

# Computable de Finetti measures

Cameron E. Freer<sup>a</sup>, Daniel M. Roy<sup>b</sup>

<sup>a</sup>*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA*

<sup>b</sup>*Computer Science and Artificial Intelligence Laboratory,  
Massachusetts Institute of Technology, Cambridge, MA, USA*

---

## Abstract

We prove a uniformly computable version of de Finetti's theorem on exchangeable sequences of real random variables. As a consequence, exchangeable stochastic processes in probabilistic functional programming languages can be automatically rewritten as procedures that do not modify non-local state. Along the way, we prove that a distribution on the unit interval is computable if and only if its moments are uniformly computable.

*Key words:* de Finetti's theorem, exchangeability, computable probability theory, probabilistic programming languages, mutation

*2010 MSC:* 03D78, 60G09, 68Q10, 03F60, 68N18

---

## 1. Introduction

The classical de Finetti theorem states that an exchangeable sequence of real random variables is a mixture of independent and identically distributed (i.i.d.) sequences of random variables. Moreover, there is an (almost surely unique) measure-valued random variable, called the *directing random measure*, conditioned on which the random sequence is i.i.d. The distribution of the directing random measure is called the *de Finetti measure* or the *mixing measure*.

This paper examines the *computable* probability theory of exchangeable sequences of real-valued random variables. We prove a uniformly computable version of de Finetti's theorem, which implies that *computable* exchangeable sequences of real random variables have *computable* de Finetti measures. The classical *proofs* do not readily effectivize; instead, we show how to directly compute the de Finetti measure (as characterized by the classical theorem) in terms of a computable representation of the distribution of the exchangeable sequence. Along the way, we prove that a distribution on  $[0, 1]^\omega$  is computable if and only if its moments are uniformly computable, which may be of independent interest.

A key step in the proof is to describe the de Finetti measure in terms of the moments of a set of random variables derived from the exchangeable sequence.

October 27, 2009

When the directing random measure is (almost surely) continuous, we can show that these moments are computable, which suffices to complete the proof of the main theorem in this case. In the general case, we give a proof inspired by a randomized algorithm which succeeds with probability one.

### 1.1. Computable Probability Theory

These results are formulated in the Turing-machine-based bit-model for computation over the reals (for a general survey, see Braverman and Cook [1]). This computational model has been explored both via the type-2 theory of effectivity (TTE) framework for computable analysis, and via effective domain-theoretic representations of measures.

Computable analysis has its origins in the study of recursive real functions, and can be seen as a way to provide “automated numerical analysis” (for a tutorial, see Brattka, Hertling, and Weihrauch [2]). Effective domain theory has its origins in the semantics of programming languages, where it continues to have many applications (for a survey, see Edalat [3]). Here we use methods from these approaches to transfer a representational result from probability theory to a setting where it can directly transform statistical objects as represented on a computer.

The computable probability measures in the bit-model coincide with those distributions from which we can generate exact samples to arbitrary precision on a computer. As such, our results have direct implications for programming languages which manipulate probability measures on real numbers via exact interfaces. In many areas of statistics and computer science, especially machine learning, one is often concerned with distributions on data structures that are higher-order or are defined using recursion. Probabilistic functional programming languages provide a convenient setting for describing and manipulating such distributions.

Exchangeable sequences play a fundamental role in both statistical models and their implementation on computers. Given a *sequential* description of an exchangeable process, in which one uses previous samples or sufficient statistics to sample the next element in the sequence, a direct implementation in these languages would need to use non-local communication (to record new samples or update sufficient statistics). This is often implemented by modifying the program’s internal state directly (i.e., using *mutation*), or via some indirect method such as a state monad. The classical de Finetti theorem implies that (for such sequences over the reals) there is an alternative description in which samples are conditionally independent (and so could be implemented without non-local communication), thereby allowing parallel implementations. But the classical result does not imply that there is a *program* which computes the sequence according to this description. Even when there is such a program, the classical theorem does

not provide a method for finding it. The computable de Finetti theorem states that such a program *does* exist. Moreover, the proof itself provides the *method* for constructing the desired program. In Section 6 we describe how an implementation of the computable de Finetti theorem uniformly transforms procedures which induce exchangeable stochastic processes into equivalent procedures which do not modify non-local state.

This transformation is of interest beyond its implications for programming language semantics. In statistics and machine learning, it is often desirable to know the representation of an exchangeable stochastic process in terms of its de Finetti measure (for several examples, see Section 6.3). Many such processes in machine learning have very complicated (though computable) distributions, and it is not always feasible to find the de Finetti representation by hand. The computable de Finetti theorem provides a method for automatically obtaining such representations.

## 2. de Finetti's Theorem

We assume familiarity with the standard measure-theoretic formulation of probability theory (see, e.g., Billingsley [4] or Kallenberg [5]). Fix a basic probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and let  $\mathcal{B}_{\mathbf{R}}$  denote the Borel sets of  $\mathbf{R}$ . Note that we will use  $\omega$  to denote the set of nonnegative integers (as in logic), rather than an element of the basic probability space  $\Omega$  (as in probability theory). By a *random measure* we mean a random element in the space of Borel measures on  $\mathbf{R}$ , i.e., a kernel from  $(\Omega, \mathcal{F})$  to  $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$ . An event  $A \in \mathcal{F}$  is said to occur *almost surely* (a.s.) if  $\mathbb{P}A = 1$ . We denote the indicator function of a set  $B$  by  $\mathbf{1}_B$ .

**Definition 2.1 (Exchangeable sequence).** Let  $X = \{X_i\}_{i \geq 1}$  be a sequence of real random variables. We say that  $X$  is *exchangeable* if, for every finite set  $\{k_1, \dots, k_j\}$  of distinct indices,  $(X_{k_1}, \dots, X_{k_j})$  is equal in distribution to  $(X_1, \dots, X_j)$ .

**Theorem 2.2 (de Finetti [6, Chap. 1.1]).** *Let  $X = \{X_i\}_{i \geq 1}$  be an exchangeable sequence of real-valued random variables. There is a random probability measure  $\nu$  on  $\mathbf{R}$  such that  $\{X_i\}_{i \geq 1}$  is conditionally i.i.d. with respect to  $\nu$ . That is,*

$$\mathbb{P}[X \in \cdot \mid \nu] = \nu^\infty \quad a.s. \tag{1}$$

Moreover,  $\nu$  is a.s. unique and given by

$$\nu(B) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{1}_B(X_i) \quad a.s., \tag{2}$$

where  $B$  ranges over  $\mathcal{B}_{\mathbf{R}}$ . □

The random measure  $\nu$  is called the *directing random measure*. Its distribution (a measure on probability measures), which we denote by  $\mu$ , is called the *de Finetti measure* or the *mixing measure*. As in Kallenberg [6, Chap. 1, Eq. 3], we may take expectations on both sides of (1) to arrive at a characterization

$$\mathbb{P}\{X \in \cdot\} = \mathbb{E}\nu^\infty = \int m^\infty \mu(dm) \quad (3)$$

of an exchangeable sequence as a mixture of i.i.d. sequences.

A Bayesian perspective suggests the following interpretation: exchangeable sequences arise from independent observations from some latent random measure. Posterior analysis follows from placing a prior distribution on  $\nu$ . For further discussion of the implications of de Finetti's theorem for the foundations of statistical inference, see Dawid [7] and Lauritzen [8].

In 1931, de Finetti [9] proved the classical result for binary exchangeable sequences, in which case the de Finetti measure is simply a mixture of Bernoulli distributions; the exchangeable sequence is equivalent to repeatedly flipping a coin whose weight is drawn from some distribution on  $[0, 1]$ . In the 1950s, Hewitt and Savage [10] and Ryll-Nardzewski [11] extended the result to arbitrary real-valued exchangeable sequences. We will refer to this more general version as the *de Finetti theorem*. Hewitt and Savage [10] provide a history of the early developments, and a discussion of some subsequent extensions can be found in Kingman [12], Diaconis and Freedman [13], and Aldous [14]. A recent book by Kallenberg [6] provides a comprehensive view of the area of probability theory that has grown out of de Finetti's theorem, stressing the role of invariance under symmetries.

### 2.1. Examples

Consider an exchangeable sequence of  $[0, 1]$ -valued random variables. In this case, the de Finetti measure is a distribution on the (Borel) measures on  $[0, 1]$ . For example, if the de Finetti measure is a Dirac measure on the uniform distribution on  $[0, 1]$  (i.e., the distribution of a random measure which is almost surely the uniform distribution), then the induced exchangeable sequence consists of independent, uniformly distributed random variables on  $[0, 1]$ .

As another example, let  $p$  be a random variable, uniformly distributed on  $[0, 1]$ , and let  $\nu := \delta_p$ . Then the de Finetti measure is the uniform distribution on Dirac measures on  $[0, 1]$ , and the corresponding exchangeable sequence is  $p, p, \dots$ , i.e., a constant sequence, marginally uniformly distributed.

As a further example, we consider the stochastic process  $\{X_i\}_{i \geq 1}$  whose finite marginals are given by

$$\mathbb{P}\{X_1 = x_1, \dots, X_n = x_n\} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + S_n)\Gamma(\beta + (n - S_n))}{\Gamma(\alpha + \beta + n)}, \quad (4)$$

4

where  $S_n := \sum_{i \leq n} x_i$ , and where  $\Gamma$  is the Gamma function and  $\alpha, \beta$  are positive real numbers. (One can verify that these marginals satisfy Kolmogorov's extension theorem [5, Theorem 6.16], and so there is a stochastic process  $\{X_i\}_{i \geq 1}$  with these finite marginals.) Clearly this process is exchangeable, as  $n$  and  $S_n$  are invariant to order. This process can also be described by a sequential scheme known as Pólya's urn [15, Chap. 11.4]. Each  $X_i$  is sampled in turn according to the conditional distribution

$$\mathbb{P}\{X_{n+1} = 1 \mid X_1 = x_1, \dots, X_n = x_n\} = \frac{\alpha + S_n}{\alpha + \beta + n}. \quad (5)$$

This is often described as a urn which starts with  $\alpha$  red balls and  $\beta$  black balls, where at each stage, a ball is drawn uniformly at random, and is returned to the urn along with an additional ball of the same color. By de Finetti's theorem, there exists a random coin weight  $\theta$  with respect to which this process is conditionally independent and  $\mathbb{P}\{X_i = 1 \mid \theta\} = \theta$  for each  $i$ . In fact,

$$\mathbb{P}[X_1 = x_1, \dots, X_n = x_n \mid \theta] = \prod_{i \leq n} \mathbb{P}[X_i = x_i \mid \theta] = \theta^{S_n} (1 - \theta)^{(n - S_n)}. \quad (6)$$

Furthermore, one can show that here  $\theta$  is  $\text{Beta}(\alpha, \beta)$ -distributed, and so the process given by the marginals (4) is called the Beta-Bernoulli process. Here the de Finetti measure is the distribution of the Bernoulli measure with random coin weight  $\theta$ .

## 2.2. The Computable de Finetti Theorem

In each of these examples, the de Finetti measure is a *computable measure*. (In Section 3, we make this and related notions precise. For an explicit computable representation of the Beta-Bernoulli process, see in Section 6.) A natural question to ask is whether computable exchangeable sequences always arise from computable de Finetti measures. In fact, computable de Finetti measures give rise to computable distributions on exchangeable sequences (see Proposition 5.1). Our main result is the converse: every computable distribution on real-valued exchangeable sequences arises from a computable de Finetti measure.

**Theorem 2.3 (Computable de Finetti).** *Let  $\chi$  be the distribution of a real-valued exchangeable sequence  $X$ , and let  $\mu$  be the distribution of its directing random measure  $\nu$ . Then  $\mu$  is uniformly computable in  $\chi$ , and  $\chi$  is uniformly computable in  $\mu$ . In particular,  $\chi$  is computable if and only if  $\mu$  is computable.*

The directing random measure is classically given a.s. by the explicit limiting expression (2). Without a computable handle on the rate of convergence, the limit is not directly computable, and so we cannot use this limit directly to compute the de Finetti measure. However, we are able to reconstruct the de Finetti measure using the moments of a set of derived random variables.

### 2.2.1. Outline of the Proof

Recall that  $\mathcal{B}_{\mathbf{R}}$  denotes the Borel sets of  $\mathbf{R}$ . Let  $\mathcal{I}_{\mathbf{R}}$  denote the set of open intervals, and let  $\mathcal{I}_{\mathbf{Q}}$  denote the set of open intervals with rational endpoints. Then  $\mathcal{I}_{\mathbf{Q}} \subsetneq \mathcal{I}_{\mathbf{R}} \subsetneq \mathcal{B}_{\mathbf{R}}$ . For  $k \geq 1$  and  $\beta \in \mathcal{B}_{\mathbf{R}}^k = \mathcal{B}_{\mathbf{R}} \times \cdots \times \mathcal{B}_{\mathbf{R}}$ , we write  $\beta(i)$  to denote the  $i$ th coordinate of  $\beta$ .

We will denote by  $\mathcal{A}_{\mathbf{R}^k}$  the algebra generated by  $\mathcal{I}_{\mathbf{R}}^k$  (i.e., finite unions of open rectangles in  $\mathbf{R}^k$ ), and denote by  $\mathcal{A}_{\mathbf{Q}^k}$  the algebra generated by  $\mathcal{I}_{\mathbf{Q}}^k$ . Note that  $\mathcal{I}_{\mathbf{Q}} \subsetneq \mathcal{A}_{\mathbf{Q}} \subsetneq \mathcal{A}_{\mathbf{R}} \subsetneq \mathcal{B}_{\mathbf{R}}$ .

Let  $X = \{X_i\}_{i \geq 1}$  be an exchangeable sequence of real random variables, with distribution  $\chi$  and directing random measure  $\nu$ . For every  $\gamma \in \mathcal{B}_{\mathbf{R}}$ , we define a  $[0, 1]$ -valued random variable  $V_\gamma := \nu\gamma$ . A classical result in probability theory [5, Lem. 1.17] implies that a Borel measure on  $\mathbf{R}$  is uniquely characterized by the mass it places on the open intervals with rational endpoints. Therefore, the distribution of the stochastic process  $\{V_\tau\}_{\tau \in \mathcal{I}_{\mathbf{Q}}}$  determines the de Finetti measure  $\mu$  (the distribution of  $\nu$ ).

**Definition 2.4 (Mixed moments).** Let  $C \subseteq \mathcal{B}_{\mathbf{R}}$ . The *mixed moments* of the variables  $\{V_\beta\}_{\beta \in C}$ , are the set of all expectations  $\mathbb{E}(\prod_{i=1}^k V_{\beta(i)})$ , for  $k \geq 1$  and  $\beta \in C^k$ .

We can now restate the consequence of de Finetti's theorem described in Eq. (3), in terms of the finite-dimensional marginals of the exchangeable sequence  $X$  and the mixed moments of  $\{V_\beta\}_{\beta \in \mathcal{B}_{\mathbf{R}}}$ .

**Corollary 2.5.**  $\mathbb{P}(\bigcap_{i=1}^k \{X_i \in \beta(i)\}) = \mathbb{E}(\prod_{i=1}^k V_{\beta(i)})$  for  $k \geq 1$  and  $\beta \in \mathcal{B}_{\mathbf{R}}^k$ .  $\square$

As we will show in Lemma 3.5, when  $\chi$  is computable, we can enumerate all rational lower bounds on quantities of the form

$$\mathbb{P}(\bigcap_{i=1}^k \{X_i \in \sigma(i)\}), \tag{7}$$

where  $\sigma \in \mathcal{A}_{\mathbf{Q}}^k$ . We can also enumerate rational upper bounds on (7), provided that  $X$  places no mass on the boundary of any  $\sigma(i)$ . In particular, if  $\nu$  is a.s. continuous (i.e., with probability one,  $\nu(\{x\}) = 0$  for every  $x \in \mathbf{R}$ ), then we can use  $\chi$  to compute the mixed moments of  $\{V_\tau\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$ .

In Section 4, we show how to computably recover a distribution from its moments. This suffices to recover the de Finetti measure when  $\nu$  is a.s. continuous, as we show in Section 5.1. In the general case, fixed point masses in  $\nu$  prevent us from computing the mixed moments. Here we use a proof inspired by a randomized algorithm which almost surely avoids the point masses and recovers the de Finetti measure. For the complete proof, see Section 5.3.

### 3. Computable Representations

Before beginning the proof of the computable de Finetti theorem, we first define computable probability measures on various spaces. These definitions follow from more general TTE notions, though we will sometimes derive simpler equivalent representations for the concrete spaces we need (such as the real numbers, Borel measures on reals, and Borel measures on Borel measures on reals). For details, see the original papers, as noted.

We assume familiarity with the standard notions of computability theory and computably enumerable (c.e.) sets (see, e.g., Rogers [16] or Soare [17]). Recall that  $r \in \mathbf{R}$  is called a c.e. real when the set of all rationals less than  $r$  is a c.e. set. Similarly,  $r$  is a co-c.e. real when the set of all rationals greater than  $r$  is c.e. A real  $r$  is a computable real when it is both a c.e. and co-c.e. real.

To represent more general spaces, we work in terms of an effectively presented topology. Suppose that  $S$  is a second-countable  $T_0$  topological space with subbasis  $\mathcal{S}$ . For every point in  $x \in S$ , define the set  $C_x := \{B \in \mathcal{S} : x \in B\}$ . Because  $S$  is  $T_0$ , we have  $C_x \neq C_y$  when  $x \neq y$ , i.e.,  $x$  is the unique point that is contained in those subbasis sets that are elements of  $C_x$  and no others. Since every point in a  $T_0$  space is determined by the subbasis in this way, it is convenient to define representations on topological spaces under the assumption that the space is  $T_0$ . In the specific cases below, we often have much more structure, which we use to simplify the representations.

We now develop these definitions more formally.

**Definition 3.1 (Computable topological space).** Let  $S$  be a second-countable  $T_0$  topological space with a countable subbasis  $\mathcal{S}$ . Let  $s : \omega \rightarrow \mathcal{S}$  be an enumeration of  $\mathcal{S}$ . We say that  $S$  is a *computable topological space (with respect to  $s$ )* when the set

$$\{(A, B) \in \mathcal{S}^2 : A = B\} \tag{8}$$

is c.e. in terms of the  $s$ -indices for  $A, B \in \mathcal{S}$ , i.e.,

$$\{\langle a, b \rangle : s(a) = s(b)\} \tag{9}$$

is a c.e. subset of  $\omega$ , where  $\langle \cdot, \cdot \rangle$  is the standard pairing function.

Computable topological spaces, as defined here, are instances of the computable  $T_0$  spaces defined in Grubba, Schröder, and Weihrauch [18, §3], when the subbasis is replaced by the basis it generates (and the enumeration is extended in a canonical way).

It is often possible to pick a subbasis  $\mathcal{S}$  (and enumeration  $s$ ) for which the elemental “observations” that one can computably observe are those of the form

$x \in B$ , where  $B \in \mathcal{S}$ . Then the set  $C_x = \{B \in \mathcal{S} : x \in B\}$  is computably enumerable (with respect to  $s$ ) when the point  $x$  is such that it is eventually noticed to be in each basic open set containing it; we will call such a point  $x$  *computable*. This is one motivation for the definition of computable point in a  $T_0$  space below.

Note that in a  $T_1$  space, two computable points are computably distinguishable, but in a  $T_0$  space, computable points will be, in general, distinguishable only in a computably enumerable fashion. However, this is essentially the best that is possible, if the open sets are the those we can “observe”. (For more details on this approach to considering datatypes as topological spaces, in which basic open sets correspond to “observations”, see Battenfeld, Schröder, and Simpson [19, §2].) Note that the choice of topology and subbasis are essential; for example, we can recover both computable reals and c.e. reals as instances of “computable point” for appropriate computable topological spaces, as we describe in Section 3.1.

**Definition 3.2 (Computable points).** Let  $(S, \mathcal{S})$  be a computable topological space with respect to an enumeration  $s$ . We say that a point  $x \in S$  is *computable (with respect to  $s$ )* when the set

$$\{B \in \mathcal{S} : x \in B\} \tag{10}$$

is c.e. (in terms of the  $s$ -indices for  $B \in \mathcal{S}$ ). We call the set (10) the *representation* of  $x$ .

Suppose that  $A$  and  $B$  are computable objects (either points, as defined here, or computable functions or measures, as defined below). We will say that  $A$  is *uniformly computable in  $B$* , or that  $A$  is *uniformly computable relative to  $B$* , when there is a single program (even as  $B$  varies) that computes the representation of  $A$  using the representation of  $B$  as an oracle (and similarly for computable enumerability). These representations, in turn, can be made into explicit subsets of  $\omega$  using the relevant enumerations, as in (9).

### 3.1. Representations of Reals

We will use both the standard topology and right order topology on the real line  $\mathbf{R}$ . The reals under the standard topology are a computable topological space using the basis  $\mathcal{I}_{\mathbf{Q}}$  with respect to the canonical enumeration. The reals under the *right order topology* are a computable topological space using the basis

$$\mathcal{R}_{<} := \{(c, \infty) : c \in \mathbf{Q}\}, \tag{11}$$

under the canonical enumeration. Note that the computable points of  $(\mathbf{R}, \mathcal{I}_{\mathbf{Q}})$  are precisely the computable reals, and the computable points of  $(\mathbf{R}, \mathcal{R}_{<})$  are precisely the c.e. reals.

### 3.2. Continuous Real Functions

We now consider computable representations for continuous functions on the reals.

**Definition 3.3 (Computable real function).** Let  $\mathcal{S}$  and  $\mathcal{T}$  each be either of  $\mathcal{I}_{\mathbf{Q}}$  or  $\mathcal{R}_{<}$ , under their canonical enumeration  $s$  and  $t$ . Let  $\bar{A}$  denote the closure of the set  $A$ . Fix  $k \geq 1$ . We say that a continuous function  $f : (\mathbf{R}^k, \mathcal{S}^k) \rightarrow (\mathbf{R}, \mathcal{T})$  is *computable* when

$$\{(A, B) \in \mathcal{S}^k \times \mathcal{T} : f(\bar{A}) \subseteq B\} \quad (12)$$

is a c.e. set (in terms of the  $s'$ -indices for  $A \in \mathcal{S}^k$  and  $t$ -indices for  $B \in \mathcal{T}$ , where  $s'$  is the canonical enumeration determined by  $s$ ). We call the set (12) the *representation* of  $f$ .

This definition is computably equivalent to the canonical construction of computable functions between two represented spaces [20, Ch. 6]. Note that when  $\mathcal{S} = \mathcal{T} = \mathcal{I}_{\mathbf{Q}}$ , this recovers the standard definition of a computable real function. When  $\mathcal{S} = \mathcal{I}_{\mathbf{Q}}$  and  $\mathcal{T} = \mathcal{R}_{<}$ , this recovers the standard definition of a lower-semicomputable real function [21].

### 3.3. Representations of Borel Probability Measures

The following representations for probability measures on computable topological spaces are devised from more general TTE representations in Schröder [22] and Bosserhoff [23], and agree with Weihrauch [24] in the case of the unit interval. In particular, the representation for  $\mathcal{M}_1(S)$  below is admissible with respect to the weak topology, hence computably equivalent (see Weihrauch [20, Chap. 3]) to the canonical representation for Borel measures given in Schröder [22].

Schröder [22] has also shown the equivalence of this representation for probability measures (as a computable space under the weak topology) with *probabilistic processes*. A probabilistic process (see Schröder and Simpson [25]) formalizes the notion of a program which uses randomness to sample points in terms of their representations of the form (10).

Let  $\mathcal{M}_1(S)$  denote the set of Borel probability measures on a second-countable  $T_0$  topological space  $S$  (i.e., the probability measures on the  $\sigma$ -algebra generated by  $\mathcal{S}$ ). Provided that the subbasis  $\mathcal{S}$  is closed under finite intersection, such measures are determined by the measure they assign to elements of  $\mathcal{S}$ . Note that  $\mathcal{M}_1(S)$  is a second-countable  $T_0$  space.

Let  $S$  be a computable topological space. We will describe a subbasis for  $\mathcal{M}_1(S)$  which makes it a computable topological space. Let  $\mathcal{A}_{\mathcal{S}}$  denote the algebra generated by the subbasis  $\mathcal{S}$  of  $S$  (i.e., the closure of  $\mathcal{S}$  under finite union and complementation). Then, the class of sets

$$\{\gamma \in \mathcal{M}_1(\mathbf{R}) : \gamma\sigma > q\}, \quad (13)$$

where  $\sigma \in \mathcal{A}_S$  and  $q \in \mathbf{Q}$ , is a subbasis for the weak topology on  $\mathcal{M}_1(S)$ . An effective enumeration of this subbasis can be constructed in a canonical fashion from the enumeration of  $S$  and the rationals, making  $\mathcal{M}_1(S)$  a computable topological space.

**Definition 3.4 (Computable distribution).** Let  $S$  be a computable topological space with respect to  $s$ , and let  $\mathcal{A}_S$  be the algebra generated by the subbasis  $S$ , under its canonical enumeration  $s'$  induced by  $s$ . A Borel probability measure  $\eta \in \mathcal{M}_1(S)$  is *computable (with respect to  $s$ )* when  $\eta B$  is a c.e. real, uniformly in the  $s'$ -index of  $B \in \mathcal{A}_S$ . We call the set  $\{(B, q) \in \mathcal{A}_S \times \mathbf{Q} : \eta B > q\}$  the *representation* of  $\eta$ .

In particular, this implies that the measure of a c.e. open set (i.e., the c.e. union of basic open sets) is a c.e. real (uniformly in the enumeration of the terms in the union), and that the measure of a co-c.e. closed set (i.e., the complement of a c.e. open set) is a co-c.e. real (similarly uniformly); see, e.g., [26, §3.3] for details. Note that on a discrete space, where singletons are both c.e. open and co-c.e. closed, the measure of each singleton is a computable real. But for a general space, it is too strong to require that even basic open sets have computable measure (as it is more than is needed to ensure that exact samples be computably described to arbitrary accuracy).

We will be interested in computable measures  $\eta \in \mathcal{M}_1(S)$  where  $S$  is either  $\mathbf{R}^\omega$ ,  $[0, 1]^k$ , or  $\mathcal{M}_1(\mathbf{R})$ , topologized as described below.

### 3.3.1. Measures on Real Vectors and Sequences under the Standard Topology

Consider  $\mathbf{R}^\omega$  under the product topology, where  $\mathbf{R}$  is endowed with its standard topology. Note that  $\mathbf{R}^\omega$  is a computable topological space with respect to the canonical enumeration on the basis given by the cylinders  $\sigma \times \mathbf{R}^\omega$  for  $\sigma \in \bigcup_{k \geq 1} \mathcal{I}_{\mathbf{Q}}^k$ . Using Definition 3.4, we can characterize the class of computable distributions on real sequences.

Let  $\vec{x} = \{x_i\}_{i \geq 1}$  be a sequence of real-valued random variables (e.g., the exchangeable sequence  $X$ , or the derived random variables  $\{V_\tau\}_{\tau \in \mathcal{I}_{\mathbf{Q}}}$  under the canonical enumeration of  $\mathcal{I}_{\mathbf{Q}}$ ). Thus, the joint distribution  $\eta$  of  $\vec{x}$  is computable when  $\eta(\sigma \times \mathbf{R}^\omega) = \mathbb{P}\{x \in \sigma \times \mathbf{R}^\omega\}$  is a c.e. real, uniformly in  $\sigma \in \mathcal{A}_{\mathbf{Q}^k}$ . The following simpler representation was shown to be equivalent by Müller [27, Thm. 3.7].

**Lemma 3.5 (Computable distribution under the standard topology).**

Let  $\vec{x} = \{x_i\}_{i \geq 1}$  be a sequence of real-valued random variables with joint distribution  $\eta$ . Then  $\eta$  is computable if and only if

$$\eta(\tau \times \mathbf{R}^\omega) = \mathbb{P}\left(\bigcap_{i=1}^k \{x_i \in \tau(i)\}\right) \tag{14}$$

is a c.e. real, uniformly in  $k \geq 1$  and  $\tau \in \mathcal{I}_{\mathbf{Q}}^k$ . □

Therefore knowing the measure of the sets in  $\bigcup_k \mathcal{I}_{\mathbf{Q}}^k \subsetneq \bigcup_k \mathcal{A}_{\mathbf{Q}^k}$  is sufficient. Note that (14) is precisely the form of the first expression in Corollary 2.5. Note also that one obtains a characterization of the computability of a finite-dimensional vector by embedding it in an initial segment of a sequence.

### 3.3.2. Measures on Real Vectors and Sequences under the Right Order Topology

Distributions on c.e. reals play an important role when representing measures on measures, because, as Definition 3.4 indicates, measures are themselves represented by collections of c.e. reals lower-bounding the measure of an element of an algebra. The set  $\mathcal{R}_{<}^k$  is a basis for the product of the right order topology on  $\mathbf{R}^k$  that makes  $(\mathbf{R}^k, \mathcal{R}_{<}^k)$  a computable topological space (under the canonical enumeration of  $\mathcal{R}_{<}^k$ ).

#### Corollary 3.6 (Computable distribution under the right order topology).

For  $m, k \geq 1$ , let  $\vec{w} = (w_1, \dots, w_k)$  be a random vector in  $\mathbf{R}^k$ , and let  $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$ . The joint distribution of  $\vec{w}$  is computable under the right order topology when  $\mathbb{P}(\bigcup_{i=1}^m \bigcap_{j=1}^k \{w_j > c_{ij}\})$  is a c.e. real, uniformly in  $C$ .  $\square$

Note that the joint distribution of a random vector in  $\mathbf{R}^\omega$  is computable under the right order topology when all its restrictions to  $\mathbf{R}^k$  are computable under the right order topology, uniformly in  $k$ . Note also that if a distribution on  $\mathbf{R}^k$  is computable under the standard topology, then it is clearly computable under the right order topology. The above representation is used in the next section as well as in Proposition 5.1, where we must compute an integral with respect to a topology that is weaker than the standard topology.

### 3.3.3. Measures on Borel Measures

The de Finetti measure  $\mu$  is the distribution of the directing random measure  $\nu$ , an  $\mathcal{M}_1(\mathbf{R})$ -valued random variable. Recall the definition  $V_\beta := \nu\beta$ , for  $\beta \in \mathcal{B}_{\mathbf{R}}$ . Then  $\mu$  is computable (with respect to an enumeration  $s$  of the weak topology) when

$$\mu(\bigcup_{i=1}^m \bigcap_{j=1}^k \{\gamma \in \mathcal{M}_1(\mathbf{R}) : \gamma\sigma(j) > c_{ij}\}) = \mathbb{P}(\bigcup_{i=1}^m \bigcap_{j=1}^k \{V_{\sigma(j)} > c_{ij}\}) \quad (15)$$

is a c.e. real, uniformly in  $\sigma \in \mathcal{A}_{\mathbf{Q}}^k$  and  $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$ . As an immediate consequence of (15) and Corollary 3.6, we obtain the following representation for the de Finetti measure.

**Corollary 3.7 (Computable de Finetti measure).** *The de Finetti measure  $\mu$  is computable if and only if the joint distribution of  $\{V_\tau\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  is computable under the right order topology.*  $\square$

### 3.3.4. Integration

The following lemma is a restatement of an integration result by Schröder [22, Prop. 3.6], which itself generalizes integration results on standard topologies of finite-dimensional Euclidean spaces by Müller [27] and the unit interval by Weihrauch [24].

Define

$$\mathbb{I} := \{A \cap [0, 1] : A \in \mathcal{I}_{\mathbf{Q}}\}, \quad (16)$$

which is a basis for the standard topology on  $[0, 1]$ , and define

$$\mathbb{I}_{<} := \{A \cap [0, 1] : A \in \mathcal{R}_{<}\}, \quad (17)$$

which is a basis for the right order topology on  $[0, 1]$ .

**Lemma 3.8 (Integration of bounded lower-semicomputable functions).**

*Let  $k \geq 1$  and let  $\mathcal{S}$  be either  $\mathcal{I}_{\mathbf{Q}}$  or  $\mathcal{R}_{<}$ . Let*

$$f : (\mathbf{R}^k, \mathcal{S}^k) \rightarrow ([0, 1], \mathbb{I}_{<}) \quad (18)$$

*be a computable function and let  $\mu$  be a computable distribution on  $(\mathbf{R}^k, \mathcal{S}^k)$ . Then*

$$\int f d\mu \quad (19)$$

*is a c.e. real, uniformly in  $f$  and  $\mu$ .  $\square$*

The following result of Müller [27] is an immediate corollary.

**Corollary 3.9 (Integration of bounded computable functions).** *Let*

$$g : (\mathbf{R}^k, \mathcal{I}_{\mathbf{Q}}^k) \rightarrow ([0, 1], \mathbb{I}) \quad (20)$$

*be a computable function and let  $\mu$  be a computable distribution on  $(\mathbf{R}^k, \mathcal{I}_{\mathbf{Q}}^k)$ . Then*

$$\int g d\mu \quad (21)$$

*is a computable real, uniformly in  $g$  and  $\mu$ .  $\square$*

#### 4. The Computable Moment Problem

One often has access to the moments of a distribution, and wishes to recover the underlying distribution. Let  $\vec{x} = (x_i)_{i \in \omega}$  be a random vector in  $[0, 1]^\omega$  with distribution  $\eta$ . Classically, the distribution of  $\vec{x}$  is uniquely determined by the mixed moments of  $\vec{x}$ . We show that the distribution is in fact *computable* from the mixed moments.

One classical way to pass from the moments of  $\vec{x}$  to its distribution is via the Lévy inversion formula, which maps the characteristic function  $\phi_{\vec{x}} : \mathbf{R}^\omega \rightarrow \mathbf{C}$ , given by

$$\phi_{\vec{x}}(t) = \mathbb{E}(e^{i\langle t, \vec{x} \rangle}), \quad (22)$$

to the distribution of  $\vec{x}$ . However, even in the finite-dimensional case, the inversion formula involves a limit for which we have no direct handle on the rate of convergence, and so the distribution it defines is not obviously computable. Instead, we use computable versions of Urysohn's lemma and the Weierstrass approximation theorem to compute a representation of a distribution from its moments.

To show that  $\eta$  is computable, it suffices to show that  $\eta(\sigma \times [0, 1]^\omega) = \mathbb{E}(\mathbf{1}_\sigma(x_1, \dots, x_k))$  is a c.e. real, uniformly in  $\sigma \in \bigcup_{k \geq 1} \mathcal{I}_{\mathbf{Q}}^k$ . We begin by showing how to build sequences of polynomials that converge pointwise from below to indicator functions of the form  $\mathbf{1}_\sigma$  for  $\sigma \in \bigcup_{k \geq 1} \mathcal{A}_{\mathbf{Q}^k}$ .

**Lemma 4.1 (Polynomial approximations).** *There is a computable array*

$$\left\{ p_{n,\sigma} : n \in \omega, \sigma \in \bigcup_{k \geq 1} \mathcal{A}_{\mathbf{Q}^k} \right\} \quad (23)$$

*of rational polynomials where, for  $n \in \omega$  and  $\sigma \in \mathcal{A}_{\mathbf{Q}^k}$ , the polynomial  $p_{n,\sigma}$  is in  $k$  variables and, for each  $\vec{x} \in [0, 1]^k$ , we have*

$$-1 \leq p_{n,\sigma}(\vec{x}) \leq \mathbf{1}_\sigma(\vec{x}) \quad \text{and} \quad \lim_{m \rightarrow \infty} p_{m,\sigma}(\vec{x}) = \mathbf{1}_\sigma(\vec{x}). \quad (24)$$

**PROOF.** Fix  $n \in \omega$  and  $\sigma \in \mathcal{A}_{\mathbf{Q}^k}$ . Let  $U_{n,\sigma}$  denote the set of  $\vec{x} \in \sigma$  for which the  $1/n$ -ball centered at  $\vec{x}$  is contained in  $\sigma$ . We will construct (using a procedure that is uniform in  $n$  and  $\sigma$ ) a rational polynomial  $p_{n,\sigma}$  which satisfies

$$-1 \leq p_{n,\sigma}(\vec{x}) \leq \mathbf{1}_\sigma(\vec{x}) \quad (25)$$

for all  $\vec{x} \in [0, 1]^k$ , and which satisfies

$$\mathbf{1}_\sigma(\vec{x}) - p_{n,\sigma}(\vec{x}) \leq 1/n \quad (26)$$

for all  $\vec{x} \in U_{n,\sigma}$  and for all  $\vec{x} \in [0, 1]^k \setminus \sigma$ . Then the values  $\{p_{n,\sigma}(\vec{x})\}_{n \geq 1}$  all equal  $\mathbf{1}_\sigma(\vec{x})$  on  $\vec{x} \in [0, 1]^k \setminus \sigma$ , and also equal  $\mathbf{1}_\sigma(\vec{x})$  for  $\vec{x}$  on an increasing sequence of subsets of  $\sigma$  whose union is  $\sigma$ .

For  $n = 1$  we may take  $p_{1,\sigma} = 0$ . Thus assume  $n \geq 2$ . Using a computable version of Urysohn's lemma (see Weihrauch [28]), we can find (uniformly in  $n$  and  $\sigma$ ) a computable real function  $f_{n,\sigma}$  which equals  $1 - 1/n$  for all  $\vec{x} \in U_{n,\sigma}$ , equals  $-1/n$  on  $\vec{x} \in [0, 1]^k \setminus \sigma$ , and is in between these two values on  $\sigma \setminus U_{n,\sigma}$ .

Then by the effective Weierstrass approximation theorem (see Pour-El and Richards [29, p. 45]), we can find (uniformly in  $n$  and  $\sigma$ ) a polynomial  $p_{n,\sigma}$  with rational coefficients which uniformly approximates  $f_{n,\sigma}$  to within  $1/2n$  on  $[0, 1]^k$ . This polynomial  $p_{n,\sigma}$  has the desired properties.  $\square$

Using these polynomials, we can compute the distribution from the moments. The other direction follows from computable integration results.

**Theorem 4.2 (Computable moments).** *Let  $\vec{x} = (x_i)_{i \in \omega}$  be a random vector in  $[0, 1]^\omega$  with distribution  $\eta$ . Then  $\eta$  is uniformly computable in the mixed moments of  $\{x_i\}_{i \in \omega}$ , and vice versa. In particular,  $\eta$  is computable if and only if the mixed moments of  $\{x_i\}_{i \in \omega}$  are uniformly computable.*

PROOF. Any monic monomial in  $x_1, \dots, x_k$ , considered as a real function, computably maps  $[0, 1]^k$  into  $[0, 1]$  (under the standard topology). Furthermore,  $\eta$  restricts to a computable measure on  $\{x_i\}_{i \leq k}$ . Therefore, by Corollary 3.9, the mixed moments are uniformly computable in  $\eta$  and the index of the monomial.

To prove that  $\eta$  is computable in the mixed moments, we give a proof which relativizes to representation of the mixed moments, and so we assume that the mixed moments of  $\{x_i\}_{i \in \omega}$  are uniformly computable, without loss of generality. Let  $k \geq 1$  and  $\sigma \in \mathcal{I}_{\mathbf{Q}}^k$ . To establish the computability of  $\eta$ , it suffices to show that

$$\eta(\sigma \times [0, 1]^\omega) = \mathbb{E}(\mathbf{1}_{\sigma \times [0, 1]^\omega}(\vec{x})) = \mathbb{E}(\mathbf{1}_\sigma(x_1, \dots, x_k)). \quad (27)$$

is a c.e. real, uniformly in  $\sigma$ . By Lemma 4.1, we are given a uniformly computable sequence of polynomials  $(p_{n,\sigma})_{n \in \omega}$  which converge pointwise from below to the indicator  $\mathbf{1}_\sigma$ . By the dominated convergence theorem,

$$\mathbb{E}(\mathbf{1}_\sigma(x_1, \dots, x_k)) = \sup_n \mathbb{E}(p_{n,\sigma}(x_1, \dots, x_k)). \quad (28)$$

The expectation  $\mathbb{E}(p_{n,\sigma}(x_1, \dots, x_k))$  is a  $\mathbf{Q}$ -linear combination of moments, hence a computable real. Thus their supremum is a c.e. real.  $\square$

## 5. Proof of the Computable de Finetti Theorem

In the remainder of the paper, let  $X$  be a real-valued exchangeable sequence with distribution  $\chi$ , let  $\nu$  denote its directing random measure, and let  $\mu$  denote the de Finetti measure.

Classically, the joint distribution of  $X$  is uniquely determined by the de Finetti measure (see Equation 3). We now show that the joint distribution of  $X$  is in fact *computable* from the de Finetti measure.

**Proposition 5.1.** *The distribution  $\chi$  is uniformly computable in  $\mu$ .*

PROOF. We give a proof that relativizes to  $\mu$ ; without loss of generality, assume that  $\mu$  is computable. In order to show that  $\chi$ , the distribution of  $X$ , is computable, we must show that  $\mathbb{P}(\bigcap_{i=1}^k \{X_i \in \sigma(i)\})$  is a c.e. real, uniformly in  $\sigma \in \mathcal{I}_{\mathbf{Q}}^k$ . Fix  $\sigma \in \mathcal{I}_{\mathbf{Q}}^k$ . Note that, by Corollary 2.5,

$$\mathbb{P}(\bigcap_{i=1}^k \{X_i \in \sigma(i)\}) = \mathbb{E}(\prod_{i=1}^k V_{\sigma(i)}). \quad (29)$$

Let  $\eta$  be the joint distribution of  $(V_{\sigma(i)})_{i \leq k}$  and let  $f : [0, 1]^k \rightarrow [0, 1]$  be defined by

$$f(x_1, \dots, x_k) := \prod_{i=1}^k x_i. \quad (30)$$

To complete the proof, we now show that

$$\int f d\eta = \mathbb{E}(\prod_{i=1}^k V_{\sigma(i)}) \quad (31)$$

is a c.e. real. Note that the computability of  $\mu$  implies that  $\eta$  is computable under the right order topology. Furthermore,  $f$  is order-preserving, and so is a continuous (and obviously computable) function from  $([0, 1]^k, \mathbb{I}_{<}^k)$  to  $([0, 1], \mathbb{I}_{<})$ . Therefore, by Lemma 3.8, we have that  $\int f d\eta$  is a c.e. real.  $\square$

We will first prove the main theorem under the additional hypothesis that the directing random measure is almost surely continuous. We then sketch a randomized argument which succeeds with probability one. Finally, we present the proof of the main result, which can be seen as a derandomization.

### 5.1. Almost Surely Continuous Directing Random Measures

For  $k \geq 1$  and  $\psi \in \mathcal{A}_{\mathbf{R}}^k$ , we say that  $\psi$  is an  $X$ -continuity set when  $X_i \notin \partial\psi(i)$  a.s. for  $1 \leq i \leq k$  (informally,  $X$  places no mass on the boundary of  $\psi$ ).

**Lemma 5.2.** *Relative to  $\chi$ , the mixed moments of  $\{V_{\tau}\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  are uniformly c.e. reals and the mixed moments of  $\{V_{\bar{\tau}}\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  are uniformly co-c.e. reals; in particular, if  $\sigma \in \mathcal{A}_{\mathbf{Q}}^k$  is an  $X$ -continuity set, then the mixed moment  $\mathbb{E}(\prod_{i=1}^k V_{\sigma(i)})$  is a computable real.*

PROOF. Without loss of generality, we may assume that  $\chi$  is computable. Let  $k \geq 1$  and  $\sigma \in \mathcal{A}_{\mathbf{Q}}^k$ . By Corollary 2.5,

$$\mathbb{E}(\prod_{i=1}^k V_{\sigma(i)}) = \mathbb{P}(\bigcap_{i=1}^k \{X_i \in \sigma(i)\}), \quad (32)$$

which is a c.e. real because  $\chi$  is computable. The set  $\bar{\sigma}$  is a co-c.e. closed set in  $\mathbf{R}^k$  because we can computably enumerate all  $\tau \in \mathcal{A}_{\mathbf{Q}}^k$  contained in the complement of  $\sigma$ . Therefore,

$$\mathbb{E}(\prod_{i=1}^k V_{\overline{\sigma(i)}}) = \mathbb{P}(\bigcap_{i=1}^k \{X_i \in \overline{\sigma(i)}\}) \quad (33)$$

is the measure of a co-c.e. closed set, hence a co-c.e. real. When  $\sigma$  is an  $X$ -continuity set,

$$\mathbb{E}(\prod_{i=1}^k V_{\sigma(i)}) = \mathbb{E}(\prod_{i=1}^k V_{\overline{\sigma(i)}}), \quad (34)$$

and so the expectation is a computable real.  $\square$

**Proposition 5.3 (Almost surely continuous case).** *Assume that  $\nu$  is continuous with probability one. Then  $\mu$  is uniformly computable in  $\chi$ .*

PROOF. We give a proof that relativizes to  $\chi$ ; assume without loss of generality that  $\chi$  is computable. Let  $k \geq 1$  and consider  $\sigma \in \mathcal{A}_{\mathbf{Q}}^k$ . The almost sure continuity of  $\nu$  implies that  $X_i \notin \partial\sigma(i)$  a.s., i.e.,  $\sigma$  is an  $X$ -continuity set. Therefore, by Lemma 5.2, the moment  $\mathbb{E}(\prod_{i=1}^k V_{\sigma(i)})$  is a computable real. The computable moment theorem (Theorem 4.2) then implies that joint distribution of the variables  $\{V_\tau\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  is computable under the standard topology, and so  $\mu$  is computable under the weaker right order topology. By Corollary 3.7, this implies that  $\mu$  is computable.  $\square$

## 5.2. “Randomized” Proof Sketch

In general, the joint distribution of  $\{V_\sigma\}_{\sigma \in \mathcal{A}_{\mathbf{Q}}}$  is not computable under the standard topology because the directing random measure  $\nu$  may, with nonzero probability, have a fixed point mass on a rational. In this case, the mixed moments of  $\{V_\tau\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  are c.e., but not co-c.e., reals (relative to  $\chi$ ). In this case, the computable moment theorem (Theorem 4.2) is inapplicable. For arbitrary directing random measures, we give a proof of the computable de Finetti theorem which works regardless of the location of point masses.

Consider the following sketch of a “randomized algorithm”: We independently sample a countably infinite set of real numbers  $\mathbf{A} \subseteq \mathbf{R}$  from a computable, absolutely continuous distribution which has support everywhere on the real line (e.g., a Gaussian or Cauchy). Let  $\mathcal{A}_{\mathbf{A}}$  denote the algebra generated by open intervals with endpoints in  $\mathbf{A}$ . Note that, with probability one (over the draw),

$\mathbf{A}$  will be dense in  $\mathbf{R}$  and  $\nu(\mathbf{A}) = 0$  almost surely, and so  $X$  will place no mass on the boundary of any  $\psi \in \mathcal{A}_{\mathbf{A}}$ ; suppose we have chosen such an  $\mathbf{A}$ . Now we can proceed analogously to the case where  $\nu$  is almost surely continuous, relative to an oracle  $A$  which encodes  $\mathbf{A}$ .

We begin by proving an extension of Lemma 5.2 that shows that the mixed moments of variables in terms of the new basis are  $A$ -computable, relative to  $\chi$ .

**Definition 5.4.** We call  $\psi \in \mathcal{A}_{\mathbf{R}}^k$  a *refinement* of  $\varphi \in \mathcal{A}_{\mathbf{R}}^k$ , and write  $\psi \triangleleft \varphi$ , when

$$\overline{\psi(i)} \subseteq \varphi(i) \tag{35}$$

for all  $i \leq k$ .

**Lemma 5.5.** *Let  $\mathbf{A}$  be sampled as above. Relative to  $\chi$ , the mixed moments of  $\{V_{\zeta}\}_{\zeta \in \mathcal{A}_{\mathbf{A}}}$  are uniformly  $A$ -c.e. reals and the mixed moments of  $\{V_{\bar{\zeta}}\}_{\zeta \in \mathcal{A}_{\mathbf{A}}}$  are uniformly  $A$ -co-c.e. reals. In particular, if  $\psi \in \mathcal{A}_{\mathbf{A}}^k$  is an  $X$ -continuity set, then the mixed moment  $\mathbb{E}(\prod_{i=1}^k V_{\psi(i)})$  is an  $A$ -computable real relative to  $\chi$ .*

PROOF. Fix  $\psi \in \mathcal{A}_{\mathbf{A}}^k$ . All computability claims are relative to  $\chi$ . We may  $A$ -compute a sequence

$$\sigma_1, \sigma_2, \dots \in \mathcal{A}_{\mathbf{Q}}^k \tag{36}$$

such that for each  $n \geq 1$ ,

$$\sigma_n \triangleleft \sigma_{n+1} \quad \text{and} \quad \bigcup_m \sigma_m = \psi. \tag{37}$$

For each  $n$ ,

$$\mathbb{E}(\prod_{i=1}^k V_{\sigma_n(i)}) \tag{38}$$

is a c.e. real, and so their supremum is an  $A$ -c.e. real.

If  $\zeta, \varphi \in \mathcal{A}_{\mathbf{R}}$  satisfy  $\zeta \triangleleft \varphi$  then  $V_{\zeta} \leq V_{\varphi}$  (a.s.). Multiplication is continuous, and so the dominated convergence theorem gives us

$$\mathbb{E}(\prod_{i=1}^k V_{\psi(i)}) = \sup_n \mathbb{E}(\prod_{i=1}^k V_{\sigma_n(i)}), \tag{39}$$

which we have already noted is an  $A$ -c.e. real. The co-c.e. result follows similarly from a sequence of nested unions of rational intervals whose intersection is  $\overline{\psi}$ . The final claim is immediate as in Lemma 5.2.  $\square$

The proof of the computable moment theorem (Theorem 4.2) relativizes, and so the joint distribution of  $\{V_{\psi}\}_{\psi \in \mathcal{A}_{\mathbf{A}}}$  is  $A$ -computable in  $\chi$ . This joint distribution also classically determines the de Finetti measure. Moreover, we can  $A$ -compute (relative to  $\chi$ ) the desired representation with respect to the original basis.

**Lemma 5.6.** *Let  $\mathbf{A}$  be sampled as above. The de Finetti measure  $\mu$  is uniformly  $A$ -computable in  $\chi$ .*

PROOF. We show that the joint distribution of  $\{V_\tau\}_{\tau \in \mathcal{A}_{\mathbf{Q}}}$  is  $A$ -computable under the right order topology. Let  $m, k \geq 1$ , let  $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$ , and let  $\tau \in \mathcal{A}_{\mathbf{Q}}^k$ . Note that  $\tau$  is an  $A$ -c.e. open set with respect to the basis  $\mathcal{A}_{\mathbf{A}}^k$ , and so we can  $A$ -computably enumerate a sequence  $\sigma_1, \sigma_2, \dots \in \mathcal{A}_{\mathbf{A}}^k$  such that  $\cup_n \sigma_n = \tau$ . Without loss of generality, we may assume that  $\sigma_n \subseteq \sigma_{n+1}$ . Note that  $V_{\sigma_n(j)} \leq V_{\tau(j)}$  (a.s.) for all  $n \geq 1$  and  $j \leq k$ . By the continuity of  $\mathbb{P}$ ,

$$\mathbb{P}(\cup_{i=1}^m \cap_{j=1}^k \{V_{\tau(j)} > c_{ij}\}) = \sup_n \mathbb{P}(\cup_{i=1}^m \cap_{j=1}^k \{V_{\sigma_n(j)} > c_{ij}\}). \quad (40)$$

The right hand side is the supremum of a uniformly  $A$ -computable set of  $A$ -computable reals, relative to  $\chi$ , hence an  $A$ -c.e. real relative to  $\chi$ .  $\square$

If we consider  $A$  to be a random oracle, then the representation of  $\mu$  that was computed (relative to  $\chi$ ) using  $A$  is also a random variable. However, while  $A$  is almost surely noncomputable, the random representation of  $\mu$  is almost surely constant. Recall that the distribution of the oracle  $A$  is computable because we started with a computable distribution on  $\mathbf{A}$ . Hence we could compute  $\mu$  by composing the distribution of  $\mathbf{A}$  with the (almost surely constant) representation of  $\mu$ .

A proof along these lines could be made precise using a simulation argument to calculate the pushforward measure of the random representation of  $\mu$ . The basis elements assigned positive probability (and therefore probability one) are precisely those which hold of  $\mu$ . Instead, in Section 5.3, we complete the proof by explicitly computing the representation of  $\mu$  in terms of our rational basis. This construction can be seen as a “derandomization” of the above algorithm.

Alternatively, the above sketch could be interpreted as a degenerate *probabilistic process* (see Schröder and Simpson [25]) which returns a representation of the de Finetti measure with probability one. Schröder [22] shows that representations in terms of probabilistic processes are computably reducible to representations of computable distributions.

The structure of the derandomized argument occurs in other proofs in computable analysis and probability theory. Weihrauch [24, Thm. 3.6] proves a computable integration result via an argument that could likewise be seen as a derandomization of an algorithm which densely subdivides the unit interval at random locations to find continuity sets. Bosserhoff [23, Lem. 2.15] uses a similar argument to compute a basis for a computable metric space, for which every basis element is a continuity set; this suggests an alternative approach to completing our proof. Müller [27, Thm. 3.7] uses a similar construction to find open hypercubes such that for any  $\epsilon > 0$ , the probability on their boundaries is less than  $\epsilon$ .

These arguments also resemble the classic proof of the Portmanteau theorem [5, Thm. 4.25], in which an uncountable family of sets with disjoint boundaries is defined, almost all of which are continuity sets.

### 5.3. “Derandomized” Construction

Fix  $m, k \geq 1$  and let  $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$ . Define

$$\mathbf{1}_C : [0, 1]^k \rightarrow [0, 1] \quad (41)$$

to be the indicator function for the set

$$\bigcup_{i=1}^m (c_{i1}, 1] \times \cdots \times (c_{ik}, 1]. \quad (42)$$

For  $n \in \omega$ , we denote by  $p_{n,C}$  the polynomial  $p_{n,\sigma}$  (as defined in Lemma 4.1), where

$$\sigma := \bigcup_{i=1}^m (c_{i1}, 2) \times \cdots \times (c_{ik}, 2) \in \mathcal{A}_{\mathbf{Q}^k}. \quad (43)$$

Here, we arbitrarily chose  $2 > 1$  so that the sequence of polynomials  $\{p_{n,C}\}_{n \in \omega}$  converges pointwise from below to  $\mathbf{1}_C$  on  $[0, 1]^k$ .

Let  $\vec{x} = (x_1, \dots, x_k)$  and  $\vec{y} = (y_1, \dots, y_k)$ . We can write

$$p_{n,C}(\vec{x}) = p_{n,C}^+(\vec{x}) - p_{n,C}^-(\vec{x}), \quad (44)$$

where  $p_{n,C}^+$  and  $p_{n,C}^-$  are polynomials with positive coefficients. Define the  $2k$ -variable polynomial

$$q_{n,C}(\vec{x}, \vec{y}) := p_{n,C}^+(\vec{x}) - p_{n,C}^-(\vec{y}). \quad (45)$$

We denote

$$q_{n,C}(V_{\psi(1)}, \dots, V_{\psi(k)}, V_{\zeta(1)}, \dots, V_{\zeta(k)}) \quad (46)$$

by  $q_{n,C}(V_{\psi}, V_{\zeta})$ , and similarly with  $p_{n,C}$ .

**Proposition 5.7.** *Let  $n \in \omega$ , let  $k, m \geq 1$ , let  $\sigma \in \mathcal{A}_{\mathbf{Q}^k}$ , and let  $C \in \mathbf{Q}^{m \times k}$ . Then  $\mathbb{E}_{q_{n,C}}(V_{\sigma}, V_{\bar{\sigma}})$  is a c.e. real relative to  $\chi$ , uniformly in  $n$ ,  $\sigma$  and  $C$ .*

PROOF. By Lemma 5.2, relative to  $\chi$ , each monomial of  $p_{n,C}^+(V_{\sigma})$  has a c.e. real expectation, and each monomial of  $p_{n,C}^-(V_{\bar{\sigma}})$  has a co-c.e. real expectation, and so by the linearity of expectation,  $\mathbb{E}_{q_{n,C}}(V_{\sigma}, V_{\bar{\sigma}})$  is a c.e. real.  $\square$

We are now ready to prove the main theorem.

PROOF OF THEOREM 2.3 (COMPUTABLE DE FINETTI). The distribution  $\chi$  is uniformly computable in  $\mu$  by Proposition 5.1. We now give a proof of the other direction, showing that the joint distribution of  $\{V_\sigma\}_{\sigma \in \mathcal{A}_\mathbf{Q}}$  is computable, relative to  $\chi$ , under the right order topology.

Let  $k, m \geq 1$ , let  $\pi \in \mathcal{A}_\mathbf{Q}^k$  and let  $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$ . For  $\zeta \in \mathcal{A}_\mathbf{R}^k$ , let  $V_\zeta$  denote the  $k$ -tuple  $(V_{\zeta(1)}, \dots, V_{\zeta(k)})$  and similarly for  $V_{\bar{\zeta}}$ . Take  $\mathbf{1}_C$  to be defined as above in (41). It suffices to show that

$$\mathbb{P}\left(\bigcup_{i=1}^m \bigcap_{j=1}^k \{V_{\pi(j)} > c_{ij}\}\right) = \mathbb{E}\mathbf{1}_C(V_\pi) \quad (47)$$

is a c.e. real relative to  $\chi$ , uniformly in  $\pi$  and  $C$ . We do this by a series of reductions, which results in a supremum over quantities of the form  $\mathbb{E}q_{n,C}(V_\sigma, V_{\bar{\sigma}})$  for  $\sigma \in \mathcal{A}_\mathbf{Q}^k$ . By Proposition 5.7, these quantities are c.e. reals relative to  $\chi$ .

Note that (42) is an open set in the right order topology on  $[0, 1]^k$  and so  $\mathbf{1}_C$  is an order preserving function. In particular, if  $\zeta, \varphi \in \mathcal{A}_\mathbf{R}^k$  satisfy  $\zeta \triangleleft \varphi$ , then  $V_\zeta \leq V_\varphi$  (a.s.), and so

$$\mathbb{E}\mathbf{1}_C(V_\psi) \leq \mathbb{E}\mathbf{1}_C(V_\pi) \quad (48)$$

for all  $\psi \in \mathcal{A}_\mathbf{R}^k$  such that  $\psi \triangleleft \pi$ . Therefore, by the density of the reals and the dominated convergence theorem, we have that

$$\mathbb{E}\mathbf{1}_C(V_\pi) = \sup_{\psi \triangleleft \pi} \mathbb{E}\mathbf{1}_C(V_\psi) \quad (49)$$

where  $\psi$  ranges over  $\mathcal{A}_\mathbf{R}^k$ . Recall that the polynomials  $\{p_{n,C}\}_{n \in \omega}$  converge pointwise from below to  $\mathbf{1}_C$  in  $[0, 1]^k$ . Therefore, by the dominated convergence theorem,

$$\mathbb{E}\mathbf{1}_C(V_\psi) = \sup_n \mathbb{E}p_{n,C}(V_\psi). \quad (50)$$

Multiplication is continuous, and so the dominated convergence theorem gives us

$$\mathbb{E}\left(\prod_{i=1}^k V_{\psi(i)}\right) = \sup_{\sigma \triangleleft \psi} \mathbb{E}\left(\prod_{i=1}^k V_{\sigma(i)}\right) \quad \text{and} \quad (51)$$

$$\mathbb{E}\left(\prod_{i=1}^k V_{\bar{\psi}(i)}\right) = \inf_{\tau \triangleright \psi} \mathbb{E}\left(\prod_{i=1}^k V_{\tau(i)}\right), \quad (52)$$

where  $\sigma$  and  $\tau$  range over  $\mathcal{A}_\mathbf{Q}^k$ . Therefore, by the linearity of expectation,

$$\mathbb{E}p_{n,C}^+(V_\psi) = \sup_{\sigma \triangleleft \psi} \mathbb{E}p_{n,C}^+(V_\sigma) \quad \text{and} \quad (53)$$

$$\mathbb{E}p_{n,C}^-(V_\psi) = \inf_{\tau \triangleright \psi} \mathbb{E}p_{n,C}^-(V_\tau). \quad (54)$$

If  $\psi$  is an  $X$ -continuity set, then  $V_{\psi(i)} = V_{\overline{\psi(i)}}$  a.s. for  $i \leq k$ , and so

$$\mathbb{E}p_{n,C}(V_\psi) = \mathbb{E}q_{n,C}(V_\psi, V_\psi) \quad (55)$$

$$= \mathbb{E}q_{n,C}(V_\psi, V_{\overline{\psi}}) \quad (56)$$

$$= \mathbb{E}p_{n,C}^+(V_\psi) - \mathbb{E}p_{n,C}^-(V_{\overline{\psi}}) \quad (57)$$

$$= \sup_{\sigma \triangleleft \psi} \mathbb{E}p_{n,C}^+(V_\sigma) - \inf_{\tau \triangleright \overline{\psi}} \mathbb{E}p_{n,C}^-(V_{\overline{\tau}}) \quad (58)$$

$$= \sup_{\sigma \triangleleft \psi \triangleleft \tau} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\tau}}). \quad (59)$$

Because  $\nu$  has at most countably many point masses, the  $\psi \in \mathcal{I}_{\mathbf{R}}^k$  which are  $X$ -continuity sets are dense in  $\mathcal{I}_{\mathbf{Q}}^k$ . Therefore,

$$\sup_{\psi \triangleleft \pi} \mathbb{E}p_{n,C}(V_\psi) = \sup_{\psi \triangleleft \pi} \sup_{\sigma \triangleleft \psi \triangleleft \tau} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\tau}}). \quad (60)$$

Note that  $\{(\sigma, \tau) : (\exists \psi \triangleleft \pi) \sigma \triangleleft \psi \triangleleft \tau\} = \{(\sigma, \tau) : \sigma \triangleleft \pi \text{ and } \sigma \triangleleft \tau\}$ . Hence

$$\sup_{\psi \triangleleft \pi} \sup_{\sigma \triangleleft \psi} \sup_{\tau \triangleright \overline{\psi}} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\tau}}) = \sup_{\sigma \triangleleft \pi} \sup_{\tau \triangleright \sigma} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\tau}}). \quad (61)$$

Again by dominated convergence we have

$$\sup_{\tau \triangleright \sigma} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\tau}}) = \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\sigma}}). \quad (62)$$

Combining (47), (49), (50), (60), (61), and (62), we have

$$\mathbb{E}\mathbf{1}_C(V_\pi) = \sup_n \sup_{\sigma \triangleleft \pi} \mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\sigma}}). \quad (63)$$

Finally, by Proposition 5.7,

$$\{\mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\sigma}}) : \sigma \triangleleft \pi \text{ and } n \in \omega\} \quad (64)$$

is a set of uniformly c.e. reals, relative to  $\chi$ .  $\square$

## 6. Exchangeability in Probabilistic Functional Programming Languages

The computable de Finetti theorem has implications for the semantics of probabilistic functional programming languages, and in particular, gives conditions under which it is possible to remove uses of mutation (i.e., code which modifies a program's internal state). Furthermore, an implementation of the computable de Finetti theorem itself performs this code transformation automatically.

For context, we provide some background on probabilistic functional programming languages. We then describe the code transformation performed by the computable de Finetti theorem, using the example of the Pólya urn and Beta-Bernoulli process discussed earlier. Finally, we discuss partial exchangeability and its role in recent machine learning applications.

### 6.1. Probabilistic Functional Programming Languages

Functional programming languages with probabilistic choice operators have recently been proposed as universal languages for statistical modeling (e.g., IBAL [30],  $\lambda_{\circ}$  [31], Church [32], and HANSEI [33]). Within domain theory, idealized functional languages that can manipulate exact real numbers, such as Escardó’s REALPCF+ [34] and Plotkin’s PCF++ [35], have also been extended by probabilistic choice operators (e.g., by Escardó [36] and Saheb-Djahromi [37]).

The semantics of probabilistic programs have been studied extensively in theoretical computer science in the context of randomized algorithms, probabilistic model checking, and other areas. However, the application of probabilistic programs to universal statistical modeling has a somewhat different character from much of the other work on probabilistic programming languages.

In Bayesian analysis, the goal is to use observed data to understand hidden variables in a probabilistic model. This type of inductive reasoning, from evidence to hypothesis, can be thought of as inferring the hidden states of a program that generates the observed output. One speaks of the *conditional execution* of probabilistic programs, in which they are “run backwards” to sample from the conditional probability distribution given the observed data.

One important difference is the algorithms used for conditional inference. Goodman et al. [32] describe the language Church, which extends Scheme, and which implements approximate conditional execution via Markov chain Monte Carlo (which can be thought of as a random walk over computational histories). Park, Pfenning, and Thrun [31] describe the language  $\lambda_{\circ}$ , which extends OCaml, and which implements approximate conditional execution by Monte Carlo importance sampling. Ramsey and Pfeffer [38] describe an implementation of the probability monad in Haskell, and develop “measure terms” suitable for implementing the sum-product algorithm, giving a more efficient method for calculating expectations.

These languages also stress the flexibility of representations. In statistics and especially nonparametric statistics, there is an emphasis on higher-order distributions (e.g., distributions on distributions, or distributions on trees), and so it is crucial to work in a language which can express these types. Functional languages with randomness are therefore a natural choice.

The de Finetti theorem gives two different representations for exchangeable sequences, each of which has its own advantages with respect to space and time complexity. In Section 6.3 we provide several examples of the representational changes provided by the de Finetti transformation. Some of these representational issues are also examined in Roy et al. [39]. Masinghka [40] has also described some uses of exchangeable sequences, and situations where one can exploit conditional independence for improved parallel execution.

## 6.2. Code Transformations

We now describe the code transformation performed by the computable de Finetti theorem. For concreteness, we will explore this connection using Church, a probabilistic functional programming language. Church extends Scheme (a dialect of LISP) with a binary-valued `flip` procedure, which denotes the Bernoulli distribution. In Church, an expression denotes the distribution induced by evaluation. For example,

```
(+ (flip) (flip) (flip))
```

denotes a Binomial( $n = 3, p = \frac{1}{2}$ ) distribution and

```
(λ (x) (if (= 1 (flip)) x 0))
```

denotes the probability kernel  $x \mapsto \frac{1}{2}(\delta_x + \delta_0)$ , where  $\delta_r$  denotes the Dirac measure concentrated at the real  $r$ . Church is call-by-value and so

```
(= (flip) (flip))
```

denotes a Bernoulli distribution on `{true, false}`, while the application of the procedure

```
(λ (x) (= x x))
```

to the argument `(flip)`, written

```
((λ (x) (= x x)) (flip)),
```

denotes  $\delta_1$ . (For more examples, see [32].) In Scheme, one can modify the state of a non-local variable using mutation via the `set!` procedure. In other functional programming languages, non-local state may be implemented via other methods. (For example, in Haskell, one could use the state monad.) If an expression modifies its environment (using mutation or otherwise), it might not denote a fixed distribution. For example, a procedure may keep a counter variable and return an increasing sequence of integers on repeated calls.

Consider the Beta-Bernoulli process and the Pólya urn scheme written in Church. (For their mathematical characterization, see Section 2.1.) While these two processes look different, they induce the same distribution on sequences. We define two procedures, `sample-beta-coin` and `sample-pólya-coin`, such that the application of either procedure returns a procedure of no arguments whose application returns (random) binary values. Fix  $a, b > 0$ . Consider the following definitions of `sample-beta-coin` and `sample-pólya-coin` (and recall that the `(λ () ...)` special form creates a procedure of no arguments):

<p>(i)</p> <pre>(define (sample-beta-coin)   (let ((weight (beta a b)))     (lambda () (flip weight)) ) )</pre>	<p>(ii)</p> <pre>(define (sample-pólya-coin)   (let ((red a)         (total (+ a b)) )     (lambda () (let ((x (flip <math>\frac{\text{red}}{\text{total}}</math>)))                   (set! red (+ red x))                   (set! total (+ total 1))                   x ) ) )</pre>
---	--

The definitions

```
(define my-beta-coin (sample-beta-coin))
(define my-pólya-coin (sample-pólya-coin))
```

define a procedure `my-beta-coin` such that repeated applications (`my-beta-coin`) induce a random binary sequence, and similarly with `my-pólya-coin`.

Evaluating (`my-beta-coin`) returns a 1 with probability `weight` and a 0 otherwise, where the shared `weight` parameter is itself drawn from a  $\text{Beta}(a, b)$  distribution on  $[0, 1]$ . Note that the sequence of values obtained by evaluating (`my-beta-coin`) is exchangeable but not i.i.d. (e.g., an initial sequence of ten 1's leads one to predict that the next draw is more likely to be 1 than 0). However, conditioned on the `weight` (a random variable within the opaque procedure `my-beta-coin`) the sequence is i.i.d. A second random coin constructed by

```
(define your-beta-coin (sample-beta-coin))
```

will have a different weight (a.s.) and will generate a sequence that is independent of that generated by `my-beta-coin`. The sequence induced by repeated applications of `my-beta-coin` is exchangeable because applications of `flip` return independent samples.

The code in (ii) implements the Pólya urn scheme with  $a$  red balls and  $b$  black balls (see [15, Chap. 11.4]), and so the sequence of return values from `my-pólya-coin` is exchangeable. (One can see that the sequence is exchangeable by noting that its joint distribution depends only on the number of red and black balls and not on their order.)

Because the sequence induced by repeated applications of `my-pólya-coin` is exchangeable, de Finetti's theorem implies that its distribution is equivalent to that induced by i.i.d. draws from *some* random measure (in particular, the directing random measure). In the case of the Pólya urn scheme, the directing random measure is a random Bernoulli whose weight parameter has a  $\text{Beta}(a, b)$  distribution. Therefore (`sample-beta-coin`) denotes the de Finetti measure of the

sequence of repeated applications of `my-beta-coin` (and also of `your-beta-coin`, incidentally). The distributions of the sequences induced by repeated applications of `my-beta-coin` and `my-pólya-coin` are identical.

However, there is an important difference between these two implementations. While `(sample-beta-coin)` denotes the de Finetti measure, `(sample-pólya-coin)` does not, because `my-pólya-coin` does not denote a fixed distribution: the internal state of the procedure `my-pólya-coin` changes after each iteration, as the sufficient statistics are updated (using the mutation operator `set!`). Therefore, each element of the sequence of repeated applications of `my-pólya-coin` is generated from a different distribution. Even though the sequence of applications of `my-pólya-coin` has the same distribution as those given by repeated applications of `my-beta-coin`, the procedure `my-pólya-coin` denotes a probability kernel which depends on the state.

In contrast, `my-beta-coin` does not modify itself via mutation; after it is randomly initialized, the value of `weight` does not change during the execution of the program. Therefore, in each run of the program, `my-beta-coin` denotes a fixed distribution — a particular Bernoulli. The procedure `my-beta-coin` is precisely the directing random measure of the Beta-Bernoulli process produced by repeated applications of `my-beta-coin`.

An implementation of the computable de Finetti theorem (Theorem 2.3) transforms (ii) into a procedure which does not use mutation and whose application is equivalent in distribution to the evaluation of `(sample-beta-coin)`. In fact, assuming we know that the code generates a binary sequence (and therefore that the directing random measure will be a random Bernoulli), we can transform the de Finetti measure into a procedure that does not use mutation and whose application is equivalent in distribution to the evaluation of `(beta a b)`.

In the general case, given a program that generates an exchangeable sequence of reals, an implementation of the computable de Finetti proof can generate a mutation-free procedure `generated-code` such that

```
(define (sample-directing-random-measure)
  (let ((shared-randomness (uniform 0 1)))
    (lambda () (generated-code shared-randomness)) ) )
```

defines the de Finetti measure.

In addition to their simpler semantics, mutation-free procedures are often desirable for practical reasons. For example, having sampled the directing random measure, an exchangeable sequence of random variables can be efficiently sampled in parallel without the overhead necessary to communicate sufficient statistics.

### 6.3. Partial Exchangeability of Arrays and Other Data Structures

The example above involved binary sequences, but the computable de Finetti theorem can be used to transform implementations of real exchangeable sequences. Consider the following exchangeable sequence whose combinatorial structure is known as the Chinese restaurant process (see Aldous [14]). Let  $\alpha > 0$  be a computable real and let  $H$  be a computable distribution on  $\mathbf{R}$ . For  $n \geq 1$ , each  $X_n$  is sampled in turn according to the conditional distribution

$$\mathbb{P}[X_{n+1} \mid X_1, \dots, X_n] = \frac{n}{n + \alpha} \sum_{i=1}^n \delta_{X_i} + \frac{\alpha}{n + \alpha} H. \quad (65)$$

The sequence  $\{X_n\}_{n \geq 1}$  is exchangeable and the directing random measure is a Dirichlet process with parameter  $\alpha H$ . Given such a program, we can automatically recover the underlying Dirichlet process prior, samples from which produce random measures whose discrete structure was characterized by Sethuraman’s “stick-breaking construction” [41]. Note that the random measure is not produced in the same manner as Sethuraman’s construction and certainly is not of closed form. But the resulting mathematical objects have the same structure and distribution.

Sequences of random objects other than reals can sometimes be given de Finetti-type representations. For example, the Indian buffet process, defined by Griffiths and Ghahramani [42], is the combinatorial process underlying a *set-valued* exchangeable sequence that can be written in a way analogous to the Pólya urn in (ii). Just as the Chinese restaurant process gives rise to the Dirichlet process, the Indian buffet process gives rise to a non-homogeneous Beta-Bernoulli process. Unlike in the Chinese restaurant process example, which was a sequence of random reals, the computable de Finetti theorem is not applicable to exchangeable sequences of random sets. The directing random measure was identified as a random Bernoulli process, whose base measure is a Beta process (see Thibaux and Jordan [43]). A non-homogeneous Bernoulli process is a sequence of independent (but not necessarily identically distributed) Bernoulli random variables  $\{Z_i\}_{i \geq 1}$ , and the *Bernoulli sequence* is the random set  $\{n : Z_n = 1\}$ . For the Beta-Bernoulli process, the Bernoulli sequence is almost surely finite. A “stick-breaking construction” of the Beta-Bernoulli process given by Teh, Görür, and Ghahramani [44] is analogous to the code in (i), but gives only a  $\Delta_1$ -index for the Bernoulli sequence, rather than its canonical index (see Soare [45, II.2]). This observation was first noted by Roy et al. [39]. The computability of sampling independent Bernoulli sequences of a Beta-Bernoulli process remains open. However, an Indian buffet process on a *discrete* base measure can be classically transformed into an exchangeable sequence of *integer* indices (representing the finite subsets of the discrete support). If we are given such a representation, the

computable de Finetti theorem implies the existence of a computable de Finetti measure that gives canonical indices for the sets.

Combinatorial structures other than sequences have been given de Finetti-type representational theorems. For example, an array of random variables is called *separately* (or *jointly*) exchangeable when its distribution is invariant under (simultaneous) permutations of the rows and columns and their higher-dimensional analogues. Nearly fifty years after de Finetti's result, Aldous [46] and Hoover [47] showed that the entries of an infinite array satisfying either separate or joint exchangeability are conditionally i.i.d. These results have been connected with the theory of graph limits by Diaconis and Janson [48] and Austin [49] by considering the adjacency matrix of an exchangeable random graph.

As in the case of the Indian buffet process, structured probabilistic models in machine learning can often be represented in multiple ways, each of which has different advantages (e.g., efficient inference procedures, representational simplicity, compositionality, etc.). The infinite relational model of Kemp et al. [50] can be viewed as a partially exchangeable array, while the hierarchical stochastic block model built with the Mondrian process of Roy and Teh [51] is described using the Aldous-Hoover representation, making the conditional independence explicit. If the computable de Finetti theorem could be extended to partially exchangeable settings, it would provide analogous uniform transformations on a wider range of data structures.

### *Acknowledgements*

The research of Roy has been supported in part by an NSF Graduate Research Fellowship. Some of the results in this paper were presented at the *Computability in Europe* conference in Heidelberg, Germany, July 19–24, 2009, and an extended abstract [52] was published in the proceedings. The authors would like to thank Nate Ackerman, Vikash Mansinghka, Hartley Rogers, and the anonymous referees of the extended abstract for helpful comments.

### **References**

- [1] M. Braverman, S. Cook, Computing over the reals: foundations for scientific computing, *Notices Amer. Math. Soc.* 53 (3) (2006) 318–329. 2
- [2] V. Brattka, P. Hertling, K. Weihrauch, A tutorial on computable analysis, in: S. B. Cooper, B. Löwe, A. Sorbi (Eds.), *New computational paradigms: changing conceptions of what is computable*, Springer, Berlin, 2008. 2

- [3] A. Edalat, Domains for computation in mathematics, physics and exact real arithmetic, *Bull. Symbolic Logic* 3 (4) (1997) 401–452. [2](#)
- [4] P. Billingsley, *Probability and measure*, John Wiley & Sons Inc., New York, third edn., 1995. [3](#)
- [5] O. Kallenberg, *Foundations of modern probability*, Springer, New York, second edn., 2002. [3](#), [5](#), [6](#), [19](#)
- [6] O. Kallenberg, *Probabilistic symmetries and invariance principles*, Springer, New York, 2005. [3](#), [4](#)
- [7] A. P. Dawid, Intersubjective statistical models, in: *Exchangeability in probability and statistics* (Rome, 1981), North-Holland, Amsterdam, 217–232, 1982. [4](#)
- [8] S. L. Lauritzen, Extreme point models in statistics, *Scand. J. Statist.* 11 (2) (1984) 65–91. [4](#)
- [9] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio, *Atti della R. Accademia Nazionale dei Lincei, Ser. 6. Memorie, Classe di Scienze Fisiche, Matematiche e Naturali* 4 (1931) 251–299. [4](#)
- [10] E. Hewitt, L. J. Savage, Symmetric measures on Cartesian products, *Trans. Amer. Math. Soc.* 80 (1955) 470–501. [4](#)
- [11] C. Ryll-Nardzewski, On stationary sequences of random variables and the de Finetti’s equivalence, *Colloq. Math.* 4 (1957) 149–156. [4](#)
- [12] J. F. C. Kingman, Uses of exchangeability, *Ann. Probability* 6 (2) (1978) 183–197. [4](#)
- [13] P. Diaconis, D. Freedman, Partial exchangeability and sufficiency, in: *Statistics: applications and new directions* (Calcutta, 1981), Indian Statist. Inst., Calcutta, 205–236, 1984. [4](#)
- [14] D. J. Aldous, Exchangeability and related topics, in: *École d’été de probabilités de Saint-Flour, XIII—1983, Lecture Notes in Math.*, vol. 1117, Springer, Berlin, 1–198, 1985. [4](#), [26](#)
- [15] B. de Finetti, *Theory of probability*. Vol. 2, John Wiley & Sons Ltd., London, 1975. [5](#), [24](#)
- [16] H. Rogers, Jr., *Theory of recursive functions and effective computability*, MIT Press, Cambridge, MA, second edn., 1987. [7](#)
- [17] R. I. Soare, *Recursively enumerable sets and degrees*, *Perspectives in Mathematical Logic*, Springer-Verlag, Berlin, 1987. [7](#)
- [18] T. Grubba, M. Schröder, K. Weihrauch, Computable metrization, *Math. Logic Q.* 53 (4-5) (2007) 381–395. [7](#)
- [19] I. Battenfeld, M. Schröder, A. Simpson, A convenient category of domains, in: *Computation, meaning, and logic: articles dedicated to Gordon Plotkin*, vol. 172 of *Electron. Notes Theor. Comput. Sci.*, Elsevier, Amsterdam, 69–99, 2007. [8](#)
- [20] K. Weihrauch, *Computable analysis: an introduction*, Springer, Berlin, 2000. [9](#)
- [21] K. Weihrauch, X. Zheng, Computability on continuous, lower semi-continuous and upper semi-continuous real functions, *Theoret. Comput. Sci.* 234 (1-2) (2000) 109–133. [9](#)
- [22] M. Schröder, Admissible representations for probability measures, *Math. Logic Q.* 53 (4-5) (2007) 431–445. [9](#), [12](#), [18](#)
- [23] V. Bosserhoff, Notions of probabilistic computability on represented spaces, *J. of Universal Comput. Sci.* 14 (6) (2008) 956–995. [9](#), [18](#)
- [24] K. Weihrauch, Computability on the probability measures on the Borel sets of the unit interval, *Theoret. Comput. Sci.* 219 (1–2) (1999) 421–437. [9](#), [12](#), [18](#)
- [25] M. Schröder, A. Simpson, Representing probability measures using probabilistic processes, *J. Complex.* 22 (6) (2006) 768–782. [9](#), [18](#)
- [26] V. Brattka, G. Presser, Computability on subsets of metric spaces, *Theor. Comput. Sci.* 305 (1-3) (2003) 43–76. [10](#)
- [27] N. T. Müller, Computability on random variables, *Theor. Comput. Sci.* 219 (1-2) (1999) 287–299. [10](#), [12](#), [18](#)

- [28] K. Weihrauch, On computable metric spaces Tietze-Urysohn extension is computable, in: J. Blanck, V. Brattka, P. Hertling (Eds.), *Computability and Complexity in Analysis*, 4th International Workshop, CCA 2000, Swansea, UK, September 17-19, 2000, Selected Papers, vol. 2064 of *Lecture Notes in Comput. Sci.*, Springer, 357–368, 2000. [14](#)
- [29] M. B. Pour-El, J. I. Richards, *Computability in analysis and physics*, Springer, Berlin, 1989. [14](#)
- [30] A. Pfeffer, IBAL: A probabilistic rational programming language, in: *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence*, Morgan Kaufmann Publ., 733–740, 2001. [22](#)
- [31] S. Park, F. Pfenning, S. Thrun, A probabilistic language based on sampling functions, *ACM Trans. Program. Lang. Syst.* 31 (1) (2008) 1–46. [22](#)
- [32] N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, J. B. Tenenbaum, Church: a language for generative models, in: *Uncertainty in Artificial Intelligence*, 2008. [22](#), [23](#)
- [33] O. Kiselyov, C. Shan, Embedded probabilistic programming, in: W. M. Taha (Ed.), *Domain-Specific Languages*, IFIP TC 2 Working Conference, DSL 2009, Oxford, UK, July 15-17, 2009, *Proceedings*, vol. 5658 of *Lecture Notes in Comput. Sci.*, Springer, 360–384, 2009. [22](#)
- [34] M. Escardó, T. Streicher, Induction and recursion on the partial real line with applications to Real PCF, *Theoret. Comput. Sci.* 210 (1) (1999) 121–157. [22](#)
- [35] G. D. Plotkin, LCF considered as a programming language, *Theoret. Comput. Sci.* 5 (3) (1977/78) 223–255. [22](#)
- [36] M. Escardó, Semi-decidability of may, must and probabilistic testing in a higher-type setting, *Electron. Notes in Theoret. Comput. Sci.* 249 (2009) 219–242. [22](#)
- [37] N. Saheb-Djahromi, Probabilistic LCF, in: *Mathematical foundations of computer science*, 1978 (Proc. Seventh Sympos., Zakopane, 1978), vol. 64 of *Lecture Notes in Comput. Sci.*, Springer, Berlin, 442–451, 1978. [22](#)
- [38] N. Ramsey, A. Pfeffer, Stochastic lambda calculus and monads of probability distributions, *Proc. of the 29th ACM SIGPLAN-SIGACT Symp. on Principles of Program. Lang.* (2002) 154–165. [22](#)
- [39] D. M. Roy, V. K. Mansinghka, N. D. Goodman, J. B. Tenenbaum, A stochastic programming perspective on nonparametric Bayes, in: *Nonparametric Bayesian Workshop*, Int. Conf. on Machine Learning, 2008. [22](#), [26](#)
- [40] V. K. Mansinghka, Natively probabilistic computing, Ph.D. thesis, Massachusetts Institute of Technology, 2009. [22](#)
- [41] J. Sethuraman, A constructive definition of Dirichlet priors, *Statistica Sinica* 4 (1994) 639–650. [26](#)
- [42] T. L. Griffiths, Z. Ghahramani, Infinite latent feature models and the Indian buffet process, in: *Adv. in Neural Inform. Processing Syst.* 17, MIT Press, Cambridge, MA, 475–482, 2005. [26](#)
- [43] R. Thibaux, M. I. Jordan, Hierarchical beta processes and the Indian buffet process, in: *Proc. of the 11th Conf. on A.I. and Stat.*, 2007. [26](#)
- [44] Y. W. Teh, D. Görür, Z. Ghahramani, Stick-breaking construction for the Indian buffet process, in: *Proc. of the 11th Conf. on A.I. and Stat.*, 2007. [26](#)
- [45] R. I. Soare, *Recursively enumerable sets and degrees*, Springer, Berlin, 1987. [26](#)
- [46] D. J. Aldous, Representations for partially exchangeable arrays of random variables, *J. Multivariate Analysis* 11 (4) (1981) 581–598. [27](#)
- [47] D. N. Hoover, Relations on probability spaces and arrays of random variables, preprint, Institute for Advanced Study, Princeton, NJ, 1979. [27](#)
- [48] P. Diaconis, S. Janson, Graph limits and exchangeable random graphs, *Rendiconti di Matematica, Ser. VII* 28 (1) (2008) 33–61. [27](#)
- [49] T. Austin, On exchangeable random variables and the statistics of large graphs and hypergraphs, *Probab. Surv.* 5 (2008) 80–145. [27](#)

- [50] C. Kemp, J. Tenenbaum, T. Griffiths, T. Yamada, N. Ueda, Learning systems of concepts with an infinite relational model, in: Proc. of the 21st Nat. Conf. on Artificial Intelligence, 2006. [27](#)
- [51] D. M. Roy, Y. W. Teh, The Mondrian process, in: Adv. in Neural Inform. Processing Syst. 21, 2009. [27](#)
- [52] C. E. Freer, D. M. Roy, Computable exchangeable sequences have computable de Finetti measures, in: K. Ambos-Spies, B. Löwe, W. Merkle (Eds.), Mathematical Theory and Computational Practice (CiE 2009), Proc. of the 5th Conf. on Computability in Europe, vol. 5635 of *Lecture Notes in Comput. Sci.*, Springer, 218–231, 2009. [27](#)