1}^k w_j^{\bar\ell_{\bullet j}-\sigma_0-1}}{\prod_{j=1}^k \Gamma(\bar\ell_{\bullet j} - \sigma_0)}\, \frac{1}{(1-|\boldsymbol{w}|)^{|\boldsymbol{\ell}|-k\sigma_0+1}} \Big(\mathbb{E}(\eta_0^*(\mathbb{X}))^{|\boldsymbol{\ell}|-k\sigma_0}\, \mathrm{e}^{-\frac{v|\boldsymbol{w}|}{1-|\boldsymbol{w}|}\,\eta_0^*(\mathbb{X})}\Big). \end{aligned}$$

Since η_0^* is a generalized gamma CRM with parameters (σ_0, V_0) and base measure P_0, its probability distributions $\mathbb{P}^*$ is absolutely continuous with respect to the probability distribution $\mathbb{P}_{\sigma_0}$ of a σ_0–stable CRM and

$$\frac{\mathrm{d}\mathbb{P}^*}{\mathrm{d}\mathbb{P}_{\sigma_0}}(m) = \exp\{-vm(\mathbb{X}) + v^{\sigma_0}\},$$

then

$$\mathbb{E}(\eta_0^*(\mathbb{X}))^{|\boldsymbol{\ell}|-k\sigma_0}\, \mathrm{e}^{-\frac{v|\boldsymbol{w}|}{1-|\boldsymbol{w}|}\,\eta_0^*(\mathbb{X})} = \mathrm{e}^{v^{\sigma_0}}\, \mathbb{E}\,(\tilde\mu_{\sigma_0}(\mathbb{X}))^{|\boldsymbol{\ell}|-k\sigma_0}\, \mathrm{e}^{-\frac{v}{1-|\boldsymbol{w}|}\,\tilde\mu_{\sigma_0}(\mathbb{X})}$$

In view of this, one can now marginalize f_v with respect to v and and obtain a density of $(W_1, \ldots, W_k)$. Indeed, one has

$$f(w_1, \ldots, w_k) = \frac{\sigma_0}{\Gamma\Big(\frac{\theta_0}{\sigma_0} + k\Big)} \int_0^\infty v^{\theta_0+k\sigma_0-1}\, \mathrm{e}^{-v^{\sigma_0}}\, f_v(w_1, \ldots, w_k)\, \mathrm{d}v$$

$$= \frac{\sigma_0}{\Gamma\left(\frac{\theta_0}{\sigma_0}+k\right)} \frac{\prod_{j=1}^k w_j^{\bar{\ell}_{\bullet j}-\sigma_0-1}}{\prod_{j=1}^k \Gamma(\bar{\ell}_{\bullet j}-\sigma_0)} \frac{1}{(1-|\boldsymbol{w}|)^{|\boldsymbol{\ell}|-k\sigma_0+1}}$$

$$\times\ \mathbb{E}\left(\tilde{\mu}_{\sigma_0}(\mathbb{X})\right)^{|\boldsymbol{\ell}|-k\sigma_0} \int_0^\infty v^{\theta_0+|\boldsymbol{\ell}|-1}\, \mathrm{e}^{-\frac{v}{1-|\boldsymbol{w}|}\tilde{\mu}_{\sigma_0}(\mathbb{X})} \mathrm{d}v$$

$$= \frac{\sigma_0\, \Gamma(\theta_0+|\boldsymbol{\ell}|)}{\Gamma\left(\frac{\theta_0}{\sigma_0}+k\right)} \frac{\left(\mathbb{E}\left(\tilde{\mu}_{\sigma_0}(\mathbb{X})\right)^{-\theta_0-k\sigma_0}\right)}{\prod_{j=1}^k \Gamma(\bar{\ell}_{\bullet j}-\sigma_0)} (1-|\boldsymbol{w}|)^{\theta_0+k\sigma_0-1} \prod_{j=1}^k w_j^{\bar{\ell}_{\bullet j}-\sigma_0-1}$$

$$= \frac{\Gamma(\theta_0+|\boldsymbol{\ell}|)}{\Gamma(\theta_0+k\sigma_0)\prod_{j=1}^k \Gamma(\bar{\ell}_{\bullet j}-\sigma_0)} (1-|\boldsymbol{w}|)^{\theta_0+k\sigma_0-1} \prod_{j=1}^k w_j^{\bar{\ell}_{\bullet j}-\sigma_0-1}$$

and this completes the proof of the posterior characterization of $\tilde{p}_0$.
Finally, one can proceed in a similar fashion in order to prove (25) □

B Blackwell–MacQueen urn scheme for Hierarchical Pitman–Yor processes

Here we specialize the algorithm described in Section 6.1 to hierarchies of Pitman–Yor processes. Hence we are assuming that $\boldsymbol{X}$ are partially exchangeable as in (1), where the prior Q_d is characterized as follows

$$\tilde{p}_i \,|\, \tilde{p}_0 \overset{\text{iid}}{\sim} \mathrm{PY}(\sigma,\theta;\tilde{p}_0) \quad (i=1,\ldots,d), \qquad \tilde{p}_0 \sim \mathrm{PY}(\sigma_0,\theta_0;P_0)$$

The same notations as in Section 6.1 are used. The determination of the full conditionals follows immediately from the augmented pEPPF

$$\Pi_{k+j}^{(n+m)}(\boldsymbol{n}_1,\cdots,\boldsymbol{n}_d;\boldsymbol{\ell},\boldsymbol{q})$$
$$= \frac{\prod_{r=1}^{k+j-1}(\theta_0+r\sigma_0)}{(\theta_0+1)_{|\boldsymbol{\ell}|-1}} \prod_{t=1}^{k+j}(1-\sigma_0)_{\bar{\ell}_{\bullet t}-1} \frac{\prod_{r=1}^{\bar{\ell}_{1\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_1+m_1-1}} \prod_{v=1}^{k+j}\prod_{t=1}^{\ell_{1,v}}(1-\sigma)_{q_{1,v,t}-1} \tag{38}$$
$$\times \frac{\prod_{r=1}^{\bar{\ell}_{2\bullet}-1}(\theta+r\sigma)}{(\theta+1)_{N_2+m_2-1}} \prod_{v=1}^{k+j}\prod_{t=1}^{\ell_{2,v}}(1-\sigma)_{q_{2,v,t}-1}$$

with the convention $(1-\sigma)_{-1}\equiv 1$. Based on (38) one can devise a Gibbs sampler that generates $(T_{i,1},\ldots,T_{i,N_i})$, for $i=1,2$, and $(X_{i,N_i+r},T_{i,N_i+r})$, for $r=1,\ldots,m_i$ and $i=1,2$, from their respective full conditionals. Details are provided for $i=1$, the case $i=2$ being identical with the appropriate adaptations.

(1) At $t=0$, start from an initial configuration $X_{i,N_i+1}^{(0)},\ldots,X_{i,N_i+m_i}^{(0)}$ and $T_{i,1}^{(0)},\ldots,T_{i,N_i+m_i}^{(0)}$, for $i=1,2$.

(2) At iteration $t\geq 1$

(2.a) With $X_{1,r} = X_h^*$ generate latent variables $T_{1,r}^{(t)}$, for $r = 1, \dots, N_i$, from

$$\mathbb{P}(T_{1,r} = \text{“new”} \mid \cdots) \propto w_{h,r} \frac{(\theta + \overline{\ell}_{1\bullet}^{-r} \sigma)}{(|\boldsymbol{\ell}^{-r}| + \theta_0)}$$

$$\mathbb{P}(T_{1,r} = T_{1,h,\kappa}^{*,-r} \mid \cdots) \propto (q_{1,h,\kappa}^{-r} - \sigma) \qquad \text{for } \kappa = 1, \dots, \ell_{1,h}^{-r}$$

where $w_{h,r} = \overline{\ell}_{\bullet h}^{-r} - \sigma_0$ if $\overline{\ell}_{\bullet h}^{-r} > 0$ and $w_{h,r} = 1$ otherwise. Moreover, $T_{1,h,1}^{*,-r}, \dots, T_{1,h,\ell_{1,h}^{-r}}^{*,-r}$ are the tables at the first restaurant where the h-th dish is served, after the removal of $T_{1,r}$.

(2.b) For $r = 1, \dots, m_1$, generate $(X_{N_i+r}^{(t)}, T_{N_i+r}^{(t)})$ from the following predictive distributions

$$\mathbb{P}(X_{1,r} = \text{“new”}, T_{1,r} = \text{“new”} \mid \cdots) = \frac{(\theta_0 + (k + j^{-r})\sigma_0)}{(\theta + N_1 + m_1 - 1)} \frac{(\theta + \overline{\ell}_{1\bullet}^{-r} \sigma)}{(\theta_0 + |\boldsymbol{\ell}^{-r}|)}$$

while, for any $h = 1, \cdots, k + j^{-r}$ and $\kappa = 1, \cdots, \ell_{1,h}^{-r}$,

$$\mathbb{P}(X_{1,r} = X_h^{*,-r}, T_{1,r} = \text{“new”} \mid \cdots) = \frac{(\overline{\ell}_{\bullet h}^{-r} - \sigma_0)}{(\theta + N_1 + m_1 - 1)} \frac{(\theta + \overline{\ell}_{1\bullet}^{-r} \sigma)}{(\theta_0 + |\boldsymbol{\ell}^{-r}|)}$$

$$\mathbb{P}(X_{1,r} = X_h^{*,-r}, T_{1,r} = T_{1,h,\kappa}^{*,-r} \mid \cdots) = \frac{q_{1,h,\kappa}^{-r} - \sigma}{\theta + N_1 + m_1 - 1} \mathbb{1}_{\{n_{i,h}^{-r} > 0\}}$$

recalling that $X_h^{*,-r}$, for $h = 1, \dots, k + j^{-r}$ denote the distinct dishes in the whole franchise after the removal of the r–th observation, while the condition $n_{i,h}^{-r} > 0$ entails that the h–th dish is served in the i–th restaurant.

References

[1] Adams, M., Kelley, J., Polymeropoulos, M., Xiao, H., Merril, C., Wu, A., Olde, B., Moreno, R., Kerlavage, A., McCombe, W. and Venter, J. (1991). Complementary DNA sequencing: Expressed Sequence Tags and human genome project. *Science* **252**, 1651–1656.

[2] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.* **3**, 993–1022.

[3] Camerlenghi, F., Lijoi, A. and Prünster, I. (2017). Bayesian prediction with multiple–sample information. *J. Multivariate Anal.*, **156**, 18–28.

[4] Caron, F., Davy, M., Doucet, A. (2007). Generalized Pólya urn for time-varying Dirichlet process mixtures. *23rd Conference on Uncertainty in Artificial Intelligence (UAI'2007)*, Vancouver, Canada.

[5] Caron, F., Teh, Y.W., Murphy, T.B. (2014). Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. *Ann. Appl. Statist.* **8**, 1145–1181.

[6] Charalambides, C.A. (2005). *Combinatorial methods in discrete distributions.* Hoboken, NJ: Wiley.

[7] Constantines G.M., Savits T.H. (1996). A multivariate version of the Faa di Bruno formula. *Trans. Amer. Math. Soc.* **348**, 503–520.

[8] Daley, D.J. and Vere–Jones, D. (2008). *An introduction to the theory of point processes. Volume II*, Springer, New York.

[9] De Blasi, P., Favaro, S., Lijoi, A., Mena, R.H., Ruggiero, M. and Prünster, I. (2015). Are Gibbs–type priors the most natural generalization of the Dirichlet process? *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 212–229.

[10] Dunson, D.B. (2010). Nonparametric Bayes applications to biostatistics. In *Bayesian Nonparametrics*, pp. 223–273. Cambridge University Press, Cambridge.

[11] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87–112.

[12] Favaro, S., Lijoi, A., Mena, R.H. and Prünster, I. (2009). Bayesian nonparametric inference for species variety with a two parameter Poisson-Dirichlet process prior. *J. Roy. Statist. Soc. Ser. B* **71**, 993–1008.

[13] Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.

[14] Ferguson, T.S. and Klass, M.J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.* **43**, 1634–1643.

[15] Foti, N.J. and Williamson, S.A. (2015). A survey of non-exchangeable priors for Bayesian nonparametric models. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 359–371.

[16] Gasthaus, J. and Teh, Y.W. (2010). Improvements to the sequence memoizer. *Advances in Neuronal Information Processing Systems* **23**.

[17] Gnedin, A.V. and Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. *Zap. Nauchn. Sem. S.Peterb. Otdel. Mat. Inst. Steklov (POMI)* **325**, 83–102.

[18] Griffin, J. E., Kolossiatis, M. and Steel, M. F. J. (2013) Comparing distributions by using dependent normalized random-measure mixtures. *J. R. Stat. Soc. Ser. B.* **75**, 499–529.

[19] Hardy M. (2006). Combinatorics of partial derivatives. *Electron. J. Comb.* **13,** 1–13

[20] James, L.F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33**, 1771–1799.

[21] James, L.F. (2006). Poisson calculus for spatial neutral to the right processes. *Ann. Statist.* **34**, 416–440.

[22] James, L.F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scand. J. Statist.* **33**, 105–120.

[23] James, L.F., Lijoi, A. and Prünster, I. (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Statist.* **36**, 76–97.

[24] Kallenberg, O. (2005) *Probabilistic Symmetries and Invariance Principles.* Springer, New York.

[25] Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.

[26] Kingman, J.F.C. (1982). The coalescent. *Stochastic Process. Appl.* **13**, 235–248.

[27] Kingman, J.F.C. (1993). *Poisson processes.* Oxford University Press, Oxford.

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

[29] Lijoi, A., Mena, R.H. and Prünster, I. (2007a). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. Roy. Statist. Soc. Ser. B* **69**, 715–740.

[30] Lijoi, A., Nipoti, B. and Prünster, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20**, 1260–1291.

[31] Lijoi, A., Nipoti, B. and Prünster, I. (2014). Dependent mixture models: clustering and borrowing information. *Comput. Statist. Data Anal.* **71**, 417–433.

[32] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*, pp. 80–136. Cambridge University Press, Cambridge.

[33] Lijoi, A., Prünster, I. and Walker, S.G. (2008a). Bayesian nonparametric estimators derived from conditional Gibbs structures. *Ann. Appl. Probab.* **18**, 1519–1547.

[34] MacEachern, S.N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the SBSS*. Alexandria: American Statistical Association, 50-55.

[35] MacEachern, S.N. (2000). Dependent Dirichlet processes. *Technical Report.* Department of Statistics, Ohio State University.

[36] Nguyen, X. (2010). Inference of global clusters from locally distributed data. *Bayesian Analysis* **5**, 817–846.

[37] Nguyen, X. (2016). Borrowing strength in hierarchical Bayes: convergence of the Dirichlet base measure. *Bernoulli* **22**, 1535–1571.

[38] Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T.S. Ferguson, L.S. Shapley and J.B. MacQueen, Eds.). IMS Lecture Notes Monogr. Ser., Vol. **30**. IMS, Hayward, 245–267.

[39] Pitman, J. (2006). *Combinatorial stochastic processes.* Ecole d'Eté de Probabilités de Saint-Flour XXXII. Lecture Notes in Mathematics N. 1875, Springer, New York.

[40] Pitman, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.

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

[42] Teh, Y.W., Jordan, M.I., Beal, M.J. and Blei, D.M. (2006). Hierarchical Dirichlet processes. *J. Amer. Statist. Assoc.* **101**, 1566–1581.

[43] Teh, Y.W., Jordan, M.I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics*, pp. 158-207, Cambridge Univ. Press, Cambridge.

[44] Walker, S. and Damien, P. (2000). Representations of Lévy processes without Gaussian components. *Biometrika.* **87**, 477–483.

[45] Wood, F., Gasthaus, J., Archambeau, C., James, L.F. and Teh, Y.W. (2011). The sequence memoizer. *Communications ACM* **54**, 91–98.

[46] Zhou, M. and Carin, L. (2015). Negative binomial process count and mixture modeling. *IEEE Trans. Pattern Anal. Mach. Intell.* **37**, 307-320.