A Monad for Probabilistic Point Processes

Swaraj Dash   Sam Staton

University of Oxford
Oxford, United Kingdom

swaraj.dash@cs.ox.ac.uk   sam.staton@cs.ox.ac.uk

A point process on a space is a random bag of elements of that space. In this paper we explore programming with point processes in a monadic style. To this end we identify point processes on a space X with probability measures of bags of elements in X. We describe this view of point processes using the composition of the Giry and bag monads on the category of measurable spaces and functions and prove that this composition also forms a monad using a distributive law for monads. Finally, we define a morphism from a point process to its intensity measure, and show that this is a monad morphism. A special case of this monad morphism gives us Wald's Lemma, an identity used to calculate the expected value of the sum of a random number of random variables. Using our monad we define a range of point processes and point process operations and compositionally compute their corresponding intensity measures using the monad morphism.

1 Introduction

Point processes (e.g. [5]) are random collections of points. They serve as important tools in statistical analysis, where they are used to study various phenomena in fields as diverse as ecology, astronomy, computational neuroscience, and telecommunications, and in Bayesian analysis, where they are used for probabilistic inference (e.g. [17]). As a simple example, in Figure 1 we illustrate five draws from a Poisson point process on the unit square. A Poisson point process is defined to be one in which the number of points in any two disjoint regions are independent of each other. One of the core tools of point process theory is the notion of intensity measure, which assigns to each region the average number of points that will appear in the region. In a homogeneous Poisson point process like Figure 1, the average number of points in a region is proportional to the area of the region. This is a very simple point process, but an important starting point for many models.

The centerpiece of our categorical analysis of point processes is the space

$$G(B(X))$$

which we now explain.

- X is a measurable space such as the unit square $\mathbb{I}^2$ or the natural numbers $\mathbb{N}$. We work in a category of measurable spaces so that we can discuss probability and integration in both the countable and uncountable settings. (See §2 for details.)
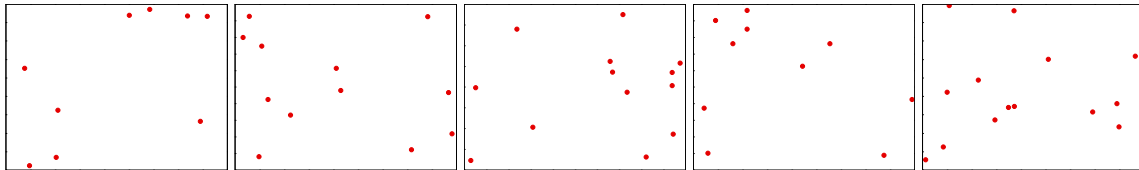

Figure 1: Five draws from a Poisson point process on the unit square with rate 10.0.
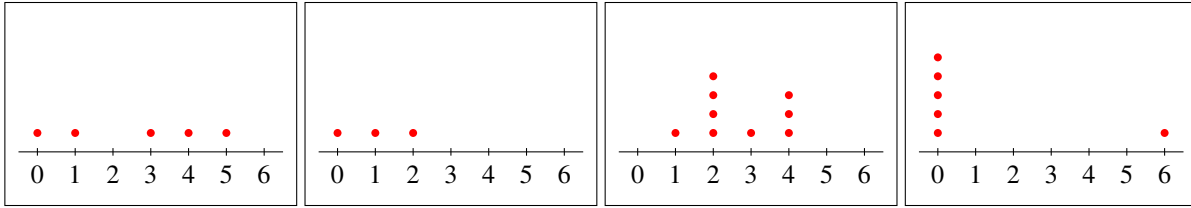
Figure 2: Four draws from a point process on the natural numbers.

- $B(X)$ is a space of all bags of points in X. A bag (aka multiset) is a finite unordered list of elements in X. For example, the first draw in Figure 1 is a bag of 9 points in $X = \mathbb{I}^2$, and the last draw in Figure 2 is a bag of 6 points in $X = \mathbb{N}$, with 5 overlapping points at 0 (multiplicity 5) and one at 6 (multiplicity 1). (See §3 for details.)

- $G(B(X))$ is a space of point processes, i.e. the space of probability measures on the space of bags of points in X. Here, G stands for Giry, who carried out early work on the category theory of spaces of probability measures.

(This space $G(B(X))$ is typically uncountable, but as is common in statistics, it is helpful to run a simulation, outputting a finite number of draws, as in Figures 1, 2, and 4.)

This paper has two main contributions:

- The construction $G(B(X))$ forms a monad (§4). This is useful because it gives us a compositional framework for point processes. One can build point processes (elements $1 \to G(B(X))$) by composing morphisms in the Kleisli category for the monad GB, using a syntax like Haskell's do-notation (§5). Our construction of a monad uses Beck's theory of distributive laws of monads.

- The construction assigning to each point process its intensity measure is a monad morphism (§6, Theorem. 15). Thus, if we build a point process in a compositional way, we can also calculate its intensity in the same compositional way. The key idea here is to regard both G and B as submonads of a monad M of all measures, so that the intensity measure function can be defined as a composite

$$GB \hookrightarrow MM \xrightarrow{\mu} M.$$

Broader context.   The broader context of this work is the idea that category theory can be a language for organizing the structure of statistical models. At one end of the spectrum, this line of work involves a foundational categorical analysis (e.g. [6, 8, 9, 10, 11, 18, 19, 23, 27]). At the other end of the spectrum is 'probabilistic programming', a popular method of statistical modelling using programs (e.g. [13, 30, 25]); in many instances this is functional programming and so heavily inspired by category theory. This full spectrum of work plays a foundational role in probability and statistics, but also addresses a practical grand challenge of interpretability in statistical models, since category theory and programming allow us to clearly organize the structure of complicated statistical models, via composition.

Our work here appears to be the first work on point processes in this context. However, point processes are widely used in statistical models in practice. We mention two programming styles whose relationship to point processes has only recently become evident:

- probabilistic logic programming, in the style of BLOG [31], is about describing random sets of points;

- probabilistic databases are random bags of records [15].

In future work we intend to set out in more detail how our categorical development in this paper can be used to inform the practice of probabilistic logic programming and probabilistic databases in a compositional way.

2 Mathematical preliminaries

We recall basic measure theory, which is the standard formulation for probability theory over uncountable spaces (e.g. [26]).

2.1 Measure theory

Definition 1. A σ-*algebra* on a set X is a nonempty family Σ_X of subsets of X that is closed under complements and countable unions. The pair (X, Σ_X) is called a *measurable space* (we just write X when Σ_X can be inferred from context).

Given (X, Σ_X), a *measure* is a function $\nu : \Sigma_X \to \mathbb{R}_+^\infty$ such that for all countable collections of disjoint sets $A_i \in \Sigma_X$, $\nu(\bigcup_i A_i) = \sum_i \nu(A_i)$. In particular, $\nu(\varnothing) = 0$. It is a *probability measure* if $\nu(X) = 1$. A *pre-measure* is defined to be this additivity condition except that it is not necessarily defined on a σ-algebra.

Examples. The Borel sets form the least σ-algebra $\Sigma_\mathbb{R}$ of subsets of $\mathbb{R}$ that contain the intervals (a, b). On a countable set X, such as $\mathbb{N}$ or the one-point set $\mathbb{1} = \{\star\}$, we will typically consider the *discrete* σ-algebra, which contains all the subsets. In this context, the measures are entirely determined by their values on singletons, $\nu(\{x\})$, and so a measure is the same thing as a function $X \to \mathbb{R}_+^\infty$.

Definition 2. Let (X, Σ_X) and (Y, Σ_Y) be two measurable spaces. A *measurable function* $f : X \to Y$ is a function such that $f^{-1}(U) \in \Sigma_X$ when $U \in \Sigma_Y$. The category **Meas** contains as objects measurable spaces with the morphisms being measurable functions between them.

For any measurable function $f : X \to \mathbb{R}_+^\infty$, and any measure $\nu : \Sigma_X \to \mathbb{R}_+^\infty$, we can define the *Lebesgue integral* or *expected value* $\int_X f \, d\nu$ of f.

Definition 3. A measurable space (X, Σ_X) is a *standard Borel space* if it is either measurably isomorphic to $(\mathbb{R}, \Sigma_\mathbb{R})$ or it is countable and discrete.

2.2 Giry monad

The Giry functor $G : \textbf{Meas} \to \textbf{Meas}$ sends a measurable space to the space of all possible probability measures on it [12]. By slight abuse of notation, let $G(X, \Sigma_X) := (GX, \Sigma_{GX})$. GX is the set of all probability measures $\nu : \Sigma_X \to [0, 1]$ on X equipped with the σ-algebra Σ_{GX} generated by the set of all evaluation maps $\mathrm{ev}_U : GX \to [0, 1]$, sending ν to $\nu(U)$ (where $U \in \Sigma_X$). In other words, it is generated by sets of probability measures $D_I^U = \{\nu : \Sigma_X \to [0, 1] \mid \nu(U) \in I\}$.

$$\Sigma_{GX} = \sigma(\{D_I^U \mid U \in \Sigma_X, I \in \mathscr{B}([0, 1])\})$$

where $\mathscr{B}([0, 1])$ is the Borel σ-algebra and σ is the closure operator which when given a family of subsets generates the required σ-algebra by closing the family under countable unions and complements. The following unit $\eta_X^G : X \to GX$ and multiplication $\mu_X^G : GGX \to GX$ make G into a monad.

$$\eta_X^G(x) = \delta_x = \lambda U. \begin{cases} 1 & \text{if } x \in U \\ 0 & \text{otherwise}^1 \end{cases} \qquad\qquad \mu_X^G(\nu) = \lambda U. \int_{GX} \mathrm{ev}_U \, d\nu$$

Given measurable $f : X \to Y$, the functorial action $Gf : GX \to GY$ sends $v \in GX$ to the *push-forward* measure of v along f, $v \circ f^{-1} \in GY$. An important property of push-forward measures is the change-of-variables formula $\int_Y g \, d(Gf)(v) = \int_X (g \circ f) \, dv$ (where $g : Y \to [0,1]$).

2.3 All-measures monad

The all-measures functor $M : \textbf{Meas} \to \textbf{Meas}$ is defined similarly to the Giry functor. It sends measurable spaces (X, Σ_X) to the space of all measures (MX, Σ_{MX}) where MX is the set of measures $\mu : \Sigma_X \to \mathbb{R}_+^\infty$ and $\Sigma_{MX} = \sigma(\{m_r^U \mid U \in \Sigma_X, r \in \mathbb{R}\})$ with $m_r^U = \{\mu : \Sigma_X \to \mathbb{R}_+^\infty \mid \mu(U) < r\}$. The same unit and multiplication maps make M into a monad. (Warning: The Giry monad is strong and commutative, but the all-measures monad M is not strong, because Fubini's Theorem does not hold for arbitrary measures. So something more refined is needed for functional programming, but that is not an issue for this paper; see e.g. [28] for details.)

3 A monad for finite bags in Meas

In this Section we discuss the construction of the bag monad in the context of measure theory. A bag (aka multiset) is a finite unordered list of elements in some set. For example, the bag $[8,8,5,8,5]$ contains 5 twice and 8 three times, and can also be written as $[5,5,8,8,8]$ since bags are unordered. We begin by recalling the bag endofunctor B in **Set**, which we show to lift to an endofunctor in **Meas** by assigning the σ-algebra Σ_{BX} to the space of bags in § 3.1 (Definition 4). In § 3.2 we prove that B, which is a monad in **Set**, lifts to a monad in **Meas** by showing that the unit and multiplication maps extend to measurable functions. Later in § 4 we will need to define probability measures on BX, i.e., functions $\Sigma_{BX} \to [0,1]$. Defining such functions entails having to define them on arbitrary combinations of unions and intersections of our generating sets A_k^U. We simplify this task by making use of Carathéodory's extension theorem in § 3.3 to show that it suffices to simply define these functions on the generating sets of Σ_{BX} without needing to define them on all of Σ_{BX}.

3.1 The bag functor in Set and Meas

Consider the well-known finite bag endofunctor $B : \textbf{Set} \to \textbf{Set}$ where BX is the set of all finite bags with elements of X. Given a function $f : X \to Y$, the function $Bf : BX \to BY$ applies f component-wise to its argument bag. The natural transformations $\eta_X^B : X \to BX$ and $\mu_X^B : BBX \to BX$ which return the singleton bag and the (multiplicity respecting) union of bags respectively make (B, η^B, μ^B) into a monad.

$$\eta_X^B(x) = [x] \qquad\qquad \mu_X^b([b_1, \ldots, b_n]) = \bigcup_i b_i$$

In order to lift $B : \textbf{Set} \to \textbf{Set}$ to $B : \textbf{Meas} \to \textbf{Meas}$ we equip a σ-algebra Σ_{BX} to our set BX.

Definition 4 (Measurable space of bags). Let BX be the set of bags on the measurable space X. Equip BX with the σ-algebra Σ_{BX} formed by the σ-closure of generating sets $A_k^U = \{b \in BX \mid b$ contains exactly k elements in $U\}$.

$$\Sigma_{BX} = \sigma(\{A_k^U \mid U \in \Sigma_X, k \in \mathbb{N}\})$$

Then (BX, Σ_{BX}) is the measurable space of bags of X.

---

[1]Throughout this paper we use λ-notation to describe functions between sets. Note that the category **Meas** is not cartesian closed and so this is not intended as a formal internal language (c.f. [18]).

It is important to note that the generating set $A_k^U \in \Sigma_{BX}$ contains bags of X of cardinality *at least k*, as each bag in it contains k elements in U, in addition to possible other elements *not* in U. The set A_k^X, on the other hand, contains all the bags of X of cardinality *exactly k* (since X is our universal set). Their intersection $A_n^X \cap A_k^U$ is then the set of *bags of cardinality n with k elements in U*. This can be extended to construct the set of *bags of cardinality n containing k_i elements in U_i* for some family of sets $U_i \in \Sigma_X$, which is then the intersection $A_n^X \cap \left(\bigcap_i A_{k_i}^{U_i} \right)$.

3.2 The bag monad in Meas

Lemma 5. *The unit and multiplication maps $\eta_X^B : X \to BX$ and $\mu_X^B : BBX \to BX$ are measurable.*

Proof. To prove the measurability of these functions it suffices to show that the pre-images of the generating sets A_k^U are measurable. Consider $U \in \Sigma_X$ and some $A_k^U \in \Sigma_{BX}$. The inverse image map of the unit $\eta_X^{B^{-1}}(A_k^U)$ evaluates to $\overline{U}$ (the complement of U) if $k = 0$, U if $k = 1$, and $\varnothing$ otherwise, all of which are elements of Σ_X, and so $\eta_X^{B^{-1}}$ is measurable. We now sketch why $\mu_X^{B^{-1}}(A_k^U) \in \Sigma_{BBX}$ and later show a detailed argument for the case $k = 4$. Call this set $\mathscr{P}$. By definition of the inverse image map, $\mathscr{P}$ is the collection of bags of bags such that the arbitrary union of each bag of bags contains exactly k elements in U. Recall that the set of *bags of cardinality n containing k_i elements in U_i* for some family of sets $U_i \in \Sigma_X$, is given by $A_n^X \cap \left(\bigcap_i A_{k_i}^{U_i} \right)$. By using this technique of describing collections of bags and considering the various partitions of the number k such that the resulting arbitrary union of bags will contain k elements in U, we can express $\mathscr{P}$ entirely using measurable sets, allowing us to conclude that μ_X^B is a measurable function. $\square$

Example 6. Let $\mathscr{P} = \mu_X^{B^{-1}}(A_4^U)$. 4 can be partitioned in five ways: $\{4, 3+1, 2+2, 2+1+1, 1+1+1+1\}$. $\mathscr{P}$ is the set of bags of bags such that the arbitrary union of each bag of bags contains exactly 4 elements in U. We start by considering elements of this set based on their cardinalities.

- There is only one collection of bags of cardinality 1 which are members of $\mathscr{P}$. These are the bags which contain a single bag which in turn contains 4 elements in U. Denote this collection as $\langle 4 \rangle$.

- There are three collections of bags of cardinality 2 which are members of $\mathscr{P}$. The first contains two bags which have 4 and 0 elements in U, the second with 3 and 1 elements in U, and the third with 2 and 2 elements in U, respectively. We write them as $\langle 4, 0 \rangle$, $\langle 3, 1 \rangle$, and $\langle 2, 2 \rangle$.

- Cardinality 3: $\langle 4, 0, 0 \rangle$, $\langle 3, 1, 0 \rangle$, $\langle 2, 2, 0 \rangle$, and $\langle 2, 1, 1 \rangle$.

- Cardinality 4: $\langle 4, 0, 0, 0 \rangle$, $\langle 3, 1, 0, 0 \rangle$, $\langle 2, 2, 0, 0 \rangle$, $\langle 2, 1, 1, 0 \rangle$, $\langle 1, 1, 1, 1 \rangle$.

- Cardinality 5: $\langle 4, 0, 0, 0, 0 \rangle$, $\langle 3, 1, 0, 0, 0 \rangle$, $\langle 2, 2, 0, 0, 0 \rangle$, $\langle 2, 1, 1, 0, 0 \rangle$, $\langle 1, 1, 1, 1, 0 \rangle$. And so on.

Each collection is definable using the generating sets, and the collections of different cardinalities are mutually disjoint. For example, $\langle 3, 1, 0 \rangle = A_3^{BX} \cap (A_1^{\mathbb{B}_3} \cap A_1^{\mathbb{B}_1} \cap A_1^{\mathbb{B}_0})$ and $\langle 2, 1, 1, 0 \rangle = A_4^{BX} \cap (A_1^{\mathbb{B}_2} \cap A_2^{\mathbb{B}_1} \cap A_1^{\mathbb{B}_0})$ where $\mathbb{B}_i = A_i^U$. Finally, $\mathscr{P}$ is the union of all these disjoint collections.

$$\mathscr{P} = \mu_X^{B^{-1}}(A_4^U) = \langle 4 \rangle \cup \langle 4, 0 \rangle \cup \langle 3, 1 \rangle \cup \langle 2, 2 \rangle \cup \langle 4, 0, 0 \rangle \cup \ldots$$

Theorem 7. $(B : \mathbf{Meas} \to \mathbf{Meas}, \eta^B, \mu^B)$ *is a monad.*

Proof. The monad laws hold as in **Set**. Furthermore, η^B and μ^B are measurable (Lemma 5). $\square$

3.3   Defining measures on *BX*

In this Section we construct a *ring* – a set of sets containing the empty set closed under pairwise unions and relative complements – of the generating sets A_k^U of Σ_{BX} in order to invoke Carathéodory's extension theorem. This allows us to define measures by defining them on just specific unions of intersections of the sets A_k^U of Σ_{BX} rather than having to define them on all the arbitrary combinations of unions and intersections of these sets.

Theorem 8 (Carathéodory's extension)**.** *Let $\mathscr{R}$ be a ring and $\nu : \mathscr{R} \to \mathbb{R}_+^\infty$ be a pre-measure. Then there exists a measure $\tilde{\nu} : \sigma(\mathscr{R}) \to \mathbb{R}_+^\infty$ such that $\tilde{\nu}(S) = \nu(S)$ for all $S \in \mathscr{R}$.*

We start by defining $\mathscr{R}'$ to be the set of countable intersections of our generating sets above such that their base sets are mutually disjoint.

$$\mathscr{R}' = \left\{ \bigcap_i A_{k_i}^{U_i} \,\middle|\, U_i \in \Sigma_X, U_i\text{'s mutually disjoint}, k_i \in \mathbb{N} \right\}$$

Now define $\mathscr{R}$ to be the closure of $\mathscr{R}'$ under countable unions. The elements of $\mathscr{R}$ are the unions of intersections of certain generating sets. In particular, any $P \in \mathscr{R}$ can be expressed as $P = \bigcup_i \bigcap_j A_{k_{i,j}}^{U_{i,j}}$.

Examples. One example of such a set P is $(A_0^\alpha \cap A_2^\beta) \cup (A_2^\beta \cap A_1^\gamma) \cup (A_2^\alpha \cap A_5^\beta \cap A_2^\delta)$ where $\alpha, \beta, \gamma, \delta \in \Sigma_X$. Note that although α and β, β and γ, and α, β, and δ are all mutually disjoint (by definition of $\mathscr{R}'$), it is still possible for α and γ to overlap. Using standard set theoretic identities we can redefine P in terms of $\alpha \setminus \gamma$ (instead of just α), β, γ, δ for *all* the base sets across the unions to be mutually disjoint.
Consider also the set $(A_1^\alpha \cap A_1^\beta) \cup (A_0^\alpha \cap A_2^\gamma)$ where $\alpha, \beta, \gamma \in \Sigma_X$ are mutually disjoint. We can rewrite $A_1^\alpha \cap A_1^\beta$ as $\bigcup_i (A_1^\alpha \cap A_1^\beta \cap A_i^\gamma)$ since $\bigcup_i A_i^\gamma$ is simply the universal set. The right half of the set above can similarly be rewritten, enabling us to reformulate it as $\bigcup_i (A_1^\alpha \cap A_1^\beta \cap A_i^\gamma) \cup \bigcup_i (A_0^\alpha \cap A_i^\beta \cap A_2^\gamma)$.

From the two examples above, we can assume without loss of generality that an arbitrary element $P \in \mathscr{R}$ will be of the form $\bigcup_i \bigcap_j A_{k_{i,j}}^{U_j}$ such that all the U_j's are mutually disjoint. Finally, note that any two sets $\bigcap_j A_{k_{m,j}}^{U_j}$ and $\bigcap_j A_{k_{n,j}}^{U_j}$ are disjoint unless for all j, $k_{m,j} = k_{n,j}$. And so, every P can be viewed as the disjoint union of a set of sets. Call this the *disjoint normal form* (it is not unique).

Lemma 9. *$\mathscr{R}$ contains the empty set and is closed under pairwise unions and relative complements.*

Proof. It is clear that $\varnothing \in \mathscr{R}$. The set $\mathscr{R}$ is by definition closed under countable unions, and so is also closed under pairwise unions. Consider P and $Q \in \mathscr{R}$ with their respective disjoint normal forms. Without loss of generality, we can express both P and Q using the same set of mutually disjoint base sets $U_j \in \Sigma_X$. This gives us $P = \bigcup_i \bigcap_j A_{a_{i,j}}^{U_j}$ and $Q = \bigcup_i \bigcap_j A_{b_{i,j}}^{U_j}$. Since both P and Q have been formed by taking the unions of a common set of disjoint sets belonging to $\mathscr{R}'$, their difference $P \setminus Q$ can also be expressed at the disjoint union of sets belonging to $\mathscr{R}'$ and so $P \setminus Q \in \mathscr{R}$. $\square$

Having shown $\mathscr{R}$ to be a ring, we have by Carathéodory's extension theorem that any pre-measure defined on $\mathscr{R}$ extends to a measure defined on $\sigma(\mathscr{R})$. And so, in order to define a measure on *BX*, it will suffice to define it on sets $\bigcup_i \bigcap_j A_{k_{i,j}}^{U_j}$. We use this fact in the next Section.

4 Point process monad

A point process on a space X is a probability measure on bags of X. By composing the Giry and bag monads we can define GBX to be the space of point processes on X. In other words, a point process $\alpha \in GBX$ is a probability measure $\alpha : \Sigma_{BX} \to [0,1]$ assigning probabilities to measurable subsets of bags A_k^U. *The probability of observing k points in the region U of the point process α is then $\alpha(A_k^U)$.*

Earlier we showed that G and B both form monads. It is well-known that the composition of two monads does not automatically yield a new monad. In this Section we prove that the composite functor GB admits a monadic structure by defining the natural transformation $l : BG \to GB$, called the *distributive law* [1] of G over B, such that the following identities hold.

$$\textbf{(Triangle I)} \quad l \circ B\eta^G = \eta^G B \qquad\qquad\qquad G\eta^B = l \circ \eta^B G \quad \textbf{(Triangle II)}$$

$$\textbf{(Pentagon I)} \quad l \circ B\mu^G = \mu^G B \circ Gl \circ lG \qquad G\mu^B \circ lB \circ Bl = l \circ \mu^B G \quad \textbf{(Pentagon II)}$$

This distributive law l then induces the GB monad with the unit defined as the horizontal composition $\eta^G * \eta^B$, and the join defined as the composition of the horizontal composition $\mu^G * \mu^B$ with GlB.

$$\eta^{GB} : 1 \xrightarrow{\eta^G * \eta^B} GB \qquad \text{and} \qquad \mu^{GB} : GBGB \xrightarrow{GlB} GGBB \xrightarrow{\mu^G * \mu^B} GB$$

4.1 Distributive law

The distributive law $l_X : BGX \to GBX$ is a function from bags of probability measures to probability measures on bags. We showed in § 3.3 that in order to define a measure on Σ_{BX} it suffices to define a pre-measure on sets of the form $\bigcup_i \bigcap_j A_{k_{i,j}}^{U_j}$. Carathéodory's extension theorem ensures this pre-measure extends to a measure.

In the definition that follows, we consider a bag of probability measures $[v_1, \ldots, v_n] \in BGX$ and $\bigcup_i \bigcap_j A_{k_{i,j}}^{U_j} \in \Sigma_{BX}$. We define the application of $l[v_1, \ldots, v_n]$ to this disjoint union of intersections as the sum of products of $l[v_1, \ldots, v_n](A_{k_{i,j}}^{U_j})$. Each of these sub-terms is in turn defined as the push-forward of the product measure along K_n, a function mapping n-tuples to bags of cardinality n. An example follows in (1).

$$l[v_1, \ldots, v_n] \left(\bigcup_i \bigcap_j A_{k_{i,j}}^{U_j} \right) \quad \overset{\text{def}}{=} \quad \sum_i \prod_j l[v_1, \ldots, v_n] \left(A_{k_{i,j}}^{U_j} \right)$$

$$\overset{\text{def}}{=} \quad \sum_i \prod_j GK_n(\otimes_i v_i) \left(A_{k_{i,j}}^{U_j} \right)$$

where $K_n : Y^n \to B_n Y$ is the measurable function that sends n-tuples to bags of cardinality n ($B_n Y \subseteq BY$). (K_n is measurable since $K_n^{-1} : \Sigma_{B_n Y} \to \Sigma_{Y^n}$ sends sets A_k^U to their corresponding disjoint unions of n-products of U and $\overline{U}$. For example, $K_2^{-1}(A_1^U) = U \times \overline{U} \uplus \overline{U} \times U$.) In the definition above Y has been instantiated to be GX.

Intuition: the term $l[v_1, \ldots, v_n]$ is a point process where the probability of observing k points in some region $U \in \Sigma_X$ is the probability of observing a total of k points landing in U after independently sampling a point each from all the v_i's $\in GX$. The following example calculation of $l[v_1, v_2](A_1^U)$ confirms this idea.

$$l[v_1, v_2](A_1^U) = GK_2(v_1 \otimes v_2)(A_1^U) = (v_1 \otimes v_2)(K_2^{-1}(A_1^U))$$
$$= (v_1 \otimes v_2)(U \times \overline{U} \uplus \overline{U} \times U) = v_1(U)v_2(\overline{U}) + v_1(\overline{U})v_2(U). \tag{1}$$

Note that $l[v_1, \ldots, v_n](A_k^U) = 0$ for $k > n$. Observe that $\sum_i l[v_1, \ldots, v_n](A_i^U) = 1$.

Showing that l is measurable is a routine calculation, and is made simpler with the knowledge that sets of constant-cardinality bags are measurable.

We noted earlier that the sets A_m^U contain bags of varying cardinalities. In the following lemma we show that the measure $l[v_1, \ldots, v_n]$ acts *only* on the subset of these sets with cardinality n. This result is simple yet very useful in providing an intuitive understanding of the distributive law, and is instrumental in proving the second pentagon identity.

Lemma 10. *For $[v_1, \ldots, v_n] \in BGX$, $l[v_1, \ldots, v_n](A_m^X \cap W) = l[v_1, \ldots, v_n](W)$ if $m = n$ and 0 if $m \neq n$.*

Proof. Consider $l[v_1, \ldots, v_n](A_m^X)$. Using the definition of l this probability can be expressed as the sum of products of a combination of $v_i(X)$ and $v_j(\overline{X})$ terms (where i, j range from 1 to n). Unless $m = n$, each summand will contain at least one factor with $v_j(\overline{X}) = v_j(\varnothing) = 0$, nullifying the entire sum. It is only non-zero when $m = n$, representing the probability of observing n points in the entire space, which has probability 1. The measure of any set $A_n^X \cap W$ is then just the measure of W. $\square$

Theorem 11. $(GB : \mathbf{Meas} \to \mathbf{Meas}, \eta^{GB}, \mu^{GB})$ *is a monad via the distributive law* $l : BG \to GB$.

Proof (sketch). We prove that the four identities for the distributive law hold. The two triangle identities follow from simple algebra. For the first pentagon identity we make use of the change-of-variables formula, as $l[v_1, \ldots, v_n] = GK_n(\bigotimes_i v_i)$ is a push-forward measure, and prove the resulting equality using standard integration identities. For the final pentagon identity we are required to work with the set $\mu_X^{B-1}(A_k^U)$, which we decompose using the method presented in Lemma 5. Invoking Lemma 10 on these constant-cardinality decompositions allows us to simplify the resulting expression by removing sets with measure zero and prove the final equality after some more algebraic manipulations. $\square$

4.2 Unit and bind

The unit η^{GB} returns the deterministic point process $\eta^{GB}(x)$ with the singular point x. When programming with monads it is often convenient to focus on Kleisli composition in a stylized form, using the function $\gg\!=_{GB} : \mathbf{Meas}(X, GBY) \to \mathbf{Meas}(GBX, GBY)$ (pronounced *bind*); we write $\alpha \gg\!=_{GB} f$ for $\gg\!=_{GB} (f)(\alpha)$ [24].

This presents us with useful intuition for programming with point processes. Let X and Y be discrete sets. Then the process of sampling points from $\alpha \gg\!=_{GB} f \in GBY$ can be viewed as the following simulation, illustrated in Figure 3. (1) Sample a bag of points from α. (2) Each point

Figure 3: Sampling from a composite process.

x_i in this bag produces a point process $f(x_i)$ from which we sample a bag of points in Y. (3) A sample of points from the overall point process is the union of these bags of points in Y. This intuition allows us to declaratively program with point processes by being able to define them simply by how they must be simulated. We make extensive use of this intuition in the next Section.
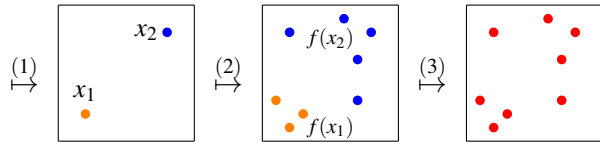
Aside about related work.   A long-term problem in programming semantics has been the combination of probability and non-determinism (e.g. [29]). In that context, it is well known that there is no distributive law between the probability monad and the powerset monad (e.g. [29, 32, 14]). It has recently become well-known that, in the set-theoretic case, it *is* possible to find a distributive law by using a bag monad instead of a powerset monad (e.g. [22, App. A], [4, 32]). This is *not* a sleight of hand, because

point processes are not necessarily to be thought of in terms of computational non-determinism. In fact, in the probability theory literature, it is most common to regard point processes as random bags rather than as random sets.

The bag construction is a free commutative monoid, and the free commutative *co*monoid also plays an important role in the theory of linear logic. Recently, bag-like exponentials have arisen in models of probabilistic linear logic [3, 16]. The precise relationship to our monad and our distributive law remains to be seen.

More broadly, bags, multisets and urns play a fundamental role in statistics and arise at various points in a categorical treatment (e.g. [19, 20, 21]).

5 Examples of point processes via the monad

Probability distributions as point processes   As a first example, we describe probability distributions on the natural numbers as point processes on the singleton space $\mathbb{1} = \{\star\}$, based on the observation that a bag of singletons is a natural number ($B\mathbb{1} \cong \mathbb{N}$). Any probability distribution $d \in G\mathbb{N}$ (so that $\sum_{i=0}^{\infty} d_i = 1$) can be presented as a point process $\underline{d} \in GB\mathbb{1}$ where we observe k copies of $\star$ with probability d_k:

$$\underline{d}\left(A_k^{\{\star\}}\right) := d_k \tag{2}$$

Building compound probability distributions.   Using our monad we define *compound distributions* as point processes on the unit type. A compound probability distribution is the probability distribution of the sum of a number of independent identically-distributed random variables, where the number of terms to be added is itself a random variable. For example, given a random variable $N \sim \text{Poisson}(\Lambda)$ and iid variables X_i, the random variable $Y = \sum_{i=0}^{N} X_i$ forms a compound Poisson distribution.

Recall the behaviour of $\gg\!\!=_{GB}$ on countable sets described in § 4.2. By considering $\underline{N} \in GB\mathbb{1}$ (say, the Poisson distribution) and $\underline{X} \in GB\mathbb{1}$ (the distribution of the iid X_i), we can express compound distributions as:

$$\gamma = \underline{N} \gg\!\!=_{GB} \lambda\star.\, \underline{X} \qquad \in GB\mathbb{1}. \tag{3}$$

A Poisson point process on the unit square   The Poisson point process on the unit square $\mathbb{I}^2$ (Fig. 1) can be simulated by first sampling a random number of points from the Poisson distribution, and then uniformly distributing these points across $\mathbb{I}^2$. Consider again a Poisson distribution $\underline{N} \in GB\mathbb{1}$ as a point process. Now consider the point process $\underline{U} \in GB\mathbb{I}^2$ which returns a single point uniformly distributed in $\mathbb{I}^2$. The Poisson point process π can be built using the monad:

$$\pi = \underline{N} \gg\!\!=_{GB} \lambda\star.\, \underline{U} \qquad \in GB\mathbb{I}^2. \tag{4}$$

Thinning a point process.   Thinning is an operation applied to the points of an underlying point process, where the points are thinned (removed) according to some probabilistic rule. Given some point process $\alpha \in GBX$ and some thinning rule $t : X \to GBX$ such that $t(x)$ probabilistically returns either $[x]$ or $\varnothing$, we can use the monad to build the thinned point process $\alpha' \in GBX$ as

$$\alpha' = \alpha \gg\!\!=_{GB} \lambda x.\, t(x) \qquad \in GBX.$$

Displacing a point process. Displacement is an operation applied to the points of an underlying point process, where the points are independently randomly displaced (translated) according to some distribution. We model this distribution as a single-point point process $\Delta \in GB\mathbb{R}$. The location of this random point is the random displacement distance. For $\alpha \in GB\mathbb{R}$ we simulate the displaced point process $\alpha' \in GB\mathbb{R}$ by sampling an x from α, a displacement distance d from Δ, and then returning the displaced point.

$$\alpha' \;=\; \alpha \gg=_{GB} \lambda x.\, (\Delta \gg=_{GB} \lambda d.\, \eta^{GB}(x+d)) \qquad \in GB\mathbb{R}.$$

Clustered point processes. Clustered point processes are useful in modelling phenomena which involve multiple points spawning from individual seeds, such as clusters of trees, galaxies, or diseases. Informally, a clustered point process is anything built using the monadic bind $\gamma_1 \gg= \lambda x.\gamma_2(x)$, where γ_1 is a point process for the initial seeds, and γ_2 is a point process that grows

Figure 4: Two draws from a clustered point process

from each seed, where the location x of a given seed may be a parameter. For a simple example, consider the point process in Fig. 4 consisting of small square clusters within the unit square. To simulate it we first sample the centres of these clusters from our Poisson point process π on the unit square (4), and for each cluster center we sample another Poisson point process. To sample the second point process, we again sample from another Poisson distribution $\underline{N}'$, whose rate now depends on the provided coordinates of the cluster – the closer to the diagonal, the higher the rate, and we uniformly distribute these points in a small square about this center using $\underline{U}'$, which is a location-dependent and scaled-down modification of $\underline{U}$ introduced earlier. This results a Poisson number of clusters of Poisson processes, with those closer to the diagonal being denser than those farther away.
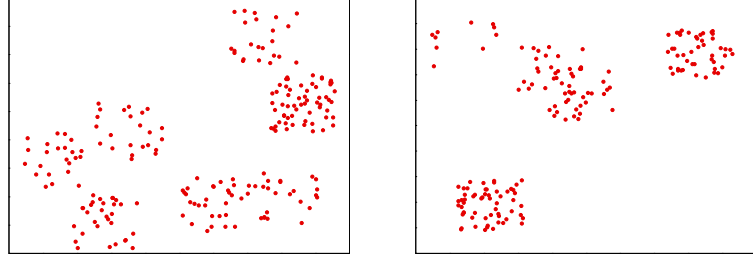
$$\beta \;=\; \pi \gg=_{GB} \lambda(x,y).\, (\underline{N}'(x,y) \gg=_{GB} \lambda \star.\, \underline{U}'(x,y)) \qquad \in GB\mathbb{I}^2.$$

Here, $\underline{N}$ is itself defined using the monad (4). This example is quite simple, but already illustrates that we can use the monad to quickly and clearly compose point processes to build complex statistical models.

6   The intensity measure as a monad morphism

A useful characteristic for describing point processes is the expected number of points in a given region. For example, in Fig. 1 we illustrated the homogeneous Poisson process with rate 10. The expected number of points in any region is proportional to $10a$ where a is the area of the region. More generally, the *intensity measure* of a point process is the measure that assigns to each measurable subset the expected number points in it.

There is a function $\mathbb{E} : GB(X) \to M(X)$ that takes a point process to its intensity measure. In Theorem. 15 we show that this function is actually a monad morphism from the point process monad (§4) to the monad of all measures (§2.3). Thus, if we build a point process using the monadic constructions (for example by composing morphisms in the Kleisli category of GB) then we can immediately read off its

intensity measure in a compositional way. This new result generalizes Wald's lemma (Ex. 17), which is a core result in probability theory.

6.1 Constructing a monad morphism

The expected number of points in a region $U \in \Sigma_X$ of a point process $\alpha \in GBX$ can be given in terms of our generating sets A_k^U for BX, as $\sum_{k=1}^{\infty} k \cdot \alpha(A_k^U)$.

We first show how to understand this in a more abstract way, by injecting both probability measures GX and bags BX each into measures MX, in a measurable and natural way. The injection $i^G : G \to M$ is straightforward because $GX \subseteq MX$. The injection $i^B : B \to M$ sends bags $[x_1, \ldots, x_n]$ to $\sum_{i=1}^{n} \delta_{x_i}$, the multiplicity-respecting sum of Dirac deltas centered around the elements x_i. It is measurable since its inverse image map sends the generating sets $m_r^U \in \Sigma_{MX}$ to $\bigcup_{i=0}^{\lfloor r \rfloor} A_i^U \in \Sigma_{BX}$. The proof that this is injective relies on X being standard Borel. This injection is familiar in point process theory, indeed many authors actually define BX as a space of integer-valued measures in the first place. We can combine the horizontal composition of these two injections $(i^G * i^B)$ with the multiplication of M in order to define $\mathbb{E}$.

$$\mathbb{E} \quad \stackrel{\text{def}}{=} \quad GB \xrightarrow{i^G * i^B} MM \xrightarrow{\mu^M} M$$

In the remainder of this paper we omit $*$ when writing horizontal compositions. This definition of $\mathbb{E}$ does indeed return the intensity measure of a point process:

Lemma 12. *For any point process $\alpha \in GBX$, $\mathbb{E}(\alpha)(U) = \sum_k k \cdot \alpha(A_k^U)$.*

Proof. Consider $\alpha \in GBX$ and $U \in \Sigma_X$. On expanding the horizontal composition $i^G i^B$ and using the change-of-variables formula for pushforward measures we have that

$$\mathbb{E}(\alpha)(U) = \int_{MX} \text{ev}_U \, \mathrm{d}M i_X^B(i_{BX}^G(\alpha)) = \int_{b \in BX} i_X^B(b)(U) \, \alpha(\mathrm{d}b).$$

We separately compute this integral on the disjoint partitions A_k^U ($k \in \mathbb{N}$) of BX. In each partition, the value of $i_X^B(b)(U)$ is equal to k (by definition). This gives us the desired infinite sum of $\sum_k k \cdot \alpha(A_k^U)$. $\square$

To show that $\mathbb{E} : GB \to M$ is a monad morphism we need to prove that

$$\textbf{(Unit)} \ \eta^M = \mathbb{E} \circ \eta^{GB} \qquad \text{and} \qquad \mu^M \circ \mathbb{E}\mathbb{E} = \mathbb{E} \circ \mu^{GB} \ \textbf{(Mult)}.$$

Our main result stems from the fact that l interacts well with i^G and i^B, which we prove next.

Lemma 13. $(\mathbb{E} \circ l =) \ \mu^M \circ i^G * i^B \circ l = \mu^M \circ i^B i^G : BG \to M$.

Proof. (Diagram chasing) Consider $[v_1, \ldots, v_n] \in BGX$ and $U \in \Sigma_X$. Going from BGX to MX along the left edge and applying the resulting map to U gives us $\sum_i v_i(U)$. Along the other edge, making use of Lemma 12, we get $\sum_i i \cdot l[v_1, \ldots, v_n](A_i^U)$. Their equality can be proved by noticing that $l[v_1, \ldots, v_n](A_k^U)$ is simply the coefficient of x^k in the polynomial $P(x) = \prod_i (v_i(\bar{U}) + v_i(U) \cdot x)$. And so equivalently $P(x) = \sum_i l[v_1, \ldots, v_n](A_i^U) \cdot x^i$. The desired equality is then arrived at by taking the derivative of $P(x)$ at $x = 1$. $\square$

$$
\begin{array}{ccc}
BGX & \xrightarrow{\ l_X\ } & GBX \\
{\scriptstyle (i^B i^G)_X} \downarrow & {\scriptstyle (i^G i^B)_X} & \downarrow \\
M^2 X \xrightarrow[\mu_X^M]{} & MX & \xleftarrow[\mu_X^M]{} M^2 X
\end{array}
$$

Lemma 14. $i^G : G \to M$ *and* $i^B : B \to M$ *are monad morphisms.*

Theorem 15. *The intensity measure* $\mathbb{E}: GB \to M$ *is a monad morphism.*

Proof. A simple calculation shows **Unit** to hold. For **Mult** consider the two diagrams below.
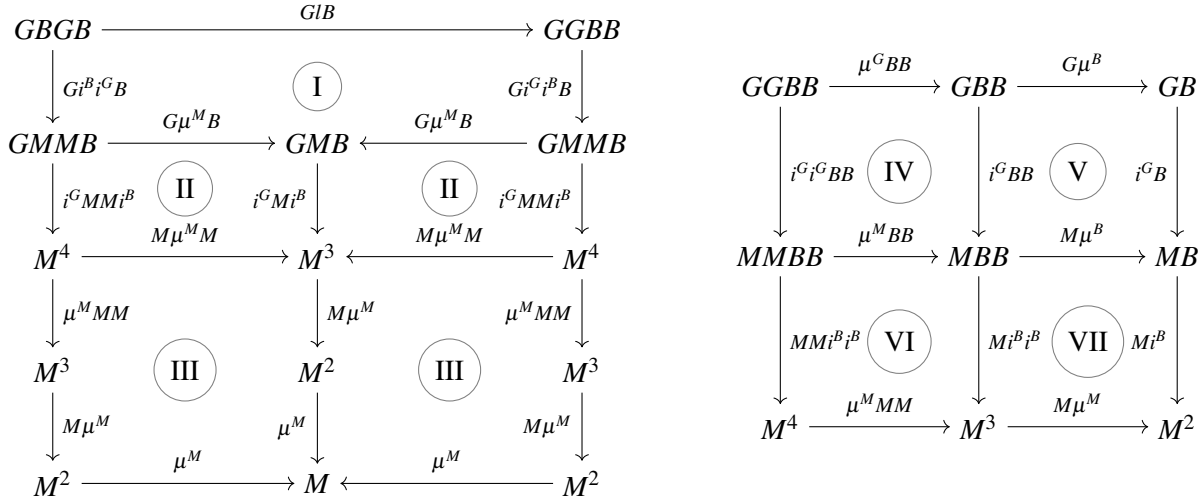
All the sub-diagrams above commute: (I) due to Lemma 13, (II) by naturality, (III) due to associativity of μ^M, (IV) and (VII) due to i^G and i^B being monad morphisms (Lemma 14), and finally (V) and (VI) by naturality. Using the commutative diagrams above we prove the required equality.

$$
\begin{aligned}
\mu^M \circ \mathbb{E}\mathbb{E} \; &: \; GBGB \to M \\
\mu^M \circ \mathbb{E}\mathbb{E} &= \mu^M \circ M\mu^M \circ \mu^M MM \circ i^G MMi^B \circ Gi^B i^G B && \text{(defn. of } \mathbb{E} + \text{naturality)} \\
&= \mu^M \circ M\mu^M \circ \mu^M MM \circ i^G MMi^B \circ Gi^G i^B B \circ GlB && \text{(left diagram)} \\
&= \mu^M \circ M\mu^M \circ \mu^M MM \circ MMi^B i^B \circ i^G i^G BB \circ GlB && \text{(naturality)} \\
&= \mu^M \circ Mi^B \circ i^G B \circ G\mu^B \circ \mu^G BB \circ GlB && \text{(right diagram)} \\
&= \mu^M \circ Mi^B \circ i^G B \circ \mu^{GB} && \text{(defn. of } \mu^{GB}) \\
&= \mathbb{E} \circ \mu^{GB} && \text{(defn. of } \mathbb{E}) \qquad \square
\end{aligned}
$$

6.2   Examples

Example 16. In §5 we simulated a Poisson point process by composing the Poisson distribution with a uniform singleton. We show this has the required intensity measure in a compositional way, using the monad morphism. Let $\underline{N} \in GB\mathbb{1} \cong G\mathbb{N}$ be the Poisson distribution with mean Λ, and let $\underline{U} \in GB\mathbb{I}^2$ be the uniformly distributed single-point process, from §5. The simulated Poisson process is $\pi = (\underline{N} \ggg_{GB} \lambda \star. \underline{U})$, and we have

$$
\begin{aligned}
\mathbb{E}(\pi) &= \mathbb{E}(\underline{N} \ggg_{GB} \lambda \star. \underline{U}) \\
&= \mathbb{E}(\underline{N}) \ggg_M \lambda \star. \mathbb{E}(\underline{U}) && \text{(Theorem 15)} \\
&= \lambda W \in \Sigma_{\mathbb{I}^2}. \mathbb{E}(\underline{N})(\star) \times \mathbb{E}(\underline{U})(W) \\
&= \lambda W \in \Sigma_{\mathbb{I}^2}. \Lambda \times |W| && \in M\mathbb{I}^2
\end{aligned}
$$

In the penultimate step we use the fact that in the discrete case the bind for M, just like for G, computes a weighted sum, which in this case is just a single term. In the final step, we substitute in the intensity

of $\underline{N}$ and the intensity measure of the uniform point process $\underline{U}$, producing the correct intensity. W is a measurable subset of the unit square and $|W|$ is its area.

Example 17 (Discrete Wald's Lemma). We regard arbitrary probability distributions $\underline{N}, \underline{X}$ on the natural numbers as point processes in $GB\mathbb{1}$ via (2). Wald's lemma says that the expected value of the compound distribution (γ, (3)) is the product of the expectations, which is immediate from the fact that $\mathbb{E}$ is a monad morphism:

$$
\begin{aligned}
\mathbb{E}(\gamma) &= \mathbb{E}(\underline{N} \ggg=_{GB} \lambda \star. \underline{X}) \\
&= \mathbb{E}(\underline{N}) \ggg=_M \lambda \star. \mathbb{E}(\underline{X}) \qquad\qquad\qquad \text{(Theorem 15)} \\
&= \lambda \star. \mathbb{E}(\underline{N})(\star) \times \mathbb{E}(\underline{X})(\star) \qquad \in M\mathbb{1}
\end{aligned}
$$

Remark. We remark that a natural transformation in the *opposite* direction ($M^+ \to GB$) has been exhibited in [2], where $M^+(X)$ is the space of finite non-empty measures. This natural transformation takes an intensity measure to the corresponding inhomogeneous Poisson process. Since M^+ is not a monad, it remains to be seen whether this natural transformation can be made into a monad morphism somehow.

Concluding remarks. We have exhibited a monad GB for point processes (§4), and shown that the intensity measure is a monad morphism (§6). This gives a compositional way of building and reasoning about increasingly complicated point processes (§5). This is further evidence towards the claim that applied category theory has the potential to be a useful tool for statistical modelling.

References

[1] J. Beck (1969): *Distributive laws*. In B. Eckmann, editor: *Seminar on Triples and Categorical Homology Theory*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 119–140, doi:10.1007/BFb0083084.

[2] F. Dahlqvist, V. Danos & I. Garnier (2016): *Giry and the Machine*. In: *Proc. MFPS 2016*, pp. 85–110, doi:10.1016/j.entcs.2016.09.033.

[3] F. Dahlqvist & D. Kozen (2020): *Semantics of Higher-Order Probabilistic Programs with Conditioning*. In: *Proc. POPL 2020*, doi:10.1017/S0960129516000426.

[4] F. Dahlqvist, L. Parlant & A. Silva (2018): *Layer by Layer – Combining Monads*. In: *Proc. ICTAC 2018*, doi:10.2168/lmcs-3(4:11)2007.

[5] D. Daley & D. Vere-Jones (2006): *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. Probability and Its Applications, Springer New York.

[6] V. Danos & I. Garnier (2015): *Dirichlet is Natural*. In: *Proc. MFPS 2015, Electr. Notes Theoret. Comput. Sci* 319, pp. 137–164, doi:10.1016/j.entcs.2015.12.010.

[7] S. Dash & S. Staton (2020): *A Monad for Point Processes*. Talk at LAFI 2020.

[8] T. Ehrhard, M. Pagani & C. Tasson (2018): *Measurable cones and stable, measurable functions: a model for probabilistic higher-order programming*. Proc. ACM Program. Lang. (POPL) 2, doi:10.1145/3158147.

[9] T. Fritz (2019): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. arxiv:1908.07021.

[10] T. Fritz, P. Perrone & S. Rezagholi (2019): *The support is a morphism of monads*. In: *Proc. ACT 2019*.

[11] R. Garner (2019): *Hypernormalisation, linear exponential monads and the Giry tricocycloid*. In: *Proc. ACT 2019*.

[12] M. Giry (1982): *A categorical approach to probability theory*. In: *Categorical aspects of topology and analysis (Ottawa, Ont., 1980), Lecture Notes in Mathematics* 915, Springer, Berlin, pp. 68–85, doi:10.1007/BFb0092872.

[13] N.D. Goodman, V.K. Mansinghka, D.M. Roy, K. Bonawitz & J.B. Tenenbaum (2008): *Church: a language for generative models*. In: *Proc. UAI 2008*, pp. 220–229.

[14] A. Goy & D. Petrisan (2020): *Combining probabilistic and non-deterministic choice via weak distributive laws*. In: *Proc. LICS 2020*, doi:10.1017/S0960129505005074.

[15] M. Grohe & P. Lindner (2019): *Probabilistic Databases with an Infinite Open-World Assumption*. In: *Proc. PODS 2019*, pp. 17–31, doi:10.1145/3294052.3319681.

[16] M. Hamano (2019): *A Linear Exponential Comonad in s-finite Transition Kernels and Probabilistic Coherent Spaces*. arxiv:1909.07589.

[17] T. Herlau, M.N. Schmidt & M. Morup (2016): *Completely random measures for modelling block-structured sparse networks*. In: *Proc. NeurIPS 2016*, pp. 4260–4268, doi:10.5555/3157382.3157574.

[18] C. Heunen, O. Kammar, S. Staton & H. Yang (2017): *A convenient category for higher-order probability theory*. In: *Proc. LICS 2017*, IEEE Press, doi:10.1109/LICS.2017.8005137.

[19] B. Jacobs (2019): *Structured Probabilitistic Reasoning*. Draft available from the author's website.

[20] B. Jacobs & S. Staton (2020): *De Finetti's construction as a categorical limit*. In: *Proc. CMCS 2020*.

[21] B. Jacobs (2019): *Learning along a Channel:the Expectation part of Expectation-Maximisation*. In: *Proc. MFPS 2019*, doi:10.1016/j.entcs.2019.09.008.

[22] K. Keimel & G. Plotkin: *Mixed powerdomains for probability and nondeterminism*. arXiv:1612.01005.

[23] P. McCullagh (2002): *What is a statistical model?* Annals of Statistics 30(5), pp. 1225–1310, doi:10.1214/aos/1035844977.

[24] E. Moggi (1991): *Notions of Computation and Monads*. Inf. Comput. 93(1), p. 55–92, doi:10.1016/0890-5401(91)90052-4.

[25] P. Narayanan, J. Carette, W. Romano, C. Shan & R. Zinkov (2016): *Probabilistic inference by program transformation in Hakaru (system description)*. In: *Proc. FLOPS 2016*, Springer, pp. 62–79, doi:10.1007/978-3-319-29604-3_5.

[26] D. Pollard (2001): *A User's Guide to Measure Theoretic Probability*. CUP, doi:10.1017/CBO9780511811555.

[27] A. Simpson (2017): *Probability Sheaves and the Giry Monad*. In: *Proc. CALCO 2017*, doi:10.4230/LIPIcs.CALCO.2017.1.

[28] S. Staton (2017): *Commutative semantics for Probabilistic Programming*. In: *Proc. ESOP 2017, Lect. Notes Comput. Sci.* 10201, Springer, pp. 855–879, doi:10.1007/978-3-662-46669-8_3.

[29] D. Varacca & G. Winskel (2006): *Distributing probability over non-determinism*. Mathematical structures in computer science 16, pp. 87–113, doi:10.1017/S0960129505005074.

[30] F. Wood, J.W. van de Meent & V. Mansinghka (2014): *A new approach to probabilistic programming inference*. In: *Proc. AISTATS 2014*.

[31] Y. Wu, S. Srivastava, N. Hay, S. Du & S.J. Russell (2018): *Discrete-Continuous Mixtures in Probabilistic Programming: Generalized Semantics and Inference Algorithms*. In: *Proc. ICML 2018*, pp. 5339–5348.

[32] M. Zwart & D. Marsden (2019): *No-Go Theorems for Distributive Laws*. In: *Proc. LICS 2019*, doi:10.1109/LICS.2019.8785707.