The Annals of Statistics
2021, Vol. 49, No. 5, 2916–2947
https://doi.org/10.1214/21-AOS2065

MEASURING DEPENDENCE IN THE WASSERSTEIN DISTANCE FOR BAYESIAN NONPARAMETRIC MODELS

BY MARTA CATALANO[1], ANTONIO LIJOI[2,*] AND IGOR PRÜNSTER[2,†]

[1]*Department of Statistics, University of Warwick, marta.catalano@warwick.ac.uk*
[2]*Department of Decision Sciences, Bocconi University, *antonio.lijoi@unibocconi.it; †igor.pruenster@unibocconi.it*

The proposal and study of dependent Bayesian nonparametric models has been one of the most active research lines in the last two decades, with random vectors of measures representing a natural and popular tool to define them. Nonetheless, a principled approach to understand and quantify the associated dependence structure is still missing. We devise a general, and not model-specific, framework to achieve this task for random measure based models, which consists in: (a) quantify dependence of a random vector of probabilities in terms of closeness to exchangeability, which corresponds to the maximally dependent coupling with the same marginal distributions, that is, the comonotonic vector; (b) recast the problem in terms of the underlying random measures (in the same Fréchet class) and quantify the closeness to comonotonicity; (c) define a distance based on the Wasserstein metric, which is ideally suited for spaces of measures, to measure the dependence in a principled way. Several results, which represent the very first in the area, are obtained. In particular, useful bounds in terms of the underlying Lévy intensities are derived relying on compound Poisson approximations. These are then specialized to popular models in the Bayesian literature leading to interesting insights.

1. Introduction. A sequence of random elements $(X_n)_{n\geq 1}$ is exchangeable when its distribution is invariant with respect to finite permutations of the indices. By de Finetti's representation theorem, this intuitive symmetry requirement is equivalent to the finite-dimensional distributions being conditionally independent and identically distributed. When the random elements are grouped in a finite number of blocks, partial exchangeability [10] is a natural generalization that amounts to assuming the invariance of their joint distribution with respect to finite permutations within each block. The corresponding representation theorem states that for partially exchangeable sequences $\{X_{1,j} \mid j \geq 1\}, \dots, \{X_{k,j} \mid j \geq 1\}$ on a Polish space $\mathbb{X}$ there exists a random vector of probability measures $(\tilde{p}_1, \dots, \tilde{p}_k) \sim Q$ s.t. for any $n_i \in \mathbb{N}$ and any Borel sets $A_i \subset \mathbb{X}^{n_i}$, for $i = 1, \dots, k$,

$$\mathbb{P}\left(\bigcap_{i=1}^{k}\{(X_{i,1}, \dots, X_{i,n_i}) \in A_i\}\right) = \int_{P_{\mathbb{X}}^k} \prod_{i=1}^{k} p_i^{(n_i)}(A_i) Q(dp_1, \dots, dp_k).$$

In particular, exchangeability is recovered when $\tilde{p}_1 = \cdots = \tilde{p}_k$ almost surely (a.s.).

In Bayesian nonparametric inference, the random elements $\{X_{1,j} \mid j \geq 1\}, \dots, \{X_{k,j} \mid j \geq 1\}$ are regarded as observables and a fundamental issue is the choice of the distribution Q for the random vector of probability measures $(\tilde{p}_1, \dots, \tilde{p}_k)$, the prior distribution. The dependence between the random probabilities is of crucial importance, since it regulates the dependence between groups of observations and, consequently, the borrowing of information across groups. The first proposal of a dependent nonparametric prior dates back to Cifarelli

Received June 2020; revised December 2020.
MSC2020 subject classifications. 62C10, 60G57, 60G09.
Key words and phrases. Bayesian nonparametrics, completely random measures, completely random vectors, compound Poisson, dependence, independent increments, Lévy copula, Wasserstein distance.

and Regazzini [5], but it was the two seminal papers of MacEachern [35, 36] that led to an impressive growth of research in this direction. Most classes of priors are defined to select a.s. discrete $\tilde{p}_i$'s, since this naturally allows for clustering at either the observations or latent level. This is true also in the exchangeable case: most notable a.s. discrete $\tilde{p}$ are obtained through either the stick-breaking construction [21, 49] or a suitable transformation of a completely random measure (CRM) $\tilde{\mu}$ [26, 33]. The former approach is particularly effective for computational purposes, whereas the latter allows us to derive important distributional properties. In particular, by using CRMs as a unifying concept, as showcased in [33], one obtains popular classes of nonparametric priors such as, for example, normalized random measures [45], neutral-to-the-right processes [11] and kernel mixtures of random measures [12, 23]. Correspondingly, in the general partially exchangeable case, one may distinguish two approaches for building dependent priors: the first approach models the dependence at the level of the atoms and/or the jumps of the stick-breaking construction of each $\tilde{p}_i$; the second models the dependence at the level of the CRMs $(\tilde{\mu}_1, \dots, \tilde{\mu}_k)$ to then obtain a dependent vector $(\tilde{p}_1, \dots, \tilde{p}_k)$ via a suitable transformation. See [15, 20, 38, 39] for extensive accounts. A crucial gap in this vast literature is the understanding and quantification of the dependence structure of a dependent nonparametric prior in order to both elicit prior parameters to achieve the desired degree of dependence and compare different priors themselves. The most natural way to approach the problem is to measure closeness to exchangeability, which corresponds to the extreme case of maximal dependence between populations. Within a parametric framework, already in 1938, de Finetti proposed to use *approximately* exchangeable priors to deal with contingency tables [10]. Recently, Bacallado, Diaconis and Holmes [1] enriched this class of examples and proposed ways to use them to test for the exchangeability assumption. However, closeness to exchangeability is left as an essentially intuitive notion. To the best of our knowledge, the only measure of dependence that has been used so far is the pairwise linear correlation of $(\tilde{p}_i(A), \tilde{p}_j(A))$, for any given set A, which is certainly useful but reducing dependencies between random probabilities to linear correlation is hardly satisfying.

Here, we tackle the problem in a general nonparametric framework adopting a principled approach in that we measure the distance to exchangeability in terms of the Wasserstein distance. Because of its intrinsically geometric definition, the Wasserstein distance is the most appropriate choice for describing the similarity between distributions. As explained in [44], this distance was first introduced by Gini [16] with this exact scope. During the past century, the Wasserstein distance was introduced and studied in many fields of research, including optimal transport theory, partial differential equations and ergodic theory. Recently, it has gained a renewed popularity in probability, statistics and the related fields of machine learning and optimization, where the distinguished theoretical properties are now supported by efficient algorithms [7]. See [42, 51] for detailed reviews. The first to use the Wasserstein distance in a Bayesian nonparametric framework, for asymptotic investigations, has been Nguyen [40] who has convincingly argued for it as an effective tool to handle discrete nonparametric priors. See also [41]. From our perspective, the Wasserstein distance is the ideal choice because it allows for a meaningful comparison between distributions with different support and without density, as the ones arising from transformations of CRMs. This property is not shared by the most common distances and divergences, such as the total variation distance, the Hellinger distance or the Kullback–Leibler divergence.

Our general setup is as follows. For simplicity, we consider the case $k = 2$, even though most of our results may be extended to a generic k with no additional cost. Since our leading purpose is to measure the closeness to exchangeability (i.e., $\tilde{p}_1 = \tilde{p}_2$ a.s), we consider random vectors $(\tilde{p}_1, \tilde{p}_2)$ with equal marginal distributions ($\tilde{p}_1 \overset{\mathrm{d}}{=} \tilde{p}_2$, where $\overset{\mathrm{d}}{=}$ stands for equality in distribution). A crucial observation is then the following: instead of measuring the distance from exchangeability of $(\tilde{p}_1, \tilde{p}_2)$, we work with *completely random vectors*

(CRVs) $(\tilde{\mu}_1, \tilde{\mu}_2)$, characterized by jointly independent increments, and measure their closeness to the comonotonic case, that is, $\tilde{\mu}_1 = \tilde{\mu}_2$ a.s. In fact, since most random vectors of discrete probabilities $(\tilde{p}_1, \tilde{p}_2)$ are obtained by a suitable componentwise transformation T of a CRV, $(T(\tilde{\mu}_1), T(\tilde{\mu}_2))$ (see [33] for details), comonotonic CRVs correspond to exchangeability. Working directly with the random measures rather than their transformed versions has two distinct advantages: (a) it provides a general and not model-specific framework for the analysis of dependence, which can then be tailored to the particular class of models one is interested in, as we do in Section 7; (b) it significantly simplifies the mathematical analysis. Closeness to the comonotonic case is then measured through the following distance on CRVs, which will be shown in Section 2 to be well defined,

$$d_{\mathcal{W}}\left(\begin{pmatrix}\tilde{\mu}_1\\ \tilde{\mu}_2\end{pmatrix}, \begin{pmatrix}\tilde{\xi}_1\\ \tilde{\xi}_2\end{pmatrix}\right) = \sup_{A\in\mathcal{X}} \mathcal{W}\left(\begin{pmatrix}\tilde{\mu}_1(A)\\ \tilde{\mu}_2(A)\end{pmatrix}, \begin{pmatrix}\tilde{\xi}_1(A)\\ \tilde{\xi}_2(A)\end{pmatrix}\right), \tag{1}$$

where $\mathcal{W}$ denotes the 2-Wasserstein distance on the Euclidean plane. The goal of this work is then to provide an analytical expression for the distance $d_{\mathcal{W}}$ in (1) with a particular focus on the distance between a CRV $(\tilde{\mu}_1, \tilde{\mu}_2)$ and the comonotonic random vector $(\tilde{\xi}_1, \tilde{\xi}_2)$ in the same Fréchet class, that is, with the same marginal distributions. We stress that our results, even though motivated by Bayesian nonparametric models, are of independent probabilistic interest with reference to the theory of multidimensional random measures and Lévy processes.

The two major challenges in the treatment of $d_{\mathcal{W}}$ in (1) may be summarized as follows. (i) The analytical computation of the Wasserstein distance needs the appointment of an optimal transport map. While this is always known in explicit form for univariate distributions, the general expression for multidimensional ones is still an open problem, with only a few known cases. Crucially, in Theorem 2 we are able to determine the optimal transport map to the comonotonic vector for any CRV $(\tilde{\mu}_1, \tilde{\mu}_2)$. This allows us to express the Wasserstein distance as an integral that involves the cumulative distribution function (cdf) of $\tilde{\mu}_1(A) + \tilde{\mu}_2(A)$, which in some cases may be computed directly, an example being when one considers the Wasserstein distance between comonotonicity and independence. (ii) The law of a CRV is usually characterized through a bivariate Lévy measure, so that the cdf of $\tilde{\mu}_1(A) + \tilde{\mu}_2(A)$ is not available in closed form. Hence, Theorem 5 is of particular importance, since we are able to find tight bounds of the distance that are expressed in terms of the Lévy measures. This is achieved through suitable compound Poisson approximations of the random vectors and by finding a new informative bound for the Wassserstein distance between multivariate compound Poisson distributions (Proposition 6). With the aim of emphasizing their role in Bayesian nonparametric inference, we then compute the bounds for $d_{\mathcal{W}}$ for laws of $(\tilde{\mu}_1, \tilde{\mu}_2)$ that correspond to well-known priors with partially exchangeable data, leading to meaningful insights and a quantification of their dependence structure in terms of the hyperparameters.

Our measure of dependence may be naturally extended to $k > 2$ groups by considering the Wasserstein distance on $\mathbb{R}^k$ from $(\tilde{\mu}_1(A), \dots, \tilde{\mu}_k(A))$ such that $\tilde{\mu}_1(A) = \cdots = \tilde{\mu}_k(A)$ a.s., that is, the comonotonic k-dimensional CRV corresponding to exchangeability. The main techniques still apply to the k-dimensional case. We underline that the natural extension to an arbitrary k represents a further benefit of our measure of dependence compared to linear correlation, since it provides an overall quantification of dependence without forcing pairwise comparisons.

The techniques that we introduce may also be used to measure the dependence directly on componentwise transformations $(T(\tilde{\mu}_1), T(\tilde{\mu}_2))$ of a CRV, which may be seen as a complementary model-specific analysis. However, this requires additional work and depends on the choice of T, since the Wasserstein distance in not transformation invariant. Here, we develop

informative bounds for a specific transformation that is widely used in Bayesian nonparametric inference for time-to-event data, namely random hazards modeled as kernel mixtures over a CRM. Since the hazards characterize the associated probability distribution, this provides a specification for the de Finetti measure. The inferential properties of this class of nonparametric priors were thoroughly studied in [12, 23, 34] for exchangeable observations and have seen interesting generalizations to a partially exchangeable setting [3, 31].

The paper is structured as follows. In Section 2, we introduce necessary concepts and notation and prove that $d_{\mathcal{W}}$ is actually a distance. In Section 3, we obtain an integral representation of the Wasserstein distance between a random vector of measures and the corresponding comonotonic one. In Section 4, we develop general bounds for the distance between CRVs in the same Fréchet class, in terms of their bivariate Lévy intensities. In Section 5, we focus on the distance from exchangeability and obtain an explicit form for the bounds of the previous section. In particular, in Section 6 we use them to bound the distance between exchangeability and the other extreme case, independence. In Section 7, the previous techniques are used to quantify the dependence of three popular nonparametric priors for partially exchangeable data, namely compound random measures [17, 46], Clayton Lévy copula [13, 29, 50] and GM-dependence [18, 32]. In Section 8, we extend the measure of dependence to random hazards that are modeled as kernel mixtures over a CRV, with a specific application to GM-dependence [31]. All proofs are deferred to Section 9.

2. Preliminaries. We first recall definitions and key properties of random vectors of measures and of the Wasserstein distance. To fix notation, let $\mathbb{R}_+ = (0, +\infty)$ and $\mathbb{R}_+^2 := [0, +\infty) \times [0, +\infty) \setminus \{(0, 0)\}$. Moreover, $\mathcal{L}(X)$ denotes the law of a random variable X.

Let $(\mathbb{X}, d_{\mathbb{X}})$ be a Polish space endowed with a distance $d_{\mathbb{X}}$ and the Borel σ-algebra $\mathcal{X}$. We denote by $(M_{\mathbb{X}}, \mathcal{M}_{\mathbb{X}})$ the Borel space of boundedly finite measures on $\mathbb{X}$ endowed with the topology of weak$^\sharp$ convergence [8]. A random vector of measures is a measurable function $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2) : \Omega \to M_{\mathbb{X}}^2$, where $(\Omega, \Sigma_\Omega, P_\Omega)$ is a generic probability space and $M_{\mathbb{X}}^2 = M_{\mathbb{X}} \times M_{\mathbb{X}}$ is endowed with the product σ-algebra. We refer to the projections $\pi_i \circ \tilde{\boldsymbol{\mu}} = \tilde{\mu}_i : \Omega \to M_{\mathbb{X}}$, for $i = 1, 2$, as the marginals of $\tilde{\boldsymbol{\mu}}$. Moreover, the random vectors evaluated on a set are denoted as $\tilde{\boldsymbol{\mu}}(A) = (\tilde{\mu}_1(A), \tilde{\mu}_2(A)) : \Omega \to [0, +\infty) \times [0, +\infty)$, for every $A \in \mathcal{X}$.

DEFINITION 1. A random vector of measures $\tilde{\boldsymbol{\mu}}$ is a *completely random vector* (CRV) if, given a finite collection of disjoint bounded Borel sets $\{A_1, \dots, A_n\}$, the random vectors $\{\tilde{\boldsymbol{\mu}}(A_1), \dots, \tilde{\boldsymbol{\mu}}(A_n)\}$ are independent.

In particular, this definition entails that the marginal distributions $\tilde{\mu}_1, \tilde{\mu}_2$ have independent increments and are thus completely random measures (CRMs) in the sense of Kingman [26]. We point out that the converse is not necessarily true: a random vector of measures whose marginals are CRMs is not necessarily a CRV. The joint independence of the increments guarantees that the distribution of $\tilde{\boldsymbol{\mu}}$ is characterized by the distribution of the random vectors evaluated on a set $\{\tilde{\boldsymbol{\mu}}(A) \mid A \in \mathcal{X}\}$. Moreover, [24], Theorem 3.19, ensures that, if $\tilde{\boldsymbol{\mu}}$ has no fixed atoms, there exists a Poisson random measure $\mathcal{N}$ on $\mathbb{R}_+^2 \times \mathbb{X}$ s.t. for every $A \in \mathcal{X}$,

$$\tilde{\boldsymbol{\mu}}(A) \overset{\mathrm{d}}{=} \int_{\mathbb{R}_+^2 \times A} \boldsymbol{s}\mathcal{N}(ds_1, ds_2, dx), \tag{2}$$

where $\boldsymbol{s} = (s_1, s_2)$. The mean measure $\nu(ds_1, ds_2, dx) = \mathbb{E}(\mathcal{N}(ds_1, ds_2, dx))$ satisfies the following properties: $\nu(\mathbb{R}_+^2 \times \{x\}) = 0$ for every $x \in \mathbb{X}$ and

$$\int_{\mathbb{R}_+^2 \times A} \min\{s_1 + s_2, \epsilon\} \nu(ds_1, ds_2, dx) < +\infty \tag{3}$$

for every bounded $A \in \mathcal{X}$ and every $\epsilon > 0$. We will focus on CRVs without fixed atoms and refer to ν as the *intensity measure* of $\tilde{\boldsymbol{\mu}}$. This will be further assumed to have no atoms. Campbell's theorem ensures that from the Lévy intensity of $\tilde{\boldsymbol{\mu}}$ one derives the Lévy intensities of the marginal CRMs $\tilde{\mu}_1$ and $\tilde{\mu}_2$, namely

$$\nu_1(ds, dx) = \int_{[0,+\infty)} \nu(ds, ds_2, dx), \qquad \nu_2(ds, dx) = \int_{[0,+\infty)} \nu(ds_1, ds, dx).$$

We underline that the marginal CRMs are not forced to have the same atoms a.s. because the measure ν may have positive mass on the axes, as it will be clear from Section 6. We say that $\tilde{\boldsymbol{\mu}}$ is *infinitely active* if for every $A \in \mathcal{X}$ both the marginal CRMs are infinitely active, that is,

$$\int_{\mathbb{R}_+ \times A} \nu_1(ds, dx) = \int_{\mathbb{R}_+ \times A} \nu_2(ds, dx) = +\infty. \tag{4}$$

Since most applications of random measures in Bayesian nonparametrics deal with infinitely active random measures, we concentrate on these.

The distribution of a CRV is characterized by the distribution of its evaluations on a set $\{\tilde{\boldsymbol{\mu}}(A) \mid A \in \mathcal{X}\}$. Thus any distance $\mathcal{D}$ on the space $P(\mathbb{R}^2)$ of probability measures on $\mathbb{R}^2$ determines a distance on the laws of CRVs by considering

$$\sup_{A \in \mathcal{X}} \mathcal{D}\big(\mathcal{L}(\tilde{\boldsymbol{\mu}}^1(A)), \mathcal{L}(\tilde{\boldsymbol{\mu}}^2(A))\big).$$

The distance $d_{\mathcal{W}}$ defined in (1) fits in this general framework, by considering the Wasserstein distance as metric $\mathcal{D}$. Given π_1, π_2 two probability measures on a Polish space $(\mathbb{X}, d_{\mathbb{X}})$, we indicate by $C(\pi_1, \pi_2)$ the Fréchet class of π_1 and π_2, that is, the set of distributions on the product space whose marginal distributions coincide with π_1 and π_2, respectively. If Z_1 and Z_2 are dependent random variables on $\mathbb{X}$ such that their joint law $\mathcal{L}(Z_1, Z_2) \in C(\pi_1, \pi_2)$, we write $(Z_1, Z_2) \in C(\pi_1, \pi_2)$.

DEFINITION 2. The Wasserstein distance of order $p \in [1, +\infty)$ between π_1 and π_2 is

$$\mathcal{W}_{p,d_{\mathbb{X}}}(\pi_1, \pi_2) = \inf_{(Z_1, Z_2) \in C(\pi_1, \pi_2)} \big\{\mathbb{E}\big(d_{\mathbb{X}}(Z_1, Z_2)^p\big)\big\}^{\frac{1}{p}}.$$

By extension, we refer to the Wasserstein distance between two random elements $X_i : \Omega \to \mathbb{X}$, $i = 1, 2$, as the Wasserstein distance between their laws, that is, $\mathcal{W}_{p,d}(X_1, X_2) = \mathcal{W}_{p,d}(\mathcal{L}(X_1), \mathcal{L}(X_2))$. An element of $C(\mathcal{L}(X_1), \mathcal{L}(X_2))$ is referred to as a coupling between X_1 and X_2.

Throughout the work, we set $p = 2$ and $(\mathbb{X}, d_{\mathbb{X}}) = (\mathbb{R}^2, \|\cdot\|)$, that is, the Euclidean plane. We will refer to such distance as the Wasserstein distance and denote it by $\mathcal{W}$, that is,

$$\mathcal{W}(\boldsymbol{X}, \boldsymbol{Y}) = \inf_{(\boldsymbol{Z}_X, \boldsymbol{Z}_Y) \in C(\boldsymbol{X}, \boldsymbol{Y})} \big\{\mathbb{E}\big(\|\boldsymbol{Z}_X - \boldsymbol{Z}_Y\|^2\big)\big\}^{\frac{1}{2}},$$

where we have used the vector notation $\boldsymbol{X} = (X_1, X_2) \in \mathbb{R}^2$. The parallelogram rule on normed spaces ensures that

$$\mathcal{W}(\boldsymbol{X}, \boldsymbol{Y})^2 \le 2\big(\mathbb{E}\big(\|\boldsymbol{X}\|^2\big) + \mathbb{E}\big(\|\boldsymbol{Y}\|^2\big)\big). \tag{5}$$

In particular, the Wasserstein distance between random elements on $\mathbb{R}^2$ with finite expected squared norm is finite. Thus, in order for $d_{\mathcal{W}}$ in (1) to be finite, we restrict to random vectors of measures with finite second moment $\mathbb{E}(\|\tilde{\boldsymbol{\mu}}(\mathbb{X})\|^2) = \mathbb{E}(\tilde{\mu}_1(\mathbb{X})^2) + \mathbb{E}(\tilde{\mu}_2(\mathbb{X})^2) < +\infty$.

Therefore, by standard properties of Poisson random measures, we ask

$$\mathbb{E}(\tilde{\boldsymbol{\mu}}(\mathbb{X})) = \int_{\mathbb{R}_+^2 \times \mathbb{X}} \boldsymbol{s}\nu(ds_1, ds_2, dx) < +\boldsymbol{\infty}, \tag{6}$$

$$\mathrm{Var}(\tilde{\boldsymbol{\mu}}(\mathbb{X})) = \int_{\mathbb{R}_+^2 \times \mathbb{X}} \boldsymbol{s}^2\nu(ds_1, ds_2, dx) < +\boldsymbol{\infty}, \tag{7}$$

where $\boldsymbol{s}^2 = (s_1^2, s_2^2)$ and $+\boldsymbol{\infty} = (+\infty, +\infty)$. We summarize our findings in the following.

PROPOSITION 1. *The function $d_{\mathcal{W}} : \mathbb{P}(M_{\mathbb{X}}^2) \times \mathbb{P}(M_{\mathbb{X}}^2) \to [0, +\infty)$ defines a distance on the laws of CRVs whose Lévy intensities satisfy* (6) *and* (7).

We conclude this section by recalling some properties of the Wasserstein distance to be used in the sequel. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two random elements in $\mathbb{R}^2$. A coupling $(\boldsymbol{Z}_X, \boldsymbol{Z}_Y) \in C(\boldsymbol{X}, \boldsymbol{Y})$ is said to be optimal if $\mathcal{W}(\boldsymbol{X}, \boldsymbol{Y}) = \mathbb{E}(\|\boldsymbol{Z}_X - \boldsymbol{Z}_Y\|^2)^{\frac{1}{2}}$. If an optimal coupling satisfies $\boldsymbol{Z}_X = \phi(\boldsymbol{Z}_Y)$ a.s. for some measurable function ϕ, we refer to ϕ as an optimal (transport) map from $\boldsymbol{X}$ to $\boldsymbol{Y}$. Optimal maps for the Wasserstein distance on the Euclidean line always exist and are explicitly available; on the contrary, on the Euclidean plane they are not guaranteed to exist if $\boldsymbol{X}$ gives nonzero mass to sets of codimension greater or equal to 1. Moreover, even when the existence is established, there is no explicit way to build such maps, except in few particular cases; see [51]. However, Knott and Smith [27] appointed derived a sufficient criterion to establish the optimality of a map, namely to express it as the gradient of a convex function. We will use this result in a reformulation provided by [47]. When an optimal transport map ϕ is available, the Wasserstein distance amounts to an expected value with respect to a degenerate distribution having support on a 2-dimensional subspace of $\mathbb{R}^4$. Nonetheless, the evaluation of such an integral is still a challenging task since bivariate integrals can be difficult to evaluate not only analytically but also numerically.

3. Distance from exchangeability. Having established conditions for $d_{\mathcal{W}}$ in (1) to be a distance on CRVs, we now use $d_{\mathcal{W}}$ to compare CRVs $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\xi}}$ in the same Fréchet class, that is, with equally distributed marginal random measures ($\tilde{\mu}_1 \overset{\mathrm{d}}{=} \tilde{\xi}_1$; $\tilde{\mu}_2 \overset{\mathrm{d}}{=} \tilde{\xi}_2$), and focus on the comparison between their dependence structures. To this end, we put particular emphasis on the Wasserstein distance from comonotonic random vectors, which induce exchangeable priors. In this section, we provide an analytical expression for the optimal transportation map from a generic CRV to the comonotonic one in the same Fréchet class. This will then be used to evaluate the exact distance between exchangeability and the other extreme case, independence.

DEFINITION 3. A random vector of measures $\tilde{\boldsymbol{\mu}}$ is said to be completely dependent or *comonotonic* if $\tilde{\mu}_1 = \tilde{\mu}_2$ a.s. We write $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\mathrm{co}}$.

In particular, we point out that every random vector of measures $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2)$ in the same Fréchet class of $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ satisfies $\tilde{\mu}_1 \overset{\mathrm{d}}{=} \tilde{\mu}_2$. For this reason, since our main interest lies in exchangeability, and thus in comonotonicity, throughout the work we deal with random vectors of measures with equal marginal distributions. It should be stressed, though, that many of our results and techniques could be easily extended to other settings.

THEOREM 2. *Let $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ be CRVs in the same Fréchet class s.t. condition* (6) *on the Lévy intensities holds. Then*

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A))^2 = 4(\mathbb{E}(\tilde{\mu}_1(A)^2) - \omega_{\tilde{\boldsymbol{\mu}},A}), \tag{8}$$

where $\omega_{\tilde{\boldsymbol{\mu}},A} = \mathbb{E}(\tilde{\mu}_1(A) F^{-1}_{\tilde{\mu}_1(A)}(F_{\tilde{\mu}_1(A)+\tilde{\mu}_2(A)}(\tilde{\mu}_1(A)+\tilde{\mu}_2(A)))$, *with* F_X *denoting the cumulative distribution function* (*cdf*) *of* X.

Upon defining $X_i = \tilde{\mu}_i(A)$ for $i = 1, 2$, it is useful to observe that the right-hand side of (8) is equal to $4(\mathbb{E}(X_1^2) - \mathbb{E}(X_1 F^{-1}_{X_1}(F_{X_1+X_2}(X_1+X_2))))$. In particular, for $\tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ one has $X_1 = X_2 = X$, so that (8) becomes $4(\mathbb{E}(X^2) - \mathbb{E}(X F^{-1}_X(F_{2X}(2X)))) = 0$, since $F_{2X}(2X) = F_X(X)$. Moreover, when the distribution of $\tilde{\boldsymbol{\mu}}$ is symmetric, that is, $\mathcal{L}(\tilde{\mu}_1, \tilde{\mu}_2) = \mathcal{L}(\tilde{\mu}_2, \tilde{\mu}_1)$, one finds the following alternative expression for $\omega_{\tilde{\boldsymbol{\mu}},A}$ in (8).

LEMMA 3. *Let* $\tilde{\boldsymbol{\mu}}$ *be a symmetric CRV satisfying the conditions of Theorem* 2. *Then*

$$\omega_{\tilde{\boldsymbol{\mu}},A} = \frac{1}{2}\mathbb{E}\big(F^{-1}_{\tilde{\mu}_1(A)+\tilde{\mu}_2(A)}(U) F^{-1}_{\tilde{\mu}_1(A)}(U)\big),$$

where $U \sim \mathrm{Unif}([0,1])$ *is a uniform random variable on* $[0,1]$.

The expression of $\omega_{\tilde{\boldsymbol{\mu}},A}$ in Theorem 2 and Lemma 3 involves the dependence structure of $\tilde{\boldsymbol{\mu}}$ and is to be evaluated case-by-case. In some specific cases, it can be computed directly leading to the exact bivariate Wasserstein distance with respect to a comonotonic random vector in the same Fréchet class, in short *the Wasserstein distance from exchangeability*. For instance, consider a CRV $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ whose marginals are independent gamma CRMs. Recall that $\tilde{\mu}$ is a gamma CRM with base measure αP_0 if the Lévy intensity is

$$\pi(ds, dx) = \alpha P_0(dx) \frac{e^{-s}}{s} \mathbb{1}_{(0,+\infty)}(s)\, ds, \tag{9}$$

where $\alpha > 0$ and P_0 is a probability distribution on $\mathbb{X}$. Moreover, $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ is a symmetric CRV, so that Proposition 3 applies. We define

$$\begin{aligned}\omega_{\alpha,P_0,A} = {} & \frac{1}{\Gamma(2\alpha P_0(A)+1)} \\ & \times \int_0^{+\infty} \mathrm{Inv}\Gamma_{\alpha P_0(A)}\left(\frac{\Gamma(\alpha P_0(A))}{\Gamma(2\alpha P_0(A))}\Gamma\big(2\alpha P_0(A), t\big)\right) e^{-t} t^{2\alpha P_0(A)}\, dt,\end{aligned}$$

where $\Gamma(a, s) = \int_s^{+\infty} e^{-t} t^{a-1}\, dt$ is the upper incomplete gamma function and $\mathrm{Inv}\Gamma_a(\cdot)$ is the inverse function of $\Gamma(a, \cdot)$.

COROLLARY 4. *Let* $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ *and* $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ *be in the same Fréchet class with marginal gamma CRM with base measure* αP_0. *Then*

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big)^2 = 4\alpha P_0(A)\big(1 + \alpha P_0(A) - \omega_{\alpha,P_0,A}\big).$$

Moreover,

$$\omega_{\alpha,P_0,A} = \frac{1}{2}\int_0^1 \mathrm{Inv}\Gamma_{2\alpha P_0(A)}(t)\, \mathrm{Inv}\Gamma_{\alpha P_0(A)}(t)\, dt.$$

For fixed values of $\alpha P_0(A)$, we can evaluate this quantity numerically. For example, Figure 1 corresponds to $\alpha = 1$ and $A = \mathbb{X}$, so that numerical simulations yield $\omega_{\alpha,P_0,A} \approx 1.70$. The analytical value is compared with the simulated Wasserstein distance between the empirical measures, which is known to converge to the Wasserstein distance between the underlying distributions as the size of the samples diverges. In many other cases, the evaluation of the expression in Theorem 2 is impossible in practice. For example, this happens if the analytical expression for $F_{\tilde{\mu}_1(A)}$ is not available in closed form, or when the dependence between

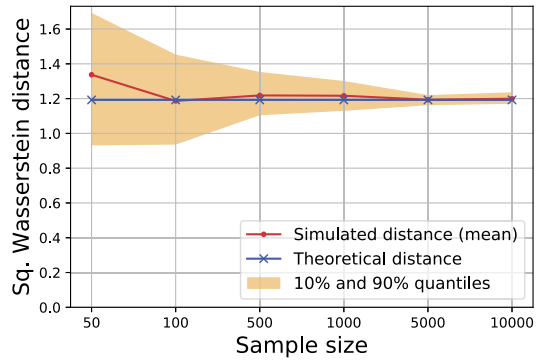

FIG. 1. *Simulation of the empirical Wasserstein distance between a bivariate distribution with independent gamma marginals with shape = scale = 1 and a bivariate distribution with a.s. equal gamma marginals of shape = scale = 1. Simulations were performed with independent samples, independent for each sample size, using the Python Optimal Transport (POT) package* [14].

the random measures is modeled through the bivariate Lévy intensity. Moreover, we observe that the quantities in Theorem 2 and Corollary 4 depend on A in a nontrivial manner, so that finding the supremum over all Borel sets as in (1) may not be an easy task. This raises the need for informative and tractable upper bounds on the distance, whose expression depends directly on the underlying Lévy intensity. Note that the upper bound in (5) only depends on the marginal distributions of the random vectors, and thus does not provide any information on their dependence structures.

4. Bounds on Fréchet classes. Given the difficulty in evaluating the integral expression of Theorem 2 for the Wasserstein distance between a CRV and a comonotonic one in the same Fréchet class, we aim at deriving suitable bounds. We first face the problem in general and develop upper bounds for the Wasserstein distance between two CRVs. Then, in the following sections, these general bounds will be specialized to the distance from exchangeability, which is the case of interest for Bayesian inference. Our general bounds rely on a compound Poisson approximation of the CRVs that is induced by certain *compatible* families of neighborhoods of the origin. Henceforth, we assume that $\tilde{\boldsymbol{\mu}}$ is an infinitely active CRV s.t. condition (6) on the Lévy intensity ν holds.

DEFINITION 4. Consider a family $B = \{B(\epsilon) \mid \epsilon \in (0, 1]\}$ of measurable neighborhoods of the origin in $\mathbb{R}^2_+$ s.t.

(B1) the family is increasing, that is, $\epsilon_1 \leq \epsilon_2$ implies that $B(\epsilon_1) \subset B(\epsilon_2)$;

(B2) the Lévy intensity gives zero mass to their intersection, that is, $\nu(\bigcap_{\epsilon\in(0,1]} B(\epsilon) \times A) = 0$ for every $A \in \mathcal{X}$;

(B3) the sets $D = \{D(\epsilon) = B(\epsilon)^c = \mathbb{R}^2_+ \setminus B(\epsilon) \mid \epsilon \in (0, 1]\}$ have continuously increasing mass, that is, there exists $r_0 = \nu(\bigcap_{\epsilon\in(0,1]} D(\epsilon))$ s.t. for every $r > r_0$ there exists $\epsilon_r = \epsilon_{r,A}$ s.t. $\nu(D(\epsilon_r) \times A) = r$.

Then we say that the family B is *compatible* with $\tilde{\boldsymbol{\mu}}$. By extension, we will also refer to the family of complementary sets D as compatible.

REMARK 1. Some technical comments are in order: (a) The choice of the index set to be $(0, 1]$ is arbitrary. Indeed, one could replace it with any neighborhood of the origin in $\mathbb{R}^+$. (b) The uncountable intersection $\bigcap_{\epsilon\in(0,1]} D(\epsilon)$ is measurable because the family is increasing. One can find more on this in Section 9. (c) Property (B2) does not contradict the continuity of the measure since ν is an infinite measure.

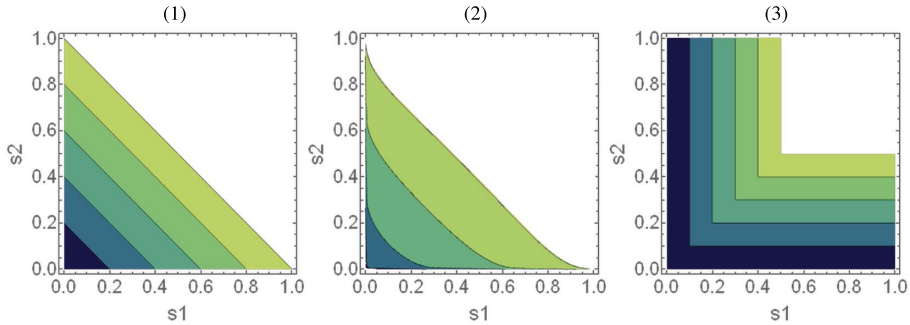

FIG. 2. *Three families of neighborhoods of the origin*: (1) $B^{+}(\epsilon) = \{(s_1, s_2) \mid s_1 + s_2 \le \epsilon\}$; (2) $B^{E_1}(\epsilon) = \{(s_1, s_2) \mid E_1(s_1)^{-\theta} + E_1(s_2)^{-\theta} \le \epsilon\}$ *with* $\theta = 0.5$; (3) $B^{\min}(\epsilon) = \{(s_1, s_2) \mid \min(s_1, s_2) \le \epsilon\}$.

REMARK 2. A standard way to find a family of measurable neighborhoods of the origin that satisfy (B1) is to consider the level sets

$$B^{g}(\epsilon) = \{(s_1, s_2) \mid g(s_1, s_2) \le \epsilon\}, \tag{10}$$

where $g : [0, +\infty) \times [0, +\infty) \to \mathbb{R}^{+}$ is a measurable function s.t. $g(0, 0) = 0$. See Figure 2. Depending on the support of ν, properties (B2) and (B3) may hold. For example, if $g(s_1, s_2) = \min(s_1, s_2)$, B^{g} is not compatible with ν having mass on the axis, whereas it is compatible with ν being absolutely continuous (a.c.) w.r.t. the Lebesgue measure.

As it will be seen in the sequel, we will mostly be interested in Lévy intensities that are a.c. w.r.t. the Lebesgue measure or have mass on lines passing through the origin. In these cases, every continuous map g s.t. $g(s_1, s_2) = 0$ if and only if $(s_1, s_2) = (0, 0)$ induces a compatible family. In particular, we will be interested in the families $B^{+} = \{(s_1, s_2) \mid s_1 + s_2 \le \epsilon\}$ and $B^{E_1} = \{(s_1, s_2) \mid E_1(s_1)^{-\theta} + E_1(s_2)^{-\theta} \le \epsilon\}$ with $\theta = 0.5$ appearing in Figure 2, where $E_1(s) = \Gamma(0, s)$ is the exponential integral.

Given a compatible family D, for every $r > r_0$ and $A \in \mathcal{X}$ we define the probability distribution $\rho_{r,A,D}$ on $\mathbb{R}^2_{+}$ as

$$\rho_{r,A,D}(ds_1, ds_2) = \frac{1}{r}\nu(ds_1, ds_2, A)\mathbb{1}_{D(\epsilon_{r,A})}(s_1, s_2), \tag{11}$$

where we use the notation $\nu(ds_1, ds_2, A) = \int_A \nu(ds_1, ds_2, dy)$. As apparent from the proof of the next theorem, this coincides with the distribution of the jumps of a compound Poisson approximation of $\tilde{\boldsymbol{\mu}}$.

THEOREM 5. *Let* $\tilde{\boldsymbol{\mu}}^1$ *and* $\tilde{\boldsymbol{\mu}}^2$ *be infinitely active CRVs in the same Fréchet class s.t. condition* (6) *on the Lévy intensities holds. Then*

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)) \le \lim_{r \to +\infty} \sqrt{r}\mathcal{W}(\rho^1_{r,A,D_1}, \rho^2_{r,A,D_2}),$$

for any D_i *compatible family for* $\tilde{\boldsymbol{\mu}}^i$, *for* $i = 1, 2$. *Moreover, the upper bound on the right-hand side is finite and does not depend on* D_1 *and* D_2.

REMARK 3. Since any CRV $\tilde{\boldsymbol{\mu}}$ has infinitely many compatible family, the above theorem holds also in the case $\tilde{\boldsymbol{\mu}}^1 = \tilde{\boldsymbol{\mu}}^2$ and $D_1 \ne D_2$. Since the limit does not depend on the families D_1 and D_2, we know that in such case it is equal to zero.

The proof is detailed in Section 9 and is based on a bound on the Wasserstein distance between compound Poisson distributions. A similar problem was treated in [37] for Lévy processes on $\mathbb{R}$. Nonetheless, the extension to $\mathbb{R}^2$ needs a new bound on the compound Poisson distributions in $\mathbb{R}^2$, summarized by the following proposition. Indeed, the arguments used in [37], Theorem 10, could be used to bound the Wasserstein distance from above with $\sqrt{r+r^2}\mathcal{W}(\rho^1_{r,A,D_1}, \rho^2_{r,A,D_2})$, which goes to $+\infty$ as $r \to +\infty$.

PROPOSITION 6. *Let $\boldsymbol{X} \overset{\mathrm{d}}{=} \sum_{i=1}^{N_x} \boldsymbol{X}^i$ and $\boldsymbol{Y} \overset{\mathrm{d}}{=} \sum_{i=1}^{N_y} \boldsymbol{Y}^i$ be two compound Poisson processes in $\mathbb{R}^2$ s.t. N_x and N_y are Poisson random variables with mean r and $\{\boldsymbol{X}^i \mid i \geq 1\}$ and $\{\boldsymbol{Y}^i \mid i \geq 1\}$ are sequences of independent and identically distributed random elements in $\mathbb{R}^2$, independent from N_x and N_y, respectively. Then*

$$\mathcal{W}(\boldsymbol{X}, \boldsymbol{Y})^2 \leq r\mathcal{W}(\boldsymbol{X}^1, \boldsymbol{Y}^1)^2 + (r^2 - r)\|\mathbb{E}(\boldsymbol{X}^1) - \mathbb{E}(\boldsymbol{Y}^1)\|^2.$$

REMARK 4. Theorem 5 bounds the Wasserstein distance between the CRVs with the Wasserstein distance between quantities that only depend on the bivariate Lévy intensities. Yet, the Wasserstein distance between these two quantities suffers from all the technical difficulties related to the Wasserstein distance in $\mathbb{R}^2$. Hence, it is complicated to evaluate it, analytically and numerically. The next sections are devoted to this task.

5. Bounds on exchangeability. Our next goal is to measure the dependence of a given CRV as the Wasserstein distance from exchangeability, which is induced by comonotonic CRVs. For this reason, we now specialize the results of the previous section, which apply to all CRVs in the same Fréchet class, to this particular framework of great importance for Bayesian inference.

In order to evaluate the bound in Theorem 5 numerically, we first need an explicit expression for the Wasserstein distance between the jumps of the compound Poisson approximations. With this goal in mind, we first dwell on the Lévy intensity ν^{co} of a comonotonic random vector $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ (Figure 3).

PROPOSITION 7. *For every $A \in \mathcal{X}$, the Lévy intensity $\nu^{\mathrm{co}}(ds_1, ds_2, A)$ has support on the bisecting line of $\mathbb{R}^2_+$, that is,*

$$\nu^{\mathrm{co}}(ds_1, ds_2, A) = \delta_{s_1}(ds_2)\nu_1(ds_1, A) = \delta_{s_2}(ds_1)\nu_2(ds_2, A).$$

It follows that every random vector $\tilde{\boldsymbol{\mu}}$ in the same Fréchet class of $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ has equal marginal Lévy intensities $\nu_1(ds, dx) = \nu_2(ds, dx)$, which we denote with $\pi(ds, dx)$. In particular, for every $A \in \mathcal{X}$, we assume that $\pi(ds, A) = \pi(s, A)\,ds$ and it is infinitely active. We denote with $U^{\pi}_A(t) = \int_{[t,+\infty)} \pi(s, A)\,ds$ its tail integral.

The following theorem provides the exact expression of the limit appearing in Theorem 5 together with a class of upper bounds. The latter are useful when the exact expression cannot

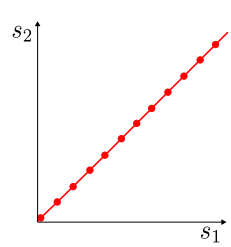

FIG. 3. *Support of the Lévy intensity of a comonotonic CRV.*

be evaluated analytically or numerically, as will be seen in Section 7.2. We first define some relevant quantities:

$$
\begin{aligned}
h^{g}_{\nu,A}(s) &= \int_{\mathbb{R}^2_+} \mathbb{1}_{(s,+\infty)}\big(g(t_1,t_2)\big)\nu(dt_1,dt_2,A);\\
K^{g}_{\nu,A} &= \sum_{i=1}^{2}\int_{\mathbb{R}^2_+} \big|s_i - \big(U^{\pi}_{A}\big)^{-1}\big(h^{g}_{\nu,A}\big(g(s_1,s_2)\big)\big)\big|^2 \nu(ds_1\,ds_2,A);
\end{aligned}
\tag{12}
$$

where $g:\mathbb{R}^2\to\mathbb{R}$ is a measurable map. When $g(s_1,s_2)=s_1+s_2$ we write $h^{+}_{\nu,A}$ and $K^{+}_{\nu,A}$. Since the sum is symmetric and ν has equal marginal measures,

$$
K^{+}_{\nu,A} = 2\int_{\mathbb{R}^2_+} \big|s_1 - \big(U^{\pi}_{A}\big)^{-1}\big(h^{+}_{\nu,A}(s_1+s_2)\big)\big|^2 \nu(ds_1\,ds_2,A).
\tag{13}
$$

THEOREM 8. *Let $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ satisfy the conditions of Theorem 5 s.t. B^{+} defined in Remark 2 is compatible with $\tilde{\boldsymbol{\mu}}$. Then*

$$
\lim_{r\to+\infty} r\mathcal{W}\big(\rho_{r,A,D},\rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}}\big)^2 = K^{+}_{\nu,A}.
\tag{14}
$$

Moreover, for every continuously differentiable $g:\mathbb{R}^2\to\mathbb{R}$ s.t. B^{g} is compatible with $\tilde{\boldsymbol{\mu}}$, $K^{+}_{\nu,A}\le K^{g}_{\nu,A}$.

Theorem 8 thus establishes that $g(s_1,s_2)=s_1+s_2$ realizes the optimal bound in the class $\{K^{g}_{\nu,A}\}$. The expression for $K^{+}_{\nu,A}$ resembles the one for the Wasserstein distance in Theorem 2 and is derived in a similar way. Nonetheless, by working at the level of the bivariate Lévy intensities rather than at the level of the evaluations on a set $\tilde{\boldsymbol{\mu}}(A)$, we overcome many of the difficulties related to its evaluation. In particular, when the Lévy intensity $\nu(\cdot,A)$ is a.c. w.r.t. the Lebesgue measure on $\mathbb{R}^2$ for any A in $\mathcal{X}$, $K^{+}_{\nu,A}$ comes in a compelling form. In such a case, we denote with $\nu(s_1,s_2,A)$ its Radon–Nikodym derivative and define

$$
K_{\nu,A} = \int_0^{+\infty} \big(U^{\pi}_{A}\big)^{-1}\big(h^{+}_{\nu,A}(t)\big)\int_0^{t} s\,\nu(s,t-s,A)\,ds\,dt,
$$

where $h^{+}_{\nu,A}$ is as in (12).

THEOREM 9. *Let $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ satisfy the conditions of Theorem 5. If the Lévy intensity of $\tilde{\boldsymbol{\mu}}$ is such that, for any $A\in\mathcal{X}$, $\nu(\cdot,A)$ is a.c. w.r.t. the Lebesgue measure on $\mathbb{R}^2$, then*

$$
\lim_{r\to+\infty} r\mathcal{W}\big(\rho_{r,A,D},\rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}}\big)^2 = 4\bigg(\int_0^{+\infty} s^2\pi(ds,A)\,ds - K_{\nu,A}\bigg).
$$

REMARK 5. We observe that the first integral in the bound only depends on the marginal distributions and provides a general upper bound for the distance. This can be seen as an improvement of the bound in (5), which amounts to

$$
\mathcal{W}\big(\tilde{\boldsymbol{\mu}}(A),\tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big)^2 \le 4\bigg(\int_0^{+\infty} s^2\pi(s,A)\,ds + \int_0^{+\infty} s\pi(s,A)\,ds\bigg),
$$

where π is the marginal Lévy intensity, as defined at the beginning of the section. On the other hand, $K_{\nu,A}$ provides relevant information on the dependence structure. In Section 7.1, this will be specialized for a concrete example.

REMARK 6. When the Lévy intensities are *homogeneous*, that is,

$$\nu(ds_1, ds_2, dx) = \alpha P_0(dx)\nu(ds_1, ds_2), \tag{15}$$

where P_0 is a probability distribution on $\mathbb{X}$ and $\alpha > 0$, also the marginal Lévy intensity takes the form $\pi(ds, dx) = \alpha\pi(s)\,ds\,P_0(dx)$ and we denote by $U_\pi(t) = \int_t^{+\infty}\pi(s)\,ds$ the tail integral. If the Lévy intensity is also diffuse, $K^+_{\nu,A} = \alpha P_0(A)K_\nu$, where

$$\begin{aligned} K_\nu &= \int_0^{+\infty}(U^\pi)^{-1}(h^+_\nu(t))\int_0^t s\nu(s, t-s)\,ds\,dt; \\ h^+_\nu(s) &= \int_{\mathbb{R}^2_+}\mathbb{1}_{(s,+\infty)}(t_1+t_2)\nu(t_1, t_2)\,dt_1\,dt_2. \end{aligned} \tag{16}$$

In particular, this entails that

$$d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^2 \le 4\alpha\left(\int_0^{+\infty} s^2\pi(ds)\,ds - K_\nu\right).$$

6. Independence. In this section, we will use Proposition 6 to bound the distance between exchangeability and the other extreme case, independence. As we shall see, in this case the Lévy intensity is not a.c. w.r.t. the Lebesgue measure, and thus the results of Theorem 9 do not apply.

Let $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ be a CRV with independent marginals and let ν^{ind} denote its Lévy intensity (Figure 4). An immediate adaptation of [25], Lemma 4.1, shows that the corresponding Lévy intensities $\nu^{\mathrm{ind}}(ds_1, ds_2, A)$ have support on the axis, namely

$$\nu^{\mathrm{ind}}(ds_1, ds_2, A) = \delta_0(ds_2)\nu^{\mathrm{ind}}_1(ds_1, A) + \delta_0(ds_1)\nu^{\mathrm{ind}}_2(ds_2, A).$$

In our setting, $\nu^{\mathrm{ind}}_1(ds_1, A) = \nu^{\mathrm{ind}}_2(ds_2, A) = \pi(s, A)\,ds$. Before stating the main result, we introduce

$$K_{\pi,A} = \int_0^{+\infty}(U^\pi_A)^{-1}(2U^\pi_A(s))s\pi(s, A)\,ds,$$

which only depends on the marginal distribution π of the CRVs.

THEOREM 10. *Let $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ be in the same Fréchet class s.t. the conditions of Theorem 5 hold. Then*

$$\lim_{r\to+\infty} r\mathcal{W}(\rho^{\mathrm{ind}}_{r,A}, \rho^{\mathrm{co}}_{r,A})^2 = 4\left(\int_0^{+\infty} s^2\pi(s, A)\,ds - K_{\pi,A}\right).$$

REMARK 7. Similar to Remark 6, when the Lévy intensities are homogeneous, $K_{\pi,A} = \alpha P_0(A)K_\pi$, where $K_\pi = \int_0^{+\infty}(U^\pi)^{-1}(2U^\pi(s))s\pi(s)\,ds$.

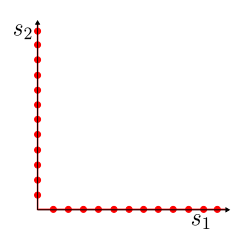

FIG. 4. *Support of the Lévy intensities of a CRV with independent marginals.*

We now apply Theorem 10 to the case where the marginal distribution is a gamma CRM with base measure αP_0, as in Corollary 4, which allows us to compare the exact Wasserstein distance with the relative bound. We first define the constant

$$\gamma = 4 - 4\int_0^{+\infty} (E_1)^{-1}\big(2E_1(s)\big)e^{-s}\,ds, \tag{17}$$

where, as before, $E_1(s) = \Gamma(0, s)$ is the exponential integral. Numerical integration show that $\gamma \approx 1.24$.

COROLLARY 11. *Let $\tilde{\boldsymbol{\mu}}^{\text{ind}}$ and $\tilde{\boldsymbol{\mu}}^{\text{co}}$ be in the same Fréchet class with marginal gamma CRM with base measure αP_0. Then*

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}^{\text{ind}}(A), \tilde{\boldsymbol{\mu}}^{\text{co}}(A)\big)^2 \le \gamma\alpha P_0(A).$$

In particular, $d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}^{\text{ind}}, \tilde{\boldsymbol{\mu}}^{\text{co}})^2 \le \gamma\alpha$.

In Figure 5, we present a graphical comparison between the exact distance in Corollary 4, the simulated empirical distance in Figure 1 as the sample size increases and the theoretical bound established in Theorem 11. We omit the noninformative bound in Remark 5 from the figure because it is out of scale (equal to 8) and point out that the theoretical bound appears to be very tight.

Similar results may be achieved for *generalized gamma* CRMs, whose Lévy intensity is

$$\pi(ds, dx) = \alpha P_0(dx)e^{-bs}s^{-1-\sigma}\mathbb{1}_{(0,+\infty)}(s)\,ds,$$

for some $\alpha > 0$, P_0 a probability distribution on $\mathbb{X}$, $b > 0$ and $\sigma \in (0, 1)$. In particular, gamma random measures as defined in (9) are achieved when $\sigma = 0$ and $b = 1$. We define

$$\gamma_{b,\sigma} = 4 - 4\frac{1}{b\Gamma(1-\sigma)}\int_0^{+\infty} \operatorname{Inv}\Gamma_{-\sigma}\big(2\Gamma(-\sigma, bs)\big)e^{-bs}s^{-\sigma}\,ds, \tag{18}$$

where $\Gamma(a, s) = \int_s^{+\infty} e^{-t}t^{a-1}\,dt$ is the upper incomplete gamma function and $\operatorname{Inv}\Gamma_a(\cdot)$ is the inverse function of $\Gamma(a, \cdot)$. Clearly, $\gamma_{1,0} = \gamma$ in (17).

COROLLARY 12. *Let $\tilde{\boldsymbol{\mu}}^{\text{ind}}$ and $\tilde{\boldsymbol{\mu}}^{\text{co}}$ be in the same Fréchet class with marginal generalized gamma CRM with parameters b, σ and base measure αP_0. Then*

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}^{\text{ind}}(A), \tilde{\boldsymbol{\mu}}^{\text{co}}(A)\big)^2 \le \gamma_{b,\sigma}\alpha P_0(A).$$

In particular, $d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}^{\text{ind}}, \tilde{\boldsymbol{\mu}}^{\text{co}})^2 \le \gamma_{b,\sigma}\alpha$.

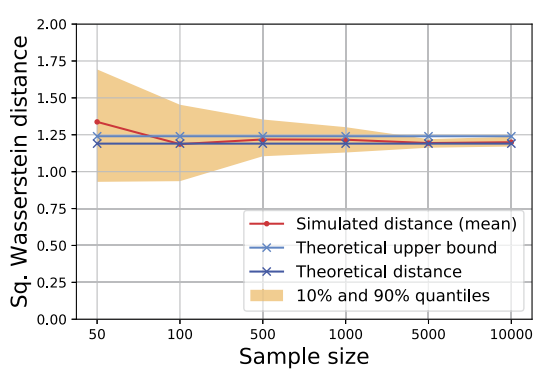

FIG. 5. *Simulations of the empirical Wasserstein distance in Figure 1 compared with the non–informative bound in Remark 5 and the informative bound in Theorem 11. As detailed in Figure 1, simulations were performed with independent samples, independent for each sample size, using the Python Optimal Transport (POT) package [14].*

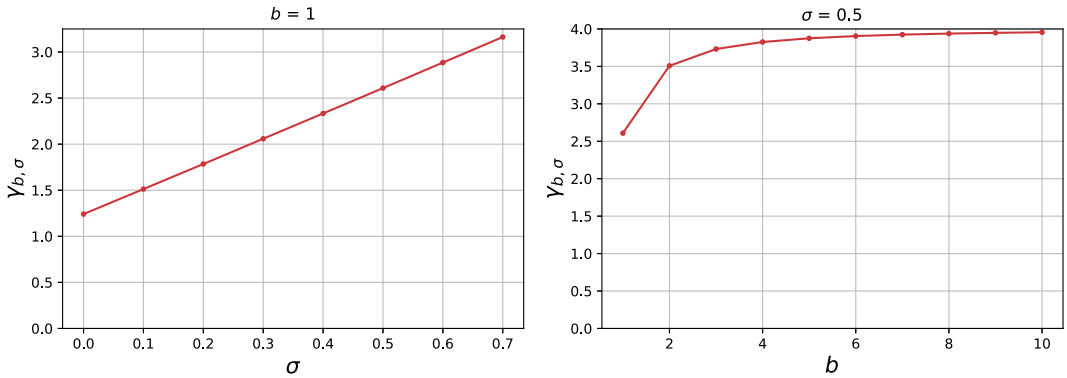

FIG. 6. *Numerical integrations of* $\gamma_{b,\sigma}$. *On the left*, $b = 1$ *and* σ *varies from* 0 *to* 0.7. *On the right*, $\sigma = 0.5$ *while* b *varies from* 1 *to* 10.

The bounds in Corollary 12 shed light on the role of the hyperparameters in the distance between the two extreme cases of independence and exchangeability. In particular, Figure 6 shows that the distance increases linearly as σ increases and logarithmically as b increases.

7. Measuring dependence in nonparametric models. We now analyze three popular procedures to model the dependence between CRMs through the choice of a hyperparameter, namely compound random measures, Clayton Lévy copula and GM-dependence. These can be seen as infinite-dimensional extensions of the *approximately* exchangeable priors suggested by de Finetti [10] for binary data, and further investigated in Bacallado, Diaconis and Holmes [1]. Our theoretical findings allow for a formal quantification of the dependence in terms of a meaningful bound on the distance from exchangeability. These bounds are expressed in terms of the models' hyperparameters leading to intuitive results, which can also guide the parameters' elicitation.

7.1. *Compound random measures.* Compound random measures, introduced in Griffin and Leisen [17], provide a general framework for building CRVs. These may be used to model the dependence between CRMs with many different marginal distributions, such as gamma, generalized gamma, beta and σ-stable random measures.

DEFINITION 5. A *compound random measure* $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2)$ is a CRV of the form

$$\begin{pmatrix} \tilde{\mu}_1 \\ \tilde{\mu}_2 \end{pmatrix} = \sum_{i=1}^{+\infty} \begin{pmatrix} m_{1,i} \\ m_{2,i} \end{pmatrix} J_i \delta_{X_i},$$

where $\tilde{\eta} = \sum_{i=1}^{+\infty} J_i \delta_{X_i}$ is a CRM with Lévy intensity $\alpha P_0(dx)\nu^*(ds)$, $(m_{1,i}, m_{2,i}) \overset{\text{iid}}{\sim} h$ and h is a bivariate density.

In [17], the authors prove that such $\tilde{\boldsymbol{\mu}}$ is a CRV with bivariate Lévy intensity

$$\nu(ds_1, ds_2, dx) = \alpha P_0(dx) \int_{\mathbb{R}^+} \frac{1}{u^2} h\left(\frac{s_1}{u}, \frac{s_2}{u}\right) \nu^*(du)\, ds_1, ds_2.$$

Specific choices for ν^* and h lead to different marginal CRMs and dependence structures. In particular, by taking h corresponding to the distribution of two independent gamma $(\phi, 1)$ random variables and $\nu^*(du) = (1-u)^{\phi-1} u^{-1} \mathbb{1}_{(0,1)}(u)\, du$, one achieves marginal gamma random measures of shape parameter 1 and base measure αP_0. We write $\tilde{\boldsymbol{\mu}} \sim \text{CoGamma}(\phi, \alpha, P_0)$. Here, we focus on the case of gamma marginal random measures,

though the techniques may be generalized. Our aim is to quantify dependence, which is controlled by the parameter ϕ. We first introduce

$$K_\phi = \int_0^{+\infty} E_1^{-1}\big(e(\phi,t)\big)\phi f(\phi, 2\phi, t)\,dt,$$

$$e(\phi,t) = \frac{1}{\Gamma(2\phi)}\int_0^1 \Gamma\left(2\phi, \frac{t}{u}\right)(1-u)^{\phi-1}u^{-1}\,du, \qquad e_{\mathbb{N}}(\phi,t) = \sum_{k=0}^{2\phi-1} f(\phi,k,t),$$

$$f(\phi,x,t) = \frac{t^x}{\Gamma(x)}\int_0^1 e^{-\frac{t}{u}}(1-u)^{\phi-1}u^{-x-1}\,du.$$

$$f_{\mathbb{N}}(\phi,n,t) = \frac{t^n}{n!}\sum_{j=0}^{\phi-1}\binom{\phi-1}{j}(-1)^j g(n,j,t),$$

where $g(n,j,t)$ is equal to

$$\begin{cases} t^{-n+j}(n-j-1)!e^{-t}\displaystyle\sum_{h=0}^{n-j-1}\frac{t^h}{h!} & \text{if } n > j, \\ \dfrac{1}{(j-n)!}\left(e^{-t}\displaystyle\sum_{j=0}^{j-n-1}(-1)^h(j-n-h-1)!t^h + (-1)^{j-n}E_1(t)\right) & \text{if } n \le j. \end{cases}$$

THEOREM 13. *Let $\tilde{\boldsymbol{\mu}} \sim \mathrm{CoGamma}(\phi, \alpha, P_0)$ and let $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ denote the comonotonic random vector in the same Fréchet class. Then*

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big)^2 \le 4\alpha P_0(A)(1-K_\phi).$$

In particular, $d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}})^2 \le 4\alpha(1-K_\phi)$. Moreover, when $\phi \in \mathbb{N}$, $e = e_{\mathbb{N}}$ and $f = f_{\mathbb{N}}$.

Theorem 13 allows us to conveniently compute the Wasserstein distance from exchangeability for ϕ an integer value. Table 1 displays some numerical results for different values of ϕ. As ϕ increases, the dependence between the induced marginal gamma random measures also increases. Moreover, we stress that the case $\phi = 1$ is of particular interest since it corresponds to the dependence structure discussed in [30].

In Figure 7 we compare the theoretical upper bounds in Theorem 13 with the simulated Wasserstein distance, as in Figure 5. As in the previous case, our upper bound appears to be tight and informative.

7.2. *Clayton Lévy copula*. Lévy copulae provide another popular way to model dependence between CRMs. Standard copulae can be seen as a means to separate the marginal components of a bivariate distribution from its dependence structure. The same happens

TABLE 1
Values of $1-K_\phi$ in Theorem 13, as ϕ varies

ϕ	$1-K_\phi$ ($\approx$)
1	0.1426
5	0.0545
10	0.0241
30	0.0081

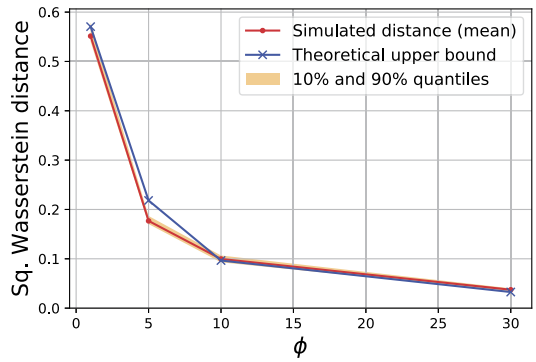

FIG. 7. *Simulation of the empirical Wasserstein distance between a random vector* $(\mu_1(\mathbb{X}), \mu_2(\mathbb{X}))$ *with marginal compound random measures of parameters* (ϕ, α, P_0), *where* $\alpha = 1$ *and* ϕ *varies, and a bivariate distribution with a.s. equal gamma marginals of shape* $=$ *scale* $= 1$. *Simulations were performed with independent samples of* 10,000 *observations using the Python Optimal Transport* (*POT*) *package* [14].

for their generalization to Lévy intensities, conceived in Tankov [50] and Cont and Tankov [6] to model the dependence structure between Lévy processes. See also [25] and [13, 29] for uses on CRMs. Given a bivariate Lévy intensity $\nu(ds_1, ds_2, A)$, we indicate by $U_{i,A}(t) = \int_t^\infty \nu_i(ds, A)$, for $i = 1, 2$, its marginal tail integrals. An analogue of Sklar's theorem states that there exists a Lévy copula $c : [0, +\infty]^2 \to [0, +\infty]$ s.t.

$$\nu\big((t_1, +\infty) \times (t_2, +\infty) \times A\big) = c\big(U_{1,A}(t_1), U_{2,A}(t_2)\big).$$

When the Lévy copula c and the tail integrals $U_{1,A}$, $U_{2,A}$ are sufficiently smooth, $\nu(ds_1, ds_2, A)$ is recovered by

$$\nu(ds_1, ds_2, A) = \frac{\partial^2}{\partial u_1\, \partial u_2} c(u_1, u_2)\bigg|_{U_{1,A}(s_1), U_{2,A}(s_2)} \nu_1(ds_1, A)\nu_2(ds_2, A). \tag{19}$$

It follows that Lévy copulae are useful to build bivariate Lévy intensities, allowing to gain insight into their dependence structure. Consider the Clayton Lévy copula, which is a smooth class of copulae with both independence and complete dependence as limiting cases:

$$c_\theta(s_1, s_2) = \big(s_1^{-\theta} + s_2^{-\theta}\big)^{-\frac{1}{\theta}},$$

for $\theta > 0$. This was used, for example, in [13, 29]. As $\theta \to +\infty$ one achieves the complete dependence copula [25] which, by taking equal marginal Lévy intensities, corresponds to the exchangeability assumption. We write $\tilde{\boldsymbol{\mu}} \sim \mathrm{Cl}(\theta, \alpha, P_0)$ for a CRV with marginal gamma random measures with base measure αP_0 and Lévy copula c_θ. Our goal is to show that, as $\theta \to +\infty$, $\tilde{\boldsymbol{\mu}}$ converges in the Wasserstein distance to the comonotonic random vector with same marginal distributions and also to provide an upper bound for the rate of convergence. Define

$$K_\theta = \frac{1+\theta}{\theta^2} \int_0^\infty \int_{y_1}^\infty E_1^{-1}\big(y_1^{-\frac{1}{\theta}}\big) E_1^{-1}\Big(\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}}\Big) y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2.$$

THEOREM 14. *Let* $\tilde{\boldsymbol{\mu}} \sim \mathrm{Cl}(\theta, \alpha, P_0)$ *and let* $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ *be in the same Fréchet class. Then*

$$d_{\mathcal{W}}\big(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}}\big)^2 \le 4\alpha(1 - K_\theta).$$

Moreover, as $\theta \to +\infty$, K_θ *goes to* 1.

7.3. *GM-dependence*. In this section we consider the model introduced in [32], which relies on CRMs whose dependence is induced by the bivariate Poisson process proposed in Griffiths and Milne [18].

DEFINITION 6. A CRV $\boldsymbol{\xi}$ is GM-dependent if

$$\begin{pmatrix}\tilde{\xi}_1\\ \tilde{\xi}_2\end{pmatrix} \stackrel{\mathrm{d}}{=} \begin{pmatrix}\tilde{\mu}_1+\tilde{\mu}_0\\ \tilde{\mu}_2+\tilde{\mu}_0\end{pmatrix}, \tag{20}$$

where $\tilde{\mu}_0$, $\tilde{\mu}_1$ and $\tilde{\mu}_2$ are three independent CRMs with Lévy intensities

$$\begin{aligned} v_1(ds,dx) &= v_2(ds,dx) = \alpha z P_0(dx)\rho(s)\,ds,\\ v_0(ds,dx) &= \alpha(1-z)P_0(dx)\rho(s)\,ds, \end{aligned}$$

where $\alpha > 0$, $z \in (0,1)$, P_0 is a probability measure on $\mathbb{R}$ and ρ is a measurable function.

Set $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}} = (\tilde{\mu}_1, \tilde{\mu}_2)$ and $\tilde{\boldsymbol{\mu}}_0^{\mathrm{co}} = (\tilde{\mu}_0, \tilde{\mu}_0)$ to underline that they are, respectively, an independent and a comonotonic CRV. While the marginal Lévy intensity $\pi(ds,dx) = \alpha P_0(dx)\rho(s)\,ds$ of the CRV $\boldsymbol{\xi}$ is available, the corresponding bivariate Lévy intensity is not known. Nonetheless, the next result provides bounds on its distance from the comonotonic and the random vector with independent marginals in the same Fréchet class, in terms of the underlying random vectors $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$, $\tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$.

PROPOSITION 15. *Let $\tilde{\boldsymbol{\xi}}$ be a* GM*-dependent CRV and let $\tilde{\boldsymbol{\xi}}^{\mathrm{co}}$ denote the comonotonic random vector in the same Fréchet class. Then*

$$\begin{aligned} d_{\mathcal{W}}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}^{\mathrm{co}}) &\le d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}, \tilde{\boldsymbol{\mu}}^{\mathrm{co}}),\\ d_{\mathcal{W}}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}^{\mathrm{ind}}) &\le d_{\mathcal{W}}(\tilde{\boldsymbol{\mu}}_0^{\mathrm{ind}}, \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}), \end{aligned}$$

where $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ is the comonotonic CRV in the same Fréchet class of $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ and $\tilde{\boldsymbol{\mu}}_0^{\mathrm{ind}}$ is the CRV with independent marginals in the same Fréchet class of $\tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$.

When the marginals are generalized gamma CRMs, the specification of the previous bounds together with Theorem 12 brings to the following. In particular, this covers the case where the marginals are gamma random measures, as in [32].

COROLLARY 16. *Let $\tilde{\boldsymbol{\xi}}$ be a* GM*-dependent CRV with marginal generalized gamma random measures with parameters b, σ and total measure α. Then*

$$d_{\mathcal{W}}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}^{\mathrm{co}})^2 \le \gamma_{b,\sigma}\alpha z, \qquad d_{\mathcal{W}}(\tilde{\boldsymbol{\xi}}, \tilde{\boldsymbol{\xi}}^{\mathrm{ind}})^2 \le \gamma_{b,\sigma}\alpha(1-z),$$

where $\gamma_{b,\sigma}$ is the constant defined in (18).

As one could expect from the construction in Definition 6, the larger the parameter z, the closer one is to the situation of independence and the farther from the one of exchangeability. Our techniques allow for the derivation of convergence rates for the approximation of exchangeability as $z \to 1$, in terms of the Wasserstein distance.

Figure 8 below shows the comparison between the simulated Wasserstein distance and our theoretical upper bound, as z increases, when the marginals are gamma CRMs ($\sigma = 0$, $b = 1$).

In this section, we have found tight upper bounds for the distance $d_{\mathcal{W}}$ from comonotonicity for notable homogeneous CRVs, leveraging on the simplifications highlighted in Remark 6:

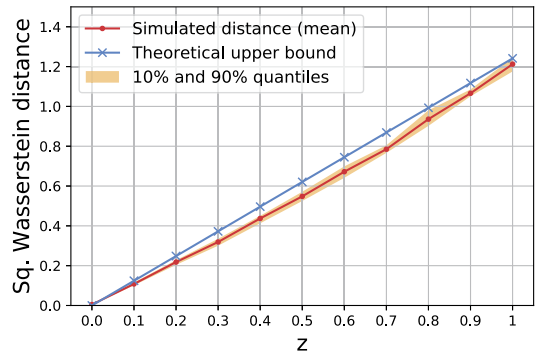

FIG. 8. *Simulation of the Wasserstein distance between a GM-dependent CRV* $(\mu_1(\mathbb{X}), \mu_2(\mathbb{X}))$ *of parameter* z *with gamma marginals of shape* $=$ *scale* $= 1$ *and a bivariate distribution with a.s. equal gamma marginals of shape* $=$ *scale* $= 1$. *Simulations were performed with independent samples of* 10,000 *observations.*

since the Lévy measure factorizes, the supremum of the Wasserstein distance over all Borel sets is always attained on the entire sample space $\mathbb{X}$. In fact, most Bayesian nonparametric models are based on homogeneous CRVs. However, studying nonhomogeneous CRVs would certainly be interesting as well, though finding the supremum could be considerably more complex.

8. Measuring dependence between random hazards. Up to now, we have investigated the dependence structure at the level of random measures, which constitute the key building block of most Bayesian nonparametric models. This has the advantage of being generic, in the sense of being independent of the particular transformation of the random measures leading to a given class of models. However, a complementary analysis tailored to such specific classes of models is also of interest. Popular transformations include normalization for modeling random probability measures [45], exponentiation to obtain a random survival function [11], simple cumulation to achieve random cumulative hazards [19], as well as kernel mixtures [12], which lead to (a.s. continuous) random hazard rates and will be the focus of this section.

For F, an absolutely continuous cumulative distribution function on $[0, +\infty)$, we recall that the hazards are defined as $h = F'/(1 - F)$ and represent the instantaneous risk of failure. Random mixture hazards are then given by $\tilde{h}(t) = \int_{\mathbb{X}} k(t|x)\, d\tilde{\mu}(x)$, with $k : \mathbb{R}^+ \times \mathbb{X} \to [0, +\infty)$ a measurable kernel and $\tilde{\mu}$ a CRM. This model was initially proposed with a specific kernel, a gamma CRM as mixing measure in Dykstra and Laud [12]. It has been further generalized to generic kernels [34] and to generic CRMs [23] and became quite popular in the survival analysis and reliability literature leading to interesting theoretical and applied contributions; see for example, [9, 22, 28, 43]. More recently, the focus has been on the construction of dependent versions of this class of models. Indeed, if $\tilde{\boldsymbol{\mu}}$ is a random vector of measures,

$$\tilde{\boldsymbol{h}}(t) = \int_{\mathbb{X}} k(t|x)\tilde{\boldsymbol{\mu}}(dx) \tag{21}$$

defines dependent hazards, which may be used as de Finetti priors for partially exchangeable sequences. Notable examples include hierarchical dependent structures [3] and GM-dependent structures [31]. The results of Section 5 and Section 7 may be adapted to quantify the dependence between the random hazards when $\tilde{\boldsymbol{\mu}}$ is a CRV. This brings to a direct measure of dependence between the de Finetti priors corresponding to different groups.

A first key result is Lemma 17 applied to the function $f(\cdot) = k(t|\cdot)$, which leads to the expression $\tilde{\boldsymbol{h}}(t) \overset{\mathrm{d}}{=} \tilde{\boldsymbol{\mu}}_t(\mathbb{X})$ for an appropriate CRV $\tilde{\boldsymbol{\mu}}_t$. Given two measure spaces $\mathbb{X}_1$ and

$\mathbb{X}_2$, we recall that if ν is a measure on $\mathbb{X}_1$ and $g : \mathbb{X}_1 \to \mathbb{X}_2$ is a measurable function, the pushforward measure $g\#\nu$ on $\mathbb{X}_2$ is defined by $(g\#\nu)(A) = \nu(g^{-1}(A))$.

LEMMA 17. *Let $\tilde{\boldsymbol{\mu}}$ be a CRV with intensity measure ν and let $f : \mathbb{X} \to \mathbb{R}^+$ be a measurable function. Then the random vector of measures $\tilde{\boldsymbol{\mu}}_f(dx) = f(x)\tilde{\boldsymbol{\mu}}(dx)$ is a CRV with Lévy intensity equal to the pushforward measure $\nu_f = p_f\#\nu$ where $p_f(s_1, s_2, x) = (s_1 f(x), s_2 f(x), x)$.*

Lemma 17 may be seen as a multivariate extension of [4], Lemma 6. In particular, we observe that the hazard rates $\tilde{\boldsymbol{h}}^{\mathrm{co}}$ induced by a comonotonic CRV $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}$ through (21) are comonotonic, that is, $\tilde{h}_1^{\mathrm{co}}(t) = \tilde{h}_2^{\mathrm{co}}(t)$ a.s. for every t. Similarly, when $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}$ is the independent CRV, the induced hazards $\tilde{\boldsymbol{h}}^{\mathrm{ind}}$ are independent. We use this observation to study the Wasserstein distance between the dependent hazards and the two extreme cases of comonotonicity and independence. Proposition 18 deals with the GM-dependent hazards of [31] when the marginals are gamma random measures and the kernel of the type of Dykstra and Laud [12], namely $k(t|x) = \beta(y)\mathbb{1}_{[0,t]}(x)$, which is a popular choice for modeling increasing hazards. For simplicity, we restrict to constant functions $\beta(s) = \beta$, which are the most common choice in applications. In such scenario, one usually considers the base measure of the gamma random measure to be equal to the Lebesgue measure on a large time interval $[0, T]$, that is, $\alpha P_0(ds) = \mathbb{1}_{[0,T]}(s)\,ds$, so that $\mathbb{X} = \mathbb{R}$. Let

$$\gamma_\beta = 4 - 4\beta \int_0^{+\infty} (E_1)^{-1}\left(2E_1\left(\frac{s}{\beta}\right)\right) e^{-\frac{s}{\beta}}\,ds. \tag{22}$$

PROPOSITION 18. *Let $\tilde{\boldsymbol{h}}$ be dependent hazards as defined in (21) s.t. $\tilde{\boldsymbol{\mu}}$ is a* GM–*dependent CRV (20) with marginal gamma CRM of base measure $\alpha P_0(ds) = \mathbb{1}_{[0,T]}(s)\,ds$ and $k(t|x) = \beta\mathbb{1}_{[0,t]}(x)$, with $\beta > 0$. If $\tilde{\boldsymbol{h}}^{\mathrm{co}}$, $\tilde{\boldsymbol{h}}^{\mathrm{ind}}$ are in the same Fréchet class as $\tilde{\boldsymbol{h}}$, for every $t \in [0, T]$,*

$$\mathcal{W}\big(\tilde{\boldsymbol{h}}(t), \tilde{\boldsymbol{h}}^{\mathrm{co}}(t)\big)^2 \le \gamma_\beta t z, \qquad \mathcal{W}\big(\tilde{\boldsymbol{h}}(t), \tilde{\boldsymbol{h}}^{\mathrm{ind}}(t)\big)^2 \le \gamma_\beta t(1-z),$$

where γ_β is the constant defined in (22).

9. Proofs.

9.1. *Background results.* We first recall some key results concerning the Wasserstein distance. See [2], Lemmas 8.6 and 8.8. If $(\boldsymbol{X}^1, \dots, \boldsymbol{X}^n)$ and $(\boldsymbol{Y}^1, \dots, \boldsymbol{Y}^n)$ are tuples of independent random vectors on $\mathbb{R}^2$, then

$$\mathcal{W}(\boldsymbol{X}^1 + \cdots + \boldsymbol{X}^n, \boldsymbol{Y}^1 + \cdots + \boldsymbol{Y}^n) \le \sum_{i=1}^n \mathcal{W}(\boldsymbol{X}^i, \boldsymbol{Y}^i). \tag{23}$$

Moreover, if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two random vectors on $\mathbb{R}^2$ with finite second moment, then

$$\mathcal{W}(\boldsymbol{X}, \boldsymbol{Y})^2 = \mathcal{W}\big(\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}), \boldsymbol{Y} - \mathbb{E}(\boldsymbol{Y})\big)^2 + \big\|\mathbb{E}(\boldsymbol{X}) - \mathbb{E}(\boldsymbol{Y})\big\|^2. \tag{24}$$

Next, if P_1, P_2, Q_1, Q_2 are probability measures, then for every $\alpha \in [0, 1]$,

$$\begin{aligned} &\mathcal{W}\big(\alpha P_1 + (1-\alpha)P_2, \alpha Q_1 + (1-\alpha)Q_2\big) \\ &\quad \le \alpha\mathcal{W}(P_1, Q_1) + (1-\alpha)\mathcal{W}(P_2, Q_2). \end{aligned} \tag{25}$$

Furthermore, we recall [47], Theorem 12, to establish the optimality of a transport map.

THEOREM 19 (Rüschendorff 1991). *If $\boldsymbol{X}$ is a random object on $\mathbb{R}^2$ and $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is continuously differentiable, then $(\boldsymbol{X}, \phi(\boldsymbol{X}))$ is an optimal coupling with respect to the 2-Wasserstein distance if and only if the following hold*:

1. *ϕ is monotone, that is, $\langle \boldsymbol{x} - \boldsymbol{y}, \phi(\boldsymbol{x}) - \phi(\boldsymbol{y})\rangle \geq 0$ for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^2$, where $\langle \cdot, \cdot \rangle$ indicates the standard scalar product on $\mathbb{R}^2$;*
2. *The matrix $D\phi = \left(\frac{\partial \phi_i}{\partial x_j}\right)_{i,j}$ is symmetric.*

9.2. *Proof of Theorem 2.* The proof of Theorem 2 is based on the following result, which will also be instrumental to further proofs. As before, F_X denotes the cdf of X.

THEOREM 20. *Let X_1, X_2, X be possibly dependent random variables whose law is a.c. w.r.t. the Lebesgue measure on $\mathbb{R}$. Then, for every continuously differentiable $g : \mathbb{R}^2 \to \mathbb{R}$, the map*

$$(x_1, x_2) \mapsto \phi_g(x_1, x_2) = \big(F_X^{-1} \circ F_{g(X_1,X_2)} \circ g(x_1, x_2), F_X^{-1} \circ F_{g(X_1,X_2)} \circ g(x_1, x_2)\big),$$

provides a transportation map between $\mathcal{L}(X_1, X_2)$ and $\mathcal{L}(X, X)$. Moreover,

$$(x_1, x_2) \mapsto \phi(x_1, x_2) = \big(F_X^{-1} \circ F_{X_1+X_2}(x_1 + x_2), F_X^{-1} \circ F_{X_1+X_2}(x_1 + x_2)\big),$$

is an optimal transport map.

PROOF. First, observe that $F_{g(X_1,X_2)} \circ g(X_1, X_2) \sim \mathrm{Unif}([0, 1])$. Since X is a.c. w.r.t., the Lebesgue measure on $\mathbb{R}$, $F_X^{-1} \circ F_{g(X_1,X_2)} \circ g(X_1 + X_2) \overset{\mathrm{d}}{=} X$. This ensures that ϕ_g is indeed a coupling between (X_1, X_2) and (X, X). In order to prove that ϕ is an optimal transport map, we refer to the sufficient conditions described in Theorem 19. Note that

$$\begin{aligned}
&\big\langle \boldsymbol{x} - \boldsymbol{y}, \phi(\boldsymbol{x}) - \phi(\boldsymbol{y})\big\rangle \\
&\quad = (x_1 - y_1 + x_2 - y_2)\big(F_X^{-1} \circ F_{X_1+X_2}(x_1 + x_2) - F_X^{-1} \circ F_{X_1+X_2}(y_1 + y_2)\big).
\end{aligned}$$

Since cdfs are nondecreasing functions, and the inverse of a nondecreasing function is nondecreasing as well, F_X^{-1} is nondecreasing. Thus $x_1 + x_2 \leq y_1 + y_2$, if and only if $F_X^{-1} \circ F_{X_1+X_2}(x_1 + x_2) \leq F_X^{-1} \circ F_{X_1+X_2}(y_1 + y_2)$. It follows that the previous expression is always nonnegative, and the monotonicity condition holds. As for the symmetry, this easily holds since the two components of ϕ are the same and are symmetric in the two arguments. □

Now consider $\tilde{\boldsymbol{\mu}}(A) = (X_1, X_2)$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A) = (X, X)$. Theorem 20 guarantees that

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big)^2 = \sum_{i=1}^{2} \mathbb{E}\big(\big|\tilde{\mu}_i(A) - F_{\tilde{\mu}_1(A)}^{-1}\big(F_{\tilde{\mu}_1(A)+\tilde{\mu}_2(A)}\big(\tilde{\mu}_1(A) + \tilde{\mu}_2(A)\big)\big)\big|\big)^2$$

and note that $F_{\tilde{\mu}_1(A)}^{-1}(F_{\tilde{\mu}_1(A)+\tilde{\mu}_2(A)}(\tilde{\mu}_1(A) + \tilde{\mu}_2(A)) \overset{\mathrm{d}}{=} \tilde{\mu}_1(A)$. Thus, we have

$$\mathcal{W}\big(\tilde{\boldsymbol{\mu}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big)^2 = 4\big(\mathbb{E}\big(\tilde{\mu}_1(A)^2\big) - \omega_{\tilde{\boldsymbol{\mu}},A}\big).$$

9.3. *Proof of Lemma 3.* Let $\tilde{\boldsymbol{\mu}}(A) = (X_1, X_2)$, so that $\omega_{\tilde{\boldsymbol{\mu}},A} = \mathbb{E}(X_1 F_{X_1}^{-1}(F_{X_1+X_2}(X_1 + X_2)))$. Since $\mathcal{L}(X_1, X_2) = \mathcal{L}(X_2, X_1)$,

$$\mathbb{E}\big(X_1 F_{X_1}^{-1}\big(F_{X_1+X_2}(X_1 + X_2)\big)\big) = \frac{1}{2}\mathbb{E}\big((X_1 + X_2) F_{X_1}^{-1}\big(F_{X_1+X_2}(X_1 + X_2)\big)\big).$$

A change of variable $U = F_{X_1+X_2}(X_1 + X_2) \sim \mathrm{Unif}([0, 1])$ leads to the conclusion.

9.4. *Proof of Corollary* 4. The proof is based on Theorem 2 and Proposition 3. First, observe that $\tilde{\mu}_1(A) \sim \text{gamma}(\alpha P_0(A), 1)$. Thus $\mathbb{E}(\tilde{\mu}_1(A)^2) = \alpha P_0(A)(1+\alpha P_0(A))$. Moreover, $\omega_{\tilde{\boldsymbol{\mu}},A}$ can be rewritten as

$$\mathbb{E}(\tilde{\mu}_1(A) S^{-1}_{\tilde{\mu}_1(A)}(S_{\tilde{\mu}_1(A)+\tilde{\mu}_2(A)}(\tilde{\mu}_1(A)+\tilde{\mu}_2(A)))),$$

where S_X denotes the survival function. Now, since $\tilde{\mu}_1(A)$ and $\tilde{\mu}_2(A)$ are independent, $\tilde{\mu}_1(A)+\tilde{\mu}_2(A) \sim \text{gamma}(2\alpha P_0(A), 1)$. Thus, we have

$$\omega_{\tilde{\boldsymbol{\mu}},A} = \int_0^{+\infty}\int_0^{+\infty} s_1 \operatorname{Inv}\Gamma_{\alpha P_0(A)}\left(\frac{\Gamma(\alpha P_0(A))}{\Gamma(2\alpha P_0(A))}\Gamma(2\alpha P_0(A), s_1+s_2)\right) \cdot \rho_{\alpha P_0(A)}(s_1)\rho_{\alpha P_0(A)}(s_2)\, ds_1\, ds_2$$

with ρ_ϕ the density function of a gamma(ϕ,1). With a change of variables $(t_1, t_2) = (s_1, s_1 + s_2)$, this is equal to

$$\int_0^{+\infty} \operatorname{Inv}\Gamma_{\alpha P_0(A)}\left(\frac{\Gamma(\alpha P_0(A))}{\Gamma(2\alpha P_0(A))}\Gamma(2\alpha P_0(A), t_2)\right) \cdot \int_0^{t_2} t_1 \rho_{\alpha P_0(A)}(t_1)\rho_{\alpha P_0(A)}(t_2-t_1)\, dt_1\, dt_2.$$

Now, $t_1\rho_{\alpha P_0(A)}(t_1) = \alpha P_0(A)\rho_{\alpha P_0(A)+1}(t_1)$, so that

$$\int_0^{t_2} t_1 \rho_{\alpha P_0(A)}(t_1)\rho_{\alpha P_0(A)}(t_2-t_1)\, dt_1\, dt_2$$

is proportional to the convolution between two gamma random variables with parameters, respectively, $(\alpha P_0(A)+1,1)$ and $(\alpha P_0(A),1)$, evaluated in t_2. This corresponds to the density of a gamma($2\alpha P_0(A)+1$,1) random variable evaluated in t_2. Thus $\omega_{\tilde{\boldsymbol{\mu}},A}$ is equal to

$$\frac{\alpha P_0(A)}{\Gamma(2\alpha P_0(A)+1)}\int_0^{+\infty} \operatorname{Inv}\Gamma_{\alpha P_0(A)}\left(\frac{\Gamma(\alpha P_0(A))}{\Gamma(2\alpha P_0(A))}\Gamma(2\alpha P_0(A), t)\right) e^{-t} t^{2\alpha P_0(A)}\, dt.$$

The alternative expression for $\omega_{\alpha,P_0,A}$ follows similarly from Proposition 3.

9.5. *Proof of Theorem* 5. We show that for every real sequence $\{r_n \mid n \in \mathbb{N}\}$ s.t. $\lim_{n\to+\infty} r_n = +\infty$,

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A)) \le \lim_{n\to+\infty} \sqrt{r_n}\,\mathcal{W}(\rho^1_{r_n,A,D_1}, \rho^2_{r_n,A,D_2}). \tag{26}$$

Since both complementary families D_1, D_2 have continuously increasing mass, there exists n_0 s.t. for every $n > n_0$ there exist $\epsilon^1_{n,A}, \epsilon^2_{n,A} > 0$ s.t.,

$$r_n = \nu^1(D(\epsilon^1_{n,A}) \times A) = \nu^2(D(\epsilon^2_{n,A}) \times A). \tag{27}$$

Before moving to the core of the proof, we show that

$$\lim_{n\to+\infty} \epsilon^i_{n,A} = 0. \tag{28}$$

We prove this by contradiction. Supposing (28) does not hold, there must be a subsequence $\{\epsilon^i_{h_n,A}\}$ converging to a (possibly infinite) limit $\epsilon^i_* \neq 0$. Since $\lim_{n\to+\infty} r_n = +\infty$, also $\lim_{n\to+\infty} r_{h_n} = +\infty$. Then there is at least one increasing subsequence $\{r_{k_n} \mid n \in \mathbb{N}\} \subset \{r_{h_n} \mid n \in \mathbb{N}\}$ s.t. $\lim_{n\to+\infty} \epsilon^i_{k_n,A} = \epsilon^i_*$ and $\lim_{n\to+\infty} r_{k_n} = +\infty$.

Since D is increasing and ν is monotone, $r_{k_n} \le r_{k_{n+1}}$ implies $D(\epsilon^i_{k_n,A}) \subset D(\epsilon^i_{k_{n+1},A})$. Thus by the monotone convergence theorem,

$$+\infty = \lim_{n\to+\infty} \nu^i(D(\epsilon^i_{k_n,A}) \times A) = \nu^i(D(\epsilon^i_*) \times A).$$

Given the Lévy intensity is finite outside of the origin by (3), $\nu^i(D(\epsilon^i_*) \times A) < +\infty$, which is a contradiction. Thus (28) holds.

Now recall that by (2) there exist Poisson random measures $\mathcal{N}^i$ s.t. for every $A \in \mathcal{X}$, $\tilde{\boldsymbol{\mu}}^i(A) = \int_{\mathbb{R}^2_+ \times A} \boldsymbol{s}\mathcal{N}^i(ds_1, ds_2, dx)$, for $i = 1, 2$. Since the evaluations of Poisson random measures on disjoint sets are independent, by (23) for every $n > 0$,

$$\mathcal{W}(\tilde{\boldsymbol{\mu}}^1(A), \tilde{\boldsymbol{\mu}}^2(A))$$

$$\leq \mathcal{W}\Big(\int_{B_1(\epsilon^1_{n,A}) \times A} \boldsymbol{s}\mathcal{N}^1(ds_1, ds_2, dx), \int_{B_2(\epsilon^2_{n,A}) \times A} \boldsymbol{s}\mathcal{N}^2(ds_1, ds_2, dx)\Big) \tag{29}$$

$$+ \mathcal{W}\Big(\int_{D_1(\epsilon^1_{n,A}) \times A} \boldsymbol{s}\mathcal{N}^1(ds_1, ds_2, dx), \int_{D_2(\epsilon^2_{n,A}) \times A} \boldsymbol{s}\mathcal{N}^2(ds_1, ds_2, dx)\Big). \tag{30}$$

We prove that the first summand (29) goes to zero as $n \to +\infty$. By bounding the Wasserstein distance with the second moments as in (5) and using the properties of Poisson random measures, (29) is bounded from above by

$$\Big(2 \sum_{\substack{i=1,2\\ j=1,2}} \int_{B_i(\epsilon^i_{n,A})} s_j^2 \nu^i(ds_1, ds_2, A) + \Big(\int_{B_i(\epsilon^i_{n,A})} s_j \nu^i(ds_1, ds_2, A)\Big)^2\Big)^{\frac{1}{2}}.$$

Thanks to the finiteness of the integrals in (6) and (7), we may apply the dominated convergence theorem and bring the limit as $n \to +\infty$ inside both integrals. In order to prove that the above expression goes to zero, we thus need to show that

$$\int_{\mathbb{R}^2_+} \mathbb{1}_{\bigcap_{n\in\mathbb{N}} B_i(\epsilon^i_{n,A})}(s_1, s_2) s_j^k \nu^i(ds_1, ds_2, A) = 0,$$

where $i, j, k = 1, 2$. By absolute continuity of the integral, it suffices to show that $\nu^i(\bigcap_{n\in\mathbb{N}} B_i(\epsilon^i_{n,A}) \times A) = 0$. Now, by assumptions on the family B, we know that $\nu^i(\bigcap_{\epsilon\in(0,1]} B_i(\epsilon) \times A) = 0$. We then prove that

$$\nu^i\big(\bigcap_{n\in\mathbb{N}} B_i(\epsilon^i_{n,A}) \times A\big) \leq \nu^i\big(\bigcap_{\epsilon\in(0,1]} B_i(\epsilon) \times A\big) = 0, \tag{31}$$

by showing that $\bigcap_{n\in\mathbb{N}} B_i(\epsilon^i_{n,A}) \subset \bigcap_{\epsilon\in(0,1]} B_i(\epsilon)$. Let $x \in \bigcap_{n\in\mathbb{N}} B_i(\epsilon^i_{n,A})$. Since $\lim_{n\to+\infty} \epsilon^i_{n,A} = 0$ by (28), for every $\epsilon \in (0, 1]$ there exists n s.t. $\epsilon^i_{n,A} < \epsilon$. Since B_i is an increasing family, $x \in B_i(\epsilon^i_{n,A}) \subset B_i(\epsilon)$. Thus $x \in \bigcap_{\epsilon>0} B_i(\epsilon)$.

As for the second summand (30), since the Lévy intensities are bounded outside of the origin by (3), $\mathbb{1}_{D_i(\epsilon^i_{n,A})}(\boldsymbol{s})\mathcal{N}^i(ds_1, ds_2, dy)$ is a Poisson random measure with finite mean for $i = 1, 2$. Thus by [48], Proposition 19.5, their integrals have a compound Poisson distribution on $\mathbb{R}^2$ with intensity measure $\int_A \mathbb{1}_{D_i(\epsilon^i_{n,A})}(\boldsymbol{s})\nu^i(ds_1, ds_2, dy)$ and same total measure r_n. Hence we have

$$\int_{D_i(\epsilon^i_{n,A}) \times A} \boldsymbol{s}\mathcal{N}^i(ds_1, ds_2, dx) \overset{\mathrm{d}}{=} \sum_{j=1}^{N^i} \boldsymbol{X}^i_j,$$

where N^i has a Poisson distribution with mean r_n and is independent of $\{\boldsymbol{X}^i_j \mid j \geq 1\}$, which are i.i.d. random variables with distribution $\rho^i_{r_n,A,D_i}$. Proposition 6 thus entails

$$\mathcal{W}\Big(\sum_{j=1}^{N^1} \boldsymbol{X}^1_j, \sum_{j=1}^{N^2} \boldsymbol{X}^2_j\Big) \leq \sqrt{r_n}\mathcal{W}(\rho^1_{r_n,A,D_1}, \rho^2_{r_n,A,D_2}) + (r_n^2 + r_n)\|\mathbb{E}(\boldsymbol{X}^1_1) - \mathbb{E}(\boldsymbol{X}^2_1)\|^2.$$

Now, $(r_n^2 + r_n)\|\mathbb{E}(\boldsymbol{X}_1^1) - \mathbb{E}(\boldsymbol{X}_1^2)\|^2$ is equal to

$$\left(1 + \frac{1}{r_n}\right) \sum_{i=1,2} \left| \int_{D_1(\epsilon_{n,A}^1)} s_i \nu^1(ds_1, ds_2, A) - \int_{D_2(\epsilon_{n,A}^2)} s_i \nu^2(ds_1, ds_2, A) \right|^2,$$

which as $n \to +\infty$ by the monotone convergence theorem converges to

$$\sum_{i=1,2} \left| \int_{\mathbb{R}_+^2} s_i \nu^1(ds_1, ds_2, A) - \int_{\mathbb{R}_+^2} s_i \nu^2(ds_1, ds_2, A) \right|^2. \tag{32}$$

Since the vectors are in the same Fréchet class, (32) is equal to 0. The bound in (26), hence, follows by taking the limit as n goes to $+\infty$. In order to prove its finiteness, it suffices to observe that by (5), $\sqrt{r_n}\mathcal{W}(\rho_{r_n,A,D_1}^1, \rho_{r_n,A,D_2}^2)$ is bounded from above by the square root of

$$2\sum_{i=1}^{2} \int_{\mathbb{R}_+^2} (s_1^2 + s_2^2)\nu^i(ds_1, ds_2, A) + \left(\int_{\mathbb{R}_+^2} (s_1 + s_2)\nu^i(ds_1, ds_2, A) \right)^2,$$

which is finite by (6) and (7).

We now show that the limit does not depend on the choice of compatible families D_1 and D_2. First, we prove that, given a bivariate Lévy intensity ν with compatible families D and D^*, one has

$$\lim_{n \to +\infty} \sqrt{r_n}\mathcal{W}(\rho_{r_n,A,D}, \rho_{r_n,A,D^*}) = 0. \tag{33}$$

For every n, consider $\epsilon_{n,A}$ and $\epsilon_{n,A}^*$ as in (27). Let then $\Omega(n) = D(\epsilon_{n,A}) \cap D^*(\epsilon_{n,A}^*)$ and denote by $q_n = \nu(\Omega(n) \times A)$. We define

$$P_n^0(ds_1, ds_2) = \frac{1}{q_n} \mathbb{1}_{\Omega(n)}(s_1, s_2)\nu(ds_1, ds_1, A),$$

$$P_n(ds_1, ds_2) = \frac{1}{r_n - q_n} \mathbb{1}_{D(\epsilon_{n,A})\setminus\Omega(n)}(s_1, s_2)\nu(ds_1, ds_1, A),$$

$$P_n^*(ds_1, ds_2) = \frac{1}{r_n - q_n} \mathbb{1}_{D^*(\epsilon_{n,A}^*)\setminus\Omega(n)}(s_1, s_2)\nu(ds_1, ds_1, A)$$

and consider the decompositions

$$\rho_{r_n,A,D} = \frac{q_n}{r_n} P_n^0 + \frac{r_n - q_n}{r_n} P_n, \qquad \rho_{r_n,A,D^*} = \frac{q_n}{r_n} P_n^0 + \frac{r_n - q_n}{r_n} P_n^*.$$

By the convexity property in (25), since P_n^0 is a shared component,

$$\mathcal{W}(\rho_{r_n,A,D}, \rho_{r_n,A,D^*}) \le \frac{r_n - q_n}{r_n} \mathcal{W}(P_n, P_n^*).$$

Hence by (5), $\sqrt{r_n}\mathcal{W}(\rho_{r_n,A,D}, \rho_{r_n,A,D^*})$ is bounded from above by the squared root of

$$\frac{r_n - q_n}{r_n} 4\left(\int_{\mathbb{R}_+^2} (s_1^2 + s_2^2)\nu(ds_1, ds_2, A) + \left(\int_{\mathbb{R}_+^2} (s_1 + s_2)\nu(ds_1, ds_2, A) \right)^2 \right).$$

Since $D(\epsilon_{n,A}) \setminus \Omega(n) \subset D^*(\epsilon_{n,A}^*)^c$, $r_n - q_n = \nu(D(\epsilon_{n,A}) \setminus \Omega(n) \times A) \le \nu(D^*(\epsilon_{n,A}^*)^c \times A)$. Thus, by reasoning as in (31),

$$\limsup_{n \to +\infty} r_n - q_n \le \limsup_{n \to +\infty} \nu\big(D^*(\epsilon_{n,A}^*)^c \times A\big) = \nu\big(\bigcap_{n \in \mathbb{N}} B^*(\epsilon_{n,A}^*) \times A\big) = 0.$$

Hence, $\lim_{n \to +\infty} r_n - q_n = 0$ and we conclude that $\lim_{r_n \to +\infty} \sqrt{r_n}\mathcal{W}(\rho_{r_n,A,D}, \rho_{r_n,A,D^*}) = 0$.

Now, consider two compatible families D_1^* and D_2^* for ν^1 and ν^2, respectively. By the triangular inequality, $\mathcal{W}(\rho^1_{r_n,A,D_1^*}, \rho^2_{r_n,A,D_2^*})$ is bounded from above by

$$\mathcal{W}(\rho^1_{r_n,A,D_1^*}, \rho^1_{r_n,A,D_1}) + \mathcal{W}(\rho^1_{r_n,A,D_1}, \rho^2_{r_n,A,D_2}) + \mathcal{W}(\rho^2_{r_n,A,D_2}, \rho^2_{r_n,A,D_2^*}).$$

Then, thanks to (33) by taking the limit as $n \to +\infty$,

$$\lim_{n\to+\infty} \sqrt{r_n}\mathcal{W}(\rho^1_{r_n,A,D_1^*}, \rho^2_{r_n,A,D_2^*}) \le \lim_{n\to+\infty} \sqrt{r_n}\mathcal{W}(\rho^1_{r_n,A,D_1}, \rho^2_{r_n,A,D_2}).$$

Equality follows by changing the role of (D_1, D_2) and (D_1^*, D_2^*) in the previous argument.

9.6. *Proof of Proposition* 6. We rely on the key identity (24). First, we observe that $\mathbb{E}(\boldsymbol{X}) = r\mathbb{E}(\boldsymbol{X}_1)$ and $\mathbb{E}(\boldsymbol{Y}) = r\mathbb{E}(\boldsymbol{Y}_1)$. By considering the couplings s.t. $N_x = N_Y$ a.s.,

$$\begin{aligned}
&\mathcal{W}(\boldsymbol{X} - r\mathbb{E}(\boldsymbol{X}^1), \boldsymbol{Y} - r\mathbb{E}(\boldsymbol{Y}^1))^2 \\
&\quad\le \inf_{C((\boldsymbol{X}^i)_{i\ge1},(\boldsymbol{Y}^i)_{i\ge1})} \mathbb{E}\left(\left\|\sum_{i=1}^{N_x} \boldsymbol{X}^i - r\mathbb{E}(\boldsymbol{X}^1) - \sum_{i=1}^{N_x} \boldsymbol{Y}^i + r\mathbb{E}(\boldsymbol{Y}^1)\right\|^2\right) \\
&\quad= \inf_{C((\boldsymbol{X}^i)_{i\ge1},(\boldsymbol{Y}^i)_{i\ge1})} \mathbb{E}\left(\operatorname{Var}\left(\sum_{i=1}^{N_x} X_1^i - r\mathbb{E}(X_1^1) - \sum_{i=1}^{N_x} Y_1^i - r\mathbb{E}(Y_1^1) \,\middle|\, N_x\right)\right) \\
&\qquad+ \mathbb{E}\left(\operatorname{Var}\left(\sum_{i=1}^{N_x} X_2^i - r\mathbb{E}(X_2^1) - \sum_{i=1}^{N_x} Y_2^i - r\mathbb{E}(Y_2^1) \,\middle|\, N_x\right)\right).
\end{aligned}$$

By considering couplings s.t. $(\boldsymbol{X}^i - \boldsymbol{Y}^i)_{i\ge1}$ are independent and identically distributed,

$$\begin{aligned}
&\le \inf_{C(\boldsymbol{X}^1,\boldsymbol{Y}^1)} \mathbb{E}(N_x \operatorname{Var}(X_1^1 - Y_1^1)) + \mathbb{E}(N_x \operatorname{Var}(X_2^1 - Y_2^1)) \\
&= r\mathcal{W}(\boldsymbol{X}^1, \boldsymbol{X}^2)^2 - r\|\mathbb{E}(\boldsymbol{X}^1 - \boldsymbol{Y}^1)\|^2.
\end{aligned}$$

Finally, by applying (24) we conclude the proof.

9.7. *Proof of Proposition* 7. A CRV $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2)$ is comonotonic if $\tilde{\mu}_1 = \tilde{\mu}_2$ a.s. By uniqueness of the Lévy intensity is suffices to show that ν^{co} induces exchangeability. Consider the set $D = \{(s_1, s_2) \mid (s_1, s_2) \in \mathbb{R}_+^2, s_1 \ne s_2\}$. By definition of Poisson random measure, for every $A \in \mathcal{X}$,

$$\mathbb{P}(\mathcal{N}(D \times A) = 0) = \exp\{-\nu^{\mathrm{co}}(D \times A)\} = 1.$$

Thus with probability 1,

$$\tilde{\mu}_1(A) = \int_{\mathbb{R}_+^2 \times A} s_1 \mathcal{N}(ds_1, ds_2, dx) = \int_{\mathbb{R}_+^2 \times A} s_2 \mathcal{N}(ds_1, ds_2, dx) = \tilde{\mu}_2(A).$$

9.8. *Proof of Theorem* 8. Let $(X_1, X_2) \sim \rho_{r,A,D}$ and $(X, X) \sim \rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}}$. For every continuously differentiable function g, we define

$$K^g_{r,\nu,A,D,D^{\mathrm{co}}} = \sum_{i=1}^{2} \int_{\mathbb{R}_+^2} \left|s_i - F_X^{-1} \circ F_{g(X_1,X_2)} \circ g(s_1, s_2)\right|^2 \rho_{r,A}(ds_1\, ds_2).$$

Theorem 20 guarantees that $\mathcal{W}(\rho_{r,A,D}, \rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}})^2 \le K^g_{r,\nu,A,D,D^{\mathrm{co}}}$, and the equality holds for $g(s_1, s_2) = s_1 + s_2$. In order to find the limit of $rK^g_{r,\nu,A,D,D^{\mathrm{co}}}$ as $r \to +\infty$, we must establish the conditions for the monotone convergence theorem. Since by Theorem 5 the limit does not

depend on the compatible families D and D^{co}, we choose $D = D^g = (B^g)^c$ defined in (10), and $D^{\mathrm{co}} = D^+$ as in (1) of Figure 2. First, rewrite the bound as

$$\sum_{i=1}^{2} \int_{\mathbb{R}_+^2} \left|s_i - S_X^{-1} \circ S_{g(X_1,X_2)}\big(g(s_1,s_2)\big)\right|^2 \mathbb{1}_{(\epsilon_{r,A},+\infty)}\big(g(s_1,s_2)\big)\nu(ds_1,ds_2,A),$$

where S_X is the survival function of X. The choice $D^{\mathrm{co}} = D^+$ guarantees that $S_X(t) = \frac{1}{r}U_A^\pi(t)\mathbb{1}_{(\epsilon/2,+\infty)}(t)$; see Figure 3. Thus $\forall s \in (0,1]$, $S_X^{-1}(s) = (U_A^\pi)^{-1}(rs)$. On the other hand, $S_{g(X_1,X_2)}(t) = r^{-1}h^g_{r,\nu,A}(t)$, where

$$h^g_{r,\nu,A}(t) = \int_{\mathbb{R}_+^2} \mathbb{1}_{(t,+\infty)}\big(g(t_1,t_2)\big)\mathbb{1}_{(\epsilon_{r,A},+\infty)}\big(g(t_1,t_2)\big)\nu(dt_1,dt_2,A).$$

Thus rK^g_{r,ν,A,D^g,D^+} is equal to

$$\sum_{i=1}^{2} \int_{\mathbb{R}_+^2} \left|s_i - \big(U_A^\pi\big)^{-1}\big(h^g_{r,\nu,A}\big(g(s_1,s_2)\big)\big)\right|^2 \mathbb{1}_{(\epsilon_{r,A},+\infty)}\big(g(s_1,s_2)\big)\nu(ds_1,ds_2,A).$$

Since for every (s_1,s_2) in the domain of integration $g(s_1,s_2) > \epsilon_{r,A}$, every (t_1,t_2) s.t. $g(t_1,t_2) > g(s_1,s_2)$ satisfies $g(t_1,t_2) > \epsilon_{r,A}$. Thus for every (s_1,s_2) in the domain of integration,

$$h^g_{r,\nu,A}\big(g(s_1,s_2)\big) = \int_{\mathbb{R}_+^2} \mathbb{1}_{(g(s_1,s_2),+\infty)}\big(g(t_1,t_2)\big)\nu(dt_1,dt_2,A) = h^g_{\nu,A}\big(g(s_1,s_2)\big),$$

where $h^g_{\nu,A}$ is defined in (12). The statement in (14) follows by the monotone convergence theorem as $r \to +\infty$.

9.9. *Proof of Theorem* 9. We first provide a preliminary result, whose proof we report because it does not seem to be readily available in the literature.

LEMMA 21. *Let $f : \mathbb{R}_+ \to \mathbb{R}_+$ be an integrable nonincreasing function and $f^{-1}(x) = \sup\{t \mid f(t) \le x\}$ its generalized inverse. Then*

$$\int_0^{+\infty} f(x)\,dx = \int_0^{+\infty} f^{-1}(z)\,dz.$$

PROOF. Consider the change of variable $z = f(x)$. Since f is integrable, $\lim_{x\to+\infty} f(x) = 0$. Moreover, since f is monotone its derivative is well defined almost everywhere. Thus

$$\int_0^{+\infty} f(x)\,dx = -\int_0^{f(0)} z\frac{1}{f'(f^{-1}(z))}\,dz = -\int_0^{+\infty} z\big(f^{-1}\big)'(z)\,dz,$$

having set $f(0) = \lim_{x\to 0^+} f(x) \in [0,+\infty]$. By integration by parts, this is equal to

$$-zf^{-1}(z)|_0^{f(0)} + \int_0^{f(0)} f^{-1}(z)\,dz.$$

If $f(0) < +\infty$, the first summand is clearly 0. Otherwise, we observe that

$$-zf^{-1}(z)|_0^{+\infty} = xf(x)|_0^{+\infty} = 0,$$

because of the integrability assumption. Thus in either case,

$$\int_0^{+\infty} f(x)\,dx = \int_0^{f(0)} f^{-1}(z)\,dz = \int_0^{+\infty} f^{-1}(z)\,dz,$$

since if $f(0) < +\infty$, f^{-1} is equal to zero on the interval $(f(0),+\infty)$. $\square$

Theorem 8 ensures that the limit, as $r \to +\infty$, of $r\mathcal{W}(\rho_{r,A,D}, \rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}})$ is equal to

$$2\int_{\mathbb{R}_+^2}\left|s_1 - (U_A^\pi)^{-1}\left(\int_{\mathbb{R}_+^2} \mathbb{1}_{(s_1+s_2,+\infty)}(t_1+t_2)\nu(t_1,t_2)\,dt_1\,dt_2\right)\right|^2 \nu(s_1,s_2,A)\,ds_1\,ds_2.$$

By expanding the square of the binomial, the integral is divided in three summands. We treat them separately. First,

$$\int_{\mathbb{R}_+^2} s_1^2\nu(s_1,s_2,A)\,ds_1\,ds_2 = \int_{\mathbb{R}_+} s_1^2\pi(s_1,A)\,ds_1.$$

Next, with a change of variable $(z_1,z_2) = (s_1, s_1+s_2)$,

$$\begin{aligned}\int_{\mathbb{R}_+^2} &(U_A^\pi)^{-1}\left(\int_{\{t_1+t_2>s_1+s_2\}} \nu(t_1,t_2,A)\,dt_1\,dt_2\right)^2 \nu(s_1,s_2,A)\,ds_1\,ds_2 \\ &= \int_0^{+\infty}(U_A^\pi)^{-1}\left(\int_{\{t_1+t_2>z_2\}} \nu(t_1,t_2,A)\,dt_1\,dt_2\right)^2\left(\int_0^{z_2}\nu(z_1,z_2-z_1,A)\,dz_1\right)dz_2.\end{aligned} \tag{34}$$

Simple calculations on the derivative of an integral lead to

$$\frac{d}{dz}\int_{\{t_1+t_2>z\}}\nu(t_1,t_2,A)\,dt_1\,dt_2 = \int_0^z \nu(t_1,z-t_1,A)\,dt_1.$$

Thus with a change of variable $s = \int_{\{t_1+t_2>z_2\}}\nu(t_1,t_2,A)\,dt_1\,dt_2$, the integral in (34) is equal to $\int_0^{+\infty}(U^\pi)^{-1}(s)^2\,ds$. The function $U_A^\pi(\sqrt{s})$ is nondecreasing and has inverse $|(U_A^\pi)^{-1}(s)|^2$. By applying Lemma 21, its integral on $(0,+\infty)$ is equal to

$$\int_0^{+\infty} U_A^\pi(\sqrt{s})\,ds = \int_0^{+\infty}\int_{\sqrt{s}}^{+\infty}\pi(dt,A)\,ds = \int_0^{+\infty} t^2\pi(dt,A).$$

Finally, the expression of the third summand, which is equal to $K_{\nu,A}$ in the statement, derives from the same change of variables.

9.10. *Proof of Theorem* 10. The proof is similar to the one of Theorem 9. By looking at the support of the Lévy intensity in Figure 4, Theorem 8 ensures that the limit as $r \to +\infty$ of $r\mathcal{W}(\rho_{r,A,D}, \rho^{\mathrm{co}}_{r,A,D^{\mathrm{co}}})$ is equal to

$$2\int_{\mathbb{R}_+^2}\left|s_1 - (U_A^\pi)^{-1}(2U_A^\pi(s_1+s_2))\right|^2\nu(s_1,s_2,A)\,ds_1\,ds_2.$$

As in the previous proof, the integral is divided in three summands, which we treat separately.

$$\int_{\mathbb{R}_+^2} s_1^2\nu(s_1,s_2,A)\,ds_1\,ds_2 = \int_{\mathbb{R}_+} s_1^2\pi(s_1,A)\,ds_1.$$

Next, by looking at the support of the Lévy intensity,

$$\begin{aligned}\int_{\mathbb{R}_+^2} &(U_A^\pi)^{-1}(2U_A^\pi(s_1+s_2))^2\nu(s_1,s_2,A)\,ds_1\,ds_2 \\ &= 2\int_0^\infty (U_A^\pi)^{-1}(2U_A^\pi(s))^2\pi(s,A)\,ds.\end{aligned}$$

Since $\frac{d}{ds}U_A^\pi(s) = -\pi(s,A)$, with a change of variable $s = 2U_A^\pi(s)$, it is equal to $\int_0^{+\infty}(U^\pi)^{-1}(s)^2\,ds$. By reasoning as in Theorem 9, this is equal to $\int_{\mathbb{R}_+} s_1^2\pi(s_1,A)\,ds_1$ as well. Finally, since the integrand is equal to zero on the vertical axis, we have

$$\begin{aligned}\int_{\mathbb{R}_+^2} &s_1(U_A^\pi)^{-1}(2U_A^\pi(s_1+s_2))\nu(s_1,s_2,A)\,ds_1\,ds_2 \\ &= \int_0^\infty s_1(U_A^\pi)^{-1}(2U_A^\pi(s_1))\pi(s_1,A)\,ds_1.\end{aligned}$$

9.11. *Proof of Theorem* 13. The proof is based on Theorem 9. Since the Lévy intensities are homogeneous, we apply (16). The marginals are gamma random measures of shape parameter 1, thus $U^{\pi}(t) = E_1(t)$ and

$$\int_0^{+\infty} s^2 \pi(s)\, ds = \int_0^{+\infty} s e^{-s}\, ds = 1.$$

As for the other quantities appearing in (16), we observe that if ρ_ϕ is the density of a gamma$(\phi, 1)$ distribution,

$$\begin{aligned}
&\int_{\mathbb{R}_+^2} \mathbb{1}_{(t,+\infty)}(z_1+z_2)\nu(z_1,z_2)\, dz_1\, dz_2 \\
&\quad = \int_0^1 \left(\int_{\mathbb{R}_+^2} \mathbb{1}_{(t,+\infty)}(z_1+z_2)\rho_\phi\left(\frac{z_1}{u}\right)\rho_\phi\left(\frac{z_2}{u}\right)\frac{1}{u^2}\, dz_1\, dz_2 \right) \frac{(1-u)^{\phi-1}}{u}\, du.
\end{aligned}$$

With a change of variables $(v_1, v_2) = (\frac{z_1}{u}, \frac{z_2}{u})$,

$$\begin{aligned}
&= \int_0^1 \left(\int_{\mathbb{R}_+^2} \mathbb{1}_{(\frac{t}{u},+\infty)}(v_1+v_2)\rho_\phi(v_1)\rho_\phi(v_2)\, dv_1\, dv_2 \right) \frac{(1-u)^{\phi-1}}{u}\, du \\
&= \int_0^1 \mathbb{P}\left\{ X_1 + X_2 > \frac{t}{u} \right\} \frac{(1-u)^{\phi-1}}{u}\, du,
\end{aligned}$$

where $X_1, X_2 \overset{\text{iid}}{\sim}$ gamma$(\phi, 1)$ random variables. Thus $X_1 + X_2 \sim$ gamma$(2\phi, 1)$ and its survival function in $\frac{t}{u}$ is equal to $\frac{\Gamma(2\phi, \frac{t}{u})}{\Gamma(2\phi)}$. Next, we observe that

$$\begin{aligned}
&\int_0^t s\nu(s, t-s)\, ds \\
&\quad = \int_0^1 \frac{(1-u)^{\phi-1}}{u^3} \left(\int_0^t s\rho_\phi\left(\frac{s}{u}\right)\rho_\phi\left(\frac{t-s}{u}\right) ds \right) du.
\end{aligned}$$

With a change of variable $v = \frac{s}{u}$,

$$= \int_0^1 \frac{(1-u)^{\phi-1}}{u} \left(\int_0^{\frac{t}{u}} v\rho_\phi(v)\rho_\phi\left(\frac{t}{u} - v\right) ds \right) du.$$

Now, $v\rho_\phi(v) = \phi\rho_{\phi+1}(v)$. Thus the inner integral is ϕ times the convolution of ρ_ϕ and $\rho_{\phi+1}$ evaluated in $\frac{t}{u}$. Now, if $X \sim$ gamma$(\phi, 1)$ is independent from $Y \sim$ gamma$(\phi+1, 1)$, $X + Y \sim$ gamma$(2\phi+1, 1)$. Thus

$$\int_0^{\frac{t}{u}} v\rho_\phi(v)\rho_\phi\left(\frac{t}{u} - v\right) ds = \frac{\phi}{\Gamma(2\phi+1)} e^{-\frac{t}{u}} \left(\frac{t}{u}\right)^{2\phi},$$

from which the final expression for the integral easily follows. We now sketch the proof for ϕ integer:

$$\Gamma\left(2\phi, \frac{t}{u}\right) = (2\phi-1)! e^{-\frac{t}{u}} \sum_{k=0}^{2\phi-1} \frac{1}{k!}\left(\frac{t}{u}\right)^k.$$

Thus $e(\phi, t)$ is equal to

$$\sum_{k=0}^{2\phi-1} \frac{t^k}{k!} \int_0^1 e^{-\frac{t}{u}} (1-u)^{\phi-1} u^{-k-1}\, du,$$

which coincides with a sum over k of $f(\phi, k, t)$. In order to derive the expression for $f(\phi, n, t)$ for a generic integer n, we apply the binomial formula

$$(1-u)^{\phi-1} = \sum_{j=0}^{\phi-1} \binom{\phi-1}{j} (-u)^j,$$

from which we easily derive

$$g(n, j, t) = \int_0^1 e^{-\frac{t}{u}} u^{-n-1+j}\, du = t^{-n+j}\Gamma(n-j, t).$$

The final expression derives from writing $\Gamma(k, t)$ as a sum, both when k is a positive integer and when it is a negative one.

9.12. *Proof of Theorem* 14. By resorting to (19), one derives the expression for $\nu(ds_1, ds_2, A)$:

$$\alpha P_0(A)(1+\theta)\big(E_1(s_1)^{-\theta} + E_1(s_2)^{-\theta}\big)^{-\frac{1}{\theta}-2} E_1(s_1)^{-\theta-1} E_1(s_2)^{-\theta-1} \frac{e^{-(s_1+s_2)}}{s_1 s_2}\, ds_1\, ds_2.$$

We obtain the bound by applying Theorem 8 to the function

$$g(s_1, s_2) = E_1(s_1)^{-\theta} + E_1(s_2)^{-\theta},$$

which trivially satisfies the necessary conditions. Since g is symmetric, $K^g_{\nu,A}$ is equal to

$$2\alpha \int_{\mathbb{R}_+^2} \left| s_1 - (U^\pi)^{-1}\left(\int_{\mathbb{R}_+^2} \mathbb{1}_{[g(s_1,s_2),+\infty)}\big(g(t_1,t_2)\big)\nu(t_1,t_2)\, dt_1\, dt_2\right)\right|^2 \nu(s_1,s_2)\, ds_1\, ds_2.$$

With a change of variables $(x_1, x_2) = (E_1(s_1)^{-\theta}, E_1(s_2)^{-\theta})$,

$$\begin{aligned}
&\int_{\mathbb{R}_+^2} \mathbb{1}_{(g(s_1,s_2),+\infty)}\big(g(t_1,t_2)\big)\nu(t_1,t_2)\, dt_1\, dt_2 \\
&\quad = \frac{1+\theta}{\theta^2} \int_0^{+\infty} \int_{\max(g(s_1,s_2)-t_1,0)}^{+\infty} (t_1+t_2)^{-\frac{1}{\theta}-2}\, dt_1\, dt_2 = \frac{1+\theta}{\theta} g(s_1,s_2)^{-\frac{1}{\theta}}.
\end{aligned}$$

Then, with the same change of variable, the bound can be rewritten as

$$\begin{aligned}
&2\alpha \frac{1+\theta}{\theta^2} \int_0^\infty \int_0^\infty \left| E_1^{-1}(x_1^{-\frac{1}{\theta}}) - E_1^{-1}\left(\frac{1+\theta}{\theta}(x_1+x_2)^{-\frac{1}{\theta}}\right)\right|^2 (x_1+x_2)^{-\frac{1}{\theta}-2}\, dx_1\, dx_2 \\
&\quad = 2\alpha \frac{1+\theta}{\theta^2} \int_0^\infty \int_{y_1}^\infty \left| E_1^{-1}(y_1^{-\frac{1}{\theta}}) - E_1^{-1}\left(\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}}\right)\right|^2 y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2.
\end{aligned}$$

We expand the binomial and treat the three terms separately:

$$\int_0^\infty \int_{y_1}^\infty E_1^{-1}(y_1^{-\frac{1}{\theta}})^2 y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2 = \frac{\theta^2}{1+\theta} \int_0^{+\infty} E_1^{-1}(x)^2\, dx = \frac{\theta^2}{1+\theta}.$$

Similarly,

$$\int_0^\infty \int_{y_1}^\infty E_1^{-1}\left(\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}}\right)^2 y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2 = \frac{\theta^2}{1+\theta}.$$

Thus $d_{\mathcal{W}}(\tilde{\mu}, \tilde{\mu}^{\mathrm{co}})^2 \le 4c$. In order to conclude it suffices to show that

$$\lim_{\theta\to+\infty} \frac{1+\theta}{\theta^2} \int_0^\infty \int_{y_1}^\infty E_1^{-1}(y_1^{-\frac{1}{\theta}}) E_1^{-1}\left(\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}}\right) y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2 = 1.$$

Since for every $a, b \in \mathbb{R}$, $(a-b)^2 \geq 0$, the integral is smaller or equal to 1. Thus it is enough to prove it to be greater or equal to 1. We observe that since $y_1 \leq y_2$,

$$\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}} \leq \frac{1+\theta}{\theta} y_1^{-\frac{1}{\theta}}.$$

Since E_1 is a decreasing function, so is its inverse. Thus

$$\begin{aligned}
&\frac{1+\theta}{\theta^2} \int_0^\infty \int_{y_1}^\infty E_1^{-1}(y_1^{-\frac{1}{\theta}}) E_1^{-1}\left(\frac{1+\theta}{\theta} y_2^{-\frac{1}{\theta}}\right) y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2 \\
&\quad \geq \frac{1+\theta}{\theta^2} \int_0^\infty \int_{y_1}^\infty E_1^{-1}(y_1^{-\frac{1}{\theta}}) E_1^{-1}\left(\frac{1+\theta}{\theta} y_1^{-\frac{1}{\theta}}\right) y_2^{-\frac{1}{\theta}-2}\, dy_1\, dy_2 \\
&\quad = \frac{1}{\theta} \int_0^\infty E_1^{-1}(y_1^{-\frac{1}{\theta}}) E_1^{-1}\left(\frac{1+\theta}{\theta} y_1^{-\frac{1}{\theta}}\right) y_1^{-\frac{1}{\theta}-1}\, dy_1 \\
&\quad = \frac{\theta}{1+\theta} \int_0^\infty E_1^{-1}\left(\frac{\theta}{1+\theta} x\right) E_1^{-1}(x)\, dx \\
&\quad \geq \frac{\theta}{1+\theta} \int_0^\infty E_1^{-1}(x)^2\, dx = \frac{\theta}{1+\theta},
\end{aligned}$$

which by taking the limit as $\theta \to +\infty$ is equal to 1.

9.13. *Proof of Proposition* 15. We point out that $\tilde{\boldsymbol{\xi}} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{\mu}}^{\mathrm{ind}} + \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$ and $\tilde{\boldsymbol{\xi}}^{\mathrm{co}} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{\mu}}^{\mathrm{co}} + \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$. Since by construction $\tilde{\boldsymbol{\mu}}^{\mathrm{ind}} \perp \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$ and $\tilde{\boldsymbol{\mu}}^{\mathrm{co}} \perp \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}$, by (23),

$$\mathcal{W}\big(\tilde{\boldsymbol{\xi}}(A), \tilde{\boldsymbol{\xi}}^{\mathrm{co}}(A)\big) \leq \mathcal{W}\big(\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A)\big) + \mathcal{W}\big(\tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}(A), \tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}(A)\big),$$

which is equal to $\mathcal{W}(\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}(A), \tilde{\boldsymbol{\mu}}^{\mathrm{co}}(A))$. By taking the supremum over $A \in \mathcal{X}$ we achieve the first statement. A very similar proof can be carried on for the second, by observing that $\tilde{\boldsymbol{\xi}}^{\mathrm{ind}} \overset{\mathrm{d}}{=} \tilde{\boldsymbol{\mu}}^{\mathrm{ind}} + \tilde{\boldsymbol{\mu}}_0^{\mathrm{ind}}$.

9.14. *Proof of Lemma* 17. Let $\{A_1, \dots A_n\}$ in $\mathcal{X}$ be disjoint sets. Then for $i = 1, \dots, n$ the random vectors $\tilde{\boldsymbol{\mu}}_f(A_i) = \int_{A_i} f(x) \tilde{\boldsymbol{\mu}}(dx)$ are independent since f is deterministic and $\tilde{\boldsymbol{\mu}}(A_1), \dots \tilde{\boldsymbol{\mu}}(A_n)$ are independent. This proves that $\tilde{\mu}_f$ is a CRV. The Lévy intensity ν_f may be found through the joint Laplace functional transform, $\mathbb{E}(e^{-\int g_1(x)\tilde{\mu}_1(dx) - \int g_2(x)\tilde{\mu}_2(dx)})$ for every pair of measurable functions $g_1, g_2 : \mathbb{X} \to \mathbb{R}^+$, which characterizes the law of a CRV $\tilde{\boldsymbol{\mu}} = (\tilde{\mu}_1, \tilde{\mu}_2)$:

$$\begin{aligned}
&\mathbb{E}\big(e^{-\int_{\mathbb{X}} g_1(x)\tilde{\mu}_{f,1}(dx) - \int_{\mathbb{X}} g_2(x)\tilde{\mu}_{f,2}(dx)}\big) \\
&\quad = \exp\left\{-\int_{\mathbb{R}_+^2 \times \mathbb{X}} [1 - e^{-(s_1 g_1(x) - s_2 g_2(x)) f(x)}] \nu(ds_1, ds_2, dx)\right\} \\
&\quad = \exp\left\{-\int_{\mathbb{R}_+^2 \times \mathbb{X}} [1 - e^{-s_1 g_1(x) - s_2 g_2(x)}] (p_f \# \nu)(ds_1, ds_2, dx)\right\},
\end{aligned}$$

where $p_f(s_1, s_2, x) = (s_1 f(x), s_2 f(x), x)$.

9.15. *Proof of Proposition* 18. Denote $\tilde{\boldsymbol{\mu}}_t = \tilde{\boldsymbol{\mu}}_{k(t|\cdot)}$ in the notation of Lemma 17, so that $\tilde{\boldsymbol{h}}(t) \overset{\mathrm{d}}{=} \tilde{\boldsymbol{\mu}}_t(\mathbb{R})$. By definition of GM-dependence, $\tilde{\boldsymbol{\mu}}_t(dy) = k(t|y)\tilde{\boldsymbol{\mu}}^{\mathrm{ind}}(dy) + k(t|y)\tilde{\boldsymbol{\mu}}_0^{\mathrm{co}}(dy)$. If $k(t|x) = \beta \mathbb{1}_{[0,t]}(x)$ and $\tilde{\mu}$ is a gamma CRM, by [4], Lemma 6 and Example 3, $k(t|x)\tilde{\mu}(dx)$ has Lévy measure

$$\pi(ds, dx) = \frac{e^{-\frac{s}{\beta}}}{s} \mathbb{1}_{(0,+\infty)}(s) \mathbb{1}_{(0,t)}(x)\, ds,$$

which corresponds to the generalized gamma CRM with parameters $b = \beta^{-1}$, $\sigma = 0$, $\alpha = t$ and $P_0 = \mathrm{Unif}([0, t])$. Thus, $\tilde{\boldsymbol{\mu}}_t$ is a special case of GM-dependent CRV with generalized gamma marginals. We conclude by Corollary 16.

Acknowledgments. The authors are grateful to the Editor and two referees for insightful comments and remarks, which led to a substantial improvement of the manuscript. Special thanks are due to one of referees who suggested the relationship reported in Lemma 3. The paper was completed while Marta Catalano was a Ph.D. student at Bocconi University. The authors are also affiliated to the Bocconi Institute for Data Science and Analytics (BIDSA) and the Collegio Carlo Alberto.

Funding. Antonio Lijoi and Igor Prünster are partially supported by MIUR, PRIN Project 2015SNS29B.

REFERENCES

[1] BACALLADO, S., DIACONIS, P. and HOLMES, S. (2015). de Finetti priors using Markov chain Monte Carlo computations. *Stat. Comput.* **25** 797–808. MR3360493 https://doi.org/10.1007/s11222-015-9562-9

[2] BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

[3] CAMERLENGHI, F., LIJOI, A. and PRÜNSTER, I. (2021). Survival analysis via hierarchically dependent mixture hazards. *Ann. Statist.* **49** 863 – 884. https://doi.org/10.1214/20-AOS1982

[4] CATALANO, M., LIJOI, A. and PRÜNSTER, I. (2020). Approximation of Bayesian models for time-to-event data. *Electron. J. Stat.* **14** 3366–3395. MR4148234 https://doi.org/10.1214/20-EJS1746

[5] CIFARELLI, D. M. and REGAZZINI, E. (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.

[6] CONT, R. and TANKOV, P. (2004). *Financial Modelling with Jump Processes. Chapman & Hall/CRC Financial Mathematics Series*. CRC Press/CRC, Boca Raton, FL. MR2042661

[7] CUTURI, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* **26** 2292–2300.

[8] DALEY, D. J. and VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I: Elementary Theory and Methods*, 2nd ed. *Probability and Its Applications* (*New York*). Springer, New York. MR1950431

[9] DE BLASI, P., PECCATI, G. and PRÜNSTER, I. (2009). Asymptotics for posterior hazards. *Ann. Statist.* **37** 1906–1945. MR2533475 https://doi.org/10.1214/08-AOS631

[10] DE FINETTI, B. (1938). Sur la condition d'Equivalence partielle. *Actual. Sci. Ind.* **739** 5–18.

[11] DOKSUM, K. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Probab.* **2** 183–201. MR0373081 https://doi.org/10.1214/aop/1176996703

[12] DYKSTRA, R. L. and LAUD, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9** 356–367. MR0606619

[13] EPIFANI, I. and LIJOI, A. (2010). Nonparametric priors for vectors of survival functions. *Statist. Sinica* **20** 1455–1484. MR2777332

[14] FLAMARY, R. and COURTY, N. (2017). POT Python Optimal Transport library.

[15] GHOSAL, S. and VAN DER VAART, A. (2017). *Fundamentals of Nonparametric Bayesian Inference. Cambridge Series in Statistical and Probabilistic Mathematics* **44**. Cambridge Univ. Press, Cambridge. MR3587782 https://doi.org/10.1017/9781139029834

[16] GINI, C. (1914). Di una misura delle relazioni tra le graduatorie di due caratteri. *Saggi Monografici del Comune di Roma, Tip. Cecchini*.

[17] GRIFFIN, J. E. and LEISEN, F. (2017). Compound random measures and their use in Bayesian nonparametrics. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 525–545. MR3611758 https://doi.org/10.1111/rssb.12176

[18] GRIFFITHS, R. C. and MILNE, R. K. (1978). A class of bivariate Poisson processes. *J. Multivariate Anal.* **8** 380–395. MR0512608 https://doi.org/10.1016/0047-259X(78)90061-1

[19] HJORT, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.* **18** 1259–1294. MR1062708 https://doi.org/10.1214/aos/1176347749

[20] HJORT, N. L., HOLMES, C. C., MÜLLER, P. and WALKER, S. G. (2010). *Bayesian Nonparametrics*. *Camb. Ser. Stat. Probab. Math*. **28**. Cambridge University Press. Cambridge. MR2722987 https://doi.org/10.1017/CBO9780511802478

[21] ISHWARAN, H. and JAMES, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc*. **96** 161–173. MR1952729 https://doi.org/10.1198/016214501750332758

[22] ISHWARAN, H. and JAMES, L. F. (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** 175–190. MR2054297 https://doi.org/10.1198/016214504000000179

[23] JAMES, L. F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist*. **33** 1771–1799. MR2166562 https://doi.org/10.1214/009053605000000336

[24] KALLENBERG, O. (2017). *Random Measures*, *Theory and Applications*. *Probability Theory and Stochastic Modelling* **77**. Springer, Cham. MR3642325 https://doi.org/10.1007/978-3-319-41598-7

[25] KALLSEN, J. and TANKOV, P. (2006). Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivariate Anal*. **97** 1551–1572. MR2275419 https://doi.org/10.1016/j.jmva.2005.11.001

[26] KINGMAN, J. F. C. (1967). Completely random measures. *Pacific J. Math*. **21** 59–78. MR0210185

[27] KNOTT, M. and SMITH, C. S. (1984). On the optimal mapping of distributions. *J. Optim. Theory Appl*. **43** 39–49. MR0745785 https://doi.org/10.1007/BF00934745

[28] LAU, J. W., CRIPPS, E. and HUI, W. (2020). Variational inference for multiplicative intensity models. *Statist. Probab. Lett*. **161** 108720, 6. MR4065484 https://doi.org/10.1016/j.spl.2020.108720

[29] LEISEN, F. and LIJOI, A. (2011). Vectors of two-parameter Poisson–Dirichlet processes. *J. Multivariate Anal*. **102** 482–495. MR2755010 https://doi.org/10.1016/j.jmva.2010.10.008

[30] LEISEN, F., LIJOI, A. and SPANÓ, D. (2013). A vector of Dirichlet processes. *Electron. J. Stat*. **7** 62–90. MR3020414 https://doi.org/10.1214/12-EJS764

[31] LIJOI, A. and NIPOTI, B. (2014). A class of hazard rate mixtures for combining survival data from different experiments. *J. Amer. Statist. Assoc*. **109** 802–814. MR3223751 https://doi.org/10.1080/01621459.2013.869499

[32] LIJOI, A., NIPOTI, B. and PRÜNSTER, I. (2014). Bayesian inference with dependent normalized completely random measures. *Bernoulli* **20** 1260–1291. MR3217444 https://doi.org/10.3150/13-BEJ521

[33] LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*. *Camb. Ser. Stat. Probab. Math*. **28** 80–136. Cambridge Univ. Press, Cambridge. MR2730661

[34] LO, A. Y. and WENG, C.-S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math*. **41** 227–245. MR1006487 https://doi.org/10.1007/BF00049393

[35] MACEACHERN, S. N. (1999). Dependent nonparametric processes. In *ASA Proceedings of the Section on Bayesian Statistical Science*. American Statistical Association, Alexandria, VA.

[36] MACEACHERN, S. N. (2000). Dependent Dirichlet processes. Technical Report, The Ohio State Univ.

[37] MARIUCCI, E. and REISS, M. (2018). Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Stat*. **12** 2482–2514. MR3833470 https://doi.org/10.1214/18-EJS1456

[38] MÜLLER, P., QUINTANA, F. A. and PAGE, G. (2018). Nonparametric Bayesian inference in applications (with discussion). *Stat. Methods Appl*. **27** 175–251. MR3807363 https://doi.org/10.1007/s10260-017-0405-z

[39] MÜLLER, P., QUINTANA, F. A., JARA, A. and HANSON, T. (2015). *Bayesian Nonparametric Data Analysis*. *Springer Series in Statistics*. Springer, Cham. MR3309338 https://doi.org/10.1007/978-3-319-18968-0

[40] NGUYEN, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist*. **41** 370–400. MR3059422 https://doi.org/10.1214/12-AOS1065

[41] NGUYEN, X. (2016). Borrowing strengh in hierarchical Bayes: Posterior concentration of the Dirichlet base measure. *Bernoulli* **22** 1535–1571. MR3474825 https://doi.org/10.3150/15-BEJ703

[42] PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl*. **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

[43] PECCATI, G. and PRÜNSTER, I. (2008). Linear and quadratic functionals of random hazard rates: An asymptotic analysis. *Ann. Appl. Probab*. **18** 1910–1943. MR2462554 https://doi.org/10.1214/07-AAP509

[44] RACHEV, S. (1985). The Monge–Kantorovich mass transference problem and its stochastic applications. *Theory Probab. Appl*. **29** 647–676. MR0773434

[45] REGAZZINI, E., LIJOI, A. and PRÜNSTER, I. (2003). Distributional results for means of normalized random measures with independent increments. *Ann. Statist*. **31** 560–585. MR1983542 https://doi.org/10.1214/aos/1051027881

[46] RIVA-PALACIO, A. and LEISEN, F. (2021). Compound vectors of subordinators and their associated positive Lévy copulas. *J. Multivariate Anal*. **183** 104728. MR4219330 https://doi.org/10.1016/j.jmva.2021.104728

[47] RÜSCHENDORF, L. (1991). Fréchet-bounds and their applications. In *Advances in Probability Distributions with Given Marginals* (*Rome*, 1990). *Math. Appl.* **67** 151–187. Kluwer Academic, Dordrecht. MR1215951

[48] SATO, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. *Cambridge Studies in Advanced Mathematics* **68**. Cambridge Univ. Press, Cambridge. Translated from the 1990 Japanese original, Revised by the author. MR1739520

[49] SETHURAMAN, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4** 639–650. MR1309433

[50] TANKOV, P. (2003). Dependence structure of spectrally positive multidimensional Lévy processes. Unpublished Manuscript.

[51] VILLANI, C. (2008). *Optimal Transport*: *Old and New*. Springer, Berlin Heidelberg.