Unit-rate Poisson representations of completely random measures

Peter Orbanz* and Sinead Williamson

University of Cambridge and Carnegie Mellon University

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 statistics and related fields. We consider the case of completely random measures and obtain a constructive representation for completely random measures 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. As examples, we provide "stick-breaking" representations of the gamma process, the stable process and the beta process.

1. Introduction. Kingman’s [15] 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 NRMs) [7]. 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 [17]. NRMs have similar generic properties [10].

A completely random measure on a space $\Omega_\theta$ can be represented, in a sense to be made precise in Sec. 2, by a discrete random measure $\xi_r$ on $\Omega_\theta$, and hence as

$$\xi_r(\cdot) = \sum_{k=1}^{\infty} S_k \delta_{\Theta_k}(\cdot).$$

The random variables $S_k$ and $\Theta_k$ take values in $\mathbb{R}_+$ and $\Omega_\theta$, 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 $\Theta_k$ can be generated from two sequences of i.i.d. random variables in a simple procedure known as stick-breaking [9, 20]. Similar representations are known for the beta process [2, 18, 21]. In

*Research supported by EPSRC grant EP/F028628/1.
the following, we derive an analogous constructive representation for completely random measures, which represents the variables $S_k$ and $\Theta_k$ 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 [4].

2. Results. Let $\Omega_\theta$ be a Polish space with Borel sets $\mathcal{B}(\Omega_\theta)$, and let $\mathcal{M}_+(\Omega_\theta)$ be the set of measures on $\Omega_\theta$. Let $\xi$ be a completely random measure on $\Omega_\theta$, that is, a $\mathcal{M}_+(\Omega_\theta)$-valued random variable for which $\xi(A) \perp \xi(A')$ whenever $A, A' \in \mathcal{B}(\Omega_\theta)$ are disjoint sets. According to Kingman [15, Theorem 1], $\xi$ admits a unique decomposition

$$\xi = \xi_d + \xi_f + \xi_r$$

into a deterministic (non-random) measure $\xi_d$, a random, purely atomic measure $\xi_f$ with fixed atom locations, and a random discrete measure $\xi_r$. The random component $\xi_r$ can be described by a Poisson process with mean measure $\mu_\xi$ on the Borel sets $\mathcal{B}(\Omega_\theta \times \mathbb{R}_+)$ [16, §8]. For $\Omega_\theta = \mathbb{R}_+$, this process is a pure-jump Lévy process, and slightly abuse terminology and refer to the measure $\nu_\xi(dt, dc) := \mu_\xi(dt \times dc)$ as the Lévy measure of $\xi$, regardless of the choice of $\Omega_\theta$.

Notation. If $Y$ is a subspace of a Euclidean space, $\lambda_Y$ denotes Lebesgue measure on $Y$. The cumulative distribution function of a measure $\rho$ 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).

2.1. Poisson representation of CRMs. The class of completely random measures to which our results are applicable is characterized by the following conditions.

Definition 2.1. Say that a CRM $\xi$ on $\Omega_\theta$ is nice if it satisfies:

(C1) It has no deterministic component, $\xi = \xi_f + \xi_r$.
(C2) It is $\Sigma$-finite in the sense of Kingman [16, §8.1]: There is a countable partition of $\Omega_\theta$ into disjoint sets $D_j$ with $P\{\xi(D_j) < \infty\} > 0$ for all $j$.
(C3) There are no jumps of size 0, that is, $\nu_\xi(\Omega_\theta, \{0\}) = 0$.
(C4) The Lévy measure $\nu_\xi$ of $\xi$ satisfies $\nu_\xi(\Omega_\theta, (s, \infty)) < \infty$.

Conditions (C1)–(C3) entail no loss of generality: (C1) merely simplifies notation. (C2) is generally assumed for CRMs [see e.g. 15, 16], and (C3) ensures that each jump location in the representation (2.16) of $\xi$ below corresponds to an
actual jump. The only critical condition is therefore (C4), which ensures that the tail $T$ of the Lévy measure, defined in (2.2) below, is well-defined. For some types of completely random measures, (C4) requires the domain $\Omega_\theta$ to be restricted to a compact subset of the possibly non-compact space $\Omega_\theta$.

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

\[
T : \mathbb{R}_+ \to \mathbb{R}_+ \\
s \mapsto T(s) := \nu_{\xi}(\Omega_\theta, (s, \infty))
\]

and by $\{\theta_1, \theta_2, \ldots \} \subset \Omega_\theta$ the set of fixed jump locations.

(i) There is a probability kernel $p : \Omega_\theta \to \mathcal{M}(\Omega_\theta)$ such that

\[
(2.3) \quad \xi(\cdot) = \xi_f(\cdot) + \xi_r(\cdot) + \sum_i J_i \delta_{\theta_i}(\cdot) + \sum_k T^{-1}(U_k) \delta_{V_k}(\cdot)
\]

where $U_k \sim \Pi(\lambda)$ is a unit rate Poisson process on $\Omega_\theta = \mathbb{R}_+$ and $V_1, V_2, \ldots$ are independent random variables with $V_k \sim p(., T^{-1}(U_k))$. The random variables $J_i$ are independent of each other and of $\xi_r$.

(ii) The regular conditional probability $p$ is unique up to equivalence and determined by

\[
(2.4) \quad \nu_{\xi}(A, B) = \int_B p(A, s) d\nu_{\xi}(\Omega_\theta, s)
\]

even if $\nu_{\xi}(\Omega_\theta, \cdot)$ is not $\sigma$-finite.

If $\nu_{\xi}(\Omega_\theta, \cdot)$ is $\sigma$-finite, (2.4) simply states that $p$ is given by the densities

\[
(2.5) \quad p(A, s) := \frac{d\nu_{\xi}(A, \cdot)}{d\nu_{\xi}(\Omega_\theta, \cdot)}(s) \quad \text{for } A \in \mathcal{B}(\Omega_\theta).
\]

The intuition underlying Theorem 2.2 is straightforward: As a completely random measure, $\xi_r$ can represented by a Poisson process with mean measure $\mu_{\xi}$ (see [16, §8] for a detailed discussion). Specifically,

\[
(2.6) \quad \xi_r \overset{d}{=} \sum_i S_i \delta_{\Theta_i} \quad \text{if and only if} \quad (S_i, \Theta_i) \sim \Pi(\mu_{\xi}).
\]

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 [16, Chapter 2.3], the image of the Poisson process $\Pi(\gamma)$ under $\phi$ satisfies

\[
(2.7) \quad \phi(\Pi(\gamma)) = \Pi(\phi(\gamma))
\]
for any \( \sigma \)-finite measure \( \gamma \) on \( \Omega_\theta \times \mathbb{R}_+ \). We can thus posit a simple mean measure \( \gamma \)—for example Lebesgue measure, if \( \Omega_\theta \) is Euclidean—and reduce the representation of \( \xi_r \) to a transformation of \( \gamma \). To determine this transformation, the equation

\[
\nu_\xi = \phi(\gamma) \quad \text{s.t. } \phi \text{ measurable}
\]

has to be solved for \( \phi \). Problems of the form (2.8) are known as transport problems in applied analysis [23]. In one dimension, these problems admit a simple solution: If \( \nu \) and \( \gamma \) are measures on the real line with distribution functions \( F_\nu \) and \( F_\gamma \), respectively, then the transport problem \( \nu = \phi(\gamma) \) is obviously solved by

\[
\phi := F_\nu^{-1} \circ F_\gamma.
\]

If \( \gamma \) is in particular Lebesgue measure, then \( \phi = F_\nu^{-1} \). Theorem 2.2 substitutes the tail \( T \) for the distribution function \( F_\nu \), since \( F_\nu \) 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 \( \mu_\xi \) into the pair \((p, \nu_\xi(\Omega_\theta, .))\), where \( p \) is a conditional probability on \( \Omega_\theta \) given \( S \in \mathbb{R}_+ \), and \( \nu_\xi(\Omega_\theta, .) \) is a measure on \( \mathbb{R}_+ \). The transport problem is then solved only on \( \mathbb{R}_+ \)—without further assumptions on the structure of \( \Omega_\theta \), a general solution for \( p \) is not feasible. Thus, \( \xi_r \) is sampled by sampling \((S_k, \Theta_k)\) from a Poisson process with mean measure given by \( p(d\theta, T^{-1}(u)) \lambda_{\mathbb{R}_+}(du) \), and only the weights \( S_k \) of \( \xi_r \) are reduced to unit-rate sampling.

2.2. Representation on Euclidean space. If \( \Omega_\theta \) is contained in Euclidean space, it is possible to construct more elaborate couplings and extend the representation in Theorem 2.2 to the variables \( \Theta_k \), thus fully reducing \( \xi_r \) 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 \( \Omega_\theta \), the disintegration approach used above to separate \( \Omega_\theta \) and \( \mathbb{R}_+ \) is now in turn applied repeatedly to \( p(., s) \). Each application separates off a one-dimensional component, to which (2.9) 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 \( p \) has been obtained by application of Theorem 2.2. For \( d = 1, \ldots, D \), denote by \( p_d \) the marginal measure of \( p \) on the subspace \( \mathbb{R}^d = \mathbb{R}^{(1)} \times \cdots \times \mathbb{R}^{(d)} \). Let \( q_d \) be the probability kernel obtained by disintegrating \( p_d \) with respect to \( p_{d-1} \). That is,

\[
p_d(d\theta_1 \cdots d\theta_d, s) = q_d(d\theta_d | \theta_1, \ldots, \theta_{d-1}, s)p_{d-1}(d\theta_1 \cdots d\theta_{d-1}, s).
\]
As above, let $F_q \theta_d(\cdot | \theta_1, \ldots, \theta_{d-1}, s)$ denote the cumulative distribution function of the one-dimensional measure $q_d(\cdot | \theta_1, \ldots, \theta_{d-1}, s)$.

**Theorem 2.3 (Successive subspace sampling).** Let $\xi$ be a nice completely random measure on $\Omega_\theta$, and let $T$ be defined as in (2.2). For unit-rate Poisson samples

$$\begin{align*}
(U_1, W^{(1)}_k, \ldots, W^{(d)}_k) &\sim \Pi(\lambda_{\mathbb{R}_+} \otimes \lambda^{D}_{[0,1]}),
\end{align*}$$

define

$$\begin{align*}
S_k := T^{-1}(U_k) \quad \text{and} \quad \Theta^{(d)}_k := F_{q_d}^{-1}(W^{(d)}_k | \Theta^{(1)}_k, \ldots, \Theta^{(d-1)}_k, S_k).
\end{align*}$$

Then the purely random component $\xi_r$ of $\xi$ is distributed as

$$\begin{align*}
\xi_r &:= \sum_{k=1}^{\infty} S_k \delta(\Theta^{(1)}_k, \ldots, \Theta^{(d)}_k).
\end{align*}$$

This reduction to the one-dimensional solution (2.9) by successive disintegration is an example of a general approach to the determination of couplings on Euclidean spaces, due originally to Rosenblatt [19].

**Remark 2.4.** In (2.5), the disintegration $p$ 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 $p_d$ always describes a probability measure. The probability kernel $q_d$ in Theorem 2.3 can therefore be represented as densities

$$\begin{align*}
q_d(A | \theta^{(1)} \ldots, \theta^{(d-1)}, s) &= \frac{P_d(d\theta^{(1)} \ldots d\theta^{(d-1)} \times A \times ds)}{P_d(d\theta^{(1)} \ldots d\theta^{(d-1)} \times \mathbb{R}^d \times ds)},
\end{align*}$$

in direct analogy to (2.5).

**2.3. Representation on the line.** If $\Omega_\theta$ is in particular an interval in $\mathbb{R}_+$, the collection of functions $F_{q_d}^{-1}$ reduces to $F_{q_d}^{-1}(w | s) = (p([0, \cdot], s)^{-1})(w)$. Therefore, nice completely random measures on the positive reals can be represented as follows:

**Corollary 2.5 (Ferguson and Klass [4]).** Let $\xi$ be a nice completely random measure on $\Omega_\theta = \mathbb{R}_+$ or $\Omega_\theta = [0, \theta_{\text{max}}]$, where $\theta_{\text{max}} \in \mathbb{R}_+$. Denote the inverse of the distribution function $\theta \mapsto p([0, \theta], s)$ by

$$\begin{align*}
m : [0, 1] \times \mathbb{R}_+ &\rightarrow \Omega_\theta \\
(w, s) &\mapsto m(w, s) := (p([0, \cdot], s)^{-1})(w).
\end{align*}$$
Then

\[ \xi(\cdot) = \xi_f(\cdot) + \xi_r(\cdot) \]

\[ \overset{d}{=} \sum_i J_i \delta \theta_i(\cdot) + \sum_k T^{-1}(U_k) \delta m(W_k, T^{-1}(U_k))(\cdot) \]

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

\[ \phi(u, w) = (T^{-1}(u), m(w, T^{-1}(u))) . \]

**Remark 2.6 (Distribution of \(\xi_f\)).** If the distribution \(\xi_r\) serves as a prior distribution in a nonparametric Bayesian model, fixed atoms arise in the posterior, where the locations \(\theta_k\) in Theorem 2.2 correspond to observations or latent observations. The distributions of the random variables \(J_i\) can then be derived explicitly from \(\nu_\xi\) [17]. For Bayesian models with a conjugate posterior, the representation of \(\xi_f\) can be absorbed into that of \(\xi_r\), by adding a suitable atomic measure to the Lévy measure \(\nu_\xi\).

3. Examples.

3.1. The homogeneous case. In analogy to Lévy processes, a CRM is called *homogeneous* if the measure \(\nu_\xi\) factorizes as \(\nu_\xi(d\theta, ds) = H_0(d\theta) \nu_\xi^{(s)}(ds)\). Consequently, the conditional probability \(p\) of \(\Theta\) in Theorem 2.2 becomes independent of \(S\), and hence \(p(A, s) = H_0(A)/H_0(\Omega_\theta)\).

**Example 1 (Gamma CRM).** The *gamma CRM* is the completely random measure given by

\[ \nu_\xi(d\theta, ds) = s^{-1} e^{-cs} ds H_0(d\theta) , \]

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

\[ T(s) = H_0(\Omega_\theta) E_1(cs) . \]

**Example 2 (Stable CRM).** For \(\alpha \in (0,1)\), the CRM with Lévy measure

\[ \nu_\xi(d\theta, ds) = \frac{\alpha}{\Gamma(1 - \alpha)s^{1+\alpha}} ds H_0(d\theta) . \]

is called an \(\alpha\)-stable CRM. The tail is given by

\[ T(s) = H_0(\Omega_\theta)(\Gamma(1 - \alpha)z^\alpha)^{-1} . \]
3.2. An inhomogeneous case: Beta processes. The beta process, introduced by Hjort [8], is a CRM with
\[ \nu_\xi(d\theta, ds) = c(\theta)s^{-1}(1-s)^{c(\theta)-1}dsdH_0(\theta), \]
where \( H_0 \) is a totally finite measure on \( \Omega_\theta \) 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 \( \Omega_\theta = [0, \theta_{\text{max}}) \) 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 [24] describe and approximate method for evaluating the degenerate incomplete beta function. For certain choices of \( c \) and \( H_0 \), \( p([0, \theta], s) \) can be obtained analytically:

**Example 3** \( (c(\theta) = \exp(-H_0(\theta))) \). Consider the beta CRM with \( c(\theta) = \exp(-H_0(\theta)) \). Then
\[
p([0, \theta], s) = \frac{1 - s - (1-s)^{\exp(-H_0(\theta))}}{1 - s - (1-s)^{\exp(-H_0(\theta_{\text{max}})))}},
\]
and we can obtain \( p([0, \cdot], s)^{-1}(u) \) if \( H_0 \) is invertible.

Another application of the beta process derives from the Indian buffet process [6], 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 process which is given by \( c(\theta) = 1 \) and hence homogeneous [22].

**Example 4** \( (c(\theta) = 1) \). The Lévy measure reduces to \( \nu_\xi(d\theta, ds) = \frac{ds}{s}dH_0(\theta) \), and the tail function is
\[ T(s) = -H_0(\Omega_\theta) \log(z). \]
Let \( \tilde{U}_1 < \tilde{U}_2 < \ldots \) 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 \( \text{Expon}(1) \), therefore \( \exp\{-\tilde{U}_{n+1} + \tilde{U}_n\} \) is distributed according to \( \text{Uniform}(0,1) \) and \( \exp\{(\tilde{U}_n - \tilde{U}_{n+1})/H_0(\Omega_\theta)\} \) is distributed according to \( \text{Beta}(\alpha, 1) \). The strictly-ordered atom sizes \( s_1 > s_2 > \ldots \) of a beta CRM with \( c(\theta) = 1 \) can therefore be generated as
\[ s_n = \prod_{i=1}^{n} b_n \quad \text{with} \quad b_n \sim \text{Beta}(H_0(\Omega_\theta)). \]
Equation (3.7) is precisely the stick-breaking construction of the beta process derived by Teh et al. [21].

4. Related work. The notion of a completely random measure, due to Kingman [15], and has been studied in a variety of contexts in Bayesian statistics (see e.g. the recent surveys by Lijoi and Prünster [17] and by Jordan [11]).

On $\Omega = \mathbb{R}_+$, the independence property $\xi(A) \perp \xi(A')$ implies that the random component of a completely random measure $\xi$ is representable as a stochastic process with independent increments. The random measure $\xi_r$ is therefore equivalent to a (possibly non-stationary) pure-jump Lévy process. In particular, the marginal distribution of the jump sizes is infinitely divisible. Khintchine [14, Hauptsatz III] shows that infinitely divisible laws admit representations of the form (2.16). Ferguson and Klass [4] 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.

The proof of Ferguson and Klass [4] implicitly assumes the marginal Lévy measure $\nu_\xi(\Omega_\theta, \cdot)$ to be $\sigma$-finite (cf. the Radon-Nikodym derivative defined on p. 1636 of [4]). Though their result carries over immediately to completely random measures on the interval, the $\sigma$-finiteness assumption excludes some important CRMs [16, Chapter 9.4]. Representations of the form (2.16), usually without explicit transformations, exist more generally for exchangeable increment processes [12, Theorem 16.21], which is closely related to the fact that such processes can be characterized as mixtures of Lévy processes [13, Theorem 1.19].

Kingman [16, 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 $\nu_\xi(\Omega_\theta, \cdot)$ is $\sigma$-finite. Thus, the beta process of Hjort [8] is a marked Poisson process; the gamma process is not. Motivated by applications in nonparametric statistics, the beta process has recently received much attention in statistics and computer science. Analogues of the stick-breaking construction of the Dirichlet process have been derived for the beta process by Teh et al. [21] (cf. Sec. 3), and more recently by Paisley et al. [18] and Broderick et al. [2], who both emphasize the Poisson representation.

5. Proofs. Verifying the existence of the disintegration $(p, \nu_\xi(\Omega_\theta, \cdot))$ raises some technical issues, since the measure $\nu_\xi(\Omega_\theta, \cdot)$ is not generally $\sigma$-finite. The intuition is, once again, very simple: If $\nu_\xi(\Omega_\theta, \cdot)$ is $\sigma$-finite, $p(A, \cdot)$ is given by the density of $\nu_\xi(A, \cdot)$ with respect to $\nu_\xi(\Omega_\theta, \cdot)$, and it is easy to see that $p(\emptyset, s) = 0$, that $p(\Omega_\theta, s) = 1$, and that $A \mapsto p(A, s)$ is increasing in $A$. In the general case, the analogous result is expressed by the following lemma.
**Lemma 5.1.** Let $\xi$ be a nice completely random measure on $\Omega_\theta$ whose Lévy measure $\nu_\xi$ satisfies $\nu_\xi(\Omega_\theta, (s, \infty)) < \infty$. Then there is a probability kernel $p : \mathbb{R}_+ \to M(\Omega_\theta)$ satisfying

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

Each function $s \mapsto p(A, s)$ is uniquely determined up to a $\nu_\xi(\Omega_\theta, \cdot)$-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 $\Sigma$-finite CRM $\xi$ 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,

$$(5.2) \quad \xi = \xi_f + \xi_r = \sum_i J_i \delta_{\theta_i} + \sum_k S_k \delta_{\Theta_k},$$

where $\theta_i$ are fixed atoms, the random variables $J_i$ are mutually independent and do not depend on $\xi_r$, and $\Theta_k, S_k$ follow a Poisson process $\Pi(\mu)$ [3, Theorem 10.1.III]. Since the existence of $p$ and Theorem 2.2(ii) follow from Lemma 5.1, what remains to be shown is that $\sum S_k \delta_{\Theta_k} \overset{d}{=} \sum T^{-1} (U_k) \delta_{\Theta_k}$.

To this end, consider first the measure $\nu_s(\cdot) := \nu_\xi(\Omega_\theta, \cdot)$ on jump sizes. Since $\xi$ is $\Sigma$-finite, the tail $T$ of $\nu_s$ is finite on $(0, \infty)$ [16]. It is straightforward to verify, for Lebesgue measure $\lambda$ on $\mathbb{R}_+$,

$$(5.3) \quad \lambda(T([a, b])) = \nu_s[a, b]$$

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

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

$$(5.4) \quad \mu(A \times B) = \int_B p(A, s)\nu_s(ds) = \int_B p(A, s)[\lambda^{-1} \lambda](ds)$$

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

$$(5.5) \quad p(A, s) = P[\Theta \in A| S = s]$$

under the law of the Poisson process, and the proof is complete. $\square$

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 $\Theta_k = \theta_k$ are sampled as in (2.12). By construction, each $\Theta_k$ has distribution

\[
\text{Law}(\Theta_k) = q(k | \theta_k, \ldots, \theta_k, s_k).
\]

Since $q_1(k | s_k) = p_1(k | s_k)$, we have

\[
q_D(d\theta(1) | \theta_k, \ldots, \theta_k, s_k) \cdots q_2(d\theta(2) | \theta_k, s_k) \cdots q_3(d\theta(3) | s_k) = p(d\theta(1) \cdots d\theta(D), s_k)
\]

and the joint law of $\Theta_k = (\Theta_k(1), \ldots, \Theta_k(D))$ is thus $\text{Law}(\Theta_k) = p(\ldots, s_k)$. An application of Theorem 2.2 yields the representation (2.13) of $\xi_r$.

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 $\nu(\Omega, \cdot)$ is not in general $\sigma$-finite.

Lemma 5.2 (Generalized Radon-Nikodym Theorem [5, 232E and 232B(b)]). Let $\nu, \nu'$ be measures on a measurable space $(X, A)$. There is a measurable function $f : X \to \mathbb{R}^+$ satisfying $\nu(A) = \int_A f d\nu'$ for all $A \in A$ if and only if:

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

In this case, $f$ is uniquely determined $\nu'$-a.e.

If $\nu'$ is $\sigma$-finite, absolute continuity implies condition (ii), and the lemma reduces to the Radon-Nikodym theorem.

Proof of Lemma 5.1. For the proof, abbreviate $\mu_A := \nu(\Omega, \cdot)$. We proceed in two steps:

Step (1). We first show that, for every $A \in B(\Omega, \cdot)$, there exists a measurable function $p(A, \cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ which satisfies (5.1). By Lemma 5.2, this is the case if the measures $\mu_A$ and $\nu(\Omega, \cdot)$ satisfy conditions (i) and (ii) of Lemma 5.2. Absolute continuity clearly holds since $\nu(A, D) \leq \nu(\mathbb{R}^+, D)$ for every $D \in B(\mathbb{R}^+)$ by construction. To verify condition (ii), observe that (2.5) implies $\mu_A(\varepsilon, \infty) < \infty$ 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 $\mu_A = \mu_A^\uparrow + \mu_A^\downarrow$ the decomposition of $\mu_A$ into its purely atomic component $\mu_A^\uparrow$ and the atomless measure $\mu_A^\downarrow$. Suppose $\mu_A(D) > 0$ as in (ii). We distinguish two cases:

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

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

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

\[
\int_B p(\cup A_n, s) \nu(\Omega, ds) = \sum_{n=1}^\infty \nu(A_n, B) = \sum_{n=1}^\infty \int_B p(A_n, s) \nu(\Omega, ds).
\]

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

\[
\sum_{n=1}^\infty \int_B p(A_n, s) \nu(\Omega, ds) = \int_B \left( \sum_{n=1}^\infty p(A_n, s) \right) \nu(\Omega, ds),
\]

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

\[
0 = \nu(\emptyset, B) = \int_B p(\emptyset, s) \nu(\Omega, ds)
\]

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

References.