Equivalences between learning of data and probability distributions, and their applications *

George Barmpalias Nan Fang Frank Stephan

August 2, 2019

Abstract. Algorithmic learning theory traditionally studies the learnability of effective infinite binary sequences (reals), while recent work by (Vitányi and Chater, 2017) and (Bienvenu et al., 2014) has adapted this framework to the study of learnability of effective probability distributions from random data. We prove that for certain families of probability measures that are parametrized by reals, learnability of a subclass of probability measures is equivalent to learnability of the class of the corresponding real parameters. This equivalence allows to transfer results from classical algorithmic theory to learning theory of probability measures. We present a number of such applications, providing many new results regarding EX and BC learnability of classes of measures, thus drawing parallels between the two learning theories.

George Barmpalias

State Key Lab of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China.
E-mail: barmpalias@gmail.com. *Web:* http://barmpalias.net

Nan Fang

Institut für Informatik, Ruprecht-Karls-Universität Heidelberg, Germany.
E-mail: nan.fang@informatik.uni-heidelberg.de. *Web:* http://fangnan.org

Frank Stephan

Department of Mathematics and School of Computing, National University of Singapore, Republic of Singapore.
E-mail: fstephan@comp.nus.edu.sg. *Web:* http://www.comp.nus.edu.sg/~fstephan

*Barmpalias was supported by the 1000 Talents Program for Young Scholars from the Chinese Government No. D1101130, NSFC grant No. 11750110425 and grant No. ISCAS-2015-07 from the Institute of Software. Frank Stephan is supported in part by the Singapore Ministry of Education Academic Research Fund MOE2016-T2-1-019 / R146-000-234-112. The work was partially done while the authors were visiting (and being supported by) the Institute for Mathematical Sciences, National University of Singapore in 2017.

1 Introduction

The present work concerns the following informally stated general problem, which we study in the context of formal language identification and algorithmic learning theory:

> Given a probability distribution $\mathcal{P}$ and a sufficiently large sample of randomly chosen data from the given distribution, learn or estimate a probability distribution with respect to which the sample has been randomly sampled. (1)

Problem (1) has a long history in statistics (e.g. see (Vapnik, 1982)) and has more recently been approached in the context of computational learning, in particular the probably approximately correct (PAC) learning model, starting with (Kearns et al., 1994). The same problem was recently approached in the context of Algorithmic Learning Theory, in the tradition of (Gold, 1967), and Kolmogorov complexity by (Vitányi and Chater, 2017).[1]

The learning concepts discussed in (Vitányi and Chater, 2017) are very similar in nature to the classic concepts of algorithmic learning which are motivated by the problem of language learning in the limit (see (Pinker, 1979)) but they differ in two major ways. In the classic setting, one starts with a class of languages or functions which have a finite description (e.g. they are computable) and the problem is to find an algorithm (often called *a learner*) which can infer, given a sufficiently long text from any language in the given class, or a sufficiently long segment of the characteristic sequence of any function in the given class, a description of the language or function in the form of a grammar or a program. More precisely, the desired algorithm makes successive predictions given longer and longer segments of the input sequence, and is required to converge to a correct grammar or program for the given infinite input.

If we apply the concept of identification in the limit to Problem (1), according to (Vitányi and Chater, 2017), one starts with a class $\mathcal{V}$ of finitely describable probability distributions (say, the computable measures on the Cantor space) and we have the following differences with respect to the classic setting:

- the inputs on which the learner is supposed to succeed in the limit are random sequences with respect to some probability distribution in the given class $\mathcal{V}$, and not elements of $\mathcal{V}$;
- success of the learner $\mathcal{L}$ on input X means that $\mathcal{L}(X \upharpoonright_n)$ converges, as $n \to \infty$, to a description of some element of $\mathcal{V}$ with respect to which X is random.

First, note that just as in the context of computational learning theory, here too we need to restrict the probability distributions in Problem (1) to a class of 'feasible' distributions, which in our case means computable distributions in the Cantor space. Second, in order to specify the learning concept we have described, we need to define what we mean by random inputs X with respect to a computable distribution $\mathcal{P}$ in the given class $\mathcal{V}$ on which the learner is asked to succeed. (Vitányi and Chater, 2017) ask the learner to succeed on every real X which is *algorithmically random*, in the sense of Martin-Löf (Martin-Löf, 1966), with respect

[1] Probabilistic methods and learning concepts in formal language and algorithmic learning theory have been studied long before (Vitányi and Chater, 2017), see (Pitt, 1989) and the survey (Ambainis, 2001). However most of this work focuses on identifying classes of languages or functions using probabilistic strategies, rather than identifying probability distributions as Problem (1) asks. Bienvenu and Monin (Bienvenu and Monin, 2012, Section IV) do study a form of (1) through a concept that they call *layerwise learnability* of probability measures in the Cantor space, but this is considerably different than (Vitányi and Chater, 2017) and the concepts of Gold (Gold, 1967), the most important difference being that it refers to classes of probability measures that are not necessarily contained in the computable probability measures.

to some computable probability measure.[2] Then the interpretation of Problem (1) through the lenses of algorithmic learning theory and in particular, the ideas of (Vitányi and Chater, 2017) is as follows:

> Given a computable measure μ and an algorithmically random stream X with respect to μ, learn in the limit (by reading the initial segments of X) a computable measure μ' with respect to which X is algorithmically random. (2)

This formulation invites many different formalizations of learning concepts which are parallel to the classic theory of algorithmic learning[3], and although we will comment on some of them later on, this article is specifically concerned with EX-learning (*explanatory learning*, one of the main concepts in Gold (Gold, 1967)), which means that in (2) we require the learner to eventually converge to a specific description of the computable measure[4] with the required properties.

A *learner* is simply a function $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$. We refer to infinite binary streams (sequences) as reals. According to (Gold, 1967), a class C of computable elements of 2^ω is EX-learnable if there exists a learner $\mathcal{L}$ such that for each $Z \in C$ we have that $\lim_n \mathcal{L}(Z \upharpoonright_n)$ exists and equals an index of Z as a computable function.[5] Similarly, C is BC-learnable if there exists a learner $\mathcal{L}$ such that for each $Z \in C$ there exists some n_0 such that for all $n > n_0$ the value of $\mathcal{L}(Z \upharpoonright_n)$ is an index of Z.

In this paper we study explanatory (EX) learning, behaviorally correct (BC) learning and partial learning of probability measures, based on the classic notion of algorithmic randomness by (Martin-Löf, 1966). Given a measure μ on the reals and a real X, we say that X is μ-random if it is algorithmically random with respect to μ. We review algorithmic randomness with respect to arbitrary measures in Section 2.3.

Definition 1.1 (EX learning of measures)**.** A class C of computable measures is EX-learnable if there exists a computable learner $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$ such that for every $\mu \in C$ and every μ-random real X the limit $\lim_n \mathcal{L}(X \upharpoonright_n)$ exists and equals an index of a measure $\mu' \in C$ such that X is μ'-random.

(Vitányi and Chater, 2017) introduced this notion and observed that any uniformly computable family of measures is EX-learnable. On the other hand, (Bienvenu et al., 2014) showed that the class of computable measures is not EX-learnable, and also not even BC-learnable in the following sense.

Definition 1.2 (BC learning of measures)**.** A class C of computable measures is BC-learnable if there exists a computable learner $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$ such that for every $\mu \in C$ and every μ-random real X there exists n_0 and $\mu' \in C$ such that for all $n > n_0$ the value $\mathcal{L}(X \upharpoonright_n)$ is an index of μ' such that X is μ'-random.

One could consider a stronger learnability condition, namely that given $\mu \in C$ and any μ-random X the learner identifies μ in the limit, when reading initial segments of X. Note that such a property would only be realizable in classes C where any $\mu, \mu' \in C$ are *effectively orthogonal*, which means that the classes of μ-random and μ'-random reals are disjoint. In this case we call C *effectively orthogonal*, and Definitions 1.1, 1.2 are equivalent with the versions where μ' is replaced by μ. On the other hand we could consider a weakened notion of learning of a class C of computable measures, where given $\mu \in C$ and any μ-random X,

[2] From this point on we will use the term (probability) *measure* instead of distribution, since the literature in algorithmic randomness that we are going to use is mostly written in this terminology.

[3] EX-learning, BC-learning, BC*-learning etc. In (Odifreddi, 1999, Chapter VII.5) the reader can find a concise and accessible introduction to these basic learning concepts and results.

[4] as opposed to, for example, eventually giving different indices of the same measure, or even different measures all of which satisfy the required properties.

[5] Here we identify subsets of natural numbers with their characteristic function and assume a fixed Gödel numbering (φ_e) of all partial computable functions with binary values (also called *reals*), which gives an 'index' to member of this class.

the learner identifies *some* computable measure μ (possibly not in C) in the limit, with respect to which X is random, when reading initial segments of X.

Definition 1.3 (Weak EX learning of measures)**.** A class C of computable measures is weakly EX-learnable if there exists a computable learner $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$ such that for every $\mu \in C$ and every μ-random real X the limit $\lim_n \mathcal{L}(X \upharpoonright_n)$ exists and equals an index of a computable measure μ' such that X is μ'-random.

Definition 1.4 (Weak BC learning of measures)**.** A class C of computable measures is weakly BC-learnable if there exists a computable learner $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$ such that for every $\mu \in C$ and every μ-random real X there exists n_0 and a computable measure μ', such that for all $n > n_0$ the value $\mathcal{L}(X \upharpoonright_n)$ is an index of μ' such that X is μ'-random.

We note that the notions in Definitions 1.1 and 1.2 are not closed under subsets. In the following proof and the rest of this article, we use '$*$' to denote concatenation of strings.

Proposition 1.5. *There exist classes $C \subsetneq \mathcal{D}$ of measures such that $\mathcal{D}$ is EX-learnable and C is not even BC-learnable.*

Proof. Let (σ_i) be a prefix-free sequence of strings, let μ_i be the measure with $\mu_i(\sigma_{2i} * 0^\omega) = \mu_i(\sigma_{2i+1} * 0^\omega) = 1/2$ and let ν_i be the measure such that $\nu_i(\sigma_i * 0^\omega) = 1$. Define $C = \{\mu_i, \nu_{2j} \mid i \in \emptyset''' \ \wedge\ j \in \mathbb{N} - \emptyset'''\}$ and $\mathcal{D} = \{\mu_i, \nu_i \mid i \in \mathbb{N}\}$. Clearly $C \subsetneq \mathcal{D}$. If C was BC-learnable then $\emptyset'''$ could be decided in $\emptyset''$: to decide if $n \in \emptyset'''$ we just need to check the limit guess of the learner on σ_{2n}. This is a contradiction. On the other hand the learner which guesses ν_i on each extension of σ_i is an EX-learner for $\mathcal{D}$. □

On the other hand, the weaker notions of Definitions 1.3 and 1.4 clearly are closed under subsets. In Section 1.1 we also consider an analogue of the notion of partial learning from (Osherson et al., 1986) for measures, and prove an analogue of the classic result from the same book that the computable reals are partially learnable.

1.1 Our main results

The aim of this paper is to establish a connection between the above notions of learnability of probability measures, with the corresponding classical notions of learnability of reals in the sense of (Gold, 1967). To this end, we prove the following equivalence theorem, which allows to transfer positive and negative learnability results from reals to probability measures that are parametrized by reals, and vice-versa. Let $\mathcal{M}$ denote the Borel measures on 2^ω.[6]

Theorem 1.6 (The first equivalence theorem)**.** *Given a computable $f : 2^\omega \to \mathcal{M}$ let $\mathcal{D} \subseteq 2^\omega$ be an effectively closed set such that for any $X \neq Y$ in $\mathcal{D}$ the measures $f(X), f(Y)$ are effectively orthogonal. If $\mathcal{D}^* \subseteq \mathcal{D}$ is a class of computable reals, $\mathcal{D}^*$ is EX-learnable if and only if $f(\mathcal{D}^*)$ is EX-learnable. The same is true of the BC learnability of $\mathcal{D}^*$.*

As a useful and typical example of a parametrization f of measures by reals as stated in Theorem 1.6, consider the function that maps each real $X \in 2^\omega$ to the Bernoulli measure with success probability the real in the unit interval $[0, 1]$ with binary expansion X. Note that the Bernoulli measures[7] are an effectively

[6] Formal background definitions regarding the metric space of Borel measures are given in Section 2.

[7] By Bernoulli measure we mean the product measure on the space of infinite binary strings, of the 'biased coin' measure on $\{0, 1\}$ that gives probability $q \in [0, 1]$ on 0 and probability $1 - q$ on 1.

orthogonal class (e.g. consider the law of large numbers regarding the frequency of 0s in the limit). The proof of Theorem 1.6 is given in Section 3. We note that It is possible to relax the hypothesis of the 'if' direction of Theorem 1.6 for the case of EX-learning – we give this extension in Section 3.4.

The next equivalence theorem concerns weak learnability.[8]

Theorem 1.7 (The second equivalence theorem)**.** *There exists a map $Z \to \mu_Z$ from 2^ω to the continuous Borel measures on 2^ω, such that for every class C of computable reals, C is EX/BC learnable if and only if $\{\mu_Z \mid Z \in C\}$ is a weakly EX/BC learnable class of computable measures, respectively.*

Finally we give a positive result in terms of partial learning. Let (μ_e) be a uniform enumeration of all partial computable measures (see Section 2.1). We say that a learner $\mathcal{L}$ *partially succeeds* on a computable measure μ if for all μ-random X there exists a j_0 such that (a) there are infinitely many n with $\mathcal{L}(X \upharpoonright_n) = j_0$; (b) if $j \neq j_0$ then there are only finitely many n with $\mathcal{L}(X \upharpoonright_n) = j$; (c) μ_{j_0} is a computable measure such that X is μ_{j_0}-random.

Theorem 1.8. *There exists a computable learner which partially succeeds on all computable measures.*

Theorems 1.6 and 1.7 allow the transfer of learnability results from the classical theory on the reals to probability measures. Detailed background on the notions that are used in our results and their proofs is given in Section 2.

1.2 Applications of our main results

The equivalences in Theorems 1.6 and 1.7 have some interesting applications, some of which are stated below, deferring their proofs to Section 4.

(Adleman and Blum, 1991) showed that an oracle can EX-learn all computable reals if and only if it is *high*, i.e. it computes a function that dominates all computable functions. Using Theorem 1.7 we may obtain the following analogue for measures.

Corollary 1.9. *The computable (continuous) measures are (weakly) EX-learnable with oracle A if and only if A is high.*

We may write EX[A] to indicate that the EX-learner is computable in A. A class C of measures is (weakly) EX$*$[A]-learnable for an oracle A, if there exists an EX-learner $\mathcal{L} \leq_T A$ for C such that for each X, the function $n \to \mathcal{L}(X \upharpoonright_n)$ uses finitely many queries to A. The following is an analogue of a result from (Kummer and Stephan, 1996) about EX$*$[A] learning of reals.

Corollary 1.10. *The class of computable measures is EX$*$[A]-learnable if and only if $\emptyset'' \leq_T A \oplus \emptyset'$.*

If we apply Theorem 1.6 we obtain an analogue of the (Adleman and Blum, 1991) characterization with respect to Bernoulli measures.

Corollary 1.11. *An oracle can EX-learn all computable Bernoulli measures if and only if it is high.*

[8]For the special case where we allow measures with atoms in our classes, Theorem 1.7 has a somewhat easier proof than the one given in Section 3.5.

(Blum and Blum, 1975) showed the so-called non-union theorem for EX-learning, namely that EX-learnability of classes of computable reals is not closed under union. We may apply our equivalence theorem in order to prove an analogue for measures.

Corollary 1.12 (Non-union for measures)**.** *There are two EX-learnable classes of computable (Bernoulli) measures such that their union is not EX-learnable.*

One can find applications of Theorem 1.6 on various more complex results in algorithmic learning theory. As an example, we mention the characterization of low oracles for EX-learning that was obtained in (Gasarch and Pleszkoch, 1989; Slaman and Solovay, 1991) (also see (Fortnow et al., 1994)). An oracle A is low for EX-learning of classes of computable measures, if any class of computable measures that is learnable with oracle A, is learnable without any oracle. The characterization mentioned above is that, an oracle is low for EX-learning if and only if it is 1-generic and computable from the halting problem. This argument consisted of three steps, first showing that 1-generic oracles computable from the halting problem are low for EX-learning, then that oracles that are not computable from the halting problem are not low for EX-learning, and finally that oracles that are computable from the halting problem but are not 1-generic are not low for EX-learning. The last two results can be combined with Theorem 1.6 in order to show one direction of the characterization for measures:

> if an oracle A is either not computable from the halting problem or not 1-generic, then there exists a class of computable (Bernoulli) measures which is not EX-learnable but which is EX-learnable with oracle A. (3)

In other words, low for EX-learning oracles for measures are 1-generic and computable from the halting problem.

Corollary 1.13. *If an oracle is low for EX-learning for measures, then it is also low for EX-learning for reals.*

We do not know if the converse of Corollary 1.13 holds.

1.3 Notions of learnability of probability measures

(Bienvenu et al., 2014) say that a learner $\mathcal{L}$ EX-succeeds on a real X if $\lim_n \mathcal{L}(X \upharpoonright_n)$ equals an index of a computable measure with respect to which X is random. Similarly, $\mathcal{L}$ BC-succeeds on X if there exists a measure μ such that X is μ-random, and for all sufficiently large n, the value of $\mathcal{L}(X \upharpoonright_n)$ is an index of μ. The results in (Bienvenu et al., 2014, 2018) are of the form 'there exists (or not) a learner which succeeds on all reals that are random with respect to a computable measure'. Hence (Bienvenu et al., 2014, 2018) refer to the weak learnability of Definitions 1.3 and 1.4.

(Bienvenu and Monin, 2012) introduced and studied *layerwise learnability*, in relation to uniform randomness extraction from biased coins. This notion is quite different from learnability in the sense of algorithmic learning theory, but it relates to the 'only if' direction of Theorem 1.6. Let $\mathcal{M}$ denote the class of Borel measures on 2^ω.[9] A class $C \subseteq \mathcal{M}$ of measures (not necessarily computable) is layerwise learnable if there is a computable function $F : 2^\omega \times \mathbb{N} \to \mathcal{M}$ which, given any $\mu \in C$ and any μ-random real X, if the μ-randomness deficiency of X is less than c then $F(X, c) = \mu$. In other words, this notion of learnability of a class $C \subseteq \mathcal{M}$

[9]Refer to Section 2 for background on the notions used in this discussion.

requires to be able to compute (as an infinite object) any measure $\mu \in C$ from any μ-random real and a guarantee on the level of μ-randomness of the real. Hence the main difference with the notions in Definitions 1.1 and 1.2 is that (a) we also learn incomputable measures and (b) learning does not identify a finite program describing the measure, but it computes a measure as an oracle Turing machine infinite computation with oracle the random real. As a concrete example of the difference between the two notions, consider the class of the computable Bernoulli measures which is layerwise learnable (Bienvenu and Monin, 2012) but is not (weakly) EX-learnable or even (weakly) BC-learnable by (Bienvenu et al., 2014).

2 Background

We briefly review the background on the Cantor space 2^ω and the space of Borel measures that is directly relevant for understanding our results and proofs. We focus on effectivity properties of these concepts and the notion of algorithmic randomness. This is textbook material in computable analysis and algorithmic randomness, and we have chosen a small number of references where the reader can obtain more detailed presentations that are similar in the way we use the notions here.

2.1 Representations of Borel measures on the Cantor space

We view 2^ω and the space $\mathcal{M}$ of Borel measures on 2^ω as computable metric spaces.[10] The distance between two reals is 2^{-n} where n is the first digit where they differ, and the basic open sets are $[\![\sigma]\!] =: \{X \in 2^\omega \mid \sigma \preceq X\}$, $\sigma \in 2^{<\omega}$, where $\preceq$ denotes the prefix relation. If $V \subseteq 2^{<\omega}$ then $[\![V]\!] := \cup_{\sigma \in V}[\![\sigma]\!]$. The distance between $\mu, \nu \in \mathcal{M}$ is given by

$$d(\mu, \nu) = \sum_n 2^{-n} \cdot \left(\max_{\sigma \in 2^n} |\mu(\sigma) - \nu(\sigma)| \right)$$

The basic open sets of $\mathcal{M}$ are the balls of the form

$$[(\sigma_0, I_0), \ldots, (\sigma_n, I_n)] = \{\mu \in \mathcal{M} \mid \forall i \leq n,\ \mu(\sigma_i) \in I_i\}$$

where σ_i are binary strings (which we identify with the open balls $[\![\sigma]\!]$ of 2^ω) and I_i are the *basic open intervals in* $[0, 1]$.[11] Define the size of a basic open set C of $\mathcal{M}$ by

$$\left|C\right| = \sup\{d(\mu, \nu) \mid \mu, \nu \in C\}, \quad \text{for } C \in \mathcal{M}^*$$

and note that this is a computable function. By the Caratheodory theorem, each $\mu \in \mathcal{M}$ is uniquely determined by its values on the basic open sets of 2^ω, namely the values $\mu(\sigma) := \mu([\![\sigma]\!])$, $\sigma \in 2^{<\omega}$. Also, each $\mu \in \mathcal{M}$ is uniquely determined by the basic open sets that contain it, and the same is true for 2^ω. A subset of $\mathcal{M}$ is effectively open if it is the union of a computably enumerable set of basic open sets.

We represent measures in $\mathcal{M}$ as the functions $\mu : 2^{<\omega} \to [0, 1]$ such that $\mu(\emptyset) = 1$ (here $\emptyset$ is the empty string) and $\mu(\sigma) = \mu(\sigma * 0) + \mu(\sigma * 1)$ for each $\sigma \in 2^{<\omega}$. We often identify a measure with its representation. A measure μ is computable if its representation is computable as a real function. There are two

[10] All of the notions and facts discussed in this section are standard in computable analysis and are presented in more detail in (Bienvenu and Monin, 2012; Bienvenu et al., 2018). More general related facts, such that the fact that for any computable metric space C the set of probability measures over C is itself a computable metric space, can be found in (Gács, 2005).

[11] These are the intervals $(q, p), [0, q), (p, 1]$ for all dyadic rationals $p, q \in (0, 1)$.

equivalent ways to define what an index (or description) of a computable measure is. One is to define it as a computable approximation to it with uniform modulus of convergence. For example, we could say that a *partial computable measure* is a c.e. (abbreviation for 'computably enumerable') set W of basic open sets (σ, I) of $\mathcal{M}$, where $\sigma \in 2^{<\omega}$, I is a basic open interval of $[0,1]$;[12] In this case we can have a uniform enumeration (μ_e) of all partial computable measures, which could contain non-convergent approximations. Then μ_e, represented by the c.e. set W_e, is total and equal to some measure μ if $\mu \in [(\sigma, I)]$ for all σ, I with $(\sigma, I) \in W_e$, and for each σ we have $\inf\{|I| \mid (\sigma, I) \in W_e\} = 0$. Alternatively, one could consider the fact that for every computable measure μ there exists a computable measure ν which takes dyadic values on each string σ, and such that $\mu = \Theta(\nu)$ (i.e. the two measures are the same up to a multiplicative constant – see (Juedes and Lutz, 1995)). Moreover, from μ one can effectively define ν, and the property $\mu = \Theta(\nu)$ implies that the μ-random reals are the same as the ν-random reals. This means that we may restrict our considerations to the computable measures with dyadic rational values on every string, without loss of generality. Then we can simply let (μ_e) be an effective list of all partial computable functions from $2^{<\omega}$ to the dyadic rationals such that $\mu(\sigma) = \mu(\sigma * 0) + \mu(\sigma * 1)$ for each σ such that the values $\mu(\sigma), \mu(\sigma * 0), \mu(\sigma * 1)$ are defined.

The two formulations are effectively equivalent, in the sense that from one we can effectively obtain the other, so we do not explicitly distinguish them. In any case, an index of a computable measure μ is a number e such that μ_e is total and equals μ. An important exception to this equivalence is when we consider subclasses of computable measures, such as the computable Bernoulli measures which feature in Section 4. In this case we have to use the first definition of (μ_e) above, since it is no longer true that every computable Bernoulli measure can be replaced with a computable Bernoulli measure with dyadic values which has the same random reals.

2.2 Computable functions and metric spaces

There is a well-established notion of a computable function f between computable metric spaces from computable analysis, e.g. see (Bienvenu and Monin, 2012; Weihrauch, 1993). The essence of this notion is effective continuity, i.e. that for each x and a prescribed error bound ϵ for an approximation to $f(x)$, one can compute a neighborhood radius around x such that all of the y in the neighborhood are mapped within distance ϵ from $f(x)$. Here we only need the notion of a computable function $f : 2^\omega \to \mathcal{M}$, which can be seen to be equivalent to the following (due to the compactness of 2^ω). Let $\mathcal{M}^*$ denote the collection of the basic open sets of $\mathcal{M}$.

Definition 2.1. A function $f : 2^\omega \to \mathcal{M}$ is computable if there exists a computable function $f^* : 2^{<\omega} \to \mathcal{M}^*$ which is monotone in the sense that $\sigma \preceq \tau$ implies $f^*(\sigma) \subseteq f^*(\tau)$, and such that for all $Z \in 2^\omega$ we have $f(Z) \in f^*(Z \upharpoonright_n)$ for all n, and $\lim_s |f^*(Z \upharpoonright_s)| = 0$.

More generally, a computable metric space is a tuple $(\mathcal{X}, d_x, (q_i))$ such that $(\mathcal{X}, d_x)$ is a complete separable metric space, (q_i) is a countable dense subset of $\mathcal{X}$ and the function $(i, j) \mapsto d_x(q_i, q_j)$ is computable. A function $f : \mathcal{X} \to \mathcal{Y}$ between two computable metric spaces $(\mathcal{X}, d_x, (q_i^x))$, $(\mathcal{Y}, d_y, (q_i^y))$ is computable if there exists computable function g such that for every $n \in \mathbb{N}$ and every $w, z \in \mathcal{X}$ such that $d_x(w, z) < 2^{-g(n)}$ we have $d_y(f(w), f(z)) < 2^{-n}$; equivalently, if for all $n, i, j \in \mathbb{N}$, such that $d_x(q_i^x, q_j^x) < 2^{-g(n)}$ we have $d_y(f(q_i^x), f(q_j^x)) < 2^{-n}$. In this way, as it is illustrated in Definition 2.1, computable functions between

[12] If one wishes to ensure that in case of convergence the property $\mu(\sigma) = \mu(\sigma * 0) + \mu(\sigma * 1)$ holds, we could also require that if $(\sigma, I), (\sigma * 0, J_0), (\sigma * 1, J_1) \in W$ then $I \cap [\inf J_0 + \inf J_1, \sup J_0 + \sup J_1] \neq \emptyset$.

$2^\omega, \mathbb{N}, \mathcal{M}$ and their products can be thought of as induced by monotone computable functions between the corresponding classes of basic open sets, such that the sizes of the images decrease uniformly as a function of the size of the arguments.

2.3 Algorithmic randomness with respect to arbitrary measures

There is a robust notion of algorithmic randomness with respect to an arbitrary measure μ on 2^ω, which was manifested in approaches by (Levin, 1976, 1984) and (Gács, 2005) in terms of uniform tests, and in (Reimann and Slaman, 2015) in terms of representations of measures, all of which were shown to be equivalent by (Day and Miller, 2013). In this paper we will mainly use the specific case when the measure is computable, which is part of the classic definition of (Martin-Löf, 1966). Given a computable measure μ, a Martin-Löf μ-test is a uniformly c.e. sequence (U_i) of sets of strings (viewed as the sets of reals with prefixes the strings in the sets) such that $\mu(U_i) < 2^{-i}$ for each i. A real Z is μ-random if it is not contained in the intersection of any Martin-Löf μ-test. By (Martin-Löf, 1966) there exists a universal Martin-Löf μ-test (uniformly in μ) i.e. a Martin-Löf μ-test (U_i) with the property that the set of μ-random reals is $2^\omega - \cap_i [\![U_i]\!]$ (where $[\![U_i]\!]$ denotes the set of reals with prefixes in U_i). Equivalently, if K is the prefix-free Kolmogorov complexity function, Z is μ-random if there exists $c \in \mathbb{N}$ such that $\forall n\ K(Z \upharpoonright_n) > -\log \mu(Z \upharpoonright_n) - c$. Occasionally it is useful to refer to the *randomness deficiency* of a real, which can be defined in many equivalent ways.[13] For example, we could define μ-deficiency to be the least i such that $Z \notin [\![U_i]\!]$ where (U_i) is the universal Martin-Löf μ-test, or $\sup_n(\lceil -\log \mu(Z \upharpoonright_n)\rceil - K(Z \upharpoonright_n))$. Clearly Z is μ-random if and only if it has finite μ-deficiency. Randomness with respect to arbitrary measures only plays a role in Section 3.1. We define it in terms of randomness deficiency, following (Bienvenu et al., 2018). We define the (uniform) *randomness deficiency function* to be the largest, up to an additive constant, function $\mathbf{d} : 2^\omega \times \mathcal{M} \to \mathbb{N} \cup \{\infty\}$ such that[14]

- the sets $\mathbf{d}^{-1}((k, \infty))$ are effectively open uniformly in k;
- $\mu(\{X \mid \mathbf{d}(X, \mu) > k\}) < 2^{-k}, \forall X \in 2^\omega, \mu \in \mathcal{M}, k \in \mathbb{N}$.

Given any $\mu \in \mathcal{M}$ and $Z \in 2^\omega$, the μ-deficiency of Z is $\mathbf{d}(Z, \mu)$ and Z is μ-random if it has finite μ-deficiency. This definition is based on the uniform tests approach as mentioned before, and is equivalent to Martin-Löf randomness for computable measures. Moreover, the deficiency notions are equivalent in the sense of footnote 13. The reader may find additional background on algorithmic randomness in the monographs (Li and Vitányi, 1997) and (Downey and Hirschfeldt, 2010).

3 Proof of Theorem 1.6 and Theorem 1.7

We start with Theorem 1.6. Let $\mathcal{D} \subseteq 2^\omega$ be an effectively closed set and let $\mathcal{D}^* \subseteq \mathcal{D}$ contain only computable reals. Also let $f : 2^\omega \to \mathcal{M}$ be a computable function such that for any $X \neq Y$ in $\mathcal{D}$ the measures $f(X), f(Y)$

[13] Equivalent in the sense that from an upper bound of one notion with respect to a real, we can effectively obtain an upper bound on another notion with respect to the same real.

[14] We can get a precise definition of $\mathbf{d}$ by starting with a universal enumeration $W_e(k)$ all uniform c.e. sequences of sets $W(k)$, where each $W(k)$ is a set of pairs (σ, I) of basic open sets of $2^\omega, \mathcal{M}$ respectively (viewed as basic open set of the product space $2^\omega \times \mathcal{M}$) with the property that for each $\mu \in I$ $\mu([\![\{\sigma \mid (\sigma, I) \in W(k)\}]\!]) < 2^{-k}$. Then define $\mathbf{d}(X, \mu) = \sum_e 2^{-e} \cdot w_e(X, \mu)$ where $w_e(X, \mu)$ is the maximum k such that (X, μ) is in the open set $W_e(k)$.

are effectively orthogonal. The easiest direction of Theorem 1.6 is that if $\mathcal{D}^*$ is (EX or BC) learnable then $f(\mathcal{D}^*)$ is (EX or BC, respectively) learnable, and is proved in Section 3.1. We stress that the effective orthogonality property of f, and hence the fact that it is injective, is used in a crucial way in the argument of Section 3.1. Sections 3.2 and 3.3 prove the 'if' direction of Theorem 1.6 for EX and BC learnability respectively, and are the more involved part of this paper. In Section 3.5 we prove Theorem 1.7.

3.1 From learning reals to learning measures

We show the 'only if' direction of Theorem 1.6, first for EX learning and then for BC learning. Let $f, \mathcal{D}, \mathcal{D}^*$ be as in the statement of Theorem 1.6. Since f maps distinct reals in $\mathcal{D}$ to effectively orthogonal measures, given $X \in 2^\omega$ there exists at most one $\mu \in f(\mathcal{D})$ such that X is μ-random. By the properties of f, there is also at most one $Z \in \mathcal{D}$ such that X is $f(Z)$-random. It follows from the definition of deficiency in Section 2.3, that for each $X \in 2^\omega$, $c \in \mathbb{N}$, the class of $Z \in \mathcal{D}$ such that X is $f(Z)$-random with deficiency $\leq c$ is a $\Pi^0_1(X)$ class $P(X, c)$ (uniformly in X, c). By the effective orthogonality of the image of $\mathcal{D}$ under f, the latter class either contains a unique real, or is empty. Moreover, the latter case occurs if and only if there is no $\mu \in f(\mathcal{D})$ with respect to which X is μ-random with deficiency $\leq c$. Now note that given a $\Pi^0_1(X)$ class $P \subseteq 2^\omega$, by compactness the emptiness of P is a $\Sigma^0_1(X)$ event, and if P contains a unique path, this path is uniformly computable from X and an index of P.

It follows that there exists a computable function $h : 2^{<\omega} \to 2^{<\omega}$ such that for all X which is $f(Z)$-random for some $Z \in \mathcal{D}$,

- $\lim_s |h(X \upharpoonright_s)| = \infty$;
- there exists n_0 such that for all $m > n > n_0$ we have $h(X \upharpoonright_n) \preceq h(X \upharpoonright_m)$;
- as $n \to \infty$ the prefixes $h(X \upharpoonright_n)$ converge to the unique real $Z \in \mathcal{D}$ such that X is $f(Z)$-random.

Indeed, on the initial segments of X, the function h will start generating the classes $P(X, c)$ as we described above, starting with $c = 0$ and increasing c by 1 each time that the class at hand becomes empty. While this process is fixed on some value of c, it starts producing the initial segments of the unique path of $P(X, c)$ (if there are more than one path, this process will stop producing longer and longer strings, reaching a finite partial limit). In the special case that X is $f(Z)$-random for some $Z \in \mathcal{D}$, such a real $Z \in \mathcal{D}$ is unique, and the process will reach a limit value of c, at which point it will produce a monotone sequence of longer and longer prefixes of Z. [15]

Note that since $f : 2^\omega \to \mathcal{M}$ is computable, there exists a computable $g : \mathbb{N} \to \mathbb{N}$ such that for each e, if e is an index of a computable $Z \in 2^\omega$, then $g(e)$ is an index of the computable measure $f(Z)$.

We are ready to define an EX-learner $\mathcal{V}$ for $f(\mathcal{D}^*)$, given an EX-learner $\mathcal{L}$ for $\mathcal{D}^*$ and the functions h, g that we defined above. For each σ we let $\mathcal{V}(\sigma) = g(\mathcal{L}(h(\sigma)))$. It remains to verify that for each X which is μ-random for some computable $\mu \in f(\mathcal{D}^*)$, the limit $\lim_s \mathcal{V}(X \upharpoonright_s)$ exists and equals an index for (the

[15] Alternatively, in order to obtain h, one can make use of a result from (Bienvenu and Monin, 2012). Since f is computable, 2^ω is compact and $\mathcal{D}$ is effectively closed, the image $f(\mathcal{D}^*)$ is compact and effectively closed, and the set of indices of computably enumerable sets of basic open sets of $\mathcal{M}$ whose union contains $f(\mathcal{D}^*)$ is itself computably enumerable. In the terminology of (Bienvenu and Monin, 2012), the image $f(\mathcal{D}^*)$ is *effectively compact*. Bienvenu and Monin (Bienvenu and Monin, 2012) showed that if a class C of effectively orthogonal measures is effectively compact then there exists a computable function $F : 2^\omega \times \mathbb{N} \to \mathcal{M}$ such that $\forall \mu \in C\ \forall X \in 2^\omega\ \forall c \in \mathbb{N}\ \ u(X, \mu) < c \Rightarrow F(X, c) = \mu$ where $u(X, \mu)$ is the μ-deficiency of X. One can derive the existence of h from this result.

unique such) μ. By the choice of X and h we have that there exists some s_0 such that for all $s > s_0$, the string $h(X \upharpoonright_s)$ is an initial segment of the unique $Z \in \mathcal{D}^*$ such that $f(Z) = \mu$; moreover, $\lim_s |h(X \upharpoonright_s)| = \infty$ and since μ is computable and $\mathcal{D}$ is effectively closed, it follows that Z is computable. Hence, since $\mathcal{L}$ learns all reals in $\mathcal{D}^*$, we get that $\lim_s \mathcal{L}(h(X \upharpoonright_s))$ exists and is an index of Z. Then by the properties of g we get that $g(\lim_s \mathcal{L}(h(X \upharpoonright_s))) = \lim_s g(\mathcal{L}(h(X \upharpoonright_s)))$ is an index for μ. Hence $\lim_s \mathcal{V}(X \upharpoonright_s)$ is an index of the unique computable $\mu \in f(\mathcal{D}^*)$ with respect to which X is random, which concludes the proof.

Finally we can verify that the same argument shows that if $\mathcal{D}^*$ is BC-learnable, then $f(\mathcal{D}^*)$ is BC-learnable. The definitions of h, g remain the same. The only change is that now we assume that $\mathcal{L}$ is a BC-learner for $\mathcal{D}^*$. We define the BC-learner $\mathcal{V}$ for $f(\mathcal{D}^*)$ in the same way: $\mathcal{V}(\sigma) = g(\mathcal{L}(h(\sigma)))$. As before, given X such that there exists (a unique) $Z \in \mathcal{D}^*$ such that X is $f(Z)$-random, we get that there exists some s_0 such that for all $s > s_0$, the string $h(X \upharpoonright_s)$ is an initial segment of the unique computable $Z \in \mathcal{D}^*$ such that $f(Z) = \mu$, and moreover, $\lim_s |h(X \upharpoonright_s)| = \infty$. Since $\mathcal{L}$ is a BC-learner for $\mathcal{D}^*$, there exists some s_1 such that for all $s > s_1$ the integer $\mathcal{L}(h(X \upharpoonright_s))$ is an index for the computable real Z. Then by the properties of g we get that for all $s > s_1$, the integer $g(\mathcal{L}(h(X \upharpoonright_s)))$ is an index for the computable measure $f(Z)$. Since X is $f(Z)$-random, this concludes the proof of the BC clause of the 'only if' direction of Theorem 1.6.

3.2 From learning measures to learning reals: the EX case

We show the 'if' direction of the EX case of Theorem 1.6. Let $f, \mathcal{D}, \mathcal{D}^*$ be as given in the theorem and suppose that $f(\mathcal{D}^*)$ is EX-learnable. This means that there exists a computable learner $\mathcal{V}$ such that for every $Z \in \mathcal{D}^*$ and every $f(Z)$-random X, the limit $\lim_s \mathcal{V}(X \upharpoonright_s)$ exists and is an index of $f(Z)$. We are going to construct a learner $\mathcal{L}$ for $\mathcal{D}^*$ so that for each $Z \in \mathcal{D}^*$ the limit $\lim_s \mathcal{L}(Z \upharpoonright_s)$ exists and is an index for Z. Since $\mathcal{D}$ is effectively closed and f is computable and injective on $\mathcal{D}$, by the compactness of 2^ω,

> there exists a computable $g : \mathbb{N} \to \mathbb{N}$ such that for each e, if e is an index of a computable $\mu \in f(\mathcal{D})$, the image $g(e)$ is an index of the unique and computable $Z \in \mathcal{D}$ such that $f(Z) = \mu$. (4)

Hence it suffices to

> construct a computable function $\mathcal{L}^* : 2^{<\omega} \to \mathbb{N}$ with the property that for each $Z \in \mathcal{D}^*$ the limit $\lim_s \mathcal{L}^*(Z \upharpoonright_s)$ exists and is an index for $f(Z)$ (5)

because then the function $\mathcal{L}(\sigma) = g(\mathcal{L}^*(\sigma))$ will be a computable learner for $\mathcal{D}^*$.

Since $f : 2^\omega \to \mathcal{M}$ is computable, there exists a computable $f^* : 2^{<\omega} \to \mathcal{M}^*$ (where $\mathcal{M}^*$ is the set of basic open sets of $\mathcal{M}$) and a computable increasing $h : \mathbb{N} \to \mathbb{N}$ such that:

- $\sigma \preceq \tau$ implies $f^*(\sigma) \subseteq f^*(\tau)$;
- for all $Z \in 2^\omega$, $\lim_s f^*(Z \upharpoonright_s) = f(Z)$;
- for all n and all $\sigma \in 2^{h(n)}$ the size of $f^*(\sigma)$ is at most 2^{-3n}.

Note that by the properties of f^* we have

> for each $Z \in 2^\omega$, each $n \in \mathbb{N}$ and any measures $\mu, \nu \in f^*(Z \upharpoonright_{h(n)})$ we have $\sum_{\sigma \in 2^n} |\mu(\sigma) - \nu(\sigma)| < 2^{-n}$. (6)

Below we will also use the fact that

> there is a computable function that takes as input any basic open interval I of $\mathcal{M}$ and returns (an index of) a computable measure (say, as a measure representation) $\mu \in I$. (7)

Proof idea. Given $Z \in \mathcal{D}^*$ we have an approximation to the measure $\mu^* = f(Z)$. Given μ^* and $\mathcal{V}$ we get a majority vote on each of the levels of the full binary tree, where each string σ votes for the index $\mathcal{V}(\sigma)$ and its vote has weight $\mu^*(\sigma)$. In search for the index of $Z \in \mathcal{D}^*$ we approximate the weights of the various indices as described above, and aim to chose an index with a positive weight. If $\mathcal{V}$ EX-learns μ^*, it follows that such an index will indeed be an index of μ^*. One obvious way to look for such an index is at each stage to choose the index whose current approximated weight is the largest. This approach has the danger that there may be two different indices with the same weight, in which case it is possible that the said approximation $\lim_n \mathcal{L}^*(X\upharpoonright_n)$ does not converge. We deal with this minor issue by requiring a sufficient difference on the current weights for a change of guess.

Construction of $\mathcal{L}^*$. We let $\mathcal{L}^*$ map the empty string to index 0 and for every other string σ we define $\mathcal{L}^*(\sigma)$ as follows. If $\sigma \notin \{2^{h(n)} \mid n \in \mathbb{N}\}$ then let $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\tau)$ where τ is the longest prefix of σ in $\{2^{h(n)} \mid n \in \mathbb{N}\}$. So it remains to define $\mathcal{L}^*$ in steps, where at step n we define $\mathcal{L}^*$ on all strings $\sigma \in 2^{h(n)}$. Since $f^*(\sigma)$ is basic open interval in $\mathcal{M}$, we may use (7) in order to get a computable function $\sigma \to \mu_\sigma$ from strings to computable measures, such that for each σ the measure μ_σ belongs to $f^*(\sigma)$.[16]

Given n and $\sigma \in 2^{h(n)}$, for each e define

$$\mathtt{wgt}_n(e) = \mu_\sigma(\{\tau \in 2^n \mid \mathcal{V}(\tau) = e\}).$$

Let e^* be the least number with the maximum $\mathtt{wgt}_n(e)$.[17] Let σ^- denote the first $|\sigma| - 1$ bits of σ. If $\mathtt{wgt}_n(e^*) > 3 \cdot \mathtt{wgt}_n(\mathcal{L}^*(\sigma^-))$, let $\mathcal{L}^*(\sigma) = e^*$; otherwise let $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\sigma^-)$.

Properties of $\mathcal{L}^*$. It remains to show (5), so let $Z \in \mathcal{D}^*$. First we show the claimed convergence and then that the limit is an index for $f(Z)$. Let $\mu^* := f(Z)$ and for each e define

$$\begin{aligned} w_e &= \mu^*(\{X \mid \lim_s \mathcal{V}(X\upharpoonright_s) = e\}) \\ w_e[n] &= \mu^*(\{\sigma \in 2^n \mid \mathcal{V}(\sigma) = e\}) \end{aligned}$$

Since $Z \in \mathcal{D}^*$ it follows that $\mathcal{V}$ learns μ^*. Hence the μ^*-measure of all the reals X such that $\lim_s \mathcal{V}(X\upharpoonright_s)$ exists and equals an index of a measure with respect to which X is random, is 1. If we take into account that $f(\mathcal{D})$ is effectively orthogonal, it follows that

> the μ^*-measure of all the reals X such that $\lim_s \mathcal{V}(X\upharpoonright_s)$ exists and equals an index of μ^* is 1. Hence there exists an index t of $\mu^* = f(Z)$ such that $w_t > 0$, and moreover, each e with $w_e > 0$ is an index of μ^*. (8)

Lemma 3.1. *For each e, $\lim_n w_e[n] = w_e$.*

Proof. Since $\mathcal{V}$ learns μ^*, the μ^*-measure of the reals on which $\mathcal{V}$ reaches a limit is 1. For each n let Q_n be the open set of reals on which $\mathcal{V}$ changes value after n bits. Then $Q_{n+1} \subseteq Q_n$ and $\lim_n \mu^*(Q_n) = \mu^*(\cap_n Q_n) =$

[16] The reader should not confuse this notation with the notation (μ_e) that we used for the universal list of all computable measures.

[17] Note that there are at most 2^n many e with $\mathtt{wgt}_n(e) \neq 0$ so this maximum exists. Moreover, we can compute the set of these numbers e, the maximum and e^*, by computing $\mu_\sigma(\tau)$ and $\mathcal{V}(\tau)$ for each $\tau \in 2^n$.

0. Let $P_e[n]$ be the closed set for reals X such that $\mathcal{V}(X\restriction_i) = e$ for all $i \geq n$. Then $P_e[n] \subseteq P_e[n+1]$ for all n and w_e is the μ^*-measure of $\cup_n P_e[n]$. Hence $w_e = \lim_n \mu^*(P_e[n])$.

Given n_0, for each $n \geq n_0$ we have $w_e[n] \leq \mu^*(P_e[n_0]) + \mu^*(Q_{n_0})$. This shows that $\limsup_n w_e[n] \leq \limsup_n \mu^*(P_e[n]) = w_e$. On the other hand $P_e[n_0] \subseteq [\![\{\sigma \in 2^n \mid \mathcal{V}(\sigma) = e)\}]\!]$ for all $n \geq n_0$. So $w_e = \lim_n(\mu^*(P_e[n])) \leq \liminf_n w_e[n]$. It follows that $\lim_n w_e[n] = w_e$. □

Now, given Z consider the sequence of computable measures $\mu_{Z\restriction_{h(n)}} \in f^*(Z\restriction_{h(n)})$ that are defined by the function $\sigma \mapsto \mu_\sigma$ applied on Z, and let

$$w_e^*[n] = \mu_{Z\restriction_{h(n)}}(\{\sigma \in 2^n \mid \mathcal{V}(\sigma) = e\}).$$

From (6) we get that for each n, e,

$$|w_e[n] - w_e^*[n]| < 2^{-n}. \tag{9}$$

In particular, by Lemma 3.1, $w_e = \lim_n w_e[n] = \lim_n w_e^*[n]$. Let m be some index such that $w_m = \max_e w_e$.

Lemma 3.2. *There exists n_0 such that for all $n \geq n_0$ and all e, $|w_e - w_e^*[n]| < w_m/5$.*

Proof. By (8) we have $\sum_e w_e = 1$ and $0 < w_m \leq 1$. Then there exists e_0 such that $\sum_{e<e_0} w_e > 1 - w_m/20$. Since we also have for all e, $\lim_n w_e^*[n] = w_e$, there exists n_0 such that for all $n \geq n_0$, $\sum_{e<e_0} |w_e - w_e^*[n]| < w_m/20$. Then for $e < e_0$, it is clear that for all $n \geq n_0$, $|w_e - w_e^*[n]| < w_m/5$. On the other hand, we have $\sum_{e \geq e_0} w_e = 1 - \sum_{e<e_0} w_e < w_m/20$. Moreover, for all $n \geq n_0$,

$$\sum_{e \geq e_0} w_e^*[n] = 1 - \sum_{e<e_0} w_e^*[n] \leq 1 - \Big(\sum_{e<e_0} w_e - w_m/20\Big) < w_m/10.$$

So for all $e \geq e_0$, $0 \leq w_e < w_m/20$ and $0 \leq w_e^*[n] < w_m/10$, and thus, $|w_e - w_e^*[n]| < w_m/5$. □

Let us now fix the constant n_0 of Lemma 3.2.

Lemma 3.3 (The limit exists)**.** *The value of $\mathcal{L}^*(Z\restriction_n)$ will converge to some index i with $w_i > 0$.*

Proof. Let $\mathcal{L}^*(Z\restriction_{h(n_0)}) = e_0$. In the case that there is some $n \geq h(n_0)$ with $\mathcal{L}^*(Z\restriction_n) \neq e_0$, there is some $n_1 \geq n_0$ such that $\mathcal{L}^*(Z\restriction_{h(n_1)}) = e_1 \neq e_0$. It then follows from the construction of $\mathcal{L}^*$ that $w_{e_1}^*[n_1] \geq w_m^*[n_1] > 4w_m/5$. Then by Lemma 3.2 for all $n \geq n_1$,

$$w_{e_1}^*[n] > w_{e_1} - w_m/5 > w_{e_1}^*[n_1] - 2w_m/5 = 2w_m/5$$

and on the other hand for all e, $w_e^*[n] < w_e + w_m/5 \leq 6w_m/5 < 3w_{e_1}^*[n]$. This means that after step n_1 the value of $\mathcal{L}^*(Z\restriction_n)$ will not change and thus, $\lim_n \mathcal{L}^*(Z\restriction_n) = e_1$ and $w_{e_1} > 4w_m/5 > 0$. In the case that for all $n \geq h(n_0)$ we have $\mathcal{L}^*(Z\restriction_n) = e_0$, we only need to show that $w_{e_0} > 0$. Assuming that $w_{e_0} = 0$, there will be some $n_2 > n_0$ such that for all $n \geq n_2$, $w_{e_0}^*[n] < w_m/4$. Note that $w_m^*[n] > 4w_m/5 > 3w_{e_0}^*[n]$, so by the construction of $\mathcal{L}^*$ the value of $\mathcal{L}^*(Z\restriction_{h(n_2)})$ will change. This is a contradiction. □

The above lemma together with (8) concludes the proof of (5) and the ‘only if’ direction of Theorem 1.6.

3.3 From learning measures to learning reals: the BC case

We show the 'if' direction of the BC case of Theorem 1.6. So consider $f : 2^\omega \to \mathcal{M}$, $\mathcal{D}, \mathcal{D}^* \subseteq 2^\omega$ as given and assume that $f(\mathcal{D}^*)$ is a BC-learnable class of computable measures. This means that there exists a learner $\mathcal{V}$ such that for all $\mu \in f(\mathcal{D}^*)$ and μ-random X

$$\text{there exists some } s_0 \text{ such that for all } s > s_0 \text{ the value of } \mathcal{V}(X \upharpoonright_s) \text{ is an index of } \mu. \tag{10}$$

We use the expression $\lim_n \mathcal{V}(X \upharpoonright_n) \approx \mu$ in order to denote property (10). Hence our hypothesis about $\mathcal{V}$ is

$$\text{for all } \mu \in f(\mathcal{D}^*) \text{ and } \mu\text{-random } X \text{ we have } \lim_n \mathcal{V}(X \upharpoonright_n) \approx \mu. \tag{11}$$

Proof idea. We would like to employ some kind of majority argument as we did in Section 3.2. The problem is that now, given $Z \in \mathcal{D}^*$, there is no way to assign weight on the various indices suggested by $\mathcal{V}$, in a way that this weight can be consistently approximated. The reason for this is that $\mathcal{V}$ is only a BC-learner and at each step the index guesses along the random reals with respect to $\mu^* = f(Z)$ may change. However there is a convergence in terms of the actual measures that the various indices represent, so we use a function that takes any number of indices, and as long as there is a majority with respect to the measures that these indices describe, it outputs an index of this majority measure. With this modification, the rest of the argument follows the structure of Section 3.2.

The formal argument.

Definition 3.4 (Weighted sets)**.** A weighted set is a finite set $A \subset \mathbb{N}$ along with a computable function $(i, s) \mapsto w_i[s]$ from $A \times \mathbb{N}$ to the dyadic rationals such that $w_i[s] \leq w_i[s+1]$ and $\sum_{i \in A} w_i[s] \leq 1$ for all s. Given such a weighted set, the weight of any subset $B \subseteq A$ is $\sum_{i \in B} w_i$, where $w_i := \lim_s w_i[s]$.[18]

In the following we regard each partial computable measure μ_e as a c.e. set of tuples (σ, I) such that I is a basic open set of $[0, 1]$ and $\mu_e(\sigma) \in I$ (see Section 2.1).

Definition 3.5 (Majority measures)**.** Given a weighted set A and a partial computable measure μ, if the weight of $A \cap \{e \mid \mu_e = \mu\}$ is more than $1/2$ we say that μ is the majority partial computable measure of A.

Note that there can be at most one majority partial computable measure of a weighted set. In the case that μ of Definition 3.5 is total, we call it the majority measure of A.

Lemma 3.6. *There is a computable function that maps any index of a weighted set A to an index of a partial computable measure μ with the property that if A has a majority partial computable measure ν then $\mu = \nu$.*[19]

Proof. Given a weighted set A we effectively define a partial computable measure μ and then verify its properties. We view partial computable measures as c.e. sets of tuples (σ, I) where $\sigma \in 2^{<\omega}$ and I is an open rational interval of $[0, 1]$ and $(\sigma, I) \in \mu$ indicates that $\mu(\sigma) \in I$. Define the weight of tuple (σ, I) to be the weight of $\{i \in A : (\sigma, I) \in \mu_i\}$. Then define μ as the tuples (σ, I) of weight $> 1/2$.

[18] It follows from Definition 3.4 that there is a uniform enumeration of all weighted sets as programs, so we may refer to *an index of a weighted set*. Just like in any uniform enumeration of programs, we can fix a numbering such that any $e \in \mathbb{N}$ may be regarded as an index of a weighted set.

[19] More formally, there exists a computable function $g : \mathbb{N} \to \mathbb{N}$ such that for each index e of a weighted set A, $g(e)$ is an index of a partial computable measure, and in the case that A has a majority partial computable measure ν, $\mu_{g(e)} = \nu$.

It remains to verify that if A has a majority partial computable measure then μ is the majority partial computable measure of A. If ν is the majority partial computable measure of A it is clear that for each $(\sigma, I) \in \nu$ we have $(\sigma, I) \in \mu$. Conversely, if $(\sigma, I) \in \mu$, there would be a subset $B \subseteq A$ of weight $> 1/2$ such that $(\sigma, I) \in \mu_i$ for all $i \in B$. Since ν is the majority partial computable measure of A, it follows that there is an index of ν in B (otherwise the weight of A would exceed 1). Hence $(\sigma, I) \in \nu$, which concludes the proof. □

Recall the function g from (4). It suffices to show that

$$\text{there exists a computable function } \mathcal{L}^* : 2^{<\omega} \to 2^{<\omega} \text{ such that for each } Z \in \mathcal{D}^* \text{ we have } \lim_s \mathcal{L}^*(Z \upharpoonright_s) \approx f(Z) \tag{12}$$

because then the function $\mathcal{L}(\sigma) = g(\mathcal{L}^*(\sigma))$ will be a computable BC-learner for $\mathcal{D}^*$.

Definition of $\mathcal{L}^*$. We let $\mathcal{L}^*$ map the empty string to index 0 and for every other string σ we define $\mathcal{L}^*(\sigma)$ as follows. If $\sigma \notin \{2^{h(n)} \mid n \in \mathbb{N}\}$ then let $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\tau)$ where τ is the longest prefix of σ in $\{2^{h(n)} \mid n \in \mathbb{N}\}$. So it remains to define $\mathcal{L}^*$ in steps, where at step n we define $\mathcal{L}^*$ on all string $\sigma \in 2^{h(n)}$. Since $f^*(\sigma)$ is basic open interval in $\mathcal{M}$, we may use (7) in order to get a computable function $\sigma \to \mu_\sigma$ from strings to computable measures, such that for each σ the measure μ_σ belongs to $f^*(\sigma)$. Given n and $\sigma \in 2^{h(n)}$, for each e define

$$\mathtt{wgt}_n(e) = \mu_\sigma(\{\tau \in 2^n \mid \mathcal{V}(\tau) = e\}).$$

Let A_n be the weighted set of all e such that $\mathtt{wgt}_n(e) > 0$ (clearly there are at most 2^n many such numbers e) where the weight of $e \in A_n$ is $\mathtt{wgt}_n(e)$. Then apply the computable function of Lemma 3.6 to A_n and let $\mathcal{L}^*(\sigma)$ be the resulting index.

Properties of $\mathcal{L}^*$. We show that $\mathcal{L}^*$ satisfies (12), so let Z be a computable member of $\mathcal{D}^*$.

Let $\mu^* := f(Z)$ and define

$$\begin{aligned} w &= \mu^*\Big(\{X \mid \lim_s(\mathcal{V}(X \upharpoonright_s) \approx \mu^*)\}\Big) \\ w_n &= \mu^*\Big(\{\sigma \in 2^n \mid \mathcal{V}(\sigma) \text{ is an index of } \mu^*)\}\Big). \end{aligned}$$

Lemma 3.7. $\lim_n w_n = w = 1$.

Proof. Since $Z \in \mathcal{D}^*$ it follows that $\mathcal{V}$ learns μ^*, hence $w = 1$. It remains to show that $\lim_n w_n = w$. For each n let Q_n be the open set of reals X with the property that there exists some $t > n$ such that $\mathcal{V}(X \upharpoonright_t)$ is not an index of μ^*. Then $Q_{n+1} \subseteq Q_n$ and since $\mathcal{V}$ learns μ^* we have $\lim_n \mu^*(Q_n) = \mu^*(\cap_n Q_n) = 0$. Let P_n be the closed set for reals X such that for all $i \geq n$ the value of $\mathcal{V}(X \upharpoonright_i)$ is an index of μ^*. Then $P_n \subseteq P_{n+1}$ for all n and w is the μ^*-measure of $\cup_n P_n$. Hence $w = \mu^*(\cup_n P_n) = \lim_n \mu^*(P_n)$.

Given n_0, for each $n \geq n_0$ we have $w_n \leq \mu^*(P_{n_0}) + \mu^*(Q_{n_0})$. This shows that $\limsup_n w_n \leq \limsup_n \mu^*(P_n) = w_e$. On the other hand $P_{n_0} \subseteq [\![\{\sigma \in 2^n \mid \mathcal{V}(\sigma) \text{ is an index of } \mu^*)\}]\!]$ for all $n \geq n_0$. So $w = \lim_n(\mu^*(P_n)) \leq \liminf_n w_n$. It follows that $\lim_n w_n = w$. □

Lemma 3.8. *For each $Z \in \mathcal{D}^*$, there exists some n_0 such that for all $n > n_0$ the value of $\mathcal{L}^*(Z \upharpoonright_n)$ is an index of $f(Z) = \mu^*$.*

Proof. Given $Z \in \mathcal{D}^*$ consider the definition of $\mathcal{L}^*(Z \upharpoonright_{h(n)})$ during the various stages n, and the associated weighted sets A_n . According to the construction of $\mathcal{L}^*$ and Lemma 3.6 it suffices to show that

$$\text{there exists } n_0 \text{ such that for all } n > n_0 \text{ the weighted set } A_n \text{ in the definition of } \mathcal{L}^*(Z \upharpoonright_{h(n)}) \text{ has a majority measure which equals } \mu^*. \tag{13}$$

Consider the sequence $\mu_{Z\upharpoonright_{h(n)}} \in f^*(Z \upharpoonright_{h(n)})$ of computable measures that are defined by the function $\sigma \mapsto \mu_\sigma$ applied on Z, and let

$$w_n^* = \mu_{Z\upharpoonright_{h(n)}}\Big(\{\sigma \in 2^n \mid \mathcal{V}(\sigma) \text{ is an index of } \mu^*)\}\Big).$$

From (6) we get that for each n, $|w_n - w_n^*| < 2^{-n}$. In particular, by Lemma 3.7, $\lim_n w_n = \lim_n w_n^* = 1$. For (13) it suffices to consider any n_0 such that for all $n > n_0$ we have $w_n^* > 1/2$. Then by the construction of $\mathcal{L}^*$ at step n and the definition of w_n^* it follows that for each $n > n_0$, the majority measure of the weighted set A_n is μ^*. □

Lemma 3.8 shows that $\mathcal{L}^*$ satisfies (12), which concludes the BC case of the proof of the 'if' direction of Theorem 1.6.

3.4 From learning measures to learning reals: an extension

There is a way in which we can relax the hypotheses of the 'if' direction of Theorem 1.6 for EX-learning, which concerns the strength of learning as well as the orthogonality hypothesis.

Definition 3.9 (Partial EX-learnability of classes of computable measures)**.** A class C of computable measures is partially EX-learnable if there exists a computable learner $\mathcal{V} : 2^{<\omega} \to \mathbb{N}$ such that

(a) C is weakly EX-learnable via $\mathcal{V}$ (recall Definition 1.3);

(b) for every $\mu \in C$ there exists a μ-random X such that $\lim_n \mathcal{V}(X \upharpoonright_n)$ is an index of μ.

The idea behind this notion is that not only for each $\mu \in C$ the learner eventually guesses a correct measure (possibly outside C) along each μ-random real, but in addition every measure $\mu \in C$ is represented as a response of the learner along *some* μ-random real.

Theorem 3.10 (An extension)**.** *Suppose that a computable function $f : 2^\omega \to \mathcal{M}$ is injective on an effectively closed set $\mathcal{D} \subseteq 2^\omega$, and $\mathcal{D}^* \subseteq \mathcal{D}$ is a set of computable reals. If $f(\mathcal{D}^*)$ is a partially EX-learnable class of computable measures then $\mathcal{D}^*$ is an EX-learnable class of computable reals.*

Proof idea. We would like to follow the argument of Section 3.2, but now we have a weaker assumption which allows the possibility that given $Z \in \mathcal{D}^*$, $\mu^* = f(Z)$, there are indices e with positive weight, which do not describe μ^*. In order to eliminate these guesses from the approximation $n \to \mathcal{L}^*(Z \upharpoonright_n)$ to an index of $f(Z)$, we compare how near the candidate measures are to our current approximation to μ^*. Using this approach, combined with the crucial fact (to be proved) that indices with positive weight correspond to total measures, allows us to eliminate the incorrect total measures (eventually they will be contained in basic open sets that are disjoint from the open ball $f(Z \upharpoonright_n)$ containing $f(Z)$) and correctly approximate an index of μ^*.

The formal argument. Recall the argument from Section 3.2 and note that (4) continues to hold under the hypotheses of Theorem 3.10. Hence it suffices to construct a computable $\mathcal{L}^* : 2^{<\omega} \to \mathbb{N}$ such that (5) holds.

Since $f(\mathcal{D}^*)$ is a partially EX-learnable class of computable measures, there exists $\mathcal{V}$ with the properties of Definition 3.9 with respect to $C := f(\mathcal{D}^*)$.

Lemma 3.11. *Every measure $\mu^* \in f(\mathcal{D}^*)$ has an index e such that $\lim_n \mathcal{V}(X \upharpoonright_n) = e$ for a positive μ^*-measure of reals X.*

Proof. Let $\mu^* \in f(\mathcal{D}^*)$ and consider a μ^*-random X such that $\lim_n \mathcal{V}(X \upharpoonright_n)$ is an index e of μ^*. Consider the Σ^0_2 class $\mathcal{F}$ of reals Z with the property that $\lim_n \mathcal{V}(Z \upharpoonright_n) = e$. It remains to show that $\mu^*(\mathcal{F}) > 0$. Since $\mathcal{F}$ is the union of a sequence of Π^0_1 classes and $X \in \mathcal{F}$, there exists a Π^0_1 class $P \subseteq \mathcal{F}$ which contains X. Since X is μ^*-random, it follows that $\mu^*(P) > 0$, so $\mu^*(\mathcal{F}) \geq \mu^*(P) > 0$. □

Given $\mu^* \in f(\mathcal{D}^*)$ define $w_e, w_e[n]$ as we did in Section 3.2. Note that Lemma 3.1 still holds by the same argument, since it only uses the hypotheses we presently have about $\mathcal{D}, f, \mathcal{V}$.

Lemma 3.12. *For every $\mu^* \in f(\mathcal{D}^*)$ there exists an index e of μ^* such that $w_e > 0$. Conversely, if $w_e > 0$ then e is an index of a computable measure μ'.*

Proof. The first claim is Lemma 3.11. For the second claim, if $w_e > 0$ it follows from clause (a) of Definition 3.9 applied on $\mathcal{V}$ that e is an index of a computable measure μ' such that all reals in some set Q with $\mu^*(Q) = w_e > 0$ are μ'-random. □

Let H be a partial computable predicate such that for every basic open set B of $\mathcal{M}$ and every e such that μ_e is total, we have $H(B, e) \downarrow$ if and only if $\mu_e \notin B$.[20] Hence

$$\text{if } \mu_e \text{ is total then, } \exists n\, H(f^*(X \upharpoonright_n), e)[n] \downarrow \iff \mu_e \neq \lim_n f^*(X \upharpoonright_n). \tag{14}$$

where the suffix '[n]' indicates the state of H after n steps of computation.

Construction of $\mathcal{L}^*$. We let $\mathcal{L}^*$ map the empty string to index 0 and for every other string σ we define $\mathcal{L}^*(\sigma)$ as follows. If $\sigma \notin \{2^{h(n)} \mid n \in \mathbb{N}\}$ then let $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\tau)$ where τ is the longest prefix of σ in $\{2^{h(n)} \mid n \in \mathbb{N}\}$. So it remains to define $\mathcal{L}^*$ in steps, where at step n we define $\mathcal{L}^*$ on all string $\sigma \in 2^{h(n)}$. Since $f^*(\sigma)$ is basic open interval in $\mathcal{M}$ so we may use (7) in order to get a computable function $\sigma \to \mu_\sigma$ from strings to computable measures, such that for each σ the measure μ_σ belongs to $f^*(\sigma)$.

Given n and $\sigma \in 2^{h(n)}$, for each e define $\texttt{wgt}_n(e) = \mu_\sigma(\{\tau \in 2^n \mid \mathcal{V}(\tau) = e\})$. Let σ^- denote the first $|\sigma| - 1$ bits of σ and define $t = \mathcal{L}^*(\sigma^-)$. Let e^* be the least number with the maximum $\texttt{wgt}_n(e)$ such that $H(f^*(\sigma), \mu_e)[n] \uparrow$; if this does not exist, define $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\sigma^-)$. Otherwise, if one of the following holds

(a) $\texttt{wgt}_n(e^*) > 3 \cdot \texttt{wgt}_n(t)$

(b) $H(f^*(\sigma), t)[n] \downarrow$

let $\mathcal{L}^*(\sigma) = e^*$. In any other case let $\mathcal{L}^*(\sigma) = \mathcal{L}^*(\sigma^-)$.

Properties of $\mathcal{L}^*$. We show that (5) holds, i.e. that for each $Z \in \mathcal{D}^*$ the limit $\lim_s \mathcal{L}^*(Z \upharpoonright_s)$ exists and is an index for $f(Z)$. Let $Z \in \mathcal{D}^*$, $\mu^* = f(Z)$ and consider the sequence of computable measures $\mu_{Z \upharpoonright_{h(n)}} \in$

[20] The machine for H starts producing a sequence of basic open sets A_s converging to μ_e based on the program e, and stops at the first stage s such that $B \cap A_s = \emptyset$, at which point it halts.

$f^*(Z\upharpoonright_{h(n)})$ that are defined by the function $\sigma \mapsto \mu_\sigma$ applied on Z, and are used in the steps n of the definition of $\mathcal{L}^*$ with respect to Z. Let

$$w_e^*[n] = \mu_{Z\upharpoonright_{h(n)}}(\{\sigma \in 2^n \mid \mathcal{V}(\sigma) = e\}).$$

and note that these are the weights used in the definition of $\mathcal{L}^*$ at step n with respect to $Z\upharpoonright_{h(n)}$.

Lemma 3.13. *For each e, $w_e = \lim_n w_e[n] = \lim_n w_e^*[n]$.*

Proof. The first equality is Lemma 3.1. From (6) we get that for each n, e, $|w_e[n] - w_e^*[n]| < 2^{-n}$, which establishes the second equality. □

Next, we show that $\lim_s \mathcal{L}^*(Z\upharpoonright_s)$ exists. Let $H_e[n]$ denote $H(f^*(Z\upharpoonright_{h(n)}), e)[n]$. Let $T = \{e : e \text{ is an index of } \mu^*\}$ and m be some index such that $w_m = \max\{w_e : e \in T\}$. By Lemma 3.11 $w_m > 0$. By (14), $e \in T$ if and only if for all n, $H_e[n] \uparrow$.

Lemma 3.14. *There exists n_0 such that for all $n \geq n_0$ and all e,*

(i) $|w_e - w_e^*[n]| < w_m/5$.

(ii) If $w_e > 4w_m/5$ and $e \notin T$ then $H_e[n] \downarrow$.

Proof. Since $\sum_e w_e = 1$ and $0 < w_m \leq 1$, then there exists e_0 such that $\sum_{e<e_0} w_e > 1 - w_m/20$. We also have that for all e, $\lim_n w_e^*[n] = w_e$, so there exists n^* such that for all $n \geq n^*$, $\sum_{e<e_0} |w_e - w_e^*[n]| < w_m/20$. Then for $e < e_0$, it is clear that for all $n \geq n^*$, $|w_e - w_e^*[n]| < w_m/5$. On the other hand, we have $\sum_{e \geq e_0} w_e = 1 - \sum_{e<e_0} w_e < w_m/20$. Moreover, for all $n \geq n^*$,

$$\sum_{e \geq e_0} w_e^*[n] = 1 - \sum_{e<e_0} w_e^*[n] \leq 1 - \Big(\sum_{e<e_0} w_e - w_m/20\Big) < w_m/10.$$

So for all $e \geq e_0$, $0 \leq w_e < w_m/20$ and $0 \leq w_e^*[n] < w_m/10$, and thus, $|w_e - w_e^*[n]| < w_m/5$. If $w_e > 4w_m/5$, it must be case that $e < e_0$, and thus, there are only finitely many such indices e. For every such index e, if $e \notin T$, there will be some k_e such that for all $n \geq k_e$, $H_e[n] \downarrow$. If n_0 is the largest number amongst k_e and n^*, clauses (i) and (ii) in the statement of the lemma hold. □

Let us now fix the constant n_0 of Lemma 3.14.

Lemma 3.15 (The limit exists)**.** *The value of $\mathcal{L}^*(Z\upharpoonright_n)$ will converge to some index $i \in T$.*

Proof. Let $\mathcal{L}^*(Z\upharpoonright_{h(n_0)}) = e_0$. In the case that there is some $n \geq h(n_0)$ with $\mathcal{L}^*(Z\upharpoonright_n) \neq e_0$, there should be some $n_1 > n_0$ such that $\mathcal{L}^*(Z\upharpoonright_{h(n_1)}) = e_1 \neq e_0$. It then follows from the construction of $\mathcal{L}^*$ that $w_{e_1}^*[n_1] \geq w_m^*[n_1] > 4w_m/5$ and $H_{e_1}[n_1] \uparrow$. Then by Lemma 3.14 for all $n \geq n_1$,

$$w_{e_1}^*[n] > w_{e_1} - w_m/5 > w_{e_1}^*[n_1] - 2w_m/5 = 2w_m/5$$

and $e_1 \in T$. On the other hand if $e \in T$, for all $n \geq n_1$ we have $w_e^*[n] < w_e + w_m/5 \leq 6w_m/5 < 3w_{e_1}^*[n]$. If $e \notin T$ but $w_e[n] > 6w_m/5$, then $w_e > w_m$, so for all $n \geq n_1$ we have $H_e[n] \downarrow$. This means that after step n_1 the value of $\mathcal{L}^*(Z\upharpoonright_n)$ will not change and thus, $\lim_n \mathcal{L}^*(Z\upharpoonright_n) = e_1 \in T$. In the case that for all $n \geq h(n_0)$ we have $\mathcal{L}^*(Z\upharpoonright_n) = e_0$, we only need to show that $e_0 \in T$. Assuming that $e_0 \notin T$, there exists some step $n_2 \geq n_0$ such that $H_{e_0}[n_2] \downarrow$. Since $m \in T$, for all $n \geq n_0$ we have $H_m[n] \uparrow$. By the construction of $\mathcal{L}^*$ the value of $\mathcal{L}^*(Z\upharpoonright_{h(n_2)})$ will change. This is a contradiction. □

The above lemma concludes the proof of Theorem 3.10.

3.5 Proof of Theorem 1.7

It is well known that if Z is computable and μ-random for some computable measure μ, then Z is an atom of μ and $\mu(Z\upharpoonright_n * Z(n))/\mu(Z\upharpoonright_n)$ tends to 1. Here is a generalization.

Lemma 3.16. *If Z is computable and $Z\oplus Y$ is μ-random for some computable measure μ, then $\mu(Z\upharpoonright_n \oplus Y\upharpoonright_n * Z(n))/\mu(Z\upharpoonright_n \oplus Y\upharpoonright_n) \to 1$ as $n\to\infty$.*

Proof. We prove the contrapositive: fix computable μ, Z, and suppose that for some Y there exists a rational $q\in(0,1)$ such that

$$\mu((Z\upharpoonright_n \oplus Y\upharpoonright_n) * Z(n)) < q\cdot\mu(Z\upharpoonright_n \oplus Y\upharpoonright_n) \tag{15}$$

for infinitely many n. For each t consider the set V_t of the strings of the form $(Z\upharpoonright_j \oplus X\upharpoonright_j) * Z(j)$ for some j, X, such that j is minimal with the property that there exist at least t many $n\leq j$ with (15) by replacing Y with X. For each nonempty string σ, let σ^- denote the largest proper prefix of σ. By the minimality of the choice of n above, we have that (a) V_t is prefix-free; (b) each string $\tau\in V_{t+1}$ extends a string $\sigma\in V_t$; (c) if $\sigma\in V_t$ then $\mu(\sigma)<q\cdot\mu(\sigma^-)$; (d) if $V_{t+1}(\sigma)$ is the set of all the strings in V_{t+1} extending $\sigma\in V_t$ then $\mu(V_{t+1}(\sigma))<q\cdot\mu(\sigma)$. It follows that $\mu(V_{t+1})<q\cdot\mu(V_t)$ so there exists a computable sequence (m_j) such that $\mu(V_{m_j})<2^{-j}$ for each j. So (V_{m_j}) is a μ-test and by its definition, if Y satisfies (15) for infinitely many n, then $Z\oplus Y$ has a prefix in V_t for each t, and so in V_{m_j} for each j. Hence in this case $Z\oplus Y$ is not μ-random. □

For each Z define μ_Z as follows: for each σ of odd length let $\mu_Z(\sigma * i)=\mu_Z(\sigma)/2$ for $i=0,1$; for each σ of even length let $j_\sigma = Z(|\sigma|/2)$ and define $\mu_Z(\sigma * j_\sigma)=\mu_Z(\sigma)$, $\mu_Z(\sigma*(1-j_\sigma))=0$. Hence for each real X and each n, all $\mu_Z(X\upharpoonright_{2n})$ goes to $X\upharpoonright_{2n} * Z(n)$ while $\mu_Z(X\upharpoonright_{2n+1})$ is split equally to $X\upharpoonright_{2n+1} *0$ and $X\upharpoonright_{2n+1} *1$. Note that for each Z the measure μ_Z is continuous. Also, the map $Z\mapsto\mu_Z$ from 2^ω to $\mathcal{M}$ is continuous.

Lemma 3.17. *Given any computable Z, a real X is μ_Z-random if and only if it is of the form $Z\oplus Y$ for some random Y with respect to the uniform measure.*

Proof. "⇒: " If X is of the form $W\oplus Y$ for some $W\neq Z$ then by the definition of μ_Z we have $\mu_Z((W\oplus Y)\upharpoonright_n)=0$ for sufficiently large n, so $W\oplus Y$ is not μ_Z-random. If X is of the form $Z\oplus Y$ and Y is not random with respect to the uniform measure λ, let (V_i) be a λ-test such that $Y\in\cap_i[\![V_i]\!]$. For each i let $U_i=\{Z\upharpoonright_{|\sigma|}\oplus\sigma \mid \sigma\in V_i\}$. By the definition of μ_Z we have $\mu_Z(U_i)=\lambda(V_i)\leq 2^{-i}$ so (U_i) is a μ_Z-test. Since $Y\in\cap_i[\![V_i]\!]$ we have $Z\oplus Y\in\cap_i[\![U_i]\!]$ hence $Z\oplus Y$ is not μ_Z-random.
"⇐: " If $Z\oplus Y$ is not μ_Z-random, then there is a μ_Z-test (U_i) such that $Z\oplus Y\in\cap_i[\![U_i]\!]$. For each i let $V_i=\{\sigma(1)\sigma(3)\cdots\sigma(2n-1) \mid \sigma\in U_i \text{ and } n=\lfloor|\sigma|/2\rfloor\}$. By the definition of μ_Z we have $\lambda(V_i)=\mu_Z(U_i)\leq 2^{-i}$ and $Y\in\cap_i[\![V_i]\!]$. So Y is not random with respect to the uniform measure. □

Hence if $Z\neq X$ are computable, the measures μ_Z,μ_X are effectively orthogonal. Then the 'only if' direction of Theorem 1.7 follows from the 'only if' direction of Theorem 1.6 (with $\mathcal{D}:=2^\omega$ and $\mathcal{D}^*:=\mathcal{C}$). The following concludes the proof of Theorem 1.7.

Lemma 3.18. *For each class C of computable reals, if $\{\mu_Z \mid Z \in C\}$ is a weakly EX/BC learnable class of measures then C is EX/BC learnable.*

Proof. We first show the EX case. Fix C and let $\mathcal{V}$ be a learner which EX-succeeds on all measures in $\{\mu_Z \mid Z \in C\}$. It remains to construct an EX-learner $\mathcal{L}$ for C.

Proof idea. Given a computable Z, in order to define $\mathcal{L}(Z \upharpoonright_n)$ we use $\mathcal{V}$ on the strings $Z \upharpoonright_n \oplus \sigma$, $\sigma \in 2^n$ and take a majority vote in order to determine $Z(n)$. According to Lemmas 3.16 and 3.17, eventually the correct value of $Z(n)$ will be the j such that $(Z \upharpoonright_n \oplus \sigma) * j$ gets most of the measure on $(Z \upharpoonright_n \oplus \sigma)$, with respect to any measure correctly guessed by $\mathcal{V}(Z \upharpoonright_n \oplus \sigma)$, for the majority of $\sigma \in 2^n$.

Construction of $\mathcal{L}$. First, define a computable $g_0 : 2^{<\omega} \to \mathbb{N}$ as follows, taking a majority vote via $\mathcal{V}$. For each Z, n we define $g_0(Z \upharpoonright_n)$ to be an index of the following partial computable real X. For each $m < n$ we let $X(m) = Z(m)$. If $m \geq n$, suppose inductively that it has already defined $X \upharpoonright_m$. In order to define $X(m)$, it calculates the measure-indices $\mathcal{V}(X \upharpoonright_m \oplus \sigma) = e$ for all $\sigma \in 2^m$ and waits until, for some $j \in \{0, 1\}$, at least $2/3$ these partial computable measures μ_e have the property $\mu_e((X \upharpoonright_n \oplus \sigma) * j) \downarrow > \mu_e(X \upharpoonright_n \oplus \sigma)/2$. If and when this happens it defines $X(m) = j$.

Fix $Z \in C$. By Lemma 3.16, if $\mathcal{V}$ weakly EX-learns μ_Z, for all sufficiently large n the value of $g_0(Z \upharpoonright_n)$ will be an index of Z (possibly different for each n). In order to produce a stable guess, define the function $\mathcal{L} : 2^{<\omega} \to \mathbb{N}$ as follows. In order to define $\mathcal{L}(Z \upharpoonright_n)$, consider the least $n_0 \leq n$ such that

(i) at least proportion $2/3$ of the strings $\sigma \in 2^n$ have not changed their $\mathcal{V}$-guess since n_0, i.e. $\mathcal{V}(Z \upharpoonright_i \oplus \sigma \upharpoonright_i) = \mathcal{V}(Z \upharpoonright_{n_0} \oplus \sigma \upharpoonright_{n_0})$ for all integers $i \in (n_0, n]$;

(ii) no disagreement between $Z \upharpoonright_n$ and the reals defined by the indices $\mathcal{L}(Z \upharpoonright_i)$, $i \in (n_0, n)$ has appeared up to stage n; formally, if $(\varphi_e[n])$ is the fixed effective list of all partial computable reals at stage n of the universal computation, there exists no $j < n$ and $i \in (n_0, n)$ such that $Z(j) \neq \varphi_i(j)[n]$.

Then let $\mathcal{L}(Z \upharpoonright_n)$ be $g_0(Z \upharpoonright_{n_0})$. Given $Z \in C$ we have that $\mathcal{V}$ weakly learns μ_Z, so $\mathcal{V}(Z \upharpoonright_n \oplus Y \upharpoonright_n)$ converges for almost all Y (with respect to the uniform measure). Hence in this case (i) will cease to apply for large enough n. Moreover, by the properties of g_0, clause (ii) will also cease to apply for sufficiently large n. Hence the n_0 in the definitions of $\mathcal{L}(Z \upharpoonright_n)$ will stabilize for large enough n, and $\mathcal{L}(Z \upharpoonright_n)$ will reach the limit $g_0(Z \upharpoonright_{n_0})$ which is an index for Z.

For the BC case, assume instead that $\mathcal{V}$ BC-succeeds on all measures in $\{\mu_Z \mid Z \in C\}$. We define g_0 exactly as above, and the BC-learner $\mathcal{L}$ by $\mathcal{L}(Z \upharpoonright_n) = g_0(Z \upharpoonright_n)$. Given $Z \in C$ we have that $\mathcal{V}$ weakly BC-learns μ_Z, so for almost all Y (with respect to the uniform measure), $\mathcal{V}(Z \upharpoonright_n \oplus Y \upharpoonright_n)$ eventually outputs indices of a computable measure μ (dependent on $Y \upharpoonright_n$) with the property that $\mu((Z \upharpoonright_n \oplus Y \upharpoonright_n) * Z(n)) > 2/3 \cdot \mu(Z \upharpoonright_n \oplus Y \upharpoonright_n)$. By the definition of g_0, this means that for sufficiently large n, the value of $\mathcal{L}(Z \upharpoonright_n)$ is an index of Z. Hence $\mathcal{L}$ is a BC-learner for C. □

3.6 Proof of Theorem 1.8

A stage s is called i-expansionary if $\ell_i[t] < \ell_i[s]$ for all i-expansionary stages $t < s$. By the padding lemma let p be a computable function such that for each i, j we have $\mu_{p(i,j)} \simeq \mu_i$ and $p(i, j) < p(i, j + 1)$.

Define the eth randomness deficiency function by setting $d_e(\sigma)$ to be $\lceil -\log \mu_e(\sigma) \rceil - K(\sigma)$ for each string σ, where K is the prefix-free complexity of σ. Define the eth randomness deficiency on a real X as:

$\mathbf{d}_e(X) = \sup_n d_e(X \upharpoonright_n)$ where the supremum is taken over the n such that $d_e(X \upharpoonright_n) \downarrow$. By (Levin, 1984), if μ_e is total then X is μ_e-random if and only if $\mathbf{d}_e(X) < \infty$.

At stage s, we define $\mathcal{L}(\sigma)$ for each σ of length s as follows. For the definition of $\mathcal{L}(\sigma)$ find the least i such that s is i-expansionary and $d_i(\sigma)[s] \leq i$. Then let j be the least such that $p(i, j)$ is larger than any k-expansionary stage $t < |\sigma|$ for any $k < i$ such that $d_k(\sigma \upharpoonright_k)[t] \leq k$, and define $\mathcal{L}(\sigma_t) = p(i, j)$.

Let X be a real. Note that $\mathcal{L}(X \upharpoonright_n) = x$ for infinitely many n, then $x = p(i, j)$ for some i, j, which means that $\mu_i = \mu_x$ is total and there are infinitely many x-expansionary stages as well as infinitely many i-expansionary stages. This implies that there are at most x many y-expansionary stages t for any $y < x$ with $d_y(\sigma \upharpoonright_y)[t] \leq y$. Moreover, for each $z > x$ there are at most finitely may n such that $\mathcal{L}(X \upharpoonright_n) = z$. Indeed, for each z if n_0 is an i-expansionary stage then $\mathcal{L}(X \upharpoonright_n) \neq z$ for all $n > n_0$. Moreover, if $\mathcal{L}(X \upharpoonright_n) = x$ for infinitely many n, then $d_x(X) = d_i(X) \leq i$ and μ_i is total, so X is μ_i-random. We have shown that for each X there exists at most one x such that $\mathcal{L}(X \upharpoonright_n) = x$ for infinitely many n, and in this case μ_x is total and X is μ_x-random.

It remains to show that if X is μ-random for some computable μ, then there exists some x such that $\mathcal{L}(X \upharpoonright_n) = x$ for infinitely many n. If X is μ_i-random for some i such that μ_i is total, let i be the least such number with the additional property that $\mathbf{d}_i(X) \leq i$ (which exists by the padding lemma). Also let j be the least number such that $p(i, j)$ is larger than any stage t which is k-expansionary for any $k < i$ with $d_k(\sigma \upharpoonright_k)[t] \leq k$. Then the construction will define $\mathcal{L}(X \upharpoonright_n) = p(i, j)$ for each i-expansionary stage n after the last k-expansionary stage t for any $k < i$ with $d_k(\sigma \upharpoonright_k)[t] \leq k$. We have shown that $\mathcal{L}$ partially succeeds on every μ-random X for any computable measure μ.

4 Applications

For the 'if' direction of Corollaries 1.9 and 1.11 we need the following lemma.

Lemma 4.1. *If A is high then the class of all computable measures and the class of all computable Bernoulli measures are both $EX[A]$-learnable.*

Proof. We first show the part for the computable Bernoulli measures. The function which maps a real $X \in 2^\omega$ to the measure representation $\mu : 2^\omega \to \mathbb{R}$ of the Bernoulli measure with success probability the real in $\mathbb{R}$ with binary expansion X is computable. Hence, given an effective list (μ_e) of all partial computable measures in $\mathcal{M}$ and an effective list (φ_e) of all partial computable reals in 2^ω, there exists a computable function $g : \mathbb{N} \to \mathbb{N}$ such that for each e such that φ_e is total, $\mu_{g(e)}$ is total and is the measure representation of the Bernoulli measure with success probability the real with binary expansion φ_e. Let $\mu_e[s]$ represent the state of approximation to μ_e at stage s of the universal approximation, so $\mu_e[s]$ is a basic open set of $\mathcal{M}$. Then the function $\sigma \to \sup\{\mu(\sigma) \mid \mu \in \mu_e[s]\}$ is computable and the function

$$d(e, \sigma)[s] = \lceil -\log\big(\sup\{\mu(\sigma) \mid \mu \in \mu_e[s]\}\big)\rceil - K(\sigma)[s]$$

is a computable approximation to the μ_e-deficiency of σ, in the case that μ_e is total. Since A is high there exists a function $h : \mathbb{N} \times \mathbb{N} \to \{0, 1\}$, $h \leq_T A$ such that for each e, φ_e is total if and only if $\lim_s h(e, s) = 1$. Define $cost(e, \sigma) = e + d(g(e), \sigma)[|\sigma|]$. We define an A-computable learner $\mathcal{V}$ as follows: for each σ let $\mathcal{V}(\sigma)$ be $g(e)$ for the least index $e \leq |\sigma|$ which minimizes $cost(e, \sigma)$ subject to the condition $h(e, |\sigma|) = 1$.

It remains to show that for each $X \in 2^\omega$ which is random with respect to a computable Bernoulli measure μ, $\lim_n \mathcal{V}(X \upharpoonright_n)$ exists and equals an index of μ. According to our working assumption about X, there exist numbers e such that φ_e is total and $\sup_n cost(e, X \upharpoonright_n) < \infty$. These numbers e are the indices of reals in 2^ω which are the binary representations of the success probability of the Bernoulli measure with respect to which X is random. Now consider the least e with this property, and which minimizes $\sup_n cost(e, X \upharpoonright_n)$. Note that, by the definition of $cost(e, \sigma)$, for each k, σ there are only finitely many e such that $cost(e, \sigma) < k$. It follows by the construction of $\mathcal{V}$ that $\lim_n \mathcal{V}(X \upharpoonright_n) = e$.

The proof for the class of all computable measures is the same as above, except that we take g to be the identity function. □

For the other direction of Corollary 1.9, let C be the class of all computable reals, and assume that the computable measures are weakly EX[A]-learnable. Then $\{\mu_Z \mid Z \in C\}$ is also weakly EX[A]-learnable, and by Theorem 1.7 we get that C is EX[A]-learnable. Then by (Adleman and Blum, 1991) it follows that A is high.

4.1 Applying Theorem 1.6 to classes of Bernoulli measures

Perhaps the most natural parametrization of measures on 2^ω by reals is the following.

Definition 4.2. Consider the function $f_b : 2^\omega \to \mathcal{M}$ mapping each $X \in 2^\omega$ to the Bernoulli measure with success probability the real whose binary expansion is X.

Clearly f_b is computable, but it is not injective since dyadic reals have two different binary expansions. In order to mitigate this inconvenience, we consider the following transformation.

Definition 4.3. Given any $\sigma \in 2^{<\omega}$ or $X \in 2^\omega$, let $\hat{\sigma}, \hat{X}$ be the string or real respectively obtained from σ, X by the digit replacement $0 \to 01$, $1 \to 10$. Fore each class $C \subseteq 2^\omega$ let $\hat{C} = \{\hat{X} \mid X \in C\}$.

Since no real in $\hat{C}$ is dyadic, f_b is injective on $\hat{C}$. Moreover, $\hat{C}$ has the same effectivity properties as C; for example it is effectively closed if and only if C is. Hence the hypotheses of Theorem 1.6 are satisfied for $f := f_b$ and $\mathcal{D} := \hat{C}$ for any effectively closed $C \subseteq 2^\omega$.

Lemma 4.4 (Invariance under computable translation)**.** *A class $C \subseteq 2^\omega$ of computable reals is EX-learnable if and only if the class $\hat{C}$ is EX-learnable. The same is true of BC-learnability.*

Proof. Clearly $C, \hat{C}$ are computably isomorphic. Suppose that C is EX or BC learnable by $\mathcal{L}$. Let g be a computable function that maps each index e of computable real X to an index $g(e)$ of the computable real $\hat{X}$. For each $\sigma \in 2^\omega$ define $\mathcal{L}^*(\hat{\sigma}) = g(L(\sigma))$. Moreover, for each τ which is a prefix of a real in $\widehat{2^{<\omega}}$ but whose length is not a multiple of 2, define $\mathcal{L}^*(\tau) = \mathcal{L}^*(\rho)$ where ρ is the largest initial segment of τ which is a multiple of 2. If τ is not a prefix of a real in $\widehat{2^{<\omega}}$ then let $\mathcal{L}^*(\tau) = 0$. In the case of EX learning, since for each real $X \in C$ the values $\mathcal{L}(X \upharpoonright_n)$ converge to an index e of X, it follows that the values $\mathcal{L}^*(\hat{X} \upharpoonright_n)$ converge to the index $g(e)$ of $\hat{X}$, so $\mathcal{L}^*$ is an EX-learner for $\hat{C}$. The case for BC learning as well as the converse are entirely similar. □

Lemma 4.5. *A class of computable reals $C \subseteq 2^\omega$ is EX-learnable if and only if the class $f_b(\hat{C})$ of Bernoulli measures is EX-learnable. The same is true for BC learnability.*

Proof. By Lemma 4.4, C is EX-learnable if and only if $\hat{C}$ is. If we consider $\hat{C}$ as a subset of the effectively closed set $\mathcal{D} = \widehat{2^\omega}$ and apply Theorem 1.6 for f_b we get that $\hat{C}$ is EX-learnable if and only if $f_b(\hat{C})$ is. □

4.2 Proofs of the corollaries of Section 1.2

We conclude the proof of Corollary 1.11 by showing that if an oracle can EX-learn all computable Bernoulli measures then it is high. Note that learnability of an effectively orthogonal class of measures is closed under subsets. Hence it suffices to show that if an oracle A can EX-learn all computable Bernoulli measures with success probabilities that have a binary expansion in $\widehat{2^\omega}$, then it is high. By a direct relativization of Theorem 1.6 and Lemma 4.5, the above working assumption on A implies that the class of computable reals is EX-learnable with oracle A. Then by (Adleman and Blum, 1991) it follows that A is high.

Next, we prove Corollary 1.12, which says that there exist two EX-learnable classes of computable (Bernoulli) measures such that their union is not EX-learnable. Blum and Blum (Blum and Blum, 1975) defined two classes S, T of computable functions which are EX-learnable but their union is not. Consider the classes $\hat{S}, \hat{T}, \hat{S} \cup \hat{T} = (\widehat{S \cup T})$. By Corollary 4.5 the classes $f_b(\hat{S}), f_b(\hat{T})$ are learnable but the class $f_b(\hat{S} \cup \hat{T})$ is not. The result follows by noticing that $f_b(\hat{S}) \cup f_b(\hat{T}) = f_b(\hat{S} \cup \hat{T})$.

Next, we show (3) which says that oracles that are not Δ^0_2 or are not 1-generic, are not low for EX-learning for measures. If $A \not\leq_T \emptyset'$ then by (Fortnow et al., 1994) there exists a class C of computable reals which is EX[A]-learnable but not EX-learnable. If $A \leq_T \emptyset'$ and A is not 1-generic, then by (Kummer and Stephan, 1996) there exists a class C of computable reals which is EX[A]-learnable but not EX-learnable. Then (3) follows by these results, combined with Corollary 4.5.

Finally we prove Corollary 1.10, which says that a learner can EX-learn all computable measures with finitely many queries on an oracle A if and only if $\emptyset'' \leq_T A \oplus \emptyset'$. We need the following lemma.

Lemma 4.6. *If $A \leq_T B'$ then every class of computable measures which is EX-learnable by A with finitely many queries, is also EX-learnable by B.*

Proof. This is entirely similar to the analogous result for EX-learning of classes computable reals from (Fortnow et al., 1994). By $A \leq_T B'$ one can obtain a B-computable function that approximates A. Given an A-computable learner and replacing the oracle with the approximation given via oracle B, the resulting learner will converge along every real on which the original learner converges and uses finitely many queries on A. Moreover, in this case, the limit will agree with the limit with respect to the original A-computable learner. This shows that any class that is EX-learnable via the A-computable learner will also be EX-learned by the new B-computable learner. □

Now given an oracle A, by the jump-inversion theorem, since $\emptyset' \leq_T A \oplus \emptyset'$, there exists some B such that $B' \equiv_T A \oplus \emptyset'$. So $A \leq_T B'$. By Lemma 4.6, if the computable measures are EX-learnable with oracle A and finitely many queries, then they will also be EX-learnable by B. Then by Corollary 1.9 it follows that B is high, so $B' \geq_T \emptyset''$ and $\emptyset'' \leq_T A \oplus \emptyset'$ as required.

Conversely, assume that $\emptyset'' \leq_T A \oplus \emptyset'$. Let (μ_e) be a universal enumeration of all partial computable measure representations with dyadic values and note that by the discussion of Section 2.1 it is sufficient to restrict our attention to these measures, which may not include some measures with non-dyadic values. By Jocksuch (Jockusch, 1972) there exists a function $h \leq_T A$ such that $(\mu_{h(e)})$ is a universal enumeration of all *total* computable measure representations with dyadic values. The fact that uniformly computable families of measures are EX-learnable (originally from (Vitányi and Chater, 2017)) relativizes to any oracle. Since $(\mu_{h(e)})$ contains all computable measure representations with dyadic values, it follows that the class of all computable measures is EX-learnable with oracle A.

5 Conclusion and open questions

We have presented tools which allow to transfer many of the results of the theory of learning of integer functions or reals based on (Gold, 1967), to the theory of learning of probability distributions which was recently introduced in (Vitányi and Chater, 2017) and studied in (Bienvenu et al., 2014, 2018). We demonstrated the usefulness of this result with numerous corollaries that provide parallels between the two learning theories. We also identified some differences; we found that although in the special case of effectively orthogonal classes, the notions of Definitions 1.1 and 1.2 are closed under the subset relation, in general they are not so. Intuitively, if we wish to learn a subclass of a given class of computable measures, the task (compared to learning the original class) becomes easier in one way and harder in another way: it is easier because we only need to consider success of the learner on μ-random reals for a smaller collection of measures μ; it is harder because the learner has fewer choices of indices that are correct answers along each real, since the class of measures at hand is smaller.

We showed that the oracles needed for the EX-learning of the computable measures are exactly the oracles needed for the EX-learning of the computable reals, which are the high oracles. In the classic theory there exists no succinct characterization of the oracles that BC-learn the computable functions. On the other hand, Theorem 1.7 shows that if an oracle can BC-learn the class of computable (continuous) measures, then it can also BC-learn the class of computable functions.

Open problem. If an oracle can BC-learn the class of computable functions, is it necessarily the case that it can learn the class of computable (continuous) measures?

Another issue discussed is the low for EX-learning oracles for learning of measures. We showed that every such oracle is also low for EX-learning in the classical learning theory of reals. We do not know if the converse holds.

References

Lenore M. Adleman and Manuel Blum. Inductive inference and unsolvability. *J. Symb. Log.*, 56(3):891–900, 1991.

Andris Ambainis. Probabilistic inductive inference: A survey. *Theor. Comput. Sci.*, 264(1):155–167, Aug. 2001.

Laurent Bienvenu and Benoit Monin. Von Neumann's Biased Coin Revisited. In *Proceedings of the 27th Annual IEEE/ACM Symposium on Logic in Computer Science*, LICS '12, pages 145–154, Washington, DC, USA, 2012. IEEE Computer Society. ISBN 978-0-7695-4769-5.

Laurent Bienvenu, Benoit Monin and Alexander Shen. Algorithmic identification of probabilities is hard. In *Algorithmic Learning Theory: 25th International Conference, Bled, Slovenia, October 8-10 2014. Proceedings*, ALT 2014, pages 85–95, 2014.

Laurent Bienvenu, Santiago Figueira, Benoit Monin and Alexander Shen. Algorithmic identification of probabilities is hard. *Journal of Computer and System Sciences*, 95:98–108, 2018.

L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28 (2):125–155, 1975.

Adam R. Day and Joseph S. Miller. Randomness for non-computable measures. *Trans. Amer. Math. Soc.*, 365:3575–3591, 2013.

Rodney G. Downey and Denis Hirschfeldt. *Algorithmic Randomness and Complexity*. Springer, 2010.

Lance Fortnow, William Gasarch, Sanjay Jain, Efim Kinber, Martin Kummer, Stuart Kurtz, Mark Pleszkovich, Theodore A. Slaman, Robert Solovay and Frank Stephan. Extremes in the degrees of inferability. *Annals of Pure and Applied Logic*, 66(3):231–276, 1994.

Peter Gács. Uniform test of algorithmic randomness over a general space. *Theor. Comput. Sci.*, 341(1): 91–137, Sept. 2005.

William I. Gasarch and Mark B. Pleszkoch. Learning via queries to an oracle. In R. Rivest, D. Haussler, and M. K. Warmuth, editors, *Proceedings of the Second Annual Workshop on Computational Learning Theory*, pages 214–229. Morgan Kaufmann, San Francisco (CA), 1989.

E. Mark Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

Carl G. Jockusch, Jr. Degrees in which the recursive sets are uniformly recursive. *Canad. J. Math.*, 24: 1092–1099, 1972.

David W. Juedes and Jack H. Lutz. Weak completeness in e_1 and e_2. *Theoretical Computer Science*, 143: 149–158, 1995.

Michael Kearns, Yishay Mansour, Dana Ron, Ronitt Rubinfeld, Robert E. Schapire and Linda Sellie. On the learnability of discrete distributions. In *Proceedings of the Twenty-sixth Annual ACM Symposium on Theory of Computing*, STOC '94, pages 273–282, New York, NY, USA, 1994. ACM.

Martin Kummer and Frank Stephan. On the structure of degrees of inferability. *Journal of Computer and System Sciences*, 52(2):214–238, 1996.

Leonid A. Levin. Uniform tests for randomness. *Dokl. Akad. Nauk SSSR*, 227(1):33–35, 1976.

Leonid A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

Ming Li and Paul M. Vitányi. *An introduction to Kolmogorov complexity and its applications*. Graduate Texts in Computer Science. Springer-Verlag, New York, second edition, 1997. ISBN 0-387-94868-6.

Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

Piergiorgio Odifreddi. *Classical recursion theory. Vol. II*. North-Holland Publishing Co., Amsterdam, 1999.

Daniel Osherson, Michael Stob and Scott Weinstein. *Systems That Learn. First Edition.* MIT Press, Cambridge, MA, 1986.

Steven Pinker. Formal models of language learning. *Cognition*, 7(3):217–283, 1979.

Leonard Pitt. Probabilistic inductive inference. *J. ACM*, 36(2):383–433, Apr. 1989.

Jan Reimann and Theodore A. Slaman. Measures and their random reals. *Trans. Amer. Math. Soc.*, 367(7): 5081–5097, 2015.

Theodore A. Slaman and Robert Solovay. When oracles do not help. In *Proceedings of the Fourth Annual Workshop on Computational Learning Theory*, COLT '91, pages 379–383, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc. ISBN 1-55860-213-5.

Vladimir Vapnik. *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1982. ISBN 0387907335.

Paul M. B. Vitányi and Nick Chater. Identification of probabilities. *Journal of Mathematical Psychology*, 76(Part A):13–24, 2017.

Klaus Weihrauch. Computability on computable metric spaces. *Theoretical Computer Science*, 113(2): 191–210, 1993.