arXiv:2503.24004v2 [math.ST] 14 Jul 2025

Multivariate Species Sampling Models

Beatrice Franzolini, Antonio Lijoi, Igor Prünster
Bocconi Institute for Data Science and Analytics, Bocconi University

Giovanni Rebaudo
ESOMAS Department, University of Torino

Abstract

Species sampling processes have long served as the fundamental framework for modeling random discrete distributions and exchangeable sequences. However, data arising from distinct but related sources require a broader notion of probabilistic invariance, making partial exchangeability a natural choice. Countless models for partially exchangeable data, collectively known as dependent nonparametric priors, have been proposed. These include hierarchical, nested and additive processes, widely used in statistics and machine Learning. Still, a unifying framework is lacking and key questions about their underlying learning mechanisms remain unanswered.

We fill this gap by introducing multivariate species sampling models, a new general class of nonparametric priors that encompasses most existing finite- and infinite-dimensional dependent processes. They are characterized by the induced partially exchangeable partition probability function encoding their multivariate clustering structure. We establish their core distributional properties and analyze their dependence structure, demonstrating that borrowing of information across groups is entirely determined by shared ties. This provides new insights into the underlying learning mechanisms, offering, for instance, a principled rationale for the previously unexplained correlation structure observed in existing models. Beyond providing a cohesive theoretical foundation, our approach serves as a constructive tool for developing new models and opens novel research directions to capture richer dependence structures beyond the framework of multivariate species sampling processes.

Keywords: Bayesian nonparametrics, Dependent nonparametric prior, Hierarchical process, Multi-armed bandit, Partial exchangeability, Random partition.

1 Introduction

A fundamental homogeneity assumption in Bayesian inference is the (infinite) exchangeability of observables, which corresponds to distributional invariance with respect to permutations of the data. According to de Finetti's representation Theorem, there is a one-to-one correspondence between an exchangeable sequence $(X_i)_{i\geq 1}$ and a random probability measure, conditionally on which the X_i's are independent and identically distributed (i.i.d.). This foundational result supports the Bayesian framework, based on likelihood and prior, via a probabilistic symmetry assumption on the data. At the same time, it establishes a conceptual bridge to the classical i.i.d. assumption in frequentist inference. In a parametric setup, the random probability measure associated with $(X_i)_{i\geq 1}$ is indexed by a finite-dimensional parameter, whereas in a nonparametric setting, no such restriction is imposed. A cornerstone of the latter is represented by the Dirichlet process (DP) introduced in Ferguson (1973). Its full weak support property implies remarkable flexibility compared to parametric counterparts, and several popular Bayesian nonparametric (BNP) models can be seen as extensions of the DP itself.

In a seminal work, Pitman (1996) introduced a unifying framework for studying almost surely discrete random probability measures under the assumption that weights and locations are independent. The resulting class, known as species sampling processes (SSPs) (Ghosal et al., 2017), includes the DP as a special case and satisfies several structural properties that have been pivotal to understanding and constructing discrete priors for modeling exchangeable data. Entire classes of popular nonparametric and parametric priors, such as homogeneous normalized random measures with independent increments (Regazzini et al., 2003; James, Lijoi, et al., 2009), Gibbs-type priors (Gnedin and Pitman, 2006; De Blasi et al., 2015), and stick-breaking processes (Ishwaran et al., 2001; Gil–Leyva et al., 2023) fall within the framework of SSPs, and henceforth also their notable special cases, which include Pitman-Yor (Pitman and Yor, 1997), normalized inverse Gaussian (Lijoi, Mena, et al., 2005) and normalized generalized gamma (Lijoi, Mena, et al., 2007b) processes. Moreover, relevant special cases also include finite-dimensional processes such as the finite Dirichlet Multinomial (Green et al., 2001) and mixture of finite-dimensional processes with a prior on the number of locations (Nobile, 1994; Richardson et al., 1997; Nobile and Fearnside, 2007; Gnedin, 2010; De Blasi et al., 2013; Miller et al., 2018). A summary of J. Pitman's theory on univariate SSP (Pitman, 1996) can be found in Section S1 of the Supplementary Material.

However, exchangeability is often too restrictive an assumption in applied settings. The pioneering contributions of MacEachern (1999) and MacEachern (2000) opened a new research line in both the statistics and machine learning literature, whose goal is to develop models that accommodate heterogeneity across data sources or experimental conditions. More specifically, when data originate from J distinct populations, such as in meta-analysis, topic modeling or multi-center studies, the exchangeability assumption becomes overly restrictive, since it fails to account for heterogeneity across distinct groups. Conversely, assuming independence between

populations precludes information sharing across experiments, which is often a key goal in multi-sample studies (see, for instance, Woodcock et al., 2017; Chen et al., 2019; Ouma et al., 2022; Su et al., 2022). A natural compromise between these extremes is the probabilistic framework of partial exchangeability (de Finetti, 1938), which entails exchangeability within but not across different populations, while still allowing for dependence among them. Consider a random array $\mathbf{X}$ with J rows and infinite columns. Then $\mathbf{X}$ is partially exchangeable if and only if its distribution is invariant with respect to finite permutations within each row but not across columns. This means that elements within each population, i.e., belonging to the same row, are exchangeable, but permuting elements across populations, i.e., belonging to different rows, would alter the distribution of $\mathbf{X}$. For instance, suppose to have partially exchangeable binary data from $J = 2$ groups and that we observe $X_{1,1}$ and $X_{1,2}$ from the first group, and $X_{2,1}$ from the second. Then, $\mathbb{P}[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1] = \mathbb{P}[X_{1,2} = 1, X_{1,1} = 0, X_{2,1} = 1]$ as this is a permutation within group 1. However, it is possible that $\mathbb{P}[X_{1,1} = 1, X_{1,2} = 0, X_{2,1} = 1] \neq \mathbb{P}[X_{1,1} = 1, X_{2,1} = 0, X_{1,2} = 1]$, since invariance with respect to permutations across groups is not preserved. Similarly to the exchangeable case, partial exchangeability implies the existence of a vector of (dependent) random probability measures $(P_1, \ldots, P_J)$, such that $X_{j,i} \mid P_1, \ldots, P_J \overset{ind}{\sim} P_j$, for $i \geq 1$ and $j = 1, \ldots, J$. From a Bayesian perspective, modeling a partially exchangeable array is equivalent to defining a prior distribution for a vector of dependent probability measures. Countless approaches have been proposed in the literature and several success stories have been recorded. Notable instances are hierarchical DPs (Teh et al., 2006), hierarchical normalized completely random measures (Camerlenghi, Lijoi, Orbanz, et al., 2019), hierarchical species sampling models (Bassetti et al., 2020), nested constructions (Rodríguez et al., 2008; Camerlenghi, Dunson, et al., 2019), additive constructions (Müller et al., 2004; Lijoi, Nipoti, et al., 2014), copula constructions (Leisen et al., 2011), normalized compound random measures (Griffin et al., 2017), normalized completely random vectors (Lijoi, Nipoti, et al., 2014; Catalano, Lijoi, et al., 2021), single-atoms dependent processes (MacEachern, 1999; MacEachern, 2000; Quintana et al., 2022), compositions of the some of the previous (Camerlenghi, Dunson, et al., 2019; Beraha et al., 2021; Lijoi, Prünster, et al., 2023; Balocchi, George, et al., 2023; Denti et al., 2023), and many others (e.g. Horiguchi et al., 2024; Yan et al., 2023; Bi et al., 2023; Lee et al., 2025).

The main goal of these approaches is to flexibly model dependence across populations to facilitate the sharing of information. This is achieved by defining a collection of dependent latent random probabilities $(P_1, \ldots, P_J)$, which in turn induce dependence among the observables $\mathbf{X}$. Clearly, having a way to quantify the dependence between these random probability measures is crucial to understanding and guiding such modeling strategies. The most widely adopted measure of inter-population dependence is the pairwise correlation between P_j and P_k, for $j \neq k$, evaluated on the same set A, namely

$$\mathbb{C}\text{or}[P_j(A), P_k(A)]. \tag{1}$$

The main reasons this measure has become the benchmark for quantifying dependence are

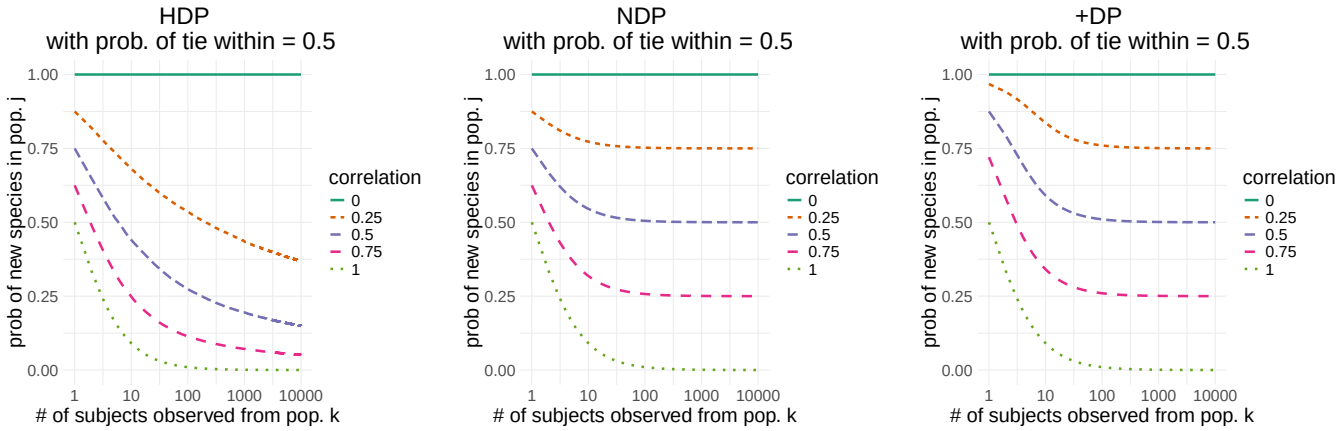

Figure 1: Probability of observing a new species at the first draw from population j, after n observations from population k, i.e., $\mathbb{P}[X_{j,1} \notin \{X_{k,1}, \ldots, X_{k,n}\}]$, plotted as a function of n. Results are shown (left to right) for hierarchical Dirichlet processes (HDP, Teh et al., 2006), nested Dirichlet processes (NDP, Rodríguez et al., 2008), and additive Dirichlet processes (+DP, Müller et al., 2004) for varying values of the correlation measure $\mathbb{C}\text{or}[P_j(A), P_k(A)]$.

threefold. First, for most models, it does not depend on the choice of the set A, allowing it to be interpreted as a global measure of dependence. Second, it is often computable either analytically or numerically, which greatly enhances its practical appeal. Third, although the correlation in (1) relies only on the first two moments, it effectively summarizes the dependence structure. On the one hand, it tends to agree with alternative, more complex measures of dependence that account for the infinite-dimensional nature of the P_i's whenever these alternatives can be computed (see Catalano, Lijoi, et al., 2021; Catalano, Lavenant, et al., 2024). On the other hand, different correlation values correspond to markedly different behaviours of the distribution of observable quantities; this is illustrated in Figure 1, which displays the probability of observing a new species for three popular dependent processes as the correlation varies. However, several fundamental questions remain open, and their resolution is crucial for establishing a rigorous foundation for correlation as a measure of dependence in this context. First, it is still obscure why (1) typically does not depend on the choice of the set A; related to this, one would like to understand what conditions on the prior ensure that (1) does not to depend on the set A. Second, the broader issue of how the properties of the latent random probability measures P_i's translate to the observable quantities is still unexplored. This can be decomposed into two key open questions in terms of correlation: (a) How does the correlation among the latent P_i's translate into dependence among observations? (b) Are correlations among the P_i's or observable quantities $\mathbf{X}$ reliable indicators of dependence? Specifically, are there models for which $\mathbb{C}\text{or}[P_j(A), P_k(A)] = 1$ if and only if $P_j = P_k$ almost surely (i.e., observations are exchangeable), and $\mathbb{C}\text{or}[P_j(A), P_k(A)] = 0$ if and only if $P_j \perp P_k$?

These questions point to an even more general theme: many structural properties appear to

be shared across several classes of dependent processes (e.g., additive, hierarchical, nested, or combinations thereof), and this naturally raises the question of whether it is possible to identify a unifying framework, one that would: (i) encompass most existing models; (ii) allow their structural properties to be studied in a unified way; (iii) provide a foundation for the principled development of new models.

In this work, we provide comprehensive answers to the open problems outlined above. Specifically, we introduce a new unifying framework for partially exchangeable data that encompasses most existing dependent nonparametric processes. The resulting class of nonparametric prior processes is termed multivariate species sampling processes (mSSPs), as they play the same foundational role for vectors $(P_1, \ldots, P_J)$ as species sampling processes (SSPs) do in the univariate case. We characterize mSSPs through their partially exchangeable partition probability function, which encodes the induced multivariate clustering structure. We show that BNP models currently used for partially exchangeable data belong to a notable subclass of mSSPs, which we refer to as *regular*, which enjoys additional appealing properties. We also analyze the dependence structure of these processes and prove that borrowing of information across groups is entirely determined by shared ties. This leads to new insights into the learning mechanisms, offering a principled explanation for the correlation structure discussed above. Beyond providing a cohesive theoretical foundation, our approach serves as a constructive tool for developing new models and opens new research directions aimed at capturing even richer dependence structures beyond the mSSP framework.
Finally, while mSSPs generalize SSPs, their essence lies in their multivariate nature: in particular, in the dependence induced across populations and, consequently, across elements of the vector $(P_1, \ldots, P_J)$. This feature is obviously absent in classical SSPs, and the structural unification of multivariate, infinite-dimensional objects represents a key innovation of this work.

The paper is structured as follows. mSSP and the notable subclass of *regular* mSSP are defined in Section 2. In Section 3, we derive expressions for marginal and mixed moments of these latent processes in terms of observable quantities. A substantial part of this section is devoted to analyzing the correlation between the random measures, and we provide the theoretical foundations for it to be regarded as the benchmark measure of dependence within the class of regular mSSPs, where uncorrelation implies independence. Sections 4 and 5 are devoted to the study of the random partitions induced by mSSPs and their corresponding predictive distributions, respectively. In Section 6, we compare the performance of different regular mSSPs in the context of a multi-armed bandit problem aimed at maximizing species discoveries, when sampling sequentially across multiple sites. Finally, Section 7 outlines future research directions. The Supplementary Material contains a review of univariate species sampling processes, all proofs, and further details on the application. Code to reproduce the experiments is available at `https://github.com/GiovanniRebaudo/MSSP`.

2 Multivariate species sampling processes

2.1 General multivariate species sampling processes

When extending a univariate random probability P to the multivariate setting involving a vector $(P_1, \ldots, P_J)$ of random probabilities, the species sampling framework of Pitman (1996) can be naturally generalized to multiple populations according to the following definition. Recall that $\boldsymbol{\pi} = (\pi_h)_{h\geq 1}$ is a sub-probability sequence if $\pi_h \in [0,1]$, for any h, and $\sum_{h\geq 1} \pi_h \leq 1$.

Definition 1. A vector of random probability measures $(P_1, \ldots, P_J)$ is a *multivariate species sampling process* (mSSP) if

$$P_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) P_0, \qquad \text{for } j = 1, \ldots, J,$$

where P_0 is a non-atomic (deterministic) distribution on a space $\mathbb{X}$, $\boldsymbol{\pi}_j = (\pi_{j,h})_{h\geq 1}$ is a random sub-probability sequence, for any j, and $\boldsymbol{\theta} = (\theta_h)_{h\geq 1}$ are i.i.d from P_0 and independent of $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_J) \sim \mathscr{L}_{\boldsymbol{\pi}}$. We write $(P_1, \ldots, P_J) \sim \text{mSSP}(\mathscr{L}_{\boldsymbol{\pi}}, P_0)$. Moreover, if $\sum_{h\geq 1} \pi_{j,h} \overset{a.s.}{=} 1$, for any j, $(P_1, \ldots, P_J)$ is said *proper*.

According to the standard terminology (see, for instance Ghosal et al., 2017), we refer to the elements in $\boldsymbol{\theta}$ as *atoms*, *labels*, *locations*, or *species*, interchangeably, and to the elements in $\boldsymbol{\pi}$ as the *weights* of the mSSP. The structural independence assumption underlying SSPs of independence between the locations and a single sequence of weights, is replaced by independence between the locations $(\theta_h)_{h\geq 1}$ and an array of weights $(\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_J)$; the dependence among the $\boldsymbol{\pi}_j$'s then induces a multivariate structure across populations. From Definition 1, the link between mSSPs and SSPs is apparent: it is indeed straightforward to prove that each coordinate of an mSSP is marginally an SSP. More generally, the class of mSSPs is closed under marginalization.

Proposition 1. *If* $(P_1, \ldots, P_J) \sim$ *mSSP and* $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, J\}$ *then* $(P_{j_1}, \ldots, P_{j_k}) \sim$ *mSSP.*

At first glance, Definition 1 might lead one to think of an mSSP as arising from several univariate SSPs sharing the same atoms, i.e., the atoms are given by the same sequence, $(\theta_h)_{h\geq 1}$ for each group. While this is certainly a possibility, the class of mSSPs is much more general. In fact, each of the $\pi_{j,h}$'s might be zero almost surely, resulting in mSSPs where the random probabilities $P_1, \ldots, P_J$ share only a handful or even none of the atoms with positive probability. Hence, from an intuitive point of view, one should think of an mSSP as arising from several SSPs, which *potentially* share atoms. Moreover, when investigating the pairwise dependence and quantifying the extent to which species are shared between two random probabilities, being able to distinguish between shared atoms (with positive probability) and idiosyncratic ones is often helpful. Definition 1 does not convey this information explicitly, which motivates

the need for an alternative representation. Consider the bivariate mSSP (P_1, P_2), which is obtained marginalizing a J-variate mSSP from Definition 1. Then, the following representation is equivalent (almost surely).

Proposition 2. *$(P_1, P_2) \sim$ mSSP iff (P_1, P_2) admits the following almost sure equivalent representation*

$$P_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}^{(1,2)} \delta_{\theta_{0,h}} + \sum_{h'\geq 1} \pi_{j,h'}^{(j)} \delta_{\theta_{j,h'}} + \pi_{j,0}^{(j)} P_0, \qquad \text{for } j = 1, 2. \tag{2}$$

with $\mathbb{P}[\pi_{1,h}^{(1,2)}\, \pi_{2,h}^{(1,2)} > 0] > 0$, for $h = 1, 2, \ldots$ and $\sum_{h\geq 1} \pi_{j,h}^{(1,2)} + \sum_{h'\geq 0} \pi_{j,h'}^{(j)} = 1$, for $j = 1, 2$, and where all the weights are independent of the atoms, which are distributed as $\theta_{j,h} \overset{iid}{\sim} P_0$, for $j = 0, 1, 2,\ \ h = 1, 2, \ldots$. By convention $\sum_{h=1}^{0} x_h = 0$, for any (x_h).

Importantly, and in contrast to Definition 1, we now require $\mathbb{P}[\pi_{1,h}^{(1,2)}\, \pi_{2,h}^{(1,2)} > 0] > 0$, which is clearly equivalent to $\mathbb{P}[\pi_{1,h}^{(1,2)} > 0, \pi_{2,h}^{(1,2)} > 0] > 0$, for any h in the first sum. This condition implies that the atoms in the first sum in (2) are common to both random elements with positive probability, thus singling out atoms shared with positive probability. Conversely, the terms in the second summation are almost surely specific to each P_j, meaning that the corresponding atoms cannot be shared across the two populations. Note that the non-atomicity of P_0 implies $\mathbb{P}(\theta_{j,h'} = \theta_{j',\ell'}) = 0$ for all $(j, h') \neq (j', \ell')$. Finally, $\pi_{j,0}^{(j)}$ is the cumulative frequency of almost surely non-shared species that are observed just one time in the infinite population sampled from P_j. The advantage of the representation in (2), compared to Definition 1, lies in the ability to immediately distinguish between shared and non-shared species. This representation naturally extends to general J-variate processes with $J \geq 2$, enabling us to distinguish between species shared across any subset of the P_j's. However, as J grows, the notation becomes increasingly cumbersome, so we omit the explicit formulation here. More importantly, representation (2) enables us to identify a notable subclass of mSSPs, which we term *regular* and examine in the next Section. Although this subclass encompasses J-variate processes $(P_1, \ldots, P_J)$, with $J \geq 2$, it suffices to adopt the representation in (2) for any pair (P_j, P_k) to fully characterize it. Finally, we define the pair consisting of a collection of random variables $\mathbf{X}$ and the mSSP from which these observations are drawn as follows.

Definition 2. A partially exchangeable array $\mathbf{X} = (X_{j,i} : i \in \mathbb{N}, j \in [J])$, for some $J \in \mathbb{N}$, follows a multivariate species sampling model (mSSM) if its de Finetti measure is an mSSP. That is, for every $j \in [J]$ and for every $i = 1, 2, \ldots$

$$X_{j,i} \mid (P_1, \ldots, P_J) \overset{\text{ind}}{\sim} P_j, \qquad (P_1, \ldots, P_J) \sim \text{mSSP}(\mathcal{L}_{\boldsymbol{\pi}}, P_0) \tag{3}$$

2.2 Regular multivariate species sampling processes

A notable subclass of mSSPs, which we term *regular*, arises by imposing a simple independence condition on the weights associated with the non-shared atoms. First, consider a bivariate

mSSP (P_1, P_2) and define

$$\boldsymbol{\pi}^{(j)} = \left(\frac{\pi^{(j)}_{j,h'}}{\sum_{\ell \geq 0} \pi^{(j)}_{j,\ell}} \right)_{h' \geq 0} \qquad \text{for } j = 1, 2$$

where the weights $\pi^{(j)}_{j,h'}$ are as defined in (2) and by convention $0/0 = 0$.

Definition 3. A bivariate mSSP (P_1, P_2) is *regular* (rmSSP) if $\boldsymbol{\pi}^{(1)} \perp \boldsymbol{\pi}^{(2)}$. A J-variate mSSP $(P_1, \ldots, P_J)$, with $J > 2$, is *regular* if (P_j, P_k) is a rmSSP for any $j, k \in [J]$, with $j \neq k$.

Intuitively, regularity requires that the relative frequencies of non-shared species are independent for each pair (P_j, P_k) of populations. Note that if either P_j, P_k, or both have no non-shared species in the representation (2), that is, if $\sum_{h' \geq 0} \pi^{(\ell)}_{\ell,h'} \overset{a.s.}{=} 0$, then regularity is trivially satisfied. From a statistical modeling standpoint, it is important to note that the independence condition required by Definition 3 is relatively mild and, in most applied contexts, reasonable. This condition implies that the relative frequencies of non-shared species, i.e., $\boldsymbol{\pi}^{(j)}$, should not influence the sharing of information across groups, once the total frequency $\sum_{k \geq 0} \pi^{(j)}_{j,k}$ of these idiosyncratic species has been accounted for. Given that these species are (almost surely) not shared among groups, this assumption seems very reasonable.

In the following, we focus specifically on rmSSPs and their use within BNP models, while leaving a broader probabilistic study of general mSSPs for future work. Special emphasis on the regular subclass is warranted for two main reasons. First, rmSSPs differ from non-regular mSSPs due to the distinctive dependence structure that enables a remarkable characterization in terms of correlation between the P_j's. This result does not extend to general mSSPs, highlighting a fundamental difference between regular and non-regular mSSPs. Second, the regular subclass is particularly relevant in statistics, as it encompasses all mSSPs currently employed in BNP. Some of these will be illustrated in the examples below.

Henceforth, $\text{DP}(\alpha, P_0)$ denotes the law of a Dirichlet process with concentration parameter α and base measure P_0 (Ferguson, 1973), and $\text{GEM}(\alpha)$ denotes a Griffiths-Engen-McCloskey distribution (Sethuraman, 1994). $\text{PYP}(\sigma, \alpha, P_0)$ stands for the law of a Pitman-Yor process with discount parameter σ, concentration parameter α and base measure P_0 (Pitman and Yor, 1997). $\text{CRM}(\rho, c, P_0)$ and $\text{NRMI}(\rho, c, P_0)$ indicate, respectively, the laws of a completely random measure and a normalized completely random measure with intensity ρ, total mass parameter c, and base measure P_0 (Regazzini et al., 2003; James, Lijoi, et al., 2009). $\text{GN}(\gamma, P_0)$ denotes the law of a Gnedin process with parameter γ, and base measure P_0 (Gnedin, 2010), $\text{DM}_M(\tau, P_0)$ is the law of a symmetric Dirichlet-Multinomial process with M number of categories, concentration parameter τ, and base measure P_0 (Richardson et al., 1997). Finally, $\text{SSP}(\mathcal{L}_{\boldsymbol{\pi}}, P_0)$ denotes the law of a (univariate) species sampling process with weights distribution defined by $\mathcal{L}_{\boldsymbol{\pi}}$ and base measure P_0 (Pitman, 1996). Recall that a concise account of SSPs, including their associated exchangeable probability partition function (EPPF), prediction rule, and notable special cases of SSPs, is provided in Section S.1 of the Supplement. In the following, P_0 always indicates a generic non-atomic deterministic distribution.

In order to highlight the comprehensive nature of mSSP and its subclass given by rmSSPs, we now show that several popular classes of dependent models are subclasses of rmSSPs and hence, a fortiori, mSSP.

Example 1 (Hierarchical processes)**.** Assume the distribution of $(P_1, \ldots, P_J)$ coincides with any of the hierarchical specifications listed in Table 1.

Table 1: Hierarchical processes (Teh et al., 2006; Camerlenghi, Lijoi, Orbanz, et al., 2019; Bassetti et al., 2020)

Hierarchical Dirichlet Process (HDP)	$P_j \mid Q \overset{iid}{\sim} \mathrm{DP}(\alpha, Q), \quad Q \sim \mathrm{DP}(\alpha_0, P_0)$
Hierarchical Pitman-Yor Process (HPY)	$P_j \mid Q \overset{iid}{\sim} \mathrm{PYP}(\sigma, \alpha, Q), \quad Q \sim \mathrm{PYP}(\sigma_0, \alpha_0, P_0)$
Hierarchical normalized completely random measure (HNRMI)	$P_j \mid Q \overset{iid}{\sim} \mathrm{NRMI}(\rho, c, Q), \quad Q \sim \mathrm{NRMI}(\rho_0, c_0, P_0)$
Hierarchical Dirichlet-Multinomial (HDM):	$P_j \mid Q \overset{iid}{\sim} \mathrm{DM}_M(\tau, Q), \quad Q \sim \mathrm{DM}_{M_0}(\tau_0, P_0)$
Hierarchical Gnedin Process (HGN)	$P_j \mid Q \overset{iid}{\sim} \mathrm{GN}(\gamma, Q), \quad Q \sim \mathrm{GN}(\gamma_0, P_0)$
Hierarchical Species Sampling Process (HSSP)	$P_j \mid Q \overset{iid}{\sim} \mathrm{SSP}(\mathcal{L}_{\pi,j}, Q), \quad Q \sim \mathrm{SSP}(\mathcal{L}_{\pi,0}, P_0)$

Since $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$, for any $j \in [J]$, $(P_1, \ldots, P_J)$ is trivially a rmSPP.

Remark 1. SSPs are defined in terms of a non-atomic base measure P_0 (Pitman, 1996). Hence, writing $\mathrm{SSP}(\mathcal{L}_{\pi,j}, Q)$ in Table 1 represents an abuse of notation. However, since the extension to the case where the "base measure" Q can be an atomic discrete random measure is immediate, we will also employ it in the sequel. Furthermore, after marginalizing out Q, we are back to the definition of mSSP with a non-atomic base measure.

Example 2 (Nested processes)**.** Assume the distribution of $(P_1, \ldots, P_J)$ corresponds to any of the nested constructions listed in Table 2.

Table 2: Nested processes (Rodríguez et al., 2008; Zuanetti et al., 2018)

Nested Dirichlet Process (NDP)	$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \sim \mathrm{DP}(\alpha, \mathrm{DP}(\beta, P_0))$
Nested Pitman-Yor Process (NPY)	$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \sim \mathrm{PYP}(\sigma_\alpha, \alpha, \mathrm{PYP}(\sigma_\beta, \beta, P_0))$
Nested Dirichlet-Multinomial (NDM)	$Q \overset{iid}{\sim} Q, \quad Q \sim \mathrm{DM}_{M_\alpha}(\tau_\alpha, \mathrm{DM}_{M_\beta}(\tau_\beta, P_0))$
Nested Gnedin Process (NGN)	$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \sim \mathrm{GN}(\gamma_\alpha, \mathrm{GN}(\gamma_\beta, P_0))$
Nested Species Sampling Process (NSSP)	$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \sim \mathrm{SSP}(\mathcal{L}_{\pi,0}, \mathrm{SSP}(\mathcal{L}_{\boldsymbol{\pi}}, P_0))$

Since $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$, for any $j \in [J]$, $(P_1, \ldots, P_J)$ is a rmSPP.

Example 3 (Additive processes)**.** Assume the distribution of $(P_1, \ldots, P_J)$ coincides with any of the additive specifications listed in Table 3.

Table 3: Additive processes (Müller et al., 2004)

Additive Dirichlet Process (+DP)	$P_j = \epsilon_j\, Q_0 + (1-\epsilon_j) Q_j, \quad Q_j \overset{ind}{\sim} \mathrm{DP}(\alpha_j, P_0),\ j = 0, 1, \ldots, J$
Additive Pitman-Yor Process (+PY)	$P_j = \epsilon_j\, Q_0 + (1-\epsilon_j) Q_j, \quad Q_j \overset{ind}{\sim} \mathrm{PYP}(\sigma_j, \alpha_j, P_0),\ j = 0, 1, \ldots, J$
Additive Dirichlet-Multinomial (+DM)	$P_j = \epsilon_j\, Q_0 + (1-\epsilon_j) Q_j, \quad Q_j \overset{ind}{\sim} \mathrm{DM}_M(\tau_j, P_0),\ j = 0, 1, \ldots, J$
Additive Gnedin Process (+GN)	$P_j = \epsilon_j\, Q_0 + (1-\epsilon_j) Q_j, \quad Q_j \overset{ind}{\sim} \mathrm{GN}(\gamma_j, P_0),\ j = 0, 1, \ldots, J$
Additive Species Sampling Process (+SSP)	$P_j = \epsilon_j\, Q_0 + (1-\epsilon_j) Q_j, \quad Q_j \overset{ind}{\sim} \mathrm{SSP}(\mathcal{L}_{\pi,j}, P_0),\ j = 0, 1, \ldots, J$

In this case, the idiosyncratic components are non-zero with positive probability. However, it is simple to see that for any pair (P_j, P_k) the required independence condition holds, i.e., $\boldsymbol{\pi}^{(j)} \perp \boldsymbol{\pi}^{(k)}$. Thus, $(P_1, \ldots, P_J)$ is a rmSSP.

Example 4. (Completely random vectors, Catalano, Lijoi, et al., 2021). If $(P_1, \ldots, P_J)$ is distributed according to any of the following:

- GM-dependent DP (GM-DP, Lijoi, Nipoti, et al., 2014):
$$P_j = \frac{\mu_0 + \mu_j}{\mu_0(\mathbb{X}) + \mu_j(\mathbb{X})}, \quad \mu_0 \sim \mathrm{CRM}((1-z)\tfrac{\exp\{-s\}}{s}, c, P_0), \quad \mu_j \overset{ind}{\sim} \mathrm{CRM}(z\tfrac{\exp\{-s\}}{s}, c, P_0)$$
- GM-dependent σ-stable (GM-σ, Lijoi, Nipoti, et al., 2014):
$$P_j = \frac{\mu_0 + \mu_j}{\mu_0(\mathbb{X}) + \mu_j(\mathbb{X})}, \quad \mu_0 \sim \mathrm{CRM}((1-z)\tfrac{\sigma s^{-1-\sigma}}{\Gamma(1-\sigma)}, c, P_0), \quad \mu_j \overset{ind}{\sim} \mathrm{CRM}(z\tfrac{\sigma s^{-1-\sigma}}{\Gamma(1-\sigma)}, c, P_0)$$

then, for any pair (P_j, P_k) we have $\boldsymbol{\pi}^{(j)} \perp \boldsymbol{\pi}^{(k)}$ and, hence, $(P_1, \ldots, P_J)$ is a rmSPP. Moreover, if $(P_1, \ldots, P_J)$ is distributed according to a normalized compound random measures vector (Griffin et al., 2017), then $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$ for any $j \in [J]$ and, hence, $(P_1, \ldots, P_J)$ is a rmSPP.

Example 5 (Hidden hierarchical DP)**.** $(P_1, \ldots, P_J)$ is distributed as a Hidden Hierarchical Dirichlet Process (HHDP, James, 2008; Lijoi, Prünster, et al., 2023) if
$$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \mid Q_0 \sim \mathrm{DP}(\alpha, \mathrm{DP}(\beta, Q_0)) \quad Q_0 \sim \mathrm{DP}(\beta_0, P_0)$$
then $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$ holds for any $j \in [J]$, and $(P_1, \ldots, P_J)$ is a rmSPP.

Example 6 (Semi hierarchical DP)**.** $(P_1, \ldots, P_J)$ is distributed according to a semi– hierarchical Dirichlet Process (semi-HDP, Beraha et al., 2021) if
$$P_j \mid Q \overset{iid}{\sim} Q, \quad Q \mid Q_0 \sim \mathrm{DP}(\alpha, \mathrm{DP}(\beta, \kappa P_0 + (1-\kappa) Q_0)), \quad Q_0 \sim \mathrm{DP}(\beta_0, P_0).$$
Also in this case, we have $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$ for any $j \in [J]$ and $(P_1, \ldots, P_J)$ is a rmSPP.

Example 7 (Processes based on stick-breaking constructions)**.** Assume the distribution of $(P_1, \ldots, P_J)$ coincides with any of the following:

- nested common atoms process (nCAM, Denti et al., 2023), which is given by
$$P_j \mid Q \overset{iid}{\sim} Q,\ Q = \sum_{s \geq 1} \pi_s \delta_{G_s},\ G_s = \sum_{t \geq 1} \omega_{t,s} \delta_{\theta_t},\ (\pi_s)_{s \geq 1} \sim \mathrm{GEM}(\alpha),\ (\omega_{t,s})_{t \geq 1} \overset{iid}{\sim} \mathrm{GEM}(\beta);$$

- tree stick-breaking process with covariates (treeSB, Horiguchi et al., 2024), which correspond to
  $P_j \sim \text{treeSB}(P_0, \{F_{j,\epsilon}\}, \mathscr{T})$.

Since $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$ for any $j \in [J]$, one has that $(P_1, \ldots, P_J)$ is a rmSPP.

Example 8 (Vectors of normalized independent finite point processes)**.** If the distribution $(P_1, \ldots, P_J)$ coincides with a Vectors of finite Dirichlet process (Vec-FDP, Colombi et al., 2025), i.e., $(P_1, \ldots, P_J) \sim \text{Vec-FDP}(\Lambda, \gamma, P_0)$, then $\sum_{h' \geq 0} \pi^{(j)}_{j,h'} \overset{a.s.}{=} 0$ for any $j \in [J]$ and, thus, $(P_1, \ldots, P_J)$ is a rmSPP.

Example 9 (Independent processes)**.** If $(P_1, \ldots, P_J)$ are independent SSPs, one trivially has $\boldsymbol{\pi}^{(j)} \perp \boldsymbol{\pi}^{(k)}$ for any $j \neq k$. Thus, $(P_1, \ldots, P_J)$ is a rmSSP.

3 Dependence structure and moments of mSSPs

3.1 Correlation and dependence

While mSSPs generalize SSPs, their essence lies in their multivariate nature and investigating the dependence between elements of the vector $(P_1, \ldots, P_J)$ is a crucial task, even more so since this aspect is obviously absent in univariate SSPs. Here we provide a solid foundation for the use of correlation as measure of dependence for mSSPs: we derive interpretable expressions for the correlation between pairs of random probability measures in terms of observable variables, prove that the correlation equals one if and only if the data are fully exchangeable, and, furthermore, show how zero correlation characterizes independence for *regular* mSSP. First, we compute the marginal expected value and variance of the P_j's, which will turn out to be helpful in the sequel.

Proposition 3. *If $(P_1, \ldots, P_J)$ is an mSSP and $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$, then*

$$\mathbb{E}[P_j(A)] = P_0(A), \qquad \mathbb{V}\text{ar}[P_j(A)] = \mathbb{P}(X_{j,1} = X_{j,2}) P_0(A)\big[1 - P_0(A)\big].$$

Note that by marginal exchangeability, the tie probability $\mathbb{P}(X_{j,i} = X_{j,m})$ between observations from the same population j does not depend on the indexes (i, m). Moreover, using representation (2), one can rewrite the tie probability of an mSSP as

$$\mathbb{P}(X_{j,1} = X_{j,2}) = \sum_{h \geq 1} \mathbb{E}\left[\left(\pi^{(j,k)}_{j,h}\right)^2\right] + \sum_{h' \geq 1} \mathbb{E}\left[\left(\pi^{(j)}_{j,h'}\right)^2\right]. \tag{4}$$

The link between tie probability and variance of a single homogeneous NRMI was first noted in James, Lijoi, et al. (2006). Here we have established it for general mSSPs: since mSSPs do not require specifying a law for the weights, this means that the link between variance and tie probability is structural. However, this represents only our starting point in uncovering the

crucial role played by tie probabilities for mSSPs.
The following simple, yet important, step consists of looking not only at the tie probabilities within each population but also across. Also in this case, the tie probability across populations j and k, $\mathbb{P}(X_{j,i} = X_{k,m})$, does not depend on the indexes (i, m) and equals

$$\mathbb{P}(X_{j,1} = X_{k,1}) = \sum_{h \geq 1} \mathbb{E}\left[\pi_{j,h}^{(j,k)} \pi_{k,h}^{(j,k)}\right]. \tag{5}$$

In section 4, we will also express the tie probability in terms of the more general law of the partition induced at the level of the observables. We are now ready to compute the correlation of mSSPs, a major highlight of this paper in terms of both understanding dependent models and methodological implications.

Proposition 4. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$ and A a (measurable) set such that $0 < P_0(A) < 1$. Then we have*

$$\mathbb{C}\text{or}[P_j(A), P_k(A)] = \frac{\mathbb{P}(X_{j,1} = X_{k,1})}{\sqrt{\mathbb{P}(X_{j,1} = X_{j,2})}\sqrt{\mathbb{P}(X_{k,1} = X_{k,2})}} \qquad \forall j \neq k \in [J].$$

This key result has several intertwined ramifications. First, it solves the open problem of identifying the reason for the correlation not to depend on the evaluation set A, which was observed on a case by case basis in most currently employed models: by Proposition 4 for any mSSM, the pairwise correlation between its elements is expressed exclusively in terms of tie probabilities within and across groups; hence, by (4)–(5) it depends only on the weights of the mSSP and it cannot depend on set A, which is characterized in terms of locations. Second, Proposition 4 implies that correlation between random probabilities is a consequence uniquely of ties between the observable species. Thus, the dependence boils down to ties across populations and the learning mechanism runs exclusively through the ties. Further insights on correlation as a measure of global dependence are collected in the following corollary.

Corollary 1. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$, and A is a (measurable) set such that $0 < P_0(A) < 1$. Then, for any $j \neq k \in [J]$, we have*

(c-i) $\mathbb{C}\text{or}[P_j(A), P_k(A)] \geq 0$;

(c-ii) $\mathbb{C}\text{or}[P_j(A), P_k(A)] = 0$ *iff* $\mathbb{P}(X_{j,1} = X_{k,1}) = 0$ *iff* $\mathbb{E}[\pi_{j,h}\pi_{k,h}] = 0$, *for any h;*

(c-iii) If P_j and P_k are equal in distribution, then $\mathbb{C}\text{or}[P_j(A), P_k(A)] = \mathbb{P}(X_{j,1} = X_{k,1})/\mathbb{P}(X_{j,1} = X_{j,2})$

Remark 2. The third statement is particularly appealing from an intuitive standpoint: in the common situation of equal marginals, one can think of correlation as the ratio of the probabilities of, respectively, "tie across groups" and "tie within a group". Implicitly, this also ensures that the tie probability "across groups" is always smaller than, or equal to, the one "within a group", which is a reasonable and appealing feature. See also Durante et al. (2025)

for a discussion in the context of multilayer networks, where this ordering can be recast as a desirable generalized homophily property.

Proposition 1 links correlation among pairs of P_j's with properties of observable quantities, providing both an intuitive and rigorous foundation to the use of correlation as a measure of dependence. However, this leaves an important question unanswered: how well does the correlation capture dependence among different processes? Are the extreme situations of full exchangeability, i.e., maximal dependence, and independence recovered when the correlation equals one and zero, respectively? The next Proposition shows that a correlation equal to one implies maximal dependence, that is, full exchangeability, of the observables.

Proposition 5. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$. Then, for any $j \neq k \in [J]$, we have*

$$\mathbb{C}\text{or}[P_j(A), P_k(A)] = 1 \quad \textit{if and only if} \quad P_j \overset{a.s.}{=} P_k$$

and $\mathbf{X} = (X_{\ell,i},\, i \geq 1,\, \ell \in \{j, k\})$ is exchangeable.

Corollary 1 and Proposition 5 jointly provide a straightforward interpretation of what happens when the probability of a tie across groups approaches the probability of a tie within: the correlation increases towards one and the dependence among the observations shifts from partial exchangeability towards the extreme of full exchangeability.

The other extreme case, namely independence, is harder to recover from a situation of zero correlation. However, if we restrict attention to rmSSPs, we are able to show that it is impossible to construct zero-correlated rmSSPs whose components are not pairwise independent. This yields the desired characterization, but also highlights the natural role played by rmSSPs within the general class of mSSP.

Theorem 6. *Let $(P_1, \ldots, P_J)$ be an rmSSP. Then, for any $j \neq k \in [J]$, we have*

$$\mathbb{C}\text{or}[P_j(A), P_k(A)] = 0 \quad \textit{if and only if} \quad P_j \perp P_k.$$

Hence, within the class of rmSSP, on one hand, correlation equal to one implies exchangeability and, on the other hand, correlation equal to zero implies independence among the P_j's and across groups of observations.

Remark 3. Not all rmSSPs can achieve correlation exactly equal to zero or one, at least in their standard definitions. To fix ideas, consider rmSSPs lacking idiosyncratic and improper components (i.e., $\sum_{h \geq 1} \pi_{j,h}^{(i,j)} \overset{a.s.}{=} 1$); one remarkable instance is given by hierarchical constructions. For these processes, the situation of independence across groups is to be interpreted as a limiting case. For example, J independent DPs arise from the HDP only if we let $\alpha_0 \to \infty$. To make things concrete and highlight how much literature we cover with mSSPs, Table 4 presents the correlation, probability of ties, and the values of hyperparameters to attain independence and exchangeability for a wide variety of models.

From an inferential perspective, the dependence among the latent $(P_1, \ldots, P_J)$ plays a key instrumental role, since it induces dependence among the observations. We have already recovered the extreme cases of exchangeability and independence of the observables as corresponding to, respectively, correlation one and zero of pairs of P_j's. Nonetheless, the following results, which hold for the entire class of mSSPs, highlight how the correlation among observables coincides with the tie probability. Further, we stress the implications on the induced dependence among the data $\mathbf{X}$.

Proposition 7. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$, and assume $\mathbb{X} = \mathbb{R}$. Then, for any $j, k \in [J]$ and any i, m, we have*

$$\mathbb{C}\text{or}(X_{j,i}, X_{k,m}) = \mathbb{P}(X_{j,i} = X_{k,m}).$$

Note that Proposition 7 holds true both within (i.e., $j = k$) and across (i.e., $j \neq k$) groups, and thus, also for (univariate) SSP.

Corollary 2. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$, and assume $\mathbb{X} = \mathbb{R}$. Then for any $j \neq k \in [J]$*

(c-i) $\mathbb{C}\text{or}(X_{j,i}, X_{k,m}) \geq 0$;

(c-ii) $\mathbb{C}\text{or}(X_{j,i}, X_{k,m}) = 0$ *iff* $\mathbb{P}(X_{j,i} = X_{k,m}) = 0$ *iff* $X_{j,i} \perp X_{k,m}$;

(c-iii) $\mathbb{C}\text{or}(X_{j,i}, X_{k,m}) = 0$ *iff* $\mathbb{E}[\pi_{j,h}\pi_{k,h}] = 0$, *for any* h.

3.2 Higher moments of mSSPs

We now derive both marginal and mixed moments of any order. These can also be seen as generalizations, to all SSPs and mSSPs, of the powerful results on joint moments of normalized completely random measures (James, Lijoi, et al., 2006) and of hierarchical normalized completely random measures (Camerlenghi, Lijoi, Orbanz, et al., 2019), which leverage the Laplace functional characterization of completely random measures. Here we show that moments can be computed in the class of mSSPs even for elements unrelated to completely random measures and/or to hierarchical processes. The following two propositions provide the expressions for the marginal moments.

Proposition 8. *Let $(P_1, \ldots, P_J)$ be an mSSP, $X_{j,i} \mid (P_1, \ldots, P_J) \overset{ind}{\sim} P_j$, for $i = 1, 2, \ldots$ and $j = 1, \ldots, J$. Then, for every $q \in \mathbb{N}$,*

$$\mathbb{E}[P_j(A)^q] = \mathbb{E}\left[P_0(A)^{K^{(j)}_{1:q}}\right],$$

where $K^{(j)}_{1:q}$ is the random number of unique species in a sample of size q from P_j.

Table 4: Correlation, tie probabilities and extreme cases. From left to right: type of mSSP (notation defined in Examples of Section 2.2); pairwise correlation; probability of tie across and within groups; values to which the hyperparameters should converge for the correlation to converge, respectively, to 0 and 1 (while P(Ties Within) does not converge to 0 or 1).

Process	Correlation	P(Ties Across)	P(Ties Within)	Indep.	Exchang.
HDP	$\frac{1+\alpha}{1+\alpha+\alpha_0}$	$\frac{1}{1+\alpha_0}$	$\frac{1+\alpha+\alpha_0}{(1+\alpha)(1+\alpha_0)}$	$\alpha_0 \to +\infty$	$\alpha \to +\infty$
HPY	$\frac{(1+\alpha)(1-\sigma_0)}{(1-\sigma\sigma_0)+\alpha(1-\sigma_0)+\alpha_0(1-\sigma)}$	$\frac{1-\sigma_0}{1+\alpha_0}$	$\frac{(1-\sigma\sigma_0)+\alpha(1-\sigma_0)+\alpha_0(1-\sigma)}{(1+\alpha)(1+\alpha_0)}$	$\alpha_0 \to +\infty$ or $\sigma_0 \to 1$	$\alpha \to +\infty$ or $\sigma \to 1$
HDM	$\frac{(1+\tau_0)(1+\tau M)}{(1+\tau M)(1+\tau_0 M_0)-\tau\tau_0(M-1)(M_0-1)}$	$\frac{1+\tau_0}{1+\tau_0 M_0}$	$\frac{(1+\tau M)(1+\tau_0 M_0)-\tau\tau_0(M-1)(M_0-1)}{(1+\tau M)(1+\tau_0 M_0)}$	$M_0 \to +\infty$	$M \to +\infty$
HGN	$\frac{\gamma_0(\gamma+1)}{(\gamma+\gamma_0)}$	$\frac{2\gamma_0}{\gamma_0+1}$	$\frac{2(\gamma+\gamma_0)}{(\gamma+1)(\gamma_0+1)}$	$\gamma_0 \to 0$	$\gamma \to 0$
HSSP	$\frac{\mathrm{EPPF}^{(2)}_{1,0}(2)}{\mathrm{EPPF}^{(2)}_{1,1}(2)+\mathrm{EPPF}^{(2)}_{2,1}(1,1)\mathrm{EPPF}^{(2)}_{1,0}(2)}$ *	$\mathrm{EPPF}^{(2)}_{1,0}(2)$	$\mathrm{EPPF}^{(2)}_{1,1}(2)+\mathrm{EPPF}^{(2)}_{2,1}(1,1)\mathrm{EPPF}^{(2)}_{1,0}(2)$	$\mathrm{EPPF}^{(2)}_{1,0}(2)=0$	$\mathrm{EPPF}^{(2)}_{1,1}(2)=0$
NDP	$\frac{1}{1+\alpha}$	$\frac{1}{(1+\alpha)(1+\beta)}$	$\frac{1}{1+\beta}$	$\alpha \to +\infty$	$\alpha \to 0$
NPY	$\frac{1-\sigma_\alpha}{1+\alpha}$	$\frac{(1-\sigma_\alpha)(1-\sigma_\beta)}{(1+\alpha)(1+\beta)}$	$\frac{1-\sigma_\beta}{1+\beta}$	$\alpha \to +\infty$ or $\sigma_\alpha \to 1$	$(\alpha,\sigma_\alpha) \to$ $\to (0,0)$
NDM	$\frac{1+\tau_\alpha}{1+\tau_\alpha M_\alpha}$	$\frac{(1+\tau_\alpha)(1+\tau_\beta)}{(1+\tau_\alpha M_\alpha)(1+\tau_\beta M_\beta)}$	$\frac{1+\tau_\beta}{1+\tau_\beta M_\beta}$	$M_\alpha \to +\infty$	$M_\alpha \to 1$
NGN	$\frac{2\gamma_\alpha}{\gamma_\alpha+1}$	$\frac{4\gamma_\alpha\gamma_\beta}{(\gamma_\alpha+1)(\gamma_\beta+1)}$	$\frac{2\gamma_\beta}{\gamma_\beta+1}$	$\gamma_\alpha \to 0$	$\gamma_\alpha \to 1$
NSSP	$\mathrm{EPPF}^{(2)}_{1,0}(2)$	$\mathrm{EPPF}^{(2)}_{1,0}(2)\mathrm{EPPF}^{(2)}_{1,1}(2)$	$\mathrm{EPPF}^{(2)}_{1,1}(2)$	$\mathrm{EPPF}^{(2)}_{1,0}(2)=0$	$\mathrm{EPPF}^{(2)}_{1,0}(2)=1$
+DP	$\frac{\frac{\epsilon_j\epsilon_k}{1+\alpha_0}}{\sqrt{\left(\frac{\epsilon_j^2}{1+\alpha_0}+\frac{(1-\epsilon_j)^2}{1+\alpha_j}\right)\left(\frac{\epsilon_k^2}{1+\alpha_0}+\frac{(1-\epsilon_k)^2}{1+\alpha_k}\right)}}$	$\frac{\epsilon_j\epsilon_k}{1+\alpha_0}$	$\frac{\epsilon_j^2}{1+\alpha_0}+\frac{(1-\epsilon_j)^2}{1+\alpha_j}$	$\epsilon=0$	$\epsilon=1$
+PY	$\frac{\frac{\epsilon_j\epsilon_k(1-\sigma_0)}{1+\alpha_0}}{\sqrt{\left(\frac{\epsilon_j^2(1-\sigma_0)}{1+\alpha_0}+\frac{(1-\epsilon_j)^2(1-\sigma_j)}{1+\alpha_j}\right)\left(\frac{\epsilon_k^2(1-\sigma_0)}{1+\alpha_0}+\frac{(1-\epsilon_k)^2(1-\sigma_k)}{1+\alpha_k}\right)}}$	$\frac{\epsilon_j\epsilon_k(1-\sigma_0)}{1+\alpha_0}$	$\frac{\epsilon_j^2(1-\sigma_0)}{1+\alpha_0}+\frac{(1-\epsilon_j)^2(1-\sigma_j)}{1+\alpha_j}$	$\epsilon=0$	$\epsilon=1$
+DM	$\frac{\frac{\epsilon_j\epsilon_k(1+\tau_0)}{1+\tau_0 M_0}}{\sqrt{\left(\frac{\epsilon_j^2(1+\tau_0)}{1+\tau_0 M_0}+\frac{(1-\epsilon_j)^2(1+\tau_j)}{1+\tau_j M_j}\right)\left(\frac{\epsilon_k^2(1+\tau_0)}{1+\tau_0 M_0}+\frac{(1-\epsilon_k)^2(1+\tau_k)}{1+\tau_k M_k}\right)}}$	$\frac{\epsilon_j\epsilon_k(1+\tau_0)}{1+\tau_0 M_0}$	$\frac{\epsilon_j^2(1+\tau_0)}{1+\tau_0 M_0}+\frac{(1-\epsilon_j)^2(1+\tau_j)}{1+\tau_j M_j}$	$\epsilon=0$	$\epsilon=1$
+GN	$\frac{\frac{\epsilon_j\epsilon_k 2\gamma_0}{\gamma_0+1}}{\sqrt{\left(\frac{\epsilon_j^2 2\gamma_0}{\gamma_0+1}+\frac{(1-\epsilon_j)^2 2\gamma_j}{\gamma_j+1}\right)\left(\frac{\epsilon_k^2 2\gamma_0}{\gamma_0+1}+\frac{(1-\epsilon_k)^2 2\gamma_k}{\gamma_k+1}\right)}}$	$\frac{\epsilon_j\epsilon_k 2\gamma_0}{\gamma_0+1}$	$\frac{\epsilon_j^2 2\gamma_0}{\gamma_0+1}+\frac{(1-\epsilon_j)^2 2\gamma_j}{\gamma_j+1}$	$\epsilon=0$	$\epsilon=1$
+SSP	$\frac{\epsilon_j\epsilon_k\mathrm{EPPF}^{(2)}_{1,0}(2)}{\sqrt{(\epsilon_j^2\mathrm{EPPF}^{(2)}_{1,0}(2)+(1-\epsilon_j)^2\mathrm{EPPF}^{(2)}_{1,1}(2))(\epsilon_k^2\mathrm{EPPF}^{(2)}_{1,0}(2)+(1-\epsilon_k)^2\mathrm{EPPF}^{(2)}_{1,1}(2))}}$	$\epsilon_j\epsilon_k\mathrm{EPPF}^{(2)}_{1,0}(2)$	$\epsilon_j^2\mathrm{EPPF}^{(2)}_{1,0}(2)+(1-\epsilon_j)^2\mathrm{EPPF}^{(2)}_{1,1}(2)$	$\epsilon=0$	$\epsilon=1$
GM-DP	$\frac{(1-z)c}{1+c}\,{}_3F_2(a,1,1;b,b;1)$*	$\frac{(1-z)c}{(1+c)^2}\,{}_3F_2(a,1,1;b,b;1)$**	$\frac{1}{1+c}$	$z=1$	$z=0$
GM-σ	$(1-z)\mathcal{G}(c,z)$***	$(1-z)(1-\sigma)\mathcal{G}(c,z)$	$1-\sigma$		
HHDP	$1-\frac{\alpha\beta_0}{(1+\alpha)(\beta_0+\beta+1)}$	$\frac{1}{\beta_0+1}+\frac{\beta_0}{(1+\alpha)(1+\beta)(1+\beta_0)}$	$\frac{1+\beta+\beta_0}{(1+\beta)(1+\beta_0)}$	$(\alpha,\beta_0) \to$ $\to (+\infty,+\infty)$	$\alpha \to 0$
nCAM	$1-\frac{\beta\alpha}{(2\beta+1)(1+\alpha)}$	$\frac{1}{1+\alpha}\left(\frac{1}{1+\beta}+\frac{\alpha}{2\beta+1}\right)$	$\frac{1}{1+\beta}$	None	$\alpha \to 0$

* $\mathrm{EPPF}^{(2)}_{\cdot,1}$ and $\mathrm{EPPF}^{(2)}_{\cdot,0}$ are induced by $\mathcal{L}_{\pi,1}=\ldots=\mathcal{L}_{\pi,J}$ and $\mathcal{L}_{\pi,0}$, respectively. ** ${}_3F_2$ is the generalized hypergeometric function and $a=\alpha(1-z)+2, b=\alpha+2$ *** $\mathcal{G}(c,z)=\frac{1}{\sigma}\int_0^1 \frac{w^{1/\sigma-1}}{[1+z(1-w^{1/\sigma})^\sigma-z(1-\omega)]}\mathrm{d}\omega$

Proposition 9. *Let $(P_1,\ldots,P_J)$ be an mSSP and $\{A_1,\ldots,A_h\}$ be pairwise disjoint sets. Then, for any sequence $q_1,q_2,\ldots,q_h$, with $q_i \in \mathbb{N}$ for $i=1\ldots,h$, we have*

$$\mathbb{E}[P_j(A_1)^{q_1}\cdots P_j(A_h)^{q_h}] = \mathbb{E}\left[P_0(A_1)^{K^{(j)}_{1:q_1}}\, P_0(A_2)^{K^{(j)}_{q_1+1:q_2}}\cdots P_0(A_h)^{K^{(j)}_{q_{h-1}+1:q_h}} \mid E_{\neq}\right]\mathbb{P}(E_{\neq}),$$

where $K^{(j)}_{a:b}$ is the random number of species in the "block of observations" from the a-th to the b-th observation, in a sample of size $q_1+\cdots+q_h$ from P_j, and $E_{\neq}$ is an event that occurs if no shared species are recorded across the different groups of observations.

The following two theorems provide the expressions for the mixed moments.

Theorem 10. *Let* $(P_1, \dots, P_J)$ *be an mSSP,* $X_{j,i} \mid (P_1, \dots, P_J) \overset{ind}{\sim} P_j$*, for* $i = 1, 2, \dots$ *and* $j = 1, \dots, J$*. Then, for any sequence* $q_1, q_2, \dots, q_J$*, with* $q_i \in \mathbb{N}$ *for* $i = 1 \dots, J$*, we have*

$$\mathbb{E}[P_1(A)^{q_1} \cdots P_J(A)^{q_J}] = \mathbb{E}\big[P_0(A)^{K_{q_1,\dots,q_J}}\big],$$

where $K_{q_1,\dots,q_J}$ *is the overall number of species observed in a sample that contains* q_j *observations from* P_j*, for* $j = 1, \dots, J$*.*

Theorem 11. *Let* $(P_1, \dots, P_J)$ *be an mSSP and* $\{A_1, \dots, A_J\}$ *be pairwise disjoint (measurable) sets. Then, for any sequence* $q_1, q_2, \dots, q_J$*, with* $q_i \in \mathbb{N}$ *for* $i = 1 \dots, J$*, we have*

$$\mathbb{E}[P_1(A_1)^{q_1} \cdots P_J(A_J)^{q_J}] = E\left[P_0(A_1)^{K^{(1)}_{1:q_1}} \cdots P_0(A_J)^{K^{(J)}_{1:q_J}} \mid E_{\neq}\right] \mathbb{P}(E_{\neq}),$$

where $K^{(j)}_{1:q_j}$ *is the number of observed species from population* j*, in a sample which contains* q_j *observations from* P_j*, for* $j = 1, \dots, J$*.*

Importantly, these results showcase that higher-order moments of an mSSP evaluated on (measurable) sets can be meaningfully interpreted in terms of simple observable quantities like the random number of observed species within and across groups. Hence, interpretability is not unique to correlation.

4 Partially exchangeable partition function

In the exchangeable case, the random partition induced by a discrete nonparametric prior is characterized by the exchangeable partition probability function (EPPF), a concept introduced in Pitman (1995). The EPPF plays a fundamental role across several domains, including combinatorics, stochastic process theory, population genetics, Bayesian statistics and machine learning; see Pitman (2006) and references therein. It also underpins models in ecology and natural language processing, where exchangeable clustering structures naturally arise. The EPPF associated with the DP corresponds to Ewens' sampling formula (Antoniak, 1974; Ewens, 1990); see Crane (2016) and Tavaré (2021) accounts of its widespread applications.
Moving from exchangeability to partial exchangeability, the counterpart of the EPPF is the partially exchangeable partition probability function (pEPPF), which characterizes the random partition induced by a partially exchangeable array. This concept is different from the one introduced in Pitman (1995), which is unrelated to the original definition of partial exchangeability in the sense of de Finetti that we adopt here. The notion of pEPPF first appeared in Leisen et al. (2011) and Lijoi, Nipoti, et al. (2014) for specific instances of dependent processes, and it started being leveraged in a systematic way for other subclasses of dependent processes only recently in, e.g., Camerlenghi, Lijoi, and Prünster (2017), Camerlenghi, Lijoi, Orbanz, et al. (2019), Camerlenghi, Dunson, et al. (2019), Beraha et al. (2021), Lijoi, Prünster, et al.

(2023), and Denti et al. (2023). Its absence from the classical probabilistic literature may stem from the fact that, unlike in the exchangeable setting where the EPPF can often be defined directly without invoking an associated exchangeable sequence, the partially exchangeable case lacks a similarly tractable direct construction. Instead, the pEPPF arises naturally by marginalizing a partially exchangeable array of random elements, which represents the canonical approach to deriving the corresponding random partition within the BNP framework. In multi-population species sampling problems, where the values sampled from P_0 serve solely as species labels with no numerical meaning, the pEPPF uniquely determines the marginal likelihood of the observations. Similarly, in the context of model-based clustering or latent multi-level modeling, the pEPPF encapsulates the underlying clustering mechanism. Importantly, from a computational perspective, the pEPPF also plays a pivotal role as it provides the key ingredient for deriving marginal posterior sampling schemes.

Let us now formally introduce the pEPPF induced by any vector $(P_1, \ldots, P_J)$ of random probability measures with possibly discrete components. A sample $(X_{j,i} : i \in [I_j], , j \in [J])$, where I_j denotes the sample size of group j and $n = \sum_{j=1}^{J} I_j$ is the total sample size, induces a random partition of the integers $[n]$ based on the ties among the observations. To see this, let the integers label the observations according to their *order of arrival by group*, that is, observations are indexed first by group $j = 1, \ldots, J$, and then by within-group order of arrival. Specifically, observation $X_{j,i}$ is associated with the label $\sum_{j'=1}^{j-1} I_{j'} + i$, for any $i = 1, \ldots, I_j$. The resulting random partition can be usefully characterized by the corresponding pEPPF. To this end, let D be the number of distinct values among the $n = \sum_{j=1}^{J} I_j$ observations in the sample $(X_{j,i} : i \in [I_j], j \in [J])$. For each group j, define the vector of frequency counts $\boldsymbol{n}_j = (n_{j,1}, \ldots, n_{j,D})$, where $n_{j,d}$ indicates the number of observations in the jth group that coincide with the dth distinct value, indexed according to the order of arrival by groups. Clearly, $n_{j,d} \geq 0$ and by construction $\sum_{i=1}^{J} n_{i,d} \geq 1$, since each distinct value must appear in at least one group. The count $n_{j,d} = 0$ indicates that the dth distinct value does not occur in group j, while it is shared between groups k and l if and only if $n_{k,d}\, n_{l,d} \geq 1$. The law of the resulting random partition is characterized by its pEPPF, defined as

$$\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J) = \mathbb{E}\left[\int_{\mathbb{X}_*^D} \prod_{d=1}^{D} P_1(\mathrm{d}x_d)^{n_{1,d}} \ldots P_J(\mathrm{d}x_d)^{n_{J,d}} \right], \tag{6}$$

under the constraint that $\sum_{d=1}^{D} n_{j,d} = I_j$ for each $j = 1, \ldots, J$, and where $\mathbb{X}$ denotes the space in which the $X_{j,i}$'s take values, while $\mathbb{X}_*^D$ denotes the subset of $\mathbb{X}^D$ consisting of vectors with all distinct entries. We stress that the expectation in (6) is taken with respect to the joint distribution of the vector of random probability measures $(P_1, \ldots, P_J)$, that is, the de Finetti measure associated with the partially exchangeable array. An important special case is immediately recovered when $J = 1$, namely the single population setting: indeed, the pEPPF in (6) reduces to a standard EPPF. Moreover, if $J = 2$, the probability of a tie across groups coincides with $\text{pEPPF}_1^{(2)}(1, 1)$.

Clearly, if $(P_1, \ldots, P_J)$ is an mSSP, it induces a pEPPF as defined by (6). But what about

the converse? Given a pEPPF, does there exist an mSSP, up to the choice of an independent, non-atomic base measure P_0, that generates it? This amounts to asking whether every pair of pEPPF and independent non-atomic P_0 determines a unique mSSP. The next result provides an affirmative answer and, as a by-product, yields an intuitive generative construction.

Theorem 12. *Let Π_n be any pEPPF as in (6) and P_0 be a non-atomic (deterministic) probability measure. Consider the partially exchangeable array $\mathbf{X} = (X_{j,i} : i \in \mathbb{N}, j \in [J])$ such that for any non-negative integers $I_1, \ldots, I_J$ the variables $(X_{j,i} : i \in [I_j], j \in [J])$ follow the generative scheme:*

1. *sample the random partition Π_n from the given pEPPF;*
2. *conditionally on the partition Π_n, sample from P_0 the iid unique values associated with each partition set.*

Then, the de Finetti measure associated with $\mathbf{X}$ is the law of an mSSP.

Remark 4. At first glance, the previous result may seem surprising: any pEPPF in (6), regardless of whether $(P_1, \ldots, P_J)$ generating it is an mSSP or not, identifies an mSSP by pairing it with an independent non-atomic base measure P_0. The core idea behind such a fundamental result is that the distribution of the weights of the directing probability measure is what characterizes the partition induced by the ties of any (infinite) partially exchangeable array. Hence, if the dependent process $(P_1, \ldots, P_J)$ inducing the pEPPF is not an mSSP (for instance, due to dependence between weights and locations), one can still identify an mSSP, say $(P_1^*, \ldots, P_J^*)$, based on the same pEPPF. The two vectors $(P_1, \ldots, P_J)$ and $(P_1^*, \ldots, P_J^*)$ will generally have different distributions, although they share the same pEPPF. In other words, the pEPPF characterizes the multivariate partition structure, but not the law of the partially exchangeable array. Recovering the latter requires specifying a mechanism for atom assignment, with the independence required by mSSPs representing a simple and tractable choice. Crucially, this implies that the random partition structure of any partially exchangeable array can be studied via the pEPPF of an mSSP, making mSSPs the natural framework for analyzing and understanding the discrete structure of arbitrary dependent vectors and partially exchangeable partition models.

The notion of pEPPF also enables us to restate the pairwise correlation results from Section 3.1 within its more general structure. For instance, one can express

$$\mathbb{C}\mathrm{or}[P_j(A), P_k(A)] = \frac{\mathrm{pEPPF}_1^{(2)}(1,1)}{\sqrt{\mathrm{EPPF}_{j,1}^{(2)}(1)}\sqrt{\mathrm{EPPF}_{k,1}^{(2)}(1)}},$$

where EPPF_j denotes the marginal EPPF corresponding to P_j. Similarly, the correlation between observations satisfies $\mathbb{C}\mathrm{or}(X_{j,i}, X_{k,m}) = \mathrm{pEPPF}_1^{(2)}(1,1)$.

Finally, we record an alternative representation of the pEPPF in terms of the weights associated with a proper mSSP.

Proposition 13. *Let $(P_1, \ldots, P_J)$ be a proper mSSP. Then*

$$\mathrm{pEPPF}_D^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J) = \mathbb{E}\Big[\sum_{h_1 \neq \ldots \neq h_D} \prod_{j=1}^{J} \prod_{d=1}^{D} \pi_{j,h_d}^{n_{j,d}}\Big]. \tag{7}$$

5 Predictive structure and inference

In the exchangeable case, the predictive distribution of an SSP admits a simple and elegant representation, with weights expressed as ratios of the associated EPPF (Pitman, 1996). The sequential mechanism that generates these prediction rules is known as the generalized Chinese restaurant process: observations correspond to customers entering a restaurant, each choosing to sit either at an already occupied table or at a new one. Each table serves a unique dish drawn independently from P_0. The predictive distribution is given by

$$\mathbb{P}\big(X_{n+1} = x \mid \boldsymbol{X}\big) = \begin{cases} \frac{\mathrm{EPPF}_K^{(n+1)}(n_1,\ldots,n_k+1,\ldots,n_K)}{\mathrm{EPPF}_K^{(n)}(n_1,\ldots,n_k,\ldots,n_K)} & \text{if } x = X_k^* \quad \text{and } k = 1,\ldots,K \\ \frac{\mathrm{EPPF}_{K+1}^{(n+1)}(n_1,\ldots,n_k,\ldots,n_K,1)}{\mathrm{EPPF}_K^{(n)}(n_1,\ldots,n_k,\ldots,n_K)} & \text{if } x = X_{K+1}^*, \end{cases} \tag{8}$$

where $(X_k^* : k = 1, \ldots, K)$ are the K distinct values observed among $X_1, \ldots, X_n$, appearing with frequencies $(n_1, \ldots, n_K)$ and drawn i.i.d. from P_0. These probabilities follow from conditioning on the observed partition: the EPPF in the numerator is updated either by increasing the count of an existing cluster or by adding a new singleton cluster. See Section S1 of the Supplementary Material for further details.

In the partially exchangeable framework, the pEPPF associated with an mSSP naturally gives rise to a multivariate generative mechanism, where predictive distributions are again expressed as ratios of pEPPFs. We refer to this construction as the multivariate generalized Chinese restaurant process (mgCRP). It differs both from the classical generalized Chinese restaurant process and from common multi-population extensions typically modeled as restaurant franchises. Unlike the latter, we do not introduce multiple restaurants. Unlike the former, although we retain a single restaurant in which each table serves a unique dish and a customer at a new table receives a previously unserved dish, the allocation mechanism is more intricate. Specifically, the probability that a customer sits at a given table depends not only on the current seating configuration, but also on the group of the incoming customer and the group membership of those already seated. Remarkably, these allocation probabilities can still be expressed as ratios of pEPPFs, a fact that is both natural and striking. The resulting mgCRP is formalized in the next proposition.

Proposition 14. *Let $\boldsymbol{X}$ be a partially exchangeable array directed by a de Finetti measure given by the law of an mSSP $(P_1, \ldots, P_J)$. For any $j \in [J]$, the corresponding predictive*

distributions are characterized by an mgCRP of the form

$$X_{j,I_j+1} \mid (\boldsymbol{X}_{j,1:I_j})_{j=1}^{J} = \begin{cases} X_l^* & w.p. \ \frac{\text{pEPPF}_D^{(n+1)}(\boldsymbol{n}_1,\ldots,[n_{j,1},\ldots,n_{j,l}+1,\ldots,n_{1,D}],\ldots,\boldsymbol{n}_J)}{\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1,\ldots,[n_{j,1},\ldots,n_{j,l},\ldots,n_{j,D}],\ldots,\boldsymbol{n}_J)} \\ X_{new}^* & w.p. \ \frac{\text{pEPPF}_{D+1}^{(n+1)}(\boldsymbol{n}_1,\ldots,[n_{j,1},\ldots,n_{j,l},\ldots,n_{j,D},1],\ldots,[\boldsymbol{n}_J,0])}{\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1,\ldots,[n_{j,1},\ldots,n_{j,l},\ldots,n_{j,D}],\ldots,\boldsymbol{n}_J)} \end{cases}$$

where $(X_1^*,\ldots,X_D^*)$ *are the* D *unique values in* $(\boldsymbol{X}_{j,1:I_j})_{j=1}^J$ *listed in order of arrival by group,* $n=\sum_j I_j$*, and* X_{new}^* *represents a new species sampled independently from* P_0*.*

Although the predictive scheme in Proposition 14 follows naturally from the structure of the pEPPF and stands out for its theoretical elegance, its computational feasibility depends heavily on the ability to evaluate ratios of pEPPFs. Unlike the case of univariate SSPs, where such ratios are sometimes available in closed form, the multivariate setting rarely admits simple analytic expressions. Exceptions are limited to trivial cases that reduce to univariate specifications, such as independent or almost surely identical Gibbs-type priors. Section S.1 details explicit predictive schemes for specific univariate SSPs within the Gibbs-type family (Gnedin and Pitman, 2006; Lijoi, Mena, et al., 2007a), arguably the most tractable generalization of the DP (De Blasi et al., 2015). Nevertheless, it is important to note that mSSPs used in Bayesian modeling give rise to pEPPFs that are obtained as mixtures of EPPFs. This includes models such as the HSSP, NSSP, +SSP, and various combinations thereof. Thus, implementable predictive sampling schemes can typically be derived through data augmentation strategies that exploit the tractability of the underlying EPPFs. These techniques leverage latent variables that simplify ratios of pEPPFs to ratios of products of EPPFs, greatly simplifying computations. A prominent example is the Chinese restaurant franchise representation for the HDP (Teh et al., 2006). The following examples present such augmented formulations of the pEPPF, which enable tractable predictive schemes and facilitate the design of marginal Gibbs samplers, for three large classes of regular mSSPs, namely HSSP, NSSP, and +SSP. Let $\text{pEPPF}_{D,\text{aug}}^{(n)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_J,\boldsymbol{\ell},\boldsymbol{q})$ denote the augmented pEPPF, in the sense that the original pEPPF can be recovered by summing over all possible values of the latent variables $\boldsymbol{\ell}$ and $\boldsymbol{q}$, i.e., $\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_J)=\sum_{\boldsymbol{\ell},\boldsymbol{q}}\text{pEPPF}_{D,\text{aug}}^{(n)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_J,\boldsymbol{\ell},\boldsymbol{q})$.

Example 1 (Continue). If $(P_1,\ldots,P_J)$ is an HSSP, then

$$\text{pEPPF}_{D,\text{aug}}^{(n)}(\boldsymbol{n}_1,\ldots,\boldsymbol{n}_J,\boldsymbol{\ell},\boldsymbol{q}) = \text{EPPF}_{D,0}^{(\ell_{\cdot,\cdot})}(\ell_{\cdot,1},\ldots,\ell_{\cdot,D})\prod_{j=1}^J \text{EPPF}_{\ell_{j,\cdot},j}^{(I_j)}(q_{j,1},\ldots,q_{j,\ell_{j,\cdot}}), \tag{9}$$

where for $j=1,\ldots,J$, $\text{EPPF}_{\ell_{j,\cdot},j}^{(I_j)}(q_{j,1},\ldots,q_{j,\ell_{j,\cdot}})$ denotes the EPPF induced by $\mathcal{L}_{\pi,j}$, which characterizes a latent partition of the I_j observations of group j into $\ell_{j,\cdot}$ blocks of cardinalities $q_{j,1},\ldots,q_{j,\ell_{j,\cdot}}$. Conditionally on these partitions, all the $\ell_{\cdot,\cdot}=\sum_{j=1}^J\ell_{j,\cdot}$ blocks (we use the $\cdot$ notation indicates summation over the corresponding index set) are grouped into a coarser partition of D blocks, each corresponding to a distinct observed species. The distribution of this coarser partition is characterized by the $\text{EPPF}_{D,0}^{(\ell_{\cdot,\cdot})}(\ell_{\cdot,1},\ldots,\ell_{\cdot,D})$ induced by Q.

Example 2 (Continue). If $(P_1, \ldots, P_J)$ is an NSSP, then

$$\text{pEPPF}^{(n)}_{D,\text{aug}}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J, \boldsymbol{\ell}, \boldsymbol{q}) = \text{EPPF}^{(J)}_{R,0}(\ell_1, \ldots, \ell_R) \prod_{r=1}^{R} \text{EPPF}^{(I^\star_r)}_{D_r}(q_{1,\cdot}, \ldots, q_{D_r,}), \tag{10}$$

where $\text{EPPF}^{(J)}_{R,0}(\ell_1, \ldots, \ell_R)$ denotes the EPPF induced by $\mathcal{L}_{\pi,0}$ that controls the clustering of the group labels $j = 1, \ldots, J$ into R blocks (obtained from the ties among the P_j's). Let $P^\star_r \overset{\text{iid}}{\sim} \text{SSP}(\mathcal{L}_{\boldsymbol{\pi}}, P_0)$ $r = 1, \ldots, R$ be the unique values of $(P_j)_{j=1}^J$ in order of arrival and let $I^\star_r = \sum_{j:P_j=P^\star_r} I_j$ be the number of observations from the ℓ_r groups assigned to $P^\star_r$. Conditionally on this clustering of the groups, for $r = 1, \ldots, R$, the $\text{EPPF}^{(I^\star_r)}_{D_r}(q_{1,\cdot}, \ldots, q_{D_r,})$ describes the distribution induced by $P^\star_r$ characterizing the partition of the $I^\star_r$ observations assigned to $P^\star_r$ into D_r distinct species. Since the $P^\star_r$'s do not share atoms a.s., it follows that the total number of distinct species is given by $D = \sum_{r=1}^{R} D_r$.

Example 3 (Continue). If $(P_1, \ldots, P_J)$ is a +SSP, then

$$\text{pEPPF}^{(n)}_{D,\text{aug}}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J, \boldsymbol{\ell}, \boldsymbol{q}) = \prod_{j=1}^{J} \epsilon_j^{\ell_0}(1-\epsilon_j)^{\ell_j} \prod_{j=0}^{J} \text{EPPF}^{(\ell_j)}_{D_j,j}(q_{j,1}, \ldots, q_{j,D_j}), \tag{11}$$

where ℓ_0 and $\ell_j = I_j - \ell_0$ denote, for each $j \in [J]$, the number of observations assigned to the shared SSP Q_0 and to the idiosyncratic SSP Q_j, respectively, while $\epsilon_j^{\ell_0}(1-\epsilon_j)^{\ell_j}$ is the probability of the i.i.d. latent assignment of the I_j observations via $\text{Bern}(\epsilon_j)$. Conditionally on these latent assignments, for $j = 0, \ldots, J$, $\text{EPPF}^{(\ell_j)}_{D_j,j}(q_{j,1}, \ldots, q_{j,D_j})$ is the EPPF induced by Q_j that governs the clustering of the ℓ_j observations assigned to Q_j into D_j unique species. Since the Q_j's do not share species a.s., the total number of distinct species across all groups is given by $D = \sum_{j=0}^{J} D_j$.

The hierarchical representations of the pEPPF derived in (9), (10), and (11), expressed as products of simple EPPFs in an augmented space, allow for simplifying the ratio in Proposition 14 into a product of tractable predictive expressions, analogous to those of the Chinese restaurant process.

6 Multi-armed bandits for species discovery

Among the numerous application areas of dependent processes, which include density regression, spatio-temporal analysis, functional data, survival analysis, topic modeling, hierarchical and multi-level clustering, and ANOVA-type models, here we focus on a multi-armed bandit problem connected to species sampling. Embedding the analysis within the framework of mSSMs offers a principled way to compare different classes of models. Our structural results for mSSMs allow model parameters to be chosen so that relevant prior quantities coincide, enabling fair performance comparisons among competing approaches. A systematic investigation of which models are preferable in specific settings is beyond the scope of this paper, but

the general strategy is clear: calibrate prior parameters to match the aspects that are most critical for the task at hand, and evaluate inferential performance accordingly. This will be the focus of future work.

A Bayesian nonparametric approach to species sampling problems in the single population case, i.e., $J = 1$, was introduced by Lijoi, Mena, et al. (2007a), where Bayesian analogs of the classical Turing and Good-Toulmin estimators (Good, 1953; Good and Toulmin, 1956) were derived. In this setting, a random probability measure P models the species proportions in the population, and, given an observed sample, the main goal is to estimate the probability of discovering a new species either at the next step or after an additional m unobserved draws. Following Lijoi, Mena, et al. (2007a), there has been a rich literature exploring alternative prior specifications, estimation of diverse functionals and quantities of interest, and a wide range of applications. For detailed reviews, see De Blasi et al. (2015) and Balocchi, Favaro, et al. (2025), and references therein.

The multi-sample setup with J populations modeled through a vector of dependent random probability measures $(P_1, \ldots, P_J)$ was first studied in Camerlenghi, Lijoi, and Prünster (2017). A sequential perspective was adopted in Battiston et al. (2018) and Camerlenghi, Dumitrascu, et al. (2020), with the goal of designing an optimal sequential sampling strategy to maximize the diversity of the observed species. This involves deciding, at each step, which population to sample from, while sequentially incorporating information from previously observed species across populations. The problem naturally fits within the framework of multi-armed bandits, where each arm represents a population and a unit reward is earned upon discovering a new species. Such problems arise in ecology and biology, where sampling from diverse environments aims to uncover new species, and in genomics, where the objective is often to detect as many genetic variants as possible (see, e.g., Lijoi, Mena, et al., 2008; Masoero et al., 2022).

6.1 Real data

Here we consider a multi-armed bandit problem of trees' species discovery, using the dataset of South American tree species publicly available in the supplementary materials of Condit et al. (2002). The dataset records 41,688 trees observed across 100 plots, comprising 802 distinct species. In accordance with Battiston et al. (2018), we aggregated the 100 plots into four larger groups based on spatial location, joining columns in the dataset whose codes begin with BCI, P, S, and C, respectively. These four groups define the $J = 4$ alternative arms. The empirical distributions and empirical tie probabilities (i.e., relative frequencies) for each group are shown in Figure 2. Further details on the dataset can be found in Pyke et al. (2001), Condit et al. (2002), and Battiston et al. (2018). The four arms corresponding to the trees' populations in the four different regions are modeled as a vector of dependent random probabilities $(P_1, \ldots, P_4)$, each representing a population whose species composition is initially unknown, both in terms of presence of a species and relative abundance. Species may be shared across arms, possibly with different frequencies, making the rmSSP framework

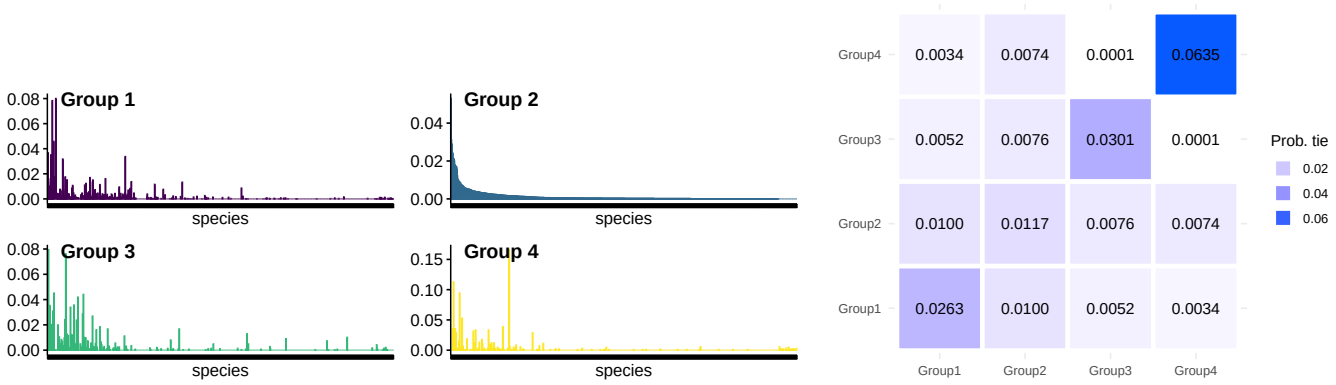

Figure 2: Empirical distribution functions of the four groups (left) and empirical tie probabilities (right), computed from the full dataset. In the left panel, species within each group are ordered according to their frequency ranking in Group 2.

a natural modeling choice.

Within this setup, we compare the performance of six rmSSPs in maximizing the number of distinct observed species in an additional sample. For each model, the sampling strategy to achieve this goal consists of selecting at step $n+1$ the arm with the highest estimated probability of discovering a new species, based on observations collected up to step n. Specifically, at each step we choose the arm j that maximizes $\mathbb{P}(X_{j,I_j+1} \notin \boldsymbol{X}_{\text{past}} \mid \boldsymbol{X}_{\text{past}})$, where $\boldsymbol{X}_{\text{past}}$ denotes the previously observed species across all sites. We also contrast these model-based approaches with a simple baseline that selects an arm uniformly at random at each step, which we refer to as the uniform model.

The six rmSSP models we compare are: independent DP and PYP, additive DP and PYP, and hierarchical DP and PYP. We assign hyperpriors to the concentration parameter in each DP-based rmSSP and to both concentration and discount parameters in each PYP-based rmSSP. To ensure a fair comparison, these hyperpriors are chosen so that the prior mean and variance of the tie probabilities, within groups for all models and across groups for those that borrow information, match across all six specifications. This calibration reflects two considerations. First, in a species sampling problem, where our objective is to maximize the species diversity, it is sensible to set the probability of not discovering a new species, i.e., the probability of ties, equal across all models. Second, by the results of the previous section, the tie probabilities effectively capture the dependence structure and information-sharing behavior for any rmSSP, regardless of the specific application at hand. Full details on model definitions, sampling algorithms, and hyperprior settings are provided in Section S.3 of the Supplementary Material. Figure 3 showcases the average cumulative number of species discovered by the two hierarchical, the two additive, and the two independent rmSSP models, each also compared to the uniform model, as a function of the number of additional samples. Table 5 reports the average number of new species discovered per sampling step. All results are averages based on 20 runs. In each

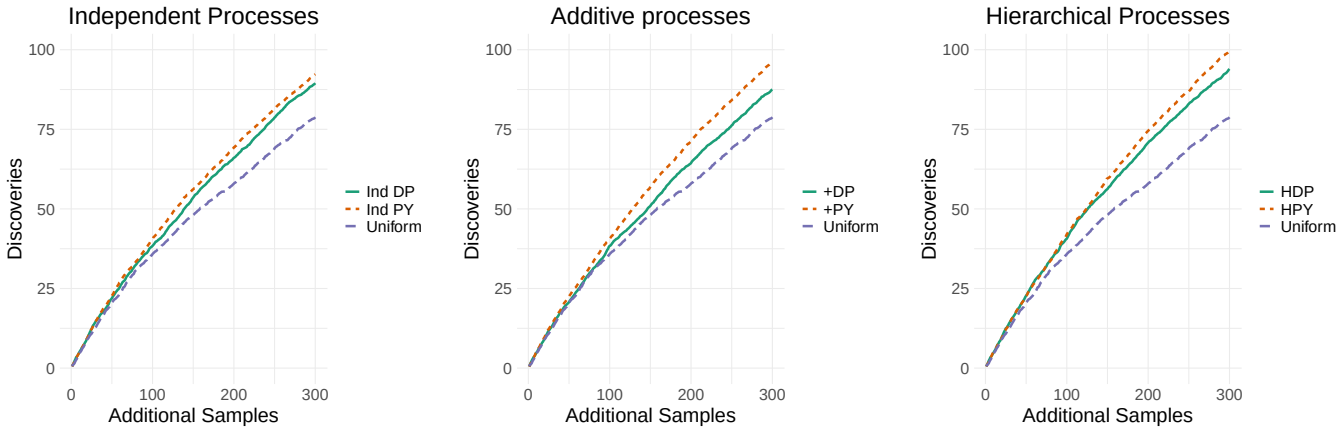

Figure 3: Tree species data: cumulative number of species discovered as a function of the additional sample size for each rmSSP model and the uniform baseline.

	Tree data						
	Uniform	DP	PY	+DP	+PY	HDP	HPY
Avg. num.	0.2608	0.2965	0.3060	0.2900	0.3186	0.3115	0.3298

Table 5: Tree species data: Average number of new species discovered per sampling step for each rmSSP model and the uniform baseline.

run, we begin with an initial sample of 30 observations per arm (drawn without replacement from the full dataset), then sample 300 further observations sequentially according to each strategy and record the species discoveries.

Several noteworthy insights emerge from this experiment: (a) All rmSSP models are clearly superior to the uniform baseline. (b) The two PYP-based rmSSPs consistently outperform their DP-based counterparts, thanks to the extra flexibility provided by the discount parameter, which governs the rate at which new species appear. (c) With the exception of the +DP, all models that borrow information across populations yield higher discovery rates than the independent specifications. The +DPs weak performance stems from its underlying assumption that shared-species frequencies are proportional across populations, which seems inappropriate in this setting, and its lower flexibility compared to PYP, which prevents it from compensating for this misspecification. To the best of our knowledge, this limitation of the +DP has not been previously noted in the literature.

6.2 Synthetic data

The *tree dataset* exhibits high probabilities of ties across samples (see the right panel of Figure 2 and recall that such probabilities are bounded above by the probability of a tie within a sample) and, thus, distributions in different samples are highly similar. This makes it quite apparent

that borrowing information across groups is advantageous. Therefore, one could argue that the setting considered is overly favourable to rmSSPs relative to the independent models. To assess whether, and under which conditions, borrowing information may become detrimental, we repeat the analysis on a simulated dataset. In the simulation experiment, we consider

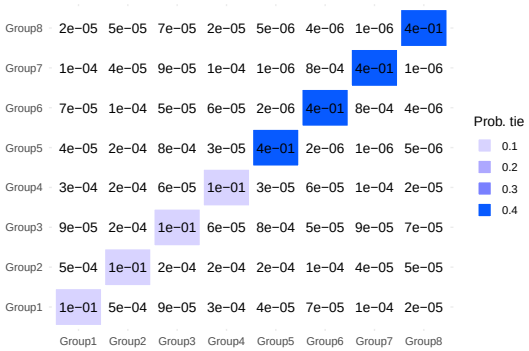

Figure 4: Probabilities of ties based on the *true* distributions in the simulated scenario.

eight populations. The true distribution of each arm is supported on a subset of 2,500 species randomly drawn from a total of 3,000, allowing for partial overlap of the supports across arms. Each arm follows a Zipf distribution, where the probability assigned to the kth most frequent species in population j is proportional to k^{-s_j}. We set $s_j = 1.3$ for $j = 1, 2, 3, 4$ and $s_j = 2$ for $j = 5, 6, 7, 8$ (cf. Battiston et al., 2018). However, before assigning the Zipf probability mass function, the 2,500 selected species in each population are randomly permuted, leading to markedly different probability mass functions and low probabilities of ties across populations. See Figure 4. This scenario represents a worst-case setting for non-independent rmSSPs: although some species are shared across populations, borrowing information is undesirable. Figure 5 and Table 6 report averages over 20 runs. Comparisons are made against both the uniform model and the oracle model, which selects the arm with the highest *true* frequency of unobserved species. The results show that even in this scenario, the hierarchical and additive rmSSPs perform on par with the independent models, and close to the oracle in terms of species discovery. This finding is reassuring, as it indicates that borrowing information, while unnecessary here, does not degrade performance.

	Simulated Scenario with low prob. of ties							
	Uniform	DP	PY	+DP	+PY	HDP	HPY	Oracle
Avg. num.	0.2335	0.3317	0.3298	0.3312	0.3262	0.3322	0.3237	0.3467
RMSE	NA	0.1563	0.0743	0.1621	0.0655	0.1929	0.0655	0

Table 6: Simulated scenario with low probability of ties across populations: Average number of species discovered per sampling step (Avg. num.) and root mean squared error (RMSE) of the estimated discovery probabilities in each population.

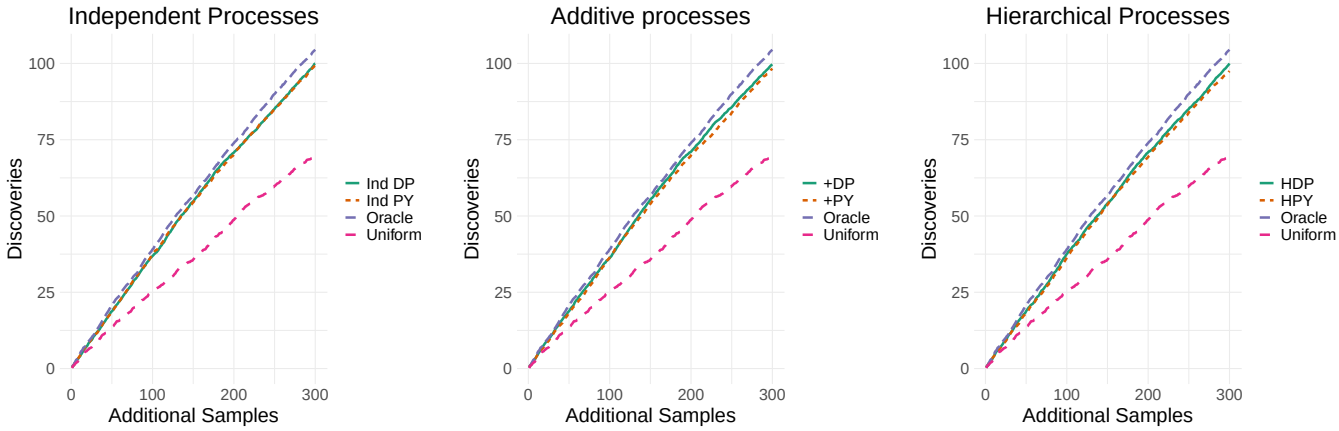

Figure 5: Simulated scenario with low probability of ties across populations: Number of species discovered as a function of the additional sample size in the rmSSPs, the uniform model, and the oracle model.

7 Conclusion

We introduced the class of mSSPs, a general framework extending Pitman's classical theory of species sampling models to the partially exchangeable setting. Our contribution is twofold. First, the mSSP framework provides a unifying perspective that encompasses most existing dependent nonparametric priors. It fundamentally advances the understanding of their behaviour by revealing that borrowing of information across groups is fully determined by ties within and across groups. These insights lead to principled strategies for prior specification, model calibration, and fair comparison across different subclasses of mSSPs. A systematic empirical comparison of competing models will be pursued in future work.

Second, our approach is constructive. It provides a modular recipe for building new models by combining EPPFs into structured dependence mechanisms. This allows for the design of both new classes of mSSPs and novel models satisfying alternative probabilistic symmetries beyond partial exchangeability. A first contribution along this path can be found in Fasano et al. (2025).

In addition to the systematic performance comparison of existing mSSPs and the development of new models, two further research directions emerge. Our finding that borrowing of information is entirely governed by ties suggests that standard dependence measures may not be well-suited to random discrete structures. This calls for a new theoretical framework based on the role of ties in generating, interpreting, and quantifying dependence. A second, more probabilistic direction is to develop a standalone framework for pEPPFs, decoupled from partially exchangeable arrays, that incorporates any sequential generative construction.

Acknowledgements

We are grateful to Jim Griffin, Fabrizio Leisen, Steven MacEachern, Li Ma, Peter Müller, Long Nguyen, Peter Orbanz, Riccardo Passeggeri, Judith Rousseau, and Aad van der Vaart for their valuable feedback and suggestions following presentations of this work at various seminars and conferences. B. Franzolini is supported by the National Recovery and Resilience Plan of Italy (PE1 FAIR - CUP B43C22000800006). A. Lijoi, I. Prünster and G. Rebaudo are partially supported by the European Union - NextGenerationEU PRIN-PNRR (project P2022H5WZ9).

References

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems". *Ann. Stat.* 2, 1152–1174.

Balocchi, C., S. Favaro, and Z. Naulet (2025). "Bayesian nonparametric inference for "species-sampling" problems". *Preprint at arXiv: 2203.06076*.

Balocchi, C., E. I. George, and S. T. Jensen (2023). "Clustering areal units at multiple levels of resolution to model crime incidence in Philadelphia". *Preprint at arXiv: 2112.02059*.

Bassetti, F., R. Casarin, and L. Rossini (2020). "Hierarchical species sampling models". *Bayesian Anal.* 15, 809–838.

Battiston, M., S. Favaro, and Y. W. Teh (2018). "Multi-armed bandit for species discovery: a Bayesian nonparametric approach". *J. Am. Stat. Assoc.* 113, 455–466.

Beraha, M., A. Guglielmi, and F. A. Quintana (2021). "The semi-hierarchical Dirichlet process and its application to clustering homogeneous distributions". *Bayesian Anal.* 16, 1187–1219.

Bi, D. and Y. Ji (2023). "A class of dependent random distributions based on atom skipping". *Preprint at arXiv: 2304.14954*, 1–75.

Camerlenghi, F., B. Dumitrascu, F. Ferrari, B. E. Engelhardt, and S. Favaro (2020). "Nonparametric Bayesian multiarmed bandits for single-cell experiment design". *Ann. Appl. Stat.* 14, 2003–2019.

Camerlenghi, F., D. B. Dunson, A. Lijoi, I. Prünster, and A. Rodríguez (2019). "Latent nested nonparametric priors (with discussion)". *Bayesian Anal.* 14, 1303–1356.

Camerlenghi, F., A. Lijoi, P. Orbanz, and I. Prünster (2019). "Distribution theory for hierarchical processes". *Ann. Stat.* 47, 67–92.

Camerlenghi, F., A. Lijoi, and I. Prünster (2017). "Bayesian prediction with multiple-samples information". *J. Multivar. Anal.* 156, 18–28.

Catalano, M., H. Lavenant, A. Lijoi, and I. P. and (2024). "A Wasserstein Index of dependence for random measures". *J. Am. Stat. Assoc.* 119, 2396–2406.

Catalano, M., A. Lijoi, and I. Prünster (2021). "Measuring dependence in the Wasserstein distance for Bayesian nonparametric models". *Ann. Stat.* 49, 2916–2947.

Chen, N. and J. J. Lee (2019). "Bayesian hierarchical classification and information sharing for clinical trials with subgroups and binary outcomes". *Biom. J.* 61, 1219–1231.

Colombi, A., R. Argiento, F. Camerlenghi, and L. Paci (2025). "Hierarchical mixture of finite mixtures". *Bayesian Anal.* in press.

Condit, R., N. Pitman, E. G. Leigh Jr., J. Chave, J. Terborgh, R. B. Foster, P. Núnez, S. Aguilar, R. Valencia, G. Villa, H. C. Muller-Landau, E. Losos, and S. P. Hubbell (2002). "Beta-diversity in tropical forest trees". *Science* 295, 666–669.

Crane, H. (2016). "The ubiquitous Ewens sampling formula". *Stat. Sci.* 31, 1–19.

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

De Blasi, P., A. Lijoi, and I. Prünster (2013). "An asymptotic analysis of a class of discrete nonparametric priors". *Stat. Sin.* 23, 1299–1321.

Denti, F., F. Camerlenghi, M. Guindani, and A. Mira (2023). "A common atom model for the Bayesian nonparametric analysis of nested data". *J. Am. Stat. Assoc.* 118, 405–416.

Durante, D., F. Gaffi, A. Lijoi, and I. Prünster (2025). "Partially exchangeable stochastic block models for (node-colored) multilayer networks". *J. Am. Stat. Assoc.* 121, forthcoming.

Ewens, W. J. (1990). "Population genetics theory - the past and the future". In: *Mathematical and Statistical Developments of Evolutionary Theory*. Ed. by S. Lessard. Vol. 299. Springer, pp. 177–227.

Fasano, A., A. Lijoi, I. Prünster, and G. Rebaudo (2025). "Probabilistic discovery of new species and homogeneous subpopulations". *Working Paper*.

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems". *Ann. Stat.* 1, 209–230.

de Finetti, B. (1938). "Sur la condition de "équivalence partielle"". *Actualités Scientifiques et Industrielles* 739, 5–18.

Ghosal, S. and A. van der Vaart (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Univ. Press.

Gil–Leyva, M. F. and R. H. Mena (2023). "Stick-breaking processes with exchangeable length variables". *J. Am. Stat. Assoc.* 118, 537–550.

Gnedin, A. V. (2010). "A species sampling model with finitely many types". *Electron. Commun. Probab.* 15, 79–88.

Gnedin, A. V. and J. Pitman (2006). "Exchangeable Gibbs partitions and Stirling triangles". *J. Math. Sci.* 138, 5674–5685.

Good, I. J. (1953). "The population frequencies of species and the estimation of population parameters". *Biometrika* 40, 237–264.

Good, I. J. and G. H. Toulmin (1956). "The number of new species, and the increase in population coverage, when a sample is increased". *Biometrika* 43, 45–63.

Green, P. J. and S. Richardson (2001). "Modelling heterogeneity with and without the Dirichlet process". *Scand. J. Stat.* 28, 355–375.

Griffin, J. E. and F. Leisen (2017). "Compound random measures and their use in Bayesian non-parametrics". *J. R. Stat. Soc. Series B Stat. Methodol.* 79, 525–545.

Horiguchi, A., C. Chan, and L. Ma (2024). "A tree perspective on stick-breaking models in covariate-dependent mixtures". *Bayesian Anal.* in press.

Ishwaran, H. and L. F. James (2001). "Gibbs sampling methods for stick-breaking priors". *J. Am. Stat. Assoc.* 96, 161–173. ISSN: 0162-1459.

James, L. F. (2008). "A discussion on: "The nested Dirichlet process" by Rodríguez, A., Dunson, D. B. and Gelfand, A." *J. Am. Stat. Assoc.* 103, 1131–1154.

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

— (2009). "Posterior analysis for normalized random measures with independent increments". *Scand. J. Stat.* 36, 76–97.

Lee, C. J., A. Zito, H. Sang, and D. B. Dunson (2025). "Logistic-beta processes for dependent random probabilities with beta marginals". *Bayesian Anal.* forthcoming.

Leisen, F. and A. Lijoi (2011). "Vectors of two-parameter Poisson–Dirichlet processes". *J. Multivar. Anal.* 102, 482–495.

Lijoi, A., R. H. Mena, and I. Prünster (2005). "Hierarchical mixture modeling with normalized inverse-Gaussian priors". *J. Am. Stat. Assoc.* 100, 1278–1291.

— (2007a). "Bayesian nonparametric estimation of the probability of discovering new species". *Biometrika* 94, 769–786.

— (2007b). "Controlling the reinforcement in Bayesian non-parametric mixture models". *J. R. Stat. Soc. Series B Stat. Methodol.* 69, 715–740.

— (2008). "A Bayesian nonparametric approach for comparing clustering structures in EST libraries". *J. Comput. Biol.* 15, 1315–1327.

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

Lijoi, A., I. Prünster, and G. Rebaudo (2023). "Flexible clustering via hidden hierarchical Dirichlet priors". *Scand. J. Stat.* 50, 213–234.

MacEachern, S. N. (1999). "Dependent nonparametric processes". In: *ASA Proc. Sect. Bayesian Stat. Sci.* Pp. 50–55.

— (2000). *Dependent Dirichlet processes*. Tech. rep. The Ohio State Univ., pp. 1–40.

Masoero, L., F. Camerlenghi, S. Favaro, and T. Broderick (2022). "More for less: predicting and maximizing genomic variant discovery via Bayesian nonparametrics". *Biometrika* 109, 17–32.

Miller, J. W. and M. T. Harrison (2018). "Mixture models with a prior on the number of components". *J. Am. Stat. Assoc.* 113, 340–356.

Müller, P., F. A. Quintana, and G. Rosner (2004). "A method for combining inference across related nonparametric Bayesian models". *J. R. Stat. Soc. Series B Stat. Methodol.* 66, 735–749.

Nobile, A. (1994). "Bayesian Analysis of Finite Mixture Distributions". PhD thesis. Carnegie Mellon Univ.

Nobile, A. and A. T. Fearnside (2007). "Bayesian finite mixtures with an unknown number of components: the allocation sampler". *Stat. Comput.* 17, 147–162.

Ouma, L. O., M. J. Grayling, J. Wason, and H. Zheng (2022). "Bayesian modelling strategies for borrowing of information in randomised basket trials". *J. R. Stat. Soc. Series C Appl. Stat.* 71, 2014–2037.

Pitman, J. (1995). "Exchangeable and partially exchangeable random partitions". *Probab. Theory Relat. Fields* 102, 145–158.

— (1996). "Some developments of the Blackwell-MacQueen urn scheme". *Lect. Notes-Monogr. Series* 30, 245–267.

— (2006). *Combinatorial Stochastic Processes*. Springer. ISBN: 3540342664.

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

Pyke, C. R., R. Condit, S. Aguilar, and S. Lao (2001). "Floristic composition across a climatic gradient in a neotropical lowland forest". *J. Veg. Sci.* 12, 553–566.

Quintana, F. A., P. Müller, A. Jara, and S. N. MacEachern (2022). "The dependent Dirichlet process and related models". *Stat. Sci.* 37, 24–41.

Regazzini, E., A. Lijoi, I. Prünster, et al. (2003). "Distributional results for means of normalized random measures with independent increments". *Ann. Stat.* 31, 560–585.

Richardson, S. and P. J. Green (1997). "On Bayesian analysis of mixtures with an unknown number of components (with discussion)". *J. R. Stat. Soc. Series B Stat. Methodol.* 59, 731–792.

Rodríguez, A., D. B. Dunson, and A. E. Gelfand (2008). "The nested Dirichlet process (with discussion)". *J. Am. Stat. Assoc.* 103, 1131–1154.

Sethuraman, J. (1994). "A constructive definition of Dirichlet priors". *Stat. Sin.* 4, 639–650.

Su, L., X. Chen, J. Zhang, and F. Yan (2022). "Comparative study of Bayesian information borrowing methods in oncology clinical trials". *JCO Precis. Oncol.* 6, 1–9.

Tavaré, S. (2021). "The magical Ewens sampling formula". *Bull. Lond. Math. Soc.* 53, 1563–1582.

Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). "Hierarchical Dirichlet processes". *J. Am. Stat. Assoc.* 101, 1566–1581.

Woodcock, J. and L. M. LaVange (2017). "Master protocols to study multiple therapies, multiple diseases, or both". *N. Engl. J. Med.* 377, 62–70.

Yan, Y. and X. Luo (2023). "Bayesian tree-structured two-level clustering for nested data Analysis". *J. Comput. Graph. Stat.* 32, 1185–1194.

Zuanetti, D. A., P. Müller, Y. Zhu, S. Yang, and Y. Ji (2018). "Clustering distributions with the marginalized nested Dirichlet process". *Biometrics* 74, 584–594.

Supplementary Materials for
Multivariate Species Sampling Models

Beatrice Franzolini, Antonio Lijoi, Igor Prünster
Bocconi Institute for Data Science and Analytics, Bocconi University

Giovanni Rebaudo
ESOMAS Department, University of Torino

S.1 Some basics on (univariate) species sampling

In classical species sampling problems, a random sample $(X_1, \dots, X_n)$ is extracted from an unknown and typically discrete distribution and each observed value corresponds to the species of a drawn individual. Denoting with P the unknown distribution of species in the population, we have

$$X_i \mid P \overset{iid}{\sim} P \qquad \text{for } i = 1, \dots, n.$$

To develop a Bayesian model for species sampling problems, a prior must be defined over the unknown distribution P. In the univariate setting, the problem can be tackled relying on the large class of priors provided by species sampling processes (SSP), introduced by Pitman (1996) as a generalization of the Dirichlet process of Ferguson (1973).

Definition S.1 (SSP)**.** A random probability measure P is a *species sampling process* (SSP) if

$$P \overset{a.s.}{=} \sum_{h\geq 1} \pi_h \delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_h\right) P_0,$$

where the atoms $(\theta_h)_{h\geq 1}$ are i.i.d. from the non-atomic distribution P_0 and are independent of the random sub-probability vector of the weights $\boldsymbol{\pi} = (\pi_h)_h$. Moreover, if $\sum_{h\geq 1} \pi_h \overset{a.s.}{=} 1$, P is said *proper*.

The corresponding model is defined once the observations are sampled independently from P given P.

Definition S.2 (SSM)**.** An infinite sequence of random variables $X_1, X_2, \dots$ follows a *species sampling model* (SSM) if it is exchangeable with an SSP directing measure. That is

$$\begin{aligned} X_i \mid P &\overset{\text{iid}}{\sim} P \quad (i = 1, 2, \dots) \\ P &\sim \text{SSP}(\mathcal{L}_\pi, P_0). \end{aligned} \tag{S.1}$$

Any sample $(X_1, \ldots, X_n)$ arising from a $P \sim \text{SSP}(\mathcal{L}_\pi, P_0)$ induces a random partition of the labels of the observations in the sample, i.e., of $[n] = \{1, \ldots, n\}$. More precisely, two observation labels i and l belong to the same block of the partition of $[n]$ (i.e., X_i and X_l are clustered together) if and only if $X_i = X_l$. The discrete part of the SSP entails that two observations are clustered together with positive probability since, unless $\sum_{h\geq 1} \pi_h \overset{a.s.}{=} 0$, $\mathbb{P}(X_i = X_l) > 0$. The law of such a random partition (denoted Π_n) of $[n]$ is characterized by the exchangeable partition probability functions (EPPF) (Pitman, 1996).

More precisely, let $\{C_1, \ldots, C_K\}$ an arbitrary partition of $[n]$ for a given $n \in \mathbb{N}$ and $n_k = |C_k|$ for $k \in [K]$ then

$$\mathbb{P}(\Pi_n = \{C_1, \ldots, C_k\}) = \text{EPPF}_K^{(n)}(n_1, \ldots, n_K). \tag{S.2}$$

In words, $\text{EPPF}_K^{(n)}(n_1, \ldots, n_K)$ can be interpreted as the probability of observing a particular (unordered) partition of n observations into K subsets of cardinalities $\{n_1, \ldots, n_K\}$. Note that the EPPF is defined on the space of the compositions of n, which can be interpreted as the space of the frequency of the partition in a given arbitrary order (e.g., the order of arrival). Let $P = \sum_{h\geq 1} \pi_h \delta_{\theta_h}$ be a proper SSP. Then the induced EPPF can be computed as

$$\text{EPPF}_K^{(n)}(n_1, \ldots, n_K) = \mathbb{E}\Big[\sum_{h_1 \neq \ldots \neq h_j} \prod_{k=1}^{K} \pi_{h_k}^{n_k} \Big]. \tag{S.3}$$

The EPPF characterizes the SSM (Pitman, 1996). For any $n \in \mathbb{N}$, if $(X_1, \ldots, X_n)$ arises from an SSM, its law can be obtained hierarchically as

1. sample the random partition Π_n from the induced EPPF obtained as in (S.3);
2. sample iid the unique values associated with each set in the partition from P_0.

The EPPF and the SSP can also be characterized by a specific sequence of predictive distributions (Pitman, 1996) also known as the *generalized Chinese restaurant process* (gCRP). In the culinary metaphor, we can think of observations corresponding to customers in a restaurant, who arrive sequentially and sit at an already occupied table or a new table and each table serves a different dish (iid sampled from P_0).

It is theoretically straightforward to derive the predictive distribution associated with any SSP via ratios of EPPFs as an application of the definition of conditional probability, leading to

$$\mathbb{P}\big(X_{n+1} = x \mid \boldsymbol{X}\big) = \begin{cases} \frac{\text{EPPF}_K^{(n+1)}(n_1,\ldots,n_k+1,\ldots,n_K)}{\text{EPPF}_K^{(n)}(n_1,\ldots,n_k,\ldots,n_K)} & \text{if } x = X_k^* \quad \text{and } k = 1, \ldots, K \\ \frac{\text{EPPF}_{K+1}^{(n+1)}(n_1,\ldots,n_k,\ldots,n_K,1)}{\text{EPPF}_K^{(n)}(n_1,\ldots,n_k,\ldots,n_K)} & \text{if } x = X_{K+1}^*, \end{cases} \tag{S.4}$$

where $(X_k^* : k = 1, \ldots, K)$ denote the K unique values of $X_1, \ldots, X_n$ that were recorded with frequency $n_1, \ldots, n_K$ and are iid sampled from P_0. See Pitman (1996), Pitman (2006), Lee et al. (2013), and Ghosal et al. (2017) for details and proofs about different characterizations of (univariate) SSM.

Although the analytical expression of the gCRP is available from the EPPF as shown in (S.4), such an expression does not reduce to simple and tractable quantities in general. However, a notable exception is the subclass of Gibbs-type prior (Gnedin et al., 2006; De Blasi et al., 2015), which, thanks to the product partition form of the EPPF, allows the ratio of EPPF in the gCRP to boil down to a simple ratio of constants for several notable examples, as in the well-known Chinese restaurant franchise (CRP) (Blackwell et al., 1973) induced by the Dirichlet process (DP) (Ferguson, 1973). The class of Gibbs-type prior is the most natural tractable generalization of the DP (De Blasi et al., 2015) and it includes the symmetric finite Dirichlet prior (Green et al., 2001), the Pitman-Yor process (PYP) (Pitman and Yor, 1997), the normalized inverse Gaussian (NIG) (Lijoi et al., 2005), the normalized generalized gamma process (NGGP) (Lijoi et al., 2007), mixture of finite symmetric Dirichlet (Nobile, 1994; Richardson et al., 1997; Nobile and Fearnside, 2007; Miller et al., 2018) and the mixture of DP (MDP) models (Antoniak, 1974). In the following sections, we recall the analytical expression of the three different characterizations (i.e., SSP, EPPF, and gCRP) of some relevant and tractable examples of Gibbs-type prior commonly used in Bayesian analysis.

S.1.1 Pitman-Yor process (PYP)

We say that an SSP($\mathcal{L}_\pi, P_0$) follow a Pitman-Yor process, i.e., $P \sim \text{PYP}(\alpha, \gamma; P_0)$, with P_0 a non-atomic measure if it is a proper SSP with $\mathcal{L}_\pi \sim \text{GEM}(\alpha, \gamma)$, where the two parameters GEM distribution, named after Griffiths, Engen, and McCloskey, can be thought as arising from the stick-breaking construction where the π_i's are such that $\pi_i = v_i \prod_{l=1}^{i-1} v_l$, with $v_i \sim \text{Beta}(1-\alpha, \gamma + i\alpha)$, $i \geq 1$, $\alpha \in [0, 1)$ and $\gamma > -\alpha$.

The following EPPF characterizes the PYP

$$\text{EPPF}_K^{(n)}(n_1, \ldots, n_K; \alpha, \gamma) = \frac{\prod_{k=1}^{K-1}(\gamma + k\,\alpha)}{(\gamma+1)_{n-1}} \prod_{k=1}^{K} (1-\alpha)_{n_k - 1}, \tag{S.5}$$

where $(x)_n = x(x+1)\cdots(x+n-1)$ is the nth ascending factorial.

Denoting with $X_1, X_2, \ldots$ an SSM from $P \sim \text{PYP}(\alpha, \gamma; P_0)$, we can derive the well-known gCRP of the PYP from the EPPF in (S.5) applying the definition of conditional probability.

$$\mathbb{P}\big(X_{n+1} = x \mid \boldsymbol{X}\big) = \begin{cases} \frac{n_k - \alpha}{\gamma + n} & \text{if } x = X_k^* \quad \text{and } k = 1, \ldots, K \\ \frac{\gamma + \alpha K}{\gamma + n} & \text{if } x = X_{K+1}^*. \end{cases} \tag{S.6}$$

S.1.2 Dirichlet process (DP)

If we consider $P \sim \text{PYP}(\alpha, \gamma; P_0)$ as in the previous section and we restrict $\alpha = 0$ and $\gamma > 0$ we obtain the relevant special case of the Dirichlet process, i.e., $P \sim \text{DP}(\gamma; P_0)$. Thus we can specialize the distribution of the weights to GEM(γ), the induced EPFF in (S.5) that boils

down to

$$\mathrm{EPPF}_K^{(n)}(n_1,\dots,n_K;\alpha,\gamma) = \frac{\gamma^K\Gamma(\gamma)}{\Gamma(\gamma+n)}\prod_{k=1}^{K}(n_k-1)!, \tag{S.7}$$

and the corresponding CRP

$$\mathbb{P}\big(X_{n+1}=x \mid \boldsymbol{X}\big) = \begin{cases} \frac{n_k}{\gamma+n} & \text{if } x = X_k^* \quad \text{and } k=1,\dots,K \\ \frac{\gamma}{\gamma+n} & \text{if } x = X_{K+1}^*. \end{cases} \tag{S.8}$$

S.1.3 Finite symmetric Dirichlet multinomial (symDM)

Here we consider an SSP P with a fixed known number M of species in the population (with $M \in \mathbb{N}$) that follow a finite-dimensional symmetric Dirichlet multinomial (symDM). That is, for a fixed $M \in \mathbb{N}$,

$$P = \sum_{h=1}^{M} \pi_h \delta_{\theta_h}, \tag{S.9}$$

where $(\pi_1,\dots,\pi_M) \sim \mathrm{Dir}(\tau,\dots,\tau) \perp \theta_h \overset{\text{iid}}{\sim} P_0$. We write $P \sim \mathrm{DM}_M(\tau, P_0)$.

Then we can derive the induced EPFF as

$$\mathrm{EPPF}_K^{(n)}(n_1,\dots,n_K) = \frac{M!}{(M-K)!}\frac{\Gamma(\tau\,M)}{\Gamma(n+\tau\,M)\Gamma(\tau)^K}\prod_{k=1}^{K}\Gamma(n_k+\tau). \tag{S.10}$$

and the corresponding gCRP

$$\mathbb{P}\big(X_{n+1}=x \mid \boldsymbol{X}\big) \propto \begin{cases} n_k+\tau & \text{if } x = X_k^* \quad \text{and } k=1,\dots,K \\ \rho(M-K)\mathbf{1}(K\neq M) & \text{if } x = X_{K+1}^*. \end{cases} \tag{S.11}$$

S.1.4 Gnedin Process (GN)

Allowing for an unknown M in a finite-dimensional symmetric Dirichlet multinomial process, the model becomes a mixture of symmetric Dirichlet models. A relevant example is the *Gnedin process* (with discount parameter equals to -1). The corresponding EPPF is

$$\mathrm{EPPF}_K^{(n)}(n_1,\dots,n_K) = \sum_{m=1}^{\infty} \mathrm{EPPF}_K^{(n)}(n_1,\dots,n_K \mid M=m)\, p(M=m), \tag{S.12}$$

where $\mathrm{EPPF}_K^{(n)}(n_1,\dots,n_K \mid M=m)$ is the EPPF of the M-symmetric Dirichlet prior in (S.10), with $\rho=1$ and $p(M=m) = \frac{\gamma(1-\gamma)_{m-1}}{m!}$, $\gamma \in (0,1)$.

The corresponding gCRP boils down to the following simple tractable expression

$$\mathbb{P}\big(X_{n+1}=x \mid \boldsymbol{X}\big) \propto \begin{cases} (n_k+1)(n-K+\gamma) & \text{if } x = X_k^* \quad \text{and } k=1,\dots,K \\ K^2-K\gamma & \text{if } x = X_{K+1}^*. \end{cases} \tag{S.13}$$

We denote the corresponding SSP with $P \sim \mathrm{GN}(\gamma, P_0)$.

S.2 Proofs

S.2.1 Proof of Proposition 1

The proof follows trivially from the Definition of mSSP.

S.2.2 Proof of Proposition 2

Proof. To prove the statement, we want to show the non-trivial implication of the iff, i.e., if (P_1, P_2) are an mSSP with non-atomic base measure P_0 they can be rewritten as in (1), that is

$$P_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}^{(1,2)} \delta_{\theta_{0,h}} + \sum_{h'\geq 1} \pi_{j,h'}^{(j)} \delta_{\theta_{j,h'}} + \pi_{j,0}^{(j)} P_0, \qquad \text{for } j = 1, 2. \tag{S.14}$$

where $\sum_{h\geq 1} \pi_{j,h}^{(1,2)} + \sum_{h'\geq 0} \pi_{j,h'}^{(j)} = 1$, for $j = 1, 2$, the atoms are independent from the weights and such that $\theta_{j,h} \overset{\text{iid}}{\sim} P_0$, for $j = 0, 1, 2,\ \ h = 1, 2, \ldots$ and $\mathbb{P}[\pi_{1,h}^{(1,2)} > 0, \pi_{2,h}^{(1,2)} > 0] > 0$.

From the definition of mSSP, we write for $j = 1, 2$

$$P_j = \sum_{h\geq 1} \pi_{j,h} \delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) P_0 = \sum_{h\in\mathcal{H}} \pi_{j,h} \delta_{\theta_h} + \left(1 - \sum_{h\in\mathcal{H}} \pi_{j,h}\right) P_0,$$

where we denote by $\mathcal{H} \coloneqq \{1, 2, \ldots\}$ the set of the indexes of the two sums. Note that $H \coloneqq \text{card}(\mathcal{H}) \in \{0\} \cup \mathbb{N} \cup \{\infty\}$, and we use the convention that, for any $(x_h)_h$, $\sum_{h=1}^{0} x_h = \sum_{h\in\emptyset} x_h = 0$. We define $\pi_{j,0}^{(j)} \coloneqq 1 - \sum_{h\in\mathcal{H}} \pi_{j,h}$ and we partition $\mathcal{H}$ in $\{\mathcal{H}_0, \bar{\mathcal{H}}_0\}$, where

$$\mathcal{H}_0 \coloneqq \{h \in \mathcal{H} :\ \Pr[\pi_{1,h} > 0, \pi_{2,h} > 0] > 0\} = \{h \in \mathcal{H} :\ \Pr[\pi_{1,h}\pi_{2,h} > 0] > 0\}$$

is the set of shared atoms and $\bar{\mathcal{H}}_0 = \mathcal{H} \setminus \mathcal{H}_0$.

Let us define, for $j = 1, 2$,

$$\left(\left(\theta_{0,h}, \pi_{j,h}^{(12)}\right) : h \in \mathcal{H}_0\right) \coloneqq \left((\theta_h, \pi_{j,h}) : h \in \mathcal{H}_0\right) \text{ and } \left(\pi_{j,h}^{(j)} : h \in \bar{\mathcal{H}}_0\right) \coloneqq \left(\pi_{j,h} : h \in \bar{\mathcal{H}}_0\right),$$

and $\theta_{j,h} \overset{\text{iid}}{\sim} P_0$, for $j = 1, 2$ and $h \in \bar{\mathcal{H}}_0$, independent from all the previous random variables, i.e., $\left((\theta_{0,h}, \pi_{j,h}^{(12)}) : h \in \mathcal{H}_0\right)$ and $\left(\pi_{j,h}^{(j)} : h \in \bar{\mathcal{H}}_0\right)$.

Then note that, for $j = 1, 2$,

$$\sum_{h\in\mathcal{H}} \pi_{j,h} \delta_{\theta_h} = \sum_{h\in\mathcal{H}_0} \pi_{j,h} \delta_{\theta_h} + \sum_{h\in\bar{\mathcal{H}}_0} \pi_{j,h} \delta_{\theta_h}$$

and

$$\sum_{h\in\mathcal{H}_0} \pi_{j,h} \delta_{\theta_h} = \sum_{h\in\mathcal{H}_0} \pi_{j,h}^{(12)}\, \delta_{\theta_{0,h}} \text{ and } \sum_{h\in\bar{\mathcal{H}}_0} \pi_{j,h} \delta_{\theta_h} = \sum_{h\in\bar{\mathcal{H}}_0} \pi_{j,h}^{(j)}\, \delta_{\theta_{j,h}}.$$

To conclude the proof, we just relabel the indexes in both $\mathcal{H}_0$ and $\bar{\mathcal{H}}_0$ such that they are ordered integers starting from 1 with no gaps and remap the elements in the corresponding sums accordingly.

□

S.2.3 Proof of Proposition 3

Proof. By the law of iterated expectations, the first and second moments of $P_j(A)$ are equal to

$$\begin{aligned}\mathbb{E}[P_j(A)] &= \mathbb{P}(X_{j,i} \in A) = P_0(A)\\ \mathbb{E}[P_j(A)^2] &= \mathbb{P}(X_{j,i} \in A, X_{j,l} \in A), \text{ with } i \neq l.\end{aligned}$$

Disintegrating with respect to $\{X_{j,i} = X_{j,l}\}$ to recover independence leads to

$$\begin{aligned}&\mathbb{P}(X_{j,i} \in A, X_{j,l} \in A)\\ &\quad= \mathbb{P}(X_{j,i} = X_{j,l})\mathbb{P}(X_{j,i} \in A, X_{j,l} \in A \mid X_{j,i} = X_{j,l})\\ &\qquad+ \mathbb{P}(X_{j,i} \neq X_{j,l})\mathbb{P}(X_{j,i} \in A, X_{j,l} \in A \mid X_{j,i} \neq X_{j,l})\\ &\quad= \mathbb{P}(X_{j,i} = X_{j,l})P_0(A) + \mathbb{P}(X_{j,i} \neq X_{j,l})P_0(A)^2.\end{aligned}$$

Finally, $\mathbb{V}\text{ar}[P_j(A)] = \mathbb{E}[P_j(A)^2] - \mathbb{E}[P_j(A)]^2 = \mathbb{P}(X_{j,i} = X_{j,l})P_0(A)[1 - P_0(A)]$. □

S.2.4 Proof of Proposition 4

Proof. For $j \neq k$, by the law of iterated expectations, we get

$$\mathbb{E}[P_j(A)P_k(A)] = \mathbb{P}(X_{j,i} \in A, X_{k,m} \in A).$$

Disintegrating with respect to $\{X_{j,i} = X_{k,m}\}$ to recover independence leads to

$$\begin{aligned}&\mathbb{P}(X_{j,i} \in A, X_{k,m} \in A)\\ &\quad= \mathbb{P}(X_{j,i} = X_{k,m})\mathbb{P}(X_{j,i} \in A, X_{k,m} \in A \mid X_{j,i} = X_{k,m})\\ &\qquad+ \mathbb{P}(X_{j,i} \neq X_{k,m})\mathbb{P}(X_{j,i} \in A, X_{k,m} \in A \mid X_{j,i} \neq X_{k,m})\\ &\quad= \mathbb{P}(X_{j,i} = X_{k,m})P_0(A) + \mathbb{P}(X_{j,i} \neq X_{k,m})P_0(A)^2.\end{aligned}$$

Thus, $\mathbb{C}\text{ov}[P_j(A), P_k(A)] = \mathbb{P}(X_{j,i} = X_{k,m})P_0(A)[1 - P_0(A)]$. The correlation is obtained using Proposition 3. □

S.2.5 Proof of Corollary 1

The proof follows trivially from Proposition 4.

S.2.6 Proof of Proposition 5

Proof. By definition of mSSP, we know that

$$P_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) P_0 \quad \text{and} \quad P_k \overset{a.s.}{=} \sum_{h\geq 1} \pi_{k,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{k,h}\right) P_0.$$

Moreover, by Cauchy-Schwarz inequality, we have a.s. that

$$\sqrt{\sum_{h\geq 1}\pi_{j,h}^2+\left(1-\sum_{h\geq 1}\pi_{j,h}\right)^2}\sqrt{\sum_{h\geq 1}\pi_{k,h}^2+\left(1-\sum_{h\geq 1}\pi_{k,h}\right)^2}$$
$$\geq \sum_{h\geq 1}\pi_{j,h}\pi_{k,h}+\left(1-\sum_{h\geq 1}\pi_{j,h}\right)\left(1-\sum_{h\geq 1}\pi_{k,h}\right).$$

Assume by contradiction that the event $\{\pi_{j,h}\neq\pi_{k,h}$ for at least one $k\}$ has positive probability. This implies that with positive probability, the above inequality is strict and thus, with positive probability, we have

$$\mathbb{P}(X_{j,1}=X_{k,1}\mid P_j,P_k)<\sqrt{\mathbb{P}(X_{j,1}=X_{j,2}\mid P_j,P_k)}\sqrt{\mathbb{P}(X_{k,1}=X_{k,2}\mid P_j,P_k)}$$

which implies

$$\mathbb{P}(X_{j,1}=X_{k,1})<\mathbb{P}(X_{j,1}=X_{j,2})\mathbb{P}(X_{k,1}=X_{k,2})$$

and

$$\mathbb{C}\mathrm{or}[P_j(A),P_k(A)]<1$$

. Therefore, we have $\pi_{j,h}=\pi_{k,h}$ a.s., for all h, and thus $P_j\overset{a.s.}{=}P_k$. □

S.2.7 Proof of Theorem 6

Proof. Clearly $P_j\perp P_k$ entails $\mathbb{C}\mathrm{or}[P_j(A),P_k(A)]=0$. We want to show that $\mathbb{C}\mathrm{or}[P_j(A),P_k(A)]=0$ entails $P_j\perp P_k$.

Let us consider the representation of (P_j,P_k) as mixtures of two components

$$P_j\overset{a.s.}{=}\omega_j^{(j,k)}\sum_{h\geq 1}\bar{\pi}_{j,h}^{(j,k)}\delta_{\theta_h}+\left(1-\omega_j^{(j,k)}\right)\left(\bar{\pi}_{j,0}^{(j)}P_0+\sum_{h'\geq 1}\bar{\pi}_{j,h'}^{(j)}\delta_{\theta_{j,h'}}\right)$$

and

$$P_k\overset{a.s.}{=}\omega_k^{(j,k)}\sum_{h\geq 1}\bar{\pi}_{k,h}^{(j,k)}\delta_{\theta_h}+\left(1-\omega_k^{(j,k)}\right)\left(\bar{\pi}_{j,0}^{(k)}P_0+\sum_{h'\geq 1}\bar{\pi}_{k,h'}^{(k)}\delta_{\theta_{k,h'}}\right).$$

where

$$\bar{\pi}_{j,h}^{(j,k)}=\frac{\pi_{j,h}^{(j,k)}}{\sum_{\ell\geq 1}\pi_{j,\ell}^{(j,k)}},\qquad \bar{\pi}_{j,h'}^{(j)}=\frac{\pi_{j,h'}^{(j)}}{\sum_{\ell\geq 0}\pi_{j,\ell}^{(j)}},\quad\text{and}\quad \omega_j^{(j,k)}=\sum_{h\geq 1}\pi_{j,h}^{(j,k)}.$$

Recall that by Proposition 4 and Corollary 1, $\mathbb{C}\mathrm{or}[P_j(A),P_k(A)]=0$ iff $\mathbb{P}(X_{1,j}=X_{1,k})=0$.

Note that

$$\mathbb{P}(X_{j,1}=X_{k,1})=\sum_{h\geq 1}\mathbb{E}\left[\omega_j^{(j,k)}\bar{\pi}_{j,h}^{(j,k)}\,\omega_k^{(j,k)}\bar{\pi}_{k,h}^{(j,k)}\right]\geq\mathbb{E}\left[\omega_j^{(j,k)}\bar{\pi}_{j,1}^{(j,k)}\,\omega_k^{(j,k)}\bar{\pi}_{k,1}^{(j,k)}\right].$$

Therefore, by Definition of rmSSP, we have that $\omega_j^{(j,k)} \overset{a.s.}{=} \omega_k^{(j,k)} \overset{a.s.}{=} 0$.

Indeed, if we assume by contradiction that (w.l.o.g.) $\mathbb{P}\big(\omega_j^{(j,k)} > 0\big) > 0$ than by Definition 3 we have that $\mathbb{P}\big(\omega_j^{(j,k)}\bar{\pi}_{j,1}^{(j,k)}\omega_k^{(j,k)}\bar{\pi}_{k,1}^{(j,k)} > 0\big) > 0$ that entails

$$\mathbb{P}(X_{j,1} = X_{k,1}) \geq \mathbb{E}\left[\omega_j^{(j,k)}\bar{\pi}_{j,1}^{(j,k)}\,\omega_k^{(j,k)}\bar{\pi}_{k,1}^{(j,k)}\right] > 0$$

that contradicts $\mathbb{P}(X_{1,j} = X_{1,k}) = 0$. Since $\omega_j^{(j,k)} \overset{a.s.}{=} \omega_k^{(j,k)} \overset{a.s.}{=} 0$ we can rewrite

$$P_j \overset{a.s.}{=} \bar{\pi}_{j,0}^{(j)} P_0 + \sum_{h' \geq 1} \bar{\pi}_{j,h'}^{(j)} \delta_{\theta_{j,h'}}$$

and

$$P_k \overset{a.s.}{=} \bar{\pi}_{j,0}^{(k)} P_0 + \sum_{h' \geq 1} \bar{\pi}_{k,h'}^{(k)} \delta_{\theta_{k,h'}}$$

and therefore $P_j \perp P_k$. □

S.2.8 Proof of Proposition 7

Proof. Define the random variable Z, so that $Z = 1$, if $X_{j,i} = X_{k,m}$, and $Z = 0$, otherwise.

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(X_{j,i}, X_{k,m}) &= \mathbb{E}\left[\mathbb{C}\mathrm{ov}(X_{j,i}, X_{k,m} \mid Z)\right] + \mathbb{C}\mathrm{ov}\left(\mathbb{E}\left[X_{j,i} \mid Z\right], \mathbb{E}\left[X_{k,m} \mid Z\right]\right) \\
&= \mathbb{E}\left[\mathbb{C}\mathrm{ov}(X_{j,i}, X_{k,m} \mid Z)\right] + 0 \\
&= \mathbb{C}\mathrm{ov}(X_{j,i}, X_{k,m} \mid Z = 1)\mathbb{P}(X_{j,i} = X_{k,m}) \\
&= \mathbb{P}(X_{j,i} = X_{k,m})\mathbb{V}\mathrm{ar}(X^*).
\end{aligned}$$

where $X^* \sim P_0$ and $\mathbb{C}\mathrm{ov}(X_{j,i}, X_{k,m} \mid Z = 1) = \mathbb{V}\mathrm{ar}(X^*)$ is obtained since the conditioning to $Z = 1$ implies that both observations are equal to the same atom, which is itself sampled from P_0. The final result follows trivially by dividing the expression of the covariance by $\sqrt{\mathbb{V}\mathrm{ar}(X_{j,i})\mathbb{V}\mathrm{ar}(X_{k,m})} = \mathbb{V}\mathrm{ar}(X_{j,i}) = \mathbb{V}\mathrm{ar}(X^*)$. □

S.2.9 Proof of Corollary 2

The proof follows trivially from Proposition 7.

S.2.10 Proof of Proposition 8

Proof. Define $\boldsymbol{X}_{j,1:q} = (X_{j,1}, \ldots, X_{j,q})$,

$$\mathbb{E}[P_j(A)^q] = \mathbb{P}(\boldsymbol{X}_{j,1:q} \in A^q).$$

Disintegrate with respect to the random partition $\Pi_q^{(j)}$ induced by the ties in $\boldsymbol{X}_{j,1:q}$ and taking values in the set $\mathscr{P}(\boldsymbol{X}_{j,1:q})$ to recover independence and aggregate by symmetry induced by

exchangeability. $K_q^{(j)}$ denotes the number of sets in $\Pi_q^{(j)}$.

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}_{j,1:q} \in A^q\right) &= \sum_{\Pi_q^{(j)} \in \mathscr{P}(\boldsymbol{X}_{j,1:q})} \mathbb{P}\big[\boldsymbol{X}_{j,1:q} \in A^q \mid \Pi_q^{(j)}\big]\, \mathbb{P}\big(\Pi_q^{(j)}\big) \\ &= \sum_{s=1}^{q} P_0(A)^s \sum_{\Pi_q^{(j)} \in \mathscr{P}(\boldsymbol{X}_{j,1:q}) : K_q^{(j)} = s} \mathbb{P}\big(\Pi_q^{(j)}\big) \\ &= \sum_{s=1}^{q} P_0(A)^s \mathbb{P}(K_q^{(j)} = s) = \mathbb{E}\big[P_0(A)^{K_q^{(j)}}\big].\end{aligned}$$

□

S.2.11 Proof of Proposition 9

Proof. For notational convenience, we prove the proposition for $h = 2$. The general case can be proven with the same argument. Notation is the same as in the proof of Proposition 7.

$$\mathbb{E}[P_j(A_1)^{q_1} P_j(A_2)^{q_2}] = \mathbb{P}\left(\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}\right),$$

where $q = q_1 + q_2$. Denote now with $\mathscr{A}_{q_1,q_2} \subset \mathscr{P}(\boldsymbol{X}_{j,1:q})$ the set of all possible partitions $\Pi_q^{(j)}$ induced by the ties in $\boldsymbol{X}_{j,1:q}$ such that the elements in $\boldsymbol{X}_{j,1:q_1}$ and in $\boldsymbol{X}_{j,q_1+1:q_2}$ do not have ties. It follows that

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}\right) &= \mathbb{P}[(\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2}) \cap (\Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2})] \\ &= \sum_{s_1=1}^{q_1} \sum_{s_2=1}^{q_2} \mathbb{P}(\Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2}, K_{q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2) \\ &\times \mathbb{P}(\boldsymbol{X}_{j,1:q} \in A_1^{q_1} \times A_2^{q_2} \mid \Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2}, K_{q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2) \\ &= \sum_{s_1=1}^{q_1} \sum_{s_2=1}^{q_2} P_0(A_1)^{s_1} P_0(A_2)^{s_2} \mathbb{P}(\Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2}, K_{q_1}^{(j)} = s_1, K_{q_1+1:q_2}^{(j)} = s_2) \\ &= \mathbb{E}\left[P_0(A_1)^{K_{q_1}^{(j)}} P_0(A_2)^{K_{q_1+1:q_2}^{(j)}} \mid \Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2}\right] \mathbb{P}(\Pi_q^{(j)} \in \mathscr{A}_{q_1,q_2}).\end{aligned}$$

□

S.2.12 Proof of Theorem 10

Proof.

$$\mathbb{E}[P_1(A)^{q_1} \cdots P_J(A)^{q_J}] = \mathbb{P}\left(\boldsymbol{X}_{j,1:q_j} \in A^{q_j} : j = 1, \ldots, J\right).$$

Disintegrate with respect to the possible partitions Π_q of $\boldsymbol{X}_{1:q_1,\dots,1:q_J}$ to recover independence and aggregate by symmetry.

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}_{j,1:q_j} \in A^{q_j} : j = 1,\dots,J\right) &= \sum_{\Pi_q \in \mathscr{P}(\boldsymbol{X}_{1:q_1,\dots,1:q_J})} \mathbb{P}\big(\boldsymbol{X}_{j,1:q_j} \in A^{q_j} : j = 1,\dots,J \mid \Pi_q\big)\mathbb{P}\big(\Pi_q\big) \\ &= \sum_{s=1}^{q} P_0(A)^s \sum_{\Pi_q \in \mathscr{P}(\boldsymbol{X}_{1:q_1,\dots,1:q_J}) : K_{q_1,\dots,q_J}=s} \mathbb{P}\big(\Pi_q\big) \\ &= \sum_{s=1}^{q} P_0(A)^s \mathbb{P}(K_{q_1,\dots,q_J} = s) = \mathbb{E}\big[P_0(A)^{K_{q_1,\dots,q_J}}\big].\end{aligned}$$

□

S.2.13 Proof of Theorem 11

Proof. First note that

$$\mathbb{E}\bigg(\prod_{j=1}^{J} P_j(A_j)^{q_j}\bigg) = \mathbb{P}\bigg(\boldsymbol{X}_{1:q_1,\dots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}\bigg).$$

Denote now with $\mathscr{A}_{q_1,\dots,q_J} \subset \mathscr{P}(\boldsymbol{X}_{1:q_1,\dots,1:q_J})$ the set of all possible partitions Π_q of the elements in $\boldsymbol{X}_{1:q_1,\dots,1:q_J}$ such that the elements in $\boldsymbol{X}_{j,1:q_1}$ and in $\boldsymbol{X}_{j',1:q_{j'}}$ do not belong to the same set, for any $j \neq j'$ according to Π_q.

$$\begin{aligned}\mathbb{P}\left(\boldsymbol{X}_{1:q_1,\dots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}\right) &= \mathbb{P}\left[\left(\boldsymbol{X}_{1:q_1,\dots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j}\right) \cap (\Pi_q \in \mathscr{A}_{q_1,\dots,q_J})\right] \\ &= \mathbb{P}\left(\Pi_q \in \mathscr{A}_{q_1,\dots,q_J}\right) \mathbb{P}\left(\boldsymbol{X}_{1:q_1,\dots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j} \mid \Pi_q \in \mathscr{A}_{q_1,\dots,q_J}\right) \\ &= \sum_{s_1=1}^{q_1} \cdots \sum_{s_J=1}^{q_J} \mathbb{P}\left(\Pi_q \in \mathscr{A}_{q_1,\dots,q_J}, K_{q_1}^{(1)} = s_1, \dots, K_{q_J}^{(J)} = s_J\right) \\ &\times \mathbb{P}\left(\boldsymbol{X}_{1:q_1,\dots,1:q_J} \in \bigtimes_{j=1}^{J} A_j^{q_j} \mid \Pi_q \in \mathscr{A}_{q_1,\dots,q_J}, K_{q_1}^{(1)} = s_1, \dots, K_{q_J}^{(J)} = s_J\right) \\ &= \sum_{s_1=1}^{q_1} \cdots \sum_{s_J=1}^{q_J} P_0(A_1)^{s_1} \cdots P_0(A_J)^{s_J} \\ &\times \mathbb{P}\left(\Pi_q \in \mathscr{A}_{q_1,\dots,q_J}, K_{q_1}^{(1)} = s_1, \dots, K_{q_J}^{(J)} = s_J\right) \\ &= \mathbb{E}\left[P_0(A_1)^{K_{q_1}^{(1)}} \cdots P_0(A_J)^{K_{q_J}^{(J)}} \mid \Pi_q \in \mathscr{A}_{q_1,\dots,q_J}\right] \mathbb{P}(\Pi_q \in \mathscr{A}_{q_1,\dots,q_J}).\end{aligned}$$

□

S.2.14 Proof of Theorem 12

Proof. To prove the theorem, we first show that when the random array follows an mSSM, then it can be obtained by sampling first the partition from the corresponding pEPPF and then associating unique values sampled independently from P_0 to each partition set. Formally, for any family of sets $(A_{j,i} : i \in [I_j], j \in [J])$,

$$\begin{aligned}
&\mathbb{P}\left[(X_{j,i} : i \in [I_j],\, j \in [J]) \in (A_{j,i} : i \in [I_j],\, j \in [J])\right] \\
&= \sum_{\Pi_n \in \mathbb{P}[\Pi_n]\, \mathscr{P}(\boldsymbol{X}_{1:I_1,\ldots,1:I_J})} \mathbb{P}[(X_{j,i} : i \in [I_j],\, j \in [J]) \in (A_{j,i} : i \in [I_j],\, j \in [J]) \mid \Pi_n]\mathbb{P}[\Pi_n] \\
&= \sum_{\substack{\{C_1,\ldots,C_D\} \\ \in \mathscr{P}(\boldsymbol{X}_{1:I_1,\ldots,1:I_J})}} \mathbb{P}[\Pi_n = \{C_1,\ldots,C_D\}] \prod_{d=1}^{D} P_0\left(\bigcap_{(j,i):\sum_{j'=1}^{j-1} I_{j'}+i \in C_d} A_{j,i}\right)
\end{aligned}$$

where $C_1, \ldots, C_D$ are the sets in Π_n, whose elements are collected according to the order of arrival by groups. Now, to complete the proof, what is left to show is that when the pEPPF is obtained from an arbitrary vector of random probability measures according to (6) in the main paper, then there always exists an mSSP that induces the same pEPPF.

To this aim, let us first consider a pair of dependent random probability measures (not necessarily mSSP) that admit the following representation

$$P_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) P_0, \qquad \text{for } j = 1, 2 \tag{S.15}$$

where P_0 is a non-atomic (deterministic) distribution on a space $\mathbb{X}$, $\boldsymbol{\pi}_j = (\pi_{j,h})_{h\geq 1}$ is a random sub-probability sequence, for $j = 1, 2$, and the sequence of $(\theta_h)_{h\geq 1}$, conditionally on $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$, follows any joint distribution such that $\mathbb{P}[\theta_h = \theta_{h'} \mid \boldsymbol{\pi}] = 0$, for any $h \neq h'$. Let define $\omega_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}$, such that

$$P_j = \omega_j \tilde{P}_j + (1 - \omega_j)P_0 \quad \text{for } j = 1, 2$$

where $\tilde{P}_j \overset{a.s.}{=} \sum_{h\geq 1} \frac{\pi_{j,h}}{\omega_j}\delta_{\theta_h} =: \sum_{h\geq 1} \tilde{\pi}_{j,h}\delta_{\theta_h}$.

$$\begin{aligned}
\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) &= \mathbb{E}\left[\int_{\mathbb{X}_*^D} \prod_{d=1}^{D} P_1(\mathrm{d}x_d)^{n_{1,d}} P_2(\mathrm{d}x_d)^{n_{2,d}}\right] \\
&= \mathbb{E}\left\{\int_{\mathbb{X}_*^D} \prod_{d=1}^{D}\prod_{j=1}^{2}\left[\omega_j \tilde{P}_j(\mathrm{d}x_d) + (1-\omega_j)P_0(\mathrm{d}x_d)\right]^{n_{j,d}}\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\int_{\mathbb{X}_*^D} \prod_{d=1}^{D}\prod_{j=1}^{2}\left[\tilde{P}_j(\mathrm{d}x_d)^{z_{j,d}} \times P_0(\mathrm{d}x_d)^{1-z_{j,d}}\right]^{n_{j,d}} \mid \omega_j\right]\right\}
\end{aligned}$$

where $z_{j,d} \mid \omega_j \overset{ind}{\sim} \text{Bernoulli}(\omega_j)$ and where the product measure of the non-atomic component P_0 equals zero on the diagonal, meaning

$$\int_{\mathbb{X}} P_0(\mathrm{d}x_d)^n = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n \geq 2 \end{cases}$$

Thus,

$$\begin{aligned}
\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1, \boldsymbol{n}_2) = \mathbb{E}\Bigg\{ \int_{\mathbb{X}_*^D} \Bigg[& \prod_{d\in[D]:n_{1,d}+n_{2,d}\geq 1} \prod_{j=1}^{2} \tilde{P}_j(\mathrm{d}x_d)^{z_{j,d}\, n_{j,d}}\, P_0(\mathrm{d}x_d)^{(1-z_{j,d})\, n_{j,d}} \\
& \times \prod_{d\in[D]:n_{1,d}+n_{2,d}=1} \prod_{j=1}^{2} \tilde{P}_j(\mathrm{d}x_d)^{z_{j,d}\, n_{j,d}}\, P_0(\mathrm{d}x_d)^{(1-z_{j,d})\, n_{j,d}} \Bigg] \Bigg\} \\
= \mathbb{E}\Bigg\{ \int_{\mathbb{X}_*^D} \Bigg[& \prod_{d\in[D]:n_{1,d}+n_{2,d}\geq 1} \prod_{j=1}^{2} \tilde{P}_j(\mathrm{d}x_d)^{z_{j,d}\, n_{j,d}} \\
& \times \prod_{d\in[D]:n_{1,d}=1,n_{2,d}=0} \tilde{P}_1(\mathrm{d}x_d)^{z_{1,d}}\, P_0(\mathrm{d}x_d)^{(1-z_{1,d})} \\
& \times \prod_{d\in[D]:n_{1,d}=0,n_{2,d}=1} \tilde{P}_2(\mathrm{d}x_d)^{z_{2,d}}\, P_0(\mathrm{d}x_d)^{(1-z_{2,d})} \Bigg] \Bigg\} \\
= \mathbb{E}\Bigg[& \left(\sum_{h_1\neq\ldots\neq h_{D'}} \prod_{j=1}^{J} \prod_{d\in[D]:n_{1,d}+n_{2,d}\geq 1} \tilde{\pi}_{j,h_d}^{n_{j,d}\, z_{j,d}} \right) \\
& \times \left(\sum_{h_1\neq\ldots\neq h_{D''}} \prod_{j=1}^{J} \prod_{d\in[D]:n_{1,d}=1,n_{2,d}=0} \tilde{\pi}_{1,h_d}^{z_{1,d}} \right) \\
& \times \left(\sum_{h_1\neq\ldots\neq h_{D'''}} \prod_{j=1}^{J} \prod_{d\in[D]:n_{1,d}=0,n_{2,d}=1} \tilde{\pi}_{2,h_d}^{z_{2,d}} \right) \Bigg]
\end{aligned}$$

Importantly, the expression and derivation above for the pEPPF also hold when (P_1, P_2) is an mSSP. Crucially, this expression depends only on the law of the weights of the random probability measures in (S.15). Since the class of mSSPs imposes no restriction on this law, one can always choose an mSSP whose pEPPF matches the expression for any prescribed weight distribution. More precisely, given a pEPPF derived from P_1 and P_2 in (S.15), the corresponding mSSP may be constructed by adopting the same weight law of P_1 and P_2 and selecting an arbitrary non-atomic base measure.

To extend the argument to arbitrary pairs of probabilities not necessarily satisfying (S.15), note that they can always be decomposed into an atomic and a non-atomic component. More precisely, let (G_1, G_2) be an arbitrary pair of random probability measures. For each $j = 1, 2$

we can rewrite

$$G_j \overset{a.s.}{=} \sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h} + \left(1 - \sum_{h\geq 1} \pi_{j,h}\right) G_{0,j}, \qquad \text{for } j = 1, 2$$

where G_{0j} is a non-atomic random distribution on a space $\mathbb{X}$, $\boldsymbol{\pi}_j = (\pi_{j,h})_{h\geq 1}$ is a random sub-probability sequence, for $j = 1, 2$, and the sequence of $(\theta_h)_{h\geq 1}$, conditionally on $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$, follows any joint distribution such that $\mathbb{P}[\theta_h = \theta_{h'} \mid \boldsymbol{\pi}] = 0$, for any $h \neq h'$. Note that the pair (G_1, G_2) induced the same pEPPF if we substitute the random non-atomic distributions $G_{0,1}$ and $G_{0,2}$ with a common deterministic arbitrary distribution P_0.

Thus, the pEPPF induced by an arbitrary pair of random probabilities (G_1, G_2) can be obtained as the pEPPF induced by a pair of random probabilities (P_1, P_2) satisfying (S.15) and thus by a pair of mSSP.

Finally, note that extending the above argument to $J \geq 2$ is a matter only of notation. □

S.2.15 Proof of Proposition 13

Proof. The proof follows by reasoning analogous to the derivation of the pEPPF for generic measures in the proof of Theorem 12. However, since here we assume the mSSP is proper, its non-atomic components vanish, and the derivation simplifies to:

$$\begin{aligned}
\text{pEPPF}_D^{(n)}(\boldsymbol{n}_1, \ldots, \boldsymbol{n}_J) &= \mathbb{E}\left[\int_{\mathbb{X}_*^D} \prod_{d=1}^{D} P_1(\mathrm{d}x_d)^{n_{1,d}} \ldots P_J(\mathrm{d}x_d)^{n_{J,d}}\right] \\
&= \mathbb{E}\left\{\int_{\mathbb{X}_*^D} \prod_{d=1}^{D}\prod_{j=1}^{J}\left[\sum_{h\geq 1} \pi_{j,h}\delta_{\theta_h}(\mathrm{d}x_d)\right]^{n_{j,d}}\right\} \\
&= \mathbb{E}\left[\sum_{h_1\neq\ldots\neq h_D} \prod_{j=1}^{J}\prod_{d=1}^{D} \pi_{j,h_d}^{n_{j,d}}\right].
\end{aligned}$$

□

S.2.16 Proof of Proposition 14

The proof follows trivially from the definition of conditional probability.

S.3 Algorithms and models details for the multi-armed bandit illustration

The algorithms used for all six strategies considered in Section 6 are Markov chain Monte Carlo marginal algorithms. These algorithms are obtained using the augmented representation of the pEPPF described in Section 5 for the additive and hierarchical processes, and the

sequential sampling schemes detailed in Section S1 for the independent processes. All models are generalized to accommodate random hyperparameters to achieve greater flexibility in the learning mechanisms, leading to the following specifications for the six strategies.

- Independent Dirichlet Process

$$\begin{aligned} X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\ P_j \mid \alpha_j &\overset{\text{ind}}{\sim} \text{DP}(\alpha_j) \\ \alpha_j &\overset{\text{iid}}{\sim} \text{Gamma}(0.75, 1). \end{aligned}$$

  where $\text{Gamma}(a, b)$ denotes a Gamma distribution with expected value equal to a/b.

- Independent Pitman-Yor Process

$$\begin{aligned} X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\ P_j \mid \sigma_j, \alpha_j &\overset{\text{ind}}{\sim} \text{PYP}(\sigma_j, \alpha_j) \\ \sigma_j \overset{\text{iid}}{\sim} \text{Beta}(1, 3) \qquad \alpha_j &\overset{\text{iid}}{\sim} \text{Gamma}(0.2, 1). \end{aligned}$$

  where Beta denotes a Beta distribution.

- Additive Dirichlet Process

$$\begin{aligned} X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\ P_j &= \epsilon_j \, Q_0 + (1 - \epsilon_j) Q_j \\ \epsilon_j &\overset{\text{iid}}{\sim} 0.15 \, \delta_0 + 0.15 \, \delta_1 + 0.7 \, \text{Uniform}(0, 1) \\ Q_j \mid \alpha_j &\overset{\text{ind}}{\sim} \text{DP}(\alpha_j) \quad \text{for } j = 0, 1, \ldots, J \\ \alpha_0 \overset{\text{iid}}{\sim} \text{Gamma}(0.5, 2), \qquad \alpha_j &\overset{\text{iid}}{\sim} \text{Gamma}(6, 2) \quad \text{for } j = 1, 2, \ldots, J. \end{aligned}$$

- Additive Pitman-Yor Process

$$\begin{aligned} X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\ P_j &= \epsilon_j \, Q_0 + (1 - \epsilon_j) Q_j \\ \epsilon_j &\overset{\text{iid}}{\sim} 0.1 \, \delta_0 + 0.1 \, \delta_1 + 0.8 \, \text{Uniform}(0, 1) \\ Q_j \mid \sigma_j, \alpha_j &\overset{\text{ind}}{\sim} \text{PYP}(\sigma_j, \alpha_j) \quad \text{for } j = 0, 1, \ldots, J \\ \sigma_0 \overset{\text{iid}}{\sim} \text{Beta}(1, 3), \qquad \sigma_j &\overset{\text{iid}}{\sim} \text{Beta}(1, 2) \quad \text{for } j = 1, 2, \ldots, J \\ \alpha_0 \overset{\text{iid}}{\sim} \text{Gamma}(0.25, 4), \qquad \alpha_j &\overset{\text{iid}}{\sim} \text{Gamma}(2, 2) \quad \text{for } j = 1, 2, \ldots, J. \end{aligned}$$

- Hierarchical Dirichlet Process

$$
\begin{aligned}
X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\
P_j \mid Q &\overset{iid}{\sim} \text{DP}(\alpha, Q), \\
Q \sim &\text{DP}(\alpha_0, P_0) \\
\alpha_0 \overset{\text{iid}}{\sim} \text{Gamma}(1, 1/3), &\qquad \alpha_j \overset{\text{iid}}{\sim} \text{Gamma}(1, 1/2) \quad \text{for } j = 1, 2, \ldots, J.
\end{aligned}
$$

- Hierarchical Pitman-Yor Process

$$
\begin{aligned}
X_{j,i} \mid (P_1, \ldots, P_J) &\overset{\text{ind}}{\sim} P_j \quad \text{for } i = 1, 2, \ldots \\
P_j \mid Q &\overset{iid}{\sim} \text{PYP}(\sigma, \alpha, Q), \\
Q \sim &\text{PYP}(\sigma_0, \alpha_0, P_0) \\
\sigma_0 \overset{\text{iid}}{\sim} \text{Beta}(1, 2), &\qquad \sigma_j \overset{\text{iid}}{\sim} \text{Beta}(1, 2) \quad \text{for } j = 1, 2, \ldots, J \\
\alpha_0 \overset{\text{iid}}{\sim} \text{Gamma}(1, 1), &\qquad \alpha_j \overset{\text{iid}}{\sim} \text{Gamma}(1, 1) \quad \text{for } j = 1, 2, \ldots, J.
\end{aligned}
$$

The choice of the parameters of the hyperpriors on discount and concentration parameters is performed as follows. We use the values suggested in Battiston et al. (2018) for the Hierarchical Pitman-Yor Process, and then we fix the ones of the other strategies by considering the probabilities of ties as a function of the hyperparameters and approximately match their expected values and variances. This selection procedure ensures a fair performance comparison, as the probabilities of ties provide an excellent summary of dependence for rmSSPs. The resulting expected probability of ties and corresponding variances are reported in Table S.1.

Model	$\mathbb{E}$[**prob tie across**]	$\mathbb{V}$[**prob tie across**]	$\mathbb{E}$[**prob tie within**]	$\mathbb{V}$[**prob tie within**]
Independent DP	0	0	0.672	0.049
Independent PYP	0	0	0.669	0.047
+DP	0.388	0.092	0.666	0.052
+PY	0.400	0.064	0.628	0.038
HDP	0.389	0.056	0.671	0.041
HPY	0.397	0.043	0.638	0.033

Table S.1: Expected probabilities of ties within and across as functions of the hyperparameters and corresponding variances. Values are obtained via Monte Carlo approximation by simulating 2000 samples of the hyperparameters from the hyperpriors of each model.

To sample the concentration parameters of the Dirichlet processes, we employed a Gibbs Sampler via an augmented representation of the full-conditional of the concentration parameter, avoiding a Metropolis within the Gibbs step. For the hyperparameters of the Pitman-Yor processes, we devised an adaptive Metropolis-Hasting, obtained via 10 repeated steps within

the main Gibbs algorithm. At each of the 30 sequential sampling steps of the multi-armed bandit problem, we perform 200 iterations of the MCMC algorithm, leading to a total of 6000 iterations (not including the Metropolis-Hasting steps, when present) per strategy. After observing a new data point in a sequential step, we initialize the MCMC for the next step with a warm start based on the last iteration of the MCMC output in the previous step. For instance, we initialize the values of the hyperparameters with the last sampled value in the previous MCMC chain that targets their posterior distribution without conditioning on the new data point.

Moreover, for hierarchical processes, we perform 1000 iterations of the MCMC before estimating the probability of discovery at the first sequential sampling step to achieve a warm start also at the first sampling step. Code for all six strategies is freely available at https://github.com/GiovanniRebaudo/MSSP.

References

Antoniak, C. E. (1974). "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems". *Ann. Stat.* 2, 1152–1174.

Battiston, M., S. Favaro, and Y. W. Teh (2018). "Multi-armed bandit for species discovery: a Bayesian nonparametric approach". *J. Am. Stat. Assoc.* 113, 455–466.

Blackwell, D. and J. B. MacQueen (1973). "Ferguson distributions via Pólya urn schemes". *Ann. Stat.* 1, 353–355.

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

Ferguson, T. S. (1973). "A Bayesian analysis of some nonparametric problems". *Ann. Stat.* 1, 209–230.

Ghosal, S. and A. van der Vaart (2017). *Fundamentals of Nonparametric Bayesian Inference*. Cambridge Univ. Press.

Gnedin, A. V. and J. Pitman (2006). "Exchangeable Gibbs partitions and Stirling triangles". *J. Math. Sci.* 138, 5674–5685.

Green, P. J. and S. Richardson (2001). "Modelling heterogeneity with and without the Dirichlet process". *Scand. J. Stat.* 28, 355–375.

Lee, J., F. A. Quintana, P. Müller, and L. Trippa (2013). "Defining predictive probability functions for species sampling models". *Stat. Sci.* 28, 209–222.

Lijoi, A., R. H. Mena, and I. Prünster (2005). "Hierarchical mixture modeling with normalized inverse-Gaussian priors". *J. Am. Stat. Assoc.* 100, 1278–1291.

— (2007). "Controlling the reinforcement in Bayesian non-parametric mixture models". *J. R. Stat. Soc. Series B Stat. Methodol.* 69, 715–740.

Miller, J. W. and M. T. Harrison (2018). "Mixture models with a prior on the number of components". *J. Am. Stat. Assoc.* 113, 340–356.

Nobile, A. (1994). "Bayesian Analysis of Finite Mixture Distributions". PhD thesis. Carnegie Mellon Univ.

Nobile, A. and A. T. Fearnside (2007). "Bayesian finite mixtures with an unknown number of components: the allocation sampler". *Stat. Comput.* 17, 147–162.

Pitman, J. (1996). "Some developments of the Blackwell-MacQueen urn scheme". *Lect. Notes-Monogr. Series* 30, 245–267.

— (2006). *Combinatorial Stochastic Processes*. Springer. ISBN: 3540342664.

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

Richardson, S. and P. J. Green (1997). "On Bayesian analysis of mixtures with an unknown number of components (with discussion)". *J. R. Stat. Soc. Series B Stat. Methodol.* 59, 731–792.