

Samplability Makes Learning Easier

Guy Blanc 


Stanford University, CA, USA

Caleb Koch 

Stanford University, CA, USA

Jane Lange 

MIT, Cambridge, MA, USA

Carmen Strassle 

Stanford University, CA, USA

Li-Yang Tan 

Stanford University, CA, USA

Abstract

The standard definition of PAC learning (Valiant 1984) requires learners to succeed under all distributions – even ones that are intractable to sample from. This stands in contrast to samplable PAC learning (Blum, Furst, Kearns, and Lipton 1993), where learners only have to succeed under samplable distributions. We study this distinction and show that samplable PAC substantially expands the power of efficient learners.

We first construct a concept class that requires exponential sample complexity in standard PAC but is learnable with polynomial sample complexity in samplable PAC. We then lift this statistical separation to the computational setting and obtain a separation relative to a random oracle. Our proofs center around a new complexity primitive, *explicit evasive sets*, that we introduce and study. These are sets for which membership is easy to determine but are extremely hard to sample from.

Our results extend to the online setting to similarly show that its landscape changes when the adversary is assumed to be efficient instead of computationally unbounded.

2012 ACM Subject Classification Theory of computation → Machine learning theory

Keywords and phrases PAC learning, Samplable distributions

Digital Object Identifier 10.4230/LIPIcs.ITCS.2026.20

Related Version *Full Version:* <https://arxiv.org/abs/2512.01276>

Funding *Guy Blanc:* Supported by NSF awards 1942123, 2211237, 2224246, a Sloan Research Fellowship, a Google Research Scholar Award, and a Jane Street Graduate Research Fellowship.

Caleb Koch: Supported by NSF awards 1942123, 2211237, 2224246, a Sloan Research Fellowship, and a Google Research Scholar Award.

Jane Lange: Supported by NSF awards 2006664 and 310818 and an NSF GRFP.

Carmen Strassle: Supported by NSF awards 1942123, 2211237, 2224246, a Sloan Research Fellowship, a Google Research Scholar Award, and an NSF GRFP.

Li-Yang Tan: Supported by NSF awards 1942123, 2211237, 2224246, a Sloan Research Fellowship, and a Google Research Scholar Award.

Acknowledgements We thank the anonymous reviewers for helpful comments and feedback.

1 Introduction

In the PAC model [25], the learner is given labeled data $(\mathbf{x}, f(\mathbf{x}))$ where \mathbf{x} is drawn from a distribution \mathcal{D} . The functions f is promised to belong to a known, often simple, concept class, but no assumptions are made about \mathcal{D} . Notably, the learner is required to succeed



© Guy Blanc, Caleb Koch, Jane Lange, Carmen Strassle, and Li-Yang Tan;
licensed under Creative Commons License CC-BY 4.0

17th Innovations in Theoretical Computer Science Conference (ITCS 2026).

Editor: Shubhangi Saraf; Article No. 20; pp. 20:1–20:12



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

20:2 Samplability Makes Learning Easier

even under distributions that are intractable to sample from. These are distributions \mathcal{D} for which any generator \mathcal{G} such that $\mathcal{G}(\text{Unif}) = \mathcal{D}$ must have superpolynomial circuit size. Consequently, making even a single draw $\mathbf{x} \sim \mathcal{D}$ takes superpolynomial time.

This is arguably an overly stringent requirement. PAC learners are expected to be efficient and yet the distribution, which can be viewed as an adversary in this context, is allowed to be computationally unbounded. This stacks the odds in favor of the adversary and is one reason why efficient PAC learning algorithms have been hard to come by even for simple concept classes. Furthermore, if \mathcal{D} is intractable to sample from then it alone renders the entire learning process inefficient, regardless of the efficiency of the learner. If the PAC model is to capture efficient end-to-end learning, from data collection to hypothesis generation, we may as well consider only samplable distributions. Relatedly, if one believes in the strong Church-Turing thesis, then all distributions occurring in practice are samplable. Quoting [16], “Presumably, real life is not so adversarial that it would solve intractable problems just to give us a hard time.”

Samplable PAC

It is therefore natural to consider a variant of the PAC model, samplable PAC, where the distribution is assumed to be samplable but is otherwise still unknown and arbitrary. This strikes the balance of imposing just enough structure on the distribution to place the learner and adversary on equal footing, while still allowing for enough expressivity to capture the complexities of real-world learning.

The samplable PAC model was first considered by Blum, Furst, Kearns, and Lipton [8]. The focus of their paper was not on the distinction between samplable and standard PAC. Rather, they showed that efficient learning in samplable PAC is tightly connected to the existence of fundamental cryptographic primitives. As their main result, they proved that if one-way functions do not exist, then every concept class is *average-case* learnable in samplable PAC.¹

1.1 This work

We study the distinction between samplable and standard PAC. We are interested in formalizing the extent to which the assumption of samplability – a seemingly mild and reasonable assumption – expands the power of efficient learners. In measuring efficiency, we focus on the two most basic resources in learning: samples and runtime.

Statistical separation

The sample complexity of learning in standard PAC is fairly well-understood, in large part due to an elegant characterization in terms of VC dimension [26, 9] – a foundational result now called “The Fundamental Theorem of PAC Learning” [23]. A VC dimension lower bound is generally viewed as an information-theoretic no-go in terms of efficient learnability: If a learning task cannot be learned with a reasonable number of samples, that trivially implies that it cannot be learned in a reasonable amount of time either.

Our first result is as follows:

¹ This is a further relaxation of samplable PAC where there is an additional distribution, this one over target functions, and the learner is only required to succeed with respect to a random target function drawn from this distribution. We do not consider this variant in our work.

► **Theorem 1.** *There is a concept class with exponential VC dimension – and hence requires exponential sample complexity in standard PAC – but is learnable with polynomial sample complexity, and in fact even in polynomial time, in samplable PAC.*

This shows that VC dimension lower bounds can be overly pessimistic. A learning task with large VC dimension may nevertheless be efficiently learnable if one assumes that these samples are generated according to a reasonable distribution. Put differently, the VC dimension lower bound may be witnessed by an extremely complicated shattering set, one that is intractable to sample from and hence arguably will not arise in the real world – indeed, this intuition is the starting point for our proof of Theorem 1.

Computational separation

The concept class in Theorem 1 has exponential circuit complexity, necessarily so since the class of size- s circuits has VC dimension $\tilde{O}(s)$. It is natural to then ask about samplable vs. standard PAC learning of classes of polynomial-size circuits. Such classes are of interest because functions that arise in practice can be assumed to be efficiently computable, and relatedly, the barriers to their efficient learnability are solely computational in nature and cannot rely on information-theoretic impossibility.

As already observed in [25], if $\text{RP} = \text{NP}$ then every concept class of polynomial-size circuits is efficiently learnable in polynomial time in standard PAC. So short of proving $\text{RP} \neq \text{NP}$, any computational separation will have to either rely on complexity assumptions or be relativized. Implicit in the work of Xiao [27] is a separation relative to a *specific* oracle. We discuss [27]’s result in Section 3, mentioning for now that this is an oracle relative to which one-way functions do *not* exist. This therefore should not be viewed as evidence as to whether such a separation exists in the unrelativized world – and [27] did not claim it as such – since presumably we believe that one-way functions do exist in the unrelativized world.

Our second result gives a computational separation conditioned on two complexity assumptions, a standard one (ironically, the existence of one-way functions) and a new one that we introduce (the existence of explicit “evasive sets”, Conjecture 11):

► **Theorem 2.** *Assume the existence of one-way functions and explicit evasive sets. There is a concept class of polynomial-size circuits that requires superpolynomial time to learn in standard PAC, but is learnable in polynomial time in samplable PAC.*

The definition of an explicit evasive set is rather technical and we defer it to Section 2. For now, we mention that it is a set $H \subseteq \{0, 1\}^n$ that is *explicit* in the sense that membership in H can be easily verified (i.e. the function $x \mapsto \mathbb{1}[x \in H]$ is computable in polynomial time), and yet is *evasive* in the sense that any samplable distribution must “mostly miss” it. The crux of the connection to learning lies in pinning down the appropriate notion of “mostly miss” (Definition 7).

Random oracle separation

Proving the existence of explicit evasive sets unconditionally is likely difficult: Like in the case of one-way functions, doing so will imply $\text{P} \neq \text{NP}$ (we show this in the full version of this paper). We nevertheless prove that they exist relative to a *random* oracle (see the full

version of this paper). Since one-way functions also exist relative to a random oracle, we obtain as a corollary a computational separation of samplable PAC from standard PAC that holds relative to a random oracle, improving on [27]’s separation for a specific oracle:

► **Corollary 3.** *The assumptions of Theorem 2, and hence the separation between samplable and standard PAC, hold relative to a random oracle.*

As is standard in complexity theory, we view Corollary 3 as saying that the speedups offered by samplable PAC over standard PAC hold not just for certain structured instances (that may have been tailored for such a separation), but even for generic, unstructured ones. See [1] for a discussion of this point and the role of random oracle separations more generally.

► **Remark 4 (Evasive sets and uniform generation).** The existence of explicit evasive sets is related to, but differs from, the hardness of *uniform generation* [17]. In uniform generation an algorithm is given the description of a circuit $C : \{0, 1\}^n \rightarrow \{0, 1\}$ and is asked to sample $\text{Unif}(C^{-1}(1))$ (either exactly or approximately). The existence of explicit evasive sets implies the hardness of uniform generation but not vice versa: It could be the case that every polynomial-size circuit C does have a corresponding polynomial-size circuit that generates $\text{Unif}(C^{-1}(1))$, but such generators are just hard to *construct* efficiently.

1.2 Extensions

Separations within samplable PAC

The techniques we use to prove Theorems 1 and 2 extend to give finer-grained separations *within* samplable PAC, showing that learning under distributions with size- s generators can be much easier than under those with size- S generators, even if s is only slightly smaller than S .

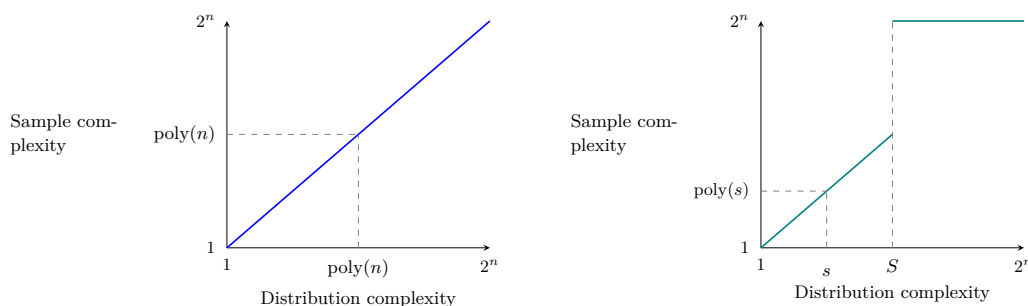
► **Theorem 5.** *For every $s \geq n$ there is a concept class that is learnable with polynomial sample complexity under distributions with size- s generators, and yet requires exponential sample complexity under those with size- S generators for $S \geq \Omega(sn \log s)$.*

While Theorem 1 shows that there are learning tasks whose sample complexity scale smoothly with the complexity of the distribution, Theorem 5 shows that there are ones for which a slight increase in the complexity of the distribution results in a dramatic increase in sample complexity. See Figure 1. We also prove a computational analogue of Theorem 5 in the full version of this paper.

Online learning

Another well-studied model of supervised learning is the online mistake bound model [21]. Learning in this model proceeds in rounds. In each round, the adversary presents the learner with an unlabeled instance $x \in \{0, 1\}^n$. The learner responds with its prediction $y \in \{0, 1\}$ and is then told whether that is correct (i.e. whether $y = f(x)$). The goal of the learner is to minimize its total number of mistakes.

Here again the standard definition allows the adversary to be computationally unbounded – it can take superpolynomial time to produce the test instance in each round. Yet, efficient online learners are expected to be efficient even against such adversaries. For the same reasons as in the PAC setting, it is therefore natural to consider the variant where the adversary is also assumed to be efficient. In the full version of this paper, we show how our



■ **Figure 1** The left and right plots illustrate how the sample complexities of the learning tasks in Theorems 1 and 5 respectively scale with the complexity of the distribution. See their formal versions for the quantitative parameters.

techniques can be extended to the online setting to similarly show how the complexity of online learning – both in terms of mistake bounds and the runtime of learners – can depend on the power of the adversary.

2 Technical overview

Our proofs center around a new notion of *evasive sets* that we introduce and study. These are sets that, as their name suggests, evade all samplable distributions. Let us now make this precise.

2.1 Defining evasive sets

Given a distribution \mathcal{D} and a set H , both over $\{0, 1\}^n$, we say that \mathcal{D} ε -misses H if it places less than ε mass on H :

► **Definition 6** (ε -miss). *A distribution \mathcal{D} ε -misses a set H if $\mathcal{D}(H) := \Pr_{\mathbf{x} \sim \mathcal{D}}[\mathbf{x} \in H] < \varepsilon$. Otherwise, we say that \mathcal{D} ε -hits H .*

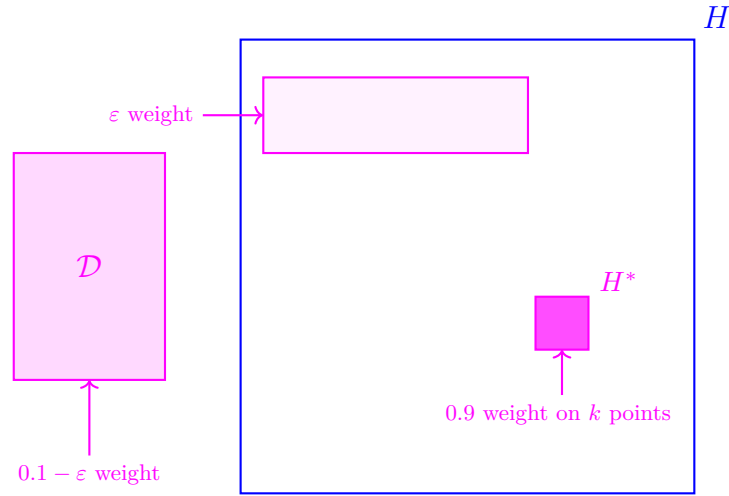
A first attempt at the definition of an evasive set H is one for which all samplable distributions ε -miss it. However, no such H can exist. For any H , a size- $O(n)$ circuit can memorize a specific $x \in H$ and generate the distribution that places all its mass on x . This distribution 1-hits H . More generally, a size- s circuit can memorize $\approx s/n$ many points in H . We therefore modify Definition 6 to exclude the heaviest elements of the distribution:

► **Definition 7** ((ε, k) -miss). *A distribution \mathcal{D} (ε, k) -misses H if there exists a set H^* of size k such that $\mathcal{D}(H \setminus H^*) < \varepsilon$. Otherwise, we say that \mathcal{D} (ε, k) -hits H .*

See Figure 2 for an example that illustrates this definition.

► **Remark 8** (Comparison with TV distance). This notion is stronger than \mathcal{D} having large TV distance from $\text{Unif}(H)$. If \mathcal{D} (ε, k) -misses H then

$$\text{dist}_{\text{TV}}(\mathcal{D}, \text{Unif}(H)) \geq 1 - \varepsilon - \frac{k}{|H|}.$$



■ **Figure 2** The weight distribution of \mathcal{D} is illustrated by the 3 pink rectangles. The sizes of these rectangles depict the number of points and their shades depict the amount of weight. Since \mathcal{D} places $0.9 + \varepsilon$ weight on H , it $(0.9 + \varepsilon)$ -hits H . However, since 0.9 amount of this weight is concentrated on the k points in H^* , it (ε', k) -misses H for any $\varepsilon' > \varepsilon$.

On the other hand, distributions with large TV distance from $\text{Unif}(H)$ can hit H . For example, for any set H and a distribution \mathcal{D} that is uniform on Ck points within H , the TV distance between \mathcal{D} and $\text{Unif}(H)$ is large, $1 - \frac{Ck}{|H|}$, and yet \mathcal{D} $(1 - \frac{1}{C}, k)$ -hits H .

We are now ready to define evasive sets. For brevity, we refer to distributions that have a size- s generator as “size- s distributions”. Samplable distributions are therefore size-poly(n) distributions.

► **Definition 9** ((ε, k) -evades size- s distributions). A set $H \subseteq \{0, 1\}^n$ (ε, k) -evades size- s distributions if every size- s distribution (ε, k) -misses H .

We will be interested in the regime where ε is small and $k \approx s$, capturing the notion that the best thing a size- s distribution can do in terms of approximating $\text{Unif}(H)$ is to simply memorize as many points in H as its size allows and output the uniform distribution over those points. If $|H| \gg s$, as will be the case in our constructions, this is a very bad approximation of $\text{Unif}(H)$.

2.2 A conjecture about explicit evasive sets

A non-explicit construction

Our statistical separation of samplable PAC from standard PAC (Theorem 1) relies on the existence of large evasive sets. Largeness will be useful for our lower bounds against standard PAC whereas the evasiveness will be useful for our upper bounds in samplable PAC. We prove:

► **Lemma 10** (Existence of an evasive set). For any $\delta \geq 2^{-n}$ there is a δ -dense set $H \subseteq \{0, 1\}^n$ that $(\varepsilon, O((s \log s)/\varepsilon))$ -evades all size- s distributions for all $s \geq n$ and $\varepsilon \geq 4\delta$.

We prove Lemma 10 using the probabilistic method. For intuition, consider the special case of flat distributions (those that are uniform over their support). For such distributions \mathcal{D} , if $\text{supp}(\mathcal{D})$ is sufficiently large, we can show that \mathcal{D} is highly unlikely to ε -hit a randomly

chosen H . The failure probability decays exponentially with $|\text{supp}(\mathcal{D})|$, allowing us to union bound over all size- s distributions with sufficiently large supports. This argument no longer works if $\text{supp}(\mathcal{D})$ is small, but in this case we can use the fact that \mathcal{D} trivially $(0, |\text{supp}(\mathcal{D})|)$ -misses every H . The parameters of Lemma 10 are near-optimal, since every H is $(1 - \lambda, \Theta(\lambda s/n))$ -hit by any size- s distribution that is uniform on $\Theta(s/n)$ many memorized points in H .

Explicitness

Our computational separation of samplable PAC from standard PAC (Theorem 2) relies on the existence of a large set that not only evasive, but is furthermore *explicit* in the sense that membership in it (i.e. the function $x \mapsto \mathbb{1}[x \in H]$) is easy to decide:

- **Conjecture 11.** *There is a set $H \subseteq \{0, 1\}^n$ satisfying:*
1. *Explicit: Membership in H is computed by a polynomial-size circuit.*
 2. *Large: H has superpolynomial size.*
 3. *Evasive: H (ε, k) -evades all size- s distributions for all $s = \text{poly}(n)$ and $\varepsilon = 1/\text{poly}(n)$, where $k \leq \text{poly}(s, 1/\varepsilon)$.*

We study Conjecture 11 in detail in the full version of this paper. We show that it implies $P \neq NP$ and that it holds relative to a random oracle. The proof of the latter strengthens that of Lemma 10. The key idea is to show that a samplable distribution \mathcal{D} is highly unlikely to hit a randomly chosen H , even if the circuit generating \mathcal{D} is allowed unit-time membership queries to H .

2.3 The connection to PAC learning

Proof overview of Theorem 1

We first describe how Lemma 10 yields a statistical separation of samplable PAC from standard PAC. For a set $H \subseteq \{0, 1\}^n$ and function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, we write $f_H : \{0, 1\}^n \rightarrow \{0, 1\}$ to denote the following restriction of f to H :

$$f_H(x) = \begin{cases} f(x) & \text{if } x \in H \\ 0 & \text{otherwise.} \end{cases}$$

For a concept class \mathcal{C} , we similarly write \mathcal{C}_H to denote the restriction of \mathcal{C} to H :

$$\mathcal{C}_H = \{f_H : f \in \mathcal{C}\}. \tag{1}$$

Now consider \mathcal{A}_H where \mathcal{A} is the class of all functions $f : \{0, 1\}^n \rightarrow \{0, 1\}$. It is easy to check that H is the largest set shattered by \mathcal{A}_H and hence the VC dimension of \mathcal{A}_H is exactly $|H|$. The sample complexity of learning \mathcal{A}_H in standard PAC is therefore governed by the size of H . In particular, if H has exponential size then learning \mathcal{A}_H in standard PAC requires exponential sample complexity. (This is why we are concerned with evasive sets of large size in Lemma 10.)

On the other hand, we are able to exploit the evasiveness of H to design an efficient algorithm for learning \mathcal{A}_H in samplable PAC:

► **Lemma 12** (Evasiveness implies efficient learners). *Let $H \subseteq \{0, 1\}^n$ be a set that (ε, k) -evades size- s distributions. Then for any concept class \mathcal{C} there is an algorithm for learning \mathcal{C}_H to error $O(\varepsilon)$ under all size- s distributions using $O(k/\varepsilon)$ samples and running in time $O(nk/\varepsilon)$.*

The intuition for Lemma 12 is simple. If H is (ε, k) -evasive, then a learner that memorizes the labels for the k heaviest points in H will only incur ε error. While a learner may not see the k heaviest points, or even know when it has seen them, we show that memorizing $O(k/\varepsilon)$ samples suffices to achieve good accuracy.

Proof overview of Theorem 2

In the proof of Theorem 1, since H is not explicit and \mathcal{A} is the class of all functions, there are no nontrivial upper bounds on circuit complexity of the functions in \mathcal{A}_H . We now describe how we extend the proof strategy so that the separating concept class is a class of polynomial-size circuits. As mentioned, the lower bounds against standard PAC will now be computational in nature and can no longer rely on the information-theoretic arguments that underlie VC dimension lower bounds.

Consider the class \mathcal{F}_H where H is an explicit evasive set given by Conjecture 11 and \mathcal{F} is a pseudorandom function family [13], the existence of which follows from the existence of one-way functions. First note that \mathcal{F}_H now does in fact have polynomial circuit complexity: every function in this class can be computed by a circuit of size $O(S_1 + S_2)$, where S_1 is the circuit complexity of deciding membership in H and S_2 is the circuit complexity of \mathcal{F} .

It is well known that pseudorandom function families are hard to learn: an efficient learner for \mathcal{F} in standard PAC can be used to break \mathcal{F} 's security guarantees [25]. We extend this to show that as long as H is sufficiently large, an efficient learner for \mathcal{F}_H suffices to break \mathcal{F} 's security guarantees. Stated in the contrapositive, \mathcal{F} 's security guarantees implies hardness of learning \mathcal{F}_H in standard PAC. (This is why we are concerned with evasive sets of large size in Conjecture 11.)

3 Related work

Xiao's separation

As mentioned, a separation between samplable PAC and standard PAC for a *specific* oracle is implicit in the work of Xiao:

- **Theorem 13** (Follows from Theorem 1.3 of [27]). *There is an oracle \mathcal{O} such that:*
1. *There is a polynomial-time algorithm A such that $A^{\mathcal{O}}$ learns $\text{SIZE}^{\mathcal{O}}(n^2)$ in samplable PAC.*
 2. *Any algorithm A such that $A^{\mathcal{O}}$ learns $\text{SIZE}^{\mathcal{O}}(n^2)$ in standard PAC must take superpolynomial time.*

An inspection of [27]'s proof shows that \mathcal{O} is an oracle relative to which one-way functions do not exist.² Since we believe that one-way functions exist in the unrelativized world, this

² We sketch the justification here. As stated in Theorem 1.3 of [27], this is an oracle relative to which the learning of all distributions with polynomial-size generators, in the sense of Kearns, Mansour, Ron, Rubinfeld, Schapire, and Sellie [19], is easy. However, as shown in [19], the hardness of this task is implied by the existence of one-way functions. Since this implication relativizes, [27]'s oracle is one relative to which one-way functions do not exist.

therefore does not shed much light on the relationship between samplable and standard PAC in the unrelativized world.³

There are several works giving learning algorithms that work in samplable PAC, and its relaxations, but are not known to work in standard PAC (see e.g. [5, 15, 12, 18] for a few recent examples).

Distribution-specific learning

Given the apparent difficulty of designing efficient algorithms in standard PAC, there has been a large body of work on *distribution-specific* learning. Here the data is promised to be drawn from a specific distribution, e.g. the uniform distribution. The main downside is that this is a stylized assumption that limits the practical relevance of the model: We want our algorithms to succeed for as broad a class of distributions as possible, not just a specific one. In the case of the uniform distribution in particular, it does not capture much of the richness real-world distributions that stem from correlations among features.

Samplable PAC can be viewed as a middle ground that simultaneously corrects for the overly stringent requirements of standard PAC and the overly strong assumptions of distribution-specific PAC.

Lifting uniform-distribution learners

In the same spirit of bridging this gap between standard PAC and distribution-specific PAC, recent works [6, 7] show how uniform-distribution PAC learners can be generically “lifted” to also succeed under various non-uniform yet still structured classes of distributions.

Compared to these works, our work attempts to bridge the gap “from the opposite direction”. While these lifters scale *up* the distribution-specific model, the samplable PAC model scales *down* the distribution-free model (i.e. standard PAC).

Computable PAC and online learning

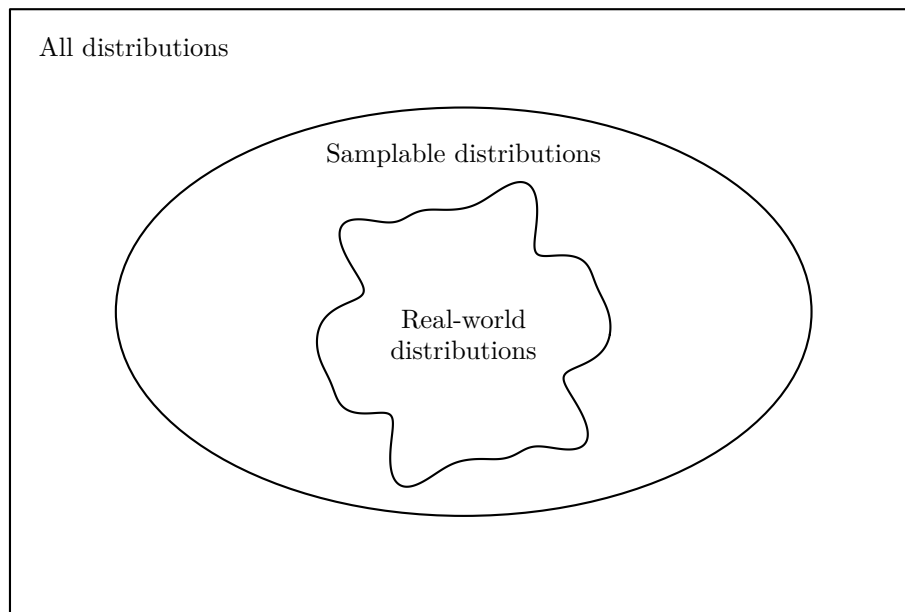
Another recent line of work [2, 3, 24, 10] studies the distinction between standard PAC and a variant known as *computable* PAC where learners are restricted to be computable. Among other results, these works show that there are classes with finite VC dimension that are not learnable in computable PAC. See also [14, 11] for the online analogue.

The focus of our work is on statistical and computational complexity in settings where computability is not an issue, rather than the distinction between computable and uncomputable learners.

Samplable distributions in average-case complexity

Outside of learning theory, samplable distributions are central to the study of average-case complexity [20, 4]. While $P \neq NP$ rules out the possibility of efficient algorithms that solve NP-hard problems on all instances, average-case complexity is concerned with the possibility of efficient algorithms that solve most instances generated by a samplable distribution (i.e. the

³ As in the case of [8], the focus of [27] was not on the distinction between standard and samplable PAC. Rather, the author had proved a result that only held for samplable PAC, and he obtained Theorem 13 on route to showing that an extension of his result to standard PAC will require nonrelativizing techniques. Similarly, see also [15], where a separation is given relative to an oracle for which every problem in PH is easy on average.



■ **Figure 3** An illustration of how samplable distributions relate to real-world distributions.

possibility that NP is hard in the worst case but easy on average). As in samplable PAC, in this context samplable distributions are taken as the formalization of distributions that actually occur in practice.

4 Discussion and future work

Samplability is generally viewed as a baseline requirement for real-world distributions, not a characterization. See Figure 3. Samplable PAC is therefore only a first cut at refining the standard PAC model, which our results show already expands the range of efficient learners.

We see two takeaways from this. First, circling back to a discussion in the introduction, lower bounds in standard PAC – be they statistical or computational – may be overly pessimistic: While certain learning tasks may have hard instances, these hard instances themselves may be hard to find. Second, a next step is to further understand the actual structure of real-world distributions beyond just samplability, and leverage them in the design of learning algorithms. This falls within the overall agenda of going beyond the worst-case analysis of algorithms [22].

More concretely, an open problem is that of characterizing the sample complexity of learning in samplable PAC. Sample complexity in standard PAC is characterized by VC dimension, but our results show that it does not capture sample complexity in samplable PAC – what is the corresponding characterization for samplable PAC? Likewise, is there a variant of Littlestone dimension that characterizes the optimal mistake bound against efficient online adversaries?

References

- 1 Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *Theory of Computing*, 10(6):133–166, 2014. doi:10.4086/TOC.2014.V010A006.

- 2 Sushant Agarwal, Nivasini Ananthakrishnan, Shai Ben-David, Tosca Lechner, and Ruth Urner. On learnability with computable learners. In *Proceedings of the 31st International Conference on Algorithmic Learning Theory (ALT)*, pages 48–60, 2020. URL: <http://proceedings.mlr.press/v117/agarwal20b.html>.
- 3 Sushant Agarwal, Nivasini Ananthakrishnan, Shai Ben-David, Tosca Lechner, and Ruth Urner. Open problem: Are all VC-classes CPAC learnable? In *Proceedings of the 34th Annual Conference on Learning Theory (COLT)*, pages 4636–4641, 2021. URL: <http://proceedings.mlr.press/v134/open-problem-agarwal21b.html>.
- 4 Shai Ben-David, Benny Chor, and Oded Goldreich. On the theory of average case complexity. In *Proceedings of the 21st Annual Symposium on Theory of Computing (STOC)*, pages 204–216, 1989.
- 5 Eric Binnendyk, Marco Carmosino, Antonina Kolokolova, Ramyaa Ramyaa, and Manuel Sabin. Learning with distributional inverters. In *Proceedings of the International Conference on Algorithmic Learning Theory (ALT)*, pages 90–106, 2022.
- 6 Guy Blanc, Jane Lange, Ali Malik, and Li-Yang Tan. Lifting uniform learners via distributional decomposition. In *Proceedings of the 55th Annual Symposium on Theory of Computing (STOC)*, pages 1755–1767, 2023. doi:10.1145/3564246.3585212.
- 7 Guy Blanc, Jane Lange, Carmen Strassle, and Li-Yang Tan. A distributional-lifting theorem for PAC learning. In *Proceedings of the 38th Annual Conference on Learning Theory (COLT)*, volume 291, pages 375–379, 2025. URL: <https://proceedings.mlr.press/v291/blanc25a.html>.
- 8 Avrim Blum, Merrick Furst, Michael Kearns, and Richard J Lipton. Cryptographic primitives based on hard learning problems. In *Annual International Cryptology Conference (CRYPTO)*, pages 278–291, 1993. doi:10.1007/3-540-48329-2_24.
- 9 Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM (JACM)*, 36(4):929–965, 1989. doi:10.1145/76359.76371.
- 10 Valentino Delle Rose, Alexander Kozachinskiy, Cristóbal Rojas, and Tomasz Steifer. Find a witness or shatter: the landscape of computable PAC learning. In *Proceedings of the 36th Annual Conference on Learning Theory (COLT)*, pages 511–524, 2023. URL: <https://proceedings.mlr.press/v195/delle-rose23a.html>.
- 11 Valentino Delle Rose, Alexander Kozachinskiy, and Tomasz Steifer. Effective Littlestone dimension. *Proceedings of the 36th International Conference on Algorithmic Learning Theory (ALT)*, 272:1–13, 2025.
- 12 Halley Goldberg and Valentine Kabanets. Improved learning from Kolmogorov complexity. In *Proceedings of the 38th Computational Complexity Conference (CCC)*, pages 12–1, 2023.
- 13 Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *Journal of the ACM (JACM)*, 33(4):792–807, 1986. doi:10.1145/6490.6503.
- 14 Niki Hasrati and Shai Ben-David. On computable online learning. In *Proceedings of the 34th International Conference on Algorithmic Learning Theory (ALT)*, pages 707–725, 2023. URL: <https://proceedings.mlr.press/v201/hasrati23a.html>.
- 15 Shuichi Hirahara and Mikito Nanashima. On worst-case learning in relativized heuristica. In *Proceedings of the 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 751–758, 2022.
- 16 Russell Impagliazzo. A personal view of average-case complexity. In *Proceedings of 10th Annual Structure in Complexity Theory Conference*, pages 134–147, 1995. doi:10.1109/SCT.1995.514853.
- 17 Mark R Jerrum, Leslie G Valiant, and Vijay V Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theoretical computer science*, 43:169–188, 1986. doi:10.1016/0304-3975(86)90174-X.
- 18 Ari Karchmer. Distributional PAC-learning from Nisan’s natural proofs. In *Proceedings of the 15th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 68–1, 2024.

20:12 Samplability Makes Learning Easier

- 19 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*, pages 273–282, 1994. doi:10.1145/195058.195155.
- 20 Leonid Levin. Average case complete problems. *SIAM Journal on Computing*, 15(1):285–286, 1986. doi:10.1137/0215020.
- 21 Nick Littlestone. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine learning*, 2:285–318, 1988. doi:10.1007/BF00116827.
- 22 Tim Roughgarden. *Beyond the worst-case analysis of algorithms*. Cambridge University Press, 2021.
- 23 Shai Shalev-Shwartz and Shai Ben-David. *Understanding Machine Learning - From Theory to Algorithms*. Cambridge University Press, 2014. URL: <http://www.cambridge.org/de/academic/subjects/computer-science/pattern-recognition-and-machine-learning/understanding-machine-learning-theory-algorithms>.
- 24 Tom Sterkenburg. On characterizations of learnability with computable learners. In *Proceedings of the 35th Annual Conference on Learning Theory (COLT)*, pages 3365–3379, 2022. URL: <https://proceedings.mlr.press/v178/sterkenburg22a.html>.
- 25 Leslie Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984. doi:10.1145/1968.1972.
- 26 Vladimir Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971. doi:10.1137/1116025.
- 27 David Xiao. Learning to create is as hard as learning to appreciate. In *Proceedings of the 23rd Annual Conference on Learning Theory (COLT)*, pages 516–528, 2010. URL: <http://colt2010.haifa.il.ibm.com/papers/COLT2010proceedings.pdf#page=524>.