Unit-rate Poisson representations of completely random measures

Peter Orbanz and Sinead Williamson

Abstract: Constructive definitions of discrete random measures, which specify a sampling procedure for the weights and atom locations of the measure, have proven to be of great value in Bayesian statistics and related fields. We consider the case of completely random measures and obtain a constructive representation on Euclidean space. For random measures on the real line satisfying a specific σ-finiteness property, the representation is equivalent to the Ferguson-Klass representation of pure-jump Lévy processes. Examples include constructive representations of the gamma process, the stable process and the beta process.

1. Introduction

Kingman's [17] notion of a completely random measure (CRM) has become a key concept in Bayesian nonparametric statistics. Many nonparametric priors describe a parameter variable which is a random measure or random probability measure. Most of these random measures are either completely random measures, or are obtained from a completely random measure by normalization (so-called normalized random measures with independent increments, or NRMIs) [9]. The characteristic decoupling properties of CRMs account for the tractable posterior distributions of such models. Several important aspects of models based on CRMs, including posterior computations, can be abstracted from the specific model in question and treated in a generic manner for the entire class of CRMs [20]; NRMIs have similar generic properties [5, 12]. In addition to immediate applications in survival analysis and mixture modeling [e.g. 9], CRMs and NRMIs can be used to construct more complex models appropriate for a number of applications including user choice models [2, 8, 26] and network models [3].

A CRM on a space Ω_θ can be represented, in a sense to be made precise in Sec. 2, by a discrete random measure ξ_r on Ω_θ, and hence as

$$\xi_r(\bullet) = \sum_{k=1}^{\infty} S_k \delta_{\Theta_k}(\bullet) \; . \tag{1.1}$$

The random variables S_k and Θ_k take values in $\mathbb{R}_+$ and Ω_θ, respectively. For the Dirichlet process—which is a NRMI rather than a CRM, and hence satisfies $\sum_k S_k = 1$ a.s.— the sequences of variables S_k and Θ_k can be generated from two sequences of i.i.d. random variables in a simple procedure known as stick-breaking [11, 24]. Analogous representations are known for the beta process [2, 21, 25]. In the following, we derive a similar constructive representation for

completely random measures, which represents the variables S_k and Θ_k as the images of unit-rate Poisson process draws under a fixed transformation mapping and can be regarded as a generalization of the Ferguson-Klass representation of pure-jump Lévy processes [6]. These representations are not generally identical to stick-breaking representations, but may coincide, as in the case of the beta process (Sec. 3.2). We refer to the well-known survey by Rosiński [23] for an overview on various related representations.

2. Results

Let Ω_θ be a Polish space with Borel sets $\mathcal{B}(\Omega_\theta)$, and let $\mathbf{M}_+(\Omega_\theta)$ be the set of measures on Ω_θ. Let ξ be a completely random measure on Ω_θ, that is, a $\mathbf{M}_+(\Omega_\theta)$-valued random variable for which $\xi(A) \perp\!\!\!\perp \xi(A')$ whenever $A, A' \in \mathcal{B}(\Omega_\theta)$ are disjoint sets. According to Kingman [17, Theorem 1], ξ admits a unique decomposition

$$\xi = \xi_d + \xi_f + \xi_r \tag{2.1}$$

into a deterministic (non-random) measure ξ_d, a random, purely atomic measure ξ_f with fixed atom locations, and a random discrete measure ξ_r. The random component ξ_r can be described by a Poisson process with mean measure μ_ξ on the Borel sets $\mathcal{B}(\Omega_\theta \times \mathbb{R}_+)$ [19, §8]. For $\Omega_\theta = \mathbb{R}_+$, this process is a subordinator—i.e. a strictly-increasing, pure-jump Lévy process. We slightly abuse terminology and refer to the measure $\nu_\xi(d\theta, ds) := \mu_\xi(d\theta \times ds)$ as the *Lévy measure* of ξ, regardless of the choice of Ω_θ.

Notation. If Y is a subspace of a Euclidean space, λ_Y denotes Lebesgue measure on Y. The cumulative distribution function of a measure ρ on a one-dimensional set Y is denoted $F_\rho(y) := \rho(\{z \in Y | z \leq y\})$. For any monotonic function $x \mapsto f(x)$ on $\mathbb{R}_+$, we write f^{-1} for the right-continuous inverse, hence $f^{-1}(y) = \inf\{x | f(x) \geq y\}$ (if f is non-decreasing) or $f^{-1}(y) = \inf\{x | f(x) \leq y\}$ (if f is non-increasing). $\Pi(\mu)$ denotes a Poisson process with mean measure μ.

2.1. Poisson representation of CRMs

The class of completely random measures to which our results are applicable is characterized as follows.

Definition 2.1. Say that a CRM ξ on Ω_θ is *nice* if it satisfies:

(C1) It has no deterministic component, $\xi = \xi_f + \xi_r$.
(C2) It is Σ-finite in the sense of Kingman [19, §8.1]: There is a countable partition of Ω_θ into disjoint sets D_j with $\mathbb{P}\{\xi(D_j) < \infty\} > 0$ for all j.
(C3) There are no jumps of size 0, that is, $\nu_\xi(\Omega_\theta, \{0\}) = 0$.
(C4) The Lévy measure ν_ξ satisfies $\int_0^\infty \min\{1, s\}\nu_\xi(\Omega_\theta, ds) < \infty$.

All these conditions are commonly assumed in the literature: (C1) merely simplifies notation. The perhaps somewhat opaque condition (C2) serves to

ensure the existence of the decomposition (2.1). If one presupposes ξ to be of the form $\xi_f + \xi_r$, its only purpose is to control the properties of the fixed-atom component ξ_f. In particular, it implies that the set of atoms of ξ_f is countable (whereas, in absence of suitable conditions, this set need not even be measurable) [17]. If (2.1) is known to hold, one can for most purposes replace (C2) by the strictly stronger condition that ξ_f has countably many atoms and $\xi_f(\Omega_\theta) < \infty$ almost surely. (C3) means each jump location in the representation (2.4) of ξ below corresponds to an actual jump. Imposing (C4) ensures the Lévy process defined by the Lévy measure $\nu_\xi(\Omega_\theta, \bullet)$ has bounded variation, and hence—since the process is non-decreasing—that the total mass $\xi_r(\Omega_\theta)$ of the random measures is almost surely finite [15, Lemma 12.13]. For our purposes, it implies

$$\nu_\xi(\Omega_\theta, (s, \infty)) < \infty \qquad \text{for all } s > 0 \,, \tag{2.2}$$

which in turn ensures the tail T of ν_ξ, in (2.3) below, is well-defined. If even $\nu_\xi(\Omega_\theta, [0, \infty)) < \infty$, then ξ_r has finitely many atoms almost surely, and can be represented as a compound Poisson process; otherwise, the number of atoms is countably infinite. The case of most interest to Bayesian nonparametrics, however, is typically the former: (2.2) holds, but ν_ξ is not totally finite. Indeed, for the arguably most important example—the gamma process, which in turn defines the Dirichlet process via normalization—ν_ξ is not even σ-finite, but nonetheless satisfies (2.2).

Theorem 2.2 (Poisson sampling of CRMs)**.** *Let ξ be a nice completely random measure on Ω_θ. Denote the tail of $\nu_\xi(\Omega_\theta, \bullet)$ by*

$$\begin{aligned} T : \mathbb{R}_+ &\to \mathbb{R}_+ \\ s &\mapsto T(s) := \nu_\xi(\Omega_\theta, (s, \infty)) \end{aligned} \tag{2.3}$$

and by $\{\theta_1, \theta_2, \dots\} \subset \Omega_\theta$ the set of fixed jump locations.
(i) There is a probability kernel $\mathbf{p} : \Omega_s \to \mathbf{M}(\Omega_\theta)$ such that

$$\xi(\bullet) \;=\; \xi_f(\bullet) + \xi_r(\bullet) \;\stackrel{d}{=}\; \sum_i J_i \delta_{\theta_i}(\bullet) + \sum_k T^{-1}(U_k)\delta_{V_k}(\bullet) \tag{2.4}$$

where $U_k \sim \Pi(\lambda)$ is a unit rate Poisson process on $\Omega_s = \mathbb{R}_+$ and $V_1, V_2, \dots$ are independent random variables with $V_k \sim \mathbf{p}(\bullet, T^{-1}(U_k))$. The random variables J_i are independent of each other and of ξ_r.
(ii) The regular conditional probability $\mathbf{p}$ is unique up to equivalence and determined by

$$\nu_\xi(A, B) = \int_B \mathbf{p}(A, s) d\nu_\xi(\Omega_\theta, s) \tag{2.5}$$

even if $\nu_\xi(\Omega_\theta, \bullet)$ is not σ-finite.

If $\nu_\xi(\Omega_\theta, \bullet)$ is σ-finite, (2.5) simply states that $\mathbf{p}$ is given by the densities

$$\mathbf{p}(A, s) := \frac{d\nu_\xi(A, \bullet)}{d\nu_\xi(\Omega_\theta, \bullet)}(s) \qquad \text{for } A \in \mathcal{B}(\Omega_\theta) \,. \tag{2.6}$$

The intuition underlying Theorem 2.2 is straightforward: As a completely random measure, ξ_r can represented by a Poisson process with mean measure ν_ξ (see [19, §8] for a detailed discussion). Specifically,

$$\xi_r \stackrel{d}{=} \sum_k S_k \delta_{\Theta_k} \quad \text{if and only if} \quad (S_k, \Theta_k) \sim \Pi(\nu_\xi) \; . \tag{2.7}$$

Let $\phi : \Omega_\theta \times \mathbb{R}_+ \to \Omega_\theta \times \mathbb{R}_+$ be a measurable mapping. By one of the basic properties of Poisson processes [19, Chapter 2.3], the image of the Poisson process $\Pi(\gamma)$ under ϕ satisfies

$$\phi(\Pi(\gamma)) = \Pi(\phi(\gamma)) \tag{2.8}$$

for any σ-finite measure γ on $\Omega_\theta \times \mathbb{R}_+$. We can thus posit a simple mean measure γ—for example Lebesgue measure, if Ω_θ is Euclidean—and reduce the representation of ξ_r to a transformation of γ. To determine this transformation, the equation

$$\nu_\xi = \phi(\gamma) \qquad \text{s.t. } \phi \text{ measurable} \tag{2.9}$$

has to be solved for ϕ. Problems of the form (2.9) are known as *transport problems* in analysis [27]. In one dimension, these problems admit a simple solution: If ν and γ are measures on the real line with distribution functions F_ν and F_γ, respectively, then the transport problem $\nu = \phi(\gamma)$ is obviously solved by

$$\phi := F_\nu^{-1} \circ F_\gamma \; . \tag{2.10}$$

If γ is in particular Lebesgue measure, then $\phi = F_\nu^{-1}$. Theorem 2.2 substitutes the tail T for the distribution function F_ν, since F_ν is not well-defined for measures whose mass is infinite in a neighborhood of 0.

In multiple dimensions, the transport problem becomes considerably more difficult, which affects the proof Theorem 2.2, since the sample space is $\Omega_\theta \times \mathbb{R}_+$ and hence at least two-dimensional. The strategy pursued in the proof is to factor out a one-dimensional problem on $\mathbb{R}_+$, by disintegrating μ_ξ into the pair $(\mathbf{p}, \nu_\xi(\Omega_\theta, \bullet))$, where $\mathbf{p}$ is a conditional probability on Ω_θ given $s \in \mathbb{R}_+$, and $\nu_\xi(\Omega_\theta, \bullet)$ is a measure on $\mathbb{R}_+$. The transport problem is then solved only on $\mathbb{R}_+$—without further assumptions on the structure of Ω_θ, a general solution for $\mathbf{p}$ is not feasible. Thus, ξ_r is sampled by sampling (S_k, Θ_k) from a Poisson process with mean measure given by $\mathbf{p}(d\theta, T^{-1}(u))\lambda_{\mathbb{R}_+}(du)$, and only the weights S_k of ξ_r are reduced to unit-rate sampling.

2.2. Representation on Euclidean space

If Ω_θ is contained in Euclidean space, it is possible to construct more elaborate couplings and extend the representation in Theorem 2.2 to the variables Θ_k, thus fully reducing ξ_r to unit-rate Poisson samples. We will assume here for simplicity that $\Omega_\theta = \mathbb{R}^D$, although the approach carries over immediately to the cone $\mathbb{R}_+^D$ or to products of closed intervals.

To solve the transport problem on Ω_θ, the disintegration approach used above to separate Ω_θ and $\mathbb{R}_+$ is now in turn applied repeatedly to $\mathbf{p}(\bullet, s)$. Each application separates off a one-dimensional component, to which (2.10) is applicable. Formalizing this approach comes at the price of some rather cumbersome notation: Label the axes of $\mathbb{R}^D$ as $\mathbb{R}^{(1)}, \mathbb{R}^{(2)}, ...$, and suppose that $\mathbf{p}$ has been obtained by application of Theorem 2.2. For $d = 1, \dots, D$, denote by $\mathbf{p}_d$ the marginal measure of $\mathbf{p}$ on the subspace $\mathbb{R}^d = \mathbb{R}^{(1)} \times \cdots \times \mathbb{R}^{(\mathrm{d})}$. Let $\mathbf{q}_d$ be the probability kernel obtained by disintegrating $\mathbf{p}_d$ with respect to $\mathbf{p}_{d-1}$; that is,

$$\mathbf{p}_d(d\theta_1 \cdots d\theta_d, s) = \mathbf{q}_d(d\theta_d|\theta_1, \dots, \theta_{d-1}, s)\mathbf{p}_{d-1}(d\theta_1 \cdots d\theta_{d-1}, s) \; . \tag{2.11}$$

As above, let $F_{\mathbf{q}_d}(\theta_d|\theta_1, \dots, \theta_{d-1}, s)$ denote the cumulative distribution function of the one-dimensional measure $\mathbf{q}_d(\bullet|\theta_1, \dots, \theta_{d-1}, s)$.

Theorem 2.3 (Successive subspace sampling)**.** *Let ξ be a nice completely random measure on Ω_θ, and let T be defined as in (2.3). For unit-rate Poisson samples*

$$(U_k, W_k^{(1)}, \dots, W_k^{(\mathrm{d})}) \sim \Pi(\lambda_{\mathbb{R}_+} \otimes \lambda_{[0,1]}^{\mathrm{D}}) \; , \tag{2.12}$$

define

$$S_k := T^{-1}(U_k) \quad \textit{and} \quad \Theta_k^{(\mathrm{d})} := F_{\mathbf{q}_d}^{-1}(W_k^{(\mathrm{d})}|\Theta_k^{(1)}, \dots, \Theta_k^{(\mathrm{d\text{-}1})}, S_k) \; . \tag{2.13}$$

Then the purely random component ξ_r of ξ is distributed as

$$\xi_r \stackrel{d}{=} \sum_{k=1}^{\infty} S_k \delta_{(\Theta_k^{(1)}, \dots, \Theta_k^{(\mathrm{D})})} \; . \tag{2.14}$$

This reduction to the one-dimensional solution (2.10) by successive disintegration is an example of a general approach to the determination of couplings on Euclidean spaces, due originally to Rosenblatt [22].

Remark 2.4. In (2.6), the disintegration of $\mathbf{p}$ is obtained as a family of densities. It is not difficult to see that the relevant argument in the proof of Theorem 2.2 also applies to the disintegration in Theorem 2.3—with some considerable simplifications, since $\mathbf{p}_d$ always describes a probability measure. The probability kernel $\mathbf{q}_d$ in Theorem 2.3 can therefore be represented as densities

$$\mathbf{q}_d(A|\theta^{(1)}, \dots, \theta^{(\mathrm{d\text{-}1})}, s) = \frac{\mathbf{p}_d(d\theta^{(1)} \cdots d\theta^{(\mathrm{d\text{-}1})} \times A \times ds)}{\mathbf{p}_d(d\theta^{(1)} \cdots d\theta^{(\mathrm{d\text{-}1})} \times \mathbb{R} \times ds)} \; , \tag{2.15}$$

in direct analogy to (2.6).

2.3. Representation on the line

If Ω_θ is an interval in $\mathbb{R}_+$, the collection of functions $F_{\mathbf{q}_d}^{-1}$ reduces to $F_{\mathbf{q}_1}^{-1}(w|s) = (\mathbf{p}([0, \bullet], s)^{-1})(w)$. Therefore, nice CRMs on the positive reals can be represented as follows:

Corollary 2.5 (Ferguson and Klass [6]). *Let ξ be a nice completely random measure on $\Omega_\theta = \mathbb{R}_+$ or $\Omega_\theta = [0, \theta_{max}]$, where $\theta_{max} \in \mathbb{R}_+$. Denote the inverse of the distribution function $\theta \mapsto \mathbf{p}([0,\theta], s)$ by*

$$\begin{aligned} m : [0,1] \times \mathbb{R}_+ &\to \Omega_\theta \\ (w,s) &\mapsto m(w,s) := (\mathbf{p}([0,\bullet], s)^{-1})(w) \ . \end{aligned} \tag{2.16}$$

Then

$$\xi(\bullet) \ = \ \xi_f(\bullet) + \xi_r(\bullet) \ \stackrel{d}{=} \ \sum_i J_i \delta_{\theta_i}(\bullet) + \sum_k T^{-1}(U_k) \delta_{m(W_k, T^{-1}(U_k))}(\bullet) \tag{2.17}$$

where $(U_k, W_k) \sim \Pi(\lambda \otimes \lambda_{[0,1]})$ is a unit rate Poisson process on $\Omega_\theta \times [0,1]$.

In other words, the transport problem $\mu = \phi(\lambda \otimes \lambda_{[0,1]})$ is in this case solved by

$$\phi(u,w) = (T^{-1}(u), m(w, T^{-1}(u))) \ . \tag{2.18}$$

Remark 2.6 (Distribution of ξ_f). If the distribution ξ_r serves as a prior distribution in a nonparametric Bayesian model, fixed atoms arise in the posterior, where the locations θ_i in Theorem 2.2 correspond to observations or latent observations. The distributions of the random variables J_i can then be derived explicitly from ν_ξ [20]. For Bayesian models with a conjugate posterior, the representation of ξ_f can be absorbed into that of ξ_r, by adding a suitable atomic measure to the Lévy measure ν_ξ.

3. Examples

We distinguish between two cases, the homogeneous and inhomogeneous case. In analogy to Lévy processes, a CRM is called *homogeneous* if the measure ν_ξ factorizes as $\nu_\xi(d\theta, ds) = H_0(d\theta)\nu_\xi^{(s)}(ds)$. Consequently, the conditional probability $\mathbf{p}$ of Θ in Theorem 2.2 becomes independent of S, and hence $\mathbf{p}(A,s) = H_0(A)/H_0(\Omega_\theta)$. See the classic work of Kingman [18] for a definitive treatment of the homogeneous case.

3.1. The homogeneous case

Example 1 (Gamma CRM). The *gamma CRM* is the completely random measure with Lévy measure

$$\nu_\xi(d\theta, ds) = s^{-1} e^{-cs} ds H_0(d\theta) \ , \tag{3.1}$$

for $c > 0$ and $H_0(d\theta)$ a finite measure on Ω_θ. The tail can therefore be represented by means of the exponential integral E_1 as

$$T(s) = H_0(\Omega_\theta)\mathrm{E}_1(cs) \ . \tag{3.2}$$

Example 2 (Stable CRM)**.** For $\alpha \in (0,1)$ and finite measure $H_0(d\theta)$ on Ω_θ, the CRM with Lévy measure

$$\nu_\xi(d\theta, ds) = \frac{\alpha}{\Gamma(1-\alpha)s^{1+\alpha}} ds H_0(d\theta) \; . \tag{3.3}$$

is called an *α-stable CRM*. The tail is given by

$$T(s) = H_0(\Omega_\theta)(\Gamma(1-\alpha)z^\alpha)^{-1} \; . \tag{3.4}$$

3.2. An inhomogeneous case: Beta CRMs

The beta CRM, introduced by Hjort [10], is a CRM with

$$\nu_\xi(d\theta, ds) = c(\theta)s^{-1}(1-s)^{c(\theta)-1} ds \, dH_0(\theta) \; , \tag{3.5}$$

where H_0 is a totally finite measure on Ω_θ and the function $c(\theta)$ is assumed to be non-negative and piecewise-continuous. In Bayesian nonparametric statistics, the beta CRM is used as a prior over cumulative hazards, in which case the corresponding distribution function is neutral-to-the-right. In this case, H_0 is a prior guess at the hazard function on $\mathbb{R}_+$ restricted to a subset $\Omega_\theta = [0, \theta_{\max})$ that meets the finiteness constraint.

In general, evaluating the tail involves evaluating a degenerate incomplete beta function, and cannot be done analytically. Wolpert and Ickstadt [28] describe an approximate method for evaluating this degenerate incomplete beta function. For certain choices of c and H_0, $\mathbf{p}([0,\theta], s)$ can be obtained analytically; two examples are given below.

Example 3 (Beta CRM with $c(\theta) = \exp(-H_0(\theta))$)**.** Consider the beta CRM with $c(\theta) = \exp(-H_0(\theta))$. Then

$$\mathbf{p}([0,\theta], s) = \frac{1 - s - (1-s)^{\exp(-H_0(\theta))}}{1 - s - (1-s)^{\exp(-H_0(\theta_{\max}))}},$$

and we can obtain $\mathbf{p}([0,\bullet], s)^{-1}(u)$ if H_0 is invertible.

Another application of the beta CRM derives from the Indian buffet process [8], a distribution over binary sequences used as a prior in nonparametric latent feature models. These random sequences are exchangeable, and the mixing measure in their de Finetti representation is a beta CRM where $c(\theta) = 1$, rendering the CRM homogeneous [26].

Example 4 (Beta CRM with $c(\theta) = 1$)**.** If $c(\theta) = 1$, the Lévy measure reduces to $\nu_\xi(d\theta, ds) = \frac{ds}{s} dH_0(\theta)$, and the tail function is

$$T(s) = -H_0(\Omega_\theta)\log(z) \; , \tag{3.6}$$

yielding $S_k = \exp\{-\frac{U_k}{H_0(\Omega_\theta)}$.

Let $\tilde{U}_1 < \tilde{U}_2 < \dots$ be the ordered arrival times of a unit rate Poisson process $U \sim \Pi(\lambda)$. The inter-arrival times $\tilde{U}_{n+1} - \tilde{U}_n$ are distributed according to Expon(1), therefore $\exp(\tilde{U}_n - \tilde{U}_{n+1})$ is distributed according to Uniform$(0,1)$ and $\exp((\tilde{U}_n - \tilde{U}_{n+1})/H_0(\Omega_\theta))$ is distributed according to Beta$(H_0(\Omega_\theta), 1)$. We can therefore directly generate the strictly-ordered atom sizes $s_1 > s_2 > \dots$ of a beta CRM with $c(\theta) = 1$ as

$$s_n = \prod_{i=1}^{n} b_n \qquad \text{with} \qquad b_n \sim \text{Beta}(H_0(\Omega_\theta), 1) \; . \tag{3.7}$$

Equation (3.7) is precisely the stick-breaking construction of the beta CRM derived by Teh, Görur, and Ghahramani [25].

4. Related work

The notion of a completely random measure is due to Kingman [17], and has been studied in a variety of contexts in Bayesian statistics (see e.g. the recent surveys by Lijoi and Prünster [20] and by Jordan [13]).

If $\Omega_\theta = \mathbb{R}_+$, the independence property $\xi(A) \perp\!\!\!\perp \xi(A')$ implies that the random component of a completely random measure ξ is representable as a stochastic process with independent increments. The random measure ξ_r is therefore equivalent to a subordinator, i.e. a pure-jump, strictly-increasing Lévy process, possibly non-stationary; two such random measures can be combined to give a general pure-jump Lévy process with finite variation. In particular, the marginal distribution of the jump sizes is infinitely divisible. Khintchine [16, Hauptsatz III] shows that infinitely divisible laws admit representations of the form (2.4). Ferguson and Klass [6] re-derive this result for the representation of pure-jump Lévy processes on the interval, and additionally give an explicit transformation equivalent to the solution of the transport problem on $\mathbb{R}_+$ as in Theorem 2.2. They conjecture that the approximating random sequence converges with probability one. Kallenberg [14] shows that this is the case, and that the sequence in fact converges uniformly almost surely. The proof of Ferguson and Klass [6] implicitly assumes the marginal Lévy measure $\nu_\xi(\Omega_\theta, \bullet)$ to be σ-finite (cf. the Radon-Nikodym derivative defined on p. 1636 of [6]). Though their result carries over immediately to completely random measures on the interval, the σ-finiteness assumption excludes some important CRMs, such as those defined by the gamma process [19, Chapter 9.4]. Representations of the form (2.4), usually without explicit transformations, exist more generally for exchangeable increment processes [15, Theorem 16.21], since such processes can be characterized as mixtures of Lévy processes [15, Theorem 11.15].

Kingman [19, Chapter 8.2] shows in detail how CRMs can represented as Poisson processes, and points out that these are indeed marked Poisson processes if the measure $\nu_\xi(\Omega_\theta, \bullet)$ is σ-finite. Thus, the beta CRM of Hjort [10] is a marked Poisson process; the gamma CRM is not. Motivated by applications in nonparametric statistics, the beta CRM has recently received renewed attention

in statistics and computer science. Analogues of the stick-breaking construction of the Dirichlet process have been derived for the beta CRM by Teh *et al.* [25] (cf. Sec. 3), and more recently by Paisley *et al.* [21] and Broderick *et al.* [2], who both emphasize the Poisson representation.

5. Proofs

Verifying the existence of the disintegration $(\mathbf{p}, \nu_\xi(\Omega_\theta, \bullet))$ raises some technical issues, since the measure $\nu_\xi(\Omega_\theta, \bullet)$ is not generally σ-finite. The intuition is, once again, very simple: If $\nu_\xi(\Omega_\theta, \bullet)$ is σ-finite, $\mathbf{p}(A, \bullet)$ is given by the density of $\nu_\xi(A, \bullet)$ with respect to $\nu_\xi(\Omega_\theta, \bullet)$, and it is easy to see that $\mathbf{p}(\emptyset, s) = 0$, that $\mathbf{p}(\Omega_\theta, s) = 1$, and that $A \mapsto \mathbf{p}(A, s)$ is increasing in A. In the general case, the analogous result is expressed by the following lemma.

Lemma 5.1. *Let ξ be a nice completely random measure on Ω_θ whose Lévy measure ν_ξ satisfies $\nu_\xi(\Omega_\theta, (s, \infty)) < \infty$. Then there is a probability kernel $\mathbf{p} : \mathbb{R}_+ \to \mathbf{M}(\Omega_\theta)$ satisfying*

$$\nu_\xi(A, ds) = \mathbf{p}(A, s)\nu_\xi(\Omega_\theta, ds) \qquad \text{for each } A \in \mathcal{B}(\Omega_\theta) \; . \tag{5.1}$$

Each function $s \mapsto \mathbf{p}(A, s)$ is uniquely determined up to a $\nu_\xi(\Omega_\theta, \bullet)$-null set.

The proof of Lemma 5.1 is more technical than instructive, and we defer it until the end of this section and first proceed with proofs of the main results.

Proof of Theorem 2.2. Any Σ-finite CRM ξ without non-random component can be represented by means of a Poisson process $\Pi(\mu)$ with mean measure $\mu(A \times B) = \nu_\xi(A, B)$. More precisely,

$$\xi \;=\; \xi_f + \xi_r \;\stackrel{d}{=}\; \sum_i J_i \delta_{\theta_i} \;+\; \sum_k S_k \delta_{\Theta_k} \; , \tag{5.2}$$

where θ_i are fixed atoms, the random variables J_i are mutually independent and do not depend on ξ_r, and (Θ_k, S_k) follow a Poisson process $\Pi(\mu)$ [4, Theorem 10.1.III]. Since the existence of $\mathbf{p}$ and Theorem 2.2(ii) follow from Lemma 5.1, what remains to be shown is that $\sum S_k \delta_{\Theta_k} \stackrel{d}{=} \sum T^{-1}(U_k)\delta_{V_k}$.

To this end, consider first the measure $\mu_{\Omega_\theta}(\bullet) := \nu_\xi(\Omega_\theta, \bullet)$ on atom sizes. Since ξ is Σ-finite, the tail T of μ_{Ω_θ} is finite on $(0, \infty]$ [19]. It is straightforward to verify, for Lebesgue measure λ on $\mathbb{R}_+$,

$$\lambda(T([a, b))) = \mu_{\Omega_\theta}[a, b) \tag{5.3}$$

for all $0 < a < b$. Since the intervals $[a, b)$ generate the Borel sets, (5.3) implies $T^{-1}\lambda = \mu_{\Omega_\theta}$, solving the one-dimensional transport problem.

Now consider the entire measure $\mu_\xi(A \times B) = \nu_\xi(A, B)$. By construction,

$$\mu_\xi(A \times B) = \int_B \mathbf{p}(A, s)\mu_{\Omega_\theta}(ds) = \int_B \mathbf{p}(A, s)[T^{-1}\lambda](ds) \tag{5.4}$$

for all $A \in \mathcal{B}(\mathbb{R}_+)$. Therefore, $\mathbf{p}$ is the conditional probability

$$\mathbf{p}(A, s) = P[\Theta \in A | S = s] \tag{5.5}$$

under the law of the Poisson process, and the proof is complete. □

Theorem 2.3 is a direct consequence of the construction of disintegrations and the representation of CRMs in Theorem 2.2. Corollary 2.5 then follows immediately as the special case $D = 1$.

Proof of Theorem 2.3. Suppose $S_k = s_k$ and $\Theta_k^{(\mathrm{d})} = \theta_k^{(\mathrm{d})}$ are sampled as in (2.13). By construction, each $\Theta_k^{(\mathrm{d})}$ has distribution

$$\textsc{Law}(\Theta_k^{(\mathrm{d})} | \Theta_k^{(1)} = \theta_k^{(1)}, \ldots, \Theta_k^{(\mathrm{d}\text{-}1)} = \theta_k^{(\mathrm{d}\text{-}1)}, S_k = s_k) = \mathbf{q}_d(\bullet | \theta_k^{(1)}, \ldots, \theta_k^{(\mathrm{d}\text{-}1)}, s_k) \,. \tag{5.6}$$

Since $\mathbf{q}_1(\bullet | s_k) = \mathbf{p}_1(\bullet | s_k)$, we have

$$\mathbf{q}_D(d\theta^{(\mathrm{D})} | \theta_k^{(1)}, \ldots, \theta_k^{(\mathrm{D}\text{-}1)}, s_k) \cdots \mathbf{q}_2(d\theta^{(2)} | \theta_k^{(1)}, s_k) \mathbf{p}_1(d\theta_k^{(1)} | s_k) = \mathbf{p}(d\theta_k^{(1)} \cdots d\theta_k^{(\mathrm{D})}, s_k)$$

and the joint law of $\Theta_k = (\Theta_k^{(1)}, \ldots, \Theta_k^{(\mathrm{D})})$ is thus $\textsc{Law}(\Theta_k) = \mathbf{p}(\bullet, s_k)$. An application of Theorem 2.2 yields the representation (2.14) of ξ_r. □

Both proofs above are contingent on Lemma 5.1, which remains to be established. The proof uses the following result to address the problem that the marginal measure $\nu_\xi(\Omega_\theta, \bullet)$ is not, in general, σ-finite.

Lemma 5.2 (Generalized Radon-Nikodym Theorem [7, 232E and 232B(b)])**.** *Let ν, ν' be measures on a measurable space $(\mathcal{X}, \mathcal{A})$. There is a measurable function $f : \mathcal{X} \to \mathbb{R}_+$ satisfying $\nu(A) = \int_A f d\nu'$ for all $A \in \mathcal{A}$ if and only if:*

(i) ν is absolutely continuous with respect to ν'.
(ii) For all $A \in \mathcal{A}$ with $\nu(A) > 0$, there exists another set $B \in \mathcal{A}$ such that $\nu'(B) < \infty$ and $\nu(A \cap B) > 0$.

In this case, f is uniquely determined ν'-a.e.

If ν' is σ-finite, absolute continuity implies condition (ii), and the lemma reduces to the Radon-Nikodym theorem.

Proof of Lemma 5.1. For the proof, abbreviate $\mu_{\mathrm{A}} := \nu_\xi(A, \bullet)$. We proceed in two steps:
Step (1). We first show that, for every $A \in \mathcal{B}(\Omega_\theta)$, there exists a measurable function $\mathbf{p}(A, \bullet) : \mathbb{R}_+ \to \mathbb{R}_+$ which satisfies (5.1). By Lemma 5.2, this is the case if the measures μ_{A} and $\nu_\xi(\Omega_\theta, \bullet)$ satisfy conditions (i) and (ii) of Lemma 5.2. Absolute continuity clearly holds since $\nu_\xi(A, D) \leq \nu_\xi(\Omega_\theta, D)$ for every $D \in \mathcal{B}(\mathbb{R}_+)$ by construction. To verify condition (ii), observe that (2.6) implies $\mu_{\mathrm{A}}(\varepsilon, \infty) < \infty$ for all $\varepsilon > 0$. The obvious strategy is therefore to show that (ii) is satisfied for a set B of the form $B = (\varepsilon, \infty)$. To this end, let $\mu_{\mathrm{A}} = \mu_{\mathrm{A}}^{\|} + \mu_{\mathrm{A}}^{\perp}$ the decomposition of μ_{A} into its purely atomic component $\mu_{\mathrm{A}}^{\perp}$ and the atomless measure $\mu_{\mathrm{A}}^{\|} = \mu_{\mathrm{A}} - \mu_{\mathrm{A}}^{\perp}$. Suppose $\mu_{\mathrm{A}}(D) > 0$ as in (ii). We distinguish two cases:

Case 1: $\mu_{\mathrm{A}}^{\parallel}(D) > 0$. Since $\mu_{\mathrm{A}}^{\parallel}$ is atomless, any set with $\mu_{\mathrm{A}}^{\parallel}(D) > 0$ has a subset D_1 with $\mu_{\mathrm{A}}^{\parallel}(D_1) > 0$ and $\mu_{\mathrm{A}}^{\parallel}(D \setminus D_1) > 0$. In particular, since $D \subset \mathbb{R}_+$, there exists $\varepsilon > 0$ such that $\mu_{\mathrm{A}}^{\parallel}(D \cap [0, \varepsilon]) > 0$ and $\mu_{\mathrm{A}}(D \cap (\varepsilon, \infty)) > 0$. Hence, $B := (\varepsilon, \infty)$ satisfies condition (2) for any D.

Case 2: $\mu_{\mathrm{A}}^{\parallel}(D) = 0$, which implies $\mu_{\mathrm{A}}^{\perp}(D) > 0$. Since $\{0\}$ is by assumption not an atom, D contains an atom $\{c\}$ of $\mu_{\mathrm{A}}^{\perp}$ with $c > 0$. Hence, $B = (c/2, \infty)$ satisfies $\mu_{\mathrm{A}}^{\perp}(D \cap B) \geq \mu_{\mathrm{A}}^{\perp}(\{c\}) > 0$.

Step (2). What remains to be shown is that the separate functions $\mathbf{p}(A, \bullet)$ for each A can be assembled into a probability kernel, i.e. we need to know that $s \mapsto \mathbf{p}(A, s)$ is measurable and $A \mapsto \mathbf{p}(A, s)$ is a measure for every $s \in \mathbb{R}_+$. Measurability follows from Lemma 5.2. To establish σ-additivity of $\mathbf{p}(\bullet, s)$, suppose that (A_n) is a sequence of disjoint sets in $\mathcal{B}(\Omega_\theta)$ with $A := \cup A_n$. By σ-additivity of ν_ξ,

$$\int_B \mathbf{p}(\cup A_n, s)\nu_\xi(\Omega_\theta, ds) = \sum_{n=1}^{\infty} \nu_\xi(A_n, B) = \sum_{n=1}^{\infty} \int_B \mathbf{p}(A_n, s)\nu_\xi(\Omega_\theta, ds) \; . \quad (5.7)$$

Since $s \mapsto \mathbf{p}(A_n, s)$ are measurable functions with values in $[0, +\infty]$, we have [1, Corollary 11.5]

$$\sum_{n=1}^{\infty} \int_B \mathbf{p}(A_n, s)\nu_\xi(\Omega_\theta, ds) = \int_B \Big(\sum_{n=1}^{\infty} \mathbf{p}(A_n, s)\Big)\nu_\xi(\Omega_\theta, ds) \; , \quad (5.8)$$

which by a.e.-uniqueness implies $\mathbf{p}(A, s) \stackrel{a.e.}{=} \sum \mathbf{p}(A_n, s)$. Moreover,

$$0 = \nu_\xi(\emptyset, B) = \int_B \mathbf{p}(\emptyset, s)\nu_\xi(\Omega_\theta, ds) \quad (5.9)$$

for all $B \in \mathcal{B}(\mathbb{R}_+)$ implies $\mathbf{p}(\emptyset, s) = 0$ for almost all s. Thus, there is a version of $\mathbf{p}$ such that $A \mapsto \mathbf{p}(A, s)$ is a probability measure for all s. □

References

[1] Bauer, H. (2001). *Measure and Integration Theory*. W. de Gruyter.

[2] Broderick, T., Jordan, M. I., and Pitman, J. (2012). Beta processes, stick-breaking and power laws. *Bayesian Anal.*, **7**(2), 439–475.

[3] Caron, F. and Fox, E. B. (2014). Sparse graphs using exchangeable random measures. Preprint. http://arxiv.org/abs/1401.1137.

[4] Daley, D. and Vere-Jones, D. (2008). *An introduction to the theory of point processes*, volume I and II. Springer, 2nd edition.

[5] Favaro, S., Lijoi, A., and Prünster, I. (2012). Conditional formulae for Gibbs-type exchangeable random partitions. *Ann. Appl. Probab.* To appear.

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

[7] Fremlin, D. H. (2000–2006). *Measure Theory*, volume I–V. Torres Fremlin.

[8] Griffiths, T. L. and Ghahramani, Z. (2006). Infinite latent feature models and the Indian buffet process. In *Adv. Neural Inf. Process. Syst. 18*, pages 475–482. MIT Press.

[9] Hjort, N., Holmes, C., Müller, P., and Walker, S., editors (2010). *Bayesian Nonparametrics*. Cambridge University Press.

[10] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, **18**, 1259–1294.

[11] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, **96**(453), 161–173.

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

[13] Jordan, M. I. (2010). Hierarchical models, nested models and completely random measures. In M.-H. Chen, D. Dey, P. Mueller, D. Sun, and K. Ye, editors, *Frontiers of Statistical Decision Making and Bayesian Analysis: In Honor of James O. Berger*. Springer.

[14] Kallenberg, O. (1974). Series of random processes without discontinuities of the second kind. *Ann. Probability*, **2**, 729–737.

[15] Kallenberg, O. (2001). *Foundations of Modern Probability*. Springer, 2nd edition.

[16] Khintchine, A. (1937). Zur theorie der unbeschränkt teilbaren verteilungsgesetze. *Mat. Sbornik*, **2**(44), 79–119.

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

[18] Kingman, J. F. C. (1975). Random discrete distributions. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **37**, 1–22.

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

[20] Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In N. L. Hjort, C. Holmes, P. Müller, and S. G. Walker, editors, *Bayesian Nonparametrics*. Cambridge University Press.

[21] Paisley, J., Blei, D., and Jordan, M. I. (2012). Stick-breaking beta processes and the Poisson process. In *Proceedings of AISTATS 2012 (JMLR W&CP 22)*, pages 850–858.

[22] Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Ann. Math. Statistics*, **23**, 470–472.

[23] Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Birkhäuser Boston, Boston, MA.

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

[25] Teh, Y.-W., Görur, D., and Ghahramani, Z. (2007). Stick-breaking construction for the Indian buffet process. In *Proceedings of AISTATS 2007 (JMLR W&CP 2)*, pages 556–563.

[26] Thibaux, R. and Jordan, M. I. (2007). Hierarchical beta processes and the Indian buffet process. In *J. Mach. Learn. Res. Proceedings*, volume 2, pages

564–571.

[27] Villani, C. (2009). *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin. Old and new.

[28] Wolpert, R. L. and Ickstadt, K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**(2), 251–267.