The Vietoris Monad in Categorical Probability

Bachelor Thesis

Victor Bloch
Advisor: Tobias Fritz

Handed in: January 26, 2023

Abstract. Fritz et al. recently developed a novel approach to probability theory using category theory, see e.g. [4, 7, 5, 8, 6]. In particular, they developed three axioms as a categorical fundament to probability theory based on Markov categories. Fritz and Rischel [8] gave a proof that the Kleisli category of the Vietoris monad fulfils one of these axioms - the existence of infinite products. In this thesis, we will discuss their approach to categorical probability and finally prove that a category similar to the one investigated by Fritz and Rischel must violate one of the other two axioms.

Contents

1 Introduction

Category theory is a branch of mathematics that investigates mathematical structures and their relations in a general manner. Its basic concepts go back to Eilenberg and MacLane [16] who applied this structural approach to mathematics to obtain a link between algebra and topology, i.e. to improve and extend algebraic topology. Supposedly, one of their main goals was to introduce functors and natural transformations to describe, for example, homology and cohomology. To do so, they needed to introduce the concept of a category.

Nevertheless, algebraic topology is by far not the only mathematical, or even scientific branch in which category theory is applied. Nowadays, category theory is an indispensable part of, for instance, physics, logics, computer science and much more. The concept of a monad has become important in functional programming, the same way category theory became fundamental to type theory. To name a few examples of applications of category theory, category theory is used in quantum information and especially monoidal categories turned out to be particularly useful in information flow.

In the 1980s there were first attempts to place probability theory onto a categorical foundation. First, Lawvere utilised Kleisli categories to describe random processes in a category theoretical setting in [13]. Later, Giry [9] described the category Mea of measurable spaces and measurable maps, and equipped it with a monad now known as the Giry monad. This monad sends these to probability measures in order to describe random processes. Both concepts will be discussed in later sections.

Undeniably, probability theory can be described well in categorical language, as - in a very broad sense - random process can be understood as information flow. Therefore, Golubtsov [10] introduced a concept similar to what is now known as Markov categories which were later further developed by Fritz et al. [4, 7, 5, 8, 6]. Notably, Markov categories provide us with the possibility to copy and discard "information", useful when handling information and hence random processes.

Nevertheless, the aforementioned approaches to probability theory are not meant to replace the classic measure theoretic background of probability theory. The analytic part still is essential when demanding to actually calculate probabilities. However, abstract probability theory yields more general or new theorems as Fritz et al. demonstrate in e.g. [5, 8, 6].

To improve their approach and hence better describe probability theory in a categorical language, Fritz et al. [8, 6, 4] developed three axioms, united in [5] in order to formulate and prove an abstraction of de Finetti's theorem, individually used in [6] and [8] to prove theorems within statistics as well as zero-one laws, on which this paper focusses. Apparently, $\mathsf{BorelStoch}$ is the only non-trivial category yet known to fulfil all three of the axioms. In order to expand their approach, it would be useful to find more interesting categories, for instance of topological or measurable spaces, on the one hand to use their yet developed theory even further, and, on the other hand, to be able to improve their approach. Specifically, a category of topological or measurable spaces satisfying their axioms might give an insight on how to adapt their approach in order to encapsulate the yet well evolved measure-theoretic fundament of probability theory even better.

The primary purpose of this thesis was to find such a category. Unfortunately, this attempt was not successful. Nevertheless, we will give an example of a category that was determined to be a good candidate - the category arising from the Vietoris monad as its Kleisli category - and prove that it cannot satisfy both of two certain axioms at once.

Therefore, in order to provide a self contained discussion of this topic, section 2 recalls a selection of basic category theoretic definitions. Section 3 provides an introduction to categorical probability and introduces the three axioms developed by Fritz et al. Section 4 discusses de Finetti's theorem in its synthetic form. Finally, section 5 discusses the Vietoris monad and the reason why it cannot have conditionals (definition 3.8) and be a.s.-compatibly representable (definition 3.5) at once.

2 Preliminaries

This section recalls some basic category-theoretic concepts, most of them to be found in any standard textbook, e.g. [15], or [19] for monoidal categories. As this section is meant to just recall some notions, we will leave out technical details and concentrate on the important parts of the definitions. Especially, we will discuss the notion of a Markov category as developed in [4].

2.1 Monoidal categories

A monoidal category is a category with some extra structure that essentially allows us to "combine" objects in a multiplicative way. Concretely, a monoidal category is equipped with a bifunctor and an unit object, as we see in the following definition.

Definition 2.1 (Monoidal category)**.** *A monoidal category* $(\mathsf{C}, \otimes, I)$ *is a category* C *together with a bifunctor* $\otimes : \mathsf{C} \times \mathsf{C} \to \mathsf{C}$ *and a specific unit object* $I \in \mathsf{C}$*, such that certain diagrams commute.*

Example 2.2. *Let* Set *be the category of sets with functions. The cartesian product - which indeed corresponds to the categorical product - equips* Set *with a monoidal structure.*

Monoidal categories are particularly useful in categorical probability since they elegantly display the idea of parallel information flow. For example, in the category FinStoch of finite sets with Markov kernels, as discussed in [4, Ex. 2.5.], the monoidal structure is used to abstract the notion of independence in probability theory.

An important property of monoidal categories is symmetry. The symmetry of a monoidal category can intuitively be understood as the word suggests.

Definition 2.3. *A monoidal category* $(\mathsf{C}, \otimes, I)$ *is called symmetric, if for all objects* $A, B \in \mathsf{C}$*, there is a natural isomorphism* $\gamma_{A,B} : A \otimes B \to B \otimes A$*, that fulfils some more conditions.*

For a detailed discussion of monoidal categories and the unmentioned conditions see [19, Def. 5.3.1, Def. 5.3.2].

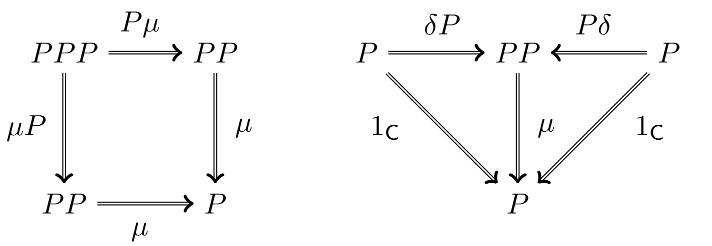
2.2 Monads

We now want to introduce the idea of monads treating them as the extension of a space to include generalised objects and morphisms. Suppose $(X, \mathcal{A})$ is a measurable space. Let PX denote the set of probability measures on $(X, \mathcal{A})$. There is a canonical σ-algebra on PX generated by functions $p_B : PX \to [0,1] : P \mapsto P(B)$, for $B \in \mathcal{A}$. This leads to an endofunctor $\mathcal{G} : \mathsf{Mea} \to \mathsf{Mea}$, where Mea is the category of measurable spaces and measurable maps, that sends every measurable space to its measurable space of probability measures, known as the Giry monad developed and discussed by Giry in [9].

Obviously, since PX is an object of Mea we can also consider PPX, i.e. the space of probability measures of the probability measures of the measurable space X. Furthermore, by integration we can embed PPX in PX, yielding a natural transformation $\mu : \mathcal{G}\mathcal{G} \Rightarrow \mathcal{G}$.

On the other hand, the Dirac-δ distribution induces a natural transformation $\delta : 1_{\mathsf{C}} \Rightarrow \mathcal{G}$ that maps every $x \in X$ to the corresponding Dirac-δ-measure δ_x. We recognize that a monad intuitively extends the objects of a category in a natural way. This motivates the following definition.

Definition 2.4 (monad)**.** *A monad on a category* C *consists of an endofunctor* $P : \mathsf{C} \to \mathsf{C}$ *and two natural transformations* $\mu : PP \Rightarrow P$ *called multiplication and* $\delta : 1_{\mathsf{C}} \Rightarrow P$ *called unit, such that the following diagrams commute:*



Two important examples of monads in this paper are the powerset monad and the Vietoris monad which will be discussed in the further sections.

Example 2.5. *Let* Set *be the category of sets with functions and* $\mathcal{P}$ *be the non-empty powerset functor, that maps every set to its corresponding non-empty powerset. For every* $X \in Obj(\mathsf{Set})$ *we define* $\mu_X : \mathcal{P}\mathcal{P}X \to \mathcal{P}X : A \mapsto \bigcup_{a \in A} a$ *which indeed defines a multiplication. Furthermore, define* $\delta_X : X \to \mathcal{P}X : x \mapsto \{x\}$ *to receive the unit.*

Example 2.6. *Let* Top *be the category of topological spaces and continuous functions. The Vietoris monad, as intensively investigated in [7, Sec. 2], is the monad which maps every topological space to the space containing its non-empty closed subsets as points equipped with the Vietoris topology, better known as its hyperspace. The multiplication is obtained analogously to example 2.5 by taking the closure of the union and similarly the unit is the closure of the singleton.*

2.3 Kleisli categories

The so-called Kleisli categories are categories that naturally arise from monads. It was first discovered by Kleisli in [12] and later discussed in the context of probability theory, also by Fritz et al. in [6], on whose ideas we will focus. We want to motivate the definition by a basic example: the category that arises from the powerset monad as in example 2.5.

Hence, let X, Y, Z be sets. Morphisms with "generalised" objects - as we introduced monads in section 2.2 - as targets $f : X \to \mathcal{P}Y$ and $g : Y \to \mathcal{P}Z$ become relations, which we would like to compose in the associated Kleisli category. To do so, we make use of the multiplication and unit transformation of the monad. By applying δ to the morphism g - sending all elements in Y and $\mathcal{P}Z$ to the corresponding singletons - we end up with $\mathcal{P}g : \mathcal{P}Y \to \mathcal{P}\mathcal{P}Z$ which can be composed with f. Then applying the multiplication - uniting the sets in $\mathcal{P}\mathcal{P}Z$ to receive an element in $\mathcal{P}Z$ - yields the composition of morphisms

$$X \xrightarrow{f} \mathcal{P}Y \xrightarrow{\mathcal{P}g} \mathcal{P}\mathcal{P}Z \xrightarrow{\mu} \mathcal{P}Z, \tag{1}$$

known as the Kleisli composition of f and g. These generalised morphisms together with the Kleisli composition are already enough to define the Kleisli category.

Definition 2.7. *Let (M, μ, δ) be a monad on a category* C. *The correlating Kleisli category $Kl(M)$ consists of*

- *objects $Obj(Kl(M)) = Obj(\mathsf{C})$,*
- *generalised morphisms $Hom_{Kl(M)}(X, Y) = Hom_{\mathsf{C}}(X, MY)$*
- *with the composition defined similar to 1.*

Remark 2.8. *For $\mathcal{P}$ the non-empty powerset monad on* Set *we define* $\mathsf{SetMulti} := Kl(\mathcal{P})$. *As explained in the introduction, the composition of two morphisms $f : X \to Y, g : Y \to Z$ is defined by*

$$(g \circ f)(x) := \bigcup_{x \in X} g(f(x)). \tag{2}$$

We now construct a topological analogue of $\mathsf{SetMulti}$. Consider CHaus the category of compact Hausdorff spaces with continuous functions. Michael proved in [14, The. 4.9.6.] that the hyperspace of a compact Hausdorff space as in example 2.6 is compact Hausdorff. Accordingly, the restriction of the Vietoris monad to compact Hausdorff spaces remains a monad. Thus, we can construct its Kleisli category.

Example 2.9. *Let V be the Vietoris monad on* CHaus. *Morphisms $X \to Y$ in $Kl(V)$ hence are continuous functions with domain X and codomain VY, the hyperspace of Y. Analogously to the thoughts for* $\mathsf{SetMulti}$ *and the the Vietoris monad in example 2.6, we receive a Kleisli composition similar to 2 yielding*

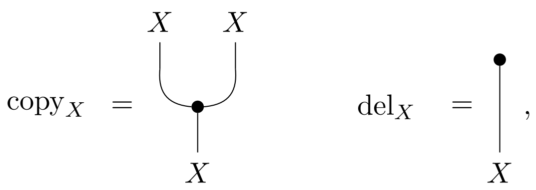
$$(g \circ f)(x) = \overline{\bigcup_{y \in f(x)} g(y)}.$$

The categories $\mathsf{SetMulti}$ and $Kl(V)$ will be further discussed in the following sections and the latter especially in section 5.

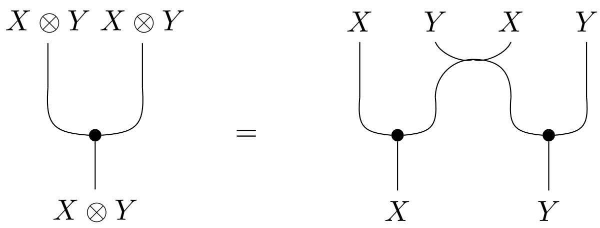
2.4 Markov categories

Fritz recently investigated Markov categories as a suitable framework for categorical probability in [4]. Basically, Markov categories are symmetric monoidal categories with some extra structure. Indeed, their morphisms behave very similar to Markov kernels.

Definition 2.10 (Markov category)**.** *A Markov category is a semicartesian, symmetric monoidal category* $(\mathsf{C}, \otimes, I)$*, where every object* X *is equipped with distinguished morphisms* $\mathrm{copy}_X : X \to X \otimes X$ *and* $\mathrm{del}_X : X \to I$*, written as*



such that they are compatible with the monoidal structure, meaning



for all objects X, Y *in* C*.*

Proposition 2.11. *Equip* $\mathsf{SetMulti}$ *with the cartesian product as monoidal structure. Then,* $\mathsf{SetMulti}$ *is a Markov category.*

Proof. For all $X \in Obj(\mathsf{SetMulti})$, there is exactly one non-empty relation between X and $\{*\}$, the monoidal unit. Hence, it is semicartesian. For $X, Y \in Obj(\mathsf{SetMulti})$ there is an obvious isomorphism $X \times Y \to Y \times X$, thus it is a symmetric monoidal category. We can trivially define the mappings $\mathrm{copy}_X : X \to X \times X : x \mapsto (x, x)$ and $\mathrm{del}_X : X \to \{*\} : x \mapsto *$ which are obviously compatible with the cartesian product. Therefore, $\mathsf{SetMulti}$ is indeed a Markov category. ☐

Similar thoughts yield the result that $Kl(V)$ also is a Markov category. Actually, it can be proven that a certain kind of monads respect the property of being a Markov category under the construction of a Kleisli category as done in [10] and discussed in [4, Sec. 3].

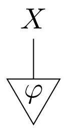
3 Categorical probability

As we have now introduced Markov categories, we want to investigate their use in categorical probability. We will start by translating some general probabilistic concepts into categorical syntax as done by Fritz in [4, Sec. 2] and later discuss the main object of interest in this paper, the three axioms developed by Fritz et al. in [4, 8, 6, 5].

While the first part of this section is merely of interest for the categorical approach to probability, these axioms already proved themselves to be an appropriate setting for categorical probability. For instance, in [5] Fritz et al. proved a synthetic version of de Finetti's theorem in purely categorical language, see section 4. Furthermore, in a slightly more general framework Fritz et al. proved abstract versions of the zero-one laws by Hewitt-Savage and Kolmogorov [8].

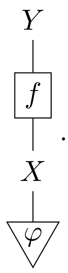
3.1 General theory

Lawvere already investigated the category of measurable spaces with Markov kernels as morphisms in [13]. He interpreted morphisms with the one-element space, i.e. the terminal object of the category, as domain as probability distributions. Indeed, the category of measurable spaces with Markov kernels is a Markov category as defined in the prior section. This motivates to define a distribution in a Markov category to be a morphism $\varphi : I \to X$, where I is the monoidal unit and X some arbitrary measurable space, denoted by

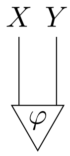


as Fritz does in [4, Sec. 2], since a distribution principally is a Markov kernel that takes no input and produces some random output.

This already motivates the definition of a random variable. If X and Y are two objects (measurable spaces in the case investigated by Lawvere [13]), we call a deterministic (definition 3.1) morphism $f : X \to Y$ random variable with distribution $f\varphi : I \to Y$, denoted by



Analogously for two objects X, Y of a Markov categeroy, the morphisms



correspond to the joint distributions of the objects X and Y.

However, the just introduced concepts are not essential for the currently developed aspects of categorical probability, but give a good understanding of how Markov categories "work". Let us now continue with the theory developed by Fritz et al. in their recent work [4, 7, 5, 8, 6] and particularly discuss their three new axioms as to be found collected in Assumption 4.3. in [5].

3.2 The axioms

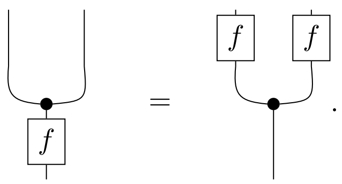
Fritz et al. [5, Sec. 3] developed three axioms defining the notions of representability, conditionals and infinite products. The first one reproduces the fact, that for every measurable space there is another measurable space of probability measures on that space. The notion of conditionals is related to the classic one in probability theory trying to generalise the concept of conditional probability. The last one axiomatises Kolmogorov's extension theorem and hence implements infinite products in this categorical approach to probability.

3.2.1 Representability

As already discussed in section 2.2, there is a canonical σ-algebra on the space of probability measures of every measurable space. The first axiom tries to implement this fact by essentially demanding the existance of a generalised object PX for every object X. For a precise definition, though, we need to define some more properties of morphisms in Markov categories.

The first one is the definition of deterministic morphisms. Imagine you roll a dice and "copy" the output. In general, this will yield a different result to rolling two dices and writing down their output. In other words, rolling a dice is not deterministic but there is some kind of uncertainty when producing the output, i.e. the rolled face on the dice. In categorical language in the setting of Markov categories this can be described as follows.

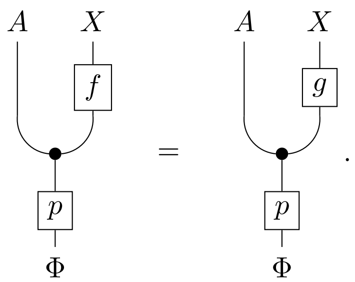
Definition 3.1 (deterministic morphism). *Let $f : A \to X$ be a morphism in a Markov category C. We call f deterministic if*



We denote the subcategory of deterministic morphisms of C by C_{det}.

For a detailed discussion of the term see section 10 in [4]. We will move on by abstracting the term of "equal almost everywhere" as done by Cho and Jacobs in [2] and developed by Fritz et al. in [6] to a more general setting. In standard measure theoretic language two functions are equal almost everywhere, if their output values are the same for all inputs in the domain but a subset of measure zero. Indeed, this property relies on a measure and its negligible subsets of measure zero. This already motivates the following definition of almost sure equality.

Definition 3.2 (a.s.-equality)**.** *Given $p : \Phi \to A$, we call two morphisms $f, g : A \to X$ p-almost surely equal, denoted by $f =_{p-a.s.} g$, if*



Example 3.3. *In the case of the Markov category* BorelStoch*, i.e. the category of standard Borel spaces with Markov kernels as morphisms, as described by Fritz in [5], this property specialises to equality of two Markov kernels with probability 1 for all values of Φ. In general, two morphisms are a.s.-equal, if they are only different for events that are negligible for p.*

Example 3.4. *Let p, f, g be morphisms in* SetMulti *according to definition 3.2. Remember that $\forall \varphi \in \Phi : p(\varphi) \subset A$ and that $\forall \varphi \in \Phi : (f \circ p)(\varphi) = \bigcup_{y \in p(\varphi)} f(y)$, $(g \circ p)(\varphi) = \bigcup_{y \in p(\varphi)} g(y)$. Thus, $f =_{p-a.s.} g$, if $\forall y \in p(\Phi) : f(y) = g(y)$. That is, f and g are p-almost surely equal, if they map the same subsets generated by p to the same subsets of X. Note that f and g may differ when applied to single elements of A.*

Now we are ready to define the notion of a Markov category to be representable respectively a.s.-compatible representability. Those notions are needed to prove the abstract version of de Finetti's theorem.

Definition 3.5 (representable Markov category, a.s.-compatible representability)**.** *Let* C *be a Markov category. We call* C *representable if for all objects X in* C *there exists a distribution object PX, such that*

$$Hom_{\mathsf{C}}(A, X) \cong Hom_{\mathsf{C}_{\text{det}}}(A, PX), \tag{3}$$

for arbitrary A and naturally in A. We denote the deterministic counterpart of $f : A \to X$ under this bijection by $f^{\#} : A \to PX$. Accordingly a representable Markov category is called a.s.-compatibly representable if for all objects A, X and all morphisms $f, g : A \to X$

$$f =_{p-a.s.} g \quad \Leftrightarrow \quad f^{\#} =_{p-a.s.} g^{\#}, \tag{4}$$

for all $p : \Phi \to A$.

Fritz et al. discuss the notion of representability extensively in [6]. For example, they argue that the category BorelStoch as described above is indeed a.s.-compatibly representable [6, Ex. 3.19.]. Another important morphism when talking about representability is the one induced by setting $A = PX$ in equation 3. It defines a correspondence to the identity on PX which we call samp since it intuitively "samples" from a given probability distribution. Furthermore, by composing the right-hand side of the equivalence in 4 with the sample morphism we immediately obtain the left-hand side, hence only the reversed direction is not trivial.

We now will discuss the category SetMulti in terms of representability.

Proposition 3.6. *A morphism f in* SetMulti *is deterministic if and only if it is single-valued, i.e.* $\mathsf{SetMulti}_{\mathsf{det}} \cong \mathsf{Set}$.

Proof. $A \xrightarrow{f} B$ is deterministic if and only if $(f \otimes f)(A \times A) \stackrel{!}{=} f(A) \times f(A)$ which holds if and only if f is single-valued. □

Proposition 3.7. SetMulti *is a.s.-compatibly representable.*

Proof. A morphism in $\mathsf{SetMulti}_{\mathsf{det}}(A, \mathcal{P}X)$ is a function $f^{\#} : A \to \mathcal{P}X$, which we again - by proposition 3.6 - can interpret as a relation and hence $\mathsf{SetMulti}_{\mathsf{det}}(A, \mathcal{P}X) \cong \mathsf{SetMulti}(A, X)$.

Now it is open to prove the a.s.-compatibility. As already described, the sufficiency in equation 4 is trivial, thus we only need to show necessity. Let $p : \Phi \to A$ and $f, g : A \to X$ be p-a.s. equal morphisms in SetMulti. We obtain the corresponding deterministic morphisms $f^{\#}, g^{\#} : A \to \mathcal{P}X$ by applying the natural transformation δ_A to f respectively g. Let $\varphi \in \Phi$:

$$(f^{\#} \circ p)(\varphi) = \bigcup_{y \in p(\varphi)} f^{\#}(y) = \bigcup_{y \in p(\varphi)} f(\{y\}) =$$
$$\stackrel{Ex.\ 3.4}{=} \bigcup_{y \in p(\varphi)} g(\{y\}) = \bigcup_{y \in p(\varphi)} g^{\#}(y) = (g^{\#} \circ p)(\varphi)$$

Hence, $f^{\#} =_{p-a.s.} g^{\#}$. □

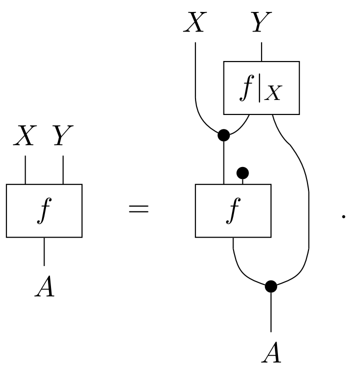
3.2.2 Conditionals

The second axiom implements a well-known fact from probability theory, the chain rule for random variables. Let f be the joint distribution of two random variables X, Y. Then,

$$f(X, Y) = f_{|X}(Y, X) f(X).$$

The idea of this axiom was first developed by Golubtsov in [10] and discussed by Cho and Jacobs in [2] and then further developed by Fritz et al. in [4, 5]. Our main focus lies on the results from the last ones. Thus, let us define the axiom.

Definition 3.8 (conditionals). *Let* C *be a Markov category. Let* $f : A \to X \otimes Y$ *be a morphism in* C. *We call* $f_{|X} : X \otimes A \to Y$ *a conditional of* f, *if*



Accordingly, we say that C *has conditionals, if a conditional exists for every such morphism.*

Intuitively, we can imagine that if we apply f to A and "delete" the Y-information, given X and A, $f_{|X}$ restores the lost information. Of course, the same holds if we exchange X and Y. Again, BorelStoch does indeed have conditionals, see [4, Ex. 11.7.]. As we will see, SetMulti also has conditionals.

Proposition 3.9. SetMulti *has conditionals.*

Proof. Let $f : A \to X \times Y$ be a morphism in SetMulti. Since SetMulti is the Kleisli category of the non-empty powerset monad, for all $a \in A$ we can interpret $f(a)$ to be a non-empty subset of $X \times Y$. If we fix one $x \in X$, we must distinguish two cases to construct $f_{|X}$.

1$^{\text{st}}$ case: $\exists y \in Y : (x, y) \in f(a)$. By $f(a)_x$ we then denote the corresponding non-empty subset of Y. Hence, we set $f_{|X}(x, a) = f(a)_x$.

2$^{\text{nd}}$ case: $\nexists y \in Y : (x, y) \in f(a)$. In this case, $f_{|X}(x, a)$ as constructed in the first case would be empty and hence $f_{|X}$ not a morphism in SetMulti. However, since they do not contain any relevant information - we only care about the x which we obtain from a together with a y - we can just extend $f_{|X}$. Thus, let M be an arbitrary non-empty subset of Y. We extend $f_{|X}$ by setting $f_{|X}(x, a) = M$.

Therefore, we have found a conditional of f and since this holds for arbitrary f, SetMulti does indeed have conditionals. □

Remark 3.10. *In the proof of proposition 3.9 we see that the conditional of a morphism needs not be unique. Precisely, Fritz proved in [4, Prop. 11.15.] that in a Markov category* C *that has conditionals, they are unique if and only if any two parallel morphisms are equal.*

3.2.3 Infinite products

One more property we would expect from an useful attempt to probability theory is the ability to handle infinite products. We thus want to introduce a way to treat infinite products of objects in Markov categories as extensively discussed by Fritz and Rischel in [8].

The main idea of this axiom is the Kolmogorov extension theorem. It states that the joint distributions of an infinite family of random variables are in bijection with the corresponding families of finite marginal distributions. The idea is to turn this fact into a definition, as done in [8, Def. 3.1., Def. 4.1.]. In the same paper Fritz and Rischel stated and proved abstract versions of the zero-one laws from Hewitt-Savage and Kolmogorov in a purely synthetic manner in terms of Markov categories using the mentioned notion of infinite products.

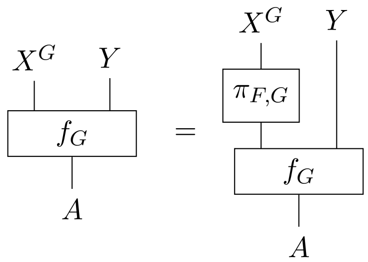
Definition 3.11 (infinite products, Kolmogorov powers)**.** *Let $(X_j)_{j\in J}$ be an arbitrary family of objects in a Markov category C. We call $X^J := \bigotimes_{j\in J} X_j$ an infinite (tensor) product of the family, if there is a natural bijection between morphisms*

$$f \in \mathsf{C}(A, X^J \otimes Y)$$

and families of morphisms

$$(f_F : A \to X^F \otimes Y),$$

where A and Y are objects in C and $F \subset \mathbb{N}$ finite. Additionally, we demand



, (5)

for $G \subset F$, where we call $\pi_{F,G} : X^F \to X^G$ finite marginalisation, acting as identity on G but del_X to the remaining factors. Furthermore, we call X^J a Kolmogorov power, if the finite marginalisation morphisms $\pi_F : X^J \to X^F$ are deterministic.

Remark 3.12. *Notice that we need to add the extra factor Y in definition 3.11 in order to preserve compatibility with the finite product. If we add a single factor we still want the infinite product to exist.*

As mentioned above, we are able to show some classical results in probability theory using this notion of infinite products. Nevertheless, in order to prove de Finetti's theorem as done in [5] it is enough to postulate an even less general case of infinite products, namely countable Kolmogorov powers of single objects, see definition 3.10. in [5]. For example, $\mathsf{BorelStoch}$ does have countable Kolmogorov powers of every object, as argued in example 3.6. in [8] making use of the Kolmogorov extension theorem. On the other hand, $\mathsf{SetMulti}$ does not have all Kolmogorov powers.

Proposition 3.13. SetMulti *does not have all Kolmogorv powers.*

Proof. Let A be a 2-element set. A morphism $I \to \prod_{i \in \mathbb{N}} A$ is a non-empty subset B of $\prod_{i \in \mathbb{N}} A$. We will see that the universal property of the product is not satisfied. Let $a, b \in \prod_{i \in \mathbb{N}} A$ be two distinct points and $B_a = \prod_{i \in \mathbb{N}} A \setminus \{a\}$, respectively $B_b = \prod_{i \in \mathbb{N}} A \setminus \{b\}$. For finite $J \subset \mathbb{N}$, the morphisms induced by these subsets with respect to the projection maps both satisfy equation 5 and are valid finite marginalisations, hence the demanded natural bijection does not exist. Therefore, the product does not exist. □

4 De Finetti‘s theorem

In this section we will concentrate on de Finetti's theorem named after Bruno de Finetti who discovered its original version in the 1930s, see [3]. Intuitively, de Finetti proved that every (infinite) exchangeable (definition 4.1) zero-one valued sequence - e.g. from a tossed coin - is a "mixture" of iid sequences of Bernoulli random variables.

The importance of this statement is easy to accept when thinking of frequently tossing a coin. While we may not decide on the independence of the joint distribution of the tosses we indeed can assume the tosses to be exchangeable. De Finetti then yields the non-trivial statement that there is a family of iid random variables which determine the distribution.

This theorem later was generalised to not only contain Bernoulli random variables but a wide range of measurable spaces. This broader version goes back to Hewitt and Savage in the 1950s, see [11], and is the one Fritz et al. adapted to a slightly different category theoretical version in [5]. Before we address the abstract statement, we want to revise the classic measure theoretic one in a more general setting than de Finetti's original valid for standard Borel spaces. To start off, we need to recall what it means to be exchangeable for a probability measure.

Definition 4.1. *Let X be a measurable space. Consider $X^{\mathbb{N}}$, the product of countable many copies of X equipped with the product σ-algebra and a probability measure p on $X^{\mathbb{N}}$. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a finite permutation and $n \in \mathbb{N}$ be the largest integer not fixed by σ. In this setting we call p exchangeable if,*

$$p(A_1 \times \ldots \times A_n) = p(A_{\sigma(1)} \times \ldots \times A_{\sigma(n)}),$$

for measurable subsets $A_1, \ldots A_n \subset X$, where $A_1 \times \ldots \times A_n$ is shorthand for the cylinder event $A_1 \times \ldots \times A_n \times X \times X \times \ldots$.

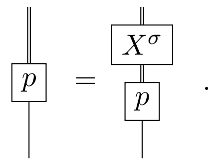
After revisiting the idea of exchangeability we are already able to state de Finetti's theorem characterising exchangeable measures on standard Borel spaces as in [5, Theo. 2.1.].

Theorem 4.2 (de Finetti's theorem)**.** *Let X be a standard Borel space and PX the (again measurable) space of measures on X. A probability measure p on $X^{\mathbb{N}}$ is exchangeable if and only if there exists a probability measure μ on PX such that for every finite collection of measurable subsets $A_1, \ldots, A_n \subset X$,*

$$p(A_1 \times \ldots \times A_n) = \int_{PX} q(A_1) \cdot \ldots \cdot q(A_n) \; d\mu(q).$$

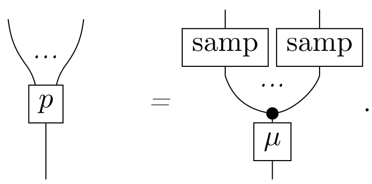
With this measure theoretic statement in mind we will now come into the realm of a synthetic adaption of theorem 4.2 as Fritz et al. devoted themselves in [5] on it. Instead of standard Borel spaces they formulated and proved a characterisation of exchangeability of morphisms in Markov categories. First of all, we need to again define the notion of exchangeability, this time in terms of morphisms in Markov categories, although, having definition 4.1 in mind, exchangeability of morphisms is very intuitive to understand as it apparently is only translated into another language.

Definition 4.3. *Let A and X be objects in a Markov category C that has countable Kolmogorov powers of every object. Let $\sigma : \mathbb{N} \to \mathbb{N}$ be a finite permutation. By $X^{\sigma} : X^{\mathbb{N}} \to X^{\mathbb{N}}$ we denote the morphism sending the n-th component of $X^{\mathbb{N}}$ to its $\sigma(n)$-th component. We then call a morphism $p : A \to X^{\mathbb{N}}$ exchangeable, if*



Again, the synthetic version characterises exchangeability but this time more generally for Markov categories. Beside being more general than theorem 4.2, the synthetic result even provides a more intuitive proof based on string diagrams, for full details, again, see [5]. Let us move on by stating the result, to be found as theorem 4.4. in [5].

Theorem 4.4 (synthetic de Finetti's theorem)**.** *Let C be a Markov category that has conditionals, is a.s.-compatibly representable and has countable Kolmogorov powers of every object. Then a morphism $p : A \to X^{\mathbb{N}}$ is exchangeable in the sense of definition 4.3 if and only if there is a morphism $\mu : A \to PX$ such that*



This result is indeed a generalisation of theorem 4.2 as we can immediately deduce it. As mentioned in section 3.2, $\mathsf{BorelStoch}$ has conditionals, is a.s.-compatibly representable and has countable Kolmogorv powers. Hence, by considering $A = I$ it is straightforward to infer theorem 4.2 as an instance from theorem 4.4 as Fritz et al. do on page 17f. in [5].

By now, there are no non-trivial Markov categories known to fulfil the presented axioms that are substantially different from BorelStoch. Of course it would be nice to find more interesting Markov categories, maybe of topological spaces or with some other interesting properties, in order to expand this novel attempt to categorical probability. For instance, an interesting category of topological or measurable spaces could help to adapt the theory in order to make it even more suitable and help it discover a proper measure theoretic background.

Although we are not yet able to provide a proper suggestion in which direction this journey will go, we will discuss one way of how we can decide on the suitability of a Markov category for this approach to categorical probability. As an (counter)example we will provide the category $Kl(V)$, introduced in example 2.9. We will discuss the details in the following section.

5 The Vietoris monad

We now will have a close look at the category that arises from the Vietoris monad on compact Hausdorff spaces as its Kleisli category as already mentioned in 2.9. Thus, we will start by recalling the concept of so-called hyperspaces.

Let (X, τ) be a topological space. One way to construct another topological space from this one is to have its non-empty closed subsets $Cl(X)$ as points. There are several ways to equip this point set with a topology that all have in common that they aim to make the embedding

$$\iota : X \to Cl(X) : x \mapsto \overline{\{x\}} \tag{6}$$

continuous. Beer [1] gives an introduction to several such hyperspaces, though we are particularly interested in one of those many possibilities - the topology now known as Vietoris topology first investigated by Vietoris in [17, 18]. Later, Fritz et al. discussed a coarser topology called lower Vietoris topology in a categorical setting in [7].

From now on, when mentioning a hyperspace, it will be equipped with the Vietoris topology. We denote the corresponding point space by VX and its topology by $\mathbb{V}$. A base for the Vietoris topology is obtained by elements of the form

$$\langle U_1, \ldots, U_k \rangle = \{C \in VX | C \subset \bigcup_{i=1}^{k} U_i \wedge \forall i = 1, \ldots, k : C \cap U_i \neq \emptyset\}, \tag{7}$$

with $U_1, \ldots, U_k \in \tau$. As already mentioned we will later concentrate on hyperspaces of compact Hausdorff spaces. Apparently, hyperspaces of compact Hausdorff spaces are again compact and Hausdorff.

Proposition 5.1. *Let (X, τ) be a compact Hausdorff space. $(VX, \mathbb{V})$ is compact and Hausdorff.*

Proof. We split the proof into two parts.

Hausdorff: Let $A, B \in VX$ be two distinct points. Without loss of generality $\exists a \in A : a \notin B$. Hence, $A \notin \langle X \setminus \{a\} \rangle$, but obviously $B \in \langle X \setminus \{a\} \rangle$. As X is Hausdorff we can find an open set U that contains a but does not intersect with B. Furthermore, again because X is Hausdorff, we can find an open set V containing all closed sets intersecting with U. We see that $A \in \langle V \rangle$ but $B \notin \langle V \rangle$. Thus, we found two disjoint neighbourhoods one containing A and the other B, making VX Hausdorff.

Compact: Let $\{c_i | i \in I\}$ be an open cover of VX. Since equation 6 is a continuous embedding, $\{\iota^{-1}(c_i) | i \in I\}$ is an open cover of X. Since X is compact, we can find a finite open subcover $\{\iota^{-1}(c_i) | i \in J\}$ of X for $J \subset I$ finite. Thus, since $\iota(X) = VX$, $\{c_i | i \in J\}$ is a finite open subcover of VX. □

From this proposition we can conclude that the functor $V : \mathsf{CHaus} \to \mathsf{CHaus}$ sending compact Hausdorff spaces to their corresponding hyperspaces is an endofunctor. Furthermore, we are able to define two natural transformations μ, δ as in definition 2.4 and obtain a monad. The multiplication is obtained analogously to example 2.5 by

$$\mu_X : VVX \to VX : A \mapsto \overline{\bigcup_{a \in A} a}$$

and

$$\delta_X : X \to VX : x \mapsto \{x\}.$$

It is easy to check that the defining diagrams commute.

We now will use this in order to construct an interesting category. All morphisms of CHaus itself are deterministic in the sense of definition 3.1, , as it is a concrete category, making it "boring" in this approach to categorical probability. A lot more exciting is the Kleisli category $Kl(V)$ arising from the monad V applied on CHaus that behaves much like SetMulti but has some advantages as we will discuss later.

First, let us construct $Kl(V)$ following definition 2.7. We again obtain compact Hausdorff spaces as objects. A morphism $f : X \to Y$ becomes a continuous function $f : X \to VY$ with the hyperspace of Y as codomain. Hence, it remains to define the corresponding Kleisli composition as in equation 1 in section 2. This is also straightforward using the just defined multiplication of the monad V and yields

$$(g \circ f)(x) = \overline{\bigcup_{y \in f(x)} g(y)} \tag{8}$$

for $f \in Kl(V)(X, Y)$, $g \in Kl(V)(Y, Z)$ and all $x \in X$. As per propositionem 3.1. in [6] $Kl(V)$ is a Markov category with the cartesian product with the product topology as monoidal product and the closure of the diagonal as the copy morphism.

We now want to discuss the usability of $Kl(V)$ in categorical probability. Although $Kl(V)$ behaves a lot like SetMulti, it indeed has Kolmogorov products of any cardinality while SetMulti does not even provide countable Kolmogorov powers of single objects, see proposition 3.13. The proof is analogous to the one by Fritz and Rischel for the lower Vietoris monad in [8, Prop. 6.4.].

In this manner, $Kl(V)$ indeed is an interesting candidate to investigate in a categorical probability setting. Thus, it would be interesting to know whether it also satisfies the other two axioms in order to, on the one hand, apply other yet known results in this approach to categorical probability and on the other hand to better understand the measure theoretic background of this approach and hence improve it.

Unfortunately, we will see that $Kl(V)$ cannot satisfy both of the other two axioms. To do so, we first take a closer look on how to handle certain morphisms of $Kl(V)$. Later we will use de Finetti's theorem in its abstract version in order to construct a counterexample showing that one of the other two axioms must be violated.

As described before, $Kl(V)$ is a lot like SetMulti. Particularly, we can regard its morphisms as some kind of generalisation of SetMulti to topological spaces. Let us demonstrate this by an example of one certain class of morphisms: $Kl(V)(I, \{0,1\}^{\mathbb{N}})$. Those morphisms are continuous functions with domain $\{*\}$ and codomain $V\{0,1\}^{\mathbb{N}}$, i.e. the hyperspace of $\{0,1\}^{\mathbb{N}}$. Since these functions are continuous and have a singleton as domain, they are in bijection with the codomain. In other words, $Kl(V)(I, \{0,1\}^{\mathbb{N}}) \cong V\{0,1\}^{\mathbb{N}}$. Hence, we can again imagine morphisms $p \in Kl(V)(I, \{0,1\}^{\mathbb{N}})$ to be elements $p \in V\{0,1\}^{\mathbb{N}}$, that is, they are nothing else but closed non-empty subsets of $\{0,1\}^{\mathbb{N}}$.

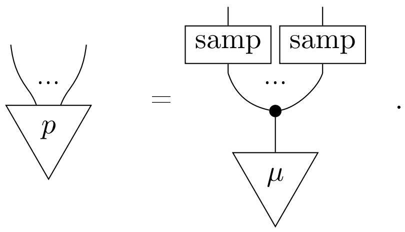
Let us now choose one of those morphisms, namely

$$q = \{(x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} | \exists_{\leq 1} i \in \mathbb{N} : x_i = 1\}. \tag{9}$$

Using a base representation (product topology) we immediately see that q is a closed subset of $\{0,1\}^{\mathbb{N}}$. We choose this morphism due to its exchangeability according to definition 4.3. As a simple logical conclusion, if $Kl(V)$ satisfied all three axioms, theorem 4.4 would be applicable to q. Thus, if it is not applicable, $Kl(V)$ cannot satisfy all three axioms.

Proposition 5.2. *$Kl(V)$ cannot have conditionals **and** be a.s.-compatible representable.*

Proof. To prove this, we translate de Finetti's theorem into this setting. In this case, theorem 4.4 states that a morphism $p : I \to \{0,1\}^{\mathbb{N}}$ is exchangeable if and only if there is a morphism $\mu : I \to V\{0,1\}$ such that



(10)

Let us have a look at the putative morphism μ. Analogous to our thoughts before, $\mu^{\#} \in VV\{0,1\} = \mathcal{PP}\{0,1\}$. Possible candidates for μ thus are representable by elements of the non-empty powerset of the non-empty powerset of $\{0,1\}$, i.e. $\mu^{\#} \in \{\{\{0\}\}, \{\{1\}\}, \{\{0,1\}\}, \{\{0\},\{1\}\}, \{\{0\},\{0,1\}\}, \{\{1\},\{0,1\}\}, \{\{0\},\{1\},\{0,1\}\}\}$.

We go on by concentrating on the right-hand side of 10. As per definitionem 3.5 the composition samp $\circ\mu$ specialises to

$$\text{samp} \circ\mu = \bigcup_{A\in\mu\subset V\{0,1\}} (A \times A \times \ldots).$$

As q from 9 is exchangeable there should exist some μ such that

$$q = \bigcup_{A\in\mu\subset V\{0,1\}} (A \times A \times \ldots).$$

However, checking the listed possibilities for μ from above yields that there is no such representation for q.

As a consequence, $Kl(V)$ cannot satisfy the three axioms. Since we already know that $Kl(V)$ has all Kolmogorov powers, it must dissatisfy at least one of the other two axioms, i.e. it either does not have conditionals or is not a.s.compatible representable. □

Hence, $Kl(V)$ is not the category we are looking for. It does not fulfil all three of the axioms and therefore will not help to improve this certain approach to categorical probability, while we do not want to note it unimportant.

Therefore, closed subsets of compact Hausdorff spaces are obviously not the right choice. Nevertheless, we can also think of restricting us to clopen subsets. Unfortunately, in some sense, this is even worse as we in general lose endofunctorality when considering the subspace topology of clopen sets of the hyperspace.

Proposition 5.3. *Let (X, τ) be a compact Hausdorff space with an infinite number of clopen subsets that do not include one another. The subspace topology $(C, \mathbb{V}_C)$ of clopen subsets of $(VX, \mathbb{V})$ is not compact.*

Proof. Since $\forall c \in C : c$ open, we can consider the base elements $\langle c\rangle$ as described in equation 7. Additionally, $\forall c \neq d \in C : c \in \langle c\rangle$ but $d \notin \langle c\rangle$. Thus, the collection $\{\langle c\rangle | c \in C\}$ is an open cover of C that has no finite subcover. Per definitionem, $(C, \mathbb{V}_C)$ hence is not compact. □

The following example shows that there actually are compact Hausdorff spaces with an infinite number of clopen subsets not containing one another.

Example 5.4. *Again consider the compact Hausdorff space $\{0,1\}^{\mathbb{N}}$ with the product topology. The sets*

$$X_J = \{(x_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}} | (x_j)_{j\in J} = (\tilde{x}_1, \cdots, \tilde{x}_{\#J})\}$$

for $J \subset \mathbb{N}$ finite and fixed sequences $(\tilde{x}_1, \cdots, \tilde{x}_{\#J}) \in \{0,1\}^{\#J}$ are clopen in $\{0,1\}^{\mathbb{N}}$. Now choose $J_n = \{n\}$ for $n \in \mathbb{N}$ and $\tilde{x}_n \in \{0,1\}$. $(X_{J_n})_{n\in\mathbb{N}}$ is an infinite family of clopen sets that do not contain one another and hence the restriction of the hyperspace to clopen sets does not preserve endofunctorality.

In summary, in this thesis we were not able to find another suitable category that helps to improve the categorical approach to probability theory developed by Fritz et al. in [4, 7, 5, 8, 6], though we were able to further investigate the category $Kl(V)$ in terms of categorical probability. The investigation of this particular question - what other categories satisfying the presented axioms there are - could include polish spaces as they are the foundation to the yet known category BorelStoch or even totally different directions. Nevertheless, category theory appears to not only be a very rich mathematical branch but also particularly useful in probability theory.

References

[1] Beer, G. *Topologies on Closed and Closed Convex Sets.* In: Mathematics and Its Applications. Springer, 1993.

[2] Cho, K. and Jacobs, B. *Disintegration and Bayesian inversion via string diagrams.* In: Mathematical Structures in Computer Science 29, 7 (2019), 938–971. https://doi.org/10.1017/S0960129518000488

[3] de Finetti, B. *La prévision : ses lois logiques, ses sources subjectives.* In: Annales de l'institut Henri Poincaré 7, 1 (1937), 1–68. http://eudml.org/doc/79004

[4] Fritz, T. *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics.* In: Advances in Mathematics 370 (2020). https://doi.org/10.1016/j.aim.2020.107239

[5] Fritz, T., Gonda, T. and Perrone, P. *De Finetti's theorem in categorical probability.* In: Journal of Stochastic Analysis 2, 4 (2021). https://doi.org/10.31390%2Fjosa.2.4.06

[6] Fritz, T., Gonda, T., Perrone, P. and Rischel, E. F. *Representable Markov categories and comparison of statistical experiments in categorical probability*, 2020. https://arxiv.org/abs/2010.07416

[7] Fritz, T., Perrone, P. and Rezagholi, S. *Probability, valuations, hyperspace: Three monads on Top and the support as a morphism.* In: Mathematical Structures in Computer Science 31, 8 (2021), 850–897. https://doi.org/10.1017%2Fs0960129521000414

[8] Fritz, T. and Rischel, E. F. *Infinite products and zero-one laws in categorical probability.* In: Compositionality 2 (2020). https://doi.org/10.32408/compositionality-2-3

[9] Giry, M. *A categorical approach to probability theory.* In: Categorical Aspects of Topology and Analysis (Berlin, Heidelberg, 1982), B. Banaschewski, Ed., Springer Berlin Heidelberg, pp. 68–85.

[10] Golubtsov, P. *Monoidal Kleisli category as a background for information transformers theory.* In: Infomration Theory and Information Processing, Volume 2, No. 1, 2002, p. 62–84. jip.ru/2002/GOLU1.pdf

[11] Hewitt, E. and Savage, L. J. *Symmetric measures on cartesian products.* In: Trans. Amer. Math. Soc. 80 (1955), 470-501. https://doi.org/10.1090/S0002-9947-1955-0076206-8

[12] Kleisli, H. *Every standard construction is induced by a pair of adjoint functors.* In: Proc. Amer. Math. Soc. 16 (1965), 544-546. https://doi.org/10.1090/S0002-9939-1965-0177024-4

[13] Lawvere, W. *The category of probabilistic mappings.* 1962. https://ncatlab.org/nlab/files/lawvereprobability1962.pdf

[14] Michael, E. *Topologies on spaces of subsets.* Trans. Amer. Math. Soc. 71 (1951), 152-182. https://doi.org/10.1090/S0002-9947-1951-0042109-4

[15] Riehl, E. *Category Theory in Context.* Dover Publications Inc., 2016.

[16] Eilenberg, S. and MacLane, S. *Eilenberg-MacLane: Collected Works.* Academic Press, 1986.

[17] Vietoris, L. *Bereiche zweiter ordnung.* In: Monatsh. f. Mathematik und Physik 32, 258–280 (1922). https://doi.org/10.1007/BF01696886

[18] Vietoris, L. *Stetige mengen.* In: Monatsh. f. Mathematik und Physik 31, 173–204 (1921). https://doi.org/10.1007/BF01702717

[19] Yanofsky, N. S. *Monoidal Categories.* unpublished. http://www.sci.brooklyn.cuny.edu/~noson/MCtext.html