Element-free probability distributions and random partitions

Victor Blanchi
victor.blanchi@ens.psl.eu
École Normale Supérieure, PSL University
Paris, France

Hugo Paquet
paquet@lipn.fr
Laboratoire d'Informatique de Paris Nord
Villetaneuse, France

ABSTRACT

An "element-free" probability distribution is what remains of a probability distribution after we forget the elements to which the probabilities were assigned. These objects naturally arise in Bayesian statistics, in situations where elements are used as labels and their specific identity is not important.

This paper develops the structural theory of element-free distributions, using multisets and category theory. We give operations for moving between element-free and ordinary distributions, and we show that these operations commute with *iid* sampling.

We then exploit this theory to prove two characterization results, establishing a precise connection between element-free distributions and two key structures in Bayesian nonparametric clustering: exchangeable random partitions, and random distributions parametrized by a base measure.

KEYWORDS

probability theory, category theory, statistics, multisets, partitions

ACM Reference Format:
Victor Blanchi and Hugo Paquet. 2024. Element-free probability distributions and random partitions. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

This paper is about the idea of "element-free" probability theory. We can take any probability distribution and forget the elements to which probabilities have been assigned, keeping only the probability coefficients. For instance, given the distribution on the set $\{a, b, c\}$ assigning $\frac{1}{5}$ to a and b, and $\frac{3}{5}$ to c, if we forget the elements a, b, and c, we are left with coefficients $\frac{1}{5}, \frac{1}{5}, \frac{3}{5}$. These coefficients are unordered and form a multiset (or bag) written $[\frac{1}{5}, \frac{1}{5}, \frac{3}{5}]$. We call this kind of multiset an element-free distribution (§3).

Clearly, the element-free approach involves a significant loss of information. But this is justified in situations where the precise identity of elements is not important. Our main motivation comes from nonparametric Bayesian statistics: element-free distributions are deeply connected to the theory of random exchangeable partitions (*e.g.* the Chinese Restaurant Process) and random probability distributions parametrized by a base measure (*e.g.* the Dirichlet process).

In this paper we develop the theory of element-free distributions. We also explain their canonical status and clarify the connection to partitions and random distributions. We sketch the main ideas in this introduction.

1.1 Element-free sampling and random partitions

We first explain the connection between element-free probability and partitions. The point is that partitions are the element-free counterpart of lists of elements.

Given a list $(x_1, \ldots, x_K) \in X^K$, where X is any set and K a positive integer, we can construct a partition of the set $\{1, \ldots, K\}$ in which i and j are in the same block whenever $x_i = x_j$. For instance, taking $X = \{a, b, c\}$, the list (a, a, b, a, c) induces the partition $\{\{1, 2, 4\}, \{3\}, \{5\}\}$. This partition is an element-free abstraction of the list: the elements themselves have been forgotten but we record the indices at which equal elements appeared.

A key operation in probability and statistics is that of drawing lists of independent and identically distributed (*iid*) samples according to a probability distribution. There is also an element-free version of iid sampling, for generating a partition directly from an element-free distribution. The two operations of forgetting elements and iid sampling commute:

$$\begin{tikzcd}
\text{distribution on } X \arrow[r, "forget"] \arrow[d, "draw"'] & \text{element-free distribution} \arrow[d, "draw"] \\
\text{iid samples} \arrow[r, "forget"'] & \text{iid partition}
\end{tikzcd}$$

A key contribution of this paper is to take this idea further, to incorporate continuous distributions on measurable spaces. In the continuous setting we can prove a new representation theorem for random partitions (§1.4, §8).

1.2 Motivation from Bayesian clustering

To motivate this work we explain the connection to statistical practice. (Our purpose is to clarify the foundations, and we do not propose any new methods, although we briefly discuss connections to probabilistic programming in §9.)

In data analysis, clustering is the problem of appropriately partitioning a dataset $y_1, \ldots, y_K$ into clusters (Figure 1). In the Bayesian approach to clustering, the goal is not to find the "best" partition, but to find a suitable probability distribution over partitions. In *nonparametric* clustering the total number of clusters is not fixed, and grows as more data is included. A Bayesian model should first

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

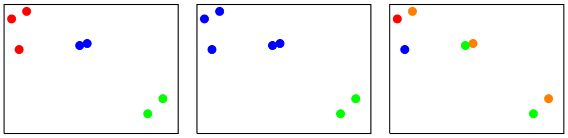

Figure 1: Three different clusterings for a set of points, with clusters represented as colours.

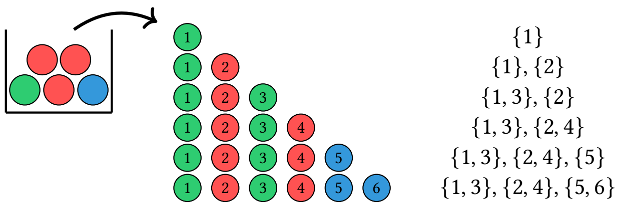

Figure 2: Iteratively sampling a partition from a probability distribution represented as an urn. (Here *forgetting elements* corresponds to forgetting the colours.)

be specified independently of the data, and so should only be concerned with partitions of the indexing set $\{1, \ldots, K\}$.

To generate a random partition, a popular method is to fix a probability distribution over a space whose elements will be used as cluster names, and then sample a name for each index $k \leq K$. This method is illustrated in Figure 2 using an urn of coloured balls. Observe how new clusters are created as k grows.

The specific names (or colours) used for clusters are not important for the partition model: substituting red balls for balls of a new colour would not affect the resulting distribution on partitions. In other words, it is only the element-free distribution that matters.

It makes sense to study this particular clustering method in depth, because it is universal: all well-behaved distributions on partitions can be generated in this way from a *random* element-free distribution. We discuss this further in §1.4.

1.3 Base measures and random distributions

Another key contribution of this paper is to show that, as well as forgetting elements, we can reconstruct them in a principled way. The idea is to sample them from a fixed distribution called a base measure. We show that, if μ is a distribution on a space X, we have a reverse situation:

$$\begin{tikzcd}
\text{distribution on } X \arrow[d, "draw"'] & \text{element-free distribution} \arrow[l, "draw"'] \arrow[d, "draw"] \\
\text{iid samples} & \text{iid partition} \arrow[l, "draw"]
\end{tikzcd}$$

These operations are all probabilistic. Given an element-free distribution, our construction does not give a fixed distribution on X, but a random distribution: a distribution on distributions.

The use of a base measure is common in nonparametric statistics. Indeed a number of key models for random distributions (*e.g.* the Dirichlet or Pitman-Yor processes [11, 28]) admit a base measure as a parameter, precisely because these models are only concerned with the element-free part of the distributions they generate. (The popular *stick-breaking* methods [31] can be regarded as generating element-free distributions.)

In §8 we study random distributions parametrized by a base measure, and give a new representation theorem (Theorem 8.4): the parametrized random distributions corresponding to random element-free distributions are in bijection with certain natural transformations between functors of distributions. This appears to be a new perspective on random distributions.

1.4 Multisets and exchangeability

This paper is heavily based on multisets. The first reason is that the coefficients in an element-free distribution are unordered, just like the elements to which they were originally assigned.

The second reason is that distributions on multisets correspond to distributions on lists that are invariant under list permutation. This kind of permutation-invariance, known as exchangeability, is a fundamental notion in Bayesian modelling, because the way we have indexed the data points $y_1, \ldots, y_K$ is usually arbitrary ([13, §1.2]).

For partitions, exchangeability means that, for instance, partitions $\{\{1, 2\}, \{3\}\}$ and $\{\{1, 3\}, \{2\}\}$ have the same probability of occurring. The model of Figure 2 is exchangeable because the colours are sampled independently and from the same urn.

By working with multisets, we implicitly enforce exchangeability of all distributions. So, in this paper:

- We model (ordinary) iid sampling as a function from distributions on X to distributions over multisets over X. (See §6.1.)
- We use *integer partitions* rather than the usual set partitions. An integer partition of K is a multiset of positive integers which sum to K, representing the "block sizes" in a partition of the set $\{1, \ldots, K\}$. (See §5.)
- We model element-free iid sampling as a function from element-free distributions to distributions over integer partitions. (See §6.2.)

We emphasize that integer partitions are the element-free counterpart of multisets of elements: for example the multiset $[a, a, a, b, c]$ gives rise to the partition $[3, 1, 1]$ of the integer 5. (There is a beautiful mathematical theory relating lists, multisets, set partitions and integer partitions [22].) In the rest of this paper, by *partition* we always mean integer partition.

Kingman's theorem. A central contribution of this paper is a new categorical proof of Kingman's representation theorem for random partitions. This theorem states that, in the limit $K = \infty$, every exchangeable random partition is induced by a unique random element-free distribution via iid sampling (as in Figure 2).

Kingman's theorem is the element-free counterpart of de Finetti's theorem [10], a foundational result in probability theory, that has recently been given various categorical presentations [12, 20, 21]. This paper makes clear the connection between the two theorems. We can derive the theorem using a purely categorical argument (Theorem 7.3), exploiting the results of §4 and §6 connecting element-free and ordinary probability theory. We emphasize that continuous distributions, and thus measure theory, are essential for this theorem to hold: discrete element-free distributions *do not* represent all exchangeable random partitions [24].

1.5 Summary of contributions and outline

We summarize the key contributions of this paper.

In §4, we construct a measurable space of element-free distributions defined as infinite multisets. We define two fundamental operations: *multiplicity count* for forgetting elements (following [16]), and drawing elements from a base measure.

In §5 we consider spaces of finite multisets and partitions, and also define multiplicity count [16] and base measures. We then make formal, in §6, the informal diagrams of §1.1 and §1.3: firstly by defining an element-free version of iid sampling for partitions, and secondly by showing that iid sampling commutes with the two fundamental operations (multiplicity count, and drawing from a base measure).

We can then prove our two main theorems. In §7 we give a new statement and proof of Kingman's theorem in terms of a categorical limit, using the categorical version of de Finetti's theorem for standard Borel spaces [20]. In §8 we prove a new representation theorem for random distributions parametrized by a base measure, in terms of a correspondence between natural transformations and random element-free distributions.

At a high level our motivation is to establish the following dictiorary as a tool for reasoning about models for clustering.

Elements in space X	Element-free
Lists $\langle x_1, \ldots, x_K \rangle$	Partitions of the set $\{1, \ldots, K\}$
Multisets $[x_1, \ldots, x_K]$	Partitions of the number K
Distributions on X	Element-free distributions
de Finetti's theorem	Kingman's theorem

We begin in §2 with a recap section on measure-theoretic probability theory. Then in §3 we introduce element-free distributions via the simple special case of distributions on $\mathbb{N}$.

2 PROBABILITY THEORY AND ATOMS

This section contains preliminary background on measure theory, the notion of probability kernel, and atoms.

A probability distribution on a set X is a function $p : X \to [0,1]$ such that $\sum_{x \in X} p(x) = 1$. This definition works well for discrete probability, but for continuous probability we need measure theory.

Definition 2.1. A measurable space is a set X equipped with a σ-algebra Σ_X: this is a set of subsets of X, containing $\emptyset$ and closed under complements, and countable unions and intersections.

Elements of Σ_X are called measurable subsets of X. We often refer to the space as X, omitting the Σ_X, when there is no ambiguity. Key examples of measurable spaces include:

- The set $\mathbb{R}$ of real numbers equipped with the Borel sets $\Sigma_{\mathbb{R}}$. This is defined as the smallest σ-algebra containing all intervals $(a, b) \subseteq \mathbb{R}$.
- Any subset of $S \subseteq \mathbb{R}$, with the σ-algebra $\{S \cap U \mid U \in \Sigma_{\mathbb{R}}\}$.
- Any set X with the discrete σ-algebra containing all subsets of X. With discrete spaces we can view distributions on sets as instances of a general notion of probability measure.

Definition 2.2. A probability measure on a measurable space (X, Σ_X) is a function $\mu : \Sigma_X \to [0,1]$ such that $\mu(X) = 1$, and for any countable family of disjoint subsets $U_i \in \Sigma_X$, we have $\mu(\bigcup_i U_i) = \sum_i \mu(U_i)$.

There is a category **Meas** of measurable spaces, whose morphisms $(X, \Sigma_X) \to (Y, \Sigma_Y)$ are measurable functions, i.e. maps $f : X \to Y$ such that for every $U \in \Sigma_Y$, $f^{-1}U \in \Sigma_X$.

We can pushforward a probability measure μ on X along a measurable function, and get a probability measure $f_*\mu$ on Y given by $(f_*\mu)(U) = \mu(f^{-1}U)$.

For any measurable function $f : X \to \mathbb{R}_+$ and probability measure μ on X, we can define its Lebesgue integral $\int_{x \in X} f(x)\mu(\mathrm{d}x) \in \mathbb{R}_+ \cup \{\infty\}$.

2.1 Kernels and the Giry monad

There is a monad of probability distributions on **Meas**, traditionally called the Giry monad [14, 26]. For a measurable space X, we write GX for the set of probability measures on X, equipped with the smallest σ-algebra which makes the maps $\mathrm{ev}_U : GX \to [0,1] : \mu \mapsto \mu(U)$ measurable for every $U \in \Sigma_X$. For a measurable function $f : X \to Y$, there is a measurable function $Gf : GX \to GY$ given by pushforward along f. This defines a functor $G : \mathbf{Meas} \to \mathbf{Meas}$.

Definition 2.3. A *(probability) kernel* from X to Y is a measurable function $X \to GY$. We use the notation $X \rightsquigarrow Y$.

A kernel $k : X \rightsquigarrow Y$ lifts to a function $k^\dagger : GX \to GY$ given by $\mu \mapsto \int_{x \in X} k(x, -)\mu(\mathrm{d}x)$, and we can use this to compose kernels: if $h : Y \rightsquigarrow Z$, then $h^\dagger \circ k : X \to GZ$ is a kernel $X \rightsquigarrow Z$.

Finally, for every space X there is a map $\eta_X : X \to GX$ which maps $x \in X$ to the Dirac distribution $\delta_x : U \mapsto [x \in U]$, which returns x with probability 1.

Lemma 2.4. *The triple $(G, (-)^\dagger, \eta)$ determines a monad on* **Meas**.

Subprobability distributions. A subprobability measure on (X, Σ_X) is defined in the same way as a probability measure, only with the weaker requirement that $\mu(X) \leq 1$. There is a space $G_\leq X$ of subprobability measures, that contains GX as a subspace; the definitions are very similar and we omit the details. ($G_\leq$ is also a monad, but we do not use this.)

2.2 Atoms and standard Borel spaces

For this paper we restrict ourselves to a class of well-behaved spaces in which it is safe to reason about elements and atoms.

Definition 2.5. A measurable space is a *standard Borel space* if it is either discrete and countable, or measurably isomorphic to $(\mathbb{R}, \Sigma_{\mathbb{R}})$.

Standard Borel spaces include any measurable subset of $(\mathbb{R}, \Sigma_{\mathbb{R}})$, and they are closed under taking subspaces and countable products. The Giry monad restricts to the subcategory **Sbs** of standard Borel spaces, i.e. if X is standard Borel then so is GX. In this paper when we write $\mathcal{K}\ell(G)$ we mean the Kleisli category for G over **Sbs**. This is the category whose objects are standard Borel spaces and whose morphisms are kernels.

In a standard Borel space, all singletons are measurable. If $X \in$ **Sbs**, and $\mu \in GX$ then we say μ has an *atom* at $x \in X$ if $\mu\{x\} > 0$.

The set of atoms of μ, written $\mathcal{A}_\mu$, must be countable and thus a measurable subset of X. Say μ is discrete or *purely atomic* if $\mu(X) = \mu(\mathcal{A}_\mu)$, and *non-atomic* if $\mathcal{A}_\mu = \emptyset$.

Product spaces and iid measures. The product of measurable spaces (X, Σ_X) and (Y, Σ_Y) is the set $X \times Y$ equipped with the σ-algebra generated by the rectangles $U \times V$ ($U \in \Sigma_X$, $V \in \Sigma_Y$).

Probability measures $\mu \in GX$ and $\nu \in GY$ determine a unique product measure $\mu \times \nu$ which satisfies $(\mu \times \nu)(U \times V) = \mu(U) \times \nu(V)$. In particular, for any $\kappa \in \mathbb{N} \cup \{\infty\}$, there is a measurable function $iid_\kappa : GX \to G(X^\kappa)$ which sends μ to the κ-fold product measure μ^κ. (We will also use *iid* to denote the sampling of multisets and partitions, to emphasize the correspondence. This should cause no confusion.)

3 DISCRETE ELEMENT-FREE DISTRIBUTIONS

This section is about a simple special case: probability distributions over the natural numbers $\mathbb{N}$, which are always discrete. For every distribution on $\mathbb{N}$, there is an induced element-free distribution: this is the family of probability coefficients, without the elements to which they were assigned.

We define discrete element-free distributions as multisets of values in $(0, 1]$ which sum to 1. We formalize multisets as functions $(0, 1] \to \mathbb{N}$, indicating the multiplicity of each value. These multisets may be (countably) infinite, because distributions on $\mathbb{N}$ can have infinite support.

Definition 3.1. A *discrete element-free distribution* is a function $\varphi : (0, 1] \to \mathbb{N}$ whose support $\mathrm{supp}(\varphi) = \{r \in (0, 1] \mid \varphi(r) > 0\}$ is (finite or) countable, and such that $\sum_{r \in \mathrm{supp}(\varphi)} \varphi(r) \cdot r = 1$.

(Note that, although the multisets may be infinite, each $r \in (0, 1]$ must have a finite multiplicity.)

A space of element-free distributions. The set of all discrete element-free distributions is denoted ∇_d. It is equipped with the σ-algebra generated by the sets $E^U_k = \{\varphi \in \nabla_d \mid \sum_{r \in U \cap \mathrm{supp}(\varphi)} \varphi(r) = k\}$ for $k \in \mathbb{N}$ and $U \in \Sigma_{(0,1]}$. Informally, E^U_k is the set of multisets having exactly k elements in U. (There is also a measurable set $E^U_\infty := \nabla_d \setminus \bigcup_{k \in \mathbb{N}} E^U_K$.)

We note that this is the smallest σ-algebra with respect to which the maps $\mathrm{ev}_U : \nabla_d \to \mathbb{N} \cup \{\infty\}$, where $\mathrm{ev}_U(\varphi) = \sum_{r \in U \cap \mathrm{supp}(\varphi)} \varphi(r)$, are measurable. (The similarity with the σ-algebra on $G(0, 1]$ is not an accident: multisets are often formalized as integer-valued measures in the probability literature (*e.g.* [25]).)

Multiplicity count. Every distribution on $\mathbb{N}$ has an underlying discrete element-free distribution. This is computed by an operation called *multiplicity count* (because it counts the multiplicities of coefficients *e.g.* [16]). We show that this is a measurable function.

Lemma 3.2. *Every distribution $\mu \in G\mathbb{N}$ induces a* multiplicity count $mc(\mu) \in \nabla_d$ *given by*

$$mc(\mu)(r) = \#\{n \in \mathbb{N} \mid \mu\{n\} = r\}$$

for $r \in (0, 1]$. The function $mc : G\mathbb{N} \to \nabla_d$ is measurable.

Proof. The preimage of a generating set $E^U_k \in \Sigma_{\nabla_d}$, where $k \in \mathbb{N}$ and $U \in \Sigma_{(0,1]}$, under the function *mc*, is the set of distributions μ such that $\mu\{n_i\} \in U$ for precisely k values $n_1, \ldots, n_k \in \mathbb{N}$. Thus $mc^{-1}(E^U_k) =$

$$\bigcup_{\substack{\langle n_1,\ldots,n_k\rangle \in \mathbb{N}^k \\ n_i \neq n_j}} \left(\bigcap_{i=1}^{k} \mathrm{ev}^{-1}_{n_i} U \right) \cap \left(\bigcap_{n_i \neq n \in \mathbb{N}} \mathrm{ev}^{-1}_n ((0, 1] \setminus U) \right),$$

a measurable set since it consists of countable unions and intersections of generating elements. ◻

Given a discrete element-free distribution φ, we write $\|\varphi\|$ for its size $\sum_{p \in \mathrm{supp}(\varphi)} \varphi(p)$, which may be infinite. For example, if $\mu \in G\mathbb{N}$, then $\|mc(\mu)\|$ equals the cardinality of the support of μ.

Element-free distributions in ordered form. It is common in the probability literature to find element-free distributions formalized as decreasing sequences over $[0, 1]$, rather than multisets ([24, 27]). We show that there is a map $ord : \nabla_d \to G\mathbb{N}$ which turns an element-free distribution into a distribution over $\mathbb{N}$ by considering the coefficients in decreasing order. This is a section (right-inverse) of multiplicity count.

Lemma 3.3. *For $\varphi \in \nabla_d$, let $ord(\varphi)$ be the distribution on $\mathbb{N}$ given by $ord(\varphi)(n) =$*

$$\begin{cases} \max \left\{ p \in \mathrm{supp}(\varphi) \mid \sum_{q \in \mathrm{supp}(\varphi), p \leq q} \varphi(q) > n \right\} & \text{if } \|\varphi\| > n \\ 0 & \text{otherwise.} \end{cases}$$

Then ord is a measurable function $\nabla_d \to G\mathbb{N}$ and a section of mc.

Example 3.4. Let φ be the multiset $[\frac{1}{5}, \frac{1}{5}, \frac{3}{5}]$ regarded as an element of ∇_d. The distribution $ord(\varphi) \in G\mathbb{N}$ is given by $0 \mapsto \frac{3}{5}$; $1, 2 \mapsto \frac{1}{5}$; and $n \mapsto 0$ for all $n \geq 3$.

Proof of Lemma 3.3. It suffices to show that $ord^{-1} V \in \Sigma_{\nabla_d}$ for every V in any basis for the σ-algebra $\Sigma_{G\mathbb{N}}$. We consider the basis consisting of the subsets $V^q_n = \{\mu \mid \mu\{n\} > q\}$, for $n \in \mathbb{N}$ and $q \in [0, 1]$. (To see why this is a basis, note that if Σ' is the σ-algebra on $G\mathbb{N}$ generated by the V^q_n, then each function $\mathrm{ev}_n : G\mathbb{N} \to [0, 1]$ is measurable, because the $(q, 1]$ generate $\Sigma_{[0,1]}$. Since $\Sigma_{G\mathbb{N}}$ is the smallest σ-algebra making the ev_n measurable, and clearly $\Sigma' \subseteq \Sigma_{G\mathbb{N}}$, they are equal.) The set $ord^{-1} V^q_n$ consists of the $\varphi \in \nabla_d$ whose n^{th} biggest element is above q, i.e.

$$ord^{-1} V^q_n = \bigcup_{k \geq n} E^{(q,1]}_k$$

which is measurable. So *mc* is measurable.

The composite $mc \circ ord$ is the identity: for $\varphi \in \nabla_d$ and $p \in (0, 1]$, we have that $\varphi(p)$ is equal to the number of $n \in \mathbb{N}$ such that

$$p = \max \left\{ p' \in \mathrm{supp}(\varphi) \mid \sum_{\substack{q \in \mathrm{supp}(\varphi) \\ q \geq p'}} \varphi(q) > n \right\}.$$

This number is precisely $mc(ord(\varphi))(p)$. ◻

As a section, the map *ord* must be injective. It has measurable image:

Lemma 3.5. *The image of $ord : \nabla_d \to G\mathbb{N}$ is the subset*

$$\{\mu \in G\mathbb{N} \mid \forall n \in \mathbb{N}.\ \mu\{n + 1\} \leq \mu\{n\}\}$$

of $G\mathbb{N}$, which is in the σ-algebra $\Sigma_{G\mathbb{N}}$.

Proof. The subset above is the countable intersection $\bigcap_{i\in\mathbb{N}}\{\mu \in G\mathbb{N} \mid \mu(i+1) \le \mu(i)\}$, each component of which has measurable characteristic function $G\mathbb{N} \xrightarrow{\langle \mathrm{ev}_i, \mathrm{ev}_{i+1}\rangle} \mathbb{R}\times\mathbb{R} \xrightarrow{\geq} \{0,1\}$. □

Measurable subsets of standard Borel spaces are standard Borel, and so ∇_d is standard Borel. As an aside, we note the following universal characterization of ∇_d and *mc*:

Proposition 3.6. *The function mc is a coequalizer for the diagram consisting of all maps* $G(\alpha)$ *for* $\alpha \in \mathrm{Aut}(\mathbb{N})$, *the symmetry group of* $\mathbb{N}$.

$$\begin{tikzcd} G\mathbb{N} \arrow[r, shift left=2, "G(\alpha)"] \arrow[r, phantom, "\cdots"] \arrow[r, shift right=2] & G\mathbb{N} \arrow[r, "mc"] & \nabla_d \end{tikzcd}$$

Proof. The map *mc* forms a cone over the diagram, because counting multiplicity is invariant under any reindexing of $\mathbb{N}$. For the universal property, let $f : G\mathbb{N} \to X$ be another invariant map into an arbitrary measurable space X. We must show that there exists a unique map $h : \nabla_d \to X$ such that $h \circ mc = f$. Uniqueness follows from the fact that *mc*, as a retraction, is epi. We set $h = f \circ ord$, and it remains to show that

$$\begin{tikzcd} G\mathbb{N} \arrow[r, "mc"] \arrow[dr, "f"'] & \nabla_d \arrow[r, "ord"] & G\mathbb{N} \arrow[dl, "f"] \\ & X & \end{tikzcd}$$

commutes. For every $\mu \in G\mathbb{N}$, the distribution $ord(mc(\mu))$ is a reindexing of μ along some permutation $\alpha \in \mathrm{Aut}(\mathbb{N})$, i.e. $mc(ord(\mu)) = G(\alpha)(\mu)$. As $f = f \circ G(\alpha)$ by the assumption that f is invariant, we are done. □

So far we have considered discrete element-free distributions: infinite multisets whose elements sum to 1. These are a natural way to represent the element-free part of a probability distribution on $\mathbb{N}$, or any other discrete space.

4 GENERAL ELEMENT-FREE DISTRIBUTIONS

In this section we introduce the general notion of element-free distribution, which also supports continuous distributions. We also develop two fundamental operations for a standard Borel space X: multiplicity count for probability measures on X (§4.1), and the sampling of new elements from a base measure (§4.2). This gives a retract situation (§4.3).

Recall that an element-free distribution is a multiset representing the weights assigned to atoms of a probability distribution. Distributions on continuous spaces can have a nonatomic part, and so those weights may not sum to 1. This is the only difference between discrete and non-discrete element-free distributions.

Definition 4.1. An *element-free probability distribution* is a function $\varphi : (0,1] \to \mathbb{N}$ whose support is countable, and such that $\sum_{p\in\mathrm{supp}(\varphi)} \varphi(p)\cdot p \le 1$.

We write ∇ for the space of element-free distributions, with a basis of measurable subsets given by

$$D_k^U = \{\varphi \in \nabla \mid \sum_{p\in U\cap\mathrm{supp}(\varphi)} \varphi(p) = k\}$$

for $U \in \Sigma_{(0,1]}$ and $k \in \mathbb{N}$. (This extends the σ-algebra of ∇_d, in particular $E_k^U = D_k^U \cap \nabla_d$.)

Keeping the nonatomic part implicit, general element-free distributions can be viewed as subprobability distributions on the natural numbers. It will sometimes be convenient to use generalized versions of the functions *mc* and *ord* from Section 3. Recall that the space of sub-probability distributions on $\mathbb{N}$ is written $G_{\leq}\mathbb{N}$, and contains $G\mathbb{N}$ as a subspace.

Lemma 4.2. *There are measurable functions* $mc : G_{\leq}\mathbb{N} \to \nabla$ *and* $ord : \nabla \to G_{\leq}\mathbb{N}$ *extending those given in Lemma 3.2 and Lemma 3.3, and defined in the same way. Thus in particular* ∇ *is standard Borel.*

We also note that ∇_d is a measurable subset of ∇ : it is the preimage of $\{1\}$ under the measurable function $\nabla \xrightarrow{ord} G_{\leq}\mathbb{N} \xrightarrow{\mathrm{ev}_{\mathbb{N}}} [0,1]$.

4.1 Multiplicity count for general spaces

By moving to general element-free distributions, we can consider arbitrary probability measures on standard Borel spaces. We thus have a new kind of multiplicity count, that forgets everything but the atom weights (and their multiplicity).

For every standard Borel space X, this is given by

$$\begin{aligned} mc : GX &\longrightarrow \nabla \\ \mu &\longmapsto (r \mapsto \#\{x \in X \mid \mu\{x\} = r\}). \end{aligned}$$

Notational remark. Here and throughout the paper we overload the notation for our two fundamental operations *mc* and $base_\mu$ (next section), to emphasize that they correspond to the same passage between element-free and ordinary probability. The types should be clear from context, and so this should pose no confusion.

Example 4.3.
- For any Dirac $\delta_x \in GX$, $mc(\delta_x) = [1]$.
- If $\mathscr{U}_{[0,1]}$ is the uniform distribution on $[0,1]$, then $mc(\mathscr{U}_{[0,1]})$ is the empty element-free distribution.
- We can have discrete-continuous mixtures. If $x \in [0,1]$, then $mc(\frac{1}{3}\delta_x + \frac{2}{3}\mathscr{U}_{[0,1]}) = [\frac{1}{3}]$, since there is a single atom with weight $\frac{1}{3}$.

It is not immediate that *mc* is measurable. We first show a preliminary lemma:

Lemma 4.4. *For* $\varepsilon > 0$, *define* $mc_\varepsilon : GX \to \nabla$ *by* $mc_\varepsilon(\mu)(r) = mc(\mu)(r)$ *if* $r \ge \varepsilon$, *and* 0 *otherwise. Then* mc_ε *is measurable.*

The measurability of *mc* ought to follow because limits of measurable functions are measurable, but to make this limit argument precise one needs the topology of ∇. To avoid this, a direct argument is as follows: let $D_{\le k}^U = \bigcup_{\ell\le k} D_\ell^U$ and note that the $D_{\le k}^U$ still generate Σ_∇. We then have that

$$\begin{aligned} &mc^{-1}D_{\le k}^U \\ &= \{\mu \in GX \mid \mu \text{ has} \le k \text{ atoms with weight in } U\} \\ &= \bigcap_{m\ge 1} \left\{\mu \in GX \mid \mu \text{ has} \le k \text{ atoms with weight in } U \cap (\tfrac{1}{m}, 1]\right\} \\ &= \bigcap_{m\ge 1} mc_{\frac{1}{m}}^{-1} D_{\le k}^U, \end{aligned}$$

a countable union of measurable sets, and so *mc* is measurable.

Proof (of Lemma 4.4). We only give details for a non-discrete X. Suppose w.l.o.g. that $X = [0,1]$. There is an increasingly-refined sequence of partitions of X: for $m \in \mathbb{N}$ and $0 \leq n < 2^m$, let $B_m^n = [\frac{n}{2^m}, \frac{n+1}{2^m})$.

We now show that, for every $\alpha \geq 0$ and $k \in \mathbb{N}$, the set $mc_\varepsilon^{-1} D_k^{[\alpha,1)}$ is measurable. We claim that this set is expressible using countable unions and intersections of basis elements, as follows:

$$\bigcup_{\ell \in \mathbb{N}} \bigcap_{m \geq \ell} \bigcup_{\substack{\{n_1,\dots,n_k\} \\ \subseteq \{0,\dots,2m-1\}}} \left(\bigcap_{i=1}^{k} \mathrm{ev}_{B_m^{n_i}}^{-1} [\alpha, 1] \right) \cap \left(\bigcap_{\substack{n \leq 2m-1 \\ \forall i.\ n \neq n_i}} \mathrm{ev}_{B_m^n}^{-1} [0, \alpha) \right).$$

To prove this claim we use standard but verbose analytic arguments; the full details are in Appendix A.1.

It remains to show that for every $U \in \Sigma_{(0,1]}$ (and not just for the $[\alpha, 1]$), the sets $mc_\varepsilon^{-1} D_k^U$ are measurable for $k \in \mathbb{N}$. Let $\mathcal{S}$ be the set of those $U \subset [0, 1)$ satisfying this property. It suffices to show that $\mathcal{S}$ is closed under relative complements and countable increasing unions: by the π-λ theorem (*e.g.* [2, Thm. 1.8]) this implies $\Sigma_{(0,1]} \subseteq \mathcal{S}$, from which we can conclude.

For relative complements, observe that, whenever $U \subseteq V \in \mathcal{S}$, $mc_\varepsilon^{-1} D_k^{V \setminus U} = \bigcup_{m \in \mathbb{N}} mc_\varepsilon^{-1} D_{k+m}^V \cap mc_\varepsilon^{-1} D_m^U$. This is a measurable set and so $V \setminus U \in \mathcal{S}$. (Here we use that $\varepsilon > 0$, to avoid infinite numbers of atoms.) For countable increasing unions, note that $mc_\varepsilon^{-1} D_k^{\bigcup_{n \in \mathbb{N}} U_n} = \bigcup_{m \in \mathbb{N}} \bigcap_{n \geq m} mc_\varepsilon^{-1} D_k^{U_n}$, and so the proof is complete. □

In summary, the map $mc : GX \to \nabla$ is measurable for every standard Borel space X. This is a deterministic way to forget elements, and we now give a probabilistic way to recover them.

4.2 Drawing elements from a base measure

We can turn an element-free distribution $\varphi \in \nabla$ into an ordinary distribution on a space X by sampling elements from a base measure $\mu \in GX$. More precisely, the weights of φ are assigned to atoms sampled independently from μ, and the nonatomic part in φ is completed using μ itself. Formally, we will define a kernel *base* : $GX \times \nabla \rightsquigarrow GX$.

For each $\varphi \in \nabla$, the set $\mathrm{supp}(\varphi) \subseteq [0, 1]$ is countable and well-ordered under $\geq$. Thus we can enumerate its elements in decreasing order: $\mathrm{supp}(\varphi) = \{r_1 \geq r_2 \geq \dots\}$. Then, given any sequence $\boldsymbol{x} = (x_i)_{i=1}^{\|\varphi\|}$ of elements in a set X, we can decompose the sequence $\boldsymbol{x}$ as

$$\boldsymbol{x} = \boldsymbol{x}^{r_1} \oplus \boldsymbol{x}^{r_2} \oplus \dots$$

where $\oplus$ denotes concatenation, and each $\boldsymbol{x}^{r_i}$ is a subsequence of size $\varphi(r_i)$. For each $r \in \mathrm{supp}(\varphi)$ we write the elements of $\boldsymbol{x}^r$ as $(x_1^r, \dots, x_{\varphi(r)}^r)$. (Note that each operation $\boldsymbol{x} \mapsto \boldsymbol{x}^{r_i}$ is simply a particular product projection, and thus a measurable function.)

Definition 4.5 (Allocation). For $\varphi \in \nabla$ and $\boldsymbol{x} \in X^{\|\varphi\|}$, let

$$alloc(\varphi, \boldsymbol{x}) = \sum_{r \in \mathrm{supp}(\varphi)} r \cdot \sum_{i=1}^{\varphi(r)} \delta_{x_i^r}$$

be the subprobability measure that results from allocating elements in $\boldsymbol{x}$ to weights in φ.

Definition 4.6 (Drawing elements from a base measure). The kernel $base_\mu : \nabla \rightsquigarrow GX$ is defined for any $\varphi \in \nabla$ and $V \in \Sigma_{GX}$ as follows:

$$base_\mu(\varphi)(V) = \int_{\boldsymbol{x} \in X^{\|\varphi\|}} [(1 - \mathrm{tt}(\varphi))\mu + alloc(\varphi, \boldsymbol{x}) \in V]\ iid_{\|\varphi\|}(\mu)(\mathrm{d}\boldsymbol{x}).$$

(We denote by $[P]$ the indicator function of a property P.)

Example 4.7. Let φ be the multiset $[\frac{1}{3}, \frac{2}{3}]$ viewed as an element of ∇. We give examples of sampling elements from discrete and continuous base measures.

(1) If $X = \{0, 1\}$ and $\mu = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1$, then the distribution $base_\mu(\varphi)$ on GX is purely atomic: it has four atoms with weights given below.

$$\left(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_1\right) \longmapsto \frac{3}{16} \qquad \delta_0 \longmapsto \frac{1}{16}$$
$$\left(\frac{2}{3}\delta_0 + \frac{1}{3}\delta_1\right) \longmapsto \frac{3}{16} \qquad \delta_1 \longmapsto \frac{9}{16}.$$

(2) If $X = [0, 1]$ and $\mu = \mathscr{U}_{[0,1]}$ is the uniform distribution, then $base_\mu(\varphi)$ is given by

$$base_\mu(\varphi)(M) = \int_{(x,y) \in [0,1]^2} \left[\left(\frac{1}{3}\delta_x + \frac{2}{3}\delta_y\right) \in M\right] \mathrm{d}x\, \mathrm{d}y$$

for $M \in \Sigma_{GX}$.

Example 4.8. We now look at a different element-free distribution: $\psi = [\frac{3}{5}] \in \nabla$. Note that ψ has a continuous part of weight $\frac{2}{5}$.

(1) If $X = \{0, 1\}$ and $\mu = \frac{1}{4}\delta_0 + \frac{3}{4}\delta_1$, then the distribution $base_\mu(\psi)$ on GX is purely atomic with atoms:

$$(\frac{2}{5}\mu + \frac{3}{5}\delta_0) \mapsto \frac{1}{4} \qquad (\frac{2}{5}\mu + \frac{3}{5}\delta_1) \mapsto \frac{3}{4}$$

(These atoms reduce to $\frac{7}{10}\delta_0 + \frac{3}{10}\delta_1$ and $\frac{1}{10}\delta_0 + \frac{9}{10}\delta_1$.)

(2) If $X = [0, 1]$ and $\mu = \mathscr{U}_{[0,1]}$ is the uniform distribution, then

$$base_\mu(\psi)(M) = \int_{x \in [0,1]} \left[\left(\frac{2}{5}\mathscr{U}_{[0,1]} + \frac{3}{5}\delta_x\right) \in M\right] \mathrm{d}x$$

for $M \in \Sigma_{GX}$.

4.3 The space ∇ as a retract

We can now move between the space of element-free distributions and spaces of distributions on X using the two constructions:

$$GX \xrightarrow{mc} \nabla \qquad \nabla \xrightarrow{base_\mu} GX \quad (\mu \in GX)$$

An important observation is that the composite

$$\nabla \xrightarrow{base_\mu} GX \xrightarrow{mc} \nabla$$

is equal to the identity on ∇ whenever the measure μ is nonatomic. The intuition is that the elements produced by iid_μ are all distinct with probability 1, and thus weights are never combined. (For a counterexample when μ is atomic: let $\mu = \delta_a$ on the set $\{a, b\}$. Then for any φ, $base_\mu(\varphi)$ assigns the element a to every weight of φ, *i.e.* $base_\mu(\varphi) = \delta_{\delta_a}$. Applying mc to δ_a returns $[1] \in \nabla$. We see that the weights of φ have all been combined.)

Proposition 4.9. *For a standard Borel space X and a nonatomic measure $\mu \in GX$, the kernel $base_\mu : \nabla \rightsquigarrow GX$ is a section of $mc : GX \to \nabla$ in the category of kernels; that is, $mc \circ base_\mu = \mathrm{id}_\nabla$.*

Proof sketch. The key idea is that, because μ is nonatomic, the measure $iid(\mu) \in G(X^{\mathbb{N}})$ is concentrated on the subspace of sequences whose elements are pairwise distinct: let $S = \{(x_n)_{n\in\mathbb{N}} \mid x_i \neq x_j \text{ if } i \neq j\}$. For any $\boldsymbol{x} \in S$ (and, say, if φ has infinite support) the distribution $alloc(\boldsymbol{x}, \sigma)$ has the exact same atom weights as found in φ. □

5 FINITE MULTISETS AND PARTITIONS

In this section we consider spaces of finite multisets and their element-free counterpart: partitions of positive integers. Just like for distributions and element-free distributions, we have two fundamental operations: multiplicity count, a deterministic operation which turns a multiset of size K into a partition of the number K by forgetting elements; and a probabilistic operation, which turns a partition into a multiset by sampling elements from a base measure. These operations will be defined in §5.2. In §5.1 we recall the basics of finite multisets and partitions.

5.1 Finite multisets and partitions

Multisets. We have previously used infinite multisets of positive reals to define element-free distributions. Here we consider finite multisets over an arbitrary space X.

Definition 5.1. For a measurable space X, let $\mathcal{M}(X)$ be the set of finite multisets of elements of X. These are formally defined as functions $X \to \mathbb{N}$ with finite support. The σ-algebra $\Sigma_{\mathcal{M}(X)}$ is generated by the sets $E^U_k = \{\varphi \in \mathcal{M}(X) \mid \sum_{U\cap\mathrm{supp}(\varphi)} \varphi(x) = k\}$ for $k \in \mathbb{N}$ and $U \in \Sigma_X$.

Notation 1. For a multiset $\varphi \in \mathcal{M}(X)$, write $\|\varphi\|$ for its size, *i.e.* $\|\varphi\| = \sum_{x\in\mathrm{supp}(\varphi)} \varphi(x)$. For $K \in \mathbb{N}$, the subspace of multisets of size K is denoted $\mathcal{M}[K](X)$. (Note that $\mathcal{M}[K](X)$ is equal to the generating set E^X_K, and thus a measurable subset of $\mathcal{M}(X)$.)

It it well-known (e.g. [9]) that if the space X is standard Borel, then so is $\mathcal{M}(X)$, and therefore so are all $\mathcal{M}[K](X)$.

Partitions. Partitions are defined as multisets of integers. Each integer represents the size of a block of (unnamed) elements. Partitions are a classical object of study in combinatorics [1]. The modern presentation we give now, in terms of multisets, is due to Jacobs [16].

Definition 5.2. For a multiset $\sigma \in \mathcal{M}(\mathbb{N}_{>0})$, the *total* of σ is an integer $\mathrm{tt}(\sigma)$ defined as $\mathrm{tt}(\sigma) = \sum_{n>0} \sigma(n) \cdot n$.

Definition 5.3. For $K \in \mathbb{N}$, the set of partitions $\mathcal{P}(K)$ is given by $\{\sigma \in \mathcal{M}(\mathbb{N}_{>0}) \mid \mathrm{tt}(\sigma) = K\}$.

Example 5.4. The set of partitions $\mathcal{P}(3)$ contains the three multisets $[1, 1, 1]$, $[1, 2]$, and $[3]$, representing the possible combinations of block sizes in partitions of a three-element set $\{a, b, c\}$: $\{\{a\}, \{b\}, \{c\}\}$, $\{\{a, b\}, \{c\}\}$, $\{\{a\}, \{b, c\}\}$, $\{\{a, c\}, \{b\}\}$, $\{\{a, b, c\}\}$.

The sets $\mathcal{P}(K)$ are all finite, and we regard them as discrete measurable spaces.

Before proceeding, we recall some combinatorial coefficients for multisets and partitions, taken from [16].

Definition 5.5. For a multiset $\varphi \in \mathcal{M}(X)$ over any set X,

- The multinomial coefficient $(\!\varphi\!)$ is defined as $\frac{\|\varphi\|!}{\prod_{x\in\mathrm{supp}(\varphi)} \varphi(x)!}$.
- For $n \geq \|\varphi\|$, the coefficient $\binom{n}{\varphi}$ is defined as

$$\binom{n}{\varphi} = (\!\varphi\!) \cdot \binom{n}{\|\varphi\|} = \frac{n!}{(n - \|\varphi\|)! \cdot \prod_x \varphi(x)!}$$

Finally, for a partition $\sigma \in \mathcal{P}(K)$, the coefficient $(\!\sigma\!)_{\mathrm{p}}$ is defined as $\frac{K!}{\prod_{n\leq K} (n!)^{\sigma(n)}}$.

5.2 Multiplicity count for multisets

Every finite multiset over a space X induces a partition, obtained by forgetting the elements. For example, for $X = \{a, b, c\}$, the multiset $[a, a, b]$ gives the partition $[2, 1]$.

In other words, partitions are element-free multisets. This operation is another form of multiplicity count [16]:

$$\begin{aligned} mc : \mathcal{M}[K](X) &\longrightarrow \mathcal{P}(K) \\ \varphi &\longmapsto (n \mapsto \#\{x \in \mathrm{supp}(\varphi) \mid \varphi(x) = n\}) \end{aligned}$$

For our applications we note that multiplicity count over standard Borel spaces is measurable.

Lemma 5.6. *For every standard Borel space X and integer $K \geq 0$, the function $mc : \mathcal{M}[K](X) \to \mathcal{P}(K)$ is measurable.*

Proof. We can leverage the fact that $mc : GX \to \nabla$ is measurable. The idea is to regard $\mathcal{M}[K](X)$ as a subspace of GX, via the injective map $f_k : \mathcal{M}[K](X) \to GX$ given by $f_K(\varphi) = \sum_{x\in\mathrm{supp}(\varphi)} \frac{\varphi(x)}{K} \delta_x$. For $U \in \Sigma_X$, the following diagram commutes

$$\begin{tikzcd} \mathcal{M}[K](X) \arrow[r, "f_K"] \arrow[d, "\mathrm{ev}_U"'] & GX \arrow[d, "\mathrm{ev}_U"] \\ \mathbb{N} \arrow[r, "\frac{\cdot}{K}"] & {[0,1]} \end{tikzcd}$$

and so each $\mathrm{ev}_U \circ f_K$ is measurable. Since the ev_U generate Σ_{GX}, f_K is measurable. Similarly, via the map $\mathcal{P}(K) \to \nabla$ which divides every element of a partition by K, we identify the set $\mathcal{P}(K)$ with the subset of ∇ consisting of multisets whose support is contained in $\{\frac{i}{K} \mid 1 \leq i \leq K\}$. This is a finite and thus measurable subset. The diagram below commutes

$$\begin{tikzcd} \mathcal{M}[K](X) \arrow[r, "mc"] \arrow[d, "f_K"'] & \mathcal{P}(K) \arrow[d, "f"] \\ GX \arrow[r, "mc"] & \nabla \end{tikzcd}$$

and so the lemma holds. □

5.3 Partitions and base measures

In the reverse direction, given a partition, we can randomly sample an element for each block to reconstruct a multiset. We give a formal definition, and then a more combinatorial characterization. For this section we fix a standard Borel space X, a probability measure $\mu \in GX$, and an integer $K \geq 0$.

We first make an observation similar to that in §4.2: for a partition $\sigma \in \mathcal{P}(K)$, any $\boldsymbol{x} \in X^{\|\sigma\|}$ is necessarily of the form

$$\boldsymbol{x} = \boldsymbol{x}^1 \oplus \cdots \oplus \boldsymbol{x}^K$$

for (possibly empty) subsequences $\boldsymbol{x}^n = (x_1^n, \ldots, x_{\sigma(n)}^n)$. (In other words, we consider the blocks of σ in increasing order of size.)

Definition 5.7 (Allocation). For $\sigma \in \mathcal{P}(K)$ and $\boldsymbol{x} \in X^{\|\sigma\|}$, define

$$alloc(\sigma, \boldsymbol{x}) = \sum_{n=1}^{K} \sum_{i=1}^{\sigma(n)} n[x_i^n] \quad \in \mathcal{M}[K](X)$$

where the sum operation and $\mathbb{N}$-action on multisets are defined pointwise on the corresponding functions $X \to \mathbb{N}$. (So, in particular, $n[x] = [x, \ldots, x]$ with x occurring n times.)

Definition 5.8 (Base measure draws for partitions). For $\sigma \in \mathcal{P}(K)$, the measure $base_\mu(\sigma)$ on $\mathcal{M}[K](X)$ is defined by

$$base_\mu(\sigma)(V) = \int_{\boldsymbol{x} \in X^{\|\sigma\|}} [alloc(\sigma, \boldsymbol{x}) \in V] \, \mathrm{d}iid_{\|\sigma\|}(\mu)(\boldsymbol{x}).$$

for $V \in \Sigma_{\mathcal{M}[K](X)}$. This defines a kernel $base_\mu : \mathcal{P}(K) \rightsquigarrow \mathcal{M}[K](X)$.

We have another retract situation between multisets over a continuous space and partitions.

Lemma 5.9. *For a non-atomic $\mu \in GX$, and $K \geq 0$, the kernels $base_\mu : \mathcal{P}(K) \rightsquigarrow \mathcal{M}[K](X)$ and $mc : \mathcal{M}[K](X) \rightsquigarrow \mathcal{P}(K)$ form a section-retraction pair,* i.e. $mc \circ base_\mu = \mathrm{id}_{\mathcal{P}(K)}$.

It will be helpful to have a more combinatorial expression of the kernel $base_\mu$. We recall the following result.

Lemma 5.10 (Dash and Staton [8]). *The family of subsets of $\mathcal{M}(X)$ of the form*

$$\bigcup_{i \in I} \bigcap_{j \in J} E_{k_{i,j}}^{U_j}$$

where I and J are countable, the U_j are all disjoint measurable subsets of X, and each $k_{i,j} \in \mathbb{N}$, forms a ring *which generates $\Sigma_{\mathcal{M}(X)}$.*

A measure on any measurable space is determined by its values on a generating ring [2], and so the following lemma suffices to characterize $base_\mu(\sigma)$ for any $\sigma \in \mathcal{P}(K)$.

Lemma 5.11. *For $\sigma \in \mathcal{P}(K)$, if $U \in \Sigma_X$ and $k \leq K$, then*

$$base_\mu(\sigma)\Big(\bigcup_{i \in I} \bigcap_{j \in J} E_{k_{i,j}}^{U_j}\Big) = \sum_{i \in I} \sum_{\substack{(\tau_j)_{j \in J}, \tau_j \leq \sigma \\ \mathrm{tt}(\tau_j) = k_{i,j}}} \mu(V)^{\|\sigma\| - \|\sum_j \tau_j\|} \left(\prod_{j \in J} \mu(U_j)^{\|\tau_j\|} \right) \prod_{n \leq K} \binom{\sigma(n)}{\tau_j(n), j \in J}$$

where V is the complement of the union $\bigcap_{j \in J} U_j$, and where the order $\leq$ on partitions is defined pointwise on the corresponding functions $\mathbb{N}_{>0} \to \mathbb{N}$.

In this section we have defined the two fundamental operations $mc : \mathcal{M}[K](X) \to \mathcal{P}(K)$ and $base_\mu : \mathcal{P}(K) \rightsquigarrow \mathcal{M}[K](X)$ for moving between the element-based and element-free settings.

6 SAMPLING MULTISETS AND PARTITIONS

The process of generating finite collections of independent and identically distributed samples is modelled by a family of kernels $iid_K : GX \rightsquigarrow \mathcal{M}[K](X)$, for $K \in \mathbb{N}$. This is the usual kind of sampling, with elements. In this section we contrast it with an element-free version, defined as a kernel $iid_K : \nabla \rightsquigarrow \mathcal{P}(K)$. As we will see, the diagrams

$$\begin{tikzcd} GX \arrow[r, "mc"] \arrow[d, "iid_K"'] & \nabla \arrow[d, "iid_K"] \\ \mathcal{M}[K](X) \arrow[r, "mc"'] & \mathcal{P}(K) \end{tikzcd} \qquad \begin{tikzcd} GX \arrow[d, "iid_K"'] & \nabla \arrow[l, "base_\mu"'] \arrow[d, "iid_K"] \\ \mathcal{M}[K](X) & \mathcal{P}(K) \arrow[l, "base_\mu"] \end{tikzcd} \tag{1}$$

both commute in $\mathcal{K}\ell(G)$.

6.1 Sampling with elements

We first discuss the generation of multisets of iid samples. The function $GX \to G(X^K)$ sending μ to the K-ary product measure μ^K is regarded as a kernel $GX \rightsquigarrow X^K$ that generates ordered lists of iid samples. We are interested in the composite

$$iid_K : GX \rightsquigarrow X^K \to M[K](X),$$

where the second (deterministic) map forgets the list order.

Lemma 6.1. *Let $\mu \in GX$, for a standard Borel space X. For the generating sets of Lemma 5.10, if $\sum_j k_{i,j} \leq K$ for all $i \in I$, then*

$$iid_K(\mu)\left(\bigcup_{i \in I} \bigcap_{j \in J} E_{k_{i,j}}^{U_j}\right) = \sum_{i \in I} \frac{K!}{(K - \sum_j k_{i,j})! \prod_j k_{i,j}!} \mu(V)^{K - \sum_j k_{i,j}} \prod_j \mu(U_j)^{k_{i,j}}.$$

6.2 Sampling without elements

We now turn to the generation of iid random partitions directly from an element-free distribution. This process was studied in depth by Jacobs [16] in the restricted setting of finitely-supported, discrete element-free distributions. Our results rely and expand on his sampling formula, which we first recall:

Definition 6.2 ([16, Def. 15]). Let φ be a discrete element-free distribution with finite support $\{r_1, \ldots, r_\ell\}$, and with $\varphi(r_i) = n_i$. The distribution $iid_K(\varphi)$ on $\mathcal{P}(K)$ is given by

$$\sigma \mapsto (\sigma)_{\mathrm{p}} \sum_{\substack{\sigma_1, \ldots, \sigma_\ell \\ \|\sigma_i\| \leq n_i, \sum_i \sigma_i = \sigma}} \prod_i \binom{n_i}{\sigma_i} \cdot r_i^{\mathrm{tt}(\sigma_i)}.$$

The intuitive justification for this formula is that we must consider all the possible ways to allocate blocks of σ to coefficients in φ. We first look at all possible ways of splitting σ into sub-partitions σ_i. The blocks of each σ_i are then allocated to distinct copies of the coefficient r_i in φ; there are $\binom{n_i}{\sigma_i}$ possible such allocations.

We now proceed to generalize this formula to the full space of element-free distributions. The first step is to account for infinitely supported (but still discrete) element-free distributions, which we can do by rewriting the above as follows:

$$\sigma \mapsto (\sigma)_{\mathrm{p}} \sum_{\substack{(\sigma_r)_{r\in\mathrm{supp}(\varphi)} \\ \|\sigma_r\|\leq\varphi(r), \sum_r \sigma_r=\sigma}} \prod_{r\in\mathrm{supp}(\varphi)} \binom{\varphi(r)}{\sigma_r} \cdot r^{\mathrm{tt}(\sigma_r)} \tag{2}$$

where the index of the sum is countable, even if $\mathrm{supp}(\varphi)$ is infinite, because all but finitely many σ_r must be empty by the condition that $\sum_r \sigma_r = \sigma$. (Similarly, in the infinite product all but finitely many factors are equal to 1.)

We now generalize this formula to general element-free distributions, whose coefficients may not sum to 1. Recall that the continuous part of φ represents an atomless distribution, and therefore only produces singleton blocks. In the definition below, the idea is to sum over the possible numbers k of singleton blocks in σ arising from this continuous part.

Definition 6.3. Let $\varphi \in \nabla$, and let $w \in [0,1]$ be the weight of its continuous part, i.e. $w = \sum_{r\in\mathrm{supp}(\varphi)} r \cdot \varphi(r)$.

For $K \geq 0$, the distribution $iid_K(\varphi)$ over $\mathcal{P}(K)$ is given by

$$\sigma \mapsto (\sigma)_{\mathrm{p}} \sum_{k=0}^{K} \frac{w^k}{k!} \sum_{\substack{(\sigma_r)_{r\in\mathrm{supp}(\varphi)} \\ \|\sigma_r\|\leq\varphi(r) \\ \sum_r \sigma_r + k[1]=\sigma}} \prod_{r\in\mathrm{supp}(\varphi)} \binom{\varphi(r)}{\sigma_r} \cdot r^{\mathrm{tt}(\sigma_r)}$$

for $\sigma \in \mathcal{P}(K)$, where $k[1] \in \mathcal{P}(k)$ is the all-singleton partition.

It is easy to see that if $w = 0$ the formula reduces to the expression in (2). We must now show that $iid_K(\varphi)$ is a well-defined probability distribution.

We apply a similar proof method as Jacobs, based on a 'partitions multinomial theorem' ([16, Theorem 16]). We first extend the theorem to an infinite setting using a straightforward limit argument.

Lemma 6.4 (Infinite partitions multinomial). *Let I be a countable set, $(r_i)_{i\in I}$ a collection of non-negative reals, and $(n_i)_{i\in I}$ a collection of non-negative integers. For $K \geq 0$,*

$$\left(\sum_{i\in I} r_i \cdot n_i\right)^K = \sum_{\sigma\in\mathcal{P}(K)} (\sigma)_{\mathrm{p}} \sum_{\substack{(\sigma_i)_{i\in I} \\ \|\sigma_i\|\leq n_i \\ \sum_{i\in I}\sigma_i=\sigma}} \prod_{i\in I} \binom{n_i}{\sigma_i} \cdot r_i^{\mathrm{tt}(\sigma_i)}.$$

Proof. Countable sums of positive reals are obtained as supremums of finite sums, and we can apply the finite partitions multinomial theorem from [16].

$$\begin{aligned}
\left(\textstyle\sum_{i\in I} r_i \cdot n_i\right)^K &\overset{(1)}{=} \sup_{J\subseteq_f I} \left(\sum_{j\in J} r_j \cdot n_j\right)^K \\
&\overset{(2)}{=} \sup_{J\subseteq_f I} \left(\sum_{\sigma\in\mathcal{P}(K)} (\sigma)_{\mathrm{p}} \sum_{\substack{(\sigma_j)_{j\in J} \\ \|\sigma_j\|\leq n_j, \sum_j \sigma_j=\sigma}} \prod_{j\in J} \binom{n_j}{\sigma_j} \cdot r_j^{\mathrm{tt}(\sigma_j)}\right) \\
&\overset{(3)}{=} \sum_{\sigma\in\mathcal{P}(K)} (\sigma)_{\mathrm{p}} \sup_{J\subseteq_f I} \left(\sum_{\substack{(\sigma_j)_{j\in J} \\ \|\sigma_j\|\leq n_j, \sum_j \sigma_j=\sigma}} \prod_{j\in J} \binom{n_j}{\sigma_j} \cdot r_j^{\mathrm{tt}(\sigma_j)}\right) \\
&\overset{(4)}{=} \sum_{\sigma\in\mathcal{P}(K)} (\sigma)_{\mathrm{p}} \sum_{\substack{(\sigma_i)_{i\in I} \\ \|\sigma_i\|\leq n_i, \sum_i \sigma_i=\sigma}} \prod_{i\in I} \binom{n_i}{\sigma_i} \cdot r_i^{\mathrm{tt}(\sigma_i)},
\end{aligned}$$

where (1) is because taking the Kth power is continuous and increasing, (2) is by the finite partitions multinomial theorem, (3) is because sums and sups can be interchanged over the positive reals, and (4) uses the fact that any $(\sigma_i)_{i\in I}$ as in the sum index is necessarily 0 outside of a finite $J \subseteq I$. □

Proposition 6.5. *For $\varphi \in \nabla$ and $K \geq 0$, $iid_K(\varphi)$ is a well-defined probability distribution.*

Proof. We first observe that for $k \leq K$ and $\tau \in \mathcal{P}(K-k)$,

$$\binom{K}{k}(\tau)_{\mathrm{p}} = \frac{1}{k!}(\tau + k[1]).$$

(This is easy to verify directly.) We then write w for the weight of the continuous part of φ. We apply the binomial theorem, Lemma 6.4, and the identity above:

$$\begin{aligned}
1 &= \left(w + \sum_{r\in\mathrm{supp}(\varphi)} r \cdot \varphi(r)\right)^K \\
&= \sum_{k=0}^{K} \binom{K}{k} w^k \left(\sum_{r\in\mathrm{supp}(\varphi)} r \cdot \varphi(r)\right)^{K-k} \\
&= \sum_{k=0}^{K} \binom{K}{k} w^k \sum_{\tau\in\mathcal{P}(K-k)} (\tau)_{\mathrm{p}} \sum_{(\tau_r)_{r\in\mathrm{supp}(\varphi)}} \prod_r \binom{\varphi(r)}{\tau_r} r^{\mathrm{tt}(\tau_r)} \\
&= \sum_{k=0}^{K} \frac{w^k}{k!} \sum_{\tau\in\mathcal{P}(K-k)} (\tau + k[1])_{\mathrm{p}} \sum_{\substack{(\tau_r)_{r\in\mathrm{supp}(\varphi)} \\ \|\tau_r\|\leq\varphi(r), \sum_r \tau_r=\tau}} \prod_r \binom{\varphi(r)}{\tau_r} r^{\mathrm{tt}(\tau_r)} \\
&= \sum_{\sigma\in\mathcal{P}(\sigma)} (\sigma)_{\mathrm{p}} \sum_{k=0}^{K} \frac{w^k}{k!} \sum_{\substack{(\sigma_r)_{r\in\varphi(r)} \\ \sigma_r\leq\varphi(r) \\ \sum_r \sigma_r + k[1]=\sigma}} \prod_r \binom{\varphi(r)}{\sigma_r} r^{\mathrm{tt}(\sigma_r)}
\end{aligned}$$

and we have recovered the sum of coefficients in $iid_K(\varphi)$. □

6.3 Relating the two forms of sampling

We can now make formal the relationship between ordinary iid sampling and element-free iid sampling. The first key result is that multiplicity count commutes with *iid*.

Proposition 6.6. *Multiplicity count commutes with iid sampling, i.e. the diagram on the left of equation* (1) *commutes.*

Proof. The crux of the argument is the same as in the the restricted setting of Jacobs [16, Prop. 17 (2)]. We give the full generalization in Appendix A.2. □

The second key result is that $base_\mu$ also commutes with *iid*.

Proposition 6.7. *Sampling elements from a base measure μ commutes with iid sampling, i.e. the diagram on the right of equation* (1) *commutes.*

Proof. This requires another combinatorial argument which we outline in Appendix A.3. □

These two propositions make precise the informal commutative diagrams in §1.1 and § 1.3, and together clarify the relationship between ordinary element-based sampling, and element-free sampling.

7 THE REPRESENTATION OF INFINITE RANDOM PARTITIONS

In this section we turn to Kingman's representation theorem for random partitions. This theorem is concerned with infinite random partitions, *i.e.* random partitions of $\mathbb{N}$, formalized as infinite sequences of random *finite* partitions subject to a consistency condition, which states that deleting a random point from the random partition at level $K+1$ should yield the random partition at level K.

7.1 Draw-delete maps for multisets and partitions

In order to state the theorem with the appropriate consistency condition, we recall the definitions of *draw-delete* kernels, both with and without elements (respectively, for multisets and for partitions). These represent the process of deleting an element from a multiset in $M[K+1](X)$, or a point from a partition in $\mathcal{P}(K+1)$, uniformly chosen among all $K+1$.

We first fix some notation. For $\varphi \in \mathcal{M}[K+1](X)$, and $x \in \mathrm{supp}(\varphi)$, the multiset $\varphi^{-x} \in \mathcal{M}[K](X)$ is given by $\varphi^{-x}(x') = \varphi(x') - [x = x']$, for $x' \in X$. Similarly, for $\sigma \in \mathcal{P}(K+1)$, and $n \in \mathrm{supp}(\sigma)$, the partition $\sigma^{n-} \in \mathcal{P}(K)$ is given for $n' > 0$ by

$$\sigma^{n-}(n') = \begin{cases} \sigma(n') - 1 & \text{if } n' = n \\ \sigma(n') + 1 & \text{if } n' = n-1 \\ \sigma(n') & \text{otherwise.} \end{cases}$$

Lemma 7.1. *Let X be a standard Borel space, and $K \geq 0$. There are kernels $dd : \mathcal{M}[K+1](X) \rightsquigarrow \mathcal{M}[K](X)$ and $pdd : \mathcal{P}(K+1) \rightsquigarrow \mathcal{P}(K)$ given by*

$$dd(\varphi) = \textstyle\sum_{x \in \mathrm{supp}(\varphi)} \frac{\varphi(x)}{\|\varphi\|} \cdot \delta_{\varphi^{-x}} \qquad pdd(\sigma) = \textstyle\sum_{n \in \mathrm{supp}(\sigma)} \frac{n \cdot \sigma(n)}{\mathrm{tt}(\sigma)} \cdot \delta_{\sigma^{n-}}$$

for $\varphi \in \mathcal{M}[K+1](X)$ and $\sigma \in \mathcal{P}(K+1)$.

Moreover the following diagram commutes in $\mathcal{K}\ell(G)$.

$$\begin{tikzcd} \mathcal{M}[K+1](X) \arrow[r, "dd"] \arrow[d, "mc"'] & \mathcal{M}[K](X) \arrow[d, "mc"] \\ \mathcal{P}(K+1) \arrow[r, "pdd"'] & \mathcal{P}(K) \end{tikzcd}$$

Proof. This is [16, Lemma 7]. We must additionally check the measurability of $dd : \mathcal{M}[K+1](X) \rightsquigarrow \mathcal{M}[K](X)$. For this we note that dd is equal to a composite $\mathcal{M}[K+1](X) \rightsquigarrow X^{K+1} \rightsquigarrow X^K \to \mathcal{M}[K](X)$ where the first kernel picks a random enumeration, the second one performs draw-and-delete on sequences, and the third map turns the resulting sequence back into a multiset. $\square$

We now show the corresponding property for $base_\mu$.

Lemma 7.2. *For every K, the following diagram commutes in $\mathcal{K}\ell(G)$.*

$$\begin{tikzcd} \mathcal{P}(K+1) \arrow[r, "pdd"] \arrow[d, "base_\mu"'] & \mathcal{P}(K) \arrow[d, "base_\mu"] \\ \mathcal{M}[K+1](X) \arrow[r, "dd"'] & \mathcal{M}[K](X) \end{tikzcd}$$

Proof. Let $\sigma \in \mathcal{P}(K+1)$ and let $\boldsymbol{x} \in X^{\|\sigma\|}$. Let $n_1 \leq \cdots \leq n_{\|\sigma\|}$ be the (unique) increasing enumeration of the blocks of σ. A straightforward calculation gives

$$dd(alloc(\boldsymbol{x}, \sigma)) = \sum_{i=1}^{\|\sigma\|} \frac{n_i}{K+1} \delta_{alloc(\boldsymbol{x},\sigma)^{-x_i}},$$

and thus for $V \in \Sigma_{\mathcal{M}[K](X)}$, $(dd \circ base_\mu)(V) =$

$$\begin{aligned} &\sum_{i=1}^{\|\sigma\|} \frac{n_i}{K+1} \int_{\boldsymbol{x} \in X^{\|\sigma\|}} \left[alloc(\boldsymbol{x}, \sigma)^{-x_i} \in V\right] iid_{\|\sigma\|}(\mathrm{d}\boldsymbol{x}) \\ &= \sum_{i=1}^{\|\sigma\|} \frac{n_i}{K+1} \int_{\boldsymbol{x} \in X^{\|\sigma^{n_i-}\|}} \left[alloc(\boldsymbol{x}, \sigma^{n_i-}) \in V\right] iid_{\|\sigma\|}(\mathrm{d}\boldsymbol{x}) \\ &= \sum_{n \in \mathrm{supp}\sigma} \frac{n \cdot \sigma(n)}{K+1} \int_{\boldsymbol{x} \in X^{\|\sigma^{n-}\|}} \left[alloc(\boldsymbol{x}, \sigma^{n-}) \in V\right] iid_{\|\sigma\|}(\mathrm{d}\boldsymbol{x}) \end{aligned}$$

which equals $(base_\mu \circ pdd)(\sigma)(V)$. $\square$

7.2 Main theorem

We consider the diagram of spaces of partitions in $\mathcal{K}\ell(G)$

$$\begin{tikzcd} \mathcal{P}(1) & \mathcal{P}(2) \arrow[l, "pdd"'] & \cdots \arrow[l, "pdd"'] & \mathcal{P}(n) \arrow[l] & \cdots \arrow[l, "pdd"'] \end{tikzcd}$$

consisting of the draw-delete kernels from Section 7.1. Our main theorem is as follows:

Theorem 7.3 (Kingman, categorical form). *The diagram*

$$\begin{tikzcd} & & & \nabla \arrow[llld, bend right, "iid_1"'] \arrow[lld, "iid_2"] \arrow[d, "iid_n"] \arrow[rd, "iid_{n+1}"] & & \\ \mathcal{P}(1) & \mathcal{P}(2) \arrow[l, "pdd"] & \cdots \arrow[l, "pdd"] & \mathcal{P}(n) \arrow[l] & \mathcal{P}(n+1) \arrow[l, "pdd"] & \cdots \arrow[l, "pdd"] \end{tikzcd} \tag{3}$$

is commutative, and the kernels $\{iid_n : \nabla \rightsquigarrow \mathcal{P}(n)\}$ form a limiting cone for the diagram of draw-delete maps.

As a consequence of this result, a consistent family of random partitions corresponds to a unique random distributions. Note that this theorem is stronger than Kingman's original result, since we could be considering a measurable family of consistent families (a cone over the diagram at an arbitrary object).

We will obtain this theorem as a corollary of its counterpart in traditional probability theory, with elements. This is a classical representation theorem for infinite exchangeable sequences due to de Finetti [10], and often regarded as a foundational result for Bayesian statistics. The categorical presentation of de Finetti's theorem is due to Jacobs and Staton [20, 21].

Theorem 7.4 (de Finetti, categorical form). *Let X be a standard Borel space. The kernels $iid_n : GX \rightsquigarrow \mathcal{M}[n](X)$ form a limiting* *cone for the diagram of draw-delete maps:*

$$\begin{tikzcd}
 & & & GX \arrow[llld, bend right, "iid_1"'] \arrow[lld, "iid_2"] \arrow[d, "iid_n"] \arrow[dr] & \\
\mathcal{M}[1](X) & \mathcal{M}[2](X) \arrow[l, "dd"] & \cdots \arrow[l, "dd"] & \mathcal{M}[n](X) \arrow[l] & \cdots \arrow[l, "dd"]
\end{tikzcd}$$

Our proof of Theorem 7.3 uses the following categorical fact[1]:

Lemma 7.5. *Let* $\mathbf{I}$ *be a small category, and let* $\mathbf{I}^{\triangleleft}$ *be the result of adding a free initial object to* $\mathbf{I}$.

- *A functor* $F : \mathbf{I}^{\triangleleft} \to \mathbf{C}$ *is the same thing as cone over the* $\mathbf{I}$*-shaped diagram* $\mathbf{I} \hookrightarrow \mathbf{I}^{\triangleleft} \xrightarrow{F} \mathbf{C}$ *in* $\mathbf{C}$.
- *For functors* $F, G : \mathbf{I}^{\triangleleft} \to \mathbf{C}$ *such that* G *is a retract of* F *in the functor category* $[\mathbf{I}^{\triangleleft}, \mathbf{C}]$, i.e. *we have natural transformations* $s : F \to G$ *and* $t : G \to F$ *such that* $t \circ s = \mathrm{id}_G$, *then if* F *is a limit cone, so is* G.

Our version of Kingman's theorem follows from this fact, because the cone over partitions is a retract of the de Finetti cone over multisets.

Proof of Theorem 7.3. We instantiate Theorem 7.4 at $X = [0, 1]$ which gives a limit cone consisting of kernels $iid_n : G[0, 1] \rightsquigarrow \mathcal{M}[n]([0, 1])$. The rest of the proof follows from the series of results in this paper, showing that the families of maps $base_{\mathscr{U}_{[0,1]}}$ and *mc*, either between partitions and multisets or between distributions and element-free distributions, form section-retraction pairs which appropriately commute with *dd*, *pdd*, and iid_K. We deduce that the diagram in (3) commutes since for each K,

$$\begin{tikzcd}
 & & \nabla \arrow[d, "base_{\mathscr{U}_{[0,1]}}"] \arrow[ddll, bend right, "iid_K"'] \arrow[ddrr, bend left, "iid_{K+1}"] & & \\
 & & G[0,1] \arrow[dl, "iid_K"'] \arrow[dr, "iid_{K+1}"] & & \\
\mathcal{P}(K) & \mathcal{M}[K]([0,1]) \arrow[l, "mc"] & & \mathcal{M}[K+1]([0,1]) \arrow[ll, "dd"] \arrow[r, "mc"'] & \mathcal{P}(K+1) \arrow[llll, bend left=20, "pdd"]
\end{tikzcd}$$

commutes. This shows that the maps iid_K form a cone over the diagram of sets of partitions. A similar argument shows that we have a section-retraction pair in the functor category $[(\mathbb{N}^{\mathrm{op}})^{\triangleleft}, \mathcal{K}\ell(G)]$, where $\mathbb{N}$ is the category $\{1 \to 2 \to \cdots\}$. We conclude by Lemma 7.5. □

8 RANDOM ELEMENT-FREE DISTRIBUTIONS AS NATURAL TRANSFORMATIONS

In this section we show how to arrive at element-free distributions from another angle. Recall that, for every standard Borel space X, we have a kernel $base_\mu : \nabla \rightsquigarrow GX$ (Definition 4.6). We observe that the same definition actually gives a kernel

$$base_X : GX \times \nabla \rightsquigarrow GX$$

where the base measure μ is regarded as an additional argument.

Given a distribution $\omega \in G\nabla$ on element-free distributions (that is, $\omega : \mathbf{1} \rightsquigarrow \nabla$), we construct a kernel denoted $(\omega \gg\!= base_X)$ as follows:

$$GX \cong \mathbf{1} \times GX \xrightarrow{\omega \times GX} \nabla \times GX \xrightarrow{base_X} GX$$

This kernel defines a natural transformation:

Proposition 8.1. *For* $\omega \in G\nabla$, *the family of measurable functions* $\omega \gg\!= base_X : GX \to GGX$, *for* $X \in \mathbf{Sbs}$, *defines a natural transformation between functors* $\mathbf{Sbs} \to \mathbf{Sbs}$. *More concretely, for every measurable function* $f : X \to Y$, *the following diagram commutes:*

$$\begin{tikzcd}
GX \arrow[r, "\omega \gg= base_X"] \arrow[d, "Gf"'] & GGX \arrow[d, "GGf"] \\
GY \arrow[r, "\omega \gg= base_Y"'] & GGY
\end{tikzcd}$$

Proof. Although we stated the lemma in this way for the purposes of the section below, this follows easily from the more primitive fact that each $base_X(-, \varphi)$ is natural, which in turn is an easy consequence of the naturality of $iid_{\|\varphi\|}$. □

Proposition 8.1 defines a function $G\nabla \to \mathrm{Nat}(G, GG)$, where $\mathrm{Nat}(G, GG)$ is the set of all natural transformations. The main theorem of this section states that this function is bijective, *i.e.* that every natural transformation $G \to GG$ is of the form $\omega \gg\!= base$ for a unique $\omega \in G\nabla$.

Our proof relies on the following well-known property of standard Borel spaces. (This is sometimes known as the randomization lemma, *e.g.* [23, Lemma 3.22].)

Lemma 8.2. *For every standard Borel space* X *and probability measure* $\mu \in GX$, *there exists* $f : [0, 1] \to X$ *such that* $\mu = G(f)(\mathscr{U}_{[0,1]})$.

We deduce that a natural transformation $G \to GG$ is determined by its action on the uniform distribution $\mathscr{U}_{[0,1]}$.

Lemma 8.3. *For natural transformations* $H, H' : G \to GG$, *if* $H_{[0,1]}(\mathscr{U}_{[0,1]}) = H'_{[0,1]}(\mathscr{U}_{[0,1]})$, *then* $H = H'$.

Proof. Let $X \in \mathbf{Sbs}$ and $\mu \in GX$. We show that $H_X(\mu) = H'_X(\mu)$. By Lemma 8.2, there is some $f : [0, 1] \to X$ such that $\mu = G(f)(\mathscr{U}_{[0,1]})$. By naturality, we have that

$$H_X(\mu) = H_X(G(f)(\mathscr{U}_{[0,1]})) = GG(f)(H_{[0,1]}(\mathscr{U}_{[0,1]})),$$

and similarly for H'_X. Thus

$$H_X(\mu) = GG(f)(H_{[0,1]}(\mathscr{U}_{[0,1]})) = GG(f)(H'_{[0,1]}(\mathscr{U}_{[0,1]})) = H'_X(\mu)$$

and we are done. □

Theorem 8.4. *There is an isomorphism* $G\nabla \cong \mathrm{Nat}(G, GG)$, *given by the pair of inverse maps* α *and* β *given below.*

$$\alpha : G\nabla \longrightarrow \mathrm{Nat}(G, GG) : \omega \longmapsto \{\omega \gg\!= base_X\}_{X \in \mathbf{Sbs}}$$
$$\beta : \mathrm{Nat}(G, GG) \longrightarrow G\nabla : H \longmapsto G(mc)(H_{[0,1]}(\mathscr{U}_{[0,1]}))$$

Proof. The fact that $\beta \circ \alpha = \mathrm{id}_{G\nabla}$ follows essentially from the retract property of Lemma 4.9, to which we apply G. We show $\alpha \circ \beta = \mathrm{id}_{\mathrm{Nat}(G,GG)}$, for which it suffices to show that β is injective. So we suppose $H, H' \in \mathrm{Nat}(G, GG)$ satisfy $G(mc)(H_{[0,1]}(\mathscr{U}_{[0,1]})) = G(mc)(H_{[0,1]}(\mathscr{U}_{[0,1]}))$.

To show $H = H'$, it suffices to show that $H_{[0,1]}(\mathscr{U}_{[0,1]}) = H_{[0,1]}(\mathscr{U}_{[0,1]}) \in GG[0, 1]$, by Lemma 8.3. By the *uniqueness* part

[1] A proof is available at https://ncatlab.org/nlab/show/retract#RetractsOfDiagrams .

of de Finetti's theorem (Theorem 7.4), it suffices to show that the family of iid measures

$$\mu_K \quad = \quad 1 \xrightarrow{H_{[0,1]}(\mathscr{U}_{[0,1]})} G[0,1] \xrightarrow{iid_K} \mathcal{M}[K]([0,1]),$$

for $K \geq 0$, coincides with the family of measures μ'_K induced by H' in the same way. So we now fix $K \geq 0$ and show $\mu'_K = \mu_K$.

Recall that $G(mc)(H_{[0,1]}(\mathscr{U}_{[0,1]})) = G(mc)(H_{[0,1]}(\mathscr{U}_{[0,1]}))$ by assumption, and that by Lemma 6.6 *mc* commutes with iid_K. From this we deduce that $G(mc)(\mu_K) = G(mc)(\mu'_K) \in G(\mathcal{P}(K))$.

At this point we observe that, by naturality of H, if $\phi : [0,1] \to [0,1]$ is a uniform-measure-preserving function, then the naturality square for ϕ says that $GG(\phi)(H_{[0,1]}(\mathscr{U}_{[0,1]})) = H_{[0,1]}(\mathscr{U}_{[0,1]})$. Then the naturality of iid_K shows that $G(\mathcal{M}[K](\phi))(\mu_K) = \mu_K$.

In summary, we have shown that the distribution μ_K on $\mathcal{M}[K]([0,1])$ is invariant under any uniform-measure-preserving transformation of $[0,1]$. This turns out to imply that μ_K is completely determined by its pushforward under *mc*, a distribution on partitions. We show this in Appendix A.4.

Hence, since $G(mc)(\mu_K) = G(mc)(\mu'_K)$, we have $\mu_K = \mu'_K$, and we are done. □

9 CONCLUSIONS

Summary. We have given new definitions and tools for the element-free distributions of Kingman [24] in the study of partition structures. Our formalization makes heavy use of multisets, showing that the coefficients of an element-free distribution do not need to be ordered.

We have studied the two key operations relating ordinary and element-free probability: multiplicity count, which forgets elements, and the operation of drawing new elements from a base measure. The latter seems new to this paper.

Our development uses technical measure-theoretic tools, but we emphasize that including continuous distributions is crucial for our main representation theorems to hold. (For an edge-case example, the infinite random partition which has all singletons with probability 1 can only be represented by a nonatomic distribution.)

Related work. A major inspiration for this paper is the line of research by Jacobs on structural probability theory (*e.g.* [17–19]). The idea of multiset-based element-free probability is due to him, and we have referred to many of his results on partitions [16]. Our version of Kingman's theorem is the answer to an open problem in [16, §12], and our proof relies on key insights about probabilistic representation theorems due to Jacobs and Staton [21]. For multisets over standard Borel spaces, another key inspiration is the monad for point processes of Dash and Staton [8].

In another line of research, Danos and Garnier have observed [6] that the Dirichlet Process determines a natural transformation. This inspired the work of Section 8, and our Theorem 8.4 should recover the stick-breaking weights for the DP, in the form of an element-free distribution.

Perspectives. There are several avenues for further work.

Nonparametric Bayesian processes influenced the early development of probabilistic programming [15]. More recently typed languages with polymorphism were designed for probabilistic modelling [7, 30], and it seems clear that a polymorphic implementation of the Dirichlet process must treat the elements and the element-free part independently. Thus an interface for element-free distributions would be convenient. Additionally a theoretical investigation relating naturality and polymorphism in a probabilistic setting (in the style of [34]) seems important for proper foundations of nonparametric probabilistic languages.

Beyond element-free distributions, we can ask whether similar constructions and theorems exist for arbitrary element-free *unnormalized* measures. These measures are fundamental in probabilistic programming [32], and random measures also play a key role in nonparametric statistics (*e.g.* as point processes [25] or as latent feature models [29, 33]), also involving base measures.

We finally note that multisets and probabilities have been combined in the context of probabilistic models of linear logic (*e.g.* [5]). A deeper connection to investigate is the construction of the free linear exponential (*e.g.* [3, 4]) using a universal categorical construction over diagrams of measures, similar to Theorem 7.3 and Theorem 7.4.

REFERENCES

[1] George E Andrews. 1998. *The theory of partitions.* Number 2. Cambridge university press.

[2] Erhan Cinlar. 2011. *Probability and stochastics.* Vol. 261. Springer.

[3] Raphaëlle Crubillé, Thomas Ehrhard, Michele Pagani, and Christine Tasson. 2017. The free exponential modality of probabilistic coherence spaces. In *International Conference on Foundations of Software Science and Computation Structures.* Springer, 20–35.

[4] Fredrik Dahlqvist and Dexter Kozen. 2019. Semantics of higher-order probabilistic programs with conditioning. *Proceedings of the ACM on Programming Languages* 4, POPL (2019), 1–29.

[5] Vincent Danos and Thomas Ehrhard. 2011. Probabilistic coherence spaces as a model of higher-order probabilistic computation. *Information and Computation* 209, 6 (2011), 966–991.

[6] Vincent Danos and Ilias Garnier. 2015. Dirichlet is natural. *Electronic Notes in Theoretical Computer Science* 319 (2015), 137–164.

[7] Swaraj Dash, Younesse Kaddar, Hugo Paquet, and Sam Staton. [n.d.]. Affine monads and lazy structures for Bayesian programming. *POPL'2022* ([n. d.]).

[8] Swaraj Dash and Sam Staton. 2020. A Monad for Probabilistic Point Processes. In *Proceedings of the 3rd Annual International Applied Category Theory Conference 2020, ACT 2020, Cambridge, USA, 6-10th July 2020 (EPTCS)*, David I. Spivak and Jamie Vicary (Eds.), Vol. 333. 19–32. https://doi.org/10.4204/EPTCS.333.2

[9] Swaraj Dash and Sam Staton. 2021. Monads for Measurable Queries in Probabilistic Databases. In *Proceedings 37th Conference on Mathematical Foundations of Programming Semantics, MFPS 2021, Hybrid: Salzburg, Austria and Online, 30th August - 2nd September, 2021 (EPTCS)*, Ana Sokolova (Ed.), Vol. 351. 34–50. https://doi.org/10.4204/EPTCS.351.3

[10] Bruno De Finetti. 1929. Characteristic function of a random phenomenon. In *Proceedings of the International Congress of Mathematicians: Bologna from 3 to 10 September 1928.* 179–190.

[11] Thomas S Ferguson. 1973. A Bayesian analysis of some nonparametric problems. *The annals of statistics* (1973), 209–230.

[12] Tobias Fritz, Tomáš Gonda, and Paolo Perrone. 2021. De Finetti's Theorem in Categorical Probability. *Journal of Stochastic Analysis* 2, 4 (2021), 6.

[13] Andrew Gelman, John B Carlin, Hal S Stern, and Donald B Rubin. 1995. *Bayesian data analysis.* Chapman and Hall/CRC.

[14] Michele Giry. 1982. A categorical approach to probability theory. In *Categorical aspects of topology and analysis.* Springer, 68–85.

[15] Noah D Goodman, Vikash K Mansinghka, Daniel Roy, Keith Bonawitz, and Joshua B Tenenbaum. 2008. Church: a language for generative models. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence.* 220–229.

[16] Bart Jacobs. 2022. Partitions and Ewens distributions in element-free probability theory. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science.* 1–9.

[17] Bart Jacobs. 2022. A Reconstruction of Ewens' Sampling Formula via Lists of Coins. In *A Journey from Process Algebra via Timed Automata to Model Learning: Essays Dedicated to Frits Vaandrager on the Occasion of His 60th Birthday.* Springer, 339–357.

[18] Bart Jacobs. 2022. Stick breaking, in coalgebra and probability. In *International Workshop on Coalgebraic Methods in Computer Science.* Springer, 176–193.

[19] Bart Jacobs. 2022. Urns and Tubes. *Compositionality* 4 (Dec. 2022), 4. https://doi.org/10.32408/compositionality-4-4

[20] Bart Jacobs and Sam Staton. 2020. Appendix to CMCS 2020: general case of de Finetti. (2020).

[21] Bart Jacobs and Sam Staton. 2020. De Finetti's construction as a categorical limit. In *International Workshop on Coalgebraic Methods in Computer Science*. Springer, 90–111.

[22] Bart Jacobs and Dario Stein. 2023. Counting and Matching. In *31st EACSL Annual Conference on Computer Science Logic, CSL 2023, February 13-16, 2023, Warsaw, Poland (LIPIcs)*, Bartek Klin and Elaine Pimentel (Eds.), Vol. 252. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 28:1–28:15. https://doi.org/10.4230/LIPICS.CSL.2023.28

[23] Olav Kallenberg. 2021. *Foundations of Modern Probability* (hardcover ed.). Springer. 958 pages.

[24] John FC Kingman. 1978. The representation of partition structures. *Journal of the London Mathematical Society* 2, 2 (1978), 374–380.

[25] Günter Last and Mathew Penrose. 2017. *Lectures on the Poisson process*. Vol. 7. Cambridge University Press.

[26] F William Lawvere. 1963. Functorial semantics of algebraic theories. *Proceedings of the National Academy of Sciences* 50, 5 (1963), 869–872.

[27] Jim Pitman. 2006. *Combinatorial stochastic processes: Ecole d'eté de probabilités de saint-flour xxxii-2002*. Springer.

[28] Jim Pitman and Marc Yor. 1997. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *The Annals of Probability* (1997), 855–900.

[29] Daniel M Roy. 2014. The continuum-of-urns scheme, generalized beta and Indian buffet processes, and hierarchies thereof. *arXiv preprint arXiv:1501.00208* (2014).

[30] Adam Ścibior, Zoubin Ghahramani, and Andrew D Gordon. 2015. Practical probabilistic programming with monads. In *Proceedings of the 2015 ACM SIGPLAN Symposium on Haskell*. 165–176.

[31] Jayaram Sethuraman. 1994. A constructive definition of Dirichlet priors. *Statistica sinica* (1994), 639–650.

[32] Sam Staton. 2017. Commutative semantics for probabilistic programming. In *Programming Languages and Systems: 26th European Symposium on Programming, ESOP 2017, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2017, Uppsala, Sweden, April 22–29, 2017, Proceedings 26*. Springer, 855–879.

[33] Romain Thibaux and Michael I Jordan. 2007. Hierarchical beta processes and the Indian buffet process. In *Artificial intelligence and statistics*. PMLR, 564–571.

[34] Philip Wadler. 1989. Theorems for free!. In *Proceedings of the fourth international conference on Functional programming languages and computer architecture*. 347–359.

A OMITTED PROOFS

A.1 Omitted details in the proof that $mc : GX \to \nabla$ is measurable

Recall the notation in the proof of Lemma 4.4: $\varepsilon > 0$, $\alpha \geq \varepsilon$, and for any $m \in \mathbb{N}$ and $n \leq 2^m - 1$ we write $B_m^n = [\frac{n}{2^m}, \frac{n+1}{2^m})$.

We claimed that

$$mc_\varepsilon^{-1} D_k^{[\alpha,1]} = \bigcup_{\ell \in \mathbb{N}} \bigcap_{m \geq \ell} \bigcup_{\substack{n_1,\ldots,n_k \\ \in \{0,\ldots,2m-1\} \\ \text{distinct}}} \left(\bigcap_{i=1}^{k} \mathrm{ev}_{B_m^{n_i}}^{-1} [\alpha, 1] \right) \cap \left(\bigcap_{\substack{n \leq 2m-1 \\ \forall i.\ n \neq n_i}} \mathrm{ev}_{B_m^n}^{-1} [0, \alpha) \right).$$

Note that the left-hand side consists of measures μ on $[0, 1]$ having exactly k atoms whose weight is in $[\alpha, 1]$. We prove inclusions in both directions.

($\subseteq$). Let $\mu \in mc_\epsilon^{-1} D_k^{[\alpha,1)}$ and call $x_1, ..., x_k$ the k distinct atoms of μ with weight in $[\alpha, 1]$.

For $i \in \{1, ..., k\}$ and $m \in \mathbb{N}$, let $n_{m,i}$ be the index, in the partition at level m, of the block containing x_i , *i.e.* $x_i \in B_{n_{m,i},m}$. For every $m \in \mathbb{N}$ we clearly have $\mu(B_{n_{m,i},m}) \in [\alpha, 1]$. Since the atoms x_i are distinct there must be some $\ell_0 \in \mathbb{N}$ such that for every $m \geq \ell_0$ the $n_{m,i}$ are all distinct.

Next, we show that there must be some $\ell \geq \ell_0$ such that, for all $m \geq \ell$ and $n \in \{0, \ldots, 2^m - 1\}$ such that $n \neq n_{m,i}$ for all i, $\mu(B_m^n) < \alpha$. For a contradiction, suppose that for all $m \in \mathbb{N}$, there exists n_m distinct from the $n_{m,i}$ such that $\mu(B_m^n) \geq \alpha$. Observe that there must be a subsequence $(n_{\phi(m)})_m$ of $(n_m)_m$ such that $B_{\phi(m)}^{n_{\phi(m)}} \supseteq B_{\phi(m+1)}^{n_{\phi(m+1)}}$, or else there would be infinitely many disjoint intervals of measure $\alpha > 0$; impossible since μ is finite. But the intersection $\bigcap_m B_{\phi(m)}^{n_{\phi(m)}}$ is a singleton $\{y\}$, and $\mu\{y\} = \lim_{m\to\infty} \mu(B_{\phi(m)}^{n_{\phi(m)}}) \geq \alpha$. Thus y is an atom of μ, distinct from the x_i and with weight $\geq \alpha$, a contradiction.

($\supseteq$). We now suppose that there is $\ell \in \mathbb{N}$ such that, for all $m \geq \ell$, there exists $n_{m,1} < \ldots < n_{m,k}$ with

$$\mu(B_m^n) \geq \alpha \iff \exists i.\ n = n_{m,i}.$$

Again we must have, for each i, a subsequence $(\phi(m, i))_m$ of $(n_{m,i})_m$, such that $B_{\phi(m,i)}^{n_{\phi(m,i),i}} \supseteq B_{\phi_{m+1,i}}^{n_{\phi_{m+1,i},i}}$. Then the intersection $\bigcap_m B_{\phi(m,i)}^{\phi(m,i)}$ must be a singleton $\{x_i\}$, and we have $\mu(x_i) \geq \alpha$.

The $n_{m,i}$ were assumed distinct and so the x_i are distinct. On the other hand, any other atom with weight $\geq \alpha$ would violate the $\iff$ property above. Hence μ has exactly k atoms whose weight is in $[\alpha, 1]$, and so the claim is true.

A.2 Proof that *mc* commutes with *iid*

Proposition 6.6. *Multiplicity count commutes with iid sampling, i.e. the diagram on the left of equation* (1) *commutes.*

Proof. Fix a probability measure $\mu \in GX$. We call $\mathcal{A}$ the set of atoms of μ, and for $r \in (0, 1]$ we write $\mathcal{A}_r$ for the subset of atoms with weight r, *i.e.* $\mathcal{A}_r = \{x \in X \mid \mu\{x\} = r\}$. We set $\omega = \mu(X \setminus \mathcal{A})$, the weight of the nonatomic part of μ.

Note that $iid_K(\mu)$ is supported on the subset of multisets $\varphi \in \mathcal{M}[K](X)$ such that $\varphi(x) \leq 1$ for every $x \in X \setminus \mathcal{A}$. Thus it suffices to show that, for all $\sigma \in \mathcal{P}(K)$,

$$iid_K(\mu)(S) = iid_K(mc(\mu))(\sigma) \tag{4}$$

where $S = \{\varphi \in mc^{-1}(\sigma) \mid \forall x \in X \setminus \mathcal{A}.\ \varphi(x) \leq 1\}$.

We proceed to reduce the LHS to the formula for iid_K on partitions (6.2). The key observation is that every $\varphi \in S$ decomposes uniquely as $\chi + \sum_{r\in R} \varphi_r$, where R is the set of weights of atoms of μ, *i.e.* $R = \mathrm{supp}(mc(\mu))$; for each $r \in R$, $\varphi_r \in \mathcal{M}(\mathcal{A}_r)$; and $\chi \in \mathcal{M}(X \setminus \mathcal{A})$ is a multiset with no duplicates (for all $x \in X \setminus \mathcal{A}$, $\chi(x) \leq 1$). Thus

$$\begin{aligned} iid_K(\mu)(S) &= iid_K(\mu) \left(\bigcup_{k=0}^{K} \bigcup_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} \left\{ \sum_r \varphi_r + \chi \mid \chi \in \mathcal{M}[k](X \setminus A), \forall x.\ \chi(x) \leq 1 \right\} \right) \\ &= \sum_{k=0}^{K} \sum_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} iid_K(\mu) \left(\left\{ \sum_r \varphi_r + \chi \mid \chi \in \mathcal{M}[k](X \setminus A), \forall x.\ \chi(x) \leq 1 \right\} \right) \end{aligned}$$

because the unions are disjoint and countable. Now, for each family $(\varphi_r)_{r\in R}$,

$$iid_K(\mu) \left(\left\{ \sum_r \varphi_r + \chi \mid \chi \in \mathcal{M}[k](X \setminus A), \forall x.\ \chi(x) \leq 1 \right\} \right) = iid_K(\mu) \left(\left\{ \sum_{r\in R} \varphi_r + \chi \mid \chi \in \mathcal{M}[k](X \setminus A) \right\} \right)$$

because the difference between the two sets is a null set for $iid_K(\mu)$. Therefore,

$$\begin{aligned}
iid_K(\mu)(S) &= \sum_{k=0}^{K} \sum_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} iid_K(\mu)\left(\left\{\sum_{r\in R} \varphi_r + \chi \mid \chi \in \mathcal{M}[k](X \setminus A)\right\}\right) \\
&= \sum_{k=0}^{K} \sum_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} iid_K(\mu)\left(D_k^{X\setminus\mathcal{A}} \cap \bigcap_{r\in R}\bigcap_{x\in\mathcal{A}_r} D^{\{x\}}_{\varphi_r(x)}\right) \\
&= \sum_{k=0}^{K} \sum_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} \frac{K!}{k!\prod_r \prod_{x\in\mathcal{A}_r} \varphi_r(x)!}\mu(X\setminus\mathcal{A})^k \prod_{r\in R}\prod_{x\in\mathcal{A}_r} \mu\{x\}^{\varphi_r(x)} \\
&= (\sigma)_{\mathrm{p}} \sum_{k=0}^{K} \frac{\omega^k}{k!} \sum_{\substack{(\varphi_r \in \mathcal{M}(\mathcal{A}_r))_{r\in R} \\ \sum_r mc(\varphi_r)+k[1]=\sigma}} \prod_{r\in R} r^{\|\varphi_r\|} \\
&= (\sigma)_{\mathrm{p}} \sum_{k=0}^{K} \frac{\omega^k}{k!} \sum_{\substack{(\sigma_r)_{r\in R} \\ \|\sigma_r\| \le |\mathcal{A}_r| \\ \sum_r \sigma_r+k[1]=\sigma}} \prod_{r\in R} \binom{|\mathcal{A}_r|}{\sigma_r} r^{\mathrm{tt}\sigma_r} \\
&= iid_K(mc(\mu))(\sigma).
\end{aligned}$$

□

A.3 Proof that $base_\mu$ commutes with *iid*

PROPOSITION 6.7. *Sampling elements from a base measure μ commutes with iid sampling, i.e. the diagram on the right of equation* (1) *commutes.*

To prove the lemma we make use of the well-known multinomial formula for computing finite powers of sums:

$$\left(\sum_{i\in I} a_i\right)^N = \sum_{\chi\in\mathcal{M}[N](I)} (\chi) \prod_i a_i^{\chi}(i).$$

PROOF. For ease of presentation we detail the case of a discrete $\varphi \in \nabla_d$, which contains the main combinatorial idea. The continuous extension adds no difficulty since it is treated deterministically by $base_\mu$.

Fix $\varphi \in \nabla_d$. For $U \in GX$ and $k \le K$, we show that $(base_\mu \circ iid_K)(\varphi)(E_k^U) = (iid_K \circ base_\mu)(\varphi)(E_k^U)$.

$$\begin{aligned}
&(iid_K \circ base_\mu)(\varphi)(E_k^U) \\
&= \int_{\boldsymbol{x}\in X^{\|\varphi\|}} iid_K\left(\sum_{r,1\le i\le\varphi(r)} r\cdot\delta_{x_i^r}\right)\left(E_k^U\right) \mathrm{d}iid_{\|\varphi\|}(\mu) && \text{(by definition of composition)} \\
&= \int_{\boldsymbol{x}\in X^{\|\varphi\|}} \binom{K}{k}\left(\Big(\sum_{r,1\le i\le\varphi(r)} r\cdot\delta_{x_i^r}\Big)(U)\right)^k \left(\Big(\sum_{r,1\le i\le\varphi(r)} r\cdot\delta_{x_i^r}\Big)(U^\complement)\right)^{K-k} \mathrm{d}iid_{\|\varphi\|}(\mu) && \text{(by definition of } iid\text{)} \\
&= \binom{K}{k}\int_{\boldsymbol{x}\in X^{\|\varphi\|}} \left(\sum_r r\cdot\#\{1\le i\le\varphi(r)\mid x_i^r\in U\}\right)^k \left(\sum_r r\cdot\#\{1\le i\le\varphi(r)\mid x_i^r\notin U\}\right)^{K-k} \mathrm{d}iid_{\|\varphi\|}(\mu) && \text{(rearranging)} \\
&= \binom{K}{k} \sum_{\substack{\chi\in\mathcal{M}[k] \\ (\{(r,i)\mid i\le\varphi(r)\})}} \sum_{\substack{\rho\in\mathcal{M}[K-k] \\ (\{(r,i)\mid i\le\varphi(r)\})}} (\chi)(\rho)\int_{\boldsymbol{x}\in X^{\|\varphi\|}} \prod_{r,1\le i\le\varphi(r)} ([x_i^r\in U]\cdot r)^{\chi(r,i)}([x_i^r\notin U]\cdot r)^{\rho(r,i)}\mathrm{d}iid_{\|\varphi\|}(\mu) && \text{(multinomial formula)} \\
&= \binom{K}{k} \sum_{\substack{\chi\in\mathcal{M}[k] \\ (\{(r,i)\mid i\le\varphi(r)\})}} \sum_{\substack{\rho\in\mathcal{M}[K-k] \\ (\{(r,i)\mid i\le\varphi(r)\}) \\ \mathrm{supp}(\rho)\cap\mathrm{supp}(\chi)=\emptyset}} (\chi)(\rho)\mu(U)^{|\mathrm{supp}(\chi)|}\mu(U^\complement)^{|\mathrm{supp}(\rho)|} \prod_{r,1\le i\le\varphi(r)} r^{\chi(r,i)+\rho(r,i)}
\end{aligned}$$

where the last step is by an easy inspection and rearranging of the *iid* integral; in particular the integral is 0 when χ and ρ have overlapping support.

Now observe that we can turn any pair (χ, ρ) (as in the sum indices above) into a pair $((\sigma_r)_{r\in\mathrm{supp}(\varphi)}, \tau)$. Formally each σ_r is mc applied to the multiset $\chi + \rho$ restricted to pairs of the form (r, i), and τ is simply defined as $mc(\chi)$. So in particular $\sum_r \sigma_r \in \mathcal{P}(K)$, $\tau \in \mathcal{P}(k)$, $\|\sigma_r\| \le \varphi(r)$, and $\tau \le \sum_r \sigma_r$.

In this change-of-variable operation, we note that $\binom{K}{k}\boldsymbol{(}\chi\boldsymbol{)(}\rho\boldsymbol{)} = \boldsymbol{(}\sum_r \sigma_r\boldsymbol{)}_{\mathrm{p}}$, and that each pair $((\sigma_r)_r, \tau)$ satisfying the conditions of the previous sentence has precisely $\binom{\sum_r \sigma_r}{\tau}\prod_r \binom{\varphi(r)}{\sigma_r}$ antecedents. We note also that $\#\mathrm{supp}(\chi) = \|\tau\|$ and $\#\mathrm{supp}(\rho) = \sum_r \|\sigma_r\| - \|\tau\|$.

Therefore we can rewrite the above expression as:

$$\begin{aligned}
&\sum_{\substack{(\sigma_r)_{r\in\mathrm{supp}\varphi}\\ \sum_r \sigma_r\in\mathcal{P}(K)\\ \|\sigma_r\|\le\varphi(r)}} \sum_{\substack{\tau\le\sum_r\sigma_r\\ \mathrm{tt}\tau=k}} \boldsymbol{(}\textstyle\sum_r\sigma_r\boldsymbol{)}_{\mathrm{p}} \binom{\sum_r \sigma_r}{\tau}\left(\prod_r \binom{\varphi(r)}{\sigma_r} r^{\mathrm{tt}(\sigma_r)}\right)\mu(U)^{\|\tau\|}\mu(U^{\complement})^{\sum_r\|\sigma_r\|-\|\tau\|} \\
&= \sum_{\sigma\in\mathcal{P}(K)} \sum_{\substack{(\sigma_r)_{r\in\mathrm{supp}\varphi}\\ \sum_r\sigma_r=\sigma\\ \|\sigma_r\|\le\varphi(r)}} \sum_{\substack{\tau\le\sigma\\ \mathrm{tt}(\tau)=k}} \boldsymbol{(}\sigma\boldsymbol{)}_{\mathrm{p}}\binom{\sigma}{\tau}(\prod_r\binom{\varphi(r)}{\sigma_r} r^{\mathrm{tt}(\sigma_r)})\mu(U)^{\|\tau\|}\mu(U^{\complement})^{\|\sigma\|-\|\tau\|} && \text{(rearranging with } \sigma = \textstyle\sum_r \sigma_r)\\
&= \sum_{\sigma\in\mathcal{P}(K)} \boldsymbol{(}\sigma\boldsymbol{)}_{\mathrm{p}} \left(\sum_{\substack{(\sigma_r)_{r\in\mathrm{supp}\varphi}\\ \sum_r\sigma_r=\sigma\\ \|\sigma_r\|\le\varphi(r)}} \prod_r\binom{\varphi(r)}{\sigma_r} r^{\mathrm{tt}\sigma_r}\right)\left(\sum_{\substack{\tau\le\sigma\\ \mathrm{tt}(\tau)=k}}\binom{\sigma}{\tau}\mu(U)^{\|\tau\|}\mu(U^{\complement})^{\|\sigma\|-\|\tau\|}\right) && \text{(rearranging)}\\
&= \sum_{\sigma\in\mathcal{P}(K)} iid_K(\varphi)(\sigma)\, base_\mu(\sigma)(E^U_k) && \text{(by definition)}\\
&= (base_\mu \circ iid_K)(\varphi)(E^U_k) && \text{(by definition)}
\end{aligned}$$

and so the argument is complete. □

A.4 Omitted details in the proof of Theorem 8.4

Let $\sigma \in \mathcal{P}(K)$ be a partition of K and $m \in \mathbb{N}$. We want to show that for every $k = k^1 \oplus \ldots \oplus k^K$ tuple of $\|\sigma\|$ distinct elements of $\{0, \ldots, 2^m - 1\}$, the measure of

$$C(\sigma, m, k) := mc^{-1}(\sigma) \cap \bigcap_{p=1}^{K}(\bigcap_{i=1}^{\sigma(p)} E_p^{[\frac{k_i^p}{2^m}, \frac{k_i^p+1}{2^m})})$$

by μ_K is independent from k. To do so, let

$$A := \{k \in \{0, \ldots, m-1\}^{\|\sigma\|}, \text{distinct} \mid \mu_K(C(\sigma, m, k)) = \mu_K(C(\sigma, m, (0, \ldots, \|\sigma\| - 1)))\}$$

We will show that A contains every tuples of $\|\sigma\|$ distinct elements in $\{0, \ldots, 2^m - 1\}$. First, consider a transposition $\tau = (ab)$ of $\{0, \ldots, 2^m - 1\}$. Let us prove that A is stable under τ. Let $k \in A$ and consider the isomorphism $f_\tau\colon [0, 1] \to [0, 1]$ defined as $f_\tau(x) = \frac{b-a}{2^m} + x$ if $x \in [\frac{a}{2^m}, \frac{a+1}{2^m})$, $f_\tau(x) = \frac{a-b}{2^m} + x$ if $x \in [\frac{b}{2^m}, \frac{b+1}{2^m})$ and $f_\tau(x) = x$ otherwise.

We have

$$G(\mathcal{M}(f_\tau))\mu_K(C(\sigma, m, k)) = \mu_K(\mathcal{M}(f_\tau)(C(\sigma, m, k))) = \mu_K(C(\sigma, m, \tau(k)))$$

Then, A is stable by any permutation as the transpositions generate the permutations. Its follows that A contains every tuples of $\|\sigma\|$ distinct elements in $\{0, \ldots, 2^m - 1\}$. The last step of the proof is to deduce that we can express $\mu_K(C(\sigma, m, k))$ just with $mc \circ \mu_K$.

First, remark that for all k and k' tuples of $\|\sigma\|$ distinct elements of $\{0, \ldots, 2^m - 1\}$, either the sets $C(\sigma, m, k)$ and $C(\sigma, m, k')$ are disjoint or equal. Moreover, they are equal if and only if for all $p \in \{1, \ldots, K\}$, $k^{\sigma(p)}$ is a permutation of $k'^{\sigma(p)}$. Then, let us define $\mathcal{B}_m$ to be the set of all lists of $\|\sigma\|$ distinct elements in $\{0, \ldots, 2^m - 1\}$ such that for every $p \in \{1, \ldots, K\}$ the list $k^{\sigma(p)}$ is ordered.

From this, we have that

$$\begin{aligned} mc \circ \mu_K(\sigma) &= \mu_K(mc^{-1}(\sigma)) \\ &= \mu_K(\bigcup_{n \in \mathbb{N}} \bigcup_{k \in \mathcal{B}_m} C(\sigma, m, k)) \\ &= \lim_{m \to \infty} \mu_K(\bigcup_{k \in \mathcal{B}_m} C(\sigma, m, k)) \\ &= \lim_{m \to \infty} \#\mathcal{B}_m \mu_K(C(\sigma, 2^m, (0, ..., \|\sigma\| - 1))) \\ &= \lim_{m \to \infty} \binom{2^m}{\sigma} \mu_k(C(\sigma, 2^m, (0, ..., \|\sigma\| - 1))) \end{aligned}$$

Then, for all m and for every k a $\|\sigma\|$ tuple of distinct elements of $\{0, ..., 2^m - 1\}$, we have

$$\mu_K(C(\sigma, m, k)) = \mu_K(C(\sigma, m, (0, ..., \|\sigma\| - 1))) \sim_{m \to \infty} \frac{\Pi_{p=1}^K !\sigma(p) \mu_K(mc^{-1}(\sigma))}{2^{\|\sigma\| m}}$$

Remark that

$$C(\sigma, m, k) = \bigcup_{r \in \mathbb{N}} \bigcup_{l \in \mathcal{B}_{m+r} C(\sigma, m+r, l) \subset C(\sigma, m, k)} C(\sigma, m + n, l)$$

Then

$$\mu_K(C(\sigma, m, k) = \lim_{r \to \infty} \mu_K(\bigcup_{l \in \mathcal{B}_{m+r} C(\sigma, m+r, l) \subset C(\sigma, m, k)} C(\sigma, m + r, l))$$

As the union is disjoint we have

$$\begin{aligned} \mu_K(C(\sigma, m, k)) &= \lim_{r \to \infty} \sum_{l \in \mathcal{B}_{m+r} C(\sigma, m+r, l) \subset C(\sigma, m, k)} \mu_K(C(\sigma, m + r, l)) \\ &= \lim_{n \to \infty} \#\{l \in \mathcal{B}_{m+r} C(\sigma, m + r, l) \subset C(\sigma, m, k)\} \mu_K(C(\sigma, m + r, (0, ..., \|\sigma\| - 1))) \\ &= \lim_{r \to \infty} 2^{\|\sigma\| n} \mu_K(C(\sigma, m + r, (0, ..., \|\sigma\| - 1))) \\ &= \lim_{r \to \infty} \frac{2^{\|\sigma\| n} \Pi_{p=1}^K !\sigma(n) \mu_K(mc^{-1}(\sigma))}{2^{\|\sigma\|(m+n)}} \\ &= \frac{\Pi_{p=1}^K !\sigma(n) mc \circ \mu_K(P)}{2^{\|\sigma\| m}} \end{aligned}$$

As $mc \circ \mu_K = mc \circ \mu'_K$, we get that μ_K and μ'_K coincide on every $C(\sigma, m, k)$ for every $\sigma \in \mathcal{P}(K)$ and $k \in \{0, ..., 2^m - 1\}^{\|\sigma\|}$. But, this is a Π-system and a base of $\Sigma_{\mathcal{M}[K]([0,1])}$, so $\mu_K = \mu'_K$ and we are done.