An Atomic Topos for Probability Theory

Adam Dauser, Adrian Marti

Abstract. WIP

Contents

1. Introduction 1
2. A combinatorial construction of $\mathfrak{Prob}$ 1
3. $\mathfrak{Prob}$ as the classifying topos of a theory of intervals 8
4. Simple functions in $\mathfrak{Prob}$ 18

References 24

1. Introduction

WIP

2. A combinatorial construction of $\mathfrak{Prob}$

In this section, we define the topos $\mathfrak{Prob}$ as a category of sheaves on the category of finite probability spaces. We then proceed by proving properties about this construction, which we will need throughout the paper.

Definition 2.1.

(i) A finite probability space is a finite set X equipped with a map (called probability measure) $\mu : X \to \mathbb{R}_{>0}$ such that
$\sum_{x \in X} \mu(x) = 1$.

(ii) For a finite probability space (X, μ) and a subset $a \subset X$, we define $\mu(a) := \sum_{x \in a} \mu(x)$.

(iii) A map of finite probability spaces $(X, \mu) \to (Y, \nu)$ is defined as a map of finite sets $f : X \to Y$ that is measure-preserving, i.e. $\mu(f^{-1}(y)) = \nu(y)$. Denote the category of finite probability spaces by $\mathsf{FinProb}$.

Note that all maps of finite probability spaces are surjective, and we do not allow elements of measure zero.

We want to consider the sheaf topos $\mathrm{Sh}(\mathsf{FinProb}, J_{at})$, where J_{at} denotes the atomic topology. Recall that the atomic topology is defined by letting every nonempty sieve be a covering sieve. We need to check that this yields a Grothendieck topology.

PROPOSITION 2.2. *Given a diagram of finite probability spaces*

$$\begin{tikzcd} & (X_2, \mu_2) \arrow[d, "g"] \\ (X_1, \mu_1) \arrow[r, "f"'] & (Y, \mu) \end{tikzcd},$$

the pullback of the finite sets carries a probability measure $\nu(x, y) := \frac{\mu_1(x)\mu_2(y)}{\mu(fx)}$ *such that the diagram*

$$\begin{tikzcd} (X_1 \times_Y X_2, \nu) \arrow[r, "p_2"] \arrow[d, "p_1"'] & (X_2, \mu_2) \arrow[d, "g"] \\ (X_1, \mu_1) \arrow[r, "f"'] & (Y, \mu) \end{tikzcd}$$

commutes.

In particular, J_{at} *is a Grothendieck topology on* $\mathsf{FinProb}$.

PROOF. The only thing to show is that the projections are measure-preserving, since that also implies that ν is actually a measure. So let $x \in X_1$ and $u := f(x)$.

$$\begin{aligned} \sum_{(x,y) \in p_1^{-1} x} \nu(x, y) &= \sum_{y \in g^{-1} u} \frac{\mu_1(x)\mu_2(y)}{\mu(u)} \\ &= \frac{\mu_1(x)}{\mu(u)} \sum_{y \in g^{-1} u} \mu_2(y) \\ &= \frac{\mu_2(x)}{\mu(u)} \mu(u) \\ &= \mu_2(x) \end{aligned}$$

Similarly, p_2 is measure-preserving.

For the second part, notice that we just proved the *right Ore condition* for $\mathsf{FinProb}$. Checking the Grothendieck topology axioms is a known straightforward formal consequence of this. □

We now introduce the protagonist of this paper.

DEFINITION 2.3. Define the topos $\mathfrak{Prob}$ to be $\mathrm{Sh}(\mathsf{FinProb}, J_{at})$.

Much of this article is dedicated to finding other equivalent ways of describing $\mathfrak{Prob}$. When we wish to leave the particular description of $\mathfrak{Prob}$ vague, we will use the notation $\mathfrak{Prob}$.

Lemma 2 in section III.4 of [**1**] gives us a criterion for checking whether a presheaf is actually a sheaf over the atomic topology:

A presheaf $\mathcal{F}$ on $\mathsf{FinProb}$ is a sheaf over the atomic topology if and only if one can check the following for any morphism $f : \alpha \to \beta$ and any $a \in \mathcal{F}(\alpha)$: If for all diagrams

$$\begin{tikzcd} \gamma \arrow[r, shift left, "g"] \arrow[r, shift right, "h"'] & \alpha \arrow[r, "f"] & \beta \end{tikzcd}$$

with $fg = fh$, we have that $\mathcal{F}(g)(a) = \mathcal{F}(h)(a)$, then there is a unique lift $b \in \mathcal{F}(\beta)$ with the property that $\mathcal{F}(f)(b) = a$.

In some cases, one can prove the following condition that is much stronger than being a sheaf. It is especially useful when the sheaf doesn't really use the measure in its definition.

PROPOSITION 2.4. *Let $\mathcal{F}$ be a presheaf on* $\mathsf{FinProb}$. *If for all morphisms $f : (X, \mu) \to (Y, \nu)$ the fork*

$$\mathcal{F}(Y, \nu) \xrightarrow{\mathcal{F}f} \mathcal{F}(X, \mu) \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} \mathcal{F}(X \times_f X, \bar{\mu})$$

is an equalizer, where $\bar{\mu}$ is as in 2.2, then $\mathcal{F}$ is a sheaf.

PROOF. This is the condition from [**1**] described above, but weakened by choosing $\gamma = (X \times_f X, \bar{\mu})$. □

We can now check the sheaf condition for the vast majority of sheaves that we are interested in:

PROPOSITION 2.5.

(i) *All morphisms in* $\mathsf{FinProb}$ *are regular epimorphisms.*
(ii) *The representable functors are sheaves, i.e. J_{at} is a subcanonical topology on* $\mathsf{FinProb}$.
(iii) *All coproducts of representable functors are sheaves.*

PROOF.

(i) We need to show that every morphism $f : (X, \mu) \to (Y, \nu)$ comes from a coequalizer. Our construction of the measure on the pullback in 2.2 gives us the following fork:

$$(X \times_Y X, \bar{\mu}) \rightrightarrows (X, \mu) \xrightarrow{f} (Y, \nu)$$

Since this is clearly a coequalizer on the underlying sets, the only thing to check is that the uniquely induced maps are measure-preserving. This is straightforward to verify.

(ii) Checking that the representable functors are sheaves by using 2.4 reduces to the fact that the fork above is a coequalizer diagram.

(iii) We check the sheaf condition 2.4 for a coproduct of representables. But since we already know this sheaf condition holds for representables, we merely need to note that equalizers of coproducts are coproducts of equalizers in Set.

□

Note that the second and third part are merely formal consequences of the first one. Any category (satisfying the *right Ore condition*) such that the first part holds, also satisfies the second and the third part.

Recall that a site $(\mathcal{C}, J)$, is called *atomic* if every nonempty sieve over an object c is covering. Furthermore, a Grothendieck topos is called *atomic* if it is equivalent to a topos of sheaves over an atomic site. Tautologically, $\mathfrak{Prob}$ can be seen to be an atomic topos. An object in an atomic topos is called an *atom* if it has precisely two subobjects (the initial object and the object itself).

We just saw that the representables are actually sheaves. In fact, since our site is atomic, these representables are atoms. Moreover, for atomic toposes, every object can be uniquely written as a coproduct of atoms. So, in some sense, we can reduce all problems to atoms. The natural question arises: Are there more atoms than the representables? In **??** we will give a complete list of the atoms and see that the representables are not all of them.

2.1. Finite Limits. Our goal in the next two sections will be gaining an understanding of $\mathfrak{Prob}$ from a more logical perspective. An absolute prerequisite for these discussions is a good understanding of finite limits in our topos. We will need products for controlling the number of free variables in our formulas and equalizers for expressing that two terms are equal.

The idea is to first understand finite limits in $\mathsf{FinProb}$, since the sheaves we are interested in are coproducts of representables. But the situation is not as simple as one may hope: finite limits in $\mathsf{FinProb}$ only very rarely exist. Naturally, the corresponding representable presheaves in $\mathsf{Set}^{\mathsf{FinProb}^{op}}$ actually do have these finite limits, and we will see that they have a straightforward form. This will give rise to a weakened form of the notion of a limit, which we will call multi-limit. It will be essential in all our discussions about logic.

The most surprising part about multi-limits will be their nontrivial combinatorial content and their connection to polytopes. We will see that many statements in this paper are proven by using computations with multi-limits and polytopes.

We introduce convenient notation to emphasize the combinatorial nature of finite probability spaces.

Definition 2.6 (U-notation).

(i) For $r_1, \cdots r_n \in [0,1]$ such that $r_1 + \cdots + r_n = 1$, denote by $U_{r_1 \cdots r_n}$ or simply by U_{r_i} the finite probability space on the set
$$\{i \mid r_i \neq 0\}$$
equipped with the measure $\mu(i) = r_i$.

(ii) We will use the shorthand $U_r := U_{r,r-1}$ for $r \in [0,1]$.

This is notation for finite probability spaces in the sense that the U_{r_i} form a skeleton of $\mathsf{FinProb}$.

Remark 2.7. Notice that for a finite probability space (X, μ), $\mathrm{Hom}((X,\mu), U_{r_i})$ precisely consists of partitions (a_i) of the set X with $\mu(a_i) = r_i$. This may seem like quite a harmless observation, but being aware of this fact makes some calculations with yU_{r_i} much more intuitive. Moreover, we will later see that the U_{r_i} in fact precisely correspond to logical formulas specifying a partition with fixed measures. We will often identify maps with partitions implicitly and without comment.

In particular, $\mathrm{Hom}((X,\mu), U_r)$ is identified with certain partitions consisting of two elements, or equivalently, subsets of X of measure r. We will identify $\mathrm{Hom}((X,\mu), U_r)$ with such subsets.

We now compute the finite limits of representables in $\mathsf{Set}^{\mathsf{FinProb}^{op}}$ (or equivalently in $\mathfrak{Prob}$).

Proposition 2.8. *Let $U_{r_1,\cdots r_n}$ and $U_{s_1,\cdots s_m}$ be finite probability spaces. For each finite probability space $U_{t_{ij}}$[1] ($0 < i \leq n$, $0 < j \leq m$) with*
$$\sum_j t_{ij} = r_i$$
and
$$\sum_i t_{ij} = s_j,$$
we define projections $p_1 : U_{t_{ij}} \to U_{r_i}$, $(x_{ij}) \mapsto (\bigvee_j x_{ij})$ and $p_2 : U_{t_{ij}} \to U_{s_i}$, $(x_{ij}) \mapsto (\bigvee_i x_{ij})$.

These projection maps induce the isomorphism
$$yU_{r_i} \times yU_{s_j} \cong \coprod_{t_{ij}} yU_{t_{ij}},$$
where the indexing t_{ij} range over the ones with the fixed sums specified above.

Proof. The t_{ij} where precisely chosen such that the projections are measure preserving. To show
$$yU_{r_i} \times yU_{s_j} \cong \coprod_{t_{ij}} yU_{t_{ij}},$$
send a pair of partitions $((a_i),(b_j))$ to the family $(a_i \wedge b_j)_{ij}$. The inverse of this map is given by joining the family with the projection maps described above. □

Notice that the isomorphism we proved is not quite a universal property in the traditional sense. Normally, we would have required $yU_{r_i} \times yU_{s_j}$ to actually be representable in order to be able to say that we have binary products in $\mathsf{FinProb}$. But this does not work in our case, as it only is a coproduct of representables.

We give a name to this weaker notion of a limit.

[1]Technically, t_{ij} has two indices, so one would have to reindex it with one index so that our notation is well-defined. Subsequently, we will ignore this issue.

DEFINITION 2.9. Given a small category $\mathcal{C}$ and a diagram $D : J \to C$, we will call a family of cones[2] $(d_i \Rightarrow D)_i$ over D a *multi-limit* of D if the cones induce an isomorphism

$$\lim_{j \in J} yD(j) \cong \coprod_i yd_i$$

in $\mathsf{Set}^{\mathcal{C}^{op}}$.

Moreover, we will use the terminology *multi-product*, *multi-equalizer* and *multi-pullback* to refer to multi-limits of particular shapes.

Alternatively, we could have presented the universal property in a more traditional style by saying that cones over our diagram factor through one of the $U_{t_{ij}}$ uniquely. In fact, this is what is done in [**2**] (Lemma 6.11.). Caramello and Lafforgue use multi-products in order to study the atoms of an atomic topos, which is also one of the things we will use them for.

Finally, we also have multi-equalizers.

PROPOSITION 2.10. *Given the diagram* $\alpha \underset{h}{\overset{g}{\rightrightarrows}} \beta$ *in* $\mathsf{FinProb}$ *we have*

$$\mathrm{eq}(\ y\alpha \underset{h}{\overset{g}{\rightrightarrows}} y\beta\ ) \cong y\alpha \textit{ if } f = g \textit{ and } \emptyset \textit{ otherwise.}$$

PROOF. The interesting case is when $f \neq g$. In that case we compute the equalizer pointwise on a finite probability space γ:

$$\mathrm{eq}(\ y\alpha \underset{h}{\overset{g}{\rightrightarrows}} y\beta\ )(\gamma) \cong \{h : \gamma \twoheadrightarrow \alpha \mid fh = gh\} \cong \emptyset.$$

□

2.2. A calculus of polytopes. We have seen that treating multi-limits requires us to talk about coproducts over certain index sets. As we progress further into the paper, we will encounter many more of these constructions. In order to describe these concisely, we introduce simplified notation for them.

DEFINITION 2.11.

(i) Let $\mathcal{A}$ denote the set

$$\{(r_1, \cdots, r_n) \mid n \in \mathbb{N}, r_i \in \mathbb{R}_{\geq 0}, \sum r_i = 1\}.$$

(ii) Define a *set of atoms* to be a pair (S, T), where S is a subset of $\mathcal{A}$ and $T(x)$ is a set for all $x \in S$.

(iii) The sheaf associated to a set of atoms (S, T) will be defined by

$$\langle\, T(x) \mid x \in S \,\rangle := \coprod_{\substack{x \in S \\ y \in T(x)}} yU_x,$$

using U-notation.

(iv) An *element* of a set of atoms (S, T) will be a tuple $(x, t) \in \coprod_{x \in S} T(x)$. If $T(x)$ is always either empty or the one-element set, we will abbreviate writing (x, t) for an element of (S, T) by merely stating that x is an element of (S, T).

One subtlety about our definitions is that the set $\mathcal{A}$ will contain multiple distinct elements corresponding to the same atom, simply because we are considering *ordered* families r_i and the space U_{r_i} does not depend on the order.

Notice that the elements of a set of atoms (S, T) index the coproduct in the definition of $\langle\, T(x) \mid x \in S \,\rangle$. Moreover, they can also be seen as the monomorphisms

$$yU_{r_i} \hookrightarrow \langle\, T(x) \mid x \in S \,\rangle$$

[2]We think of cones over D as natural transformations $c \Rightarrow D$ from a constant diagram c to the diagram D.

into the sheaf associated to (S, T).

Occasionally, we may add additional properties $\phi_1, \phi_2, \cdots, \phi_n$ to the notation, as follows:

$$\langle\, T(x) \mid x \in S, \phi_1(x), \cdots, \phi_n(x) \,\rangle := \langle\, T(x) \mid x \in \{x' \in S \mid \phi_1(x'), \cdots, \phi_n(x')\} \,\rangle$$

While this is somewhat informal(*What is a property?*), the reader most likely knows what is meant.

We have seen that when thinking about binary multi-products, subsets (t_{ij}) of $\mathcal{A}$ with fixed sums of rows and columns are essential. Since multi-products will be invaluable in our study of $\mathfrak{Prob}$, we introduce some notation for these index sets.

Definition 2.12. Let $n_1, \cdots, n_k$ be positive integers.

(i) Define:

$$P^k_{n_1,\cdots,n_k} := \{(t_{i_1\cdots i_k}) \in \mathbb{R}^{n_1\times\cdots\times n_k} \mid t_{i_1\cdots i_k} \geq 0, \sum_{i_1,\cdots,i_k} t_{i_1\cdots i_k} = 1\}$$

(ii) For each $j = 1, \cdots k$, let $r^j_1, \cdots r^j_{n_j}$ be nonnegative real numbers summing up to 1. Now let

$$\begin{aligned} P^k_{r^1_i,\cdots r^k_i} := \{(t_{i_1\cdots i_k}) \in \mathbb{R}^{n_1\times\cdots\times n_k} \mid t_{i_1\cdots i_k} \geq 0, \\ \sum_{i_2,\cdots i_k} t_{ji_2\cdots i_k} = r^1_j, \\ \cdots \\ \sum_{i_1,\cdots i_{k-1}} t_{i_1 i_2\cdots j} = r^k_j \\ \text{for any index } j\}. \end{aligned}$$

Example 2.13.

- Binary multi-products can now be written as

$$yU_{r_i} \times yU_{s_j} \cong \langle\, 1 \mid t_{ij} \in P^2_{r_i,s_j} \,\rangle,$$

  where 1 denotes the one-object set.
- Since FinProb has a terminal object, we can use our binary multi-product formula in order to compute arbitrary finite multi-products:

$$\prod_{j=1}^{n} yU_{r^j_i} \cong \langle\, 1 \mid t_{i_1\cdots i_n} \in P^n_{r^1_i\cdots r^n_i} \,\rangle$$

  In more detail, given a family $t_{i_1\cdots i_n} \in P^n_{r^1_i,\cdots r^n_i}$, the projections

$$\iota^k : U_{t_{i_1,\cdots i_n}} \to U_{r_k}$$

  given by $a_{i_1,\cdots i_n} \mapsto \bigvee_{i_1,\cdots i_{j-1},i_{j+1},\cdots i_n} a_{i_1,\cdots,i_{j-1},j,i_{j+1},\cdots i_n}$ assemble into an n-ary multi-product.
- Computing the multi-pullback of a diagram

$$\begin{tikzcd} & U_{r'_i} \arrow[d, "g"] \\ U_{r_i} \arrow[r, "f"'] & U_{s_i}, \end{tikzcd}$$

  (by noting that equalizers of coproducts are coproducts of equalizers in presheaf toposes, which can be checked pointwise), yields the formula

$$\begin{aligned} & yU_{r_i} \times_{yU_{s_i}} yU_{r'_i} \cong \\ & \quad \langle\, 1 \mid t_{ij} \in P^2_{r_i,r'_i}, t_{ij} = 0 \text{ for all } i, j \text{ with } f(i) \neq g(j) \,\rangle. \end{aligned}$$

- Define the presheaf U in the following way:

$$U \cong \langle\, 1 \mid (r_1, r_2) \in P^1_2 \,\rangle$$

We will later see that this presheaf is in fact the underlying sheaf of a universal model in $\mathfrak{Prob}$ for a suitable theory of boolean algebras with measures.

REMARK 2.14. We have seen that the elements t_{ij} of a binary multi-product[3] are the $t_{ij} \in P^2_{r_i,s_i}$. This set can be seen a subset of the set of $n \times m$-matrices, which is, in fact, a polytope in $\mathbb{R}^{n\times m}$. The combinatorics literature has called these polytopes *transportation polytopes*. Chapter 8 of [**3**] is dedicated to their study.

Notice that all the sheaves we have worked with until now can be written as sheaves associated to sets of atoms. One might hope that every sheaf can be written in this form, but as remarked earlier, this is not possible, since there are non-representable atoms. Nevertheless, our notation can be easily extended to incorporate the non-representable atoms we discuss in **??**, but we won't need this in this article.

The most convenient thing about sets of atoms, is that it is easy to send them through (the inverse image functor of) a geometric morphism, since geometric morphisms preserve coproducts. In fact, we found that the only practical way to compute geometric morphisms on a sheaf is by decomposing the sheaf into atoms. This will be done numerous times throughout the article. In fact, we will occasionally also need to send morphisms in $\mathfrak{Prob}$ through inverse image functors, so we also describe a mechanism through which to decompose morphisms:

PROPOSITION 2.15. *Let (S,T) and (S',T') be sets of atoms and let $\mathcal{F}$ and $\mathcal{G}$ be their associated sheaves, respectively. Then the maps $\mathcal{F} \to \mathcal{G}$ are precisely pairs $(g,(f_{x,t}))$, where g is a map between the sets of elements*

$$\coprod_{x\in S} T(x) \to \coprod_{x\in S'} T'(x)$$

and for each element $(x,t) \in \coprod_{x\in S} T(x)$ with $(x',t') = g(x,t)$, $f_{x,t}$ is a morphism

$$U_x \xrightarrow{f_{x,t}} U_{x'}.$$

PROOF. Using the universal property of the coproduct, we first see that a map $f : \mathcal{F} \to \mathcal{G}$ is determined by maps $\tilde{f}_{x,t} : yU_x \to \mathcal{G}$ for each element (x,t) of (S,T). By the Yoneda lemma, there exists an element $g(x,t) := (x',t')$ of (S',T') such that $\tilde{f}_{x,t}$ is determined by an $f_{x,t} \in \mathrm{Hom}(U_x, U_{x'})$. Now the g and the family $(f_{x,t})$ give us the required decomposition of f. □

2.3. A better sheaf condition. As a toy application of the facts discovered so far, we develop a better sheaf condition. Since we won't use it anywhere, the reader can safely skip this subsection.

PROPOSITION 2.16. *A presheaf $\mathcal{F}$ on* FinProb *is a sheaf over the atomic topology if and only if for all maps of finite probability spaces $f : U_{r_i} \to U_{s_i}$ the fork*

$$\mathcal{F}U_{s_i} \xrightarrow{\mathcal{F}f} \mathcal{F}U_{r_i} \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} \textstyle\prod_{t_{ij}} \mathcal{F}U_{t_{ij}}$$

is an equalizer diagram, where the t_{ij} range over the elements of the multi-pullback $U_{r_i} \times_{U_{s_i}} U_{r_i}$ (we are pulling f back along itself) and the projection π_1 and π_2 are induced by the universal 'projection maps' of multi-pullbacks.

[3]We have not specified what exactly the data (S,T) is, for which we want to consider elements. In this case, and in the future, we trust that the reader is able to infer this data on his own.

PROOF. Since for each $U_{t_{ij}}$ we have the commuting square

$$\begin{tikzcd} U_{t_{ij}} \arrow[r] \arrow[d] & U_{r_i} \arrow[d, "f"] \\ U_{r_i} \arrow[r, "f"'] & U_{s_i}, \end{tikzcd}$$

f certainly gives us a diagram

$$\mathcal{F}U_{s_i} \xrightarrow{\mathcal{F}f} \mathcal{F}U_{r_i} \underset{\pi_2}{\overset{\pi_1}{\rightrightarrows}} \prod_{t_{ij}} \mathcal{F}U_{t_{ij}}$$

with $\pi_1\mathcal{F}(f) = \pi_2\mathcal{F}(f)$. Claiming that this is an equalizer means saying that for every $a \in \mathcal{F}U_{r_i}$ with the property that for all the projections $p_1 : U_{t_{ij}} \to U_{r_i}$, $p_2 : U_{t_{ij}} \to U_{r'_i}$ the equation $\mathcal{F}(p_1)(a) = \mathcal{F}(p_2)(a)$ is satisfied, implies that there is a unique lift $b \in \mathcal{F}U_{s_i}$ of a.

So our claim is that the criterion described above needs only be checked on the $U_{t_{ij}}$. Thus, we assume that the equation above holds for the $U_{t_{ij}}$ with their projection maps. We need to show that the equation $\mathcal{F}(g)(a) = \mathcal{F}(h)(a)$ already holds for all the

$$\gamma \underset{h}{\overset{g}{\rightrightarrows}} U_{r_i}$$

with $fg = fh$. But this certainly holds, since we can factor g and h through one of the $U_{t_{ij}}$ by the universal property of the multi-pullback. □

Note that this proposition actually had little to do with finite probability spaces and is really about a sheaf condition for atomic topologies where the base category has multi-pullbacks.

3. $\mathfrak{Prob}$ as the classifying topos of a theory of intervals

In this section we will show that $\mathsf{Sh}(\mathsf{FinProb}, J_{at})$ classifies a certain theory of intervals. For the most part, we will use the notations and definitions as in part D.1 of [**4**].

DEFINITION 3.1. Define the geometric theory $\mathbb{T}_{\text{bpalg}}$ of boolean probability algebras to have one sort B, constants and function symbols

$$\begin{gathered} 1 : B \\ 0 : B \\ \wedge : B \times B \to B \\ \vee : B \times B \to B \\ \neg : B \to B \end{gathered}$$

and a unary relation symbol B_r on B for each $r \in [0, 1]$. We impose the following axioms(a, b and c denote free variables in B):

- *Boolean algebra.*

$\top \vdash a \wedge (b \wedge c) = (a \wedge b) \wedge c$	$\top \vdash a \vee (b \vee c) = (a \vee b) \vee c$	*associativity*
$\top \vdash a \wedge 1 = a$	$\top \vdash a \vee 0 = a$	*identity*
$\top \vdash a \wedge \neg a = 0$	$\top \vdash a \vee \neg a = 1$	*inverse*
$\top \vdash a \wedge (a \vee b) = a$	$\top \vdash a \vee (a \wedge b) = a$	*absorption*
$\top \vdash a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$	$\top \vdash a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$	*distributivity*

- *The B_r form a partition.* For all $r, s \in [0, 1]$ with $r \neq s$ we have an axiom

$$B_r(a) \wedge B_s(a) \vdash \bot$$

and we also have an axiom

$$\top \vdash \bigvee_{r\in[0,1]} B_r(a).$$

- *Probability measure.* For all $r, s \in [0,1]$, we further require

$$(a \wedge b = 0) \wedge B_r(a) \wedge B_s(b) \vdash B_{r+s}(a \vee b)$$

and finally we need

$$\top \vdash B_1(1)$$
$$B_0(a) \vdash a = 0.$$

Denote the category models of $\mathbb{T}_{\text{bpalg}}$ in Set by BPAlg.

Essentially, we have encoded a boolean probability algebra geometrically by partitioning the base sort B into the elements B_r of measure $r \in [0,1]$. Notice that we require that there is only one element of measure zero. This requirement may seem somewhat surprising, since the boolean algebra of measurable sets of a σ-algebra certainly can have sets of measure zero, or *nullsets*. Instead, we have a different kind of model in mind.

Example 3.2. Let (X, Σ, μ) be a probability space. Consider the equivalence relation $\sim$ on Σ with

$$A \sim B \text{ iff } \mu(A \setminus B \cup B \setminus A) = 0.$$

We now consider the boolean algebra $B := \Sigma / \sim$ (*'measurable sets up to nullset'*) together with the unary relations $B_r \subset B$ given by the equivalence classes in B that have measure r. This forms a boolean probability algebra.

Notice that the boolean probability algebras associated to the Borel and the Lebesgue intervals equipped with the Lebesgue measure are the same.

This is a well-established construction, and it yields what is known as a *measure algebra*. In this case we even get a *probability algebra*.

Definition 3.3. Call a boolean probability algebra (A, μ) a probability algebra if A is a complete boolean algebra and moreover, for every directed family a_i of elements of A, we have the additivity condition

$$\mu(\bigvee a_i) = \sup \mu(a_i).$$

The pairs (A, μ) for which there is a $c > 0$ such that $(A, c\mu)$ is a probability algebra will simply be called measure algebras. A morphism of measure algebras is a morphism of boolean algebras that is measure-preserving.

Note that we have chosen a slightly more restrictive definition of measure algebra than is standard, since we did not allow the measure to adopt infinite values.

Given a measure space (X, Σ, μ), dividing out the nullsets out of Σ (as above) is guaranteed to yield a measure algebra. This is surprising, since the σ-algebra doesn't need to have infinite joins or meets, but the measure algebra does have them. Conversely, every measure algebra (A, μ) arises this way, which is proved using the Stone space associated to A. For more details, see §321 of [**5**]. For a more category-theoretic treatment of this *'point-free'* approach to measure spaces, and it's relationship to other approaches, see [**6**] or [**7**].

The bottom line is that it is a harmless and well-established idea to replace probability spaces by probability algebras. In fact, [**5**] develops integration and the theory of L^p spaces in the setting of measure algebras. Rather, the limitation of the setting of boolean probability algebras is that we do not have actual measure algebras, and thus we cannot make statements about infinite unions and infinite sums of measures. This gives us a first glimpse how naively working with the topos $\mathfrak{Prob}$ limits us to the study of *finitary phenomena*.

3.1. Classifying the theory of boolean probability algebras. We would like to show that the presheaf topos
$\mathsf{Set}^{\mathsf{FinProb}^{op}}$ classifies the theory $\mathbb{T}_{\text{bpalg}}$. More specifically, we will prove the equivalence

$$\mathsf{Set}[\mathbb{T}_{\text{bpalg}}] \simeq \mathsf{Set}^{\mathsf{FinProb}^{op}}, \tag{3.1}$$

where $\mathsf{Set}[\mathbb{T}_{\text{bpalg}}]$ denotes the construction of the classifying topos of $\mathbb{T}_{\text{bpalg}}$ as the sheaves on a certain geometric category of formulas in context.

The reader uninterested in the proof can skip to **??**.

Remark 3.4. We have an equivalence of categories

$$\mathsf{FinProb}^{op} \simeq \mathsf{BPAlg}_f$$

between the opposite category of finite probability spaces and the category of finite boolean probability algebras. We omit the proof and the data of the equivalence, since we won't use this fact outside this remark.

If 3.1 were to be true, we would have that the points (or models) of both toposes are equivalent. This fact can be directly verified, as a consistency check, using *Diaconescu's theorem*, the fact that flat functors to Set can be given as Ind-completions and a substantial amount of property-checking for seeing that BPAlg is the Ind-completion of BPAlg_f:

$$\begin{aligned}
\text{Geom}(\mathsf{Set}, \mathsf{Set}[\mathbb{T}_{\text{bpalg}}]) &\simeq \mathbb{T}_{\text{bpalg}}\text{-mod}(\mathsf{Set}) \\
&\simeq \mathsf{BPAlg} \\
&\simeq \text{Ind}(\mathsf{BPAlg}_f) \\
&\simeq \text{Ind}(\mathsf{FinProb}^{op}) \\
&\simeq \text{Flat}(\mathsf{FinProb}, \mathsf{Set}) \\
&\simeq \text{Geom}(\mathsf{Set}, \mathsf{Set}^{\mathsf{FinProb}^{op}}),
\end{aligned}$$

where we use Geom to denote the category of geometric morphisms between two toposes, $\mathbb{T}_{\text{bpalg}}$-mod denotes the models of the theory in a topos, Flat denotes the flat functors from a small category into a left-exact category and Ind denotes the Ind-completion of a small category.

While it is useful to know this, we want to prove the stronger statement 3.1.

We will prove the equivalence 3.1 by purely syntactical means. The functor in one direction will be given by an appropriate model of $\mathbb{T}_{\text{bpalg}}$ in $\mathsf{Set}^{\mathsf{FinProb}^{op}}$. The functor in the other direction will require us to work with the syntactic category $\mathcal{C}^{\mathbb{T}_{\text{bpalg}}}$ directly, and we will see that this will lead to us needing to work formally with geometric logic.

Before we begin, we explain the intuition behind this Morita-equivalence. The idea is that, while the left topos axiomatizes a boolean probability algebra, the topos on the right directly axiomatizes the *partitions* of the boolean algebra into elements of prescribed measures. The object yU_{r_i} is supposed to represent disjoint elements $a_1, \cdots, a_n$ with measures $\mu(a_1) = r_1, \cdots, \mu(a_n) = r_n$. The maps $yU_{s_i} \to yU_{r_i}$ are supposed to represent unions of the partition (s_i) into a coarser one.

With this intuition in mind, how should we go about defining the universal model defining the geometric morphism $\mathsf{Set}^{\mathsf{FinProb}^{op}} \to \mathsf{Set}[\mathbb{T}_{\text{bpalg}}]$? Essentially, we are asking how to construct an object of *all* elements of a boolean algebra from the objects of elements with specified measures. Under this perspective, it becomes clear that we should define the underlying presheaf of the universal model U to be

$$\coprod_{r \in [0,1]} yU_r \in \mathsf{Set}^{\mathsf{FinProb}^{op}}.$$

One immediately sees that U is just the powerset functor on the underlying set of a given probability space (a map f between probability spaces gets sent to the map that takes the preimage of each subset).

LEMMA 3.5. *U is a model of the $\mathbb{T}_{\text{bpalg}}$. The sort B is given by U. The relation symbols B_r are given by the inclusions of the yU_r and the boolean algebra operations are inherited from the boolean algebra structure of $U(X,\mu)$, the powerset of X.*

This yields a left-exact, colimit-preserving functor

$$\mathsf{Set}^{\mathsf{FinProb}^{op}} \xleftarrow{\text{Lan}_{\text{y}}(U)} \mathsf{Set}[\mathbb{T}_{\text{bpalg}}],$$

where U is the functor we get by identifying models of $\mathbb{T}_{\text{bpalg}}$ with functors whose domain is the syntactic category of $\mathbb{T}_{\text{bpalg}}$.

PROOF. The last statement follows from the first part by using the universal property of classifying toposes.

Since we are looking at presheaves, all the colimit and limit conditions from the axioms can be checked pointwise, where they are immediate. More formally, we can use [**4**], Corollary D.1.2.14, to get the fact that U is a $\mathbb{T}_{\text{bpalg}}$-model in $\mathsf{Set}^{\mathsf{FinProb}^{op}}$. □

We proceed by formalizing our intuition about the representables in $\mathsf{Set}^{\mathsf{FinProb}^{op}}$ being partitions by constructing our candidate inverse functor. We diverge from the notation in D.1 of [**4**] in that we will sometimes annotate formulas with a set of variables containing all the variables used in the formula, as seen below. This is done so that when these formulas get combined with logical connectives, the big formula remains easy to read.

LEMMA 3.6. *Let $f:(X,\mu)\to(Y,\nu)$ be an arrow in $\mathsf{FinProb}$. We define the functor $F:\mathsf{FinProb}\to\mathcal{C}^{\mathbb{T}_{\text{bpalg}}}$ on the object (X,μ) and on the function f as follows: Let*

$$\{\vec{a}.\phi^{\vec{a}}_{X,\mu}\} := \left\{\vec{a}.\bigwedge_{i\in X} B_{\mu(i)}a_i \wedge \bigwedge_{\substack{i,j\in X\\ i\neq j}}(a_i\wedge a_j=0)\right\}$$

$$[\theta^{\vec{a},\vec{b}}_f] := \left[\bigwedge_{j\in Y}(b_j=\bigvee_{fi=j}a_i)\wedge\phi^{\vec{a}}_{X,\mu}\right],$$

and now define $F(X,\mu):=\{\vec{a}.\phi^{\vec{a}}_{X,\mu}\}$ and $F(f):=[\theta^{\vec{a},\vec{b}}_f]$.

This yields a left-exact, colimit-preserving functor

$$\text{Lan}_F:\mathsf{Set}^{\mathsf{FinProb}^{op}}\to\mathsf{Set}[\mathbb{T}_{\text{bpalg}}].$$

PROOF. We need to show functoriality and then verify that Lan_F is left-exact. In order to check this left-exactness, we use theorem VII.9.1 in [**1**](page 399). The cited theorem says that flat functors(defined as functors whose left Kan extension is left-exact) precisely correspond to the functors satisfying the straightforward list of properties listed in the definition of a filtering functor, which is defined in VII.8.1 (page 394).

The reader who actually reads the definition of a filtering functor will quickly notice that that definition can be checked in our case by showing that F preserves the terminal object, binary multi-products and multi-equalizers in the sense checked below.

(i) We show that F preserves identities. We compute that

$$\begin{aligned} F(\text{id}) &= [\theta^{\vec{a},\vec{b}}_{\text{id}}] \\ &= [\vec{a}=\vec{b}] \\ &= \text{id}, \end{aligned}$$

which shows our claim.

(ii) We show that F preserves composition. Given morphisms $f : \alpha \to \beta$, $g : \beta \to \gamma$ in $\mathsf{FinProb}$,

$$\begin{aligned} F(g)F(f) &= [\exists c.\theta_f^{\vec{a},\vec{c}} \wedge \theta_g^{\vec{c},\vec{b}}] \\ &= [\theta_{gf}^{\vec{a},\vec{b}}], \end{aligned}$$

where the last equality follows from associativity and commutativity of the $\wedge$ operation and by additivity of the measure.

(iii) We show that F preserves the terminal object. We simply compute this:

$$\begin{aligned} F(*) &= \{a.B_1(a)\} \\ &\cong \{[].\top\}, \end{aligned}$$

where the last isomorphism follows from the fact we can geometrically prove that there is a unique element of measure 1.

(iv) We show that F preserves multi-products. Concretely, let α, β be finite probability spaces. Then we have a polytope $P^2_{r_i,s_j}$ and for each of its elements t_{ij} we have a cone

$$\begin{tikzcd} & U_{t_{ij}} \arrow[dl, "p_{t_{ij}}"'] \arrow[dr, "q_{t_{ij}}"] & \\ \alpha & & \beta \end{tikzcd} .$$

In order to show that this multi-product gets preserved, we need to show that the map

$$\coprod_{t_{ij} \in P^2_{r_i,s_j}} \langle \theta_{p_{t_{ij}}}^{\vec{a},\vec{b}}, \theta_{q_{t_{ij}}}^{\vec{a},\vec{b}} \rangle : \coprod_{t_{ij} \in P^2_{r_i,s_j}} FU_{t_{ij}} \to F(\alpha) \times F(\beta)$$

is an isomorphism. It is crucial that there does not only exist an isomorphism, but that it is this concrete isomorphism constructed by universal property. Indeed, for showing corresponding condition VII.8.1 in [**1**] we need to show that a specific family of morphisms is jointly epimorphic. And our claim here certainly implies that.

We will show that this map is an isomorphism by computing this map as a formula. First, note that the inclusions

$$FU_{t_{ij}} \to F(\alpha) \times F(\beta)$$

can be computed (using D.1.4.2 in [**4**]) to be

$$[b_i = \bigvee_j a_{ij}, c_j = \bigvee_i a_{ij}] : \{\vec{a}.\phi_{U_{t_{ij}}}^{\vec{a}}\} \to \{\vec{b}, \vec{c}.\phi_\alpha^{\vec{b}} \wedge \phi_\beta^{\vec{c}}\}.$$

By using the partition axiom multiple times, we notice that the formulas $\{\vec{a}.\phi_{U_{t_{ij}}}^{\vec{a}}\}$ are disjoint (in the subobject lattice of $\{\vec{a}.\top\}$). Thus, the coproduct corresponds to union in that subobject lattice. Therefore, by D.1.4.10 in [**4**] we can compute the required coproduct to be

$$[b_i = \bigvee_j a_{ij}, c_j = \bigvee_i a_{ij}] : \{\vec{a}. \bigvee_{t_{ij} \in P^2_{r_i,s_j}} \phi_{U_{t_{ij}}}^{\vec{a}}\} \to \{\vec{b}, \vec{c}.\phi_\alpha^{\vec{b}} \wedge \phi_\beta^{\vec{c}}\}.$$

Interestingly, we have shown that this map is simply a substitution. But we can give another substitution that can be checked to be an inverse of this map by using the boolean probability algebra axioms:

$$[a_{ij} = b_i \wedge c_j]$$

Thus, we have shown that the required map is an isomorphism.

(v) Let $\alpha \underset{g}{\overset{f}{\rightrightarrows}} \beta$ be a diagram in $\mathsf{FinProb}$. We show that F preserves multi-equalizers. More specifically, if we have a family of forks

$$\gamma_i \xrightarrow{h_i} \alpha \underset{g}{\overset{f}{\rightrightarrows}} \beta$$

forming a multi-equalizer, then we need to show that

$$\coprod_i F(h_i) : \coprod_i F(\gamma_i) \to \mathrm{eq}(\ F(\alpha) \underset{Fg}{\overset{Ff}{\rightrightarrows}} F(\beta)\ )$$

is an isomorphism.

We already know that multi-equalizers in $\mathsf{FinProb}$ are particularly simple. If $f = g$, then the family of forks consists of only the identity map $\mathrm{id} : \alpha \to \alpha$. In this case our claim becomes trivial.

Now assume that $f \neq g$. Then the multi-equalizer is the empty family. Thus, our claim becomes that

$$\mathrm{eq}(\ \{\vec{a}.\phi_\alpha^{\vec{a}}\} \underset{[\theta_g^{\vec{a},\vec{b}}]}{\overset{[\theta_f^{\vec{a},\vec{b}}]}{\rightrightarrows}} \{\vec{b}.\phi_\beta^{\vec{b}}\}\ )$$

is the initial object $\{\vec{a}.\bot\}$. In fact, using D.1.4.2 in [**4**], we can compute this equalizer to be $\{\vec{a}.(\exists\vec{b})\theta_f^{\vec{a},\vec{b}} \wedge \theta_g^{\vec{a},\vec{b}}\}$. This means that our objective is to show the sequent

$$(\exists\vec{b})\theta_f^{\vec{a},\vec{b}} \wedge \theta_g^{\vec{a},\vec{b}} \vdash_{\vec{a}} \bot.$$

Since that $f \neq g$, we have a k be such that $f(k) \neq g(k)$. Now we can simply show that this sequent holds.

$$\begin{aligned}
(\exists\vec{b})\theta_f^{\vec{a},\vec{b}} \wedge \theta_g^{\vec{a},\vec{b}} &\vdash_{\vec{a}} (\exists\vec{b}) \bigwedge_{j\in\beta}(b_j = \bigvee_{fi=j} a_i) \wedge \bigwedge_{j\in\beta}(b_j = \bigvee_{gi=j} a_i) \wedge \phi_\alpha^{\vec{a}} \\
&\vdash_{\vec{a}} \bigwedge_{j\in\beta}(\bigvee_{fi=j} a_i = \bigvee_{gi=j} a_i) \wedge \phi_\alpha^{\vec{a}} \\
&\vdash_{\vec{a}} (\bigvee_{fi=fk} a_i = \bigvee_{gi=fk} a_i) \wedge \phi_\alpha^{\vec{a}} \\
&\vdash_{\vec{a}} (\bigvee_{fi=fk} a_i \wedge a_k = \bigvee_{gi=fk} a_i \wedge a_k) \wedge \phi_\alpha^{\vec{a}} \\
&\vdash_{\vec{a}} (a_k = 0) \wedge \phi_\alpha^{\vec{a}} \\
&\vdash_{\vec{a}} \bot
\end{aligned}$$

□

We now prove that our constructions are mutual inverses.

THEOREM 3.7 (Classifying topos of $\mathbb{T}_{\text{bpalg}}$). $\mathrm{Lan}_y(U)$ *and* Lan_F *define an equivalence of categories*

$$\mathsf{Set}^{\mathsf{FinProb}^{op}} \simeq \mathsf{Set}[\mathbb{T}_{\text{bpalg}}].$$

PROOF. For this proof, let $\mathbb{T} := \mathbb{T}_{\text{bpalg}}$.

We first check that $\mathrm{Lan}_y(U) \circ \mathrm{Lan}_F \cong \mathrm{id}$. Since these functors preserve colimits, we check only need to check this on representables. For a U_{r_i}, we get natural isomorphisms

$$\mathrm{Lan}_y(U)(\mathrm{Lan}_F(yU_{r_i})) \cong \mathrm{Lan}_y(U)(y\{\vec{a}.\phi_{U_{r_i}}^{\vec{a}}\}) \cong [\![\vec{a}.\phi_{U_{r_i}}^{\vec{a}}]\!]_U.$$

On an object (X,μ) this evaluates to the partitions (a_i) of (X,μ) with measures $\mu(a_i)=r_i$, because the pullbacks and equalizers in the formula are computed on objects. On a morphism $f:(X,\mu)\to(Y,\nu)$, the presheaf evaluates to the map sending the partition (a_i) of (Y,ν) to the partition $(f^{-1}a_i)$ of (X,μ). Thus, this presheaf can be seen to be isomorphic to yU_{r_i}.

For proving naturality, let $(Y,\nu),(X,\mu),(X',\mu')\in\mathsf{FinProb}$. The diagram

$$\begin{tikzcd}
\mathrm{Hom}((Y,\nu),(X,\mu)) \arrow[r, "\cong"] \arrow[d, "f_*"'] & [\![\vec{a}.\phi^{\vec{a}}_{X,\mu}]\!]_U(Y,\nu) \arrow[d, "{[\![\vec{a},\vec{b}.\theta_f]\!]}"] \\
\mathrm{Hom}((Y,\nu),(X',\mu')) \arrow[r, "\cong"] & [\![\vec{a}.\phi^{\vec{a}}_{X',\mu'}]\!]_U(Y,\nu)
\end{tikzcd}$$

commutes because this is the way $\{\vec{a},\vec{b}.\theta_f\}$ internalizes in a Set-model. This concludes our proof of $\mathrm{Lan}_y(U)\circ\mathrm{Lan}_F\cong\mathrm{id}$.

We now prove that $\mathrm{Lan}_F\circ\mathrm{Lan}_y(U)\cong\mathrm{id}$. Since all the functors involved are left-exact and colimit-preserving, it suffices to show that the isomorphism holds on the corresponding models of $\mathbb{T}$ inside $\mathsf{Set}[\mathbb{T}]$. This means we have to compute $\mathrm{Lan}_F\circ\mathrm{Lan}_y(U)$ on the signature of $\mathbb{T}$ and show that the yielded structure is isomorphic as a $\mathbb{T}$-model to the universal model of $\mathsf{Set}[\mathbb{T}]$. [4]

We first check this for the sort B:

$$\mathrm{Lan}_F(U)\cong\coprod_{r\in[0,1]}y\{b.B_rb\}\cong y\{b.\top\}, \tag{*}$$

where the last isomorphism holds since the $\{b.B_rb\}$ cover $\{b.\top\}$ and are disjoint[5].

Next, we need to show that the above isomorphism actually is an isomorphism of $\mathbb{T}$-models. Explicitly, we have to show that for each function symbol f our isomorphism * induces an iso

$$\begin{tikzcd}
\mathrm{Lan}_F\,U^n \arrow[d, "\mathrm{Lan}_F\,U(f)"'] \arrow[r, phantom, "\cong"] & y\{a_1,\cdots,a_n.\top\} \arrow[d, "{[f(a_1,\cdots,a_n)=a]}"] \\
\mathrm{Lan}_F\,U & y\{a.\top\}
\end{tikzcd}$$

in the category of arrows in $\mathsf{Set}[\mathbb{T}]$, where the upper isomorphism is constructed by using 3.6.

The 0-ary operation $U(0):*\to U$ comes from the inclusion into the component U_0 of $\coprod_{r\in[0,1]}U_r$. Applying Lan_F and suitably composing isomorphisms * and $\{b.B_0b\}\cong\{[].\top\}$ from 3.6 yields the morphism

$$[a=0]_*:y\{[].\top\}\to y\{a.\top\},$$

as desired.

The binary operation $U(\wedge):U\times U\to U$ can be described as a map $\coprod_{t_{ij}\in P^2}U_{t_{ij}}\to\coprod_{r\in[0,1]}U_r$ by the components (see 2.15) $U_{t_{ij}}\to U_{t_{11}}$ sending the only the element of measure t_{11} to the element of measure t_{11}. Sending this through Lan_F and again suitably composing * and the *inverse substution* $[a_{ij}=b_i\wedge c_j]$ described in (iv) of 3.6 yields the map

$$[a=a_1\wedge a_2]_*:y\{a_1,a_2.\top\}\to y\{a.\top\},$$

which happens to be just as desired.

The unary operation $U(\neg):U\to U$ is given on components $U_r\to U_{1-r}$ by the map sending 0 to 1 and 1 to 0. Applying Lan_F and our isomorphisms yields the negation

$$[a=\neg b]_*:\{b.\top\}\to\{a.\top\},$$

as we wanted to prove.

[4]One should emphasize that this reduction step is the only reason that we laboriously showed that F is left-exact.

[5]This can be explicitly proven by using the Yoneda lemma and the sheaf condition.

The symbols 1 and $\vee$ now automatically get preserved, since they are both easily expressible in terms of 0,$\vee$ and $\neg$. Sending $U_r \to \coprod U_r$ through Lan_F gives the expected subobject $\{a.B_r a\} \to \{a.\top\}$, concluding the proof. □

REMARK 3.8. There is a way to embed $\mathsf{Set}[\mathbb{T}_{\text{bpalg}}]$ into a coherent topos. We simply expand our theory to include boolean algebras B together with a map $B \to A$ into an arbitrary abelian group A with a fixed element 1, satisfying conditions analogous to the ones for the measure of a boolean probability algebra. The classifying topos of this theory is coherent, and thus it has enough points. Since it contains $\mathsf{Set}[\mathbb{T}_{\text{bpalg}}]$ as a subtopos, this might yield a strategy for proving 3.7 without the lengthy syntactical detour.

3.2. Classifying the theory of intervals. We will show that the topos $\mathrm{Sh}(\mathsf{FinProb}, J_{at})$ classifies the following theory of intervals.

DEFINITION 3.9. Define the *theory of intervals* $\mathbb{T}_{\text{int}}$ by adding the axioms

$$B_s a \vdash_a (\exists b)\big((b = a \wedge b) \wedge B_r b\big)$$

for all $r, s \in [0, 1]$ with $r \leq s$ to the theory of boolean probability algebras.

The models of $\mathbb{T}_{\text{int}}$ in Set must all be uncountable.

EXAMPLE 3.10. We consider some examples of intervals.

(i) Let (A, μ) be an atomless probability algebra. Let $a \in A$ with $\mu(a) = r$ and let $s \in [0, r]$. We use the axiom of choice to construct an element $b \leq a$ of measure s. Consider the set

$$S := \{b \leq a \mid \mu(b) \leq s\}.$$

Since every chain in S has a supremum in S, we can use Zorn's lemma to get a maximal element $b \leq a$ with

$$\mu(b) \leq s.$$

Now consider the set

$$T := \{b' \leq a \mid b' \geq b, \mu(b') \geq s\}.$$

Using Zorn's lemma again, we get a minimal element $b' \leq a$, $b' \geq b$ such that

$$\mu(b') \geq s$$

By construction, for all c with $b \leq c \leq b'$, c must either be b or b'. This implies that $b' \wedge \neg b$ is an atom, or that $b' = b$. But the former is impossible due to the atomless hypothesis. Thus, we have constructed an element of measure s, which means that (A, μ) is an interval.

(ii) Let (I, λ) denote the boolean probability algebra associated to the σ-algebra of Lebesgue-subsets of $I := [0, 1]$ equipped with the Lebesgue measure (use the construction from example 3.2). This is an atomless measure algebra and thus an interval. We will refer to (I, λ) as the *Lebesgue interval*.

(iii) Note that not all intervals are atomless probability algebras. Consider the subalgebra

$$A = \{(x_i) \in \prod_{i=1}^{\infty} I \mid \text{ finitely many } x_i \neq 0 \text{ or finitely many } x_i \neq 1\}$$

of the product of boolean algebras $\prod_{i \in \mathbb{N}} I$, where (I, λ) is the Lebesgue interval. A is the boolean algebra generated by elements that are everywhere zero except at one index. Note that A is not complete, and we have a measure on this boolean algebra defined by

$$\mu((x_i)) := \sum_{i=1}^{\infty} 2^{-i} \lambda(x_i),$$

where λ denotes the Lebesgue measure on the ith Lebesgue interval. It can be easily seen that the defined boolean probability algebra is actually an interval.

We state the main result of this section.

THEOREM 3.11. *The topos* $\mathrm{Sh}(\mathsf{FinProb}, J_{at})$ *classifies the theory* $\mathbb{T}_{\text{int}}$. *The desired equivalence is given by the universal model* U.

PROOF. We extend our result about presheaves in Theorem 3.7 to the sheaf case. By Theorem 8.1.12 in [**8**], there exists a unique quotient theory (up to syntactic equivalence) of $\mathbb{T}_{\text{bpalg}}$, such that its models (in a topos $\mathcal{E}$) are precisely the J_{at}-homogeneous models (in $\mathcal{E}$). This theory is given by adding an axiom

$$\phi_{\alpha}^{\vec{b}} \vdash_{\vec{b}} (\exists \vec{a})\theta_{f}^{\vec{a},\vec{b}}$$

for each morphism $f : \alpha \to \beta$ in $\mathsf{FinProb}$ to the theory $\mathbb{T}_{\text{bpalg}}$. One can easily see that the axioms

$$\phi_{U_{r+s,t}}^{\vec{b}} \vdash_{\vec{b}} (\exists \vec{a})\theta_{f}^{\vec{a},\vec{b}}$$

are sufficient, where $f : U_{r,s,t} \to U_{r+s,t}$ joins the first two elements. Note that r, s and t denote any elements in $[0, 1]$. These are precisely reformulations of the interval axioms.

We have shown that theory $\mathbb{T}_{\text{int}}$ classifies the topos $j : \mathrm{Sh}(\mathsf{FinProb}, J_{at}) \to \mathsf{Set}^{\mathsf{FinProb}^{op}}$. In this setting, for any Grothendieck topos $\mathcal{E}$, we get a square

$$(*) \qquad \begin{tikzcd} \mathrm{Geom}(\mathcal{E}, \mathrm{Sh}(\mathsf{FinProb}, J_{at})) \arrow[r, "\simeq"] \arrow[d, "j\circ -"] & \mathbb{T}_{\text{int}}\text{-mod}(\mathcal{E}) \arrow[d] \\ \mathrm{Geom}(\mathcal{E}, \mathsf{Set}^{\mathsf{FinProb}^{op}}) \arrow[r, "\simeq"] & \mathbb{T}_{\text{bpalg}}\text{-mod}(\mathcal{E}), \end{tikzcd}$$

of geometric morphisms that commutes up to isomorphism (as in Theorem 8.1.3 in [**8**]).

To compute the universal model, insert the identity $\mathrm{Id}_{\mathrm{Sh}(\mathsf{FinProb}, J_{at})}$ into the top-left corner. Applying the left and bottom morphisms, yields the sheafification $j^*(U)$ (as a boolean probability algebra). Since the underlying presheaf U is already a sheaf by Proposition 2.5 (iii), this must be U. Thus, the result on the top-right must be the boolean probability algebra U, as required. □

We can describe more explicitly how this result lets us view *intervals* as geometric morphisms $\mathsf{Set} \to \mathrm{Sh}(\mathsf{FinProb}, J_{at})$.

REMARK 3.12. The diagram * in the proof above has fully faithful functors on the left and on the right. This means that we describe the functor

$$\mathbb{T}_{\text{int}}\text{-mod}(\mathcal{E}) \xrightarrow{\simeq} \mathrm{Geom}(\mathcal{E}, \mathrm{Sh}(\mathsf{FinProb}, J_{at}))$$

yielding the equivalence as a restriction of the functor in 3.6.

In particular, if we have a model (A, μ) of $\mathbb{T}_{\text{int}}$ in Set, we can view it as a geometric morphism $\mathsf{Set} \to \mathrm{Sh}(\mathsf{FinProb}, J_{at})$ by specifying it as the J_{at}-flat functor $\mathsf{FinProb} \to \mathsf{Set}$ given by sending a finite probability space (X, ν) to the set

$$\{(a_i)_{i\in X} \mid a_i \in A \text{ are pairwise disjoint}, \mu(a_i) = \nu(i)\}$$

and by sending a map $(X, \nu_1) \to (Y, \nu_2)$ to the map

$$(a_j)_{j\in X} \mapsto \Big(\bigvee_{f(j)=i} a_j\Big)_{i\in Y}.$$

3.3. Completeness and quantifier elimination. The abstract categorical statement of Theorem 3.11 has some very concrete consequences for the structure of formulas-in-context up to $\mathbb{T}_{\text{int}}$-provable equivalence, since those bijectively correspond to isomorphism classes of subobjects of a power of the universal model U.

While we typically the fragment of logic preserved by (inverse images of) geometric morphisms is geometric logic, if a topos (such as $\mathfrak{Prob}$) is boolean, then every geometric morphism into it will also preserve negation. This is the case since the complement b of a subobject a can be uniquely determined by saying that $a \wedge b = 0$ and $a \vee b = 1$, which is a property that gets preserved. In other words, every geometric morphism into a boolean topos is open.

This has the pleasant consequence that every model of $\mathbb{T}_{\text{int}}$ in *any* Grothendieck topos will satisfy all $\mathbb{T}_{\text{int}}$-provable sequent possibly containing *infinitary* first-order-formulas. In this way, the fact that the classifying topos of $\mathbb{T}_{\text{int}}$ is atomic does not only have consequences for geometric sequents, but also for infinitary first-order sequents.

In the following statement, let the term x^ϵ denote x if $\epsilon = 1$ and $\neg x$ if $\epsilon = 2$.

COROLLARY 3.13. *For every infinitary first-order formula $\{\vec{x}.\psi\}$ in context over the variables $x_1, \cdots, x_n$, there is a subset $S \subset P^n_{2,\cdots,2}$ such that we have the $\mathbb{T}_{\text{int}}$-provable equivalence*

$$\psi \dashv\vdash_{\vec{x}} \bigvee_{t_{i_1,\cdots,i_n} \in S} \phi^{\vec{y}}_{U_{t_{i_1,\cdots,i_n}}}[x_1^{i_1} \wedge \cdots \wedge x_n^{i_n}/y_{i_1,\cdots,i_n}],$$

where the context $\vec{y}$ used for substitution contains 2^n variables $y_{i_1,\cdots,i_n}$ indexed by n-tuples $i_1, \cdots, i_n$ of elements $i_j \in \{1, 2\}$.

PROOF. We start with an analysis of the subobjects of U^n. Recall (from Example 2.13) that $U_{r_1} \times \cdots U_{r_n} \cong \langle * \mid t_{i_1,\cdots,i_n} \in P^n_{r_1,\cdots r_n} \rangle$ and that the maps $U_{t_{i_1,\cdots i_n}} \to U_{r_1} \times \cdots U_{r_n}$ are given by

$$a_{i_1,\cdots i_n} \mapsto (\bigvee_{i_2,\cdots i_n} a_{1,i_2,\cdots i_n}, \cdots, \bigvee_{i_1,\cdots i_{n-1}} a_{i_1,\cdots i_{n-1},1}).$$

Composing these maps with the inclusions $yU_{r_1} \times \cdots yU_{r_n} \to U^n$ yields the subobject of U^n given by the formula-in-context

$$\{\vec{x}.\phi^{\vec{y}}_{U_{t_{i_1,\cdots i_n}}}[x_1^{i_1} \wedge \cdots \wedge x_n^{i_n}/y_{i_1,\cdots,i_n}]\}.$$

Moreover, these are precisely the atoms of the subobject lattice of U^n and thus there is a subset $S \subset P^n_{2,\cdots,2}$ such that we have an isomorphism

$$\{\vec{x}.\psi\} \cong \{\vec{x}. \bigvee_{t_{i_1,\cdots,i_n} \in S} \phi^{\vec{y}}_{U_{t_{i_1,\cdots,i_n}}}[x_1^{i_1} \wedge \cdots \wedge x_n^{i_n}/y_{i_1,\cdots,i_n}]\}.$$

By the discussion above about infinitary first-order logic, this does translate to the corresponding isomorphism in any (possibly categorical) model of $\mathbb{T}_{\text{int}}$. Completeness of the deduction system for infinitary first-order logic now implies the claim. □

This is a form of quantifier elimination. Every infinitary first-order formula is $\mathbb{T}_{\text{int}}$-provably equivalent to a quantifier-free geometric formula. The quantifier-free geometric formula may very well contain proper infinitary disjunctions, so this cannot be compared to a quantifier-elimination type statement for classical, finitary, first-order logic.

The case $n = 0$ of the previous result deserves some additional attention:

COROLLARY 3.14. *Every infinitary first-order formula over the empty context is $\mathbb{T}_{\text{int}}$-provably equivalent to $\top$ or $\bot$ in $\mathbb{T}_{\text{int}}$.*

PROOF. $P^0 = 1$, which has precisely two subsets that yield the formulas $\bot$ and $\top$. □

This result is particularly remarkable, since it tells us that for the purposes of studying infinitary first-order sentences for the theory $\mathbb{T}_{\text{int}}$, *the choice of the interval does not matter*. This supports the widespread practice in probability theory of working with *some* σ-algebra, but never putting any special attention on its specific features. In fact, it tells us that, without losing generality, we might as well work with the Lebesgue interval (I, λ).

On the flip side, this also gives a theoretical guarantee that $\mathfrak{Prob}$ is a suitable *synthetic* context for probability theory: Any infinitary sentence that you are able to prove for *some* interval, will also hold for the model U in $\mathfrak{Prob}$.

The question remains whether the theory $\mathbb{T}_{\text{int}}$ (and also the topos $\mathfrak{Prob}$) are suitable settings for *stating* the types of theorems one would be interested in probability theory.

While infinitary first-order logic does give us a reasonably expressive language for formulating statements about intervals, there is one key class of phenomena that does not get modelled well by $\mathfrak{Prob}$. Concretely, sequences or *countable sets of free variables* do not get modelled well by $\mathfrak{Prob}$, as is showcased by the following proposition.

Proposition 3.15. *The internal language of* $\mathfrak{Prob}$ *proves the negation of countable choice.*

Proof. The countable multi-product of the family $U_{1/n}$ is empty, as there is no finite probability space with a sequence of elements x_n with measures $1/n$. □

This gives us evidence that it is difficult (if not impossible) to use $\mathfrak{Prob}$ as a setting for any interesting analysis.

Still, it is important to remark that while this might seem like a deficiency of the topos $\mathfrak{Prob}$, it is a manifestation of a more general problem. The theory of classifying toposes is built on the notion of a geometric morphism, which by design only preserves geometric logic. While when mapping into $\mathfrak{Prob}$, we do also preserve infinitary first-order logic, this is not the case as soon as we have infinite products.

So, while we could easily fix the problem above by enlarging $\mathsf{FinProb}$ to contain more objects (such as countable finite probability spaces, or even general polish spaces), statements involving countable products will not have the usual semantic counterparts for intervals (there is no clear soundness theorem).

Fixing this does require *rethinking what the category of Grothendieck toposes and geometric morphisms is*, by modifying it so that the inverse image functors further preserve certain infinitary limits.

4. Simple functions in $\mathfrak{Prob}$

While our logical results about $\mathbb{T}_{\text{int}}$ are certainly interesting, there is no doubt that probability theory is really about spaces of functions on σ-algebras. Formulas over $\mathbb{T}_{\text{int}}$ might not be sufficient for expressing statements about functions, however, the classifying topos $\mathbb{T}_{\text{int}}$ contains additional objects that cannot be defined by formulas in $\mathbb{T}_{\text{int}}$. Among those objects is the algebra of simple functions over the interval U.

We shall note here again, that we limit ourselves to studying combinatorial aspects of functions by studying the simple ones. The reason for this is that the study of analytic phenomena (such as integrable functions) would require an enlargement of our original site $\mathsf{FinProb}$[6]. Nevertheless, the benefit of working out the theory for the less complex topos $\mathfrak{Prob}$ is, that it gives us a relatively simple and well fleshed-out theory.

Definition 4.1. Define the presheaf $\mathcal{S}$ of simple functions on $\mathsf{FinProb}$ to be defined by sending (X, μ) to the set of maps of sets $\mathbb{R}^X$ and by sending maps between finite probability spaces to the corresponding precomposition maps.

We show that $\mathcal{S}$ is a sheaf by showing that $\mathcal{S}$ is a coproduct of atoms.

[6]The authors arrived at this conclusion after extensive trial and error.

PROPOSITION 4.2. *We have that*

$$\mathcal{S} \cong \langle \{y_1 < \cdots < y_n \mid y_i \in \mathbb{R}\} \mid (r_1, \cdots, r_n) \in \mathcal{A} \rangle$$

and in particular, $\mathcal{S}$ is a sheaf.

This is actually a way to describe the simple functions $\mathcal{S}$ through their distributions. In fact, interpreting subobjects as formulas, *the most we can say about a single simple function is stating its distribution*.

PROOF. Each pair $(r_i), (y_i)$ induces an element $(y_i) \in \mathcal{S}(U_{r_i})$ and thus induces a map

$$f_{r_i, y_i} : yU_{r_i} \to \mathcal{S}.$$

This map can be checked to be sectionwise injective and thus a monomorphism. Now assume that there is another pair r_i', y_i' such that the subobjects f_{r_i, y_i} and $f_{r_i', y_i'}$ are the same. Since the domains must match, $r_i = r'_{\sigma(i)}$ for a permutation σ and thus by Yoneda, $y_i = y'_{\sigma(i)}$. However, since there is an order on y_i and y_i', this implies that $\sigma = \mathrm{id}$. It follows that the induced map

$$\langle \{y_1 < \cdots < y_n \mid y_i \in \mathbb{R}\} \mid (r_1, \cdots, r_n) \in \mathcal{A} \rangle \to \mathcal{S}$$

is injective. One can show that this map is sectionwise surjective by describing any $s \in \mathbb{R}^X$ by looking at the ordered set of values in $\mathbb{R}$ that the function s takes. □

The aim of this chapter is convincing the reader that for a model $F : \mathcal{E} \to \mathfrak{Prob}$, $F^*(\mathcal{S})$ is a suitable definition for the simple functions over F in $\mathcal{E}$. We will prove that we can recover the simple functions over a σ-algebra with this construction. Of course, for proving meaningful statements, we want to equip $\mathcal{S}$ with suitable structure.

We take Terrance Tao's blog post [**9**] as a rough model for the following definition.

DEFINITION 4.3. Define the theory of expectation algebras to be the geometric theory of $\mathbb{R}$-algebras together with a unary relation symbol E_r for each $r \in \mathbb{R}$, adding the following geometric sequents:

(i) *E_r form a partition*. For all $r, s \in \mathbb{R}$ with $r \neq s$ we have an axiom

$$E_r(a) \wedge E_s(a) \vdash \bot$$

and we also have an axiom

$$\top \vdash \bigvee_{r \in \mathbb{R}} E_r(a).$$

(ii) *$\mathbb{R}$-linearity*. For all $r, s \in \mathbb{R}$, we require the sequents

$$E_r(a) \wedge E_s(b) \vdash E_{r+s}(a+b)$$

and for all $r \in \mathbb{R}$ and $\lambda \in \mathbb{R}$, we also need

$$E_r(a) \vdash E_{\lambda r}(\lambda \cdot a).$$

(iii) *Positivity*.

$$\top \vdash \bigvee_{r \geq 0} E_r a^2$$

(iv) *Expectation*.

$$\top \vdash E_1 1$$

(v) *No nullsets*.

$$E_0(a^2) \vdash a = 0$$

Here we made use of the same trick we used when defining the theory of boolean probability algebras. We used unary relation symbols to encode a function. The thing that now is different, is that if we have a real $\lambda \in \mathbb{R}$, we get again a canonical element of our $\mathbb{R}$-algebra A, since we can consider $\lambda \cdot 1$. In this way, we have a *subalgebra of real numbers* $R \subset A$ and an $\mathbb{R}$-linear map $A \to R$, which can be thought of *the expectation operator*.

Now under the usual functional-analytic correspondence of finitely additive measures and linear operators to $\mathbb{R}$, the fact that a measure is non-negative corresponds to our *positivity* axiom. One can see this by noticing that the indicator functions a in our expectation algebra are characterized by the equation $a^2 = a$ (at least for boolean probability algebras in Set). The fact that a measure is a probability measure, corresponds to our *expectation* axiom. The *no nullsets* axiom corresponds to the fact that there is only one element of measure zero in a boolean probability algebra. Note that we could have also axiomatized the operator $\mathbb{E}$ directly, although that wouldn't necessarily have been easier.

Remark 4.4. This theory of expectation-algebras is perfectly suitable for expressing questions about order: the predicate $\exists z, y - x = z^2$ can be viewed as an abbreviation of $x \leq y$. This yields the intended semantics for models in Set.

We now define expectation algebra structure on $\mathcal{S}$.

Definition 4.5. Equip $\mathcal{S}$ with the componentwise $\mathbb{R}$-algebra structure induced from the $\mathbb{R}$-algebras $\mathbb{R}^X$. Define $E_r \to \mathcal{S}$ to be the subsheaf that on an object (X, μ) consists of the $s \in \mathbb{R}^X$ with $\sum_{i \in X} s(i) = r$.

Using this structure, we can sectionwise verify that $\mathcal{S}$ does indeed satisfy the expectation algebra axioms, when viewed as a presheaf. It follows that it also satisfies those axioms when viewed as a sheaf. Moreover, as a sheaf there will be additional axiom analogous to the *interval axiom*, which we did not state.

Our aim is to send $\mathcal{S}$ and its structure through inverse image functors. Thus, it is imperative to figure out how our structure behaves on the atoms. The first step is decomposing $\mathcal{S} \times \mathcal{S}$:

$$\begin{aligned} S \times S &\cong \langle \{y_1 < \cdots < y_n \mid y_i \in \mathbb{R}\} \mid (r_1, \cdots, r_n) \in \mathcal{A} \rangle^2 \\ &\cong \langle \{y_1 < \cdots < y_n, z_1 < \cdots < z_m \mid y_i, z_j \in \mathbb{R}\} \mid (t_{ij})_{1 \leq i \leq n, 0 \leq j \leq m} \in \mathcal{A} \rangle \end{aligned}$$

Now we can use 2.15 to specify our structure on the elements of this set of atoms:

Proposition 4.6. (i) *We calculate addition on an element (t_{ij}, y_i, z_j). Let $x_1 < x_2 < \cdots < x_k$ denote the elements of $\{y_i + z_j \mid i, j\}$. Then addition is given on elements by*

$$(t_{ij}, y_i, z_j) \mapsto (\sum_{y_i + z_j = x_k} t_{ij}, x_k)$$

with the corresponding map on atoms

$$U_{t_{ij}} \to U_{\sum_{y_i + z_j = x_k} t_{ij}}$$

defined to be the join the two pairs (i, j), (i', j') if the equality $y_i + z_j = y_{i'} + z_{j'}$ holds.

(ii) *We calculate multiplication on an element (t_{ij}, y_i, z_j). Let $x_1 < x_2 < \cdots < x_k$ denote the elements of $\{y_i \cdot z_j \mid i, j\}$. Then, the multiplication is given on elements by*

$$(t_{ij}, y_i, z_j) \mapsto (\sum_{y_i \cdot z_j = x_k} t_{ij}, x_k)$$

with the corresponding map on atoms

$$U_{t_{ij}} \to U_{\sum_{y_i \cdot z_j = x_k} t_{ij}}$$

defined to be the join the two pairs (i, j), (i', j') if the equality $y_i \cdot z_j = y_{i'} \cdot z_{j'}$ holds.

(iii) *Scalar multiplication by $\lambda \in \mathbb{R}$ applied to an element (t_i, y_i) is $(t_i, \lambda \cdot y_i)$.*

(iv) *$E_r \to \mathcal{S}$ is the subobject*

$$\langle \{y_1 < \cdots < y_n \mid y_i \in \mathbb{R}, \sum r_i y_i = r\} \mid (r_1, \cdots, r_n) \in \mathcal{A} \rangle$$

of $\mathcal{S}$.

Having defined the simple functions in $\mathfrak{Prob}$, we can now transport this object to any setting whether there is an interval (a boolean probability algebra would suffice).

Definition 4.7. Let $F : \mathcal{E} \to \mathfrak{Prob}$ be an interval in $\mathcal{E}$. Then we can define the simple functions over F to be $F^*(\mathcal{S})$ equipped with the corresponding expectation algebra structure.

For the sake of comparing this unusual definition of the set of simple functions to the more standard one, we recall the following:

Definition 4.8. Define the expectation algebra of simple functions $S(X, \Sigma, \mu)$ on a probability space (X, Σ, μ) to be the set of measurable maps $X \to \mathbb{R}$ of finite image modulo equality almost everywhere equipped with pointwise addition, pointwise multiplication and expectation. The expectation is given by the usual expectation of a simple function.

Proposition 4.9. *Let (X, Σ, μ) be a probability space such that the corresponding measure algebra is atomless. Then $S(X, \Sigma, \mu)$ is isomorphic to the expectation algebra over the corresponding interval.*

Proof. The atomless measure algebra of (X, Σ, μ) is an interval and thus can be viewed as a geometric morphism $F : \mathsf{Set} \to \mathfrak{Prob}$. F^* sends an atom U_{r_i} to the set of partitions of the interval with measures $r_1, \cdots, r_n$. Thus, since F^* is colimit preserving, $\mathcal{S}$ gets sent to

$$\coprod_{\substack{r_1, \cdots r_n \in \mathcal{A} \\ x_1 < \cdots < x_n}} F^*(U_{r_1 \cdots r_n}).$$

It should be clear that this corresponds to the traditional set of simple functions. It is a straightforward calculation that expectation, addition and multiplication are the correct ones. The necessary setup has already been done, since addition and multiplication have already been decomposed into their components on atoms. □

This proof illustrates one of the most fascinating things about the set of atoms notation(the same goes for the $(\mathsf{FinProb}, J_{at})$ site). Partitions are treated on a very abstract level, where it does not even make sense to ask *where* the different parts of the partitions are located. There is simply no data about the location of the partition, only how big it is. The way we defined the simple functions, we never had to deal with issues about *where* exactly a simple function took a specific value, but instead just specified the measure of the preimages.

Applying the inverse image functors of models inserts that missing location information. The sets we deal with are suddenly much, much bigger. The surprising thing is that for the purposes of calculating with simple functions, that data is fully irrelevant.

We move to applications of the simple function construction in logic.

Theorem 4.10. *Let $I : \mathcal{E} \to \mathfrak{Prob}$ be an interval in a Grothendieck topos $\mathcal{E}$. Then the expectation algebra $I^*(S)$ of simple functions over I in $\mathcal{E}$ satisfies the same infinitary first-order sentences as the expectation algebra $\mathcal{S}$. In particular, the simple functions over any interval satisfy the same sentences.*

Proof. This follows from the fact that $\mathfrak{Prob}$ is two-valued. □

At it's core, this is a first promising result about *the irrelevance of the base probability space*. Given any probability space such that it's associated measure algebra is atomless, the sentences satisfied by the corresponding expectation algebra of simple functions will always be the same. This is a first taste of how

topos theory can give us *theoretical guarantees*, that the common practice of paying no attention to the base probability space is actually completely justified.

We can give an apparently stronger, topos-theoretic variation of the Theorem 4.10.

DEFINITION 4.11. We define the theory of $\mathbb{T}_{\text{sint}}$ simple interval-like expectation algebras to be the theory of expectation algebras together with the *interval axiom*

$$\top \vdash E_s a \wedge a^2 = a \wedge_a \exists b, b = b^2 \wedge b = a \cdot b \wedge E_r b$$

for any $r, s \in [0,1]$ with $r < s$ and the *idempotent generation axiom*

$$\top \vdash \bigvee_{n \in \mathbb{N}, r_1, \cdots r_n} \exists a_1, \cdots, a_n, a_1^2 = a_2 \wedge \cdots \wedge a_n^2 = a_n \wedge b = r_1 a_1 + \cdots + r_n a_n.$$

The interval axiom is to be interpreted the following way: The condition $a^2 = a$ says that a is an indicator function, and the E_r determine what we think of as its measure. The formula $b = a \cdot b$ expresses that $b \leq a$.

The idempotent generation axiom states that the set of idempotents forms a generating set of the base $\mathbb{R}$-vector space.

THEOREM 4.12. *S is a simple interval-like expectation algebra and exhibits $\mathfrak{Prob}$ as the classifying topos of the theory of simple interval-like expectation algebras $\mathbb{T}_{\text{sint}}$.*

PROOF. The expectation algebra S yields a geometric morphism $\mathfrak{Prob} \to \mathsf{Set}[\mathbb{T}_{\text{sint}}]$, with inverse image functor constructed by extending

$$F : \mathcal{C}_{\mathbb{T}_{\text{sint}}} \to \mathfrak{Prob}.$$

The key insight for this proof is that the category $\mathcal{C}_{\mathbb{T}_{\text{sint}}}$ already contains an *interval* of it's base elements. It's underlying object is given by the formula $V := \{a.a^2 = a\}$ expressing idempotence. We equip it with the following boolean probability algebra structure:

$$\begin{aligned} a \wedge b &:= a \cdot b \\ a \vee b &:= a + b - a \cdot b \\ \neg a &:= 1 - a \\ B_r a &:= E_r a \end{aligned}$$

The reader can check that this does indeed form an interval. For a finite probability space U_{r_i} we also get the corresponding object of partitions V_{r_i} interpreting the corresponding formula.

The functor F now carries V to the interval U, and we can see that all finite powers of U are in the essential image of F and so are all the maps and subobjects of the structure $(U, \vee, \wedge, \neg, U_r)$. Notice that we can express the objects U_{r_i} in terms of this structure and that we can moreover use the structure to define all morphisms $U_{r_i} \to U_{s_i}$ (both in terms of finite limits and colimits).

It follows that the U_{r_i} and the morphisms $U_{r_i} \to U_{s_i}$ are all in the essential image of F, and so is the whole of $\mathfrak{Prob}$ since those objects and arrows generate it (under finite limits and colimits).

Thus, F is full and essentially surjective on objects. It remains to show that it is fully faithful. To see that, it suffices to prove that the inverse image functor $V^* : \mathfrak{Prob} \to \mathsf{Sh}(\mathcal{C}_{\mathbb{T}_{\text{sint}}}, J)$ sends the expectation algebra $\mathcal{S}$ to the expectation algebra that corresponds to $\{a.\top\}$.

Using the decomposition of $\mathcal{S}$ into a coproduct gives us that $V^*(\mathcal{S}) \cong \coprod_{\substack{r_1, \cdots r_n \in \mathcal{A} \\ x_1 < \cdots < x_n}} V_{r_i}$ equipped with the expectation algebra structure defined on the components of the coproduct as in Proposition 4.6.

Indeed, by the idempotent generation axiom, we do have a surjection $\coprod_{\substack{r_1, \cdots r_n \in \mathcal{A} \\ x_1 < \cdots < x_n}} V_{r_i} \to \{b.\top\}$ induced by the formula $[b = x_1 a_1 + \cdots x_n a_n]$ on the component given by r_i, x_i.

To see that this function is an isomorphism, we have to prove that for two families r_i, $x_1 < \cdots < x_n$ and s_i, $y_1 < \cdots < y_n$, the corresponding subobjects are disjoint. More precisely, given that the two families

are distinct, we need to prove

$$x_1a_1 + \cdots + x_na_n = y_1b_1 + \cdots + y_mb_m \vdash \bot.$$

We consider two separate cases. In the first case, assume that $\exists i \forall j, x_i \neq y_j$. Then we have the following chain of sequents:

$$\begin{aligned} x_1a_1 + \cdots + x_na_n = y_1b_1 + \cdots + y_mb_m &\vdash \bigwedge_j x_ia_ib_j = y_ja_ib_j \\ &\vdash \bigwedge_j a_ib_j = 0 \\ &\vdash \sum_j a_ib_j = 0 \\ &\vdash a_i = 0 \\ &\vdash \bot. \end{aligned}$$

In the first line we multiply the equation with a_ib_j for all choices of j and in the second line, we use that $x_i \neq y_j$. In the last line, we get $\bot$, since we now have that E_0a_i but also $E_{r_i}a_i$.

For the second case, $\forall i \exists j, x_i = y_j$. Since both families are ordered, we must have that $n = m$ and $x_i = y_i$ for all i. If $n = 1$, there is nothing to prove now, so in the following assume $n \geq 2$.

$$\begin{aligned} \sum_i x_ia_i = \sum_i x_ib_i &\vdash x_1a_1 = a_1 \cdot \sum_i x_ib_i \\ &\vdash x_1a_1 \geq a_1x_2 \sum_i b_2 \\ &\vdash x_1a_1 \geq x_2a_1 \\ &\vdash \bot \end{aligned}$$

Here we're multiplying by a_1, and then using the partial order structure induced by the squaring operation to express and deduce $x_1a_1 \geq x_2a_1$. By taking expectations, we see that this is not possible, because the expectation of a square is always nonnegative.

It follows that we have an isomorphism

$$V^*(\mathcal{S}) \cong \coprod_{\substack{r_1, \cdots r_n \in \mathcal{A} \\ x_1 < \cdots < x_n}} V_{r_i},$$

as claimed. The reader can check that this preserves the expectation algebra structure. This concludes the proof. □

The remarkable thing about this result, is that is gives an example of geometric morphism, that is not an interpretation between theories in the classical model-theoretic sense, making this *a result that is unique to topos-theory*. It shows very concretely how Morita-equivalence of theories extends bi-interpretation of theories in such a way that we allow for constructions involving formal colimits.

In analogy to the case of intervals, the coproduct decomposition of $\mathcal{S}$ now gives a *normal form* for formulas in the theory of interval-like expectation algebras. We see that in this theory, formulas precisely describe sets of possible *joint distributions*. This mimics a very direct way how probability theorists think about random variables. The base probability space is, quite simply, a formality; and everything of relevance for random variables is their distributions and their joint distributions.

4.1. Practical considerations. We have managed to build a framework for interpreting *random variables* in $\mathfrak{Prob}$. While, as discussed previously, in this setting we are limited to finitary phenomena, there is still an interesting class of problems that we can analyze.

We can easily model a Bernoulli random variable by specifying its distribution. In fact, we can model any distribution with finitely many possible real values in this way.

Notice that there is a way to axiomatize independent random variables by an infinitary first-order formula describing the distributions that we admit. Similarly, things like Bayes' theorem can be stated and proven.

With these tools, the kind of problem in probability theory, that one might encounter in high-school. For instance, one can consider iterated coin flips, dice throws or turning of a roulette wheel and deduce the formula for the Binomial distribution that arises when we sum i.i.d Bernoulli-distributed random variables. One can also compute its expected value.

Moreover, we can also model dependencies between events or also random variables that allows us to specify the *probability tree diagrams* that we will often encounter when modeling problems involving drawing balls from an urn.

The theory of interval-like expectation algebras now provides us with a formal setting which we can use to *decide* or even *get normal forms* (in our case distributions) for this type of formulas. For an actual computer implementation, like an SMT-solver for this theory that one might then apply as tactics in an interactive theorem prover, however, some considerations would have to be made regarding how one computes with the reals. Interestingly, for this type of problems, it does not seem to be the case that *completeness* of the reals actually plays a role, so one could simply axiomatize a general linearly ordered field R instead of working with $\mathbb{R}$.

References

[1] Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory*. Springer Science & Business Media, 2012.

[2] Olivia Caramello and Laurent Lafforgue. Some aspects of topological galois theory. *Journal of Geometry and Physics*, 142:287–317, 2019.

[3] Richard A. Brualdi. *Combinatorial Matrix Classes*, volume 108. Cambridge University Press, 2006.

[4] Peter T. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium: Volume 2*. Oxford University Press, 2002.

[5] D. H. Fremlin. *Measure Theory Volume 3: Measure Algebras*. Torres Fremlin, 2012.

[6] Dmitri Pavlov. Gelfand-type duality for commutative von Neumann algebras. *Journal of Pure and Applied Algebra*, 226(4):106884, 2022.

[7] Asgar Jamneshan and Terence Tao. Foundational aspects of uncountable measure theory: Gelfand duality, Riesz representation, canonical models, and canonical disintegration. *arXiv preprint arXiv:2010.00681*, 2020.

[8] Olivia Caramello. *Theories, Sites, Toposes: Relating and studying mathematical theories through topos-theoretic'bridges'*. Oxford University Press, 2018.

[9] Terrance Tao. Algebraic probability spaces. https://terrytao.wordpress.com/2014/06/28/algebraic-probability-spaces/, 2014. Accessed: 2025-09-11.