De Finetti's Theorem in Categorical Probability

Tobias Fritz[*1], Tomáš Gonda[†2,3], and Paolo Perrone[‡4]

[1]Department of Mathematics, University of Innsbruck, Austria
[2]Perimeter Institute for Theoretical Physics, Waterloo ON, Canada
[3]School of Physics and Astronomy, University of Waterloo, Canada
[4]Department of Computer Science, University of Oxford, UK

May 10, 2021

Abstract

We present a novel proof of de Finetti's Theorem characterizing permutation-invariant probability measures of infinite sequences of variables, so-called exchangeable measures. The proof is phrased in the language of Markov categories, which provide an abstract categorical framework for probability and information flow. The diagrammatic and abstract nature of the arguments makes the proof intuitive and easy to follow. We also show how the usual measure-thereotic version of de Finetti's Theorem for standard Borel spaces is an instance of this result.

Contents

[*]tobias.fritz@uibk.ac.at

[†]tgonda@pitp.ca

[‡]paolo.perrone@cs.ox.ac.uk

1   Introduction

De Finetti's Theorem states that every permutation-invariant joint probability distribution of countably many random variables is such that these variables are conditionally independent given a suitable latent variable. Moreover, a canonical choice of the latent variable is the empirical distribution of the variables under consideration (which exists almost surely). In this paper, we state and prove a more abstract version of de Finetti's Theorem in the context of *categorical probability theory*, which is a nascent framework for a foundation of probability and statistics that is more abstract and more general than the traditional measure-theoretic approach.

Context of de Finetti's Theorem.   To provide some context for and illustrate the significance of de Finetti's Theorem, it is helpful to consider the notoriously divise debate on the subjective vs. objective view on probability [22, 32]. For our purposes, note that there are situations in which it is commonly presumed that one can give objective meaning to probability and infer its value. For example, flipping a coin and counting the relative frequency of 'heads' among the trials ought to converge to the bias of the coin. This is the content of the law of large numbers.

However, there is no free lunch and the assumption that the objective probability *exists* is crucial in the above reasoning. As de Finetti eloquently argues in Section 3 of [11], from the subjectivist point of view,

> "nothing obliges us to choose [probability of heads in the next toss] to be close to the frequency [of heads in the previously observed tosses]."

Let us give a brief account of the argument here and refer to de Finetti's classic [11] for more details.

In the coin-flipping experiment, suppose that we are given a (subjective) joint probability distribution over possible sequences of heads and tails. We cannot justify the trials to be independent without implicitly bringing in the assumption of objective probability. Nevertheless, we may require that the joint distribution is unchanged under permutations of the different trials, a property known as *exchangeability*[1] which expresses the belief that these are trials of the "same phenomenon".

De Finetti shows that with respect to any exchangeable probability measure, the individual trials are conditionally independent given some random variable Y. Moreover, there is a canonical choice for Y whose values precisely correspond to the possible biases of the coin, so that (in this example) Y takes values in the unit

---

[1]Exchangeable measures are occasionally also called "symmetric" or "equivalent", see the translator's note of [11].

interval $[0, 1]$. De Finetti's Theorem thus justifies the use of probability distributions in the subjectivist conception, since it says that exchangeable behaviors (of infinitely many trials) are indistinguishable from the behaviors of a repeated flipping of a coin with unknown, but fixed, bias.

Indeed, one can easily see that flipping a coin with an unknown bias, distributed according to some probability measure on Y, leads to a distribution of sequences of heads and tails that is exchangeable. The non-trivial part of de Finetti's Theorem is in the converse. It says that, given a countable number of trials, *every* exchangeable distribution arises in this way.

Historically speaking, the above context is the one in which the topic of our paper—de Finetti's Theorem—originates. However, besides its implications for the foundations of the concept of probability, the result also bears significance for the development of nonparametric Bayesian modelling [24, 31]. More concretely, many natural stochastic processes—such as drawing with replacement (e.g. the Pólya urn model and modifications thereof) or without replacement—give rise to exchangeable distributions. By de Finetti's Theorem, they can be equivalently described as drawing a random probability distribution q according to a specified prior *distribution over distributions*, μ (e.g. a Dirichlet process), and subsequently drawing independent samples from q. In other words, they correspond to a mixture of iid (independent and identically distributed) samplings, which is useful both as an aid for intuitive understanding and for concrete calculations.

Existing proofs and variants of de Finetti's Theorem. The original proofs by de Finetti in the binary case (as in coin-flipping) can be found in [10, 11], while a more general result is due to Hewitt and Savage [21], who have shown the analogous statement for exchangeable Radon probability measures on compact Hausdorff spaces. We give more details on the measure-theoretic formulation for standard Borel spaces in Section 2. Since then, many proofs based on different methods have appeared in various contexts. For instance, those of [26, Theorem 11.10] and [4, Theorem 3.1] use the mean ergodic theorem and conditional expectations; in [29, Theorem 12.17] and [3, Theorem 3.1] one arrives at the result via backwards martingales; while harmonic [30] and non-standard [2] analysis have also been used. The latter has recently led to a new generalization of de Finetti's Theorem for exchangeable Radon measures on *any* Hausdorff space [1]. Functional analysis and moment methods have been utilized in the proof from [15, Chapter VII], whose more elementary and explicitly calculational version that applies to binary variables can be found in [28].

Very recently, Jacobs and Staton [23] have presented a category-theoretic ap-

3

proach to de Finetti's Theorem, different from the present one, for the binary case. Therein, the statement of de Finetti's Theorem is encoded in the fact that the unit interval—the space of probability distributions—arises as the categorical limit of a sequence of multisets over $\{0, 1\}$ ("urns") related by morphisms that represent the action of drawing a random element from an urn without replacement.

Our version of de Finetti's Theorem. In Section 4, we present an abstract version of de Finetti's Theorem as a statement about morphisms in categories which admit notions of parallel composition, copying, and discarding, called *Markov categories* [17]. The framework of Markov categories is very general, and there are many Markov categories in which de Finetti's Theorem does not hold. Correspondingly, there are additional axioms that enter as ingredients in the proof of our abstract version of de Finetti's Theorem. Specifically, they correspond to the fact that one should be able to

- construct conditional probabilities,
- describe distributions on spaces of probability distributions, and
- consider countable sequences of trials.

These three requirements will be stated formally in Assumption 4.2, after introducing the basics of the formalism of Markov categories in Section 3. Within a Markov category that satisfies them, one therefore obtains a version of de Finetti's Theorem. As we show in Section 4, the measure-theoretic version for discrete and continuous random variables (described by standard Borel spaces) arises in this way as well. Additionally, we automatically get a characterization of exchangeable Markov kernels, unlike the classical results that focus exclusively on exchangeable measures.

It is likely that other Markov categories also have the required properties, and that instantiating the abstract de Finetti's Theorem in those leads to a context in which the result would be entirely new. However, we do not know of any such examples at present and leave the search for these to future investigations.

Synthetic probability theory. The process described above—identifying abstract results that follow from properties expressed within the framework and instantiating them in concrete Markov categories—is an example of the *synthetic* approach to probability theory, which is distinguished from the standard *analytic* one in terms of measure theory by encapsulating measure-theoretic statements in suitable higher-level axioms [17]. It differs from standard approaches by using more formal and abstract reasoning and it only depends on the particular measure-theoretic semantics insofar as the synthetic axioms may or may not be satisfied. A number of

concepts and theorems of classical probability and statistics have been given a synthetic treatment in recent years [7, 17, 19, 18]. We recall the ones relevant for de Finetti's Theorem in Section 3.

The proof that we present in Section 5 is inspired by several of the concrete ones mentioned above, but as far as we know it does not match either of them completely. Its abstract nature, which places the focus on the essential aspects while allowing us to ignore irrelevant details, arguably makes it easier to follow than any measure-theoretic one, although it is still far from obvious. The proof relies on the graphical calculus of string diagrams, which captures several non-trivial properties implicitly and whose connectivity explicitly depicts information flow. It is these two features that enable one to readily interpret any stage of the proof with relative ease, once some familiarity with the diagrammatic notation has been obtained.

Outlook. Given the intuitive nature of our proof, it is natural to hope that even deeper results can be proven along similar lines in a purely synthetic manner, and that one can ultimately aim at proving *new* statements that would be too difficult to obtain in the traditional measure-theoretic formalism due to its higher complexity. With this in mind, it may be worth mentioning a few extensions of de Finetti's Theorem which one can try to aim at next.

Among the most interesting variations on the exchangeability theme is arguably the notion of *partial exchangeability* [14], where the invariance under finite permutation invariance is relaxed to invariance under certain structure-preserving permutations. For example, the Aldous–Hoover Theorem [3, 4] characterizes exchangeable arrays of random variables and is closely related to random graphs. There is a similar result for Markov chains [12, 13, 6]. More recently, partial exchangeability has been generalized to hierarchical exchangeability [5, 25]. The work of Crane and Towsner provides perhaps the most general currently available results along these lines [8, 9], situated in a model-theoretic framework.

2 Measure-Theoretic Version of de Finetti's Theorem

Let us turn to a more formal exposition of de Finetti's Theorem in standard measure-theoretic language. Given a measurable space X, consider the product $X^{\mathbb{N}}$ of countably many copies of X, equipped with the product σ-algebra. A bijection $\mathbb{N} \to \mathbb{N}$ which fixes all elements apart from a finite subset is called a *finite permutation*. Given a finite permutation $\sigma \colon \mathbb{N} \to \mathbb{N}$, we denote by $n_\sigma \in \mathbb{N}$ the largest natural number not fixed by σ.

Consider a probability measure p on $X^{\mathbb{N}}$. By convention, given a finite collection of measurable subsets $S_1, \dots, S_n \subseteq X$, we write

$$p(S_1 \times \cdots \times S_n)$$

as shorthand for the probability of the "cylinder" event

$$p(S_1 \times \cdots \times S_n \times X \times X \times \ldots).$$

These probabilities specify the marginal distribution of p on the first n components of $X^{\mathbb{N}}$. We say that the measure p is *exchangeable* if for every finite permutation σ and for every finite sequence of measurable subsets $S_1, \dots, S_{n_\sigma} \subseteq X$, we have

$$p(S_1 \times \cdots \times S_{n_\sigma}) = p\big(S_{\sigma(1)} \times \cdots \times S_{\sigma(n_\sigma)}\big). \tag{1}$$

As we saw in the introduction, a somewhat trivial example of exchangeable measures is given by laws of iid random variables. They are the product measures satisfying

$$q(S_1 \times \cdots \times S_n) = q(S_1) \times \cdots \times q(S_n). \tag{2}$$

Having introduced the notion of exchangeability, we now turn to de Finetti's Theorem itself. A convenient way to express the statement, which lends itself well to category-theoretical translations, is to use the concept of *measures on a space of measures*, as done by Hewitt and Savage [21, Section 2] among others. If X is a standard Borel space, we denote by PX the set of probability measures on X. The set PX can be equipped with a canonical σ-algebra, namely the one generated by the functions $\varepsilon_f \colon PX \to \mathbb{R}$ given by

$$p \mapsto \int f(x)\, p(dx)$$

for all bounded measurable functions $f\colon X \to \mathbb{R}$. Measures *on PX* (thus elements of PPX) can be thought of as *random measures on X*, where also the specific form of the distribution is subject to uncertainty.[2]

Equivalently, measures on PX describe mixtures of measures—either in the sense of finite convex combinations or integrals. Indeed, de Finetti's Theorem can be summarized as the fact that exchangeable measures are mixtures of product measures. Here is the precise statement, in the version for standard Borel spaces.

Theorem 2.1 (de Finetti's Theorem). *Let X be a standard Borel space. S 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 $S_1, \ldots, S_n \subseteq X$, we have*

$$p(S_1 \times \cdots \times S_n) = \int_{PX} q(S_1) \times \cdots \times q(S_n)\, \mu(dq). \tag{3}$$

For example, the iid case amounts to μ being the delta measure $\delta_q \in PPX$. On the other hand, if μ is a random measure supported on delta measures $\delta_x \in PX$, then we obtain p supported on the diagonal. That is, the associated exchangeable random variables are perfectly correlated.

By convex analysis arguments, it can also be shown that given p as in Theorem 2.1, the measure μ is uniquely determined. Note that the result due to Hewitt and Savage [21] is more general than the one above, as it applies to Radon probability measures on arbitrary compact Hausdorff spaces. It includes the uniqueness of μ as well.

3 Markov Categories

We now take a detour from the discussion of de Finetti's Theorem into the realm of Markov categories. All of the concepts defined in this section have been introduced in earlier works on categorical probability [7, 17, 19, 18]. Nevertheless, in the interest of a self-contained presentation, we recall the main points here in a slightly less formal way, referring to the existing literature for full technical detail.

As mentioned in the introduction, a Markov category is a category that comes with notions of parallel composition, copying, and discarding. The idea is that contexts in which one wants to model flow of information will generally satisfy these basic requirements and thus correspond to a Markov category. In particular, one

---

[2]Such recursive ways of forming spaces can often be accurately captured by the categorical notion of a monad, and indeed these spaces of measures can be described in terms of a well-known monad, the *Giry monad* [20].

would expect that any formalisation of classical probability theory does. However, there are also many Markov categories that have nothing to do with probability theory (see [17]).

One of the simplest interesting examples to keep in mind is FinStoch, the category of finite sets and stochastic matrices. An object in FinStoch is a finite set, which can be interpreted as the set of possible values of a random variable. A morphism, say $f\colon A \to X$, assigns a probability measure on the finite set X to each element of the finite set A. It can thus be described as a *stochastic matrix* with entries $f(x|a)$ indexed by $a \in A$ and $x \in X$. Parallel composition of two stochastic maps, $f\colon A \to X$ and $g\colon B \to Y$, is just their tensor product $f \otimes g\colon A \otimes B \to X \otimes Y$, where $A \otimes B$ is the cartesian product of the underlying sets and one multiplies the probabilities in the formation of the tensor product. On the other hand, the sequential composition of $f\colon A \to X$ and $h\colon X \to Z$ to produce $h \circ f\colon A \to Z$ is given by the usual matrix multiplication. In the context of stochastic maps, it is also known as the *Chapman-Kolmogorov equation*.

In the string diagrammatic notation that we make heavy use of, objects are depicted as 'wires', while morphisms are generally drawn as 'boxes', with their domain below and their codomain above the box depicted as incident wires. Parallel and sequential composition is depicted by

$$
\begin{array}{c}
X \quad Y \\
\boxed{f} \; \boxed{g} \\
A \quad B
\end{array}
\qquad \text{and} \qquad
\begin{array}{c}
Z \\
\boxed{h} \\
\boxed{f} \\
A
\end{array}
$$
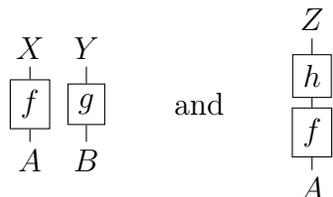
respectively.

Copying and discarding in FinStoch are likewise as one would expect. Specifically, $\mathrm{copy}_X\colon X \to X \otimes X$ is a morphism that assigns, to each element x, the delta distribution $\delta_{(x,x)}$ of the pair $(x, x) \in X \otimes X$. Discarding, $\mathrm{del}_X\colon X \to I$, is the stochastic map that corresponds to marginalization over X, where I denotes a fixed single-element set. Thus to each element of X, the stochastic map del_X assigns the unique probability measure on I. Diagrammatically, we represent these maps as

$$
\mathrm{copy}_X \;=\;
\begin{array}{c}
X \qquad X \\
\diagup \quad \diagdown \\
\bullet \\
X
\end{array}
\qquad\qquad
\mathrm{del}_X \;=\;
\begin{array}{c}
\bullet \\
| \\
X
\end{array}
$$
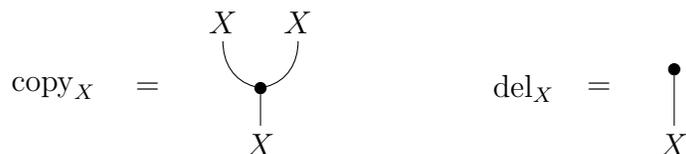
where we implicitly make use of $X \otimes I \cong X \cong I \otimes X$, so that I need not be drawn explicitly.

8

Abstracting these properties along with corresponding compatibility requirements leads to the notion of Markov category. For a more detailed discussion of the definition and its formal aspects, we refer the reader to [17].

Definition 3.1 (Markov category [17, Definition 2.1])**.** *A Markov category* C *is a symmetric monoidal category where the monoidal unit object I is a terminal object,[3] every object $X \in$ C *is equipped with distinguished morphisms* $\mathrm{copy}_X \colon X \to X \otimes X$ *and* $\mathrm{del}_X \colon X \to I$ *that make X into a commutative comonoid, and such that*

$$\tag{4}$$

holds for all $X, Y \in$ C.

Example 3.2. As far as this paper is concerned, the most relevant example is that of BorelStoch, which is the category of standard Borel spaces and measurable Markov kernels. It extends the objects of FinStoch by including both countably infinite measurable spaces as well as those isomorphic to $[0, 1]$ with its Borel σ-algebra. Morphisms coincide with those of FinStoch on finite sets, but in general they are given by Markov kernels. That is, a morphism $f \colon A \to X$ is specified by a family of probability measures $f(\_|a)$ over X, one for each $a \in A$, such that $f(S|\_) \colon A \to [0, 1]$ is a measurable map for every measurable subset $S \subseteq X$. Sequential composition of Markov kernels $f \colon A \to X$ and $h \colon X \to Z$ is given by the Chapman-Kolmogorov equation as usual, which in our notation reads

$$(h \circ f)(T|a) = \int_X h(T|x) \, f(dx|a) \tag{5}$$

for every $a \in A$ and every measurable $T \subseteq Z$. For more details, see [17, Section 4].

3.1 Conditionals

In the introduction, we mentioned that one of the ingredients in our synthetic proof of de Finetti's Theorem is the existence of conditional probability distributions. Let us make this more precise in the context of Markov categories. In FinStoch, given

---

[3]This means that for every object X in C there is a *unique* morphism $X \to I$.

a stochastic matrix $f\colon A \to X \otimes Y$ with entries $f(x, y|a)$, there is also a stochastic matrix $f_{|X}(y|x, a)$ satisfying

$$f_{|X}(y|x, a) = \frac{f(x, y|a)}{\sum_{y'} f(x, y'|a)} \tag{6}$$

whenever the denominator $\sum_{y'} f(x, y'|a)$ is non-zero, and taking arbitrary values otherwise. This gives the corresponding probability of $Y = y$ given that A and X attain values a and x, respectively. Equation (6) can be viewed as a version of Bayes' Theorem. One can also characterize $f_{|X}$ implicitly by

$$f(x, y|a) = f_{|X}(y|x, a) \sum_{y'} f(x, y'|a) \tag{7}$$

which, unlike equation (6), can be expressed in string diagrams.

Definition 3.3 (conditionals [17, Definition 11.5]). *Let $f\colon A \to X \otimes Y$ be a morphism in a Markov category* C. *A morphism $f_{|X}\colon X \otimes A \to Y$ in* C *is called a* conditional *of f with respect to X if the equation*

$$\tag{8}$$

holds. We say that C *has conditionals if such a conditional exists for all morphisms $f\colon A \to X \otimes Y$ for any objects A, X, Y in* C.

In BorelStoch, this amounts to the existence of regular conditional probabilities for measurable Markov kernels [18, Example 2.4].

As a special case of conditionals, we obtain a synthetic definition of a Bayesian inverse of $f\colon A \to X$ with respect to a prior $m\colon I \to A$.

Definition 3.4 (Bayesian inverse [7]). *Given morphisms $m\colon I \to A$ and $f\colon A \to X$, a* Bayesian inverse *of f with respect to m is a conditional of*

$$\tag{9}$$

10

with respect to X. Explicitly, it is a morphism $f^\dagger : X \to A$ satisfying

$$
\begin{array}{ccc}
\vcenter{\hbox{\includegraphics{fig1}}} & = & \vcenter{\hbox{\includegraphics{fig2}}}
\end{array}
\tag{10}
$$

For example in FinStoch, such a Bayesian inverse $f^\dagger \colon X \to A$ satisfies

$$
f^\dagger(a|x) \sum_{a'} f(x|a')\, m(a') = f(x|a)\, m(a)
\tag{11}
$$

In general, one should keep in mind that even though we denote a Bayesian inverse of f by $f^\dagger$, it does depend non-trivially on the prior m. Moreover, conditionals and Bayesian inverses are generally not unique when they exist. In FinStoch, this is because $f_{|X}(y|x, a)$ is arbitrary whenever $\sum_y f(x, y|a) = 0$ in equation (7). However, one can show that conditionals (and therefore also Bayesian inverses) *are* unique up to almost sure equality [17, Proposition 13.6].

Definition 3.5 (a.s.-equality [7]). *Given $m \colon \Theta \to A$, we say that $f, g \colon A \to X$ are m-almost surely equal, denoted by $f =_{m\text{-a.s.}} g$, if we have*

$$
\begin{array}{ccc}
\vcenter{\hbox{\includegraphics{fig3}}} & = & \vcenter{\hbox{\includegraphics{fig4}}}
\end{array}
\tag{12}
$$

We can interpret equation (12) as saying that f and g can only differ for events that are deemed irrelevant by m. For example, in BorelStoch, f and g are m-a.s. equal if and only if they are equal with probability 1 for every value of Θ, i.e. if and only if

$$
\int_S f(T|a)\, m(da|\theta) = \int_S g(T|a)\, m(da|\theta)
\tag{13}
$$

holds for all $\theta \in \Theta$ and for all measurable subsets S of A and T of X respectively.

3.2 Representability

Second on our list of proof ingredients is the ability to express a space of distributions on an object in a Markov category C as an object in C itself. For example, given an object $X = \{0, 1\}$ in BorelStoch, we would like there be an object PX isomorphic to $[0, 1]$ whose elements are themselves probability distributions over X. Indeed, BorelStoch allows for such a construction [18, Example 3.19]. However, this is not the case for FinStoch of course since PX cannot be a finite set.

Before stating the general definition of PX more formally, we need to address the question of how to refer to "elements" of an object in a Markov category. After all, the objects do not come equipped with any underlying set a priori. In FinStoch, we can identify the finite set X with those morphisms $I \to X$ that are delta distributions. That is, they are morphisms describing no randomness—the deterministic ones. More generally, deterministic morphisms in FinStoch are the $\{0, 1\}$-valued stochastic matrices. Intuitively, a deterministic f can be characterized by the fact that applying it to two independent copies of its input is guaranteed to result in the same pair of output values as applying f directly to the input and copying its output.

Definition 3.6 (deterministic morphism [17, Definition 10.1]). *Let $f\colon A \to X$ be a morphism in C. We say that f is* deterministic *if it satisfies:*

$$
\begin{array}{ccc}
\begin{gathered}
\overset{X}{|} \quad \overset{X}{|} \\
\boxed{f} \quad \boxed{f} \\
\bullet \\
| \\
A
\end{gathered}
& = &
\begin{gathered}
\overset{X}{} \quad \overset{X}{} \\
\bullet \\
\boxed{f} \\
| \\
A
\end{gathered}
\end{array}
\tag{14}
$$

The subcategory of C that consists of its deterministic morphisms is denoted by C_{det}.

Indeed, condition (14) fails for every stochastic matrix that is not $\{0, 1\}$-valued. Besides deterministic morphisms, we will also make use of the concept of m-almost surely deterministic ones.

Definition 3.7 (a.s.-deterministic morphism [17, Definition 13.11]). *A morphism f*

in C *is* m-almost surely deterministic *if it satisfies:*

$$\text{(diagram)} \quad = \quad \text{(diagram)} \tag{15}$$

Given Definition 3.6, we thus identify "elements" of the hypothetical space of distributions PX with deterministic morphisms $I \to PX$. In order for these to faithfully represent probability distributions on X (and nothing else), in the general categorical setting we therefore require there to be a bijection between $\mathsf{C}(I, X)$ and $\mathsf{C}_{\text{det}}(I, PX)$. Extending this requirement to morphisms with arbitrary domain A, which is expected to hold by the same reasoning, leads to the definition of distribution objects.

Definition 3.8 (representable Markov category [18, Definition 3.7]). *Given an object X in a Markov category* C, *a* distribution object *for X is an object PX together with natural bijections*[4]

$$\mathsf{C}(A, X) \cong \mathsf{C}_{\text{det}}(A, PX) \tag{16}$$

between morphisms into X and deterministic morphisms into PX. Given any morphism $f \colon A \to X$, we denote by $f^{\sharp} \colon A \to PX$ its counterpart under this bijective correspondence. We say that C *is* representable *if every object of* C *has a distribution object.*

It turns out that a slightly stronger version of representability is needed in the synthetic proof of de Finetti's Theorem. In particular, we require that the identification from (16) is compatible with a.s.-equality in the following sense.

Definition 3.9 (a.s.-compatible representability [18, Definition 3.18]). *A representable Markov category* C *is called* a.s.-compatibly representable *if we have*

$$\boxed{f} \;=_{m\text{-}a.s.}\; \boxed{g} \qquad \Longleftrightarrow \qquad \boxed{f^{\sharp}} \;=_{m\text{-}a.s.}\; \boxed{g^{\sharp}} \tag{17}$$

for all $m \colon \Theta \to A$ and all f, g as indicated.

---

[4]The bijections being *natural* refers to the property that the functions instantiating them for different choices of A make up the components of a natural isomorphism.

For example, BorelStoch is a.s.-compatibly representable [18, Example 3.19].

If we set $A = PX$ in bijection (16), we get a correspondence between deterministic morphisms $PX \to PX$ and generic morphisms $PX \to X$. The identity on PX corresponds to a map $PX \to X$ which we denote by samp, and which we can think of as taking a probability distribution p on X and returning a random element of X distributed according to p (hence, "sampling" from p). In BorelStoch, for instance, for every $p \in PX$ and every measurable $S \subseteq X$ we have

$$\mathsf{samp}(S|p) = p(S). \tag{18}$$

3.3 Infinite Products

The third and last ingredient that we need in the proof of our synthetic de Finetti's Theorem is the ability to construct countable products of objects. Once again we would not expect FinStoch to allow those, since its objects are merely finite sets, but BorelStoch does [19, Example 3.6]. The relevant definition is that of *Kolmogorov products* introduced in [19] in the context of 0/1-laws for Markov categories.

Given a hypothetical object $X^{\mathbb{N}}$ describing the product of countably many X, every probability measure $f \colon I \to X^{\mathbb{N}}$ should give rise to a corresponding measure $f_F \colon I \to X^F$ for each finite subset $F \subseteq \mathbb{N}$. Intuitively, f_F is given by marginalizing over those copies of X indexed by the set complement of F. Therefore, we might expect there to be a corresponding deterministic morphism $\pi_F \colon X^{\mathbb{N}} \to X^F$, which gives f_F when composed with f. Conversely, Kolmogorov extension theorem suggests that one can reconstruct f uniquely from its family of finite marginals (f_F), provided that these are suitably compatible. That is to say, if $F' \subseteq F$ is a further subset, then one can obtain $f_{F'}$ from f_F by marginalizing over the extraneous copies of X. Under this condition, there should be a unique f such that $f_F = \pi_F \circ f$ for all F.

The idea underlying Kolmogorov products is to turn these properties that one would expect from a countable product into a definition of a meaningful product of infinitely many objects in any Markov category. However, besides merely morphisms of type $I \to X^{\mathbb{N}}$, we also require similar properties of those with a non-trivial domain A and an additional codomain Y that could be correlated with $X^{\mathbb{N}}$.

Definition 3.10 (Kolmogorov powers [19, Definition 4.1])**.** *Given an object X in a Markov category C, consider an object $X^{\mathbb{N}}$ of C such that there is a natural bijection between*

- *morphisms f in $\mathsf{C}(A, X^{\mathbb{N}} \otimes Y)$ and*

14

- *families $(f_F\colon A \to X^F \otimes Y)$ of morphisms indexed by finite subsets F of $\mathbb{N}$ that are compatible in the sense that whenever F' is a subset of F, we have*

$$
\begin{array}{ccc}
\begin{array}{c}
X^{F'} \quad Y \\
\boxed{f_{F'}} \\
A
\end{array}
& = &
\begin{array}{c}
X^{F'} \quad Y \\
\boxed{\pi_{F,F'}} \\
\boxed{f_F} \\
A
\end{array}
\end{array}
\tag{19}
$$

  where $\pi_{F,F'}\colon X^F \to X^{F'}$ acts as the identity on $X^{F'}$ and applies del_X to each of the remaining factors,

for all objects A and Y of C. If the family of morphisms corresponding to the identity on $X^{\mathbb{N}}$ consists of deterministic morphisms, then $X^{\mathbb{N}}$ is termed countable Kolmogorov power *of X.*

For example, Kolmogorov's extension theorem (in the countable case) states that countable Kolmogorov powers exist in $\mathsf{BorelStoch}$ [19, Example 3.6].

Note that the family of morphisms that corresponds to $\mathrm{id}\colon X^{\mathbb{N}} \to X^{\mathbb{N}}$ under the prescribed bijection is given by the "infinite marginalization maps" $\pi_F\colon X^{\mathbb{N}} \to X^F$ mentioned before, which play the role of product projections.

Although the original definition in [19] defines Kolmogorov products of any arbitrarily large family of objects, in the present context we limit ourselves to countable powers of one and the same object, since this is all we need for de Finetti's Theorem.

4 De Finetti's Theorem for Markov Categories

Suppose that a Markov category C has countable Kolmogorov powers. We can then express what it means for a morphism with a countable power $X^{\mathbb{N}}$ as codomain to be exchangeable. Note that for any injective function $i\colon \mathbb{N} \to \mathbb{N}$ (and in particular for any permutation) we can define a morphism $X^i\colon X^{\mathbb{N}} \to X^{\mathbb{N}}$ that maps the n-th component in $X^{\mathbb{N}}$ to

- the $i^{-1}(n)$-th component of $X^{\mathbb{N}}$ whenever n is in the image of i, and
- discards it otherwise.

The construction of X^i is by the definition of $X^{\mathbb{N}}$ as the Kolmogorov power of X. That is, X^i corresponds to the family of morphisms $(X_F^i\colon X^{\mathbb{N}} \to X^F)$ as in Definition 3.10, where X_F^i is the composite of the product projection $\pi_{i(F)}\colon X^{\mathbb{N}} \to X^{i(F)}$ with the canonical isomorphism $X^{i(F)} \cong X^F$ resulting from identifying the factors as prescribed by i.

Definition 4.1 (exchangeability). *A morphism $p\colon A \to X^{\mathbb{N}}$ is said to be* exchangeable *if it is invariant under finite permutations of the components of the Kolmogorov power $X^{\mathbb{N}}$. That is, we demand that for every finite permutation $\sigma\colon \mathbb{N} \to \mathbb{N}$, we have*

$$
\boxed{p} \quad = \quad \frac{\boxed{X^{\sigma}}}{\boxed{p}}
\tag{20}
$$

Assumption 4.2. *Unless stated otherwise, throughout the rest of this paper we assume that C is a Markov category that:*

1. *has conditionals,*
2. *is a.s.-compatibly representable, and*
3. *has countable Kolmogorov powers.*

Example 4.3. $\mathsf{BorelStoch}$ satisfies these assumptions. The relevant arguments can be found in [27, Theorem 1.25], [18, Example 3.19], and [19, Example 3.6] respectively.

Theorem 4.4 (synthetic de Finetti's Theorem). *Let C be a Markov category satisfying Assumption 4.2. Then a morphism $p\colon A \to X^{\mathbb{N}}$ in C is exchangeable if and only if there is a morphism $\mu\colon A \to PX$ such that we have*

$$
\boxed{p} \quad = \quad \frac{\boxed{\text{samp}} \;\cdots\; \boxed{\text{samp}}}{\boxed{\mu}}
\tag{21}
$$

In particular, this result says that the outputs of p are conditionally independent and identically distributed. As we will see in the proof, the conditioning here is with respect to a synthetic version of the tail σ-algebra, thereby reproducing this classical aspect of de Finetti's Theorem as well. In fact, the proof uses an abstract version of the well-known result of measure-theoretic probability that the exchangeable σ-algebra and the tail σ-algebra coincide up to sets of measure zero [29, Corollary 12.18]. In the following synthetic version of this result, the role of these σ-algebras is played by all possible ways of "probing" joint distributions of $X^{\mathbb{N}}$ by composing with a deterministic (or almost surely deterministic) morphism f.

Proposition 4.5. *Let C be a Markov category with conditionals and countable Kolmogorov powers. Let $p\colon A \to X^{\mathbb{N}}$ be exchangeable, and suppose that $f\colon X^{\mathbb{N}} \to Y$ is p-a.s. deterministic. Then the following are equivalent:*

1. *Finite permutation invariance: For every finite permutation σ, we have*

$$f =_{p\text{-a.s.}} f \circ X^\sigma. \tag{22}$$

2. *Shift invariance: For the successor function $s\colon \mathbb{N} \to \mathbb{N}$ mapping n to $n+1$, we have*

$$f =_{p\text{-a.s.}} f \circ X^s. \tag{23}$$

The successor function can be equivalently the inclusion of the second component in the coproduct decomposition $\mathbb{N} \cong 1+\mathbb{N}$ and we depict its action on the Kolmogorov power $X^{\mathbb{N}}$ by either of the two following string diagrams:

This morphism effectively discards the first component of $X^{\mathbb{N}}$.

Remark 4.6. While Theorem 4.4 captures the key part of de Finetti's Theorem characterizing exchangeable morphisms, we have not been able to prove that our assumptions imply the uniqueness of the "de Finetti measure" μ in equation (21). This property *is* often proven as part of classical versions of de Finetti's Theorem including the one of Hewitt and Savage for compact Hausdorff spaces [21].

Here is how we can use Theorem 4.4 to obtain the classical de Finetti's Theorem in the form of Theorem 2.1. First of all, we take as C the category BorelStoch, which satisfies the relevant assumptions. Moreover, since the usual statement is given for the case of probability measures as opposed to Markov kernels with nontrivial domain, it suffices to consider the case of $A = I$.

Then the left-hand side of equation (21) instantiated in our case is an exchangeable probability measure on $X^{\mathbb{N}}$ in the sense of Section 2. Theorem 4.4 says that there exists a Markov kernel $\mu\colon I \to PX$ (equivalently, a probability measure on PX) such that p can be written as in the right-hand side of equation (21). Evaluating this equation on a cylinder set corresponding to a sequence of measurable subsets $S_1, \ldots, S_n$ of X gives

$$p(S_1 \times \cdots \times S_n) = \int_{PX} \left(\prod_{i=1}^n \int_{PX} \mathsf{samp}(S_i|q_i) \right) \mathsf{copy}(dq_1 \otimes \cdots \otimes dq_n|q)\, \mu(dq)$$

$$= \int_{PX} \prod_{i=1}^n \mathsf{samp}(S_i|q)\, \mu(dq)$$

$$= \int_{PX} q(S_1) \times \cdots \times q(S_n)\, \mu(dq),$$

17

where we used the fact that $\mathsf{samp}(S|q) = q(S)$. Thus, Theorem 2.1 follows from Theorem 4.4.

Similarly, here is how Proposition 4.5 recovers the classical coincidence of exchangeable σ-algebra and tail σ-algebra up to null sets. In $\mathsf{BorelStoch}$, a deterministic morphism $f\colon X^{\mathbb{N}} \to \{0,1\}$ corresponds to an event in $X^{\mathbb{N}}$, and such f satisfies the shift invariance condition if and only if the corresponding event is, up to null sets, in the tail σ-algebra. Likewise, such f satisfies finite permutation invariance if and only if the event is, up to null sets, in the exchangeable σ-algebra.

Remark 4.7. Note that by making the same argument with A an arbitrary standard Borel space, one obtains a parametric version of the classical Theorem 2.1. It says that if the exchangeable probability measure p depends measurably on a parameter, then μ can also be chosen to depend measurably on the same parameter.

Remark 4.8. It is conceivable that versions of de Finetti's Theorem applying to larger classes of spaces, such as the result of Hewitt and Savage for exchangeable Radon probability measures on compact Hausdorff spaces, can also be obtained from Theorem 4.4 by instantiating it in a suitable Markov category. The recent results of Forré [16] may be relevant here to establish the existence of conditionals.

5 Diagrammatic Proof of de Finetti's Theorem

The overall structure of the proof is presented in Figure 1. We first present a number of lemmas that may be useful in other contexts too. We assume throughout that we are in a Markov category C satisfying Assumption 4.2.

Lemma 5.1 (spreadability). *Let $p\colon A \to Y \otimes X^{\mathbb{N}}$ be exchangeable in the second factor. Then for every injective function $i\colon \mathbb{N} \to \mathbb{N}$, we also have*

$$\vcenter{\hbox{\includegraphics{diagram}}} \tag{24}$$

In the concrete setting of $\mathsf{BorelStoch}$, this statement reads as follows in the case $A = Y = I$. Consider a probability measure $p \in P(X^{\mathbb{N}})$ specifying an exchangeable joint distribution of infinitely many random variables. If we marginalize over the first component, then the resulting distribution is again p. In other words, if p is exchangeable, then it is also shift-invariant. More generally, even if we can marginalize
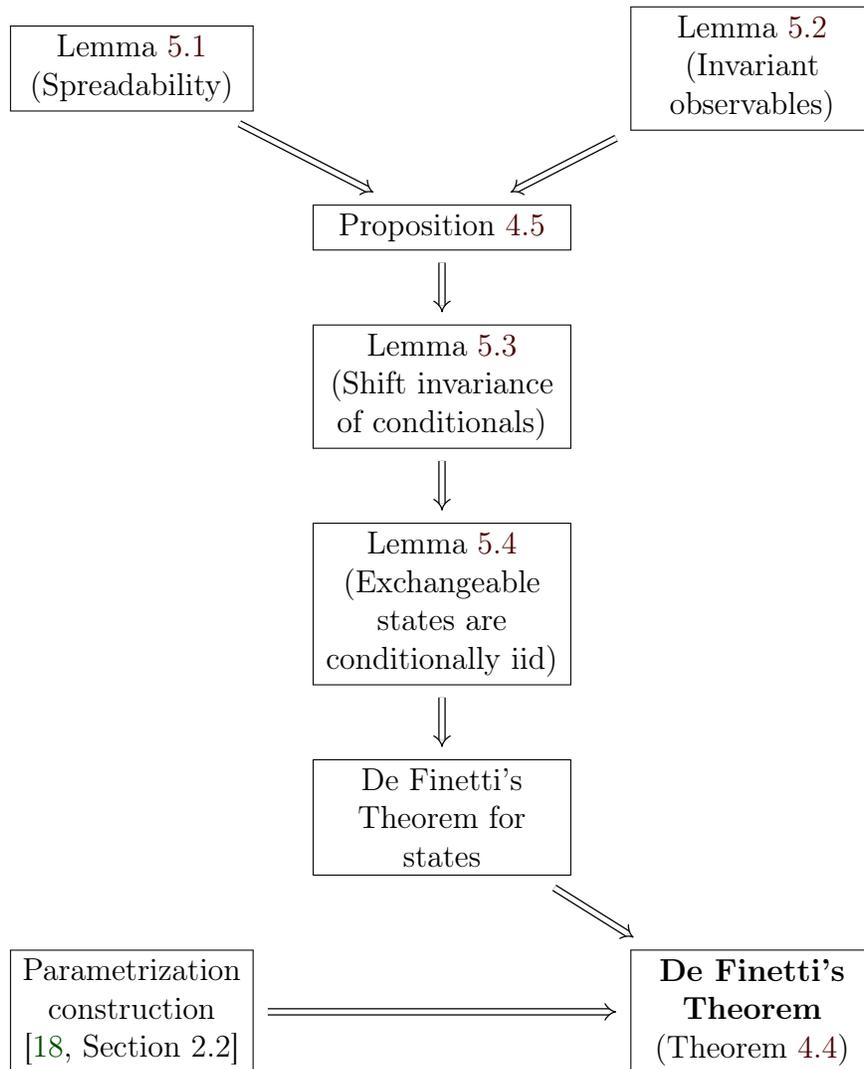
Figure 1: Overall structure of the proof of Theorem 4.4.

over any number of components whose indices are specified by the image of some injective function $i\colon \mathbb{N} \to \mathbb{N}$, the resulting distribution is still p. The same logic can be applied in the presence of a non-trivial covariate Y and parameter A, i.e. when p is a Markov kernel $A \to Y \otimes X^{\mathbb{N}}$.

The relevant argument was essentially already given in the proof of [19, Section 5], but we reproduce the argument here in the present more general context. Note that this works the same way for any other infinite set in place of $\mathbb{N}$ (if the relevant Kolmogorov power exists).

Proof. Since $X^{\mathbb{N}}$ is the Kolmogorov power of X, it suffices to prove equation (24) after composition with each of the infinite marginalization maps $\pi_F\colon X^{\mathbb{N}} \to X^F$ where F is a finite subset of $\mathbb{N}$ as before. That is, we need to prove

$$
\vcenter{\hbox{$\begin{array}{c}\boxed{\pi_F}\\ \boxed{X^i}\\ \boxed{p}\end{array}$}} \quad = \quad \vcenter{\hbox{$\begin{array}{c}\boxed{\pi_F}\\ \boxed{p}\end{array}$}} \tag{25}
$$

for each finite $F \subseteq \mathbb{N}$. For any given i and F, we can find a finite permutation $\sigma\colon \mathbb{N} \to \mathbb{N}$ whose action on F coincides with that of i. By definitions of X^σ and X^i, this implies $\pi_F \circ X^\sigma = \pi_F \circ X^i$, and the claim now follows by the assumed finite permutation invariance. $\qquad\square$

The next lemma can be thought of as a synthetic statement about stochastic dynamical systems where $t\colon \Theta \to \Theta$ is the dynamics, p is an invariant probability measure, and f is an observable.

Lemma 5.2 (dynamically invariant observables). *Let $p\colon I \to \Theta$ and $t\colon \Theta \to \Theta$ be two morphisms satisfying $t\,p = p$. If $t^\dagger$ denotes a Bayesian inverse of t with respect to p, then every p-a.s. deterministic $f\colon \Theta \to X$ satisfies*

$$
\vcenter{\hbox{$\begin{array}{c}\boxed{f}\\ \boxed{t^\dagger}\end{array}$}} =_{p\text{-}a.s.} \boxed{f} \quad \implies \quad \vcenter{\hbox{$\begin{array}{c}\boxed{f}\\ \boxed{t}\end{array}$}} =_{p\text{-}a.s.} \boxed{f} \tag{26}
$$

Before presenting the proof, let us again instantiate this for the concrete case of BorelStoch, and in particular for $X = \mathbb{R}$. Consider a standard Borel probability

20

space (Θ, p) and a measure-preserving Markov kernel $t \colon \Theta \to \Theta$. A Bayesian inverse of t is regular conditional kernel $t^\dagger$ satisfying

$$\int_S t(T|\theta) \, p(d\theta) = \int_T t^\dagger(S|\theta) \, p(d\theta) \tag{27}$$

for all measurable subsets $S, T \subseteq \Theta$. Lemma 5.2 says that if a (deterministic) function $f \colon \Theta \to \mathbb{R}$ satisfies

$$f(\theta) = \int_\Theta f(\theta') \, t^\dagger(d\theta'|\theta), \tag{28}$$

for p-almost every θ, then it also satisfies

$$f(\theta) = \int_\Theta f(\theta') \, t(d\theta'|\theta). \tag{29}$$

for p-almost every θ. By Lemma 5.2, if f is an observable invariant under the time-reversed dynamics, then it is also preserved forward in time. One may therefore expect this result to have further relevance for a synthetic treatment of ergodic theory.

Proof. Using the definition of Bayesian inverses, the assumed $t \, p = p$, and the antecedent statement in implication (26), we get

$$\tag{30}$$

The claim now follows by [19, Lemma 5.5] whose proof, even if originally stated for deterministic f only, works just as well for p-a.s. deterministic f. $\qquad\square$

Getting to the proof of Proposition 4.5, it is worth noting that neither Lemma 5.1 nor Lemma 5.2 use the assumption of representability of C. As expressed in the statement of Proposition 4.5, the following proof does not require representability either, even while making use of the previous two lemmas.

Proof of Proposition 4.5. Let us argue that shift invariance as expressed by equation (23) implies finite permutation invariance as expressed by equation (22). Indeed, for any finite permutation $\sigma\colon \mathbb{N} \to \mathbb{N}$, there is an integer n_σ beyond which all larger integers are fixed by σ. With $s\colon \mathbb{N} \to \mathbb{N}$ denoting the successor function introduced in the statement of Proposition 4.5, this entails $\sigma \circ s^{n_\sigma} = s^{n_\sigma}$. Therefore, we have

$$(X^s)^{n_\sigma} \circ X^\sigma = (X^s)^{n_\sigma}, \tag{31}$$

where the morphism on the right-hand side effectively discards the first n_σ components of $X^{\mathbb{N}}$. Thus, assuming shift invariance of f, we can infer the permutation invariance as follows:

$$\tag{32}$$

In particular, the first equality holds by exchangeability of p, the second and fourth ones correspond to shift invariance of f applied n_σ times, and the third follows from equation (31) (as well as exchangeability of p once again).

Conversely, let us now assume finite permutation invariance of f and show that it implies shift invariance. By the first equality of (32), we have that the morphism
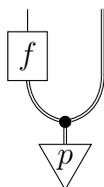
is exchangeable in the second output. Therefore, by Lemma 5.1, we also obtain

$$
\begin{array}{ccc}
\includegraphics{eq33a} & = & \includegraphics{eq33b} & = & \includegraphics{eq33c}
\end{array}
\tag{33}
$$

By Lemma 5.2, this entails

$$
f \quad =_{p\text{-a.s.}} \quad \begin{array}{c} f \\ X^s \end{array}
\tag{34}
$$

which is what we wanted to prove. $\qquad\square$

For a given exchangeable morphism $p\colon I \to X^{\mathbb{N}}$, we now consider a conditional $p_{|\mathrm{tail}}\colon X^{\mathbb{N}} \to X$ of the first output given all others, which means that it satisfies the left equation in

$$
\begin{array}{ccc}
\includegraphics{eq35a} & = & \includegraphics{eq35b} & = & \includegraphics{eq35c}
\end{array}
\tag{35}
$$

while the right one follows by exchangeability of p and Lemma 5.1.

Such a tail conditional depends only on the tail of the product $X^{\mathbb{N}}$ in the following sense.

Lemma 5.3 (shift invariance of the tail conditional). *If $p\colon I \to X^{\mathbb{N}}$ is exchangeable, then we have*

$$
p_{|\mathrm{tail}} \quad =_{p\text{-}a.s.} \quad \bullet\ \begin{array}{c} p_{|\mathrm{tail}} \end{array}
\tag{36}
$$

Here is what the statement looks like in BorelStoch. Let p be an exchangeable probability measure on $X^{\mathbb{N}}$, and let $p_{|\text{tail}}$ be a regular conditional of the first component given the other ones. Denote a generic element of $X^{\mathbb{N}}$ by $(x_1, x_2, \dots)$. Then for each measurable subset $S \subseteq X$, we have that

$$p_{|\text{tail}}(S|x_1, x_2, \dots) = p_{|\text{tail}}(S|x_2, x_3, \dots) \tag{37}$$

holds for p-almost all sequences $(x_1, x_2, \dots) \in X^{\mathbb{N}}$. Iterating this equation shows that $p_{|\text{tail}}$ only depends on the tail.

Proof. By the definition of $p_{|\text{tail}}$ and exchangeability of p, we have

$$\tag{38}$$

for any finite permutation $\sigma\colon \mathbb{N} \to \mathbb{N}$. By the a.s.-compatible representability of C (Definition 3.9), this is equivalent to

$$\left(p_{|\text{tail}} \circ X^{\sigma}\right)^{\sharp} =_{p\text{-a.s.}} p^{\sharp}_{|\text{tail}}. \tag{39}$$

Since X^{σ} is deterministic, we have

$$\left(p_{|\text{tail}} \circ X^{\sigma}\right)^{\sharp} = p^{\sharp}_{|\text{tail}} \circ X^{\sigma}, \tag{40}$$

and conclude that the deterministic morphism $p^{\sharp}_{|\text{tail}}$ is p-almost surely finite permutation invariant. Thus, by Proposition 4.5, it is also shift invariant. This proves the claim after composition with the sampling map samp_X. $\qquad\square$

Lemma 5.4 (exchangeable states are conditionally iid)**.** *If $p\colon I \to X^{\mathbb{N}}$ is an exchangeable morphism, then we have*

$$\tag{41}$$

24

Concretely, in BorelStoch, Lemma 5.4 says the following. Let p be an exchangeable probability measure on $X^{\mathbb{N}}$. Then for every cylinder defined by measurable subsets $S_1, \ldots, S_N \subseteq X$, we have

$$p(S_1 \times \cdots \times S_n) = \int_{X^{\mathbb{N}}} p_{|\text{tail}}(S_1|\xi) \times \cdots \times p_{|\text{tail}}(S_n|\xi)\, p(d\xi). \tag{42}$$

where ξ is shorthand for a generic sequence $(x_1, x_2, \ldots) \in X^{\mathbb{N}}$.

Proof. By the universal property of the Kolmogorov power $X^{\mathbb{N}}$, it is enough to prove

$$\tag{43}$$

for every $n \in \mathbb{N}$. Using induction, the base case $n = 0$ is trivial. In order to get from n to $n + 1$, we use
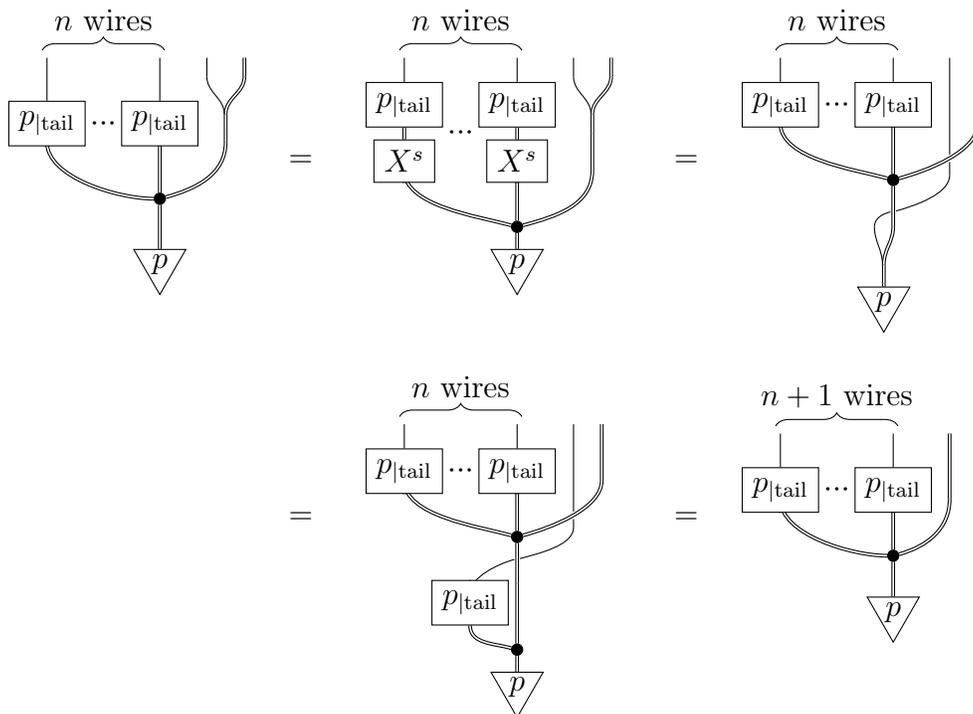
where the first step is by Lemma 5.3 and the third one by equation (35). $\qquad\square$

25

Parametrization construction. Lemma 5.4 allows us to prove de Finetti's Theorem for exchangeable morphisms out of I, but our Theorem 4.4 is meant to be applicable to exchangeable morphisms with an arbitrary domain A. In order to make this logical transition, we interpret the statement of Lemma 5.4 in a new Markov category C_A where every morphism $f\colon A \to Y$ of C can be reinterpreted as a morphism $f\colon I \to Y$ in C_A. Specifically, we use the *parametric Markov category* C_A [18, Section 2.2], which is defined in terms of C and A by taking its objects to coincide with those of C and its morphisms to be

$$\mathsf{C}_A(B,Y) := \mathsf{C}(A \otimes B, Y). \tag{44}$$

In a sense, C_A contains the same data as C; it is just organized differently. When composing morphisms in C_A we use the composition of C and distribute a copy of A to each of the morphisms in the composition. The key aspect of the parametrization construction for our proof is that as long as C satisfies Assumption 4.2, so does the parametric category C_A. The arguments for conditionals and a.s.-compatible representability have been presented in [18] as Lemma 2.10 and Lemma 3.22 respectively. Here, we also spell out the argument for Kolmogorov powers.

Lemma 5.5. *If $X^{\mathbb{N}}$ is a countable Kolmogorov power of X in C, then $X^{\mathbb{N}}$ is also a countable Kolmogorov power of X in the parametric Markov category C_A.*

Proof. Let B, Y be arbitrary objects of C_A. We have the following three bijective correspondences

$$f \in \mathsf{C}_A(B, X^{\mathbb{N}} \otimes Y) \longleftrightarrow f \in \mathsf{C}(A \otimes B, X^{\mathbb{N}} \otimes Y)$$

$$\updownarrow$$

$$\big(f_F \in \mathsf{C}_A(B, X^F \otimes Y)\big) \longleftrightarrow \big(f_F \in \mathsf{C}(A \otimes B, X^F \otimes Y)\big)$$

where the lower two are compatible families in the same sense as in Definition 3.10. The vertical one is natural in both B and X by assumption, while for the horizontal ones, naturality follows from the definition of C_A.

Since identities in C_A correspond to identities in C with A discarded, the infinite marginalization maps in C_A are likewise given by discarding A and applying the respective $\pi_F \in \mathsf{C}(X^{\mathbb{N}}, X^F)$. In particular, the infinite marginalization maps in C_A are deterministic, which means that the countable Kolmogorov power of X in C_A indeed exists and is given by $X^{\mathbb{N}}$. $\qquad\square$

Proof of Theorem 4.4. We first consider the case $A = I$. Then we get

$$
\begin{array}{c}
\boxed{p} \\
\end{array}
\quad = \quad
\begin{array}{c}
\boxed{\text{samp}} \quad \cdots \quad \boxed{\text{samp}} \\
\boxed{p^\sharp_{|\text{tail}}} \qquad \boxed{p^\sharp_{|\text{tail}}} \\
\boxed{p}
\end{array}
\quad = \quad
\begin{array}{c}
\boxed{\text{samp}} \quad \cdots \quad \boxed{\text{samp}} \\
\boxed{p^\sharp_{|\text{tail}}} \\
\boxed{p}
\end{array}
\tag{45}
$$

as a consequence Lemma 5.4. This already has the desired form of equation (21).

Now, for a general morphism $p\colon A \to X^{\mathbb{N}}$, we apply this result to the parametric Markov category C_A, in which p is represented by a morphism with domain I. By [18, Lemmas 2.10 and 3.22] and Lemma 5.5, C_A satisfies our three assumptions provided that C itself does. As is mentioned in [18, Example 3.17], the sampling map in C_A is represented by $\mathrm{del}_A \otimes \mathsf{samp}$ in C. Therefore, instantiating equation (45) in C_A gives

$$
\begin{array}{c}
X^{\mathbb{N}} \\
\boxed{p} \\
A
\end{array}
\quad = \quad
\begin{array}{c}
\boxed{\text{samp}} \quad \cdots \quad \boxed{\text{samp}} \\
\boxed{p^\sharp_{|\text{tail}}} \\
\boxed{p}
\end{array}
\tag{46}
$$

in C, which is the relevant form for what we wanted to show. $\qquad\square$

References

[1] Irfan Alam. Generalizing the de Finetti–Hewitt–Savage theorem. arXiv:2008.08754. ↑4

[2] Irfan Alam. A nonstandard proof of de Finetti's theorem for Bernoulli random variables. *J. Stoch. Anal.*, 1(4):Art. 15, 18, 2020. arXiv:1912.02784. ↑4

[3] David J. Aldous. Exchangeability and related topics. In *École d'été de probabilités de Saint-Flour, XIII—1983*, volume 1117 of *Lecture Notes in Math.*, pages 1–198. Springer, Berlin, 1985. https://www.stat.berkeley.edu/∼aldous/Papers/me22.pdf. ↑4, 5

[4] Tim Austin. On exchangeable random variables and the statistics of large graphs and hypergraphs. *Probability Surveys*, 5:80–145, 2008. ↑4, 5

[5] Tim Austin and Dmitry Panchenko. A hierarchical version of the de Finetti and Aldous–Hoover representations. *Probability Theory and Related Fields*, 159(3-4):809–823, 2014. ↑5

[6] Federico Camerlenghi, Antonio Lijoi, Peter Orbanz, Igor Prünster, et al. Distribution theory for hierarchical processes. *Annals of Statistics*, 47(1):67–92, 2019. carloalberto.org/wp-content/uploads/2018/11/no.517.pdf. ↑5

[7] Kenta Cho and Bart Jacobs. Disintegration and Bayesian inversion via string diagrams. *Math. Structures Comput. Sci.*, 29:938–971, 2019. arXiv:1709.00322. ↑5, 7, 11

[8] Harry Crane and Henry Towsner. Relatively exchangeable structures. *arXiv preprint arXiv:1509.06733*, 2015. ↑5

[9] Harry Crane and Henry Towsner. Relative exchangeability with equivalence relations. *Archive for Mathematical Logic*, 57(5):533–556, 2018. ↑5

[10] Bruno De Finetti. Funzione caratteristica di un fenomeno aleatorio. In *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*, pages 179–190, 1929. docplayer.it/24159-Funzione-caratteristica-di-un-fenomeno-aleatorio.html, English translation available at arXiv:1512.01229. ↑3

[11] Bruno De Finetti. La prévision: ses lois logiques, ses sources subjectives. In *Annales de l'institut Henri Poincaré*, volume 7, pages 1–68, 1937. eudml.org/doc/79004, English translation available at socsci.uci.edu/∼bskyrms/bio/readings/bruno_definetti_subjective_probability.pdf. ↑2, 3

[12] Bruno De Finetti. *Sur la condition d'equivalence partielle*, volume 739. 1938. English version, translated by P. Benacerraf and R. Jeffrey, available as "On the condition of partial exchangeability," in *Studies in Inductive Logic and Probability II*, 193–205, Berkeley, University of California Press, 1980. ↑5

[13] Persi Diaconis and David Freedman. De Finetti's theorem for Markov chains. *The Annals of Probability*, pages 115–130, 1980. projecteuclid.org/journals/annals-of-probability/volume-8/issue-1/De-Finettis-Theorem-for-Markov-Chains/10.1214/aop/1176994828.full. ↑5

[14] Persi Diaconis and David Freedman. De Finetti's generalizations of exchangeability. *Studies in inductive logic and probability*, 2:233–249, 1980. ↑5

[15] Willliam Feller. *An introduction to probability theory and its applications, vol 2*. John Wiley & Sons, 2008. ↑4

[16] Patrick Forré. Transitional conditional independence. arXiv:2104.11547. ↑18

[17] Tobias Fritz. A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics. *Adv. Math.*, 370:107239, 2020. arXiv:1908.07021. ↑4, 5, 7, 8, 9, 10, 11, 12, 13

[18] Tobias Fritz, Tomáš Gonda, Paolo Perrone, and Eigil Fjeldgren Rischel. Representable markov categories and comparison of statistical experiments in categorical probability. *arXiv preprint arXiv:2010.07416*, 2020. arxiv.org/abs/2010.07416. ↑5, 7, 10, 12, 13, 14, 16, 19, 26, 27

[19] Tobias Fritz and Eigil Fjeldgren Rischel. The zero-one laws of Kolmogorov and Hewitt–Savage in categorical probability. *Compositionality*, 2:3, 2020. compositionality-journal.org/papers/compositionality-2-3. ↑5, 7, 14, 15, 16, 20, 22

[20] Michèle Giry. A Categorical Approach to Probability Theory. In *Categorical aspects of topology and analysis*, volume 915 of *Lecture Notes in Mathematics*. 1982. ↑7

[21] Edwin Hewitt and Leonard J Savage. Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80(2):470–501, 1955. ↑3, 6, 7, 17

[22] Alan Hájek. Interpretations of Probability. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2019 edition, 2019. plato.stanford.edu/entries/probability-interpret/. ↑2

[23] Bart Jacobs and Sam Staton. De Finetti's construction as a categorical limit. In *International Workshop on Coalgebraic Methods in Computer Science*, pages 90–111, 2020. arXiv:2003.01964. ↑6

[24] Michael I Jordan. Bayesian nonparametric learning: Expressive priors for intelligent systems. *Heuristics, probability and causality: A tribute to*

Judea Pearl, 11:167–185, 2010. people.eecs.berkeley.edu/∼jordan/papers/pearl-festschrift.pdf. ↑3

[25] Paul Jung, Jiho Lee, Sam Staton, and Hongseok Yang. A generalization of hierarchical exchangeability on trees to directed acyclic graphs. *Annales Henri Lebesgue*, 4:325–368, 2021. ↑5

[26] Olav Kallenberg. *Foundations of modern probability*, volume 2. Springer, second edition, 2002. ↑4

[27] Olav Kallenberg. *Random measures, theory and applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2017. ↑16

[28] Werner Kirsch. An elementary proof of de Finetti's theorem. *Statistics & Probability Letters*, 151:84–88, 2019. arXiv:1809.00882. ↑4

[29] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course. ↑4, 17

[30] Paul Ressel. De Finetti-type theorems: an analytical approach. *Ann. Probab.*, 13(3):898–922, 1985. ↑4

[31] Jan-Willem Romeijn. Philosophy of Statistics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2017 edition, 2017. plato.stanford.edu/entries/statistics/. ↑3

[32] Patrick Suppes. Representation and invariance of scientific structures. 2002. web.stanford.edu/group/cslipublications/cslipublications/pdf/1575863332.rissbook.pdf. ↑2