NONCOMPUTABLE CONDITIONAL DISTRIBUTIONS

NATHANAEL L. ACKERMAN, CAMERON E. FREER, AND DANIEL M. ROY

Abstract. We study the computability of conditional probability, a fundamental notion in probability theory and Bayesian statistics. In the elementary discrete setting, a ratio of probabilities defines conditional probability. In more general settings, conditional probability is defined axiomatically, and the search for more constructive definitions is the subject of a rich literature in probability theory and statistics. However, we show that in general one cannot compute conditional probabilities. Specifically, we construct a pair of computable random variables (X, Y) in the unit interval whose conditional distribution $\mathbf{P}[Y|X]$ encodes the halting problem.

Nevertheless, probabilistic inference has proven remarkably successful in practice, even in infinite-dimensional continuous settings. We prove several results giving general conditions under which conditional distributions *are* computable. In the discrete or dominated setting, under suitable computability hypotheses, conditional distributions are computable. Likewise, conditioning is a computable operation in the presence of certain additional structure, such as independent absolutely continuous noise.

Keywords: conditional probability; computable probability theory; probabilistic programming languages; and real computation. June 10, 2011

1. Introduction

The use of probability to reason about uncertainty is fundamental to modern science and engineering, and the formation of conditional probabilities, in order to perform evidential reasoning in probabilistic models, is one of its most important computational problems.

The desire to build probabilistic models of increasingly complex phenomena has led researchers to propose new representations for joint distributions of collections of random variables. In particular, within statistical AI and machine learning, there has been renewed interest in *probabilistic programming languages*, in which practitioners can define intricate, even infinite-dimensional, models by implementing a generative process that produces an exact sample from the joint distribution. (See, e.g., PHA [1], IBAL [2], $\lambda_\circ$[3], Church [4], and HANSEI [5].) In sufficiently expressive languages built on modern programming languages, one can easily represent distributions on higher-order, structured objects, such as distributions on data structures, distributions on functions, and distributions on distributions. Furthermore, the most expressive such languages are capable of representing the same robust class of *computable distributions*, which delineates those from which a probabilistic Turing machine can sample to arbitrary accuracy.

Whereas the probability-theoretic derivations necessary to build special-purpose algorithms for probabilistic models have typically been performed by hand, implementations of probabilistic programming languages provide varying degrees of algorithmic support for computing conditional distributions. Progress has been made at increasing the scope of these implementations, and one might hope that there would eventually be a generic implementation that would support the entire class of computable distributions. What are the limits of this endeavor? Can we hope to automate probabilistic reasoning via a general inference algorithm?

Despite recent progress, support for conditioning with respect to continuous random variables has remained ad-hoc and incomplete. We demonstrate why this is the case, by showing that there are computable joint distributions with noncomputable conditional distributions.

The fact that generic algorithms do not exist for computing conditional distributions does not rule out the possibility that large classes of distributions may be amenable to automated inference. The challenge for mathematical theory is to explain the widespread success of probabilistic methods and develop a characterization of the circumstances when conditioning is possible. In this vein, we describe broadly-applicable conditions under which conditional distributions are computable.

1.1. **Conditional probability.**

For an experiment with a discrete set of outcomes, computing conditional probabilities is straightforward. However, in modern Bayesian statistics, and especially the probabilistic programming setting, it is common to place distributions on continuous or higher-order objects, and so one is already in a situation where elementary notions of conditional probability are insufficient and more sophisticated measure-theoretic notions are required. When conditioning on a continuous random variable, each particular observation has probability 0, and the elementary rule that characterizes the discrete case does not apply. Kolmogorov [6] gave an axiomatic characterization of conditional probabilities, but this definition provides no recipe for their calculation. In some situations, e.g., when joint densities exist, conditioning can proceed using a continuous version of the classic Bayes' rule; however, it may not be possible to compute the density of a computable distribution (if the density even exists classically at all). The probability and statistics literature contains many ad-hoc rules for calculating conditional probabilities in special circumstances, but even the most constructive definitions (e.g., those due to Tjur [7], [8], [9], Pfanzagl [10], and Rao [11], [12]) are often not sensitive to issues of computability.

In order to characterize the computational limits of probabilistic inference, we work within the framework of *computable probability theory*, which pertains to the computability of distributions and probability kernels, and which builds on the classical computability theory of deterministic functions. Just as the notion of a Turing machine allows one to prove results about discrete computations performed using an arbitrary (sufficiently rich) programming language, the notion of a probabilistic Turing machine likewise provides a basis for precisely describing the operations that various probabilistic programming languages are capable of performing in principle. The basic tools of this approach have been developed in the area known as *computable analysis*; in particular, computable distributions on *computable metric spaces* are a rich enough class to describe distributions on higher-order objects like distributions on distributions. In Section 2 we present the necessary definitions and results from computable probability theory.

We recall the basics of the measure-theoretic approach to conditional distributions in Section 3, and in Section 4 we consider the sense in which formation of conditional probability is a potentially computable operation. In the remainder of the paper, we provide our main positive and negative results about the computability of conditional probability, which we now summarize.

1.2. **Summary of results.** In Proposition 23, we construct a pair (X, C) of computable random variables such that every version of the conditional distribution $\mathbf{P}[\mathsf{C}|\mathsf{X}]$ is discontinuous even when restricted to a $\mathbf{P}_{\mathsf{X}}$-measure one subset. (We make these notions precise in Section 4.) Every function computable on a domain D is continuous on D, and so this construction rules out the possibility of a completely general algorithm for conditioning. A natural question is whether conditioning is a computable operation when we restrict the operator to random variables for which *some* version of the conditional distribution is continuous everywhere, or at least on a measure one set.

Our main result, Theorem 29, states that conditioning is not a computable operation on computable random variables, even in this restricted setting. We construct a pair (X, N) of computable random variables such that there is a version of the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ that is continuous on a measure one set, but no version of $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ is computable. Moreover, if some oracle A computes $\mathbf{P}[\mathsf{N}|\mathsf{X}]$, then

A computes the halting problem. In Theorem 50 we strengthen this result by constructing a pair of computable random variables whose conditional distribution is noncomputable but has an *everywhere continuous* version.

We also characterize several circumstances in which conditioning *is* a computable operation. Under suitable computability hypotheses, conditioning is computable in the discrete setting (Lemma 30) and where there is a conditional density (Corollary 35).

Finally, we characterize the following situation in which conditioning on noisy data is possible. Let U, V and E be computable random variables, and define $\mathsf{Y} = \mathsf{U} + \mathsf{E}$. Suppose that $\mathbf{P}_{\mathsf{E}}$ is absolutely continuous with a bounded computable density p_{E} and E is independent of U and V. In Corollary 36, we show that the conditional distribution $\mathbf{P}[(\mathsf{U}, \mathsf{V}) \mid \mathsf{Y}]$ is computable.

All proofs not presented in the body of this extended abstract can be found in Appendices A through F.

1.3. **Related work.** Conditional probabilities for distributions on finite sets of discrete strings are manifestly computable, but may not be efficiently so. In this finite discrete setting, there are already interesting questions of computational complexity, which have been explored through extensions of Levin's theory of average-case complexity [13]. If f is a one-way function, then it is difficult to sample from the conditional distribution of the uniform distribution of strings of some length with respect to a given output of f. This intuition is made precise by Ben-David, Chor, Goldreich, and Luby [14] in their theory of polynomial-time samplable distributions, which has since been extended by Yamakami [15] and others. Extending these complexity results to the richer setting considered here could bear on the practice of statistical AI and machine learning.

Osherson, Stob, and Weinstein [16] study learning theory in the setting of *identifiability in the limit* (see [17] and [18] for more details on this setting) and prove that a certain type of "computable Bayesian" learner fails to identify the index of a (computably enumerable) set that is computably identifiable in the limit. More specifically, a "Bayesian" learner is required to return an index for a set with the highest conditional probability given a finite prefix of an infinite sequence of random draws from the unknown set. An analysis of their construction reveals that the conditional distribution of the index given the infinite sequence is an everywhere discontinuous function (on every measure one set), hence noncomputable for much the same reason as our elementary construction involving a mixture of measures concentrated on the rationals and on the irrationals (see Section 5). As we argue, the more appropriate operator to study is that restricted to those random variables whose conditional distributions admit versions that are continuous everywhere, or at least on a measure one set.

Our work is distinct from the study of conditional distributions with respect to priors that are universal for partial computable functions (as defined using Kolmogorov complexity) by Solomonoff [19], Zvonkin and Levin [20], and Hutter [21]. The computability of conditional distributions also has a rather different character in Takahashi's work on the algorithmic randomness of points defined using universal

Martin-Löf tests [22]. The objects with respect to which one is conditioning in these settings are typically *computably enumerable*, but not computable. In the present paper, we are interested in the problem of computing conditional distributions of random variables that are *computable* (even though the conditional distribution may itself be noncomputable).

2. Computable probability theory

For a general introduction to this approach to real computation, see Braverman [23] or Braverman and Cook [24].

2.1. Computable and c.e. reals.

We first recall some elementary definitions from computability theory (see, e.g. Rogers [25, Ch. 5]). We say that a set (of rationals, integers, or other finitely describable objects with an implicit enumeration) is *computably enumerable* (c.e.) when there is a computer program that outputs every element of the set eventually. A set is co-c.e. when its complement is c.e. (and so the computable sets are precisely those that are both c.e. and co-c.e.).

We now recall basic notions of computability for real numbers (see, e.g., [26, Ch. 4.2] or [27, Ch. 1.8]). We say that a real r is a *c.e. real* when the set of rationals $\{q \in \mathbb{Q} : q < r\}$ is c.e. Similarly, a *co-c.e. real* is one for which $\{q \in \mathbb{Q} : q > r\}$ is c.e. (C.e. and co-c.e. reals are sometimes called *left-c.e.* and *right-c.e.* reals, respectively.) A real r is *computable* when it is both c.e. and co-c.e. Equivalently, a real is computable when there is a program that approximates it to any given accuracy (e.g., given an integer k as input, the program reports a rational that is within 2^{-k} of the real).

2.2. Computable metric spaces.

Computable metric spaces, as developed in computable analysis, provide a convenient framework for formulating results in computable probability theory. For consistency, we largely use definitions from [28] and [29]. Additional details about computable metric spaces can also be found in [26, Ch. 8.1] and [30, §B.3].

Definition 1 (Computable metric space [29, Def. 2.3.1])**.** A **computable metric space** is a triple $(S, \delta, \mathcal{D})$ for which δ is a metric on the set S satisfying

(1) (S, δ) is a complete separable metric space;
(2) $\mathcal{D} = \{s_i\}_{i \in \mathbb{N}}$ is an enumeration of a dense subset of S, called **ideal points**; and,
(3) the real numbers $\delta(s_i, s_j)$ are computable, uniformly in i and j (i.e., the function $(i, j) \mapsto \delta(s_i, s_j)$ is computable).

Let $B(s_i, q_j)$ denote the ball of radius q_j centered at s_i. We call $\mathscr{B}_S := \{B(s_i, q_j) : s_i \in \mathcal{D},\ q_j \in \mathbb{Q},\ q_j > 0\}$ the **ideal balls of** S, and fix the canonical enumeration of them induced by that of $\mathcal{D}$ and $\mathbb{Q}$.

For example, the set $\{0, 1\}$ is a computable metric space under the discrete metric, characterized by $\delta(0, 1) = 1$. Cantor space, the set $\{0, 1\}^\infty$ of infinite binary sequences, is a computable metric space under its usual metric and the dense set of eventually constant strings (under a standard enumeration of finite strings). The set $\mathbb{R}$ of real numbers is a computable metric space under the Euclidean metric with the dense set $\mathbb{Q}$ of rationals (under its standard enumeration).

Definition 2 (Computable point [29, Def. 2.3.2])**.** Let $(S, \delta, \mathcal{D})$ be a computable metric space. A point $x \in S$ is **computable** when there is a program that enumerates a sequence $\{x_i\}$ in $\mathcal{D}$ where $d(x_i, x) < 2^{-i}$ for all i. We call such a sequence $\{x_i\}$ a **representation** of the point x.

Remark 3. A real $\alpha \in \mathbb{R}$ is computable (as in Section 2.1) if and only if α is a computable point of $\mathbb{R}$ (as a computable metric space). Although most of the familiar reals are computable, there are only countably many computable reals, and so almost every real is not computable.

The notion of a c.e. open set (or Σ^0_1 class) is fundamental in classical computability theory, and admits a simple definition in an arbitrary computable metric space.

Definition 4 (C.e. open set [29, Def. 2.3.3])**.** Let $(S, \delta, \mathcal{D})$ be a computable metric space with the corresponding enumeration $\{B_i\}_{i \in \mathbb{N}}$ of the ideal open balls $\mathscr{B}_S$. We say that $U \subseteq S$ is a **c.e. open set** when there is some c.e. set $E \subseteq \mathbb{N}$ such that $U = \bigcup_{i \in E} B_i$.

Note that the class of c.e. open sets is closed under computable unions and finite intersections.

A computable function can be thought of as a continuous function whose local modulus of continuity is witnessed by a program. It is important to consider the computability of *partial* functions, since many natural and important random variables are continuous only on a measure one subset of their domain.

Definition 5 (Computable partial function [29, Def. 2.3.6])**.** Let $(S, \delta_S, \mathcal{D}_S)$ and $(T, \delta_T, \mathcal{D}_T)$ be computable metric spaces, the latter with the corresponding enumeration $\{B_n\}_{n \in \mathbb{N}}$ of the ideal open balls $\mathscr{B}_T$. A function $f : S \to T$ is said to be **computable on** $R \subseteq S$ when there is a computable sequence $\{U_n\}_{n \in \mathbb{N}}$ of c.e. open sets $U_n \subseteq S$ such that $f^{-1}(B_n) \cap R = U_n \cap R$ for all $n \in \mathbb{N}$.

Remark 6. Let S and T be computable metric spaces. If $f : S \to T$ is computable on some subset $R \subseteq S$, then for every *computable* point $x \in R$, the point $f(x)$ is also computable. One can show that f is computable on R when there is a program that uniformly transforms representations of points in R to representations of points in S. (For more details, see [28, Prop. 3.3.2].)

2.3. **Computable random variables and distributions.** Intuitively, a random variable maps an input source of randomness to an output, inducing a distribution on the output space. Here we will use a sequence of independent fair coin flips as our source of randomness. We formalize this via the probability space $(\{0,1\}^\infty, \mathbf{P})$, where $\{0,1\}^\infty$ is the space of infinite binary sequences whose basic clopen sets are cylinders extending some finite binary sequence, and $\mathbf{P}$ is the product measure of the uniform distribution on $\{0,1\}$.

Henceforth we will take $(\{0,1\}^\infty, \mathbf{P})$ to be the basic probability space, unless otherwise stated. We will typically use a sans serif font for random variables.

Definition 7 (Random variable and its distribution)**.** Let S be a computable metric space. A **random variable in** S is a function $\mathsf{X} : \{0,1\}^\infty \to S$ that is measurable with respect to the Borel σ-algebras of $\{0,1\}^\infty$ and S. For a measurable subset $A \subseteq S$, we let $\{\mathsf{X} \in A\}$ denote the inverse image $\mathsf{X}^{-1}[A] = \{\omega \in \{0,1\}^\infty : \mathsf{X}(\omega) \in A\}$, and for $x \in S$ we similarly define the event $\{\mathsf{X} = x\}$. The **distribution of** X is a measure on S defined to be $\mathbf{P}_\mathsf{X}(\cdot) := \mathbf{P}\{\mathsf{X} \in \cdot\}$.

Definition 8 (Computable random variable)**.** Let S be a computable metric space. Then a random variable X in S is a **computable random variable**[1] when X is computable on some $\mathbf{P}$-measure one subset of $\{0,1\}^\infty$.

Intuitively, X is a computable random variable when there is a program that, given access to an oracle bit tape $\omega \in \{0,1\}^\infty$, outputs a representation of the point $\mathsf{X}(\omega)$ (i.e., enumerates a sequence $\{x_i\}$ in $\mathcal{D}$ where $\delta(x_i, \mathsf{X}(\omega)) < 2^{-i}$ for all i), for all but a measure zero subset of bit tapes $\omega \in \{0,1\}^\infty$ (see Remark 6).

It is crucial that we consider random variables that are computable only on a $\mathbf{P}$-measure one subset of $\{0,1\}^\infty$. For a real $\alpha \in [0,1]$, we say that a binary random variable $\mathsf{X} : \{0,1\}^\infty \to \{0,1\}$ is a **Bernoulli**(α) random variable when $\mathbf{P}_\mathsf{X}\{1\} = \alpha$. There is a Bernoulli$(\frac{1}{2})$ random variable that is computable on all of $\{0,1\}^\infty$, given by the program that simply outputs the first bit of the input sequence. Likewise, when α is **dyadic** (i.e., a rational with denominator a power of 2), there is a Bernoulli(α) random variable that is computable on all of $\{0,1\}^\infty$. However, this is not possible for any other choices of α (e.g., $\frac{1}{3}$).

Proposition 9. *Let $\alpha \in [0,1]$ be a nondyadic real. Every* Bernoulli(α) *random variable $\mathsf{X} : \{0,1\}^\infty \to \{0,1\}$ is discontinuous, hence not computable on all of $\{0,1\}^\infty$.*

On the other hand, for an arbitrary computable $\alpha \in [0,1]$, a more sophisticated construction [32] produces a Bernoulli(α) random variable that is computable on every point of $\{0,1\}^\infty$ other than the binary expansion of α. These random variables are manifestly computable in an intuitive sense (and can even be shown to be optimal in their use of input bits, via classic analysis of rational-weight coins by Knuth and Yao [33]). Hence it is natural to admit as computable random variables those measurable functions that are computable only on a $\mathbf{P}$-measure one subset of $\{0,1\}^\infty$, as we have done.

Let $\mathcal{M}_1(S)$ denote the set of (Borel) probability measures on a computable metric space S. The Prokhorov metric (and a suitably chosen dense set of measures [30, §B.6.2]) makes $\mathcal{M}_1(S)$ into a computable metric space [28, Prop. 4.1.1].

Theorem 10 ([28, Thm. 4.2.1])**.** *Let $(S, \delta_S, \mathcal{D}_S)$ be a computable metric space. A probability measure $\mu \in \mathcal{M}_1(S)$ is a computable point of $\mathcal{M}_1(S)$ (under the*

[1] Even though the source of randomness is a sequence of discrete bits, there are computable random variables with *continuous* distributions, such as a uniform random variable (by subdividing the interval according to the random bittape) or an i.i.d.-uniform sequence (by splitting up the given element of $\{0,1\}^\infty$ into countably many disjoint subsequences and dovetailing the constructions). (For details, see [31, Ex. 3, 4].)

Prokhorov metric) if and only if the measure $\mu(A)$ of a c.e. open set $A \subseteq S$ is a c.e. real, uniformly in A.

Proposition 11 (Computable random variables have computable distributions [29, Prop. 2.4.2])**.** *Let X be a computable random variable in a computable metric space S. Then its distribution is a computable point in the computable metric space $\mathcal{M}_1(S)$.*

On the other hand, one can show that given a computable point μ in $\mathcal{M}_1(S)$, one can construct an i.i.d.-μ sequence of computable random variables in S.

Henceforth, we say that a measure $\mu \in \mathcal{M}_1(S)$ is computable when it is a computable point in $\mathcal{M}_1(S)$, considered as a computable metric space in this way. Note that the measure $\mathbf{P}$ on $\{0,1\}^\infty$ is a computable probability measure.

3. Conditional distributions

The notion of *conditional probability* captures the intuitive idea of how likely an event B is given the knowledge that some positive-measure event A has already occurred.

Definition 12 (Conditional probability)**.** Let S be a measurable space and let $\mu \in \mathcal{M}_1(S)$ be a probability measure on S. Let $A, B \subseteq S$ be measurable sets, and suppose that $\mu(A) > 0$. Then the **conditional probability of B given A**, written $\mu(B|A)$, is defined by

$$\mu(B|A) = \frac{\mu(B \cap A)}{\mu(A)}. \tag{1}$$

Note that for any fixed measurable $A \subseteq S$ with $\mu(A) > 0$, the function $\mu(\,\cdot\,|A)$ is a probability measure. However, this notion of conditioning is well-defined only when $\mu(A) > 0$, and so is insufficient for defining the conditional probability given the event that a *continuous* random variable takes a particular value, as such an event has measure zero.

In order to define the more abstract notion of a conditional distribution, we first recall the notion of a probability kernel. (For more details, see, e.g., [34, Ch. 3, 6].) Suppose T is a metric space. We let $\mathcal{B}_T$ denote the Borel σ-algebra on T.[2]

Definition 13 (Probability kernel)**.** Let S and T be metric spaces. A function $\kappa : S \times \mathcal{B}_T \to [0,1]$ is called a **probability kernel (from S to T)** when

(1) for every $s \in S$, the function $\kappa(s, \cdot)$ is a probability measure on T; and
(2) for every $B \in \mathcal{B}_T$, the function $\kappa(\cdot, B)$ is measurable.

Suppose X is a random variable mapping a probability space S to a measurable space T.

[2]The Borel σ-algebra of T is the σ-algebra generated by the open balls of T (under countable unions and complements). In this paper, measurable functions will always be with respect to the Borel σ-algebra of a metric space.

Definition 14 (Conditional distribution)**.** Let X and Y be random variables in metric spaces S and T, respectively, and let $\mathbf{P}_\mathsf{X}$ be the distribution of X. A probability kernel κ is called a **(regular) version of the conditional distribution** $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ when it satisfies

$$\mathbf{P}\{\mathsf{X} \in A, \mathsf{Y} \in B\} = \int_A \kappa(x, B)\, \mathbf{P}_\mathsf{X}(dx), \tag{2}$$

for all measurable sets $A \subseteq S$ and $B \subseteq T$.

Definition 15. Let μ be a measure on a topological space S with open sets $\mathcal{S}$. Then the **support of** μ, written $\mathrm{supp}(\mu)$, is defined to be the set of points $x \in S$ such that all open neighborhoods of x have positive measure, i.e., $\mathrm{supp}(\mu) := \{x \in S : \forall B \in \mathcal{S}\ (x \in B \implies \mu(B) > 0)\}$.

Given any two versions κ_1, κ_2 of $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$, the functions $x \mapsto \kappa_i(x, \cdot)$ need only agree $\mathbf{P}_\mathsf{X}$-almost everywhere, although the functions $x \mapsto \kappa_i(x, \cdot)$ will agree at points of continuity in $\mathrm{supp}(\mathbf{P}_\mathsf{X})$.

Lemma 16. *Let X and Y be random variables in topological spaces S and T, respectively, let $\mathbf{P}_\mathsf{X}$ be the distribution of X, and suppose that κ_1, κ_2 are versions of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$. Let $x \in S$ be a point of continuity of both of the maps $x \mapsto \kappa_i(x, \cdot)$ for $i = 1, 2$. If $x \in \mathrm{supp}(\mathbf{P}_\mathsf{X})$, then $\kappa_1(x, \cdot) = \kappa_2(x, \cdot)$.*

When conditioning on a discrete random variable, a version of the conditional distribution can be built using conditional probabilities.

Lemma 17. *Let X and Y be random variables mapping a probability space S to a measurable space T. Suppose that X is a discrete random variable with support $R \subseteq S$, and let ν be an arbitrary probability measure on T. Define the function $\kappa : S \times \mathcal{B}_T \to [0, 1]$ by*

$$\kappa(x, B) := \mathbf{P}\{\mathsf{Y} \in B \mid \mathsf{X} = x\} \tag{3}$$

for all $x \in R$ and $\kappa(x, \cdot) = \nu(\cdot)$ for $x \notin R$. Then κ is a version of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$.

4. Computable conditional distributions

Having defined the abstract notion of a conditional distribution in Section 3, we now define our notion of computability for conditional distributions.

Definition 18 (Computable probability kernel)**.** Let S and T be computable metric spaces and let $\kappa : S \times \mathcal{B}_T \to [0, 1]$ be a probability kernel from S to T. Then we say that κ is a **computable (probability) kernel** when the map $\phi_\kappa : S \to \mathcal{M}_1(T)$ given by $\phi_\kappa(s) := \kappa(s, \cdot)$ is a computable function. Similarly, we say that κ is computable on a subset $D \subseteq S$ when ϕ_κ is computable on D.

Recall that a lower semicomputable function from a computable metric space to $[0, 1]$ is one for which the preimage of $(q, 1]$ is c.e. open, uniformly in rationals q. Furthermore, we say that a function f from a computable metric space S to $[0, 1]$

is *lower semicomputable on* $D \subseteq S$ when there is a uniformly computable sequence $\{U_q\}_{q\in\mathbb{Q}}$ of c.e. open sets such that

$$f^{-1}\big((q,1]\big) \cap D = U_q \cap D. \tag{4}$$

We can also interpret a computable probability kernel κ as a computable map sending each c.e. open set $A \subseteq T$ to a lower semicomputable function $\kappa(\cdot, A)$.

Lemma 19. *Let S and T be computable metric spaces, let κ be a probability kernel from S to T, and let $D \subseteq S$. Then ϕ_κ is computable on D if and only if $\kappa(\cdot, A)$ is lower semicomputable on D uniformly in a c.e. open set A.*

In fact, when $A \subseteq T$ is a decidable set (i.e., A and $T \setminus A$ are both c.e. open), $\kappa(\cdot, A)$ is a computable function.

Corollary 20. *Let S and T be computable metric spaces, let κ be a probability kernel from S to T computable on a subset $D \subseteq S$, and let $A \subseteq T$ be a decidable set. Then $\kappa(\cdot, A) : S \to [0, 1]$ is computable on D.*

Although a conditional distribution may have many different versions, their computability as probability kernels does not differ (up to a change in domain by a null set).

Lemma 21. *Let κ be a version of a conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ that is computable on some $\mathbf{P}_\mathsf{X}$-measure one set. Then any version of $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is also computable on some $\mathbf{P}_\mathsf{X}$-measure one set.*

Proof. Let κ be a version that is computable on a $\mathbf{P}_\mathsf{X}$-measure one set D, and let κ' be any other version. Then $Z := \{s \in S : \kappa(s, \cdot) \neq \kappa'(s, \cdot)\}$ is a $\mathbf{P}_\mathsf{X}$-null set, and $\kappa = \kappa'$ on $D \setminus Z$. Hence κ' is computable on the $\mathbf{P}_\mathsf{X}$-measure one set $D \setminus Z$. □

Definition 22 (Computable conditional distributions)**.** We say that the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is computable when *some* version is computable on a $\mathbf{P}_\mathsf{X}$-measure one subset of S.

Intuitively, a conditional distribution is computable when for some (and hence for any) version κ there is a program that, given as input a representation of a point $s \in S$, outputs a representation of the measure $\phi_\kappa(s) = \kappa(s, \cdot)$ for $\mathbf{P}_\mathsf{X}$-almost all inputs s.

Suppose that $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is computable, i.e., there is a version κ for which the map ϕ_κ is computable on some $\mathbf{P}_\mathsf{X}$-measure one set $S' \subseteq S$.[3] The restriction of ϕ_κ to S' is necessarily continuous (under the subspace topology on S'). We say that κ is **$\mathbf{P}_\mathsf{X}$-almost continuous** when the restriction of ϕ_κ to some $\mathbf{P}_\mathsf{X}$-measure one set is continuous. Thus when $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is computable, there is some $\mathbf{P}_\mathsf{X}$-almost continuous version.

In Section 5 we describe a pair of computable random variables X, Y for which $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is not computable, by virtue of every version being not $\mathbf{P}_\mathsf{X}$-almost continuous. In Section 6 we describe a pair of computable random variables X, Y for which there is a $\mathbf{P}_\mathsf{X}$-almost continuous version of $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$, but still no version that is computable on a $\mathbf{P}_\mathsf{X}$-measure one set.

[3]As noted in Definition 18, we will often abuse notation and say that κ is computable on S'.

5. Discontinuous conditional distributions

Any attempt to characterize the computability of conditional distributions immediately runs into the following roadblock: a conditional distribution need not have *any* version that is continuous or even almost continuous (in the sense described in Section 4).

Recall that a random variable C is a **Bernoulli**(p) random variable, or equivalently, a p**-coin**, when $\mathbf{P}\{\mathsf{C}=1\} = 1-\mathbf{P}\{\mathsf{C}=0\} = p$. We call a $\frac{1}{2}$-coin a **fair coin**. A random variable N is **geometric** when it takes values in $\mathbb{N} = \{0,1,2,\dots\}$ and satisfies

$$\mathbf{P}\{\mathsf{N}=n\} = 2^{-(n+1)}, \quad \text{for } n \in \mathbb{N}. \tag{5}$$

A random variable that takes values in a discrete set is a **uniform** random variable when it assigns equal probability to each element. A continuous random variable U on the unit interval is **uniform** when the probability that it falls in the subinterval $[\ell, r]$ is $r-\ell$. It is easy to show that the distributions of these random variables are computable.

Let C, U, and N be independent computable random variables, where C is a fair coin, U is a uniform random variable on $[0,1]$, and N is a geometric random variable. Fix a computable enumeration $\{r_i\}_{i\in\mathbb{N}}$ of the rational numbers (without repetition) in $(0,1)$, and consider the random variable

$$\mathsf{X} := \begin{cases} \mathsf{U}, & \text{if } \mathsf{C}=1; \\ r_{\mathsf{N}}, & \text{otherwise.} \end{cases} \tag{6}$$

It is easy to verify that X is a computable random variable.

Proposition 23. *No version of the conditional distribution* $\mathbf{P}[\mathsf{C}|\mathsf{X}]$ *is* $\mathbf{P}_{\mathsf{X}}$*-almost continuous.*

Proof. Note that $\mathbf{P}\{\mathsf{X} \text{ rational}\} = \frac{1}{2}$ and, furthermore, $\mathbf{P}\{\mathsf{X} = r_k\} = \frac{1}{2^{k+1}} > 0$. Therefore, any two versions of the conditional distribution $\mathbf{P}[\mathsf{C}|\mathsf{X}]$ must agree on *all* rationals in $[0,1]$. In addition, any two versions must agree on *almost all* irrationals in $[0,1]$ because the support of U is all of $[0,1]$. An elementary calculation shows that $\mathbf{P}\{\mathsf{C}=0 \mid \mathsf{X} \text{ rational}\} = 1$, while $\mathbf{P}\{\mathsf{C}=0 \mid \mathsf{X} \text{ irrational}\} = 0$. Therefore, all versions κ of $\mathbf{P}[\mathsf{C}|\mathsf{X}]$ satisfy

$$\kappa(x,\{0\}) = \begin{cases} 1, & x \text{ rational;} \\ 0, & x \text{ irrational,} \end{cases} \quad \text{almost surely (a.s.),} \tag{7}$$

which, when considered as a function of x, is the *nowhere continuous* function known as the Dirichlet function.

Suppose some version κ were continuous when restricted to some $\mathbf{P}_{\mathsf{X}}$-measure one subset $D \subseteq [0,1]$. But D must contain every rational and almost every irrational in $[0,1]$, and so the inverse image of an open set containing 1 but not 0 would be the set of rationals, which is not open in the subspace topology induced on D. ☐

Discontinuity is a fundamental obstacle, but focusing our attention on settings admitting almost continuous versions will rule out this more trivial way of producing noncomputable conditional distributions. We might still hope to be able to compute the conditional distribution when there is *some* version that is almost continuous. However we will show that even this is not possible in general.

6. Noncomputable almost continuous conditional distributions

In this section, we construct a pair of random variables (X, N) that is computable, yet whose conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ is not computable, despite the existence of a $\mathbf{P}_{\mathsf{X}}$-almost continuous version.

Let $h : \mathbb{N} \to \mathbb{N} \cup \{\infty\}$ be the map given by $h(n) = \infty$ if the nth Turing machine (TM) does not halt (on input 0) and $h(n) = k$ if the nth TM halts (on input 0) at the kth step. The function h is lower semicomputable because we can compute all lower bounds: for all $k \in \mathbb{N}$, we can run the nth TM for k steps to determine whether $h(n) < k$, or $h(n) = k$, or $h(n) > k$. But h is not computable because any finite upper bound on $h(n)$ would imply that the nth TM halts, thereby solving the halting problem. However, we will define a computable random variable X such that conditioning on its value recovers h.

Let N be a computable geometric random variable, C a computable $\frac{1}{3}$-coin and U and V both computable uniform random variables on $[0, 1]$, all mutually independent. Let $\lfloor x \rfloor$ denote the greatest integer $y \leq x$. Note that $\lfloor 2^k \mathsf{V} \rfloor$ is uniformly distributed on $\{0, 1, 2, \ldots, 2^k - 1\}$. Consider the derived random variables

$$\mathsf{X}_k := \frac{2\lfloor 2^k \mathsf{V} \rfloor + \mathsf{C} + \mathsf{U}}{2^{k+1}} \tag{8}$$

for $k \in \mathbb{N}$. The limit $\mathsf{X}_\infty := \lim_{k\to\infty} \mathsf{X}_k$ exists with probability one and satisfies $\lim_{k\to\infty} \mathsf{X}_k = \mathsf{V}$ a.s. Finally, we define $\mathsf{X} := \mathsf{X}_{h(\mathsf{N})}$.

Proposition 24. *The random variable* X *is computable.*

Proof. Let $\{\mathsf{U}_n : n \in \mathbb{N}\}$ and $\{\mathsf{V}_n : n \in \mathbb{N}\}$ be the binary expansions of U and V, respectively. Because U and V are computable and almost surely irrational, it is not hard to show that these are computable random variables in $\{0, 1\}$, uniformly in n.

For each $k \geq 0$, define the random variable

$$\mathsf{D}_k = \begin{cases} \mathsf{V}_k, & h(\mathsf{N}) > k; \\ \mathsf{C}, & h(\mathsf{N}) = k; \\ \mathsf{U}_{k-h(\mathsf{N})-1}, & h(\mathsf{N}) < k. \end{cases} \tag{9}$$

Because h is lower semicomputable, $\{\mathsf{D}_k\}_{k\geq 0}$ are computable random variables, uniformly in k.

We now show that, with probability one, $\{\mathsf{D}_k\}_{k\geq 0}$ is the binary expansion of X, thus showing that X is itself a computable random variable.

There are two cases to consider:

First, conditioned on $h(\mathsf{N}) = \infty$, we have that $\mathsf{D}_k = \mathsf{V}_k$ for all $k \geq 0$. In fact, $\mathsf{X} = \mathsf{V}$ when $h(\mathsf{N}) = \infty$, and so the binary expansions match.

Condition on $h(\mathsf{N}) = m$ and let D denote the computable random real whose binary expansion is $\{\mathsf{D}_k\}_{k\geq 0}$. We must then show that $\mathsf{D} = \mathsf{X}_m$ a.s. Note that

$$\lfloor 2^m \mathsf{X}_m \rfloor = \lfloor 2^m \mathsf{V} \rfloor = \sum_{k=0}^{m-1} 2^{m-1-k} \mathsf{V}_k = \lfloor 2^m \mathsf{D} \rfloor, \tag{10}$$

and thus the binary expansions agree for the first m digits. Similarly, the next binary digit of X_m is C, followed by the binary expansion of U, thus agreeing with D for all $k \geq 0$. □

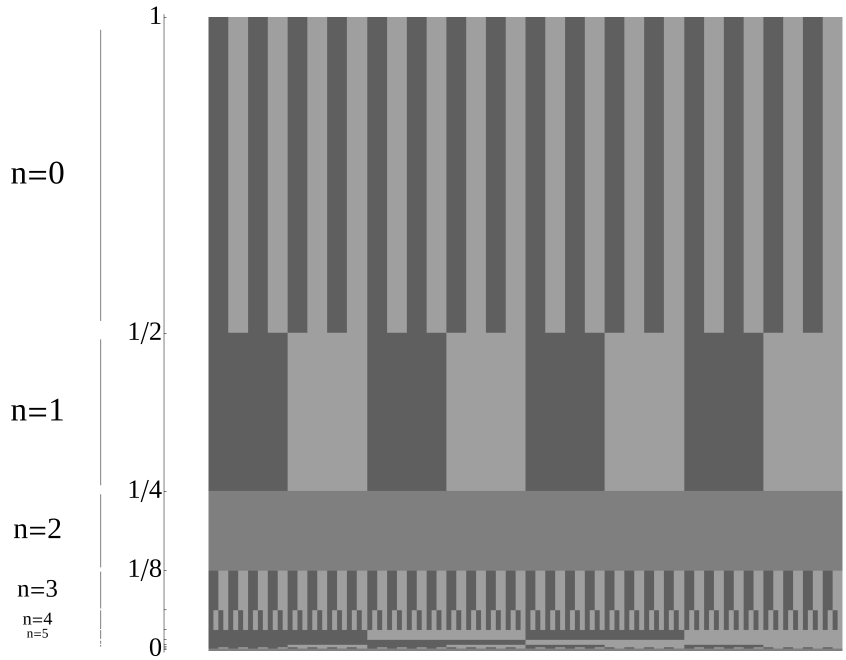

FIGURE 1. A visualization of (X, Y), where Y is uniformly distributed and $\mathsf{N} = \lfloor -\log_2 \mathsf{Y} \rfloor$. Regions that appear (at low resolution) to be uniform can suddenly be revealed (at higher resolutions) to be patterned. Deciding whether the pattern is in fact uniform (or below the resolution of this printer/display) is tantamount to solving the halting problem, but it is possible to sample from this distribution nonetheless. Note that this is not a plot of the density, but instead a plot where the darkness of each pixel is proportional to its measure.

We now show that $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ is not computable, despite the existence of a $\mathbf{P}_{\mathsf{X}}$-almost continuous version of $\mathbf{P}[\mathsf{N}|\mathsf{X}]$. We begin by characterizing the conditional density of X given N. Note that the constant function $p_{\mathsf{X}_\infty}(x) := 1$ is the density of X_∞ with respect to Lebesgue measure on $[0, 1]$.

Lemma 25. *For each $k \in \mathbb{N}$, the distribution of X_k admits a density p_{X_k} with respect to Lebesgue measure on $[0,1]$ given by*

$$p_{\mathsf{X}_k}(x) = \begin{cases} \frac{4}{3}, & \lfloor 2^{k+1}x \rfloor \text{ even;} \\ \frac{2}{3}, & \lfloor 2^{k+1}x \rfloor \text{ odd.} \end{cases} \tag{11}$$

As X_k admits a density with respect to Lebesgue measure on $[0,1]$ for all $k \in \mathbb{N} \cup \{\infty\}$, it follows that the conditional distribution of X given N admits a conditional density (with respect to Lebesgue measure on $[0,1]$) given by $p_{\mathsf{X}|\mathsf{N}}(x|n) := p_{\mathsf{X}_{h(n)}}(x)$. Each of these densities is continuous and bounded on the nondyadic reals, and so they can be combined to form an $\mathbf{P}_{\mathsf{X}}$-almost continuous version of the conditional distribution.

Lemma 26. *There is a $\mathbf{P}_{\mathsf{X}}$-almost continuous version of $\mathbf{P}[\mathsf{N}|\mathsf{X}]$.*

Lemma 27. *For all $m, n \in \mathbb{N}$ all versions κ of $\mathbf{P}[\mathsf{N}|\mathsf{X}]$, and $\mathbf{P}_{\mathsf{X}}$-almost all x, we have*

$$2^{m-n} \cdot \frac{\kappa(x, \{m\})}{\kappa(x, \{n\})} \in \begin{cases} \{\frac{1}{2}, 1, 2\}, & h(n), h(m) < \infty; \\ \{1\}, & h(n) = h(m) = \infty; \\ \{\frac{2}{3}, \frac{3}{4}, \frac{4}{3}, \frac{3}{2}\}, & \textit{otherwise.} \end{cases}$$

Let $H = \{n \in \mathbb{N} : h(n) < \infty\}$, i.e., the indices of the TMs that halt (on input 0). A classic result in computability theory [35] shows that the halting set H is not computable.

Proposition 28. *The conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ is not computable.*

Proof. Suppose the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ were computable. Let n be the index of some TM that halts (on input 0), i.e., for which $h(n) < \infty$, and consider any $m \in \mathbb{N}$.

Let κ be an arbitrary version of $\mathbf{P}[\mathsf{N}|\mathsf{X}]$, and let R be a $\mathbf{P}_{\mathsf{X}}$-measure one set on which κ is computable. Then the function

$$\tau(\cdot) := 2^{m-n} \cdot \frac{\kappa(\cdot, \{m\})}{\kappa(\cdot, \{n\})} \tag{12}$$

is also computable on R, by Corollary 20. By Lemma (27), there is a $\mathbf{P}_{\mathsf{X}}$-measure one subset $D \subseteq R$ on which τ exclusively takes values in the set $T = \{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1, \frac{4}{3}, \frac{3}{2}, 2\}$.

Although $\mathbf{P}_{\mathsf{X}}$-almost all reals in $[0,1]$ are in D, any particular real may not be. The following construction can be viewed as an attempt to compute a particular point $d \in D$ at which we can evaluate τ. In fact, we need only a finite approximation to d, because τ is computable on D and T is finite.

For each $t \in T$, let B_t be an ideal ball centered at t of radius less than $\frac{1}{6}$, so that $B_t \cap T = \{t\}$. By Definition 5, for each $t \in T$, there is a c.e. open set $U_t \subseteq [0,1]$ such that $\tau^{-1}(B_t) \cap R = U_t \cap R$. Because every open interval has positive $\mathbf{P}_{\mathsf{X}}$-measure, if U_t is nonempty, then $U_t \cap D$ is a positive $\mathbf{P}_{\mathsf{X}}$-measure set whose image is $\{t\}$. Thus, $\mathbf{P}_{\mathsf{X}}$-almost all $x \in U_t \cap R$ satisfy $\tau(x) = t$. As $\bigcup_t U_t$ has $\mathbf{P}_{\mathsf{X}}$-measure one, there is at least one $t \in T$ for which U_t is nonempty. Because each U_t is c.e. open, we can compute the index $\hat{t} \in T$ of some nonempty $U_{\hat{t}}$.

By Lemma 27 and the fact that $h(n) < \infty$, there are two cases:

(i) $\hat{t} \in \{\frac{1}{2}, 1, 2\}$, implying $h(m) < \infty$, or

(ii) $\hat{t} \in \{\frac{2}{3}, \frac{3}{4}, \frac{4}{3}, \frac{3}{2}\}$, implying $h(m) = \infty$.

Because m was arbitrary, and because the mth TM halts if and only if $h(m) < \infty$, we can use τ to compute the halting set H. Therefore if $\mathbf{P}[\mathsf{X}|\mathsf{N}]$ were computable, then H would be computable, a contradiction. □

Because this proof relativizes, we see that if the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$ is A-computable for some oracle A, then A computes the halting set H.

Computable operations map computable points to computable points, and so we obtain the following consequence.

Theorem 29. *The operation* $\mathsf{X}, \mathsf{Y} \mapsto \mathbf{P}[\mathsf{Y}|\mathsf{X}]$ *of conditioning a pair of real-valued random variables, even when restricted to pairs for which there exists a* $\mathbf{P}_{\mathsf{X}}$*-almost continuous version of the conditional distribution, is not computable.*

It is natural to ask whether this construction can be extended to produce a pair of computable random variables whose conditional distribution is noncomputable but has an *everywhere continuous* version. We provide such a strengthening in Appendix F.

Despite these results, many important questions remain: How badly noncomputable is conditioning, even restricted to these continuous settings? What is the computational *complexity* of conditioning on *efficiently* computable continuous random variables? In what restricted settings is conditioning *computable*? In the final section, we begin to address the latter of these.

7. Positive results

We now consider situations in which we can compute conditional distributions, with an aim towards explaining the widespread success of probabilistic methods. We begin with the setting of discrete random variables.

For simplicity, we will consider a computable discrete random variable to be a computable random variable in a computable metric space S where S is a countable set. Let X be such a computable random variable. Then for $x \in T$, the sets $\{\mathsf{X} = x\}$ and $\{\mathsf{X} \neq x\}$ are both c.e. open in $\{0,1\}^\infty$, disjoint, and obviously satisfy $\mathbf{P}\{\mathsf{X} = x\} + \mathbf{P}\{\mathsf{X} \neq x\} = 1$. Therefore, $\mathbf{P}\{\mathsf{X} = x\}$ is a computable real, uniformly in x. It is then not hard to show the following:

Lemma 30 (Conditioning on a discrete random variable)**.** *Let* X *and* Y *be computable random variables in computable metric spaces* S *and* T*, respectively, where* S *is a countable set. Then the conditional distribution* $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ *is computable, uniformly in* X, Y.

7.1. **Continuous, dominated, and other settings.**

The most common way to calculate conditional distributions is to use Bayes' rule, which requires the existence of a conditional density (and is thus known as the *dominated* setting within statistics). We first recall some elementary definitions.

Definition 31 (Density)**.** Let $(\Omega, \mathcal{A}, \nu)$ be a measure space and let $f : A \to \mathbb{R}^+$ be a measurable function. Then the function μ on $\mathcal{A}$ given by

$$\mu(A) = \int_A f d\nu \tag{13}$$

for $A \in \mathcal{A}$ is a measure on $(\Omega, \mathcal{A})$ and f is called a **density of** μ **(with respect to** ν**)**. Note that g is a density of μ with respect to ν if and only if $f = g$ ν-a.e.

Definition 32 (Conditional density)**.** Let X and Y be random variables in (complete separable) metric spaces, let $\kappa_{\mathsf{X}|\mathsf{Y}}$ be a version of $\mathbf{P}[\mathsf{X}|\mathsf{Y}]$, and assume that there is a measure $\nu \in \mathcal{M}(S)$ and measurable function $p_{\mathsf{X}|\mathsf{Y}}(x|y) : S \times T \to \mathbb{R}^+$ such that $p_{\mathsf{X}|\mathsf{Y}}(\cdot|y)$ is a density of $\kappa_{\mathsf{X}|\mathsf{Y}}(y, \cdot)$ with respect to ν for $\mathbf{P}_{\mathsf{Y}}$-a.e. y. That is,

$$\kappa_{\mathsf{X}|\mathsf{Y}}(y, A) = \int_A p_{\mathsf{X}|\mathsf{Y}}(x|y)\nu(dx) \tag{14}$$

for measurable sets $A \subseteq S$ and $\mathbf{P}_{\mathsf{Y}}$-almost all y. Then $p_{\mathsf{X}|\mathsf{Y}}(x|y)$ is called a **conditional density of** X **given** Y **(with respect to** ν**)**.

Common parametric families of distributions (e.g., exponential families like Gaussian, Gamma, etc.) admit conditional densities, and in these cases, the well-known Bayes' rule gives a formula for expressing the conditional distribution.

Lemma 33 (Bayes' rule [36, Thm. 1.13])**.** *Let* X *and* Y *be random variables as in Definition 14, let* $\kappa_{\mathsf{X}|\mathsf{Y}}$ *be a version of the conditional distribution* $\mathbf{P}[\mathsf{X}|\mathsf{Y}]$*, and assume that there exists a* conditional density $p_{\mathsf{X}|\mathsf{Y}}(x|y)$*. Then the function defined by*

$$\kappa_{\mathsf{Y}|\mathsf{X}}(x, B) := \frac{\int_B p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy)}{\int \; p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy)} \tag{15}$$

is a version of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$*.*

Comparing Bayes' rule (15) to the definition of conditional density (14), we see that the conditional density of Y given X (with respect to $\mathbf{P}_{\mathsf{Y}}$) is given by

$$p_{\mathsf{Y}|\mathsf{X}}(y|x) = \frac{p_{\mathsf{X}|\mathsf{Y}}(x|y)}{\int p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy)}. \tag{16}$$

Using the following well-known integration result, we can study when the conditional distribution characterized by Bayes' rule is computable.

Proposition 34 (Integration of computable functions ([28, Cor. 4.3.2]))**.** *Let* S *be a computable metric space, and* μ *a computable probability measure on* S*. Let* $f : S \to \mathbb{R}^+$ *be a bounded computable function. Then* $\int f d\mu$ *is a computable real, uniformly in* f*.*

Corollary 35 (Density and independence)**.** *Let* U*,* V*, and* Y *be computable random variables (in computable metric spaces), where* Y *is independent of* V *given* U*. Assume that there exists a conditional density* $p_{\mathsf{Y}|\mathsf{U}}(y|u)$ *of* Y *given* U *(with respect to*

ν) that is bounded and computable. Then the conditional distribution $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$ *is computable.*

Proof. Let $\mathsf{X} = (\mathsf{U},\mathsf{V})$. Then $p_{\mathsf{Y}|\mathsf{X}}(y|(u,v)) = p_{\mathsf{Y}|\mathsf{U}}(y|u)$ is the conditional density of Y given X (with respect to ν). Therefore, the computability of the integrand and the existence of a bound imply, by Proposition 34, that $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$ is computable. □

As an immediate corollary, we obtain the computability of the following common situation in probabilitic modeling: where the observed random variable has been corrupted by independent absolutely continuous noise.[4]

Corollary 36 (Independent noise)**.** *Let* U *be a computable random variable in a computable metric space and let* V *and* E *be computable random variables in* $\mathbb{R}$*. Define* $\mathsf{Y} = \mathsf{U} + \mathsf{E}$*. If* $\mathbf{P}_{\mathsf{E}}$ *is absolutely continuous with a bounded computable density* p_{E} *and* E *is independent of* U *and* V *then the conditional distribution* $\mathbf{P}[(\mathsf{U},\mathsf{V}) \mid \mathsf{Y}]$ *is computable.*

Proof. We have that

$$p_{\mathsf{Y}|\mathsf{U}}(y|u) = p_{\mathsf{E}}(y-u) \tag{17}$$

is the conditional density of Y given U (with respect to Lebesgue measure). The result then follows from Corollary 35. □

This result is analogous to a classical theorem of information theory. Hartley [39] and Shannon [40] show that the capacity of a continuous real-valued channel without noise is infinite, yet the addition of, e.g., Gaussian noise with $\epsilon > 0$ variance causes the channel capacity to drop to a finite amount. The Gaussian noise prevents too much information from being encoded in the bits of the real number. Similarly, the amount of information in a continuous observation is too much in general for a computer to be able to update a probabilistic model. However, the addition of noise with enough structure is sufficient for making conditioning possible on a computer.

The computability of conditioning with noise, coupled with the noncomputability of conditioning in general, has significant implications for our ability to recover a signal when noise is added, and suggests several interesting questions. For example, suppose we have a uniformly computable sequence of noise $\{\mathsf{E}_n\}_{n\in\mathbb{N}}$ with absolutely continuous, uniformly computable densities such that the magnitude of the densities goes to 0 in some sufficiently nice way, and consider $\mathsf{Y}_n := \mathsf{U} + \mathsf{E}_n$. Such a situation

[4]Note that Corollary 36 implies that noiseless observations cannot always be computably approximated by noisy ones. For example, even though an observation corrupted with zero mean Gaussian noise with standard deviation σ may recover the original condition as $\sigma \to 0$, by our main noncomputability result (Theorem 29) one cannot, in general, compute how small σ must be in order to bound the error introduced by noise.

Myhill [37] exhibits a computable function $[0,1] \to \mathbb{R}$ whose derivative is continuous, but not computable, and Pour-El and Richards [38, Ch. 1, Thm. 2] show that a twice continuously differentiable computable function has a computable derivative. Therefore, noise with a sufficiently smooth distribution has a computable density, and by Corollary 36, a computable random variable corrupted by such noise still admits a computable conditional distribution.

could arise, e.g., when we have a signal with noise but some way to reduce the noise over time.

When there is a continuous version of $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$, we have

$$\lim_{n\to\infty} \mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}_n] = \mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]. \tag{18}$$

However, we know that the right side is, in general, noncomputable, despite the fact that each term in the limit on the left side is computable. This suggests that we should be unable to recover any information about $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$ from $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}_n]$ for any particular n.

This raises several questions, such as: What do bounds on how fast the sequence $\{\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}_n]\}_{n\in\mathbb{N}}$ converges to $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$ tell us about the computability of $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$? What conditions on the relationship between U and the sequence $\{\mathsf{E}_n\}_{n\in\mathbb{N}}$ will allow us to recover information about $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}]$ from individual distributions $\mathbf{P}[(\mathsf{U},\mathsf{V})|\mathsf{Y}_n]$?

7.2. **Conclusion.** There is no generic algorithm for conditioning on continuous random variables, and yet there are many particular situations in which practitioners have developed algorithms (sometimes even quite efficient) for computing conditional probabilities. An important challenge for computer science theory is to characterize broadly-applicable circumstances in which conditioning on computable random variables *is* possible. The positive results in this section provide several such settings.

Freer and Roy [31] show how to compute conditional distributions in the setting of *exchangeable sequences*. A classic result by de Finetti shows that exchangeable sequences of random variables are in fact conditionally i.i.d. sequences, conditioned on a random measure, often called the *directing random measure*. Freer and Roy describe how to transform an algorithm for sampling an *exchangeable sequence* into a rule for computing the posterior distribution of the directing random measure given observations. The result is a corollary of a computable version of de Finetti's theorem [41], and covers a wide range of common scenarios in nonparametric Bayesian statistics (often where no conditional density exists). The search for additional positive results is an exciting future avenue for logic and theoretical computer science.

Acknowledgments

Some of the results in this extended abstract first appeared in [42]. The authors would like to thank Jeremy Avigad, Henry Cohn, Leslie Kaelbling, Vikash Mansinghka, Hartley Rogers, Michael Sipser, and Joshua Tenenbaum for helpful discussions, Bill Thurston for a useful comment regarding Lemma 45, and Quinn Culver, Noah Goodman, Leslie Kaelbling, Bjørn Kjos-Hanssen, Timothy O'Donnell, and Geoff Patterson for comments on a draft. C.E.F. is partially supported by the National Science Foundation under Grant No. 0901020. D.M.R. is partially supported by graduate fellowships from the National Science Foundation and MIT Lincoln Laboratory.

References

[1] D. Poole, "Representing Bayesian networks within probabilistic Horn abduction," in *Proc. of the 7th Conf. on Uncertainty in Artificial Intelligence*, 1991, pp. 271–278.

[2] A. Pfeffer, "IBAL: A probabilistic rational programming language," in *Proc. of the 17th Int. Joint Conf. on Artificial Intelligence*. Morgan Kaufmann Publ., 2001, pp. 733–740.

[3] S. Park, F. Pfenning, and S. Thrun, "A probabilistic language based on sampling functions," *ACM Trans. Program. Lang. Syst.*, vol. 31, no. 1, pp. 1–46, 2008. http://dx.doi.org/10.1145/1452044.1452048

[4] N. D. Goodman, V. K. Mansinghka, D. M. Roy, K. Bonawitz, and J. B. Tenenbaum, "Church: a language for generative models," in *Proc. of the 24th Conf. on Uncertainty in Artificial Intelligence*, 2008. http://uai2008.cs.helsinki.fi/UAI_camera_ready/goodman.pdf

[5] O. Kiselyov and C. Shan, "Embedded probabilistic programming," in *Domain-Specific Languages*, ser. Lecture Notes in Computer Science, W. M. Taha, Ed., vol. 5658. Springer, 2009, pp. 360–384. http://dx.doi.org/10.1007/978-3-642-03034-5_17

[6] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, 1933.

[7] T. Tjur, *Conditional probability distributions*, ser. Lecture Notes, no. 2. University of Copenhagen, Copenhagen: Institute of Mathematical Statistics, 1974.

[8] ——, *A Constructive definition of conditional distributions*, ser. Preprint 13. University of Copenhagen, Copenhagen: Institute of Mathematical Statistics, 1975.

[9] ——, *Probability based on Radon measures*, ser. Wiley Series in Probability and Mathematical Statistics. Chichester: John Wiley & Sons Ltd., 1980.

[10] J. Pfanzagl, "Conditional distributions as derivatives," *Ann. Probab.*, vol. 7, no. 6, pp. 1046–1050, 1979. http://www.jstor.org/pss/2243105

[11] M. M. Rao, "Paradoxes in conditional probability," *J. Multivariate Anal.*, vol. 27, no. 2, pp. 434–446, 1988. http://dx.doi.org/10.1016/0047-259X(88)90140-6

[12] ——, *Conditional measures and applications*, 2nd ed., ser. Pure and Applied Mathematics. Chapman & Hall/CRC, Boca Raton, FL, 2005, vol. 271. http://dx.doi.org/10.1201/9781420027433

[13] L. A. Levin, "Average case complete problems," *SIAM J. Comput.*, vol. 15, no. 1, pp. 285–286, 1986. http://dx.doi.org/10.1137/0215020

[14] S. Ben-David, B. Chor, O. Goldreich, and M. Luby, "On the theory of average case complexity," *J. Comput. System Sci.*, vol. 44, no. 2, pp. 193–219, 1992. http://dx.doi.org/10.1016/0022-0000(92)90019-F

[15] T. Yamakami, "Polynomial time samplable distributions," *J. Complexity*, vol. 15, no. 4, pp. 557–574, 1999. http://dx.doi.org/10.1006/jcom.1999.0523

[16] D. N. Osherson, M. Stob, and S. Weinstein, "Mechanical learners pay a price for Bayesianism," *J. Symbolic Logic*, vol. 53, no. 4, pp. 1245–1251, 1988. http://dx.doi.org/10.2307/2274619

[17] E. M. Gold, "Language identification in the limit," *Inform. and Control*, vol. 10, no. 5, pp. 447–474, 1967. http://dx.doi.org/10.1016/S0019-9958(67)91165-5

[18] H. Putnam, "Trial and error predicates and the solution to a problem of Mostowski," *J. Symbolic Logic*, vol. 30, pp. 49–57, 1965. http://dx.doi.org/10.2307/2270581

[19] R. J. Solomonoff, "A formal theory of inductive inference II," *Inform. and Control*, vol. 7, pp. 224–254, 1964. http://dx.doi.org/10.1016/S0019-9958(64)90131-7

[20] A. K. Zvonkin and L. A. Levin, "The complexity of finite objects and the basing of the concepts of information and randomness on the theory of algorithms," *Uspehi Mat. Nauk*, vol. 25, no. 6 (156), pp. 85–127, 1970. http://dx.doi.org/10.1070/RM1970v025n06ABEH001269

[21] M. Hutter, "On universal prediction and Bayesian confirmation," *Theoret. Comput. Sci.*, vol. 384, no. 1, pp. 33–48, 2007. http://dx.doi.org/10.1016/j.tcs.2007.05.016

[22] H. Takahashi, "On a definition of random sequences with respect to conditional probability," *Inform. and Comput.*, vol. 206, no. 12, pp. 1375–1382, 2008. http://dx.doi.org/10.1016/j.ic.2008.08.003

[23] M. Braverman, "On the complexity of real functions," in *FOCS'05—Proc. of the 46th Ann. IEEE Symp. on Foundations of Comput. Sci.* IEEE Computer Society, 2005, pp. 155–164. http://dx.doi.org/10.1109/SFCS.2005.58

[24] M. Braverman and S. Cook, "Computing over the reals: foundations for scientific computing," *Notices Amer. Math. Soc.*, vol. 53, no. 3, pp. 318–329, 2006. http://www.ams.org/notices/200603/fea-cook.pdf

[25] H. Rogers, Jr., *Theory of recursive functions and effective computability*, 2nd ed. Cambridge, MA: MIT Press, 1987.

[26] K. Weihrauch, *Computable analysis: an introduction.* Berlin: Springer-Verlag, 2000.

[27] A. Nies, *Computability and randomness*, ser. Oxford Logic Guides. Oxford: Oxford University Press, 2009, vol. 51. http://dx.doi.org/10.1093/acprof:oso/9780199230761.001.0001

[28] M. Hoyrup and C. Rojas, "Computability of probability measures and Martin-Löf randomness over metric spaces," *Inform. and Comput.*, vol. 207, no. 7, pp. 830–847, 2009. http://dx.doi.org/10.1016/j.ic.2008.12.009

[29] S. Galatolo, M. Hoyrup, and C. Rojas, "Effective symbolic dynamics, random points, statistical behavior, complexity and entropy," *Inform. and Comput.*, vol. 208, no. 1, pp. 23–41, 2010. http://dx.doi.org/10.1016/j.ic.2009.05.001

[30] P. Gács, "Uniform test of algorithmic randomness over a general space," *Theoret. Comput. Sci.*, vol. 341, no. 1-3, pp. 91–137, 2005. http://dx.doi.org/10.1016/j.tcs.2005.03.054

[31] C. E. Freer and D. M. Roy, "Posterior distributions are computable from predictive distributions," in *Proc. of the 13th Int. Conf. on Artificial Intelligence and Statistics (AISTATS 2010) (Y. W. Teh and M. Titterington, eds.), JMLR: W&CP 9*, 2010, pp. 233–240. http://jmlr.csail.mit.edu/proceedings/papers/v9/freer10a/freer-roy10a.pdf

[32] I. Mann, "Probabilistic recursive functions," *Trans. Amer. Math. Soc.*, vol. 177, pp. 447–467, 1973. http://dx.doi.org/10.2307/1996608

[33] D. E. Knuth and A. C. Yao, "The complexity of nonuniform random number generation," in *Algorithms and complexity (Proc. Sympos., Carnegie-Mellon Univ., Pittsburgh, Pa., 1976).* New York: Academic Press, 1976, pp. 357–428.

[34] O. Kallenberg, *Foundations of modern probability*, 2nd ed. New York: Springer, 2002.

[35] A. M. Turing, "On computable numbers, with an application to the Entscheidungsproblem," *Proc. London Math. Soc.*, vol. 42, no. 1, pp. 230–265, 1936. http://dx.doi.org/10.1112/plms/s2-42.1.230

[36] M. J. Schervish, *Theory of statistics*, ser. Springer Series in Statistics. New York: Springer-Verlag, 1995.

[37] J. Myhill, "A recursive function, defined on a compact interval and having a continuous derivative that is not recursive," *Michigan Math. J.*, vol. 18, pp. 97–98, 1971. http://projecteuclid.org/euclid.mmj/1029000631

[38] M. B. Pour-El and J. I. Richards, *Computability in analysis and physics*, ser. Perspectives in Mathematical Logic. Berlin: Springer-Verlag, 1989. http://projecteuclid.org/euclid.pl/1235422916

[39] R. Hartley, "Transmission of information," *Bell System Technical J.*, pp. 535–563, July 1928.

[40] C. E. Shannon, "Communication in the presence of noise," *Proc. I.R.E.*, vol. 37, pp. 10–21, 1949. http://dx.doi.org/10.1109/PROC.1984.12998

[41] C. E. Freer and D. M. Roy, "Computable exchangeable sequences have computable de Finetti measures," in *Mathematical Theory and Computational Practice (CiE 2009), Proc. of the 5th Conf. on Computability in Europe*, ser. Lecture Notes in Comput. Sci., K. Ambos-Spies, B. Löwe, and W. Merkle, Eds., vol. 5635. Springer, 2009, pp. 218–231. http://dx.doi.org/10.1007/978-3-642-03073-4_23

[42] N. L. Ackerman, C. E. Freer, and D. M. Roy, "On the computability of conditional probability," 2010, preprint. http://arxiv.org/abs/1005.3014

[43] K.-I. Ko and H. Friedman, "Computational complexity of real functions," *Theoret. Comput. Sci.*, vol. 20, no. 3, pp. 323–352, 1982. http://dx.doi.org/10.1016/S0304-3975(82)80003-0

[44] M. Schröder, "Admissible representations for probability measures," *Math. Log. Q.*, vol. 53, no. 4-5, pp. 431–445, 2007. http://dx.doi.org/10.1002/malq.200710010

Appendix A. Proofs from Section 2

Proof of Proposition 9: Assume X is continuous. Let $Z_0 := \mathsf{X}^{-1}(0)$ and $Z_1 := \mathsf{X}^{-1}(1)$. Then $\{0,1\}^\infty = Z_0 \cup Z_1$, and so both are closed (as well as open). The compactness of $\{0,1\}^\infty$ implies that these closed subspaces are also compact, and so Z_0 and Z_1 can each be written as the finite disjoint union of clopen basis elements. But each of these elements has dyadic measure, hence their sum cannot be either α or $1-\alpha$, contradicting the fact that $\mathbf{P}(Z_1) = 1 - \mathbf{P}(Z_0) = \alpha$. □

Appendix B. Proofs from Section 3

Proof of Lemma 16: Fix a measurable set $A \subseteq Y$ and define $g(\cdot) := \kappa_1(\cdot, A) - \kappa_2(\cdot, A)$. We know that $g = 0$ $\mathbf{P}_\mathsf{X}$-a.e., and also that g is continuous at x. Assume, for the purpose of contradiction, that $g(x) = \epsilon > 0$. By continuity, there is an open neighborhood B of x, such that $g(B) \in (\frac{\epsilon}{2}, \frac{3\epsilon}{2})$. But $x \in \mathrm{supp}(\mathbf{P}_\mathsf{X})$, hence $\mathbf{P}_\mathsf{X}(B) > 0$, contradicting $g = 0$ $\mathbf{P}_\mathsf{X}$-a.e. □

Proof of Lemma 17: The function κ, given by

$$\kappa(x, B) := \mathbf{P}\{\mathsf{Y} \in B \mid \mathsf{X} = x\} \tag{19}$$

for all $x \in R$ and $\kappa(x, \cdot) = \nu(\cdot)$ for $x \notin R$, is well-defined because $\mathbf{P}\{\mathsf{X} = x\} > 0$ for all $x \in R$, and so the right hand side of Equation (19) is well-defined. Furthermore, $\mathbf{P}\{\mathsf{X} \in R\} = 1$ and so κ is characterized by Equation (19) for almost all x. Finally,

$$\int_A \kappa(x, B)\, \mathbf{P}_\mathsf{X}(dx) \tag{20}$$

$$= \sum_{x \in R \cap A} \mathbf{P}\{\mathsf{Y} \in B \mid \mathsf{X} = x\}\, \mathbf{P}\{\mathsf{X} = x\} \tag{21}$$

$$= \sum_{x \in R \cap A} \mathbf{P}\{\mathsf{Y} \in B,\ \mathsf{X} = x\}, \tag{22}$$

which is equal to $\mathbf{P}\{\mathsf{Y} \in B,\ \mathsf{X} \in A\}$, and so κ is a version of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$. □

Appendix C. Proofs from Section 4

In the following proof we use a correspondence [26, Ch. 9.4] between type-two effectivity and oracle computability: A function f is computable on D if and only if $f(x)$ is uniformly computable relative to an oracle for $x \in D$, where the oracle encodes a convergent sequence of ideal balls containing x.

Proof of Lemma 19: Assume that $\kappa(\cdot, A)$ is lower semicomputable on D uniformly in

A. In other words, for $s \in D$, the real number $\kappa(s, A)$ is uniformly c.e. relative to s, uniformly in A, and so by Theorem 10, the measure $\kappa(s, \cdot)$ is uniformly computable relative to s. Hence ϕ_κ is computable on D.

We now prove the other direction. Let ϕ_κ be as in Definition 18, fix a rational $q \in (0, 1)$ and c.e. open set A, and define $I = (q, 1]$. Then $\kappa(\cdot, A)^{-1}[I] = \phi_\kappa^{-1}[P]$, where

$$P := \{\mu \in \mathcal{M}_1(T) \,:\, \mu(A) > q\}. \tag{23}$$

This is an open set in the weak topology induced by the Prokhorov metric (see [44, Lem. 3.2]). We now show that P is, in fact, c.e. open, uniformly in q and A.

Consider the set $\mathcal{D}$ of all probability measures on $(T, \delta_T, \mathcal{D}_T)$ that are concentrated on a finite subset and where the measure of each atom is rational, i.e., every $\nu \in \mathcal{D}$ can be written as $\nu = \sum_{i=1}^{k} q_i \delta_{t_i}$ for some rationals $q_i \geq 0$ such that $\sum_{i=1}^{k} q_i = 1$ and some points $t_i \in \mathcal{D}_T$. Gács [30, §B.6.2] shows that $\mathcal{D}$ is dense in the Prokhorov metric and makes $\mathcal{M}_1(T)$ a computable metric space.

Let $\nu \in \mathcal{D}$, and let R be the finite set on which it concentrates. Gács [30, Prop. B.17] characterizes the ideal ball E centered at ν with rational radius $\epsilon > 0$ as the set of measures $\mu \in \mathcal{M}_1(T)$ such that

$$\mu(C^\epsilon) > \nu(C) - \epsilon \tag{24}$$

for all $C \subseteq R$, where $C^\epsilon = \bigcup_{t \in C} B(t, \epsilon)$.

We can write $A = \bigcup_{n \in \mathbb{N}} B(d_n, r_n)$ for a computable sequence of ideal balls in T with centers $d_n \in \mathcal{D}_T$ and rational radii r_n. Let $A_m = \bigcup_{n \leq m} B(d_n, r_n)$. Then $A_m \subseteq A_{m+1}$ and $A = \bigcup_m A_m$. Writing

$$P_m := \{\mu \in \mathcal{M}_1(T) \,:\, \mu(A_m) > q\}, \tag{25}$$

we have $P = \bigcup_m P_m$. In order to show that P is c.e. open, it suffices to show that P_m is c.e. open, uniformly in m. It is straightforward to show that $E \subseteq P_m$ if and only if $\nu(C_m) \geq q + \epsilon$, where

$$C_m := \{t \in R \,:\, B(t, \epsilon) \subseteq A_m\}. \tag{26}$$

Note that C_m is a decidable subset of R (uniformly in m and E) and that $\nu(C_m)$ is a rational and so we can decide whether $E \subseteq P_m$, showing that P is c.e. open.

Hence, by the computability of ϕ_κ, there is a c.e. open set V, uniformly computable in P (and hence I) such that $\phi_\kappa^{-1}[P] \cap D = V \cap D$. But then, we have that $\kappa(\cdot, A)^{-1}[I] \cap D = V \cap D$, and so $\kappa(\cdot, A)$ is computable on D. □

Proof of Corollary 20: If B a c.e. open set, $\kappa(\cdot, B)$ is lower semicomputable on D and $\kappa(\cdot, T \setminus B) = 1 - \kappa(\cdot, B)$ is upper semicomputable on D. Because A is decidable, both A and $T \setminus A$ are c.e. open, and so $\kappa(\cdot, A)$ is computable on D. □

Appendix D. Proofs from Section 6

Proof of Lemma 25: Let $k \in \mathbb{N}$. With probability one, the integer part of $2^{k+1}\mathsf{X}_k$ is $2\lfloor 2^k \mathsf{V} \rfloor + \mathsf{C}$ while the fractional part is U. Therefore, the distribution of $2^{k+1}\mathsf{X}_k$ (and hence X_k) admits a piecewise constant density with respect to Lebesgue measure.

In particular, $\lfloor 2^{k+1}\mathsf{X}_k \rfloor \equiv \mathsf{C} \pmod 2$ almost surely and $2\lfloor 2^k \mathsf{V} \rfloor$ is independent of C and uniformly distributed on $\{0, 2, \dots, 2^{k+1}-2\}$. Therefore,

$$\mathbf{P}\{\lfloor 2^{k+1}\mathsf{X}_k \rfloor = \ell\} = 2^{-k} \cdot \begin{cases} \frac{2}{3}, & \ell \text{ even;} \\ \frac{1}{3}, & \ell \text{ odd,} \end{cases} \tag{27}$$

for every $\ell \in \{0, 1, \dots, 2^{k+1}-1\}$. It follows immediately that the density p of $2^{k+1}\mathsf{X}_k$ with respect to Lebesgue measure on $[0, 2^{k+1}]$ is given by

$$p(x) = 2^{-k} \cdot \begin{cases} \frac{2}{3}, & \lfloor x \rfloor \text{ even;} \\ \frac{1}{3}, & \lfloor x \rfloor \text{ odd.} \end{cases} \tag{28}$$

and so the density of X_k is obtained by rescaling: $p_{\mathsf{X}_k}(x) = 2^{k+1} \cdot p(2^{k+1}x)$. □

Proof of Lemma 26: By Bayes' rule (Lemma 33), the probability kernel κ given by

$$\kappa(x, B) := \frac{\sum_{n \in B} p_{\mathsf{X}|\mathsf{N}}(x|n)\, \mathbf{P}\{\mathsf{N} = n\}}{\sum_{n \in \mathbb{N}} p_{\mathsf{X}|\mathsf{N}}(x|n)\, \mathbf{P}\{\mathsf{N} = n\}} \tag{29}$$

is a version of the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{X}]$. Every nondyadic real $x \in [0,1]$ is a point of continuity of $p_{\mathsf{X}|\mathsf{N}}$, and so the kernel κ is $\mathbf{P}_{\mathsf{X}}$-almost continuous by Lemma 44. □

Proof of Lemma 27: Let κ be as in Equation (29). Let $m, n \in \mathbb{N}$. Then

$$\begin{aligned} \tau(x) &:= 2^{m-n} \cdot \frac{\kappa(x, \{m\})}{\kappa(x, \{n\})} \\ &= 2^{m-n} \cdot \frac{p_{\mathsf{X}|\mathsf{N}}(x|m)\mathbf{P}\{\mathsf{N} = m\}}{p_{\mathsf{X}|\mathsf{N}}(x|n)\mathbf{P}\{\mathsf{N} = n\}} \\ &= \frac{p_{\mathsf{X}_{h(m)}}(x)}{p_{\mathsf{X}_{h(n)}}(x)}. \end{aligned}$$

For $k < \infty$, $p_{\mathsf{X}_k}(x) \in \{\frac{2}{3}, \frac{4}{3}\}$ for $\mathbf{P}_{\mathsf{X}}$-almost all x. Therefore, for $h(n), h(m) < \infty$, $\tau(x) \in \{\frac{1}{2}, 1, 2\}$ for $\mathbf{P}_{\mathsf{X}}$-almost all x. As $p_{\mathsf{X}_\infty}(x) = 1$ for $\mathbf{P}_{\mathsf{X}}$-almost all x, we have $\tau(x) = 1$ for $\mathbf{P}_{\mathsf{X}}$-almost all x when $h(n) = h(m) = \infty$ and $\tau(x) \in \{\frac{2}{3}, \frac{3}{4}, \frac{4}{3}, \frac{3}{2}\}$ otherwise. □

Appendix E. Proofs from Section 7

Definition 37 (Computable probability space [29, Def. 2.4.1])**.** A **computable probability space** is a pair (S, μ) where S is a computable metric space and μ is a computable probability measure on S.

Let (S, μ) be a computable probability space. The measure $\mu(A)$ of a c.e. open set A is always a c.e. real, but is not in general a computable real. It will be useful to understand those sets A whose measure $\mu(A)$ is a computable real.

Definition 38 (Almost decidable set [29, Def. 3.1.3])**.** Let S be a computable metric space and let $\mu \in \mathcal{M}_1(S)$ be a probability distribution on S. A (Borel) measurable subset $A \subseteq S$ is said to be **μ-almost decidable** when there are two c.e. open sets U and V such that $U \subseteq A$ and $V \subseteq S \setminus A$ and $\mu(U) + \mu(V) = 1$.

When μ is a computable measure and A is an arbitrary c.e. open set, then $\mu(A)$ is merely a c.e. real. However, when A is a μ-almost decidable set, then $\mu(A)$ is also a co-c.e. real, hence computable ([29, Prop. 3.1.1]).

We now show that every c.e. open set is the union of a computable sequence of almost decidable subsets.

Lemma 39 (Almost decidable subsets)**.** *Let (S, μ) be a computable probability space and let V be a c.e. open set. Then, uniformly in (the index of) V, we can compute a sequence of μ-almost decidable sets $\{V_k\}_{k\in\mathbb{N}}$ such that, for each k, $V_k \subseteq V_{k+1}$, and $\bigcup_{k\in\mathbb{N}} V_k = V$.*

Proof. Note that the finite union or intersection of almost decidable sets is almost decidable. By [29, Thm. 3.1.2] there is a computable sequence $\{r_j\}_{j\in\mathbb{N}}$ of reals, dense in $\mathbb{R}^+$ and for which the balls $\{B(d_i, r_j)\}_{i,j\in\mathbb{N}}$ form a basis of μ-almost decidable sets. Let $E \subseteq \mathbb{N}$ be a c.e. set such that $V = \bigcup_{i\in E} B_i$, where $\{B_i\}_{i\in\mathbb{N}}$ is the enumeration of the ideal balls of S. Consider the set $F = \{(i, j) : \exists k \in E \text{ with } \overline{B}(d_i, r_j) \subseteq B_k\}$ of indices (i, j) such that the closure of the ball $B(d_i, r_j)$ lies strictly within an ideal ball within V. Then F is c.e. and, by the density of the sequence $\{r_j\}$, we have $V = \bigcup_{(i,j)\in F} B(d_i, r_j)$. Consider the finite union $V_k := \bigcup_{\{(i,j)\in F\,:\,i,j\le k\}} B(d_i, r_j)$, which is almost decidable. By construction, for each k, $V_k \subseteq V_{k+1}$, and $\bigcup_{k\in\mathbb{N}} V_k = V$. ☐

Conversely, we have the following characterization of computable measures.

Corollary 40. *Let S be a computable metric space and let $\mu \in \mathcal{M}_1(S)$ be a probability measure on S. Then μ is computable if the measure $\mu(A)$ of every μ-almost decidable set A is a computable real, uniformly in A.*

Proof. Let V be a c.e. open set of S. By Theorem 10, it suffices to show that $\mu(V)$ is a c.e. real, uniformly in V. By Lemma 39, we can compute a nested sequence $\{V_k\}_{k\in\mathbb{N}}$ of μ-almost decidable sets whose union is V. Because V is open, $\mu(V) = \sup_{k\in\mathbb{N}} \mu(V_k)$. By hypothesis, $\mu(V_k)$ is a computable real for each k, and so the supremum is a c.e. real, as desired. ☐

Recall the definition of conditional probability (Definition 12). When μ is computable and A is an almost decidable set, the conditional probability given A is computable.

Lemma 41 (Conditional probability given an almost decidable set [29, Prop. 3.1.2])**.** *Let (S, μ) be a computable probability space and let A be an almost decidable subset of S satisfying $\mu(A) > 0$. Then $\mu(\cdot|A)$ is computable.*

Proof. By Corollary 40, it suffices to show that $\frac{\mu(B\cap A)}{\mu(A)}$ is computable for an almost decidable set B. But then $B \cap A$ is almost decidable and so its measure, the numerator, is a computable real. The denominator is likewise the measure of an almost decidable set, hence a computable real. Finally, the ratio of two computable reals is computable. □

The equation $\mathbf{P}\{\mathsf{Y} \in A \mid \mathsf{X} = x\} = \frac{\mathbf{P}\{\mathsf{Y}\in A,\ \mathsf{X}=x\}}{\mathbf{P}\{\mathsf{X}=x\}}$ gives a recipe for calculating the conditional distribution of a discrete random variable. However, the event $\{\mathsf{X} = x\}$ is not necessarily even an open set, and so in order to compute the conditional distribution given a discrete random variable, we need additional computability hypotheses on its support.

Definition 42 (Computably discrete set)**.** Let S be a computable metric space. We say that a (finite or countably infinite) subset $D \subseteq S$ is **computably discrete** when, for some enumeration $d_0, d_1, \ldots$ of D (possibly with repetition) there is a computable function $f : \mathbb{N} \to \mathbb{N}$ such that each d_j is the unique point of D in the ideal ball $B_{f(j)}$.

Lemma 43. *Let (S, μ) be a computable probability space, and let D be a computably discrete subset of S. Define $D_+ := \{d \in D : \mu(\{d\}) > 0\}$. There is a partial function $g : S \to \mathbb{N}$, computable on D_+, such that for $d \in D_+$, the integer $g(d)$ is (the index of) a μ-almost decidable set containing d and no other points of D.*

Proof. Immediate from Lemma 39. □

Proof of Lemma 30: We will actually prove the result in the slightly more general setting where the "discrete" random variable takes values in a possibly uncountable space S, but that we have a computable handle on the discrete subspace on which the random variable concentrates. In particular, assume that $\mathbf{P}_{\mathsf{X}}$ is concentrated on a computably discrete set D (see Definition 42).

Define $D_+ := \{d \in D : \mathbf{P}_{\mathsf{X}}(\{d\}) > 0\}$, and let g be a computable partial function that assigns each point in D_+ a $\mathbf{P}_{\mathsf{X}}$-almost decidable set covering it, as in Lemma 43. Let $A_{g(d)}$ denote the $\mathbf{P}_{\mathsf{X}}$-almost decidable set coded by $g(d)$.

Because X is also concentrated on D_+, a version κ of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is an arbitrary kernel $\kappa(\cdot, \cdot)$ that satisfies

$$\kappa(d, \cdot) = \mathbf{P}\{\mathsf{Y} \in \cdot \mid \mathsf{X} = d\} \tag{30}$$

for every $d \in D_+$ (as in Lemma 17).

Let $d \in D_+$ be arbitrary. The set $A_{g(d)}$ contains exactly one point of positive $\mathbf{P}_{\mathsf{X}}$-measure, and so the events $\{\mathsf{X} = d\}$ and $\{\mathsf{X} \in A_{g(d)}\}$ are positive $\mathbf{P}_{\mathsf{X}}$-measure sets that differ by a $\mathbf{P}_{\mathsf{X}}$-null set. Hence

$$\mathbf{P}\{\mathsf{Y} \in \cdot \mid \mathsf{X} = d\} = \mathbf{P}\{\mathsf{Y} \in \cdot \mid \mathsf{X} \in A_{g(d)}\}. \tag{31}$$

The event $\{\mathsf{X} \in A_{g(d)}\}$ is $\mathbf{P}$-almost decidable, and so the measure $\mathbf{P}\{\mathsf{Y} \in \cdot \mid \mathsf{X} \in A_{g(d)}\}$ is computable, by Lemma 41.

Thus the partial function mapping $S \to \mathcal{M}_1(T)$ by

$$x \mapsto \mathbf{P}\{\mathsf{Y} \in \cdot \mid \mathsf{X} \in A_{g(x)}\} \tag{32}$$

is computable on D_+, a subset of S of $\mathbf{P}_{\mathsf{X}}$-measure one, and so the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$ is computable. □

Proof of Lemma 33: By Definition 14 and Fubini's theorem, for Borel sets $A \subseteq S$ and $B \subseteq T$, we have that

$$\mathbf{P}\{\mathsf{X} \in A, \mathsf{Y} \in B\} = \int_B \kappa_{\mathsf{X}|\mathsf{Y}}(y, A)\mathbf{P}_{\mathsf{Y}}(dy) \tag{33}$$

$$= \int_B \left(\int_A p_{\mathsf{X}|\mathsf{Y}}(x|y)\nu(dx) \right) \mathbf{P}_{\mathsf{Y}}(dy) \tag{34}$$

$$= \int_A \left(\int_B p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy) \right) \nu(dx). \tag{35}$$

Taking $B = T$, we have

$$\mathbf{P}_{\mathsf{X}}(A) = \int_A \left(\int p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy) \right) \nu(dx). \tag{36}$$

Therefore,

$$\int_A \kappa_{\mathsf{Y}|\mathsf{X}}(x, B)\mathbf{P}_{\mathsf{X}}(dx) \tag{37}$$

$$= \int_A \kappa_{\mathsf{Y}|\mathsf{X}}(x, B) \left(\int p_{\mathsf{X}|\mathsf{Y}}(x|y)\mathbf{P}_{\mathsf{Y}}(dy) \right) \nu(dx). \tag{38}$$

Finally, using the definition of $\kappa_{\mathsf{Y}|\mathsf{X}}$, and by Equation (35), we see that $\kappa_{\mathsf{Y}|\mathsf{X}}$ is a version of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$. □

Lemma 44. *Let $R \subseteq S$ be a $\mathbf{P}_{\mathsf{X}}$-measure one subset. If the conditional density $p_{\mathsf{X}|\mathsf{Y}}(x|y)$ of X given Y is continuous on $R \times T$ and bounded, then there is a $\mathbf{P}_{\mathsf{X}}$-almost continuous version of the conditional distribution $\mathbf{P}[\mathsf{Y}|\mathsf{X}]$.*

Proof. Fix an open set $B \subseteq T$. We will show that for fixed B, the map $x \mapsto \kappa_{\mathsf{Y}|\mathsf{X}}(x, B)$ given by Bayes' rule is a lower semicontinuous by demonstrating that the numerator is lower semicontinuous, while the denominator is continuous.

Let $\mathbf{P}_{\mathsf{Y}}$ be the distribution of Y. By hypothesis, the map $\phi : S \to \mathcal{C}(T, \mathbb{R}^+)$ given by $\phi(x) = p_{\mathsf{X}|\mathsf{Y}}(x|\cdot)$ is continuous on R, while the indicator function $\mathbf{1}_B$ is lower semicontinuous. Because the integration operator $f \mapsto \int f d\mu$ of a lower semicontinuous function f with respect to a probability measure μ is itself lower semicontinuous, the map $x \mapsto \int \mathbf{1}_B \phi(x) d\mathbf{P}_{\mathsf{Y}}$ is lower semicontinuous on R.

Now let $B = T$ and note that for every $x \in R$, the function $\phi(s)$ is bounded by hypothesis. Integration of a bounded continuous function with respect to a probability measure is a continuous operation, and so the map $x \mapsto \int \phi(x) d\mathbf{P}_{\mathsf{Y}}$ is continuous on R. Therefore, $\kappa_{\mathsf{Y}|\mathsf{X}}$ is $\mathbf{P}_{\mathsf{X}}$-almost continuous. □

Appendix F. Noncomputable everywhere continuous conditional distributions

As we saw in Section 5, discontinuity poses a fundamental obstacle to the computability of conditional probabilities. As such, it is natural to ask whether we can construct a pair of random variables (Z, N) that are computable and admit an *everywhere* continuous version of the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$, yet for which every version is noncomputable. In fact, this is possible using a construction similar to that of (X, N) in Section 6.

In particular, if we think of the construction of the kth bit of X as an iterative process, we see that there are two distinct stages. During the first stage, which occurs so long as $k < h(\mathsf{N})$, the bits of X simply mimic those of the uniform random variable V. Then during the second stage, once $k \geq h(\mathsf{N})$, the bits mimic that of $\frac{1}{2}(\mathsf{C} + \mathsf{U})$.

Our construction of Z will differ in the second stage, where the bits of Z will instead mimic those of a random variable S specially designed to smooth out the rough edges caused by the biased coin C. In particular, S will be absolutely continuous and its density will be infinitely differentiable.

We will now make the construction precise. We begin by defining several random variables from which we will construct S.

Lemma 45. *There is a distribution F on $[0,1]$ with the following properties:*

- F *is computable.*
- F *admits a density p_{F} with respect to Lebesgue measure (on $[0,1]$) which is infinitely differentiable on all of $[0,1]$.*
- $p_{\mathsf{F}}(0) = \frac{2}{3}$ *and* $p_{\mathsf{F}}(1) = \frac{4}{3}$.
- $\frac{d_+^n}{dx^n} p_{\mathsf{F}}(0) = \frac{d_-^n}{dx^n} p_{\mathsf{F}}(1) = 0$, *for all $n \geq 1$ (where $\frac{d_-^n}{dx^n}$ and $\frac{d_+^n}{dx^n}$ are the left and right derivatives respectively).*

(See Figure 2 for one such random variable.) Note that F is almost surely nondyadic and so the r-th bit F_r of F is a computable random variable.

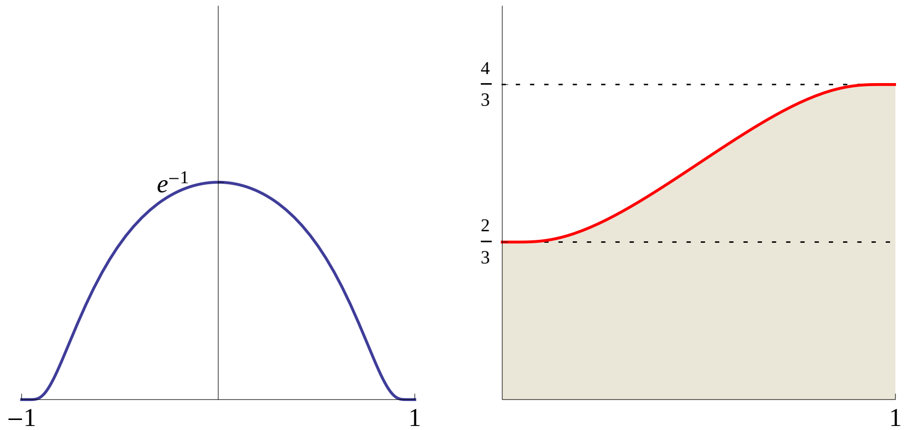

FIGURE 2. (left) $f(x) = e^{-\frac{1}{1-x^2}}$, for $x \in (-1, 1)$, and 0 otherwise, a C^∞ bump function whose derivatives at ± 1 are all 0. (right) A density $p(y) = \frac{2}{3}\Big(\frac{\Phi(2y-1)}{\Phi(1)} + 1\Big)$, for $y \in (0, 1)$, of a random variable satisfying Lemma 45, where $\Phi(y) = \int_{-1}^{y} e^{-\frac{1}{1-x^2}}\, dx$ is the integral of the bump function.

Let $t \in \{0,1\}^3$. For $r \in \mathbb{N}$, define

$$\mathsf{S}^{000}_r := \begin{cases} 0, & r < 3; \\ \mathsf{F}_{r-3}, & r \geq 3; \end{cases}$$

$$\mathsf{S}^{100}_r := \begin{cases} 1, & r = 0; \\ 0, & 1 \leq r < 3; \\ 1 - F_{r-3}, & r \geq 3; \end{cases}$$

$$\mathsf{S}^{t}_r := \begin{cases} C, & r = \emptyset; \\ t(r), & 1 \leq r < 3; \\ \mathsf{U}_{r-3}, & \text{otherwise}; \end{cases}$$

when $t \notin \{000, 100\}$. It is straightforward to show that S^t_r are computable random variables, uniformly in t and r.

Finally, let T be a uniformly random element in $\{0,1\}^3$, and let the r-th bit of S be $\mathsf{S}^{\mathsf{T}}_r$.

It is straightforward to show that

(i) S admits a density p_S with respect to Lebesgue measure on $[0, 1]$.

(ii) p_S is infinitely differentiable everywhere with $\frac{d^n_+}{dx^n} p_\mathsf{S}(0) = \frac{d^n_-}{dx^n} p_\mathsf{S}(1)$, for all $n \geq 0$.

(For a visualization of the density p_S see Figure 3.)

We say a real $x \in [0, 1]$ is **valid for** S if $x \in (\frac{1}{8}, \frac{4}{8}) \cup (\frac{5}{8}, \frac{8}{8})$. (For nondyadic x, this is equivalent to the first 3 bits of the binary expansion of x not being 000 or

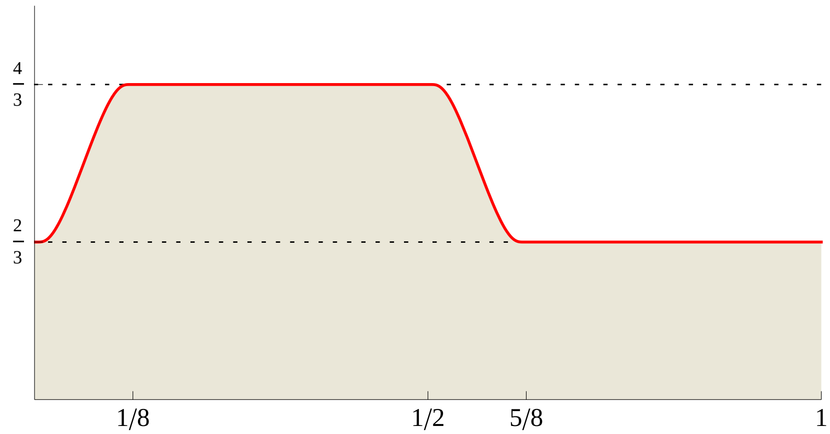

FIGURE 3. Graph of the density function p_{S}.

100.) The following are then straightforward consequences of the construction of S and the definition of valid points:

(iii) If x is valid for S then $p_{\mathsf{S}}(x) \in \{\frac{2}{3}, \frac{4}{3}\}$.

(iv) The Lebesgue measure (and $\mathbf{P}_{\mathsf{S}}$-measure) of the collection of valid x is $\frac{3}{4}$.

Next we define, for every $k \in \mathbb{N}$, the random variables Z_k mimicking the construction of X_k. Specifically, for $k \in \mathbb{N}$, define

$$\mathsf{Z}_k := \frac{\lfloor 2^k \mathsf{V} \rfloor + \mathsf{S}}{2^k}, \tag{39}$$

and let $\mathsf{Z}_\infty := \lim_{k\to\infty} \mathsf{Z}_k = \mathsf{V}$. Then the nth bit of Z_k is

$$(\mathsf{Z}_k)_n = \begin{cases} \mathsf{V}_n, & n < k; \\ \mathsf{S}_{n-k}, & n \geq k \end{cases} \qquad \text{a.s.} \tag{40}$$

For $k < \infty$, we say that $x \in [0, 1]$ is **valid for** Z_k if the fractional part of $2^k x$ is valid for S, and we say that x is **valid for** Z_∞ for all x. Let A_k be the collection of x valid for Z_k. It follows from *(iv)* that the Lebesgue measure of A_k is $\frac{3}{4}$ for all $k < \infty$.

It is straightforward to show from *(i)* and *(ii)* above that Z_k admits a density p_{Z_k} with respect to Lebesgue measure on $[0, 1]$ and that this density is infinitely differentiable.

To complete the construction, we define $\mathsf{Z} := \mathsf{Z}_{h(\mathsf{N})}$. The following results are analogous to those in the almost continuous construction:

Lemma 46. *The random variable* Z *is computable.*

Lemma 47. *There is an everywhere continuous version of* $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$.

Proof. The density p_{Z} is everywhere continuous and positive. □

Lemma 48. *For all $m, n \in \mathbb{N}$, all version κ of the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$ and $\mathbf{P}_{\mathsf{Z}}$-almost all x, if x is valid for $\mathsf{Z}_{h(n)}$ and for $\mathsf{Z}_{h(m)}$ then*

$$2^{m-n} \cdot \frac{\kappa(x, \{m\})}{\kappa(x, \{n\})} \in \begin{cases} \{\frac{1}{2}, 1, 2\}, & h(n), h(m) < \infty; \\ \{1\}, & h(n) = h(m) = \infty; \\ \{\frac{2}{3}, \frac{3}{4}, \frac{4}{3}, \frac{3}{2}\}, & \textit{otherwise.} \end{cases}$$

We now show that one can compute the halting set from any version of the conditional distribution.

Proposition 49. *The conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$ is not computable.*

Proof. Suppose the conditional distribution $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$ were computable. Let n be the index of some TM that does not halt (on input 0), i.e., for which $h(n) = \infty$. Consider any $m \in \mathbb{N}$. Notice that all $x \in [0, 1]$ are valid for $\mathsf{Z}_{h(n)}$ and so $A_{h(n)} \cap A_{h(m)} = A_{h(m)}$.

Let κ be an arbitrary version of $\mathbf{P}[\mathsf{N}|\mathsf{Z}]$, and let R be a $\mathbf{P}_{\mathsf{Z}}$-measure one set on which κ is computable. Then the function

$$\tau(\cdot) := 2^{m-n} \cdot \frac{\kappa(\cdot, \{m\})}{\kappa(\cdot, \{n\})} \tag{41}$$

is also computable on R. Define $T_\infty := \{1\}$, $T_{<\infty} := \{\frac{2}{3}, \frac{4}{3}\}$ and $T := T_\infty \cup T_{<\infty}$.

By equation (41), there is a $\mathbf{P}_{\mathsf{Z}}$-measure one subset $D \subseteq R$ such that whenever $x \in D \cap A_{h(m)}$ then $\tau(x)$ is in T.

For $t \in T$, let B_t be an ideal ball of radius less than $\frac{1}{6}$ about t, and let U_t be a c.e. open set such that $\tau^{-1}(B_t) \cap R = U_t \cap R$. Define $U_\infty := U_1$ and $U_{<\infty} := U_{\frac{2}{3}} \cup U_{\frac{4}{3}}$. Notice these are both c.e. open sets and $D \cap U_\infty \cap U_{<\infty} = \emptyset$.

We now consider two cases. First, assume $h(m) = \infty$. In this case $A_{h(m)} = [0, 1]$ and $A_{h(m)} \cap D \subseteq \tau^{-1}(T_\infty) \cap D = U_\infty \cap D$. Hence

(a) The Lebesgue measure of U_∞ is $1 > \frac{1}{2}$.

If, however, $h(m) < \infty$ then $A_{h(m)}$ has Lebesgue measure $\frac{3}{4}$ and

$$A_{h(m)} \subseteq \tau^{-1}(T_{<\infty}) \cap D = U_{<\infty} \cap D, \tag{42}$$

and so

(b) The Lebesgue measure of $U_{<\infty}$ is at least $\frac{3}{4} > \frac{1}{2}$.

In particular, for each $m \in \mathbb{N}$ exactly one of *(a)* or *(b)* must hold. But it is clear that the collection of m for which *(b)* holds a c.e. set and the collection of m for which *(b)* does not hold (i.e., for which *(a)* holds) is also a c.e. set. So, as *(b)* holds of m if and only if $m \in H = \{m : h(m) < \infty\}$, we have H is a computable set, which we know is a contradiction.

Therefore κ must not be computable. □

In conclusion, we obtain the following strengthening of Theorem 29.

Theorem 50. *Let X and Y be computable real-valued random variables. Then operation $\mathsf{X}, \mathsf{Y} \mapsto \mathbf{P}[\mathsf{X}|\mathsf{Y}]$ of conditioning a pair of real-valued random variables, even when restricted to pairs for which there exists an everywhere continuous version of the conditional distribution, is not computable.*

Department of Mathematics, Harvard University, One Oxford Street, Cambridge, MA 02138

E-mail address: nate@math.harvard.edu

Department of Mathematics, University of Hawai'i at Mānoa, 2565 McCarthy Mall, Honolulu, HI 96815

E-mail address: freer@math.hawaii.edu

Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32 Vassar Street, Cambridge, MA 02139

E-mail address: droy@csail.mit.edu