#P-COMPLETE CONDITIONAL DISTRIBUTIONS

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

Abstract. We study conditional probability from the perspective of complexity theory of functions and operators in analysis, building on work by Ko (1983), Friedman (1984), and Kawamura and Cook (2010). For some random variable X in $\{0,1\}^{\mathbb{N}}$ whose distribution is continuous and polynomial-time computable, and some polynomial-time computable function $f : \{0,1\}^{\mathbb{N}} \to [0,1]$ for which the random variable $f(\mathsf{X})$ is "polynomially-diffuse", the function taking (integers encoding) $A \in \{0,1\}^*$, an open rational interval B, and an accuracy 2^{-i} to a rational within 2^{-i} of the conditional probability that $\mathsf{X} \in A$ given $f(\mathsf{X}) \in B$ is shown to be #P-complete. On the other hand, all such functions computing conditional probabilities are in #P.

1. Introduction

Recent work [1] by the authors (to be presented at LICS 2011) studies conditional probability from the perspective of *computability* in analysis using notions from the Type-2 Theory of Effectivity (TTE). In short, computable joint distributions of continuous random variables may have noncomputable conditional distributions. That work shows that the operation of conditioning is computationally difficult. In this extended abstract, we present aspects of preliminary work on complexity-theoretic analogues of these questions, and show that even in a restricted setting where the conditional distribution is computable, it can be highly nonefficient.

The study of the computational complexity and computability of conditional probability not only reveals fundamental limitations on important operations in probability theory and Bayesian statistics, but is increasingly relevant to probabilistic modeling within science and engineering. In particular, proposals within machine learning and artificial intelligence to use *probabilistic programs* as effective representations for complex joint distributions and probabilistic inference motivate the investigation of the class of computable distributions (see [1] for references).

1.1. **Related work.** In the elementary setting, the conditional probability $\mu(A|B)$ of an event A given an event B is given by the ratio $\mu(A \cap B)/\mu(B)$, provided $\mu(B) > 0$. Because we have $\mu(B) = \int 1_B d\mu$, one should expect that the complexity of conditional probability should relate to that of integration. In particular, if we assume that μ admits a density p with respect to Lebesgue measure, then $\mu(B) = \int_B p(x)dx$ and results by Friedman [2] and Ko [3] (and as reframed in [4, §3]) on #P-computable integrals are relevant, assuming that p is efficiently computable. Of course, a computable distribution need not admit a density and even so, need not admit a computable (nevermind efficiently computable) density [1].

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 [5]. If f is a one-way function, then it is difficult to compute the conditional distribution of the uniform distribution on strings of some length with respect to a given output of f. This intuition is made precise by Ben-David, Chor, Goldreich, and Luby [6] in their theory of polynomial-time computable distributions, which has since been extended by Yamakami [7], Aaronson [8], and others.

However, such work is concerned with families of increasingly long bitstrings, which need not cohere in any uniform way. Our approach is directly motivated by the problem of conditioning computable joint distributions in order to form conditional distributions. In this extended abstract, we build on notions of computational complexity for continuous real functions in order to explore the polynomial-time computability of conditional distributions using essentially the standard notions for complexity in analysis via TTE [9].

2. Computational complexity of distributions

Let $\mathcal{I}$ be the collection of open intervals in $[0,1]$ of the form $(\frac{i}{2^n}, \frac{j}{2^n})$ for $n > 0$ and $0 \leq i < j < 2^n$, and consider a standard encoding of an element of $\mathcal{I}$ as the string $\langle i, j, 0^n \rangle$. For our purposes here, we are already able to find a #P-complete problem even restricting to open intervals in $\mathcal{I}$.

The following notions could also be formulated in terms of type-two machines with infinite strings, though here we express the necessary notions explicitly in terms of rapidly-converging approximations.

Definition 1. We say that a **witness** to a probability distribution μ on $[0,1]$ is a map $M_\mu : \mathcal{I} \times \mathbb{N} \to \mathbb{Q}$ satisfying

$$\left|\mu(I) - M_\mu(I,i)\right| < 2^{-i}. \tag{1}$$

We say that a probability distribution is **polynomial-time computable** when it has a witness that is computable in time polynomial in the encoding of I and in i itself (not the length of i in binary). (We define notions similarly for probability distributions on $\{0,1\}^{\mathbb{N}}$.)

Note that when μ is discrete (with a finite or countable number of atoms), this notion is similar to that of a P-computable distribution in Yamakami [7].

Definition 2. Let X be a random variable on $\{0,1\}^{\mathbb{N}}$. and let $f : \{0,1\}^{\mathbb{N}} \to [0,1]$ be a continuous function. A **witness** to the conditional distribution $\mathbf{P}[\mathsf{X} \mid f(\mathsf{X})]$ is a function $w : \{0,1\}^* \times \mathcal{I} \times \mathbb{N} \to \mathbb{Q}$ such that

$$\left|\mathbf{P}(\mathsf{X} \in I \mid f(\mathsf{X}) \in J) \; - \; w(I,J,i)\right| < 2^{-i}. \tag{2}$$

We say that a conditional distribution is **polynomial-time computable** when it has a witness that is computable in time polynomial in the length of I, the encoding of J, and in i itself.

Note that functions $(\{0,1\}^*)^m \times \mathcal{I}^k \times \mathbb{N}^\ell \to \mathbb{Q}$, such as a (conditional) distribution witness, or functions $\mathbb{N} \to \mathbb{N}$, such as #SAT, can be coded on an oracle tape in a standard way (e.g., via the graph of the function, with a standard encoding of $\mathbb{Q}$, spread out by a delimiter symbol).

Definition 3. Let Y be a random variable in $[0,1]$. We say that Y is **polynomially-diffuse** when there is some polynomial r such that for every dyadic open interval $J \in \mathcal{I}$, we have

$$\frac{\mathbf{P}\{\mathsf{Y} \in J\}}{|J|} < r\big(\mathrm{size}(J)\big),$$

where $\mathrm{size}(J)$ is the length of the standard encoding of J, and $|J|$ is length of the interval.

Note that if Y is absolutely continuous with bounded density, then it is polynomially-diffuse.

3. #P-completeness of conditional distributions

We now state the main results. Recall the standard notion [9] of a polynomial-time computable function $\{0,1\}^{\mathbb{N}} \to [0,1]$.

Theorem 4 (Conditioning is in #P)**.** *For every random variable X on $\{0,1\}^{\mathbb{N}}$ with a polynomial-time computable distribution and every polynomial-time computable function $f : \{0,1\}^{\mathbb{N}} \to [0,1]$ for which $f(\mathsf{X})$ is a polynomially-diffuse continuous random variable, the conditional distribution $\mathbf{P}[\mathsf{X} \mid f(\mathsf{X})]$ has a witness that is polynomial-time computable with respect to a #P-complete oracle.*

We now sketch the proof informally. Consider $I \in \{0,1\}^*$ and an interval $J \in \mathcal{I}$, and let f and X be as in the theorem. The proof begins by subdividing the domain $\{0,1\}^{\mathbb{N}}$ into many small blocks (some of which are in I, and others of which are not) such that the images of most blocks (according to where $f(\mathsf{X})$ assigns mass) under f are either completely contained in J or disjoint from it. By the particular choice of blocks, and the fact that $f(\mathsf{X})$ is polynomially-diffuse, the property of whether or not the image of each block lands in the desired range J is shown to be polynomial-time computable. Using the ability to count provided by our #P-complete oracle, along with the polynomial-time computability of X, we can count the appropriately-scaled measure of such blocks, and from this compute the desired conditional distribution.

Theorem 5 (Conditioning is #P-hard)**.** *There are a random variable X on $\{0,1\}^{\mathbb{N}}$ with a polynomial-time computable distribution and a polynomial-time computable function $f : \{0,1\}^{\mathbb{N}} \to [0,1]$ for which $f(\mathsf{X})$ is a polynomially-diffuse continuous random variable, such that given a witness to the conditional distribution $\mathbf{P}[\mathsf{X} \mid f(\mathsf{X})]$ as an oracle, we can compute any #P function via a classical polynomial-time Turing reduction.*

The proof of Theorem 5 is more involved. We build a random variable X on $\{0,1\}^{\mathbb{N}}$ whose distribution is polynomial-time computable, a polynomial-time computable

function f for which $f(\mathsf{X})$ is polynomially-diffuse, a set $I \in \{0,1\}^*$, and a nested series of intervals $(J_k)_{k\in\mathbb{N}}$, such that given a #P function to reduce, we can determine the answer to the counting problem for any particular input by querying our conditional distribution on I and some J_k (where k is selected in polynomial-time), followed by a polynomial-time computation.

Acknowledgments

The authors would like to thank Vikash Mansinghka, André Nies, and Michael Sipser for helpful discussions. C.E.F. is partially supported by the National Science Foundation under Grant No. 0901020. D.M.R. is supported by a Newton International Fellowship.

References

[1] N. L. Ackerman, C. E. Freer, and D. M. Roy, "Noncomputable conditional distributions," in *Proc. 26th Ann. IEEE Symp. on Logic in Comp. Sci.* IEEE Computer Society, 2011.

[2] H. Friedman, "The computational complexity of maximization and integration," *Adv. in Math.*, vol. 53, no. 1, pp. 80–98, 1984. http://dx.doi.org/10.1016/0001-8708(84)90019-7

[3] K.-I. Ko, "On the computational complexity of ordinary differential equations," *Inform. and Control*, vol. 58, no. 1-3, pp. 157–194, 1983. http://dx.doi.org/10.1016/S0019-9958(83)80062-X

[4] A. Kawamura, "Lipschitz continuous ordinary differential equations are polynomial-space complete," *Comput. Complexity*, vol. 19, no. 2, pp. 305–332, 2010. http://dx.doi.org/10.1007/s00037-010-0286-0

[5] 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

[6] 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

[7] 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

[8] S. Aaronson, "The Equivalence of Sampling and Searching," *Proc. of Int. Comput. Sci. Symp. in Russia (CSR 2011)*, 2011. http://arxiv.org/abs/1009.5104

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

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

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

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

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

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

E-mail address: d.roy@eng.cam.ac.uk