Sankhyā : The Indian Journal of Statistics
2024, Volume 86-A, Supplement 1, pp. S255-S287

A Unified Approach to Hierarchical Random Measures

Marta Catalano
Luiss University, Rome, Italy
Claudio Del Sole, Antonio Lijoi and Igor Prünster
Bocconi University, Milan, Italy

Abstract

Hierarchical models enjoy great popularity due to their ability to handle heterogeneous groups of observations by leveraging on their underlying common structure. In a Bayesian nonparametric framework, the hierarchy is introduced at the level of group-specific random measures, and then translated to the observations' level via suitable transformations. In this work, we propose a new strategy to derive closed-form expressions for the marginal and posterior distributions of each group. Indeed, by directly inserting a suitable set of latent variables into the generative model for the data, we unravel a common core shared by the different hierarchical constructions proposed in the Bayesian nonparametric literature. Specifically, we identify a key identity that underlies these models and highlight its role in the derivation of quantities of interest.

AMS (2000) subject classification. 62F15; 62G05; 60G57.
Keywords and phrases. Completely random measure, dependence structure, hierarchical process, mixture hazard, normalized random measure, partial exchangeability.

1 Introduction

In a Bayesian framework, $\mathbb{X}$–valued observations from a homogeneous population are modeled as an (infinitely extendable) exchangeable sequence, which means that their distribution does not depend on the order of appearance (de Finetti, 1937). By de Finetti's representation theorem, this is equivalent to stating that the observations $(X_n)_{n\geq 1}$ are conditionally independent and identically distributed (i.i.d.) given a random probability measure $\tilde{P}$, taking values in the space of probability measures on $\mathbb{X}$, denoted as $\mathscr{P}$. Hence, one has

$$\begin{aligned} X_1, \dots, X_n \mid \tilde{P} &\overset{\text{i.i.d.}}{\sim} \tilde{P}, \qquad n \geq 1, \\ \tilde{P} &\sim \mathcal{Q}, \end{aligned}$$

where $\mathcal{Q}$ is known as the de Finetti measure and acts as prior distribution for Bayesian inference. If $\mathcal{Q}$ does not degenerate on a subset of $\mathscr{P}$ indexed by a finite-dimensional parameter, we are considering a nonparametric setting, the one of interest to the present paper. The model choice then reduces to identifying a suitable distribution for the random probability measure $\tilde{P}$. Probability measures may be characterized in many different ways, including their probability mass or density function, cumulative distribution function or survival function, cumulative hazard, and hazard rate function. Each representation highlights different features of the distribution, and may be preferred to the others depending on the main target of the analysis. For this reason, different Bayesian nonparametric models have focused on each of these representations (with seminal contributions in Doksum (1974), Dykstra and Laud (1981), Ferguson (1973, 1974), Hjort (1990), Lo (1984), Walker and Muliere (1997)), which share one common feature: the random probability measure $\tilde{P}$ is modeled as a suitable transformation of a completely random measure $\tilde{\mu}$. Indeed, completely random measures (Kingman, 1967) are a remarkable class of discrete random measures that lends itself to modeling both discrete functions, thanks to their almost-sure discrete nature, and continuous ones, typically through kernel smoothing. Moreover, completely random measures feature an infinite number of random atoms and random weights, which guarantees full-modeling flexibility of many quantities of interest; see Lijoi and Prünster (2010) for a review using completely random measures as unifying concept. We will consider almost surely finite completely random measures without drift and fixed atoms, which can be conveniently represented as

$$\tilde{\mu}(dx) = \sum_{i \geq 1} J_i \, \delta_{\theta_i}(dx),$$

where $(J_i)_{i\geq 1}$ is a sequence of random jumps and $(\theta_i)_{i\geq 1}$ is a sequence of i.i.d. atoms (points of discontinuity), whose common distribution is termed *base probability measure*. In view of the upcoming discussion, it is relevant to point out that, when such a measure is diffuse, the atoms are distinct almost surely.

Recent developments in Bayesian nonparametrics are focused on flexible ways to account for different forms of heterogeneity across observations; see Cifarelli and Regazzini (1978) and MacEachern (1999, 2000) for pioneering works and Quintana et al. (2022) for a recent review. A particularly relevant framework is that of partial exchangeability (de Finetti, 1938), which allows to model multiple groups of observations sharing similar features, with homogeneity (exchangeability) holding only within each group but not across groups. Typical instances include patients with the same disease treated in different hospitals, or children of the same age raised in different countries. In this setting, the distribution in each group can be modeled by means of a group-specific completely random measure, and since the groups share similar features, it is important to incorporate borrowing of information across different groups. This goal is naturally met by the Bayesian paradigm: if dependence among the random measures is introduced a priori, the posterior distribution for each group will also make use of the information contained in the other groups. This typically induces a shrinkage effect that makes the estimates more reliable and disappears as the number of observations diverges.

The previous discussion raises the fundamental question of how to introduce dependence among random measures, which has been addressed from several different perspectives in the vast literature on the topic. De Iorio et al. (2004) develop the dependent Dirichlet process framework of MacEachern (1999, 2000) by imposing an ANOVA-type structure on the atoms, while Dunson and Park (2008) and Rodríguez and Dunson (2011) model predictor-dependent weights via kernel and probit transformations. As for the partially exchangeable setting, proposals inducing dependence across different groups of observations include additive (Müller et al., 2004, Lijoi et al., 2014), nested (Rodriguez et al., 2008, Camerlenghi et al., 2019a, Lijoi et al. 2023) and hierarchical structures (Teh et al., 2006, Camerlenghi et al., 2019b, 2021). Further interesting constructions of dependent completely random measures, based on multivariate Lévy intensities, can be found in Epifani and Lijoi (2010), Griffin and Leisen (2017), Lau and Cripps (2022), Riva-Palacio and Leisen (2021). See also the recent review by Quintana et al. (2022) and references therein.

Among these constructions, hierarchical forms of dependence are arguably the most natural ones for a Bayesian statistician: being used to introduce dependence among the observations through conditional independence, it is conceptually straightforward to introduce dependence among the random measures through conditional independence as well. Thanks to de Finetti's representation theorem, this leads to an (infinitely extendable) exchangeable sequence of random measures. A compelling strategy to define conditionally independent completely random measures consists in assuming a random base measure, which is modeled either through a normalized completely random measure or directly through a completely random measure. The first approach has been mainly used to model dependent random discrete probability measures (Teh et al., 2006, Camerlenghi et al., 2019b), whereas the second has been used to provide the main ingredients to model dependent random hazard functions (Camerlenghi et al., 2021), though it is probably interesting to remark that, in principle, they could both be used to model both quantities.

These two classes of hierarchical models entail different dependence assumptions on the random measures; however, they also present significant conceptual and mathematical similarities. In this paper, we investigate these similarities and propose a unifying framework that sheds light on their common structure and on intriguing analogies in their posterior and predictive representations. Indeed, even if dealing with hierarchical models may appear more challenging than treating simpler exchangeable ones, they both rely on the same identity, which can be applied recursively to reduce the analysis of the multi-group framework to the easier single-group scenario. Specifically, consider a completely random measure $\tilde{\mu}$ with a diffuse base probability measure, i.e. whose atoms are distinct almost surely. For any non-negative measurable function f, and mutually disjoint balls $B_\varepsilon(x_j^*) = \{x : d(x, x_j^*) < \varepsilon\}$, one can explicitly characterize the limiting behavior of

$$\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(x)\,\tilde{\mu}(dx)\right\}\prod_{j=1}^{k}\tilde{\mu}(B_\varepsilon(x_j^*))^{n_j}\right], \tag{1}$$

as $\varepsilon \to 0$, as recalled in (14) of Section 5. From this identity, one can derive the law of the random partition of the exchangeable observations according to their ties, which in turn determines both the predictive and posterior distributions, as shown in James et al. (2006, 2009).

At first sight, it seems difficult to extend this strategy to the multi-group hierarchical framework. Indeed, the base probability measure of each exchangeable completely random measure is modeled as an almost-surely discrete random measure itself. This implies that the atoms of such measures display ties with positive probability and thus do not fit into the above framework. Camerlenghi et al. (2019b, 2021) work around this issue by exploiting the celebrated Faà di Bruno's formula, expressing higher order derivatives of compositions of functions: eventually, the inherent combinatorial structure induced by this formula turns out to be effectively represented by introducing suitable latent variables. Leveraging on this observation, here we propose an alternative approach that bypasses the Faà di Bruno Formula by directly inserting the latent variables into the data generative model. From an analytical point of view, this strategy substantially mitigates the combinatorial burden connected to the Faà di Bruno's formula, at the negligible cost of an augmented Lévy intensity measure. In addition, the description of the induced random partition structure becomes more transparent, as it can be considered a mere consequence of the ties within sequences of latent (unobserved) variables.

The intuition behind our proposal is the following: by adding a diffuse independent mark to each atom of the exchangeable random measure, one derives a completely random measure on the joint space of the atoms and the marks with the compelling property of not displaying ties almost surely. One can, then, use (1) on the augmented random measure and possibly remove the auxiliary latent marks through marginalization. Remarkably, this strategy leads to posterior and predictive representations for both classes of hierarchical models, thanks to a fundamental identity that extends (1) to hierarchical random measures. In Basu and Tiwari (1982), which stands out as an insightful contribution to the foundations of the Dirichlet process, the authors state a crucial desirable property for nonparametric models: "If a prior is selected from this class, then the posterior distribution given a sample of observations from P is manageable analytically, and it belongs to the class, i.e. the class is closed under 'Bayesian operation'." Our paper shows that the two considered classes of hierarchical nonparametric priors meet this *desideratum* by developing the necessary analytical tools and suitably extending the notion of closure to fit the more complex partially exchangeable framework.

The paper is structured as follows. Section 2 recalls the definition of completely random measure (CRM), along with two transformations that are important for the following developments, namely, normalization of CRMs to model discrete random probability measures and kernel mixtures of CRMs to model random hazards. Section 3 introduces dependence between CRMs through two related hierarchical constructions. Section 4 provides some intuition on the partition structure and introduces a convenient set of latent variables, representing the marks of the atoms in the previous discussion. Its formal treatment can be found in Section 5, together with the key identity for closed-form computations involving CRMs, which is then exploited to derive marginal distributions. The posterior characterization of CRMs is described in Section 6 for both classes of hierarchical models. In light of these results, the generalized gamma CRM arises as the natural conjugate prior, as discussed in Section 7. Section 8 provides further insights on the dependence structure between hierarchical random measures, together with some intuition on how to enhance its flexibility.

2 Background on Completely Random Measures

Let $(\mathbb{X}, \mathcal{X})$ be a complete and separable metric space. Denote by $\mathbb{M}$ the space of boundedly finite measures on $(\mathbb{X}, \mathcal{X})$ equipped with the corresponding Borel σ-algebra $\mathcal{M}$, that is, the smallest σ-algebra that makes the projections $A \mapsto \mu(A)$ measurable for every measure μ and every bounded set A; see Daley and Vere-Jones (2007) for details. A random element $\tilde{\mu}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $(\mathbb{M}, \mathcal{M})$ is termed *random measure.*

A Completely Random Measure (CRM) $\tilde{\mu}$ is a random measure on $(\mathbb{X}, \mathcal{X})$ such that, for any collection of $n \geq 1$ bounded and pairwise disjoint sets $A_1, \ldots, A_n \in \mathcal{X}$, the random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$ are mutually independent (Kingman, 1967). CRMs represent a very natural and convenient choice of discrete random measures: their key role in Bayesian nonparametrics is extensively discussed in Lijoi and Prünster (2010).

In this work, we consider CRMs without fixed points of discontinuity and without deterministic drift: a CRM belonging to this class can be represented as a linear functional of a Poisson Random Measure (PRM) $\tilde{N}$ on $\mathbb{R}^+ \times \mathbb{X}$,

$$\tilde{\mu}(dx) \overset{\mathrm{d}}{=} \int_{\mathbb{R}^+} s\, \tilde{N}(ds, dx). \tag{2}$$

Therefore, its realizations are almost surely discrete, and its law is characterized by the Laplace functional transform at any non-negative measurable function $f\colon \mathbb{X} \mapsto \mathbb{R}^+$, namely

$$\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(x)\, \tilde{\mu}(dx)\right\}\right] = \exp\left\{-\int_{\mathbb{R}^+ \times \mathbb{X}} (1 - e^{-s\, f(x)})\, \nu(ds, dx)\right\}, \tag{3}$$

where ν is the Lévy intensity measure uniquely identifying $\tilde{\mu}$. Note that ν is the mean intensity measure of the PRM $\tilde{N}$ in (2), and must satisfy the condition

$$\int_{\mathbb{R}^+ \times \mathbb{X}} \min(s, 1)\, \nu(ds, dx) < \infty.$$

This characterization motivates the notation $\tilde{\mu} \sim \mathrm{CRM}(\nu)$. The Lévy intensity measure ν can be always disintegrated as

$$\nu(ds, dx) = \rho(ds|x)\, \alpha(dx), \tag{4}$$

where $\rho\colon \mathcal{B}(\mathbb{R}^+) \times \mathbb{X} \mapsto \mathbb{R}^+$ is a transition kernel and α is a diffuse σ-finite measure on $(\mathbb{X}, \mathcal{X})$. Lévy intensities, and the corresponding CRMs, are termed *homogeneous* if $\rho(\cdot|x) = \rho(\cdot)$ is a measure not depending on $x \in \mathbb{X}$, and *non-homogeneous* otherwise. An exhaustive account on CRMs can be found in Kingman (1993).

In the following, homogeneous CRMs are considered for their simplicity and tractability. Intuitively, this is equivalent to the independence between atoms and jumps of the random measure. Moreover, the measure α appearing in (4) is assumed to be finite, which ensures that the random measure is finite almost surely. This implies that the total mass of α can be included into the transition kernel ρ, and the Lévy intensity measure is uniquely disintegrated as

$$\nu(ds, dx) = \rho(ds)\, P_0(dx),$$

with P_0 a diffuse probability measure on $(\mathbb{X}, \mathcal{X})$ termed *base probability measure*; later on, it will be sometimes useful to rely on this representation rather than on (4). The *infinite activity* property of the Lévy intensity measure is also assumed, namely $\rho(\mathbb{R}^+) = \infty$, which implies that the corresponding random measure has an infinite number of jumps on any bounded set, and thus ensures it is non-zero almost surely. Further details can be found in Regazzini et al. (2003).

In Bayesian Nonparametrics, (completely) random measures can be effectively used as the basic building block for the construction of discrete nonparametric priors. This work focuses on random probability measures and random hazards obtained from two suitable transformations of random measures: normalization and kernel mixtures.

Normalized random measures An almost surely discrete random probability measure $\tilde{p}$ can be defined via normalization of a random measure $\tilde{\mu}$ as

$$\tilde{p}(dx) := \frac{\tilde{\mu}(dx)}{\tilde{\mu}(\mathbb{X})}, \tag{5}$$

provided that $0 < \tilde{\mu}(\mathbb{X}) < \infty$ almost surely. In case $\tilde{\mu}$ is a homogeneous CRM, these conditions are guaranteed by the finiteness of α and the infinite activity property, as proved in Regazzini et al. (2003), where this normalization procedure has first been introduced. A posterior characterization of normalized CRMs has been derived in James et al. (2009) for the exchangeable setting. Popular special instances include the Dirichlet process (Ferguson, 1973), which arises from normalization of gamma CRMs (Ferguson, 1974), the normalized σ–stable process (Kingman, 1975), the normalized inverse Gaussian process (Lijoi et al., 2005) and the normalized generalized gamma process (Lijoi et al., 2007). Their use in mixture models is reviewed in Barrios et al. (2013). Henceforth, whenever $\tilde{\mu} \sim \mathrm{CRM}(\nu)$, its normalization $\tilde{p}$ in (5) is denoted by $\tilde{p} \sim \mathrm{NCRM}(\nu)$.

Random mixture hazards In survival analysis, the time to failure is usually modeled by a random variable T, taking values on $\mathbb{R}^+$, whose probability distribution is here assumed to be absolutely continuous with respect to the Lebesgue measure. The hazard rate function of T represents the instantaneous risk of failure, and is defined as

$$h(t) := \frac{f(t)}{S(t)}, \qquad t \in \mathbb{R},$$

where f is the density function of T and S is its survival function. A random hazard rate $\tilde{h}$ can be defined mixing a non-negative kernel k over a random measure $\tilde{\mu}$ as

$$\tilde{h}(t) := \int_{\mathbb{X}} k(t;x)\,\tilde{\mu}(dx). \tag{6}$$

Such mixture structure was introduced in pioneering papers by Dykstra and Laud (1981) and Lo and Weng (1989), and developed in full generality by James (2005), where the posterior characterization is derived for exchangeable data.

The random survival function associated to this random hazard rate is

$$\tilde{S}(t) = \exp\left\{ -\int_0^t \int_{\mathbb{X}} k(s;x)\,\tilde{\mu}(dx)\,ds \right\}, \tag{7}$$

which is a proper survival function whenever $\lim_{t\to\infty} \tilde{S}(t) = 0$ almost surely. In case $\tilde{\mu}$ is a homogeneous CRM, this condition is guaranteed by infinite activity and the condition

$$\int_{\mathbb{R}^+} k(s;x)\,ds = \infty, \qquad P_0 - \text{a.s.}$$

The most popular kernel was proposed in Dykstra and Laud (1981), where the authors assume $\mathbb{X} = \mathbb{R}^+$ and define $k(t;x) = \beta(x)\,\mathbb{1}_{t\geq x}$, for β positive and right-continuous function. Such kernel satisfies the condition above and represents the reference model for increasing hazard rates. Further methodological, computational and asymptotic investigations with different kernels and choices of CRMs can be found, e.g., in Ishwaran and James (2004), Nieto-Barajas and Walker (2004), Peccati and Prünster (2008), De Blasi et al. (2009), Donnet et al. (2017), Catalano et al. (2020).

3 Hierarchical Random Measures

Hierarchical structures represent a natural way to construct vectors of dependent random measures. Indeed, dependence in a vector of random measures $\tilde{\boldsymbol{\mu}}$ can be induced through the following hierarchical scheme:

$$\begin{aligned} \tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_D) \mid \tilde{\mu}_0 &\overset{\text{i.i.d.}}{\sim} \tilde{\mathcal{G}}, \\ \tilde{\mu}_0 &\sim \mathcal{G}_0, \end{aligned} \tag{8}$$

where $\tilde{\mathcal{G}}$ is the (random) conditional distribution of each $\tilde{\mu}_d$, for $d = 1, \dots, D$, given the random measure $\tilde{\mu}_0$, which represents the root of the hierarchy and is distributed according to $\mathcal{G}_0$.

Completely random measures are particularly well-suited to this hierarchical structure, as the random measure $\tilde{\mu}_0$ at the root of the hierarchy can be easily incorporated into the Lévy intensity measure characterizing the distribution $\tilde{\mathcal{G}}$ of the vector at the lower level of the hierarchy. We distinguish two different hierarchical constructions that are commonly used to model dependent random probabilities and dependent hazards, respectively.

A vector of dependent discrete random probability measures can be defined through the hierarchical structure

$$\begin{aligned} \tilde{\boldsymbol{p}} = (\tilde{p}_1, \dots, \tilde{p}_D) \mid \tilde{p}_0 &\overset{\text{i.i.d.}}{\sim} \text{NCRM}(\tilde{\nu}), \\ \tilde{p}_0 &\sim \text{NCRM}(\nu_0), \end{aligned} \tag{9}$$

where $\tilde{p}_1, \dots, \tilde{p}_D$ are conditionally independent normalized CRMs with random Lévy intensity measure

$$\tilde{\nu}(ds, dx) = \rho(ds)\, \tilde{p}_0(dx),$$

and $\tilde{p}_0$ is a normalized CRM with Lévy intensity measure $\nu_0(ds, dx) = \rho_0(ds)\, P_0(dx)$. This hierarchical nonparametric construction was first introduced in Teh et al. (2006) for the special case of hierarchical Dirichlet processes, and extensively studied in Camerlenghi et al. (2019b) for the more general class of processes considered here. Further investigations beyond the Dirichlet setup can be found, e.g., in Teh and Jordan (2010), Camerlenghi et al. (2017, 2018), Catalano et al. (2022).

In the case of random mixture hazard models, normalization of random measures is not needed to define the nonparametric prior, since it results from the combination of dependent non-normalized random measures with suitable kernels. The hierarchical structure defining the vector of underlying dependent random measures is

$$\begin{aligned} \tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_D) \mid \tilde{\mu}_0 &\overset{\text{i.i.d.}}{\sim} \text{CRM}(\tilde{\nu}), \\ \tilde{\mu}_0 &\sim \text{CRM}(\nu_0), \end{aligned} \tag{10}$$

where $\tilde{\mu}_1, \ldots, \tilde{\mu}_D$ are conditionally independent CRMs with random Lévy intensity measure

$$\tilde{\nu}(ds, dx) = \rho(ds)\, \tilde{\mu}_0(dx),$$

and $\tilde{\mu}_0$ is a CRM with Lévy intensity measure $\nu_0(ds, dx) = \rho_0(ds)\, P_0(dx)$. This class of mixture hazard rates based on a hierarchical dependence structure of the underlying random measures was introduced in Camerlenghi et al. (2021). An alternative approach to define dependent mixture hazards can be found in Lijoi and Nipoti (2014).

4 Random Partitions and Latent Variables

Consider an array of partially exchangeable sequences with de Finetti measure featuring the hierarchically dependent specification discussed in the previous section. The almost-sure discreteness of (normalized) CRMs naturally induces a random partition structure, with groups including elements both within and across such partially exchangeable sequences. We illustrate this partition structure through the hierarchical prior specification in (9), based on normalized random measures. However, it characterizes every Bayesian nonparametric model built upon hierarchical CRMs priors, being a property of the hierarchical structure itself rather than of the specific model.

Consider the array of D partially exchangeable samples

$$\begin{aligned} \mathbf{X}_d = (X_{d1}, \ldots, X_{dN_d}) \mid \tilde{p}_d &\overset{\text{i.i.d.}}{\sim} \tilde{p}_d, \qquad d = 1, \ldots, D, \\ \tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_D) &\sim \mathcal{Q}_D, \end{aligned} \tag{11}$$

for integers $N_1, \ldots, N_D \geq 1$, where $\mathcal{Q}_D$ is the hierarchical prior in (9). The almost-sure discreteness and dependence of the random probability measures in $\tilde{\boldsymbol{p}}$ imply that tied values occur with positive probability both within and across samples, that is $\mathbb{P}(X_{\ell i} = X_{\kappa j}) > 0$ for any ℓ and κ. Therefore, a random partition of the observations is naturally induced, whereby two elements are in the same partition group if and only if they have the same value. Denote by $X_1^*, \ldots, X_k^*$ the k distinct values assumed in the D partially exchangeable samples, with respective multiplicities $n_1, \ldots, n_k$.

As a consequence, elements belonging to the same partition group may or may not belong to the same sample, which means that there is no structural relationship between the random partition induced by tied values and the natural one determined by the D samples. Let n_{dj} be the number of elements in sample d belonging to group j,

$$n_{dj} = \sum_{i=1}^{N_d} \mathbb{1}_{X_{di}=X_j^*}, \qquad d = 1, \ldots, D, \ j = 1, \ldots, k.$$

However, given the two-fold nature of the partition structure, observed values are not enough to fully characterize its complexity. It is therefore convenient to introduce corresponding sequences of latent variables

$$\mathbf{Z}_d = (Z_{d1}, \ldots, Z_{dN_d}), \qquad d = 1, \ldots, D,$$

taking values on a complete and separable metric space $(\mathbb{T}, \mathcal{T})$, which themselves admit ties with positive probability. These latent variables are formally introduced in the next section for both normalized random measures and random mixture hazards models, and allow to describe a finer partition structure, featuring ties within each sample (but not across samples); this greatly simplifies the learning scheme. In particular, for each sample d and each group j, consider the n_{dj} elements in $\mathbf{X}_d$ for which $X_{di} = X_j^*$ holds. The corresponding elements in $\mathbf{Z}_d$ may themselves show ties: denote by $Z_{dj1}^*, \ldots, Z_{djr_{dj}}^*$ their r_{dj} distinct values, with multiplicities $q_{dj1} + \cdots + q_{djr_{dj}} = n_{dj}$. Notice that, whenever $Z_{di} = Z_{di'}$ then $X_{di} = X_{di'}$, i.e. a tie among values in $\mathbf{Z}_d$ implies a tie among the corresponding elements in $\mathbf{X}_d$, while the converse is not necessarily true. Moreover, let r_j be the partial sum of elements in $(r_{dj})_{dj}$ with respect to d, and let r be their total sum.

An intuitive description of the partition structure introduced in this section is provided by the well-known *Chinese restaurant franchise* metaphor, first presented in Teh et al. (2006) for the hierarchical Dirichlet process. According to this scheme, a franchise consists of D restaurants sharing the same menu, which includes an infinite number of dishes; each restaurant has infinitely many tables and the customers seated at the same table eat the same dish. Customers arriving at each restaurant may choose to either sit at a table with other customers, thus eating the dish already served at that table, or sit at an empty table, either eating a dish already served at other tables in the franchise or eating a new dish from the menu. Notice that, in contrast to the simple Chinese restaurant metaphor, the same dish can be served at different tables within the same restaurant and across different restaurants. Embedding the partition structure described above in the metaphor, element X_{di} represents the dish served in restaurant d to customer i, and the distinct values $X_1^*, \ldots, X_k^*$ represent the k distinct dishes served in the franchise, with n_{dj} being the number of customers eating dish j at restaurant d. Likewise, the latent variable Z_{di} represents the table in restaurant d at which customer i is seated, with r_{dj} being the number of tables in restaurant d at which dish j is served, and q_{djh} being the number of customers in restaurant d eating dish j at table h.

The sequences of latent variables introduced above can be explicitly included in the hierarchical prior specifications by extending the random measures at the lower level of the hierarchies to a larger space (as clarified in the next section). Specifically, given the root of the hierarchy, the conditionally independent (normalized) CRMs appearing in (9) and (10), namely $\tilde{p}_1, \ldots, \tilde{p}_D$ and $\tilde{\mu}_1, \ldots, \tilde{\mu}_D$, can be extended as random measures on $\mathbb{T} \times \mathbb{X}$, and characterized, respectively, by the random Lévy intensity measures

$$\tilde{\nu}(ds, dz, dx) = \rho(ds)\, H(dz)\, \tilde{p}_0(dx), \qquad \tilde{\nu}(ds, dz, dx) = \rho(ds)\, H(dz)\, \tilde{\mu}_0(dx),$$

where H is an arbitrary diffuse probability measure on $(\mathbb{T}, \mathcal{T})$. The random measures on $(\mathbb{X}, \mathcal{X})$ introduced in the original definition are easily recovered via marginalization.

An interesting feature of this analytical device is the extension of the (random) *discrete* components, defined on $\mathbb{X}$, of the Lévy intensity measures characterizing CRMs at the lower level of the hierarchies, to *diffuse* components, defined on $\mathbb{T} \times \mathbb{X}$, as a consequence of H being diffuse. Such property turns out to be fundamental for the recursive application of the result in (14) in presence of hierarchical schemes, as discussed in the following section.

5 The Key Identity for Random Measures

The availability of a framework to describe the random partition structure induced by discrete hierarchical priors is a prerequisite for the determination of marginal and posterior distributions. In this respect, the possibility to marginalize quantities of interest with respect to the prior represents the cornerstone of computations with CRMs, and relies on a specific core structure, which leads to closed-form and tractable expressions.

Let $\tilde{\mu}$ be a homogeneous CRM with Lévy intensity measure $\nu(ds, dx) = \rho(ds)\alpha(dx)$, where α is a finite diffuse measure on $\mathbb{X}$; both its Laplace exponent and cumulants, defined respectively by

$$\psi(u) = \int_{\mathbb{R}^+} \left(1 - e^{-us}\right) \rho(ds), \qquad \tau(m; u) = \int_{\mathbb{R}^+} s^m e^{-us} \rho(ds), \tag{12}$$

play a crucial role in the investigation of the distributional properties of $\tilde{\mu}$. The core quantity representing the building block for every computation with CRMs is

$$\exp\left\{-\int_{\mathbb{X}} f(x)\tilde{\mu}(dx)\right\} \prod_{j=1}^{k} \tilde{\mu}(B_\varepsilon(x_j^*))^{n_j} \tag{13}$$

where we recall that $B_\varepsilon(x_j^*) = \{x \in \mathbb{X} : d(x, x_j^*) < \varepsilon\}$, for $j = 1, \dots, k$, are ε-balls centered at distinct values $x_1^*, \dots, x_k^* \in \mathbb{X}$, with multiplicities $n_1, \dots, n_k \geq 1$ such that $\sum_{j=1}^{k} n_j = n$, and $f : \mathbb{X} \mapsto \mathbb{R}^+$ is any non-negative measurable function. Computing the expectation of the quantity in (13), one can show that

$$\lim_{\varepsilon \to 0} \frac{\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(x)\tilde{\mu}(dx)\right\} \prod_{j=1}^{k} \tilde{\mu}(B_\varepsilon(x_j^*))^{n_j}\right]}{\prod_{j=1}^{k} \alpha(B_\varepsilon(x_j^*))}$$

$$= \exp\left\{-\int_{\mathbb{X}} \psi(f(x))\, \alpha(dx)\right\} \prod_{j=1}^{k} \tau(n_j; f(x_j^*)),$$

which can be informally rewritten as

$$\mathbb{E}\left[\exp\left\{-\int_{\mathbb{X}} f(x)\, \tilde{\mu}(dx)\right\} \prod_{j=1}^{k} \tilde{\mu}(dx_j^*)^{n_j}\right]$$

$$= \exp\left\{-\int_{\mathbb{X}} \psi(f(x))\, \alpha(dx)\right\} \prod_{j=1}^{k} \tau(n_j; f(x_j^*))\, \alpha(dx_j^*). \tag{14}$$

Formally, we can interpret the left-hand side of (14) as the (first) moment measure of an exponentially tilted n-fold product measure of the CRM $\tilde{\mu}$, restricted to a specific subset of $\mathbb{X}^n$. Indeed, $\tilde{\mu}(\cdot)^{n_j}$ may be regarded as the restriction of the n_j-fold product measure of $\tilde{\mu}$ to the diagonal

$\mathit{\Delta}_{n_j} = \{(x_1, \ldots, x_{n_j}) \in \mathbb{X}^{n_j} : x_1 = \cdots = x_{n_j}\}$, through the correspondence

$$\tilde{\mu}(A)^{n_j} = \tilde{\mu}^{n_j}\big(\{(\underbrace{x, \ldots, x}_{n_j}) : x \in A\}\big) = \tilde{\mu}^{n_j}\big((\underbrace{A \times \cdots \times A}_{n_j}) \cap \mathit{\Delta}_{n_j}\big),$$

which is non-zero due to the almost-sure discreteness of $\tilde{\mu}$. Similarly, $\prod_{j=1}^{k} \tilde{\mu}(dx_j^*)^{n_j}$ may be identified with the restriction of the n-fold product measure of $\tilde{\mu}$ to a particular subspace of $\mathbb{X}^n$. Specifically, for any $1 \leq k \leq n$, consider the linear subspaces of $\mathbb{X}^n$ of dimension k for which the n coordinates can be partitioned into k groups with multiplicities $n_1, \ldots, n_k$, where coordinates in the same group take on the same value. The moment measure introduced above, restricted to each of such k-dimensional subspaces, is shown in (14) to be absolutely continuous with respect to the product measure α^k, i.e. the k-fold product of α with itself, with Radon-Nikodym derivative

$$\exp\left\{-\int_{\mathbb{X}} \psi(f(x))\,\alpha(dx)\right\} \prod_{j=1}^{k} \tau(n_j; f(x_j^*)),$$

where $x_1^*, \ldots, x_k^* \in \mathbb{X}$ are the k distinct values assumed by coordinates belonging to the same group. Note that this density only depends on the groups' multiplicities $n_1, \ldots, n_k$ and distinct values $x_1^*, \ldots, x_k^*$, while the coordinates' ordering identifies the specific k-dimensional subspace where the measure in (14) is concentrated. Moreover, such measure can be decomposed into the product of a constant exponential term and k independent measures on $\mathbb{X}$, each absolutely continuous with respect to the diffuse measure α.

This result is particularly suited to hierarchical structures of random measures. Indeed, if $\tilde{\mu}$ is defined through a hierarchical scheme and α is itself a random measure, the same result can be applied recursively, since the structure in (13) reappears for measure α in the right-hand side of (14). Note that the equality above holds true only for diffuse choices of the finite measure α, making it potentially useless if α is a CRM (thus having almost surely discrete realizations). However, this issue is successfully addressed by replacing the random measure $\tilde{\mu}$ with its extended counterpart, as formally described hereunder.

Let $\tilde{\mu}$ be a homogeneous CRM with random Lévy intensity measure

$$\nu(ds, dz, dx) = \rho(ds)\, H(dz)\, \tilde{\mu}_0(dx),$$

where H is a diffuse probability measure on $\mathbb{T}$ and $\tilde{\mu}_0$ is itself a homogeneous CRM with Lévy intensity measure $\nu_0(ds, dx) = \rho_0(ds)\, P_0(dx)$, for P_0 a diffuse probability measure on $\mathbb{X}$. The core quantity of interest, playing the role of (13) for hierarchical schemes, is

$$\exp\left\{-\int_{\mathbb{T}\times\mathbb{X}} f(x)\,\tilde{\mu}(dz, dx)\right\} \prod_{j=1}^{k}\prod_{h=1}^{r_j} \tilde{\mu}(dz_{jh}^*, dx_j^*)^{q_{jh}}, \tag{15}$$

where $x_1^*, \dots, x_k^* \in \mathbb{X}$ are distinct values, $z_{j1}^*, \dots, z_{jr_j}^* \in \mathbb{T}$ are the $r_j \geq 1$ distinct values corresponding to the same x_j^* with multiplicities $q_{j1}, \dots, q_{jr_j} \geq 1$ and such that $\sum_{j=1}^{k} r_j = r$, and $f\colon \mathbb{X} \mapsto \mathbb{R}^+$ is any non-negative measurable function. The expectation of such quantity with respect to the random measure $\tilde{\mu}$ is computed by recursive application of (14), first at the lower level and then at the root of the hierarchy, obtaining

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{-\int_{\mathbb{T}\times\mathbb{X}} f(x)\,\tilde{\mu}(dz, dx)\right\} \prod_{j=1}^{k}\prod_{h=1}^{r_j} \tilde{\mu}(dz_{jh}^*, dx_j^*)^{q_{jh}}\right] \\
&= \exp\left\{-\int_{\mathbb{X}} \psi_0(\psi(f(x)))\, P_0(dx)\right\} \prod_{j=1}^{k} \tau_0(r_j; \psi(f(x_j^*)))\, P_0(dx_j^*) \\
&\times \prod_{j=1}^{k}\prod_{h=1}^{r_j} \tau(q_{jh}; f(x_j^*))\, H(dz_{jh}^*),
\end{aligned} \tag{16}$$

where ψ_0 and τ_0 are defined as in (12) replacing ρ with ρ_0. Similarly to the non-hierarchical case, (16) defines a (first) moment measure of an exponentially tilted n-fold product measure of the random measure $\tilde{\mu}$ with itself, which is here a measure on the space $(\mathbb{T} \times \mathbb{X})^n = \mathbb{T}^n \times \mathbb{X}^n$. Specifically, consider the linear subspaces of $\mathbb{T}^n \times \mathbb{X}^n$ of dimension $r \times k$ for which the $2n$ coordinates can be grouped according to the partition structure encoded into elements $(q_{jh})_{jh}$. This moment measure, restricted to each of such subspaces, is shown in (16) to be absolutely continuous with respect to the product measure $H^r \times P_0^k$, with Radon-Nikodym derivative

$$\exp\left\{-\int_{\mathbb{X}} \psi_0(\psi(f(x)))\, P_0(dx)\right\} \prod_{j=1}^{k} \tau_0(r_j; \psi(f(x_j^*))) \left(\prod_{h=1}^{r_j} \tau(q_{jh}; f(x_j^*))\right),$$

where $x_1^*, \ldots, x_k^* \in \mathbb{X}$ are the k distinct values assumed by $\mathbb{X}$-valued coordinates belonging to the same group. Again, this measure can be decomposed into the product of a constant exponential term, the r-fold product of H with itself, and k independent measures on $\mathbb{X}$, each absolutely continuous with respect to the diffuse probability measure P_0.

The rest of this section is devoted to the analysis of the likelihood functions associated to normalized random measures and random mixture hazards, assuming partially exchangeable models with a hierarchical prior specification. In particular, the introduction of latent variables specific to each model, together with suitable analytical manipulations, recovers the core structure introduced in (15), which in turn allows one to compute marginal distributions explicitly via the recursive application of (14).

Normalized random measures Consider the array of D partially exchangeable samples

$$\begin{aligned} \mathbf{X}_d = (X_{d1}, \ldots, X_{dN_d}) \mid \tilde{p}_d &\overset{\text{i.i.d.}}{\sim} \tilde{p}_d, \qquad d = 1, \ldots, D, \\ \tilde{\boldsymbol{p}} = (\tilde{p}_1, \ldots, \tilde{p}_D) \mid \tilde{p}_0 &\overset{\text{i.i.d.}}{\sim} \ \text{NCRM}(\tilde{\nu}), \\ \tilde{p}_0 &\sim \ \text{NCRM}(\nu_0), \end{aligned} \tag{17}$$

for integers $N_1, \ldots, N_D \geq 1$ and hierarchical prior (9). Introducing the corresponding latent variables, which represent the tables in the restaurant franchise metaphor, the likelihood function associated to the augmented sample $(\mathbf{X}_d, \mathbf{Z}_d)$ is

$$\mathcal{L}(\tilde{\mu}_d; \mathbf{X}_d, \mathbf{Z}_d) = \prod_{i=1}^{N_d} \tilde{p}_d(dZ_{di}, dX_{di}) = \tilde{\mu}_d(\mathbb{T}, \mathbb{X})^{-N_d} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tilde{\mu}_d(dZ_{djh}^*, dX_j^*)^{q_{djh}}.$$

By using a simple analytical manipulation based on the density of a gamma random variable, this can be rewritten as

$$\begin{aligned} \mathcal{L}(\tilde{\mu}_d; \mathbf{X}_d, \mathbf{Z}_d) = \frac{1}{\Gamma(N_d)} \int_{\mathbb{R}^+} u_d^{N_d - 1} \exp\left\{ - \int_{\mathbb{T}\times\mathbb{X}} u_d \, \tilde{\mu}_d(dz, dx) \right\} du_d \\ \times \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tilde{\mu}_d(dZ_{djh}^*, dX_j^*)^{q_{djh}}, \end{aligned}$$

where u_d is an additional latent variable, thanks to which the core structure in (15) is successfully recovered in the likelihood. Therefore, the result in

(14) can be applied to marginalize the expression with respect to the lower level of the hierarchical prior, i.e. conditionally on $\tilde{p}_0$, obtaining

$$\mathbb{P}(\mathbf{X}_d, \mathbf{Z}_d \,|\, \tilde{p}_0) = \frac{1}{\Gamma(N_d)} \int_{\mathbb{R}+} u_d^{N_d - 1} \, e^{-\psi(u_d)} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tau(q_{djh}; u_d) \, du_d \\ \times \prod_{j=1}^{k} \tilde{p}_0(dX_j^*)^{r_{dj}} \prod_{h=1}^{r_{dj}} H(dZ_{djh}^*).$$

Note that only the random partition induced by ties in the sequence $\mathbf{Z}_d$, which is encoded into groups multiplicities $(q_{djh})_{jh}$, is relevant in the expression above, while their specific values are sampled independently from the measure H. Since these values are not even observed in the model, they can be safely ignored, and $\mathbf{Z}_d^\pi$ can be written instead of $\mathbf{Z}_d$ to indicate that results depend on the partition induced by latent variables, rather than on their values.

The same analytical manipulation can be performed recursively for the random probability measure $\tilde{p}_0$, considering the likelihood associated to the D partially exchangeable samples. The joint marginal distribution of the observations $\mathbf{X}$ and the latent variables $\mathbf{Z}$ can be effectively expressed in terms of the random partitions they induce, respectively denoted by $\mathbf{X}^\pi$ and $\mathbf{Z}^\pi$, and of the vectors of their distinct values, denoted by $\mathbf{X}^*$ and $\mathbf{Z}^*$, as

$$\mathbb{P}(\mathbf{X}^\pi, \mathbf{X}^*, \mathbf{Z}^\pi, \mathbf{Z}^*) = \\ = \prod_{d=1}^{D} \left(\frac{1}{\Gamma(N_d)} \int_{\mathbb{R}+} u_d^{N_d - 1} \, e^{-\psi(u_d)} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tau(q_{djh}; u_d) \, du_d \cdot \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} H(dZ_{djh}^*) \right) \\ \times \frac{1}{\Gamma(r)} \int_{\mathbb{R}+} u_0^{r-1} \, e^{-\psi_0(u_0)} \prod_{j=1}^{k} \tau_0(r_j; u_0) \, du_0 \cdot \prod_{j=1}^{k} P_0(dX_j^*). \tag{18}$$

In analogy with the comment above, the distinct values in $\mathbf{X}^*$ do not enter the partition function as well, and are sampled independently from the base probability measure P_0. Therefore, marginalizing out the contribution of the distinct values $(\mathbf{X}^*, \mathbf{Z}^*)$ in (18), one obtains the partially exchangeable partition probability function (pEPPF):

$$\mathbb{P}(\mathbf{X}^\pi, \mathbf{Z}^\pi) = \prod_{d=1}^{D} \frac{1}{\Gamma(N_d)} \int_{\mathbb{R}+} u_d^{N_d - 1} \, e^{-\psi(u_d)} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tau(q_{djh}; u_d) \, du_d$$

$$\times \frac{1}{\Gamma(r)} \int_{\mathbb{R}^+} u_0^{r-1}\, e^{-\psi_0(u_0)} \prod_{j=1}^{k} \tau_0(r_j; u_0)\, du_0. \quad (19)$$

This expression clearly highlights the way random partitions are composed at the two levels of the hierarchy. At the level of single samples (i.e. restaurant level), the partition function depends on the number of customers seated at each table, that is on the partition induced by each sequence $\mathbf{Z}_d$. On the other hand, at the root level (i.e. franchise level), the partition function depends on the number of tables eating each dish, which describes how the finer partition induced by latent variables $\mathbf{Z}$ is related to the coarser partition induced by observations $\mathbf{X}$.

According to the interpretation discussed in the previous section, the joint marginal distribution in (18) represents a moment measure on the joint space $\mathbb{T}^n \times \mathbb{X}^n$, restricted to the linear subspace of dimension $r \times k$ which reflects the partition structure induced by observations and latent variables and encoded into elements $(q_{djh})_{djh}$. Specifically, this restricted measure is absolutely continuous with respect to the product measure $H^r \times P_0^k$, and has constant Radon-Nikodym derivative expressed by the pEPPF in (19).

The structure of the pEPPF represents the cornerstone of computational developments: indeed, full conditional distributions derived from it can be exploited to devise marginal Gibbs sampling schemes, as extensively discussed in Camerlenghi et al. (2019b) (Section 6.1).

Random mixture hazards Consider the array of D partially exchangeable samples

$$\begin{aligned} \mathbf{T}_d = (T_{d1}, \ldots, T_{dN_d}) \mid \tilde{p}_d &\overset{\text{i.i.d.}}{\sim} \tilde{p}_d, \qquad d = 1, \ldots, D, \\ \tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \ldots, \tilde{\mu}_D) \mid \tilde{\mu}_0 &\overset{\text{i.i.d.}}{\sim} \text{CRM}(\tilde{\nu}), \\ \tilde{\mu}_0 &\sim \text{CRM}(\nu_0), \end{aligned} \quad (20)$$

for integers $N_1, \ldots, N_D \geq 1$ and hierarchical prior (10). The random probability measure $\tilde{p}_d$ is recovered from the expression of the random hazard rate in (6) and is expressed in terms of the random measures $\tilde{\mu}_d$ as

$$\tilde{p}_d(dt) = \int_{\mathbb{X}} k(t;x)\, \tilde{\mu}_d(dx)\, \exp\left\{ -\int_0^t \int_{\mathbb{X}} k(s;x)\, \tilde{\mu}_d(dx)\, ds \right\} dt.$$

The further level of complexity represented by the kernel mixture requires the introduction of additional latent variables, namely the ones representing the latent samples from the random measures, i.e. the dishes in the restaurant franchise metaphor, which instead are directly observed in the previously discussed case of normalized random measures. The augmented random probability measure takes the more tractable form

$$\tilde{p}_d(dt, dx) = k(t; x)\tilde{\mu}_d(dx) \exp\left\{-\int_0^t \int_{\mathbb{X}} k(s; y)\, \tilde{\mu}_d(dy)\, ds\right\} dt.$$

Introducing the latent variables representing the tables in the restaurant franchise metaphor, the likelihood function associated to the augmented sample $(\mathbf{T}_d, \mathbf{X}_d, \mathbf{Z}_d)$ is

$$\begin{aligned}
&\mathcal{L}(\tilde{\mu}_d;\, \mathbf{T}_d, \mathbf{X}_d, \mathbf{Z}_d) = \\
&\quad = \prod_{i=1}^{N_d} k(T_{di}; X_{di})\, \tilde{\mu}_d(dZ_{di}, dX_{di}) \exp\left\{-\int_0^{T_{di}} \int_{\mathcal{T}\times\mathbb{X}} k(s; x)\, \tilde{\mu}_d(dz, dx)\, ds\right\} dT_{di} \\
&\quad = Q(\mathbf{T}_d, \mathbf{X}_d) \exp\left\{-\int_{\mathcal{T}\times\mathbb{X}} K_d(x)\, \tilde{\mu}_d(dz, dx)\right\} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tilde{\mu}_d(dZ^*_{djh}, dX^*_j)^{q_{djh}},
\end{aligned}$$

where, in order to ease the notation, the following quantities have been defined:

$$Q(\mathbf{T}_d, \mathbf{X}_d) = \prod_{i=1}^{N_d} k(T_{di}; X_{di})\, dT_{di}, \qquad K_d(x) = \sum_{i=1}^{N_d} \int_0^{T_{di}} k(s; x)\, ds.$$

Again, the structure in (15) is recovered in the likelihood, with the constant term u_d for normalized random measures replaced by the function K_d. Therefore, the result in (14) can be applied recursively at the lower and root levels of the hierarchical prior, so that the joint marginal distribution of both the observations $\mathbf{T}$ and the latent variables $\mathbf{X}$ and $\mathbf{Z}$ becomes

$$\begin{aligned}
\mathbb{P}(\mathbf{T}, \mathbf{X}^\pi, \mathbf{X}^*, \mathbf{Z}^\pi, \mathbf{Z}^*) = Q(\mathbf{T}, \mathbf{X}) \exp&\left\{-\int_{\mathbb{X}} \psi_0\left(\sum_{d=1}^{D} \psi(K_d(x))\right) P_0(dx)\right\} \\
&\times \prod_{d=1}^{D} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tau(q_{djh}; K_d(X^*_j))\, H(dZ^*_{djh}) \\
&\times \prod_{j=1}^{k} \tau_0\left(r_j; \sum_{d=1}^{D} \psi(K_d(X^*_j))\right) P_0(dX^*_j). \qquad (21)
\end{aligned}$$

The similarities with the joint distribution derived in (18) are apparent, as the composition of random partitions at the two levels of the hierarchy follows the same structure. An important difference is represented by the dependence of the partition function from the specific distinct values $\mathbf{X}^*$ through the functions $K_1, \dots, K_D$, so that their contribution cannot be marginalized out, and a proper pEPPF cannot be defined. This dependence on the specific values is reflected by a non-constant Radon-Nikodym derivative, when (21) is regarded as a moment measure.

The lack of a proper pEPPF also represents a computational drawback in this context, as the Gibbs resampling step for the latent distinct values $\mathbf{X}^*$ involves both the kernel term $Q(\mathbf{T}, \mathbf{X})$ and the value of the partition function. On the other hand, analytical tractability greatly benefits from the absence of the integrals with respect to $u_1, \dots, u_D$ and u_0 appearing in the expression (19) for normalized random measures, which are merely a byproduct of analytical manipulations and need to be treated as additional latent variables.

6 Posterior Characterizations

Another essential result which leverages on the random partition structure is the posterior characterization of the random measures $\tilde{\mu}_1, \dots, \tilde{\mu}_D$ and $\tilde{\mu}_0$. Specifically, posterior distributions of CRMs are recovered via the determination of their conditional Laplace functional transforms: in these expressions, one identifies the distributions of jumps at fixed locations and the Lévy intensities of CRMs without fixed jump points. A structural conjugacy property is shown to hold, that is, a posteriori, the vector of random measures $\tilde{\boldsymbol{\mu}}$ retains its hierarchical form, with random measures at the lower level of the hierarchy being conditionally independent given the random measure at the root.

In the following, posterior updates of hierarchical CRMs priors are explicitly described for the partially exchangeable models based on both normalized random measures and random mixture hazards. For convenience, the intensities of the jump components ρ and ρ_0 at both levels of the hierarchy are assumed to be absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^+$, and suitably written as $\rho(ds) = \rho(s)\,ds$ and $\rho_0(ds) = \rho_0(s)\,ds$, respectively. Moreover, the results in this section consider the original definitions of random measures at the lower level of the hierarchies, as introduced in (9) and (10): the adaptation to their extended versions is straightforward.

Normalized random measures Consider the partially exchangeable model described in (17). The posterior distributions of the non-normalized random measures $\tilde{\mu}_1, \dots, \tilde{\mu}_D$ and $\tilde{\mu}_0$ are characterized conditionally on the observations $\mathbf{X}$, the latent variables $\mathbf{Z}$, and the additional latent variables $U_1, \dots, U_D$ and U_0. As already mentioned, such latent variables are a byproduct of analytical manipulations in the likelihood, and are needed to recover a (conditional) structural conjugacy.

Let $U_1, \dots, U_D$ and U_0 be conditionally independent positive random variables with density functions

$$f_d(u \,|\, \mathbf{X}_d^\pi, \mathbf{Z}_d^\pi) \propto u^{N_d - 1}\, e^{-\psi(u)} \prod_{j=1}^{k} \prod_{h=1}^{r_{dj}} \tau(q_{djh}; u), \qquad d = 1, \dots, D,$$

$$f_0(u \,|\, \mathbf{X}^\pi, \mathbf{Z}^\pi) \propto u^{r-1}\, e^{-\psi_0(u)} \prod_{j=1}^{k} \tau_0(r_j; u),$$

where we recall that $\mathbf{X}^\pi$ and $\mathbf{Z}^\pi$ denote the random partitions induced by ties in the sequences $\mathbf{X}$ and $\mathbf{Z}$, respectively, which are encoded into groups multiplicities r_j's (coarser partition) and q_{djh}'s (finer partition). At the lower level of the hierarchy, the posterior distribution of each random measure $\tilde{\mu}_d$, given the observations $\mathbf{X}_d$, the latent variables $\mathbf{Z}_d$ and U_d, and the root random measure $\tilde{\mu}_0$ is

$$\tilde{\mu}_d(dx) \mid \mathbf{X}_d^\pi, \mathbf{X}_d^*, \mathbf{Z}_d^\pi, U_d, \tilde{\mu}_0 \;\sim\; \tilde{\mu}_d^*(dx) + \sum_{j=1}^{k} \sum_{h=1}^{r_{dj}} J_{djh}\, \delta_{X_j^*}(dx),$$

where the random elements in the sum are independent, $\tilde{\mu}_d^* \sim \mathrm{CRM}(\nu_d^*)$ with homogeneous Lévy intensity measure

$$\nu_d^*(ds, dx) = e^{-U_d\, s}\, \nu(ds, dx) = e^{-U_d\, s}\, \rho(ds)\, \tilde{p}_0(dx),$$

and each J_{djh} is a non-negative random variable with density function

$$f_{djh}(s) \propto s^{q_{djh}}\, e^{-U_d\, s}\, \rho(s), \qquad j = 1, \dots, k, \quad h = 1, \dots, r_{dj}.$$

Therefore, a posteriori and conditionally on $\tilde{\mu}_0$, each random measure $\tilde{\mu}_d$ is still a CRM, resulting from the sum of random jumps at fixed points of discontinuity and a CRM without fixed points of discontinuity. The latter is characterized by the Lévy intensity measure of the prior with an exponential updating term, while the fixed points of discontinuity correspond to the distinct values of the observations. Moreover, the random measures $\tilde{\mu}_1, \ldots, \tilde{\mu}_d$ preserve their conditional independence, given $\tilde{\mu}_0$.

Similarly, the posterior distribution of the random measure $\tilde{\mu}_0$ at the root of the hierarchy, given the observations $\mathbf{X}$ and the latent variables $\mathbf{Z}$ and U_0, is

$$\tilde{\mu}_0(dx) \mid \mathbf{X}^\pi, \mathbf{X}^*, \mathbf{Z}^\pi, U_0 \;\sim\; \tilde{\mu}_0^*(dx) + \sum_{j=1}^{k} I_j \, \delta_{X_j^*}(dx),$$

where the random elements in the sum are independent, $\tilde{\mu}_0^* \sim \mathrm{CRM}(\nu_0^*)$ with homogeneous Lévy intensity measure

$$\nu_0^*(ds, dx) = e^{-U_0\, s}\, \nu_0(ds, dx) = e^{-U_0\, s}\, \rho_0(ds)\, P_0(dx),$$

and each I_j is a non-negative random variable with density function

$$f_j(s) \propto s^{r_j}\, e^{-U_0\, s}\, \rho_0(s), \qquad j = 1, \ldots, k.$$

Again, the random measure $\tilde{\mu}_0$ is still a CRM a posteriori, given by the sum of random jumps at fixed points of discontinuity, corresponding to the distinct observed values, and an exponentially updated CRM without fixed points of discontinuity.

An interesting feature of this result is that the prior-posterior updating mechanism preserves the homogeneity of the random measures. Indeed, the density functions of additional latent variables $U_1, \ldots, U_D$ and U_0, the exponential update of the Lévy intensity and the distributions of random jumps at the fixed points of discontinuity depend only on the partition structure induced by the observations and the latent variables, encoded into $\mathbf{X}^\pi$ and $\mathbf{Z}^\pi$, while observed distinct values $\mathbf{X}^*$ only determine the fixed locations of discontinuity points. This property clearly parallels the factorization property of the marginal distribution into pEPPF and independent sampling of distinct values, and it represents a fundamental computational advantage when one needs to sample directly from the posterior distribution of hierarchical CRMs.

Random mixture hazards Consider the partially exchangeable model described in (20). In this case, the posterior distributions of random measures $\tilde{\mu}_1, \ldots, \tilde{\mu}_D$ and $\tilde{\mu}_0$ are characterized conditionally on the observations $\mathbf{T}$ and the latent variables $\mathbf{X}$ and $\mathbf{Z}$, featuring a proper structural conjugacy. Specifically, the posterior distribution of each random measure $\tilde{\mu}_d$, given the observations $\mathbf{T}_d$, the latent variables $\mathbf{X}_d$ and $\mathbf{Z}_d$, and the root measure $\tilde{\mu}_0$ is

$$\tilde{\mu}_d(dx) \mid \mathbf{T}_d, \mathbf{X}^{\pi}_d, \mathbf{X}^{*}_d, \mathbf{Z}^{\pi}_d, \tilde{\mu}_0 \;\sim\; \tilde{\mu}^{*}_d(dx) + \sum_{j=1}^{k} \sum_{h=1}^{r_{dj}} J_{djh}\, \delta_{X^*_j}(dx),$$

where the random elements in the sum are mutually independent, $\tilde{\mu}^*_d \sim \mathrm{CRM}(\nu^*_d)$ with non-homogeneous Lévy intensity measure

$$\nu^*_d(ds, dx) = e^{-K_d(x)\, s}\, \rho(ds)\, \tilde{\mu}_0(dx),$$

and each J_{djh} is a non-negative random variable with density function

$$f_{djh}(s) \propto s^{q_{djh}}\, e^{-K_d(X^*_j)\, s}\, \rho(s), \qquad j = 1, \dots, k, \quad h = 1, \dots, r_{dj}.$$

Similarly, the posterior distribution of the random measure $\tilde{\mu}_0$ at the root of the hierarchy, given the observations $\mathbf{T}$ and the latent variables $\mathbf{X}$ and $\mathbf{Z}$, is

$$\tilde{\mu}_0(dx) \mid \mathbf{T}, \mathbf{X}^{\pi}, \mathbf{X}^{*}, \mathbf{Z}^{\pi} \;\sim\; \tilde{\mu}^*_0(dx) + \sum_{j=1}^{k} I_j\, \delta_{X^*_j}(dx),$$

where the random elements in the sum are mutually independent, $\tilde{\mu}^*_0 \sim \mathrm{CRM}(\nu^*_0)$ with non-homogeneous Lévy intensity measure

$$\nu^*_0(ds, dx) = \exp\left\{ -\sum_{d=1}^{D} \psi(K_d(x))\, s \right\} \rho_0(ds)\, P_0(dx),$$

and each I_j is a non-negative random variable with density function

$$f_j(s) \propto s^{r_j}\, \exp\left\{ -\sum_{d=1}^{D} \psi(K_d(X^*_j))\, s \right\} \rho_0(s), \qquad j = 1, \dots, k.$$

As already highlighted for marginal distributions, the structural analogies with the posterior characterization for normalized random measures are apparent, and similar considerations apply. In particular, here the role of the random variables $U_1, \dots, U_D$ is played by the non-random functions $K_1, \dots, K_D$, which summarize the contribution of the observations $\mathbf{T}$ to the posterior update.

However, the analytical and computational advantage represented by the absence of additional latent variables is partially overturned by the non-homogeneity of the Lévy intensity measures characterizing, a posteriori, the continuous part of the hierarchical CRMs. The challenges represented by non-homogeneous CRMs in conditional sampling algorithms are discussed in Camerlenghi et al. (2021) (Section 6.2), where a novel general-purpose approach is proposed.

7 Generalized Gamma CRMs as Natural Conjugate Priors

The practical implementation of Bayesian procedures involving hierarchical CRMs priors requires the specification of their Lévy intensity measures. In particular, a fundamental role is played by measures ρ and ρ_0, which characterize the jump components and deeply affect the induced partition structure. Indeed, such measures directly impact the distributions of random jumps at fixed points of discontinuity in the posterior characterizations of hierarchical CRMs, and also enter the definition of key quantities in (12), which constitute the core structure of the marginal distributions (18) and (21). The availability of closed-form and tractable expressions represents a computational advantage for both marginal and conditional algorithms. On the contrary, the specification of P_0 has a far lower impact from both analytical and computational points of view.

A natural choice of ρ and ρ_0 for hierarchical constructions is represented by the *generalized gamma* hierarchical CRM, corresponding to the specifications

$$\rho(ds) = \frac{1}{\Gamma(1-\sigma)}\, s^{-\sigma-1}\, e^{-\beta s}\, ds, \qquad \rho_0(ds) = \frac{1}{\Gamma(1-\sigma_0)}\, s^{-\sigma_0-1}\, e^{-\beta_0 s}\, ds,$$

with parameters $\beta, \beta_0 \in \mathbb{R}^+$ and $\sigma, \sigma_0 \in [0, 1)$. Notable special cases are obtained setting $\sigma = \sigma_0 = 0$, which corresponds to the *gamma* hierarchical CRM, and $\beta = \beta_0 = 0$, characterizing the *σ–stable* hierarchical CRM.

The generalized gamma hierarchical CRM allows for the explicit computation of the integrals defining the Laplace exponent and its cumulants in (12), namely,

$$\psi(u) = \int_{\mathbb{R}^+} (1 - e^{-us})\, \rho(ds) = \frac{(\beta+u)^\sigma - \beta^\sigma}{\sigma} \overset{\sigma=0}{=} \log\left(1 + \frac{u}{\beta}\right),$$

$$\tau(m; u) = \int_{\mathbb{R}^+} s^m\, e^{-us}\, \rho(ds) = \frac{\Gamma(m-\sigma)}{\Gamma(1-\sigma)}\, (\beta+u)^{-m+\sigma}.$$

These quantities can be directly substituted into the expressions of the marginal distributions, from which full conditional distributions and predictive urn schemes are easily derived.

Moreover, the generalized gamma choice acts as the (conditionally) conjugate prior with respect to the posterior characterization of hierarchical CRMs for the partially exchangeable models discussed in this work. For example, considering the model in (17) based on normalized random measures, the posterior distribution of each random measure $\tilde{\mu}_d$ at the lower level of the hierarchy consists of the sum of random jumps at fixed points of discontinuity having gamma distribution, namely,

$$J_{djh} \sim \mathrm{Gamma}(q_{djh} - \sigma, \beta + U_d),$$

and a CRM without fixed points of discontinuity, which still has the Lévy intensity measure of a generalized gamma CRM, with the exponential term resulting in the parameter update $\beta \mapsto \beta + U_d$. The same structure is observed for the root measure $\tilde{\mu}_0$, and for the model in (20) based on random mixture hazards, with the usual roles swap of the variables $U_1, \dots, U_D$ and the functions $K_1, \dots, K_D$, converting real parameters β and β_0 into functional parameters and making the Lévy intensity non-homogeneous.

8 Eliciting the Induced Dependence Structure

In the previous sections, we have described two different hierarchical models that provide an intuitive and effective way to introduce dependence among the components of a vector of random measures. The amount of dependence regulates the borrowing of information across groups, that is, how much inference and prediction for each group are influenced by the observations in other groups. In an ideal setting where infinite observations for each group are available, one would not need to leverage on the information contained in the other groups of observations, and the borrowing of information would be useless (if not harmful). However, in real situations with only few observations per group available or strongly unbalanced datasets, the borrowing of information can lead to crucial improvements in the estimates and meaningful reduction of their uncertainty.

Let $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_D)$ be a vector of random measures. Two extreme situations can be identified: (i) when the random measures are equal almost surely, i.e., $\tilde{\mu}_1 = \cdots = \tilde{\mu}_D$ a.s., there is maximal dependence and, since all observations are treated as belonging to the same group, full borrowing of information; (ii) when the random measures are mutually independent, there is no borrowing of information, since the inference for each group is not affected by the observations in other groups. In the elicitation of the prior, it is crucial to understand how much dependence is introduced in the model, as this has major consequences on the learning mechanism. In this respect, it is sufficient to remark that if (i) is assumed a priori, the estimates for each group are exactly the same, without taking into account possible differences across the groups, while if (ii) is imposed a priori, the estimate for each group does not take into account the observations in other groups, with potential loss of information.

One of the most natural summaries of the dependence structure between two random measures $\tilde{\mu}_i$ and $\tilde{\mu}_j$ is their pairwise covariance structure $\mathrm{Cov}(\tilde{\mu}_i(A), \tilde{\mu}_j(A))$, for any set $A \in \mathcal{X}$, and its normalized version, the pairwise correlation $\mathrm{Corr}(\tilde{\mu}_i(A), \tilde{\mu}_j(A))$. See Catalano et al. (2023, 2021) for alternative approaches based on the Wasserstein distance. Note that, when the pairwise correlation is 1, it is sufficient to consider random measures with the same marginal distribution to prove that the random variables $\tilde{\mu}_1(A), \ldots, \tilde{\mu}_D(A)$ are equal almost surely. Moreover, even if a pairwise correlation equal to 0 does not imply, in general, that the random variables $\tilde{\mu}_1(A), \ldots, \tilde{\mu}_D(A)$ are mutually independent, it actually does for many specific models. Both these features hold for the hierarchical structures we have considered in this work, namely,

$$\tilde{\boldsymbol{\mu}}^{(1)} = \left(\tilde{\mu}_1^{(1)}, \ldots, \tilde{\mu}_D^{(1)}\right) \mid \tilde{\mu}_0 \overset{\text{i.i.d.}}{\sim} \mathrm{CRM}\left(\rho(ds)\, \frac{\tilde{\mu}_0(dx)}{\tilde{\mu}_0(\mathbb{X})}\right), \tag{22}$$

$$\tilde{\boldsymbol{\mu}}^{(2)} = \left(\tilde{\mu}_1^{(2)}, \ldots, \tilde{\mu}_D^{(2)}\right) \mid \tilde{\mu}_0 \overset{\text{i.i.d.}}{\sim} \mathrm{CRM}(\rho(ds)\, \tilde{\mu}_0(dx)), \tag{23}$$

where $\tilde{\mu}_0$ is a CRM with Lévy intensity measure $\nu_0(ds, dx) = \rho_0(ds)\, P_0(dx)$. In the previous sections, we have applied normalization to the random measures in (22) and a hazard mixture transformation to the random measures in (23). In this section, we rather focus on the comparison between the dependence structures at the level of the random measures, irrespective of the particular choice of transformation. Indeed, while it would be possible to focus on the transformed random measures as well, we believe that a direct comparison between (22) and (23) can help to disentangle the effect of the hierarchical construction from the effects of the transformations.

A decisive advantage of the pairwise covariance is represented by its plain evaluation for hierarchical models through the law of total covariance and Campbell's theorem. In particular, considering the vector of random measures $\tilde{\boldsymbol{\mu}}^{(1)} = \left(\tilde{\mu}_1^{(1)}, \dots, \tilde{\mu}_D^{(1)}\right)$ in (22), for any fixed set $A \in \mathcal{X}$ and $i \neq j$,

$$
\begin{aligned}
\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A)\right) &= \left(\int s\rho(ds)\right)\mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right),\\
\operatorname{Var}\left(\tilde{\mu}_i^{(1)}(A)\right) &= \operatorname{Cov}\left(\tilde{\mu}_i^{(1)}(A), \tilde{\mu}_j^{(1)}(A)\right) + \left(\int s^2\rho(ds)\right)\mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right),\\
\operatorname{Cov}\left(\tilde{\mu}_i^{(1)}(A), \tilde{\mu}_j^{(1)}(A)\right) &= \left(\int s\rho(ds)\right)^2 \operatorname{Var}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right).
\end{aligned}
\tag{24}
$$

The corresponding expressions for the vector $\tilde{\boldsymbol{\mu}}^{(2)} = \left(\tilde{\mu}_1^{(2)}, \dots, \tilde{\mu}_D^{(2)}\right)$ in (23) are obtained by replacing the normalized random measure $\tilde{\mu}_0(A)/\tilde{\mu}_0(\mathbb{X})$ with its non-normalized counterpart $\tilde{\mu}_0(A)$. At the root level of the hierarchy, the mean and the variance of both the CRM $\tilde{\mu}_0$ and its normalization $\tilde{\mu}_0(\cdot)/\tilde{\mu}_0(\mathbb{X})$ may be expressed in terms of their Lévy intensity measure ν_0, exploiting Campbell's theorem and the techniques developed in James et al. (2006). For illustration purposes, consider the hierarchical gamma process, where $\tilde{\mu}_1^{(h)}, \dots, \tilde{\mu}_D^{(h)}$, for $h = 1, 2$, and $\tilde{\mu}_0$ are gamma completely random measures, obtained by choosing

$$
\rho(ds) = \frac{\alpha\, e^{-s}}{s}\, ds, \qquad \rho_0(ds) = \frac{\alpha_0\, e^{-s}}{s}\, ds.
$$

This specification yields the following expressions for the quantities considered in (24), for both hierarchical structures in (22) (left column) and (23) (right column):

$$
\begin{aligned}
\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A)\right) &= \alpha\, P_0(A), & \mathbb{E}\left(\tilde{\mu}_i^{(2)}(A)\right) &= \alpha\alpha_0 P_0(A),\\
\operatorname{Cov}\left(\tilde{\mu}_i^{(1)}(A), \tilde{\mu}_j^{(1)}(A)\right) &= \frac{\alpha^2\, P_0(A)(1 - P_0(A))}{1+\alpha_0}, & \operatorname{Cov}\left(\tilde{\mu}_i^{(2)}(A), \tilde{\mu}_j^{(2)}(A)\right) &= \alpha^2\alpha_0 P_0(A),\\
\operatorname{Var}\left(\tilde{\mu}_i^{(1)}(A)\right) &= \frac{\alpha^2\, P_0(A)(1-P_0(A))}{1+\alpha_0} + \alpha P_0(A), & \operatorname{Var}\left(\tilde{\mu}_i^{(2)}(A)\right) &= \alpha(\alpha+1)\,\alpha_0 P_0(A),\\
\operatorname{Corr}\left(\tilde{\mu}_i^{(1)}(A), \tilde{\mu}_j^{(1)}(A)\right) &= \frac{\alpha(1-P_0(A))}{\alpha(1-P_0(A))+1+\alpha_0}, & \operatorname{Corr}\left(\tilde{\mu}_i^{(2)}(A), \tilde{\mu}_j^{(2)}(A)\right) &= \frac{\alpha}{1+\alpha}.
\end{aligned}
$$

In order to correctly interpret the information contained in (24), let us elaborate on two different flexibility properties that are desirable for the dependence structure of a model that induces positive association between the random measures. The first kind of flexibility ensures that, for every value $\gamma \in [0,1]$, there exists a specification of the model parameters such that the random measures have correlation equal to (or converging to) γ. This property holds for hierarchical models in general, and can be easily checked for the hierarchical gamma process considered above: in both cases, the values of α and, possibly, α_0 can be chosen so that the correlations are equal to (or converge to) every fixed value $\gamma \in [0,1]$. The second, and stronger, kind of flexibility asks that, for every marginal law of the random measures and for every value $\gamma \in [0,1]$, there exists a specification of the model parameters such that the random measures have correlation equal to (or converging to) γ. This kind of flexibility ensures that the marginal law of the random measures can be modeled separately from their dependence structure, a feature which is often desirable in practice, as they encode different aspects of the model. For simplicity, it is usually sufficient to restrict to a weaker version, whereby one fixes only the first and second moments of the random measures, instead of fixing their whole marginal laws.

Interestingly, most hierarchical models currently used in the literature do not achieve this second type of flexibility. For example, consider the vector of random measures $\tilde{\boldsymbol{\mu}}^{(2)}$ with the hierarchical gamma specification described above. As revealed by the expression of the correlation, in order to recover perfectly correlated random measures, one needs $\alpha \to +\infty$; however, in such case, the expected value diverges. This suggests that a good practice for hierarchical gamma random measures, when the root measure is not normalized, is to fix $\alpha_0 = 1/\alpha$, so that $\mathbb{E}\big(\tilde{\mu}_i^{(2)}(A)\big) = P_0(A)$ and thus the dependence structure does not affect the mean of the random measure. Nevertheless, with such choice of parameters, one obtains $\mathrm{Var}\big(\tilde{\mu}_i^{(2)}(A)\big) = (\alpha+1)\,P_0(A)$, which in turn implies that the only way to recover perfectly correlated random measures is to have (marginally) infinite variance. In conclusion, the flexibility of second kind cannot be achieved for the hierarchical structure in (23). On the other hand, such issues do not arise for the vector $\tilde{\boldsymbol{\mu}}^{(1)}$ described above, and the ultimate reason of this can be understood by looking at the expressions in (24): if $\tilde{\mu}_0$ is a gamma random measure, its mean and variance coincide, whereas the variance of the normalization $\tilde{\mu}_0(\cdot)/\tilde{\mu}_0(\mathbb{X})$ can be adjusted separately from its expected value. This fact suggests to consider other classes of random measure for $\tilde{\mu}_0$, where a hyper-parameter can be set to flexibly account for different values of the variance.

Summing up, when resorting to hierarchical constructions to model the dependence between random measures, particular attention has to be put in eliciting the dependence structure, as it will also affect the marginal distributions. For the same reason, the covariance is not a reliable measure of dependence: since changing the covariance also affects the variance, the normalization required in the expression of the correlation is not only a way to obtain values in $[0,1]$, but also provides important information about the dependence structure. In order to effectively showcase this last reasoning,

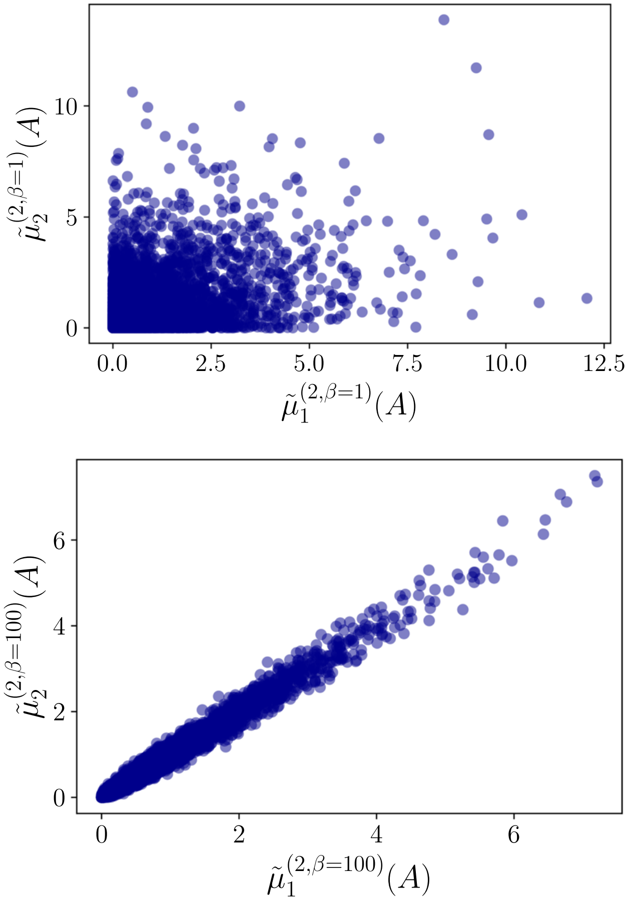

Figure 1: Samples from $\tilde{\boldsymbol{\mu}}^{(2,\beta)}(A) = \big(\tilde{\mu}_1^{(2,\beta)}(A),\, \tilde{\mu}_2^{(2,\beta)}(A)\big)$, for $\beta = 1$ (top) and $\beta = 100$ (bottom), where A is such that $P_0(A) = 0.5$. The covariance is the same for every value of $\beta > 0$, that is, $\mathrm{Cov}\big(\tilde{\mu}_1^{(2,\beta)}(A),\, \tilde{\mu}_2^{(2,\beta)}(A)\big) = 0.5$

consider the bivariate vector of hierarchical random measures

$$\begin{aligned}\tilde{\boldsymbol{\mu}}^{(2,\beta)} = \left(\tilde{\mu}_1^{(2,\beta)}, \tilde{\mu}_2^{(2,\beta)}\right) \mid \tilde{\mu}_0 &\overset{\text{i.i.d.}}{\sim} \text{CRM}\left(\frac{\beta\, e^{-\beta s}}{s}\, ds\; \tilde{\mu}_0(dx)\right), \\ \tilde{\mu}_0 &\sim \text{CRM}\left(\frac{e^{-s}}{s}\, ds\, P_0(dx)\right),\end{aligned}$$

where $\beta > 0$. Resorting to the expressions in (24), one can show that, for any $A \in \mathcal{X}$, the covariance $\text{Cov}\big(\tilde{\mu}_1^{(2,\beta)}(A), \tilde{\mu}_2^{(2,\beta)}(A)\big) = P_0(A)$ remains unchanged for every value of β; however, the dependence structure of $\tilde{\boldsymbol{\mu}}^{(2,\beta)}$ appears substantially different for different values of β, as shown in Figure 1, in the cases $\beta = 1$ and $\beta = 100$. This difference is correctly detected by the correlation, which equals $\beta/(1+\beta)$ and thus converges to 1 as β diverges.

Acknowledgements. The authors are grateful to an anonymous Referee for insightful comments and suggestions, which greatly improved the paper. Part of this work was carried out while M. Catalano was affiliated with the Department of Statistics of the University of Warwick.

Funding Open access funding provided by Università Commerciale Luigi Bocconi within the CRUI-CARE Agreement. M. Catalano was partially supported by a Research Development Fund from the University of Warwick. C. Del Sole, A. Lijoi and I. Prünster are partially supported by MIUR, PRIN Project 2022CLTYP4.

Compliance with ethical standards

Conflicts of Interest All authors declare that they have no conflicts of interest.

References

Barrios, E., A. Lijoi, L. E. Nieto-Barajas, and I. Prünster (2013). Modeling with normalized random measure mixture models. *Statistical Science 28*(3), 313–334.

Basu, D. and R. C. Tiwari (1982). A note on the Dirichlet process. In *Statistics and probability: essays in honor of C. R. Rao*, pp. 89–103. North-Holland, Amsterdam-New York.

Camerlenghi, F., D. B. Dunson, A. Lijoi, I. Prünster, and A. Rodriguez (2019a). Latent nested nonparametric priors. *Bayesian Analysis 14*(4), 1303–1356.

Camerlenghi, F., A. Lijoi, P. Orbanz, and I. Prünster (2019b). Distribution theory for hierarchical processes. *The Annals of Statistics 47*(1), 67–92.

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

Camerlenghi, F., A. Lijoi, and I. Prünster (2018). Bayesian nonparametric inference beyond the Gibbs-type framework. *Scand. J. Stat. 45*(4), 1062–1091.

Camerlenghi, F., A. Lijoi, and I. Prünster (2021). Survival analysis via hierarchically dependent mixture hazards. *The Annals of Statistics 49*, 863 – 884.

Catalano, M., P. De Blasi, A. Lijoi, and I. Prünster (2022). Posterior asymptotics for boosted hierarchical Dirichlet process mixtures. *Journal of Machine Learning Research 23*(80), 1–23.

Catalano, M., H. Lavenant, A. Lijoi, and I. Prünster (2023). A Wasserstein index of dependence for random measures. *Journal of the American Statistical Association*, forthcoming.

Catalano, M., A. Lijoi, and I. Prünster (2020). Approximation of Bayesian models for time-to-event data. *Electron. J. Stat. 14*(2), 3366–3395.

Catalano, M., A. Lijoi, and I. Prünster (2021). Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. *The Annals of Statistics 49*(5), 2916–2947.

Cifarelli, D. M. and E. Regazzini (1978). Nonparametric statistical problems under partial exchangeability: The role of associative means. *Quaderni Istituto Matematica Finanziaria dell'Università di Torino Serie III 12*, 1–36.

Daley, D. and D. Vere-Jones (2007). *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*. Probability and Its Applications. Springer New York.

De Blasi, P., G. Peccati, and I. Prünster (2009). Asymptotics for posterior hazards. *Ann. Statist. 37*(4), 1906–1945.

de Finetti, B. (1937). La prévision, ses lois logiques, ses sources subjectives. *Annales de l'Institute Henri Poincaré 7*, 1–68.

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

De Iorio, M., P. Müller, G. L. Rosner, and S. N. MacEachern (2004). An ANOVA model for dependent random measures. *Journal of the American Statistical Association 99*(465), 205–215.

Doksum, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *The Annals of Probability 2*(2), 183 – 201.

Donnet, S., V. Rivoirard, J. Rousseau, and C. Scricciolo (2017). Posterior concentration rates for counting processes with Aalen multiplicative intensities. *Bayesian Anal. 12*(1), 53–87.

Dunson, D. B. and J.-H. Park (2008). Kernel stick-breaking processes. *Biometrika 95*(2), 307–323.

Dykstra, R. L. and P. Laud (1981). A Bayesian nonparametric approach to reliability. *The Annals of Statistics 9*(2), 356–367.

Epifani, I. and A. Lijoi (2010). Nonparametric priors for vectors of survival functions. *Statistica Sinica 20*(4), 1455–1484.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics 1*, 209 – 230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics 2*, 615 – 629.

Griffin, J. E. and F. Leisen (2017). Compound random measures and their use in Bayesian non-parametrics. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 79*(2), 525–545.

Hjort, N. L. (1990). Nonparametric Bayes estimators based on Beta processes in models for life history data. *The Annals of Statistics 18*(3), 1259 – 1294.

Ishwaran, H. and L. F. James (2004). Computational methods for multiplicative intensity models using weighted gamma processes: proportional hazards, marked point processes, and panel count data. *J. Amer. Statist. Assoc. 99*(465), 175–190.

James, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *The Annals of Statistics 33*(4), 1771–1799.

James, L. F., A. Lijoi, and I. Prünster (2006). Conjugacy as a distinctive feature of the Dirichlet process. *Scandinavian Journal of Statistics 33*(1), 105–120.

James, L. F., A. Lijoi, and I. Prünster (2009). Posterior analysis for normalized random measures with independent increments. *Scandinavian Journal of Statistics 36*(1), 76–97.

Kingman, J. (1993). *Poisson Processes.* Oxford Studies in Probability. Clarendon Press.

Kingman, J. F. C. (1967). Completely random measures. *Pacific Journal of Mathematics 21*(1), 59–78.

Kingman, J. F. C. (1975). Random discrete distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 37*(1), 1–22.

Lau, J. W. and E. Cripps (2022). Thinned completely random measures with applications in competing risks models. *Bernoulli 28*(1), 638 – 662.

Lijoi, A., R. H. Mena, and I. Prünster (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the American Statistical Association 100*(472), 1278–1291.

Lijoi, A., R. H. Mena, and I. Prünster (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(4), 715–740.

Lijoi, A. and B. Nipoti (2014). A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc. 109*(506), 802–814.

Lijoi, A., B. Nipoti, and I. Prünster (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli 20*(3), 1260–1291.

Lijoi, A., I. Prünster, and G. Rebaudo (2023). Flexible clustering via hidden hierarchical Dirichlet priors. *Scandinavian Journal of Statistics 50*(1), 213–234.

Lijoi, A. and I. Prünster (2010). Models beyond the Dirichlet process. In N. L. Hjort, C. C. Holmes, P. Müller, and S. G. Walker (Eds.), *Bayesian Nonparametrics*, pp. 80–136. Cambridge University Press.

Lo, A. and C.-S. Weng (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Annals of the Institute of Statistical Mathematics 41*(2), 227–245.

Lo, A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics 12*(1), 351 – 357.

MacEachern, S. N. (1999). Dependent nonparametric processes. *in ASA Proceedings of the Section on Bayesian Statistical Science.*, Alexandria, VA: American Statistical Association.

MacEachern, S. N. (2000). Dependent Dirichlet processes. *Technical Report,*, The Ohio State University.

Müller, P., F. Quintana, and G. Rosner (2004). A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*(3), 735–749.

Nieto-Barajas, L. E. and S. G. Walker (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statist. Sinica 14*(4), 1127–1146.

Peccati, G. and I. Prünster (2008). Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Ann. Appl. Probab. 18*(5), 1910–1943.

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

Regazzini, E., A. Lijoi, and I. Prünster (2003). Distributional results for means of normalized random measures with independent increments. *The Annals of Statistics 31*, 560–585.

Riva-Palacio, A. and F. Leisen (2021). Compound vectors of subordinators and their associated positive Lévy copulas. *Journal of Multivariate Analysis 183*, 104728.

Rodríguez, A. and D. B. Dunson (2011). Nonparametric Bayesian models through probit stick-breaking processes. *Bayesian Anal. 6*(1), 145–177.

Rodriguez, A., D. B. Dunson, and A. E. Gelfand (2008). The nested Dirichlet process. *Journal of the American Statistical Association 103*(483), 1131–1154.

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

Teh, Y. W., M. I. Jordan, M. J. Beal, and D. M. Blei (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association 101*(476), 1566–1581.

Walker, S. and P. Muliere (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist. 25*(4), 1762–1780.

Marta Catalano
Department of Economics and
Finance, Luiss University, Rome,
Italy

Claudio Del Sole, Antonio Lijoi and
Igor Prünster
Bocconi Institute for Data Science
and Analytics, Bocconi University,
Milan, Italy
E-mail: igor@unibocconi.it

Paper received: 4 July 2023