A Note on the Dependence Structure of Hierarchical Completely Random Measures

Marta Catalano and Claudio Del Sole

Abstract Hierarchical models offer a principled framework to make inference and predictions on different (groups of) observations by leveraging their common features. In a nonparametric setting, the borrowing of information is controlled by the dependence structure induced on a vector of random measures. Two different hierarchical specifications are now well-established in the literature: we compare their dependence structures, provide some intuition on how to enhance their flexibility, and highlight a possibly misleading behaviour of their pairwise covariance. This note is based on some recent results in [3].

Keywords Completely random measures · Dependence · Hierarchical models · Partial exchangeability · Partial pooling

1 Introduction

Hierarchical random measures represent a key ingredient to nonparametric multi-level models. This class of models loosens the exchangeability assumption with two conceptual steps. First, the distribution of each observation, or each group of observations, is modelled through a distinct random measure $\tilde{\mu}_i$, by exploiting well-established transformations, such as normalization [21], kernel mixtures [8, 18, 19], or exponential transformations [7]; we refer to [17] for a general overview. Then, the vector of random measures $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ is modelled as a dependent vector of random measures through conditional independence and identity in distribution.

The dependence between prior random measures guarantees that the predictive and posterior distributions will borrow information across groups. This induces a

M. Catalano (✉)
University of Warwick, Coventry, UK
e-mail: marta.catalano@warwick.ac.uk

C. Del Sole
Bocconi University, Milan, Italy
e-mail: claudio.delsole@phd.unibocconi.it

A. Avalos-Pacheco et al. (eds.), *Bayesian Statistics, New Generations New Approaches*, Springer Proceedings in Mathematics & Statistics 435,
https://doi.org/10.1007/978-3-031-42413-7_8

shrinkage effect that often makes the estimates more reliable when only one or a few observations per group are observed, and experimentally disappears as the number of observations per group diverges. The positive effects of *partial pooling* are well-established in a parametric setting (see, e.g., [10]), and are now dominant in the nonparametric literature as well (see [20] for a recent review). Though the link between parametric and nonparametric models is often neglected, we underline that they both fit into the framework of *partial exchangeability*, in which the distribution of the observations is invariant with respect to overall permutations that do not move observations to different groups.

Broadly speaking, the specification of the law of a vector of random measures $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$ amounts to the specification of the marginal distributions $\mathcal{L}(\tilde{\mu}_i)$, which determine how data in each group will be used to estimate the corresponding distribution and make predictions, and their dependence structure, which determines how much influence data in other groups will have in addressing the same tasks. These two aspects are fundamentally different, and the practitioner should be able to elicit them independently. A very common choice for modelling the marginal distributions $\mathcal{L}(\tilde{\mu}_i)$ is to consider completely random measures (CRMs), a flexible class of random measures that is characterized by the independence of the evaluations on disjoint sets. As for the dependence structure, there have been several proposals in the literature (e.g., [9, 11, 15, 16]). Still, hierarchical forms of dependence are arguably the most natural ones for a Bayesian statistician: as observations are usually modelled as conditionally independent, it is conceptually straightforward to introduce dependence among the random measures through conditional independence as well. Two alternative strategies to model hierarchical completely random measures have recently emerged in the literature (see [1, 22] and [2, 12, 23], respectively). Their main difference consists in whether the dependence is introduced through a common random probability measure or through an (unnormalized) random measure, as precisely depicted in (1) and (2) in Sect. 2. In this note, we compare these two classes of models and provide some intuition on how to enhance the flexibility of their dependence structure. This is based on some recent results in [3].

2 Hierarchical Completely Random Measures

Let $\mathbb{X}$ be a Polish space and denote by $\mathbb{M}$ the space of boundedly finite measures on $\mathbb{X}$, equipped with the corresponding Borel σ-algebra, that is, the smallest σ-algebra that makes the projections $A \mapsto \mu(A)$ measurable for every measure $\mu \in \mathbb{M}$ and every bounded set A; see [6] for details. A random measure $\tilde{\mu}$ is a random element taking values in $\mathbb{M}$; we focus here on completely random measures, a very natural and convenient class of discrete random measures, first introduced in [14].

Definition 1 A completely random measure (CRM) $\tilde{\mu}$ is a random measure on $\mathbb{X}$ 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.

In the following, we consider CRMs without fixed points of discontinuity and without a deterministic drift; their law is characterized by the Laplace transform, for any $\lambda > 0$ and any Borel set A,

$$\mathbb{E}\left[e^{-\lambda\tilde{\mu}(A)}\right] = \exp\left\{-\int_{(0,+\infty)\times A}(1-e^{-\lambda s})\,\nu(ds,dx)\right\},$$

where ν is the Lévy intensity measure of the CRM, which uniquely identifies $\tilde{\mu}$ and can be any Borel measure on $(0, +\infty) \times \mathbb{X}$ that satisfies the condition

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

This characterization motivates the notation $\tilde{\mu} \sim \mathrm{CRM}(\nu)$, which denotes a CRM with Lévy intensity ν. In order to simplify the exposition, we only consider homogeneous CRMs, for which the Lévy measure ν can be disintegrated as $\nu = \rho \otimes \alpha$, where α is a σ-finite measure on $\mathbb{X}$, ρ is a Lévy measure on $(0, +\infty)$, and $\otimes$ denotes the product measure. We refer to α as the atom component and to ρ as the jump component of the Lévy intensity, though they are uniquely defined only up to multiplication by a constant.

Example 1 A random measure $\tilde{\mu} \sim \mathrm{CRM}(\nu)$ is a (homogeneous) gamma CRM with total mass $a > 0$ and base probability measure P_0 if

$$\nu(ds,dx) = a\,\frac{e^{-s}}{s}\,ds\,P_0(dx).$$

Building upon the definition of CRMs, hierarchical structures represent a natural way to construct vectors of dependent random measures $(\tilde{\mu}_1, \ldots, \tilde{\mu}_d)$. We focus on two different proposals that are now well-established in the literature:

$$\tilde{\mu}_1^{(1)}, \ldots, \tilde{\mu}_d^{(1)} \mid \tilde{\mu}_0 \sim \mathrm{CRM}\left(\rho \otimes \frac{\tilde{\mu}_0}{\tilde{\mu}_0(\mathbb{X})}\right), \qquad \tilde{\mu}_0 \sim \mathrm{CRM}(\nu_0); \tag{1}$$

$$\tilde{\mu}_1^{(2)}, \ldots, \tilde{\mu}_d^{(2)} \mid \tilde{\mu}_0 \sim \mathrm{CRM}\left(\rho \otimes \tilde{\mu}_0\right), \qquad \tilde{\mu}_0 \sim \mathrm{CRM}(\nu_0), \tag{2}$$

where $\nu_0 = \rho_0 \otimes \alpha_0$ is the Lévy intensity of $\tilde{\mu}_0$. The first construction (1) has been mainly exploited to model dependent random probability measures: it was first discussed in [22] for the special case of hierarchical Dirichlet processes, which can be recovered from gamma CRMs, and extensively studied in [1] for the more general class of CRMs. The second construction (2) has been adopted to model dependent mixture hazard rates [2] and dependent feature models [12, 23].

Note that (1) defines a law for the vector of dependent random measures building upon the normalized completely random measure $\tilde{\mu}_0/\tilde{\mu}_0(\mathbb{X})$; as shown in [21], this construction requires the atom component α_0 of ν_0 to be a finite measure, and its jump component ρ_0 to have infinite mass around the origin, i.e. $\rho_0(\mathbb{R}^+) = +\infty$. In particular, (1) allows to make inference on CRMs whose atom component is a.s. finite, whereas (2) can be adapted to make inference on a.s. infinite measures as well.

3 Dependence Structure

Hierarchical constructions introduce dependence between random measures in a very interpretable and natural fashion. The amount of dependence between the random measures introduced a priori regulates the borrowing of information across different populations a posteriori and, as such, it should be carefully elicited. In an ideal world where infinite observations for each group are available, one would not need to leverage on the information contained in other groups of observations, and the borrowing of information would be useless (if not harmful). However, it is often the case that only few observations per group are observed, or that the datasets are strongly unbalanced: in these scenarios, borrowing information from similar populations can be a valuable tool to make inference and predictions more accurate.

Let $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \dots, \tilde{\mu}_d)$ be a vector of random measures. One can depict two extreme situations: (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. This highlights how the amount of prior dependence has a major impact in the learning mechanism and should be carefully elicited. To this end, one needs a way to quantify the dependence between random measures.

One of the most natural summaries of the dependence structure between two random measures is their pairwise covariance structure $\mathrm{Cov}(\tilde{\mu}_i(A), \tilde{\mu}_j(A))$, and its normalized version, the pairwise correlation $\mathrm{Corr}(\tilde{\mu}_i(A), \tilde{\mu}_j(A))$. There have been proposals to go beyond pairwise comparisons by using the Wasserstein distance on the joint distribution of the vector of random measures [4, 5]. However, a decisive advantage of the pairwise covariance is that it can be easily evaluated for hierarchical models through the law of total covariance and Campbell's theorem.

Proposition 1 *Let* $\tilde{\boldsymbol{\mu}}^{(1)} = \big(\tilde{\mu}_1^{(1)}, \dots, \tilde{\mu}_d^{(1)}\big)$ *as defined in* (1)*. Then, for* $i \neq j$,

$$\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A)\right) = \left(\int s \,\mathrm{d}\rho(s)\right) \mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right);$$

$$\mathrm{Var}\left(\tilde{\mu}_i^{(1)}(A)\right) = \mathrm{Cov}\left(\tilde{\mu}_i^{(1)}(A),\ \tilde{\mu}_j^{(1)}(A)\right) + \left(\int s^2 \,\mathrm{d}\rho(s)\right) \mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right);$$

$$\mathrm{Cov}\left(\tilde{\mu}_i^{(1)}(A),\ \tilde{\mu}_j^{(1)}(A)\right) = \left(\int s \,\mathrm{d}\rho(s)\right)^2 \mathrm{Var}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right).$$

Proposition 2 *Let $\tilde{\boldsymbol{\mu}}^{(2)} = \big(\tilde{\mu}_1^{(2)}, \dots, \tilde{\mu}_d^{(2)}\big)$ as defined in (2). Then, for $i \neq j$,*

$$\begin{aligned}
\mathbb{E}\left(\tilde{\mu}_i^{(2)}(A)\right) &= \left(\int s\,\mathrm{d}\rho(s)\right)\mathbb{E}(\tilde{\mu}_0(A));\\
\mathrm{Var}\left(\tilde{\mu}_i^{(2)}(A)\right) &= \mathrm{Cov}\left(\tilde{\mu}_i^{(2)}(A),\ \tilde{\mu}_j^{(2)}(A)\right) + \left(\int s^2\,\mathrm{d}\rho(s)\right)\mathbb{E}(\tilde{\mu}_0(A));\\
\mathrm{Cov}\left(\tilde{\mu}_i^{(2)}(A),\ \tilde{\mu}_j^{(2)}(A)\right) &= \left(\int s\,\mathrm{d}\rho(s)\right)^2 \mathrm{Var}(\tilde{\mu}_0(A)).
\end{aligned}$$

The mean and the variance of both the CRM $\tilde{\mu}_0$ and its normalization $\tilde{\mu}_0/\tilde{\mu}_0(\mathbb{X})$ may be easily expressed in terms of the Lévy measure ν_0 by using Campbell's theorem and the techniques in [13], as in the next example.

Example 2 In the hierarchical gamma process, both $\tilde{\mu}_0$ and $\tilde{\mu}_i \mid \tilde{\mu}_0$ are gamma completely random measures, as introduced in Example 1. This implies that $\int s\,\mathrm{d}\rho(s) = \int s^2\,\mathrm{d}\rho(s) = a$ and $\mathbb{E}(\tilde{\mu}_0(A)) = \mathrm{Var}(\tilde{\mu}_0(A)) = a_0 P_0(A)$, while $\mathbb{E}(\tilde{\mu}_0(A)/\tilde{\mu}_0(\mathbb{X})) = P_0(A)$ and $\mathrm{Var}(\tilde{\mu}_0(A)/\tilde{\mu}_0(\mathbb{X})) = P_0(A)(1 - P_0(A))/(1 + a_0)$. In particular, for every Borel set A such that $0 < P_0(A) < 1$,

$$\begin{gathered}
\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A)\right) = a P_0(A); \qquad \mathrm{Var}\left(\tilde{\mu}_i^{(1)}(A)\right) = \frac{a^2 P_0(A)(1 - P_0(A))}{1 + a_0} + a P_0(A);\\
\mathrm{Cov}\left(\tilde{\mu}_i^{(1)}(A),\ \tilde{\mu}_j^{(1)}(A)\right) = \frac{a^2 P_0(A)(1 - P_0(A))}{1 + a_0};\\
\mathrm{Corr}\left(\tilde{\mu}_i^{(1)}(A),\ \tilde{\mu}_j^{(1)}(A)\right) = \frac{a(1 - P_0(A))}{a(1 - P_0(A)) + 1 + a_0};\\
\mathbb{E}\left(\tilde{\mu}_i^{(2)}(A)\right) = a a_0 P_0(A); \qquad \mathrm{Var}\left(\tilde{\mu}_i^{(2)}(A)\right) = a(a+1) a_0 P_0(A);\\
\mathrm{Cov}\left(\tilde{\mu}_i^{(2)}(A),\ \tilde{\mu}_j^{(2)}(A)\right) = a^2 a_0 P_0(A); \qquad \mathrm{Corr}\left(\tilde{\mu}_i^{(2)}(A),\ \tilde{\mu}_j^{(2)}(A)\right) = \frac{a}{1 + a}.
\end{gathered}$$

Propositions 1 and 2 contain important information about the dependence structure of hierarchical models. We elaborate on these results by first highlighting two desirable flexibility properties characterizing random measures with positive association. 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. By looking at the expressions in Example 2, one can easily check this property to be true for both hierarchical gamma processes; more generally, this flexibility property holds for both hierarchical models we have introduced. The second kind of flexibility is stronger and 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 flexibility property ensures that the marginal laws of the random measures may be elicited separately from their dependence structure. As we underlined in the introduction, this is a desirable feature since the marginal laws and

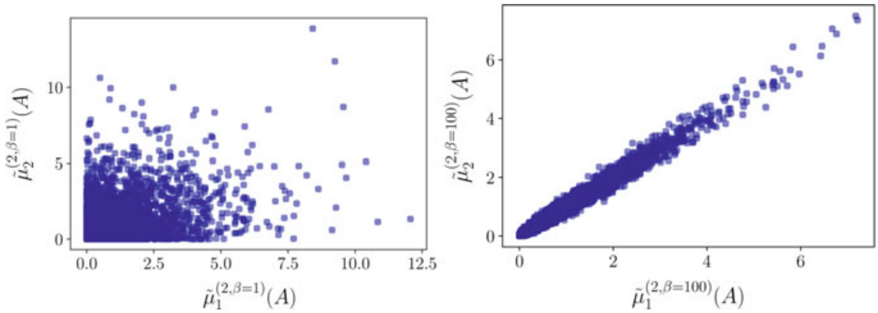

Fig. 1 Samples from $\tilde{\boldsymbol{\mu}}^{(2,\beta)}(A) = \big(\tilde{\mu}_1^{(2,\beta)}(A),\ \tilde{\mu}_2^{(2,\beta)}(A)\big)$, where $\beta = 1$ (left), $\beta = 100$ (right), and A is a set with $P_0(A) = 1/2$. The covariance is the same, $\mathrm{Cov}\big(\tilde{\mu}_1^{(2,\beta)}(A)),\ \tilde{\mu}_2^{(2,\beta)}(A)\big) = 1/2$

the dependence encode very different aspects of the model: the former shapes the distribution of a group of observations, the latter regulates the impact of the other groups. For simplicity, one can restrict to a slightly weaker form of this second kind of flexibility, where one only considers every possible value of the mean and the variance of the random measures, instead of every marginal law.

Interestingly, most hierarchical models currently used in the literature do not achieve the second type of flexibility. As an example, consider the gamma hierarchical random measures $\tilde{\boldsymbol{\mu}}^{(2)}$ defined in Example 2. As highlighted by the expression of the correlation, one can have perfect dependence, i.e. $\tilde{\mu}_1^{(2)} = \tilde{\mu}_2^{(2)}$ a.s., only if $a \to +\infty$; however, in such case, the expected value of the marginals diverges. This suggests that a good practice for hierarchical gamma random measures of type (2) is to fix $a_0 = 1/a$, 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. However, with such choice of parameters $\mathrm{Var}\big(\tilde{\mu}_i^{(2)}(A)\big) = (a+1)P_0(A)$, which in turn implies that the only way to recover perfectly correlated random measures is to have infinite variance. In short, we are not able to achieve the flexibility of second kind.

It is worth underlying that such issues do not appear for the hierarchical gamma measure $\tilde{\boldsymbol{\mu}}^{(1)}$ in Example 2, as clarified by comparing the expressions for the covariances in Propositions 1 and 2. Indeed, if $\tilde{\mu}_0$ is a gamma CRM as in Proposition 2, its mean and variance coincide, whereas the variance of the normalization $\tilde{\mu}_0/\tilde{\mu}_0(\mathbb{X})$ can be adjusted separately from its expected value, which is the situation of Proposition 1. This suggests a possible way to adjust the hierarchical models of type (2) in order to ensure the flexibility of second kind as well: we should consider other classes of random measures for $\tilde{\mu}_0$, where a suitable hyperparameter can be set to flexibly account for different values of the variance.

4 Discussion

In this note, we have discussed how hierarchical constructions represent a natural and intuitive way to model the dependence between random measures; however, its elicitation can be subtle, since adjusting the dependence often affects the marginal distributions as well. For the same reason, the covariance is not a reliable measure of dependence: given that changing the covariance also affects the variance, the normalization required to compute the correlation is not only a way to obtain values in [0, 1], but also provides important information about the dependence structure. This is showcased by the following simple example. For $\beta > 0$, consider the bivariate vectors of hierarchical random measures

$$\tilde{\mu}_1^{(2,\beta)},\ \tilde{\mu}_2^{(2,\beta)} \mid \tilde{\mu}_0 \sim \text{CRM}\left(\beta\,\frac{e^{-\beta s}}{s}\,\mathrm{d}s \otimes \tilde{\mu}_0\right), \qquad \tilde{\mu}_0 \sim \text{CRM}\left(\frac{e^{-s}}{s}\,\mathrm{d}s \otimes P_0\right).$$

One can easily prove that the covariance $\text{Cov}\big(\tilde{\mu}_1^{(2,\beta)}(A)),\ \tilde{\mu}_2^{(2,\beta)}(A)\big) = P_0(A)$ stays the same for every value of β. On the other hand, the dependence structure of $\tilde{\boldsymbol{\mu}}^{(2,\beta)}$ appears to be very different for different values of β, as depicted in Fig. 1 for $\beta = 1$ and $\beta = 100$. This feature is correctly detected by the correlation, which is equal to $\text{Corr}\big(\tilde{\mu}_1^{(2,\beta)}(A)),\ \tilde{\mu}_2^{(2,\beta)}(A)\big) = \beta/(1+\beta)$ and thus goes to 1 as $\beta \to +\infty$.

5 Proofs

Proof (Propositions 1 and 2) Thanks to the tower property,

$$\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A)\right) = \mathbb{E}\left(\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A) \mid \tilde{\mu}_0\right)\right) = \left(\int s\,\mathrm{d}\rho(s)\right)\mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right),$$
$$\mathbb{E}\left(\tilde{\mu}_i^{(2)}(A)\right) = \mathbb{E}\left(\mathbb{E}\left(\tilde{\mu}_i^{(2)}(A) \mid \tilde{\mu}_0\right)\right) = \left(\int s\,\mathrm{d}\rho(s)\right)\mathbb{E}(\tilde{\mu}_0(A)).$$

The law of total covariance yields, for $k = 1, 2$ and $i \neq j$,

$$\begin{aligned}
&\text{Cov}\left(\tilde{\mu}_i^{(k)}(A),\ \tilde{\mu}_j^{(k)}(A)\right) = \\
&= \mathbb{E}\left(\text{Cov}\left(\tilde{\mu}_i^{(k)}(A),\ \tilde{\mu}_j^{(k)}(A)\right) \mid \tilde{\mu}_0\right) + \text{Cov}\left(\mathbb{E}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right),\ \mathbb{E}\left(\tilde{\mu}_j^{(k)}(A) \mid \tilde{\mu}_0\right)\right) \\
&= \text{Cov}\left(\mathbb{E}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right),\ \mathbb{E}\left(\tilde{\mu}_j^{(k)}(A) \mid \tilde{\mu}_0\right)\right) = \text{Var}\left(\mathbb{E}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right)\right),
\end{aligned}$$

where

$$\operatorname{Var}\left(\mathbb{E}\left(\tilde{\mu}_i^{(1)}(A) \mid \tilde{\mu}_0\right)\right)=\operatorname{Var}\left(\int s\,\mathrm{d}\rho(s)\,\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right)=\left(\int s\,\mathrm{d}\rho(s)\right)^2\operatorname{Var}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right),$$

$$\operatorname{Var}\left(\mathbb{E}\left(\tilde{\mu}_i^{(2)}(A) \mid \tilde{\mu}_0\right)\right)=\operatorname{Var}\left(\int s\,\mathrm{d}\rho(s)\,\tilde{\mu}_0(A)\right)=\left(\int s\,\mathrm{d}\rho(s)\right)^2\operatorname{Var}\left(\tilde{\mu}_0(A)\right).$$

Similarly, one can also compute the variances: for $k = 1, 2$,

$$\operatorname{Var}\left(\tilde{\mu}_i^{(k)}(A)\right)=\mathbb{E}\left(\operatorname{Var}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right)\right)+\operatorname{Var}\left(\mathbb{E}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right)\right),$$

where the expressions of $\operatorname{Var}\left(\mathbb{E}\left(\tilde{\mu}_i^{(k)}(A) \mid \tilde{\mu}_0\right)\right)$ can be found above, and

$$\mathbb{E}\left(\operatorname{Var}\left(\tilde{\mu}_i^{(1)}(A) \mid \tilde{\mu}_0\right)\right)=\mathbb{E}\left(\int s^2\,\mathrm{d}\rho(s)\,\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right)=\left(\int s^2\,\mathrm{d}\rho(s)\right)\mathbb{E}\left(\frac{\tilde{\mu}_0(A)}{\tilde{\mu}_0(\mathbb{X})}\right),$$

$$\mathbb{E}\left(\operatorname{Var}\left(\tilde{\mu}_i^{(2)}(A) \mid \tilde{\mu}_0\right)\right)=\mathbb{E}\left(\int s^2\,\mathrm{d}\rho(s)\,\tilde{\mu}_0(A)\right)=\left(\int s^2\,\mathrm{d}\rho(s)\right)\mathbb{E}\left(\tilde{\mu}_0(A)\right).$$

References

1. Camerlenghi, F., Lijoi, A., Orbanz, P., Prünster, I.: Distribution theory for hierarchical processes. Ann. Stat. **47**(1), 67–92 (2019)
2. Camerlenghi, F., Lijoi, A., Prünster, I.: Survival analysis via hierarchically dependent mixture hazards. Ann. Stat. **49**, 863–884 (2021)
3. Catalano, M., Del Sole, C., Lijoi, A., Prünster, I.: A unified approach to hierarchical random measures. Submitted (2023)
4. Catalano, M., Lavenant, H., Lijoi, A., Prünster, I.: A Wasserstein index of dependence for random measures. ArXiv 2109.06646 (2022)
5. Catalano, M., Lijoi, A., Prünster, I.: Measuring dependence in the Wasserstein distance for Bayesian nonparametric models. Ann. Stat. **49**(5), 2916–2947 (2021)
6. Daley, D., Vere-Jones, D.: An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure. Probability and Its Applications. Springer, New York (2007)
7. Doksum, K.: Tailfree and neutral random probabilities and their posterior distributions. Ann. Probab. **2**(2), 183–201 (1974)
8. Dykstra, R.L., Laud, P.: A Bayesian nonparametric approach to reliability. Ann. Stat. **9**(2), 356–367 (1981)
9. Epifani, I., Lijoi, A.: Nonparametric priors for vectors of survival functions. Statistica Sinica **20**(4), 1455–1484 (2010)
10. Gelman, A., Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models. Analytical Methods for Social Research. Cambridge University Press (2006)
11. Griffin, J.E., Leisen, F.: Compound random measures and their use in Bayesian non-parametrics. J. R. Stat. Soc. Ser. B (Stat. Methodol.) **79**(2), 525–545 (2017)
12. James, L.F., Lee, J., Pandey, A.: Bayesian analysis of generalized hierarchical Indian buffet processes for within and across group sharing of latent features. ArXiv 2304.05244 (2023)
13. James, L.F., Lijoi, A., Prünster, I.: Conjugacy as a distinctive feature of the Dirichlet process. Scand. J. Stat. **33**(1), 105–120 (2006)
14. Kingman, J.F.C.: Completely random measures. Pac. J. Math. **21**(1), 59–78 (1967)
15. Lau, J.W., Cripps, E.: Thinned completely random measures with applications in competing risks models. Bernoulli **28**(1), 638–662 (2022)

16. Lijoi, A., Nipoti, B.: A class of hazard rate mixtures for combining survival data from different experiments. J. Am. Statist. Assoc. **109**(506), 802–814 (2014)
17. Lijoi, A., Prünster, I.: Models beyond the Dirichlet process. In: Hjort, N.L., Holmes, C.C., Müller, P., Walker, S.G. (eds.) Bayesian Nonparametrics, pp. 80–136. Cambridge University Press (2010)
18. Lo, A., Weng, C.S.: On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. Ann. Inst. Stat. Math. **41**(2), 227–245 (1989)
19. Lo, A.Y.: On a class of Bayesian nonparametric estimates: I. density estimates. Ann. Stat. **12**(1), 351–357 (1984)
20. Quintana, F.A., Müller, P., Jara, A., MacEachern, S.N.: The dependent Dirichlet process and related models. Stat. Sci. **37**, 24–41 (2022)
21. Regazzini, E., Lijoi, A., Prünster, I.: Distributional results for means of normalized random measures with independent increments. Ann. Stat. **31**, 560–585 (2003)
22. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical Dirichlet processes. J. Am. Stat. Assoc. **101**(476), 1566–1581 (2006)
23. Thibaux, R., Jordan, M.I.: Hierarchical beta processes and the Indian buffet process. In: Meila, M., Shen, X. (eds.) Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 2, pp. 564–571. PMLR, San Juan, Puerto Rico (2007)