Computable de Finetti measures

Cameron E. Freer[a], Daniel M. Roy[b]

[a]*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA*
[b]*Computer Science and Artificial Intelligence Laboratory,*
Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We prove a computable version of de Finetti's theorem on exchangeable sequences of real random variables. As a consequence, exchangeable stochastic processes expressed 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 computable version of de Finetti's theorem: the distribution of an exchangeable sequence of real random variables is computable if and only if its de Finetti measure is computable. 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.

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 that, with probability one, computes the de Finetti measure.

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 study of 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. Results in the bit-model also have direct implications for programs that manipulate probability distributions numerically. In many areas of statistics and computer science, especially machine learning, the objects of interest include 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, and the theory we present here is directly relevant to this setting.

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 a probabilistic functional programming language would need to use non-local communication (to access old 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* that samples 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 a *method* for constructing the desired program. In Section 6 we describe how an implementation of the computable de Finetti theorem would perform a code transformation that eliminates the use of non-local state in procedures that induce exchangeable stochastic processes.

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 ω to denote the set of nonnegative integers (as in logic), rather than an element of the basic probability space Ω (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-valued random variables. We say that X is *exchangeable* if, for every finite set $\{k_1, \ldots, k_j\}$ of distinct indices, $(X_{k_1}, \ldots, X_{k_j})$ is equal in distribution to $(X_1, \ldots, X_j)$.

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 ν on $\mathbf{R}$ such that $\{X_i\}_{i \geq 1}$ is conditionally i.i.d. with respect to ν. That is,*

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

Moreover, ν is a.s. unique and given by

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

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

The random measure ν is called the *directing random measure.*[1] Its distribution (a measure on probability measures), which we denote by μ, 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

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

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 a latent measure ν. Posterior analysis follows from placing a prior distribution on ν. 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 1937, de Finetti [10] extended the result to arbitrary real-valued exchangeable sequences. We will refer to this more general version as the *de Finetti theorem.* Later, Hewitt and Savage [11] extended the result to compact Hausdorff spaces, and Ryll-Nardzewski [12] introduced a weaker notion than exchangeability that suffices to give a conditionally i.i.d. representation. Hewitt and Savage [11] provide a history of the early developments, and a discussion of some subsequent extensions can be found in Kingman [13], Diaconis and Freedman [14], and Aldous [15]. 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 $\nu := \delta_p$, i.e., the Dirac measure concentrated on p. Then the de Finetti measure is the uniform distribution on Dirac measures on $[0, 1]$, and

---

[1] The directing random measure is only unique up to a null set, but it is customary to refer to it as if it were unique, as long as we only rely on almost-sure properties.

the corresponding exchangeable sequence is $p, p, \ldots$, i.e., a constant sequence, marginally uniformly distributed.

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

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

where $S_n := \sum_{i \leq n} x_i$, and where Γ is the Gamma function and α, β 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 [16, Chap. 11.4]. Each X_i is sampled in turn according to the conditional distribution

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

This process is often described as repeated sampling from an urn: starting with α red balls and β black balls, a ball is drawn at each stage uniformly at random, and then returned to the urn along with an additional ball of the same color. By de Finetti's theorem, there exists a random variable $\theta \in [0, 1]$ with respect to which the sequence is conditionally independent and $\mathbb{P}\{X_i = 1 \mid \theta\} = \theta$ for each i. In fact,

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

Furthermore, one can show that θ is $\text{Beta}(\alpha, \beta)$-distributed, and so the process given by the marginals (4) is called the Beta-Bernoulli process. Finally, the de Finetti measure is the distribution of the random Bernoulli measure $\theta\delta_1 + (1 - \theta)\delta_0$.

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 implementation of the Beta-Bernoulli process in a probabilistic programming language, see 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 χ be the distribution of a real-valued exchangeable sequence X, and let μ be the distribution of its directing random measure ν. Then μ is computable relative to χ, and χ is computable relative to μ. In particular, χ is computable if and only if μ 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 random variables derived from the directing random measure.

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 $\boldsymbol{\beta} \in \mathcal{B}_{\mathbf{R}}^k = \mathcal{B}_{\mathbf{R}} \times \cdots \times \mathcal{B}_{\mathbf{R}}$, we write $\boldsymbol{\beta}(i)$ to denote the ith coordinate of $\boldsymbol{\beta}$.

Let $X = \{X_i\}_{i \geq 1}$ be an exchangeable sequence of real random variables, with distribution χ and directing random measure ν. For every $\boldsymbol{\gamma} \in \mathcal{B}_{\mathbf{R}}$, we define a $[0,1]$-valued random variable $V_{\boldsymbol{\gamma}} := \nu\boldsymbol{\gamma}$. 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 μ (the distribution of ν).

Definition 2.4 (Mixed moments). Let $\{x_i\}_{i \in C}$ be a family of random variables indexed by a set C. The *mixed moments* of $\{x_i\}_{i \in C}$ are the expectations $\mathbb{E}\big(\prod_{i=1}^k x_{j(i)}\big)$, for $k \geq 1$ and $j \in C^k$.

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 $\{V_{\boldsymbol{\beta}}\}_{\boldsymbol{\beta} \in \mathcal{B}_{\mathbf{R}}}$.

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

For $k \geq 1$, let $\mathcal{L}_{\mathbf{R}^k}$ denote the set of finite unions of open rectangles in $\mathbf{R}^k$ (i.e., the lattice generated by $\mathcal{I}_{\mathbf{R}}^k$), and let $\mathcal{L}_{\mathbf{Q}^k}$ denote the set of finite unions of open rectangles in $\mathbf{Q}^k$. (Note that $\mathcal{I}_{\mathbf{Q}} \subsetneq \mathcal{L}_{\mathbf{Q}} \subsetneq \mathcal{L}_{\mathbf{R}} \subsetneq \mathcal{B}_{\mathbf{R}}$.) As we will show in Lemma 3.5, when χ is computable, we can enumerate all rational lower bounds on quantities of the form

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

where $k \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$.

In general, we cannot enumerate all rational upper bounds on (7). However, if $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$ (for $k \geq 1$) is such that, with probability one, ν places no mass on the boundary of any $\sigma(i)$, then $\mathbb{P}\big(\bigcap_{i=1}^k \{X_i \in \sigma(i)\}\big) = \mathbb{P}\big(\bigcap_{i=1}^k \{X_i \in \overline{\sigma(i)}\}\big)$, where $\overline{\sigma(i)}$ denotes the closure of $\sigma(i)$. In this case, for every rational upper bound q on (7), we have that $1 - q$ is a lower bound on

$$\mathbb{P}\big(\bigcup_{i=1}^k \{X_i \not\in \overline{\sigma(i)}\}\big), \tag{8}$$

a quantity for which we can enumerate all rational lower bounds. If this property holds for all $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$, then we can compute the mixed moments $\{V_\tau\}_{\tau \in \mathcal{L}_{\mathbf{Q}}}$. A natural condition that implies this property for all $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$ is that ν is a.s. continuous (i.e., with probability one, $\nu\{x\} = 0$ for every $x \in \mathbf{R}$).

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

3. Computable Representations

We begin by introducing notions of computability 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 standard notions of computability theory, such as computable and computably enumerable (c.e.) sets (see, e.g., Rogers [17] or Soare [18]). Recall that $r \in \mathbf{R}$ is a *c.e. real* (sometimes called a *left-c.e.* or *left-computable real*) when the set of all rationals less than r is a c.e. set. Similarly, r is a *co-c.e. real* (sometimes called a *right-c.e.* or *right-computable 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 $x \in S$, define the set $\mathcal{S}_x := \{B \in \mathcal{S} : x \in B\}$. Because S is T_0, we have $\mathcal{S}_x \neq \mathcal{S}_y$ when $x \neq y$, and so the set $\mathcal{S}_x$ uniquely determines the point x. It is therefore 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 $s : \omega \to \mathcal{S}$ be an enumeration of $\mathcal{S}$ (possibly with repetition), i.e., a total surjective (but not necessarily injective) function. We say that S is a *computable topological space (with respect to s)* when the set

$$\big\{ \langle m, n \rangle : s(m) = s(n) \big\} \tag{9}$$

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

This definition of a computable topological space is derived from Weihrauch's definition [21, Def. 3.2.1] in terms of "notations". (See also, e.g., Grubba, Schröder, and Weihrauch [19, Def. 3.1].)

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 $\mathcal{S}_x = \{B \in \mathcal{S} : x \in B\}$ 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 those that 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 [20, §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 (Names and computable points). Let $(S, \mathcal{S})$ be a computable topological space with respect to an enumeration s. Let $x \in S$. The set

$$\{n : s(n) \in \mathcal{S}_x\} = \{n : x \in s(n)\} \tag{10}$$

is called the *s-name* (or simply, *name*) of x. We say that x is computable when its s-name is c.e.

Note that this use of the term "name" is similar to the notion of a "complete name" (see [21, Lem. 3.2.3]), but differs somewhat from TTE usage (see [21, Def. 3.2.2]).

Definition 3.3 (Computable functions). Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be computable topological spaces (with respect to enumerations s and t, respectively). We say that a function $f : S \to T$ is *computable (with respect to s and t)* when there is a partial computable functional $g : \omega^\omega \to \omega^\omega$ such that for all $x \in \mathrm{dom}(f)$ and enumerations $N = \{n_i\}_{i \in \omega}$ of an s-name of x, we have that $g(N)$ is an enumeration of a t-name of $f(x)$.

(See [21, Def. 3.1.3] for more details.) Note that an implication of this definition is that computable functions are continuous.

Recall that a functional $g : \omega^\omega \to \omega^\omega$ is partial computable if there is a monotone computable function $h : \omega^{<\omega} \to \omega^{<\omega}$ mapping finite prefixes (of integer sequences) to finite prefixes, such that given increasing prefixes of an input N in the domain of g, the output of h will eventually include every finite prefix of $g(N)$. (See [21, Def. 2.1.11] for more details.) Informally, h can be used to read in an enumeration of an s-name of a point x and outputs an enumeration of a t-name of the point $f(x)$.

Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be computable topological spaces. In many situations where we are interested in establishing the computability of some function $f : S \to T$, we may refer to the function implicitly via pairs of points $x \in S$ and $y \in T$ related by $y = f(x)$. In this case, we will say that y (under the topology $\mathcal{T}$) is *computable relative to x* (under the topology $\mathcal{S}$) when $f : S \to T$ is a computable function. We will often elide one or both topologies when they are clear from context.

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 a straightforward effective enumeration; the computable points of this space are the computable reals. The reals under the *right order topology* are a computable topological space using the basis

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

under a standard enumeration; the computable points of this space are the c.e. reals.

Recall that, for $k \geq 1$, the set $\mathcal{I}_{\mathbf{Q}}^k$ is a basis for the (product of the) standard topology on $\mathbf{R}^k$ that is closed under intersection and makes $(\mathbf{R}^k, \mathcal{I}_{\mathbf{Q}}^k)$ a computable topological space (under a straightforward enumeration of $\mathcal{I}_{\mathbf{Q}}^k$). Likewise, an effective enumeration of cylinders $\sigma \times \mathbf{R}^\omega$, for $\sigma \in \bigcup_{k \geq 1} \mathcal{I}_{\mathbf{Q}}^k$, makes $\mathbf{R}^\omega$ a computable topological space. Replacing $\mathcal{I}_{\mathbf{Q}}$ with $\mathcal{R}_<$ and "standard" with "right order" above gives a characterization of computable vectors and sequences of reals under the right order topology.

We can use the right order topology to define a representation for open sets. Let $(S, \mathcal{S})$ be a computable topological space, with respect to an enumeration s. Then an open set $B \subseteq S$ is *c.e. open* when the indicator function $\mathbf{1}_B$ is computable with respect to $\mathcal{S}$ and $\mathcal{R}_<$. The c.e. open sets can be shown to be the computable points in the space of open sets under the Scott topology. Note that for the computable topological space ω (under the discrete topology and the identity enumeration) the c.e. open sets are precisely the c.e. sets of naturals.

3.2. Representations of Continuous Real Functions

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

Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ each be either of $(\mathbf{R}, \mathcal{I}_{\mathbf{Q}})$ or $(\mathbf{R}, \mathcal{R}_<)$, and let s and t be the associated enumerations. For $k \geq 1$, the compact-open topology on the space of continuous functions from S^k to T has a subbasis composed of sets of the form

$$\left\{ f \,:\, f(\overline{A}) \subseteq B \right\}, \tag{12}$$

where A and B are elements in the *bases* $\mathcal{S}^k$ and $\mathcal{T}$, respectively. An effective enumeration of this subbasis can be constructed in a straightforward fashion from s and t.

In particular, let $k \geq 1$ and let s^k be an effective enumeration of k-tuples of basis elements derived from s. Then a continuous function $f : (\mathbf{R}^k, \mathcal{S}^k) \to (\mathbf{R}, \mathcal{T})$ is computable (under the compact-open topology) when

$$\left\{ \langle m, n \rangle \,:\, f\big(\overline{s^k(m)}\big) \subseteq t(n) \right\} \tag{13}$$

is a c.e. set. The set (13) is the name of f.

A continuous function is computable in this sense if and only if it is computable according to Definition 3.3. (See [21, Ch. 6] and [21, Thm. 3.2.14]). 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 [22].

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 [23] and Bosserhoff [24], and agree with Weihrauch [25] 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 [21, Chap. 3]) to the canonical TTE representation for Borel measures given in Schröder [23].

Schröder [23] 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 [26]) formalizes a notion of a program that uses randomness to sample points in terms of their names of the form (10).

For a second-countable T_0 topological space S with subbasis $\mathcal{S}$, let $\mathcal{M}_1(S)$ denote the set of Borel probability measures on S (i.e., the probability measures on the σ-algebra generated by $\mathcal{S}$). Such measures are determined by the measure they assign to finite intersections of elements of $\mathcal{S}$. Note that $\mathcal{M}_1(S)$ is itself a second-countable T_0 space.

Now let $(S, \mathcal{S})$ be a computable topological space with respect to the enumeration s. We will describe a subbasis for $\mathcal{M}_1(S)$ that makes it a computable topological space. Let $\mathcal{L}_{\mathcal{S}}$ denote the lattice generated by $\mathcal{S}$ (i.e., the closure of $\mathcal{S}$ under finite union and intersection), and let $s^{\mathcal{L}}$ be an effective enumeration derived from s. Then, the class of sets

$$\{\gamma \in \mathcal{M}_1(S) \,:\, \gamma\sigma > q\}, \tag{14}$$

where $\sigma \in \mathcal{L}_{\mathcal{S}}$ 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 straightforward fashion from the enumeration of $\mathcal{S}$ and an effective enumeration $\{q_n\}_{n \in \omega}$ of the rationals, making $\mathcal{M}_1(S)$ a computable topological space. In particular, the name of a measure $\eta \in \mathcal{M}_1(S)$ is the set $\{\langle m, n \rangle \,:\, \eta(s^{\mathcal{L}}(m)) > q_n\}$.

Corollary 3.4 (Computable distribution). *A Borel probability measure $\eta \in \mathcal{M}_1(S)$ is* computable (under the weak topology) *if and only if ηB is a c.e. real, uniformly in the $s^{\mathcal{L}}$-index of $B \in \mathcal{L}_{\mathcal{S}}$.* □

Note that, for computable topological spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$ with enumerations s and t, a measure $\eta \in \mathcal{M}_1(T)$ is computable relative to a point $x \in S$ when ηB is a c.e. real relative to x, uniformly in the $t^{\mathcal{L}}$-index of $B \in \mathcal{L}_{\mathcal{T}}$. Corollary 3.4 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., [27, §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 (see Weihrauch [25] for a discussion; moreover, such a requirement is stronger than what is necessary to ensure that a, e.g., probabilistic Turing machine can produce exact samples to arbitrary accuracy).

We will be interested in computable measures in $\mathcal{M}_1(S)$, where S is either $\mathbf{R}^\omega$, $[0, 1]^k$, or $\mathcal{M}_1(\mathbf{R})$. In order to apply Corollary 3.4 to characterize concrete

notions of computability for $\mathcal{M}_1(S)$, we will now describe choices of topologies on these three spaces.

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

Using Corollary 3.4, we can characterize the class of computable distributions on real sequences using the computable topological spaces characterized above in Section 3.1. 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 $\mathcal{I}_\mathbf{Q}$), and let η be the joint distribution of $\vec{x}$. Then η is computable if and only if $\eta(\sigma \times \mathbf{R}^\omega) = \mathbb{P}\{x \in \sigma \times \mathbf{R}^\omega\}$ is a c.e. real, uniformly in $k \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}^k}$. The following simpler characterization was given by Müller [28, 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 η. Then η is computable if and only if

$$\eta(\tau \times \mathbf{R}^\omega) = \mathbb{P}\big(\bigcap_{i=1}^k \{x_i \in \tau(i)\}\big) \tag{15}$$

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 $\bigcup_k \mathcal{I}_\mathbf{Q}^k \subsetneq \bigcup_k \mathcal{L}_{\mathbf{Q}^k}$ is sufficient. Note that the right-hand side of (15) is precisely the form of the left-hand side of the expression in Corollary 2.5. Note also that one obtains a characterization of the computability of a finite-dimensional vector by embedding it as an initial segment of a sequence.

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

Borel measures on $\mathbf{R}$ under the right order topology play an important role when representing measures on measures, as Corollary 3.4 portends.

Corollary 3.6 (Computable distribution under the right order topology).
Let $\vec{x} = \{x_i\}_{i \geq 1}$ be a sequence of real-valued random variables with joint distribution η. Then η is computable under the (product of the) right order topology if and only if

$$\eta\big(\bigcup_{i=1}^m \big((c_{i1}, \infty) \times \cdots \times (c_{ik}, \infty) \times \mathbf{R}^\omega\big)\big) = \mathbb{P}\big(\bigcup_{i=1}^m \bigcap_{j=1}^k \{x_j > c_{ij}\}\big) \tag{16}$$

is a c.e. real, uniformly in $k, m \geq 1$ and $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$. □

Again, one obtains a characterization of the computability of a finite-dimensional vector by embedding it as an initial segment of a sequence. 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 characterization 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 coarser than the standard topology.

3.3.3. Measures on Borel Measures

The de Finetti measure μ is the distribution of the directing random measure ν, an $\mathcal{M}_1(\mathbf{R})$-valued random variable. Recall the definition $V_{\boldsymbol{\beta}} := \nu\boldsymbol{\beta}$, for $\boldsymbol{\beta} \in \mathcal{B}_{\mathbf{R}}$. From Corollary 3.4, it follows that μ is computable under the weak topology if and only if

$$\mu(\textstyle\bigcup_{i=1}^{m} \bigcap_{j=1}^{k}\{\gamma \in \mathcal{M}_1(\mathbf{R}) \,:\, \gamma\sigma(j) > c_{ij}\}) = \mathbb{P}(\textstyle\bigcup_{i=1}^{m} \bigcap_{j=1}^{k}\{V_{\sigma(j)} > c_{ij}\}) \tag{17}$$

is a c.e. real, uniformly in $k, m \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$ and $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$. As an immediate consequence of (17) and Corollary 3.6, we obtain the following characterization of computable de Finetti measures.

Corollary 3.7 (Computable de Finetti measure). *The de Finetti measure μ is computable relative to the joint distribution of $\{V_\tau\}_{\tau \in \mathcal{L}_{\mathbf{Q}}}$ under the right order topology, and vice versa. In particular, μ is computable if and only if the joint distribution of $\{V_\tau\}_{\tau \in \mathcal{L}_{\mathbf{Q}}}$ is computable under the right order topology.* $\square$

3.3.4. Integration

The following lemma is a restatement of an integration result by Schröder [23, Prop. 3.6], which itself generalizes integration results on standard topologies of finite-dimensional Euclidean spaces by Müller [28] and the unit interval by Weihrauch [25].

Define

$$\mathbb{I} := \{A \cap [0,1] \,:\, A \in \mathcal{I}_{\mathbf{Q}}\}, \tag{18}$$

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

$$\mathbb{I}_< := \{A \cap [0,1] \,:\, A \in \mathcal{R}_<\}, \tag{19}$$

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

Lemma 3.8 (Integration of bounded lower-semicontinuous functions).
Let $k \geq 1$ and let $\mathcal{S}$ be either $\mathcal{I}_{\mathbf{Q}}$ or $\mathcal{R}_<$. Let

$$f : (\mathbf{R}^k, \mathcal{S}^k) \to ([0,1], \mathbb{I}_<) \tag{20}$$

be a continuous function and let μ be a Borel probability measure on $(\mathbf{R}^k, \mathcal{S}^k)$. Then

$$\int f \, d\mu \tag{21}$$

is a c.e. real relative to f and μ. $\square$

13

The following result of Müller [28] is an immediate corollary.

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

$$g : (\mathbf{R}^k, \mathcal{I}_{\mathbf{Q}}^k) \to ([0,1], \mathbb{I}) \tag{22}$$

be a continuous function and let μ be a Borel probability measure on $(\mathbf{R}^k, \mathcal{I}_{\mathbf{Q}}^k)$. Then

$$\int g \, d\mu \tag{23}$$

is a computable real relative to g and μ. □

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 η. 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 $\phi_{\vec{x}} : \mathbf{R}^\omega \to \mathbf{C}$, given by

$$\phi_{\vec{x}}(t) := \mathbb{E}(e^{i \langle t, \vec{x} \rangle}), \tag{24}$$

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 a computable version of the Weierstrass approximation theorem to compute the distribution relative to the mixed moments.

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

Lemma 4.1 (Polynomial approximations). *Let $k \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}^k}$. There is a sequence*

$$\{p_{n,\sigma} : n \in \omega\} \tag{25}$$

of rational polynomials of degree k, computable uniformly in n, k, and σ, such that, for all $\vec{x} \in [0,1]^k$, we have

$$-2 \leq p_{n,\sigma}(\vec{x}) \leq \mathbf{1}_\sigma(\vec{x}) \qquad and \qquad \lim_{m \to \infty} p_{m,\sigma}(\vec{x}) = \mathbf{1}_\sigma(\vec{x}). \tag{26}$$

14

PROOF. Let $k \geq 1$. For $\sigma \in \mathcal{L}_{\mathbf{Q}^k}$, and $\vec{x} \in \mathbf{R}^k$, define $d(\vec{x}, [0,1]^k \setminus \sigma)$ to be the distance from $\vec{x}$ to the nearest point in $[0,1]^k \setminus \sigma$. It is straightforward to show that $d(\vec{x}, [0,1]^k \setminus \sigma)$ is a computable real function of $\vec{x}$, uniformly in k and σ.

For $n \in \omega$, define $f_{n,\sigma} : \mathbf{R}^k \to \mathbf{R}$ by

$$f_{n,\sigma}(\vec{x}) := -\frac{1}{n+1} + \min\{1, n \cdot d(\vec{x}, [0,1]^k \setminus \sigma)\}, \tag{27}$$

and note that $-1 \leq f_{n,\sigma}(\vec{x}) \leq \mathbf{1}_\sigma(\vec{x}) - \frac{1}{n+1}$ and $\lim_{m \to \infty} f_{m,\sigma}(\vec{x}) = \mathbf{1}_\sigma(\vec{x})$. Furthermore, $f_{n,\sigma}(\vec{x})$ is a computable (hence continuous) real function of $\vec{x}$, uniformly in n, k, and σ.

By the effective Weierstrass approximation theorem (see Pour-El and Richards [30, p. 45]), we can find (uniformly in n, k, and σ) a polynomial $p_{n,\sigma}$ with rational coefficients that uniformly approximates $f_{n,\sigma}$ to within $1/(n+1)$ on $[0,1]^k$. These polynomials have the desired properties. $\square$

We thank the anonymous referee for suggestions that simplified the proof of this lemma.

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 η. Then η is computable relative to the mixed moments of $\{x_i\}_{i \in \omega}$, and vice versa. In particular, η is computable if and only if the mixed moments of $\{x_i\}_{i \in \omega}$ are uniformly computable.*

PROOF. Any monic monomial in k variables, considered as a real function, computably maps $[0,1]^k$ into $[0,1]$ (under the standard topology). Furthermore, as the restriction of η to any k coordinates is computable relative to η (uniformly in the coordinates), it follows from Corollary 3.9 that each mixed moment (the expectation of a monomial under such a restriction of η) is computable relative to η, uniformly in the index of the monomial and the coordinates.

Let $k \geq 1$ and $\sigma \in \mathcal{I}_{\mathbf{Q}}^k$. To establish the computability of η, it suffices to show that

$$\eta(\sigma \times [0,1]^\omega) = \mathbb{E}\big(\mathbf{1}_{\sigma \times [0,1]^\omega}(\vec{x})\big) = \mathbb{E}\big(\mathbf{1}_\sigma(x_1, \ldots, x_k)\big). \tag{28}$$

is a c.e. real relative to the mixed moments, uniformly in k and σ. By Lemma 4.1, there is a uniformly computable sequence of polynomials $(p_{n,\sigma})_{n \in \omega}$ that converge pointwise from below to the indicator $\mathbf{1}_\sigma$. Therefore, by the dominated convergence theorem,

$$\mathbb{E}\big(\mathbf{1}_\sigma(x_1, \ldots, x_k)\big) = \sup_n \mathbb{E}\big(p_{n,\sigma}(x_1, \ldots, x_k)\big). \tag{29}$$

15

The expectation $\mathbb{E}\big(p_{n,\sigma}(x_1,\ldots,x_k)\big)$ is a $\mathbf{Q}$-linear combination of mixed moments, hence a computable real relative to the mixed moments, uniformly in n, k, and σ. Thus the supremum (29) is a c.e. real relative to the mixed moments, uniformly in k and σ. □

5. Proof of the Computable de Finetti Theorem

For the remainder of the paper, let X be a real-valued exchangeable sequence with distribution χ, let ν be its directing random measure, and let μ be the corresponding 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* relative to the de Finetti measure.

Proposition 5.1. *The distribution χ is computable relative to μ.*

PROOF. Let $k \geq 1$ and $\sigma \in \mathcal{I}_{\mathbf{Q}}^k$. All claims are uniform in k and σ. In order to show that χ, the distribution of X, is computable relative to μ, we must show that $\mathbb{P}\big(\bigcap_{i=1}^k \{X_i \in \sigma(i)\}\big)$ is a c.e. real relative to μ. Note that, by Corollary 2.5,

$$\mathbb{P}\big(\textstyle\bigcap_{i=1}^k \{X_i \in \sigma(i)\}\big) = \mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\sigma(i)}\big). \tag{30}$$

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

$$f(x_1,\ldots,x_k) := \textstyle\prod_{i=1}^k x_i. \tag{31}$$

To complete the proof, we now show that

$$\int f \, d\eta = \mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\sigma(i)}\big) \tag{32}$$

is a c.e. real relative to μ. Note that η is computable under the right order topology relative to μ. Furthermore, f is order-preserving (in each dimension) and lower-semicontinuous, i.e., 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 relative to μ. □

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 that 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 $\boldsymbol{\psi} \in \mathcal{L}_{\mathbf{R}}^k$, we say that $\boldsymbol{\psi}$ is a ν-continuity set when, for $i \leq k$, we have $\nu(\partial\boldsymbol{\psi}(i)) = 0$ a.s., where $\partial\boldsymbol{\psi}(i)$ denotes the boundary of $\boldsymbol{\psi}(i)$.

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

PROOF. Let $k \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$. All claims are uniform in k and σ. By Corollary 2.5,

$$\mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\sigma(i)}\big) = \mathbb{P}\big(\textstyle\bigcap_{i=1}^k \{X_i \in \sigma(i)\}\big), \tag{33}$$

which is a c.e. real relative to χ. The set $\overline{\sigma}$ is a co-c.e. closed set in $\mathbf{R}^k$ because we can computably enumerate all $\tau \in \mathcal{L}_{\mathbf{Q}}^k$ contained in the complement of σ. Therefore,

$$\mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\overline{\sigma(i)}}\big) = \mathbb{P}\big(\textstyle\bigcap_{i=1}^k \{X_i \in \overline{\sigma(i)}\}\big) \tag{34}$$

is the measure of a co-c.e. closed set, hence a co-c.e. real relative to χ. When σ is a ν-continuity set,

$$\mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\sigma(i)}\big) = \mathbb{E}\big(\textstyle\prod_{i=1}^k V_{\overline{\sigma(i)}}\big), \tag{35}$$

and so the expectation is a computable real relative to χ. $\qquad\square$

Proposition 5.3 (Almost surely continuous directing random measure).
Assume that ν is almost surely continuous. Then μ is computable relative to χ.

PROOF. Let $k \geq 1$ and $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$. The almost sure continuity of ν implies that σ is an ν-continuity set. Therefore, by Lemma 5.2, the moment $\mathbb{E}\big(\prod_{i=1}^k V_{\sigma(i)}\big)$ is a computable real relative to χ, uniformly in k and σ. The computable moment theorem (Theorem 4.2) then implies that the joint distribution of the variables $\{V_\tau\}_{\tau \in \mathcal{L}_{\mathbf{Q}}}$ is computable under the standard topology relative to χ, and so their joint distribution is also computable under the (coarser) right order topology relative to χ. By Corollary 3.7, this implies that μ is computable relative to χ.
$\square$

5.2. "Randomized" Proof Sketch

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

Consider the following sketch of a "randomized algorithm": We independently sample a countably infinite sequence of real numbers $\mathbf{A}$ from a computable, absolutely continuous distribution that has support everywhere on the real line (e.g., a Gaussian or Cauchy). Let $\mathcal{L}_\mathbf{A}$ denote the lattice generated by open intervals with endpoints in $\mathbf{A}$. Note that, with probability one, $\mathbf{A}$ will be dense in $\mathbf{R}$ and every $\psi \in \mathcal{L}_\mathbf{A}$ will be a ν-continuity set. If the algorithm proceeds analogously to the case where ν is almost surely continuous, using $\mathcal{L}_\mathbf{A}$ as our basis, rather than $\mathcal{L}_\mathbf{Q}$, then it will compute the de Finetti measure with probability one.

Let A be a dense sequence of reals such that $\nu(A) = 0$ a.s. Consider the variables V_ζ defined in terms of elements ζ of the new basis $\mathcal{L}_A$ (defined analogously to $\mathcal{L}_\mathbf{A}$). We begin by proving an extension of Lemma 5.2: The mixed moments of the set of variables $\{V_\zeta\}_{\zeta \in \mathcal{L}_A}$ are computable relative to A and χ.

Lemma 5.4. *Let $k \geq 1$ and $\psi \in \mathcal{L}_A^k$. The mixed moment $\mathbb{E}\big(\prod_{i=1}^k V_{\psi(i)}\big)$ is a computable real relative to A and χ, uniformly in k and ψ.*

PROOF. Let $k \geq 1$ and $\psi \in \mathcal{L}_A^k$. All claims are uniform in k and ψ. We first show that, relative to A and χ, the mixed moments of $\{V_\zeta\}_{\zeta \in \mathcal{L}_A}$ are uniformly c.e. reals. We can compute (relative to A) a sequence

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

such that componentwise for each $n \geq 1$,

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

Note that if $\zeta, \varphi \in \mathcal{L}_\mathbf{Q}$ satisfy $\zeta \subseteq \varphi$, then $V_\zeta \leq V_\varphi$ (a.s.), and so, by the continuity of measures (and of multiplication), $\prod_{i=1}^k V_{\sigma_n(i)}$ converges from below to $\prod_{i=1}^k V_{\psi(i)}$ with probability one. Therefore, the dominated convergence theorem gives us

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

Using Corollary 2.5, we see that the expectation $\mathbb{E}\big(\prod_{i=1}^k V_{\sigma_n(i)}\big)$ is a c.e. real relative to A and χ, uniformly in n, and so the supremum (38) is a c.e. real relative to A and χ.

18

Similarly, the mixed moments of $\{V_{\bar{\zeta}}\}_{\zeta \in \mathcal{L}_A}$ are uniformly co-c.e. reals relative to A and χ, as can be seen via a sequence of nested unions of rational intervals whose union has complement equal to $\overline{\psi}$. Thus, because ψ is a ν-continuity set, the mixed moment $\mathbb{E}\big(\prod_{i=1}^{k} V_{\psi(i)}\big)$ is a computable real relative to A and χ. $\qquad \square$

Lemma 5.5. *The de Finetti measure μ is computable relative to A and χ.*

PROOF. It follows immediately from Lemma 5.4 and Theorem 4.2 that the joint distribution of $\{V_\psi\}_{\psi \in \mathcal{L}_A}$ is computable relative to A and χ. This joint distribution classically determines the de Finetti measure. Moreover, as we now show, we can compute (relative to A and χ) the desired representation with respect to the (original) rational basis. In particular, we prove that the joint distribution of $\{V_\tau\}_{\tau \in \mathcal{L}_\mathbf{Q}}$ is computable under the right order topology relative to A and χ.

Let $m, k \geq 1$, let $\tau \in \mathcal{L}_\mathbf{Q}^k$, and let $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$. We will express τ as a union of elements of $\mathcal{L}_A^k$. Note that τ is an c.e. open set (relative to A) with respect to the basis $\mathcal{L}_A^k$. In particular, we can computably enumerate (relative to A, and uniformly in k and τ) a sequence $\sigma_1, \sigma_2, \ldots \in \mathcal{L}_A^k$ such that $\cup_n \sigma_n = \tau$ and $\sigma_n \subseteq \sigma_{n+1}$. Note that $V_{\tau(j)} \geq V_{\sigma_n(j)}$ (a.s.) for all $n \geq 1$ and $j \leq k$. By the continuity of measures (and of union and intersection),

$$\mathbb{P}\big(\textstyle\bigcup_{i=1}^{m} \bigcap_{j=1}^{k} \{V_{\tau(j)} > c_{ij}\}\big) = \sup_n \mathbb{P}\big(\textstyle\bigcup_{i=1}^{m} \bigcap_{j=1}^{k} \{V_{\sigma_n(j)} > c_{ij}\}\big). \tag{39}$$

The probability $\mathbb{P}\big(\bigcup_{i=1}^{m} \bigcap_{j=1}^{k} \{V_{\sigma_n(j)} > c_{ij}\}\big)$ is a c.e. real relative to A and χ, uniformly in n, m, k, τ, and C, and so the supremum (39) is a c.e. real relative to A and χ, uniformly in m, k, τ, and C. $\qquad \square$

Let Φ denote the map taking (A, χ) to μ, as described in Lemma 5.5.

Recall that $\mathbf{A}$ is a random dense sequence with a computable distribution, as defined above, and let $\hat{\mu} = \Phi(\mathbf{A}, \chi)$. Then $\hat{\mu}$ is a random variable, and moreover, $\hat{\mu} = \mu$ almost surely. However, while $\mathbf{A}$ is almost surely noncomputable, the distribution of $\mathbf{A}$ is computable, and so the distribution of $\hat{\mu}$ is computable relative to χ. Expectations with respect to the distribution of $\hat{\mu}$ can then be used to (deterministically) compute μ relative to χ.

A proof along these lines could be made precise by making $\mathcal{M}_1(\mathcal{M}_1(\mathcal{M}_1(\mathbf{R})))$ into a computable topological space. Instead, in Section 5.3, we complete the proof by explicitly computing μ relative to χ in terms of the standard 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 [26]) that samples a name of the de Finetti measure with probability one. Schröder [23] 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 [25, Thm. 3.6] proves a computable integration result via an argument that could likewise be seen as a derandomization of an algorithm that densely subdivides the unit interval at random locations to find continuity sets. Bosserhoff [24, 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 [28, 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 ϵ. These arguments also resemble the proof of the classical 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

Let $m, k \geq 1$ and $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$. By an abuse of notation, we define

$$\mathbf{1}_C : [0,1]^k \to [0,1] \tag{40}$$

to be the indicator function for the set

$$\bigcup_{i=1}^m (c_{i1}, 1] \times \cdots \times (c_{ik}, 1]. \tag{41}$$

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

$$\sigma := \bigcup_{i=1}^m (c_{i1}, 2) \times \cdots \times (c_{ik}, 2) \in \mathcal{L}_{\mathbf{Q}^k}. \tag{42}$$

Here, we have arbitrarily chosen $2 > 1$ 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 $\vec{x} = (x_1, \ldots, x_k)$ and $\vec{y} = (y_1, \ldots, y_k)$. We can write

$$p_{n,C}(\vec{x}) = p_{n,C}^+(\vec{x}) - p_{n,C}^-(\vec{x}), \tag{43}$$

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

$$q_{n,C}(\vec{x}, \vec{y}) := p_{n,C}^+(\vec{x}) - p_{n,C}^-(\vec{y}). \tag{44}$$

We denote

$$q_{n,C}(V_{\boldsymbol{\varphi}(1)}, \ldots, V_{\boldsymbol{\varphi}(k)}, V_{\boldsymbol{\zeta}(1)}, \ldots, V_{\boldsymbol{\zeta}(k)}) \tag{45}$$

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

20

Proposition 5.6. *Let $n \in \omega$, let $k, m \geq 1$, let $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$, and let $C \in \mathbf{Q}^{m \times k}$. Then $\mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\sigma}})$ is a c.e. real relative to χ, uniformly in n, k, m, σ, and C.*

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

In the final proof we use the following dense partial order on products of $\mathcal{L}_{\mathbf{R}}$.

Definition 5.7. Let $k \geq 1$. We call $\boldsymbol{\psi} \in \mathcal{L}_{\mathbf{R}}^k$ a *refinement* of $\boldsymbol{\varphi} \in \mathcal{L}_{\mathbf{R}}^k$, and write $\boldsymbol{\psi} \lhd \boldsymbol{\varphi}$, when

$$\overline{\boldsymbol{\psi}(i)} \subseteq \boldsymbol{\varphi}(i) \tag{46}$$

for all $i \leq k$.

We are now ready to prove the main theorem.

PROOF OF THEOREM 2.3 (COMPUTABLE DE FINETTI). The distribution χ (of the exchangeable sequence X) is computable relative to the de Finetti measure μ by Proposition 5.1. We now give a proof of the other direction, showing that the joint distribution of $\{V_\sigma\}_{\sigma \in \mathcal{L}_{\mathbf{Q}}}$ is computable under the right order topology relative to χ, which by Corollary 3.7 will complete the proof.

Let $k, m \geq 1$, let $\pi \in \mathcal{L}_{\mathbf{Q}}^k$, and let $C = (c_{ij}) \in \mathbf{Q}^{m \times k}$. For $\boldsymbol{\zeta} \in \mathcal{L}_{\mathbf{R}}^k$, let $V_{\boldsymbol{\zeta}}$ denote the k-tuple $(V_{\boldsymbol{\zeta}(1)}, \ldots, V_{\boldsymbol{\zeta}(k)})$ and similarly for $V_{\overline{\boldsymbol{\zeta}}}$. Take $\mathbf{1}_C$ to be defined as above in (40) and (41). It suffices to show that

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

is a c.e. real relative to χ, uniformly in k, m, π, and C. We do this by a series of reductions, which results in a supremum over quantities of the form $\mathbb{E}q_{n,C}(V_\sigma, V_{\overline{\sigma}})$ for $\sigma \in \mathcal{L}_{\mathbf{Q}}^k$.

By the density of the reals and the continuity of measures, we have that

$$V_\pi = \sup_{\boldsymbol{\psi} \lhd \pi} V_{\boldsymbol{\psi}} \quad \text{a.s.,} \tag{48}$$

where $\boldsymbol{\psi}$ ranges over $\mathcal{L}_{\mathbf{R}}^k$. It follows that

$$\mathbf{1}_C(V_\pi) = \sup_{\boldsymbol{\psi} \lhd \pi} \mathbf{1}_C(V_{\boldsymbol{\psi}}) \quad \text{a.s.,} \tag{49}$$

21

because $\mathbf{1}_C$ is lower-semicontinuous and order-preserving (in each dimension), as (41) is an open set in the right order topology on $[0,1]^k$. Therefore, by the dominated convergence theorem, we have that

$$\mathbb{E}\mathbf{1}_C(V_\pi) = \sup_{\boldsymbol{\psi} \triangleleft \pi} \mathbb{E}\mathbf{1}_C(V_{\boldsymbol{\psi}}). \tag{50}$$

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,

$$\mathbb{E}\mathbf{1}_C(V_{\boldsymbol{\psi}}) = \sup_n \mathbb{E}p_{n,C}(V_{\boldsymbol{\psi}}). \tag{51}$$

As $V_{\overline{\boldsymbol{\psi}(i)}} \geq V_{\boldsymbol{\psi}(i)}$ a.s. for $i \leq k$, we have that

$$\mathbb{E}p_{n,C}(V_{\boldsymbol{\psi}}) = \mathbb{E}p_{n,C}^+(V_{\boldsymbol{\psi}}) - \mathbb{E}p_{n,C}^-(V_{\boldsymbol{\psi}}) \tag{52}$$

$$\geq \mathbb{E}p_{n,C}^+(V_{\boldsymbol{\psi}}) - \mathbb{E}p_{n,C}^-(V_{\overline{\boldsymbol{\psi}}}). \tag{53}$$

Note that if $\boldsymbol{\psi}$ is a ν-continuity set, then $V_{\overline{\boldsymbol{\psi}(i)}} = V_{\boldsymbol{\psi}(i)}$ a.s., and so

$$\mathbb{E}p_{n,C}(V_{\boldsymbol{\psi}}) = \mathbb{E}p_{n,C}^+(V_{\boldsymbol{\psi}}) - \mathbb{E}p_{n,C}^-(V_{\overline{\boldsymbol{\psi}}}). \tag{54}$$

Again, dominated convergence theorem gives us

$$\mathbb{E}\left(\prod_{i=1}^k V_{\boldsymbol{\psi}(i)}\right) = \sup_{\sigma \triangleleft \boldsymbol{\psi}} \mathbb{E}\left(\prod_{i=1}^k V_{\sigma(i)}\right) \quad \text{and} \tag{55}$$

$$\mathbb{E}\left(\prod_{i=1}^k V_{\overline{\boldsymbol{\psi}(i)}}\right) = \inf_{\tau \triangleright \boldsymbol{\psi}} \mathbb{E}\left(\prod_{i=1}^k V_{\overline{\tau(i)}}\right), \tag{56}$$

where σ and τ range over $\mathcal{L}_{\mathbf{Q}}^k$. Therefore, by the linearity of expectation,

$$\mathbb{E}p_{n,C}^+(V_{\boldsymbol{\psi}}) = \sup_{\sigma \triangleleft \boldsymbol{\psi}} \mathbb{E}p_{n,C}^+(V_\sigma) \quad \text{and} \tag{57}$$

$$\mathbb{E}p_{n,C}^-(V_{\overline{\boldsymbol{\psi}}}) = \inf_{\tau \triangleright \boldsymbol{\psi}} \mathbb{E}p_{n,C}^-(V_{\overline{\tau}}), \tag{58}$$

and so, if $\boldsymbol{\psi}$ is a ν-continuity set, we have that

$$\mathbb{E}p_{n,C}(V_{\boldsymbol{\psi}}) = \sup_{\sigma \triangleleft \boldsymbol{\psi}} \mathbb{E}p_{n,C}^+(V_\sigma) - \inf_{\tau \triangleright \boldsymbol{\psi}} \mathbb{E}p_{n,C}^-(V_{\overline{\tau}}) \tag{59}$$

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

Because ν has at most countably many point masses, those $\boldsymbol{\psi} \in \mathcal{I}_{\mathbf{R}}^k$ that are ν-continuity sets are dense in $\mathcal{I}_{\mathbf{Q}}^k$. On the other hand, for those $\boldsymbol{\psi}$ that are not ν-continuity sets, (60) is a lower bound, as can be shown from (53). Therefore,

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

22

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

$$\sup_{\boldsymbol{\psi} \lhd \pi} \sup_{\sigma \lhd \boldsymbol{\psi}} \sup_{\tau \rhd \boldsymbol{\psi}} \mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\tau}}) = \sup_{\sigma \lhd \pi} \sup_{\tau \rhd \sigma} \mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\tau}}). \tag{62}$$

Again by dominated convergence we have

$$\sup_{\tau \rhd \sigma} \mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\tau}}) = \mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\sigma}}). \tag{63}$$

Combining (47), (50), (51), (61), (62), and (63), we have

$$\mathbb{E} \mathbf{1}_C(V_\pi) = \sup_n \sup_{\sigma \lhd \pi} \mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\sigma}}). \tag{64}$$

Finally, by Proposition 5.6, the expectation

$$\mathbb{E} q_{n,C}(V_\sigma, V_{\overline{\sigma}}) \tag{65}$$

is a c.e. real relative to χ, uniformly in σ, n, k, m, π, and C. Hence the supremum (64) is a c.e. real relative to χ, uniformly in k, m, π, and C. □

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 eliminate modifications of non-local 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 [31], $\lambda_\circ$[32], Church [33], and HANSEI [34]). Within domain theory, researchers have considered idealized functional languages that can manipulate exact real numbers, such as Escardó's REALPCF+ [35] (based on Plotkin [36]), and functional languages have also been extended by probabilistic choice operators (e.g., by Escardó [37] and Saheb-Djahromi [38]).

The semantics of probabilistic programs have been studied extensively in theoretical computer science in the context of randomized algorithms, probabilistic
23

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 unobserved 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.

A wide variety of algorithms implement conditional inference in probabilistic functional programming. Goodman et al. [33] describe the language Church, which extends a pure subset of Scheme, and whose implementation MIT-Church performs approximate conditional execution via Markov chain Monte Carlo (which can be thought of as a random walk over the execution of a Lisp machine). Park, Pfenning, and Thrun [32] describe the language $\lambda_\circ$, which extends OCaml, and they implement approximate conditional execution by Monte Carlo importance sampling. Ramsey and Pfeffer [39] describe a stochastic lambda calculus whose semantics are given by *measure terms*, which support the efficient computation of conditional expectations.

Finally, in nonparametric Bayesian statistics, higher-order distributions (e.g., distributions on distributions, or distributions on trees) arise naturally, and so it is helpful to work in a language that can express these types. Probabilistic functional programming languages are therefore a convenient choice for expressing nonparametric models.

The representation of distributions by randomized algorithms that produce samples can highlight algorithmic issues. For example, a distribution will, in general, have many different representations as a probabilistic program, each with its own time, space, and entropy complexity. For example, both ways of sampling a Beta-Bernoulli process described in Section 2.1 can be represented in, e.g., the Church probabilistic programming language. One of the questions that motivated the present work was whether there is always an algorithm for sampling from the de Finetti measure when there is an algorithm for sampling the exchangeable sequence. This question was first raised by Roy et al. [40]. The computable de Finetti theorem answers this question in the affirmative, and, furthermore, shows that one can move between these representations automatically. In the following section, we provide a concrete example of the representational change made possible by the computable de Finetti transformation, using the syntax of the Church probabilistic programming language.

6.2. Code Transformations

Church extends a pure subset of Scheme (a dialect of Lisp) with a stochastic, binary-valued[2] `flip` procedure, calls to which return independent, Bernoulli($\frac{1}{2}$)-distributed random values in $\{0, 1\}$. Using the semantics of Church, it is possible to associate every closed Church expression (i.e., one without free variables) with a distribution on values. For example, evaluations of the expression

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

produce samples from the Binomial($n = 3$, $p = \frac{1}{2}$) distribution, while evaluations of

```
(λ (x) (if (= 1 (flip)) x 0))
```

always return a procedure, applications of which behave like the probability kernel $x \mapsto \frac{1}{2}(\delta_x + \delta_0)$, where δ_r denotes the Dirac measure concentrated on the real r. Church is call-by-value and so evaluations of

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

return `true` and `false` with equal probability, while the application of the procedure

```
(λ (x) (= x x))
```

to the argument `(flip)`, written

```
((λ (x) (= x x)) (flip)),
```

always returns `true`. (For more examples, see [33].)

In Scheme, unlike Church, one can modify the state of a non-local variable using mutation via the `set!` procedure. (In functional programming languages, non-local state may be implemented via other methods. For example, in Haskell, one could use the state monad.) If we consider introducing a `set!` operator to Church, thereby allowing a procedure to modify its environment using mutation, it is not clear how one can, in a manner similar to above, associate procedures with probability kernels and closed expressions with distributions. For example, a procedure could then keep a counter variable and return an increasing sequence of integers on repeated calls. Such a procedure would not correspond with a probability kernel.

A generic way to translate code with mutation into code without mutation is to perform a state-passing transformation, where the state is explicitly threaded

---

[2]The original Church paper defined the `flip` procedure to return `true` or `false`, but it is easy to move between these two definitions.

throughout the program. In particular, a variable representing state is passed into all procedures as an additional argument, transformed in lieu of `set!` operations, and returned alongside the original return values at the end of procedures. Under such a transformation, the procedure in the counter variable example would be transformed into one that accepted the current count and returned the incremented count. One downside of such a transformation is that it obscures conditional independencies in the program, and thus complicates inference from an algorithmic standpoint.

An alternative transformation is made possible by the computable de Finetti theorem, which implies that a particular type of *exchangeable* mutation can be removed *without* requiring a state-passing transformation. Furthermore, this alternative transformation exposes the conditional independencies. The rest of this section describes a concrete example of this alternative transformation, and builds on the mathematical characterization of the Beta-Bernoulli process and the Pólya urn scheme as described in Section 2.1.

Recall that the Pólya urn scheme induces the Beta-Bernoulli process, which can also be described directly as a sequence of independent Bernoulli random variables with a shared parameter sampled from a Beta distribution. In Church it is possible to write code corresponding to both descriptions, but expressing the Pólya urn scheme without the use of mutation requires that we keep track of the counts and thread these values throughout the sequence. If instead we introduce the `set!` operator and track the number of red and black balls by mutating non-local state, we can compactly represent the Pólya urn scheme in a way that mirrors the form of the more direct description using Beta and Bernoulli random variables.

Fix $a, b > 0$, and define `sample-beta-coin` and `sample-pólya-coin` as follows:

(i)

```
(define (sample-beta-coin)
  (let ((weight (beta a b)))
    (λ () (flip weight)) ) )
```

(ii)

```
(define (sample-pólya-coin)
  (let ((red a)
        (total (+ a b)) )
    (λ () (let ((x (flip red/total)))
            (set! red (+ red x))
            (set! total (+ total 1))
            x ) ) )
```

Recall that, given a Church expression E, the evaluation of the $(\lambda \ () \ E \ )$ special form in an environment ρ creates a procedure of no arguments whose application results in the evaluation of the expression E in the environment ρ. The application of either `sample-beta-coin` or `sample-pólya-coin` returns a procedure of

no arguments whose application returns (random) binary values. In particular, if we sample two procedures `my-beta-coin` and `my-pólya-coin` via

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

then repeated applications of both `my-beta-coin` and `my-pólya-coin` produce random binary sequences that are Beta-Bernoulli processes.

Evaluating `(my-beta-coin)` returns 1 with probability `weight` and 0 otherwise, where the shared `weight` parameter is itself drawn from a $\text{Beta}(a, b)$ distribution on $[0, 1]$. The sequence induced by repeated applications of `my-beta-coin` is exchangeable because applications of *flip* return independent samples. Note that the sequence is *not* i.i.d.; for example, an initial sequence of ten 1's would lead one to predict that the next application is more likely to return 1 than 0. However, conditioned on `weight` (a variable hidden within the opaque procedure `my-beta-coin`) the sequence is i.i.d. If we sample another procedure, `my-other-beta-coin`, via

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

then its corresponding `weight` variable will be independent, and so repeated applications will generate a sequence that is independent of that generated by `my-beta-coin`.

The code in *(ii)* implements the Pólya urn scheme with a red balls and b black balls (see [16, Chap. 11.4]), and so the sequence of return values from repeated applications of `my-pólya-coin` is exchangeable. Therefore, de Finetti's theorem implies that the distribution of the sequence is equivalent to that induced by i.i.d. draws from the directing random measure. In the case of the Pólya urn scheme, we know that the directing random measure is a random Bernoulli measure whose parameter has a $\text{Beta}(a, b)$ distribution. In fact, the (random) distribution of each sample produced by `my-beta-coin` is such a random Bernoulli measure. Informally, we can therefore think of `sample-beta-coin` as the de Finetti measure of the Beta-Bernoulli process.

Although the distributions on sequences induced by `my-beta-coin` and `my-pólya-coin` are identical, there is an important semantic difference between these two implementations caused by the use of `set!`. While applications of `sample-beta-coin` produce samples from the de Finetti measure in the sense described above, applications of `sample-pólya-coin` do not; successive applications of `my-pólya-coin` produce samples from different distributions, none of which is the directing random measure for the sequence (a.s.). In particular, the distribution on return values changes each iteration as the sufficient statistics are updated (using the mutation operator `set!`). In contrast, applications

27

of `my-beta-coin` do not modify non-local state; in particular, the sequence produced by such applications is i.i.d. conditioned on the variable `weight`, which does not change during the course of execution.

An implementation of the computable de Finetti theorem (Theorem 2.3), specialized to the case of binary sequences (in which case the de Finetti measure is a distribution on Bernoulli measures and is thus determined by the distribution on $[0, 1]$ of the random probability assigned to the value 1), transforms *(ii)* into a mutation-free procedure whose return values have the same distribution as that of the samples produced by evaluating (`beta a b`).

In the general case, given a program that generates an exchangeable sequence of reals, an implementation of the computable de Finetti theorem produces a mutation-free procedure `generated-code` such that applications of the procedure `sample-directing-random-measure` defined by

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

sample from the de Finetti measure in the sense described above. In particular, *(ii)* would be transformed into a procedure `generated-code` such that the sequences produced by repeated applications of the procedures returned by `sample-beta-coin` and `sample-directing-random-measure` have the same distribution.

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. Mansinghka [41] describes some situations where one can exploit conditional independence and exchangeability in probabilistic programming languages for improved parallel execution.

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 [15]). 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

$$\mathbb{P}[X_{n+1} \mid X_1, \ldots, X_n] = \frac{1}{n + \alpha}\Big(\alpha H + \sum_{i=1}^{n} \delta_{X_i}\Big) \qquad \text{a.s.} \tag{66}$$

28

The sequence $\{X_n\}_{n \geq 1}$ is exchangeable and the directing random measure is a Dirichlet process whose "base measure" is αH. Given such a program, we can automatically recover the underlying Dirichlet process prior, samples from which are random measures whose discrete structure was characterized by Sethuraman's "stick-breaking construction" [42]. 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.

Exchangeable sequences of random objects other than reals can often be given de Finetti-type representations. For example, the Indian buffet process, defined by Griffiths and Ghahramani [43], 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 the Beta process (see Thibaux and Jordan [44] for more details).

In the case where the "base measure" of the underlying Beta process is *discrete*, the resulting exchangeable sequence of sets corresponds to an exchangeable sequence of *integer* indices (encoding finite subsets of the countable support of the discrete base measure). If we are given such a representation, the computable de Finetti theorem implies the existence of a computable de Finetti measure.

However, the case of a general base measure is more complicated. A "stick-breaking construction" of the Indian buffet process given by Teh, Görür, and Ghahramani [45] is analogous to the code in *(i)*, but samples only a Δ_1-index for the (a.s. finite) sets, rather than a canonical index (see Soare [18, II.2]); however, many applications depend on having a canonical index. These observation were first noted by Roy et al. [40]. Similar problems arise when using the Inverse Lévy Measure method [46] to construct the Indian buffet process. The computable de Finetti theorem is not directly applicable in this case because the theorem pertains only to exchangeable sequences of real random variables, not random sets, although an extension of the theorem to computable Polish spaces might suffice.

Combinatorial structures other than sequences have been given de Finetti-type representational theorems based on notions of *partial* exchangeability. 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 [47] and Hoover [48] 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 [49] and Austin [50] by considering the adjacency matrix of an exchangeable random graph.

As we have seen with the Beta-Bernoulli process and other examples, structured probabilistic models can often be represented in multiple ways, each with its own advantages (e.g., representational simplicity, compositionality, inherent parallelism, etc.). Extensions of the computable de Finetti theorem to partially exchangeable settings could provide analogous transformations between representations on a wider range of data structures, including many that are increasingly used in practice. For example, the Infinite Relational Model [51] can be viewed as an urn scheme for a partially exchangeable array, while the hierarchical stochastic block model constructed from a Mondrian process in [52] is described in a way that mirrors the Aldous-Hoover representation, making the conditional independence explicit.

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, 20

[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] B. de Finetti, La prévision : ses lois logiques, ses sources subjectives, Ann. Inst. H. Poincaré 7 (1) (1937) 1–68. 4

[11] E. Hewitt, L. J. Savage, Symmetric measures on Cartesian products, Trans. Amer. Math. Soc. 80 (1955) 470–501. 4

[12] C. Ryll-Nardzewski, On stationary sequences of random variables and the de Finetti's equivalence, Colloq. Math. 4 (1957) 149–156. 4

[13] J. F. C. Kingman, Uses of exchangeability, Ann. Probability 6 (2) (1978) 183–197. 4

[14] P. Diaconis, D. Freedman, Partial exchangeability and sufficiency, in: Statistics: applications and new directions (Calcutta, 1981), Indian Statist. Inst., Calcutta, 205–236, 1984. 4

[15] 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, 28

[16] B. de Finetti, Theory of probability. Vol. 2, John Wiley & Sons Ltd., London, 1975. 5, 27

[17] H. Rogers, Jr., Theory of recursive functions and effective computability, MIT Press, Cambridge, MA, second edn., 1987. 7

[18] R. I. Soare, Recursively enumerable sets and degrees, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1987. 7, 29

[19] T. Grubba, M. Schröder, K. Weihrauch, Computable metrization, Math. Logic Q. 53 (4-5) (2007) 381–395. 8

[20] 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

[21] K. Weihrauch, Computable analysis: an introduction, Springer, Berlin, 2000. 8, 9, 10

[22] 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. 10

[23] M. Schröder, Admissible representations for probability measures, Math. Logic Q. 53 (4-5) (2007) 431–445. 10, 11, 13, 19

[24] V. Bosserhoff, Notions of probabilistic computability on represented spaces, J. of Universal Comput. Sci. 14 (6) (2008) 956–995. 10, 20

[25] K. Weihrauch, Computability on the probability measures on the Borel sets of the unit interval, Theoret. Comput. Sci. 219 (1–2) (1999) 421–437. 10, 11, 13, 20

[26] M. Schröder, A. Simpson, Representing probability measures using probabilistic processes, J. Complex. 22 (6) (2006) 768–782. 11, 19

[27] V. Brattka, G. Presser, Computability on subsets of metric spaces, Theor. Comput. Sci. 305 (1-3) (2003) 43–76. 11

[28] N. T. Müller, Computability on random variables, Theor. Comput. Sci. 219 (1-2) (1999) 287–299. 12, 13, 14, 20

[29] 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.

[30] M. B. Pour-El, J. I. Richards, Computability in analysis and physics, Springer, Berlin, 1989. 15

[31] 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. 23

[32] S. Park, F. Pfenning, S. Thrun, A probabilistic language based on sampling functions, ACM Trans. Program. Lang. Syst. 31 (1) (2008) 1–46. 23, 24

[33] 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. 23, 24, 25

[34] 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. 23

[35] 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. 23

[36] G. D. Plotkin, LCF considered as a programming language, Theoret. Comput. Sci. 5 (3) (1977/78) 223–255. 23

[37] 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. 23

[38] 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. 23

[39] 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. 24

[40] 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. 24, 29

[41] V. K. Mansinghka, Natively probabilistic computation, Ph.D. thesis, Massachusetts Institute of Technology, 2009. 28

[42] J. Sethuraman, A constructive definition of Dirichlet priors, Statistica Sinica 4 (1994) 639–650. 29

[43] 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. 29

[44] 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. 29

[45] 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. 29

[46] R. L. Wolpert, K. Ickstadt, Simulation of Lévy random fields, in: Practical nonparametric and semiparametric Bayesian statistics, vol. 133 of *Lecture Notes in Statist.*, Springer, New York, 227–242, 1998. 29

[47] D. J. Aldous, Representations for partially exchangeable arrays of random variables, J. Multivariate Analysis 11 (4) (1981) 581–598. 29

[48] D. N. Hoover, Relations on probability spaces and arrays of random variables, preprint, Institute for Advanced Study, Princeton, NJ, 1979. 29

[49] P. Diaconis, S. Janson, Graph limits and exchangeable random graphs, Rendiconti di Matematica, Ser. VII 28 (1) (2008) 33–61. 29

[50] T. Austin, On exchangeable random variables and the statistics of large graphs and hypergraphs, Probab. Surv. 5 (2008) 80–145. 29

[51] 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. 30

[52] D. M. Roy, Y. W. Teh, The Mondrian process, in: Adv. in Neural Inform. Processing Syst. 21, 2009. 30

[53] 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. 30