Mathematical Logic and Probability

Wesley Calvert

Contents

Preface

In the late 19th and early 20th centuries, logic and probability were frequently treated as closely related disciplines. Each has, in an important sense, gone its own way, so that neither, in its modern form, is in any proper sense a systematization of the "Laws of Thought," as Boole called them.

However, the last four decades have seen a remarkable rapproachment. On the most obvious level, the various probability logics have developed as formal systems of reasoning in the modern sense of logic.

At a deeper level, though, attempts have been made to formulate logics in which model theory of random variables, stochastic processes, and randomized structures can be explored from the perspective of model theory. Continuous first-order logic as a context for stability theory on metric structures is perhaps the most conspicuous example, but others exist.

At the same time, algorithmic randomness in its various forms has come to play a core role in computability theory, while probabilistic computation of various kinds (randomized computation, interactive proofs, and others) has come to dominate major parts of computational complexity. The older recursion-theoretic program of machine learning, initiated by Gold in the 1960s, has become much more important thanks to Valiant's reformulation in probabilistic terms to allow for reasonable errors.

The model theory of random objects, Fraïssé limits, and pseudofinite structures, each of which embodies some important aspect of 0-1 laws, has been important for longer, but advances in stability, simplicity, and the transition from finite to infinite model theory have enriched this subject.

In set theory, too, the study of dynamics that respect probability measures has played a central role in the study of equivalence relations. Probability is frequently at the center of modern descriptive set theory.

Nor have these developments been independent. The PAC learning theory of Valiant is inextricably linked to the model theory of NIP theories. The dynamics of computable Polish spaces have become an important emerging area in computability. Randomized computation is the natural computation on metric structures. Notions of random structures have become intertwined with algorithmic randomness, and are naturally described in continuous first order logic.

Many of these developments have been adequately treated in isolation by various books. Probability logic has been discussed at length from various perspectives in [**10, 241, 244, 397, 422**]. Bayesian networks are well-covered, for instance, in [**221, 400, 401**], and a monograph on adapted distributions also exists [**181**]. Randomized computation has a detailed treatment in [**31**]. Algorithmic randomness is the subject of three relatively recent books, [**159, 390, 196**]. Zero-one laws are treated at length in [**166, 232**], and other places, and [**268**] includes an extended treatment of Fraïssé limits. Random graphs are extensively covered in [**75, 126, 341**]. The definitive reference on PAC learning is [**300**]. In the field of set-theoretic dynamics, there have been several treatments at several levels of detail, among which [**56, 262, 301, 306**] merit special mention. There is no shortage of book-length treatments of subjects within the range of this book.

However, a reader in a well-stocked library might well pass all these separate books without knowing that they had anything substantive in common. Indeed, one could read most of them in detail — in addition to the long papers that give strong expositions on many related subjects (the seminal paper [**62**] on continuous first-order logic comes to mind) — without finding a commonality.

It is true that [**240**] describes connections between probability logics and Bayesian networks. However, it is silent on the rest of these issues.

The present book, then, attempts to take a unified — or, at least, unifying — approach to this subject. The expanding literature in each of these fields has seen more interaction between them, so that a model theorist might well want to know more about the frontier of probabilistic work in set theory, or a computability theorist more about the relevant work in model theory.

We focus here on *mathematical* logic and probability. Probability logic and its relatives seem frequently to arise as works of *philosophical* logic, and this has implications for the questions that are asked about it. Frequently it is seen in connection with the theory of rational decision, as in [**244**]. Mathematical logic, by contrast, asks about computability and undecidability; about theories and their models; about reducibilities and regularity of sets. Alternate logics are of interest to mathematical logic inasmuch as they provide the necessary infrastructure for carrying out this program in interesting settings. Applications of logic to artificial intelligence and other modeling contexts are important, but they arise as applications of the theory, not as its defining elements.

Chapter 1 begins to lay out the central thesis of the book: that all the other chapters have something to say to one another. This is done by identifying several important cross-cutting themes that come up in several of the other chapters.

In the next chapter, we begin the technical section of the book by describing the various logics useful for probability. Continuous first-order logic has a central role, not least because it generalizes many others. Probability logic is extensively studied, and is explored here as well, as are some other approaches.

In a third chapter, we will consider the theory of algorithmic randomness, with special attention to normal numbers, Martin-Löf randomness, and their relation to computation. This treatment will not be complete, of course — the subject is well-covered elsewhere. Rather, the focus will be on those aspects of algorithmic randomness that interact with other areas of advance in the logic and probability community.

The chapter on randomized computation involves the leap of reasoning that computability and complexity still have something to say to one another. Recent work on generic and coarse computability, as well as that on derandomization, descriptive complexity, and continuous first-order logic support this hypothesis.

The following two chapters will take up the various approaches to random structures. The investigation of random structures seems to have arisen historically from the study of random graphs, which invited generalization to 0-1 laws, and which connected with the earlier beginnings of Fraïssé limits. More recent approaches consider the "random" structure as a single structure that somehow embodies the possible variation — graphons, Keisler randomizations, invariant random subgroups, and the like. Others use algorithmic randomness to define the structure.

In taking up the problem of learning theory, there is a fair viewpoint from which learning after the tradition of Gold, probably approximately correct (PAC) learning after the tradition of Valiant, and the model theory of NIP structures are wildly different fields. The chapter devoted to these topics takes the opposite view. Valiant's definition is a natural extension of Gold's framework, and the theory of Vapnik-Chervonenkis dimension governs both PAC learning and NIP theories.

The final chapter surveys the general area of dynamics. An introduction to orbit equivalence relations and Borel cardinality is given, and several topics on the relation of measure to equivalence relations are considered, including the implications of ergodicity and Hjorth's notion of turbulence. Recent model-theoretic approaches to Szemeredi Regularity and Furstenberg Correspondence belong here, too, as does the characterization of 1-randomness by the Ergodic Theorem and the emerging theory of computable Polish spaces.

Of course, some limits must be set on the content of such a book. For instance, a new line of thought has arisen in recent years over categorical treatments of probability [**201, 202, 398**]. In view of traditional [**353**] and recent [**120, 250, 249**] work on connections between category theory and logic, this work is certainly interesting and relevant, but it is hard, at this stage of the theory, to explain its relationship to the other work.

The book is to be formally self-contained, but realistically anticipates a reader who has completed a first course in logic at the graduate or upper undergraduate level. Such a reader will, after reading the book, be prepared to understand the frontier of the research literature in probability-related areas of computability, model theory, set theory, and logical aspects of artificial intelligence. There is an important place in the world for a reader equipped in this way: A major part of logic in the coming years will involve connections between these fields, and those who understand something of all of them will be well-poised to contribute.

CHAPTER 7

Learning and Independence

7.1. The Learning Problem

The basic problem of machine learning is to create an algorithm that can improve its own performance on some task. This is desirable in cases where the programmer cannot predict all of the situations in which the algorithm should be able to perform well, or in which the algorithm is looking for things that the programmer does not understand at the outset.

The second situation is becoming quite common. It happens — and has happened to this author — that scientists collect an enormous body of data, and then ask a convenient mathematician, statisitician, or computer scientist, "What is important here?"

At least for a broad range of applications, we can model the learning task by the idea of learning sets. There is a large set, called the *instance space*, from which individual inputs will be drawn. The *target* is a particular subset of the instance space.

The target should, on the whole, be unknown to the programmer. Possibly we may allow some guarantees about the form of the target, but the set itself should be unknown. Then the algorithm is given some examples, called the *training set.* In some models, it is given only elements of the target. In others, it is given both elements and non-elements, labeled. These two classes are forms of *supervised learning.* In other models still, the learner receives some feedback from changes that it makes, so that it can see when it has improved its hypothesis, in what is known as *reinforcement learning.* Depending on the particular model, there may be restrictions on how the training set is selected.

The ultimate goal of the algorithm is to identify the target set from the training set. Of course, without some restriction this is impossible. Certainly if the training set is finite and the instance space (as usual) is not, then many different sets agree with it, and are equally possibilities for the target set. We can address this by allowing the training set to be infinite (usually given one element at a time), but then the difficulty is one of convergence: If, at any finite point, the algorithm halts, it has seen only a finite training set.

Consequently, most models of learning either weaken the convergence criterion for the algorithm or weaken the sense in which the algorithm must identify the target set correctly. Sometimes both weakenings are made simultaneously.

In the case of a convergence criterion weaker than halting with output specifically identifying a set, we usually ask that the algorithm output a sequence of identifiers which converge, in some sense, to a description of the target. In this sense, the convergence is the sense in which the algorithm is improving its performance. If we do require the algorithm to halt, then it has still improved its performance if the set it identifies is somehow closer to the target than any identification it could reasonably have made before the computation.

This type of learning problem, in which a set is to be learned, is called a *classification* problem. Many classical learning tasks are not transparently the learning of sets, but could be understood this way. For instance, one sometimes asks a robot to learn the layout of a maze. While there may be many more efficient representations, this is equivalent to learning a set. Indeed, we could understand the instance space to be the set of pairs of locations in the maze, and the target set is the set of pairs immediately accessible from one another, using only unobstructed paths.

This brings up another important point about the machine learning literature. One can productively think about what formal definitions to use in order to model learning and in order to prove whether a particular learning task is possible or not. On the other hand, there is a large literature that takes the possibility of learning as given, and then approaches how to do it most efficiently. Those familiar with the division of computability and complexity on one hand, and algorithms on the other will recognize a similar difference here.

While there is probably important work to be done in logic in this second area of how to make an algorithm that will learn efficiently, the bulk of the literature to date seems to be in the first direction, and that is where the present chapter will place its focus.

7.2. Learning c.e. Sets to Equality and Almost Equality

In 1967, E. Mark Gold attempted a mathematical model of the problem of child learning of grammar. He took as a starting point the (disputable) observation that children are seldom informed of their grammatical errors, and any corrections they do receive seem to have limited impact on their language learning. Intuitively, it should be very difficult — even impossible — to learn a language in this way.

Of course, the idea that a language has static underlying rules of grammar with a well-defined set of "correct" sentences might well be considered more problematic than Gold gave it credit for. In fairness, this model arises in a broader school of thought in which Chomsky [**123**], for instance, likened a grammar to a scientific theory, where the set of grammatical sentences should represent a theoretical prediction of how the natural language will work. Later work in the same theoretical framework extended it to account more completely and explicitly for the contingency of language on a particular "speech community," for the distinction between the observed external utterances used (the E-language) and the language as it is internally understood by its users (the I-language), and many other features [**124**].

He asked, "Is there enough information in a text, even one of unlimited length, to allow the identification of a context-free language?" That is, if we are presented only with many instances of grammatically correct sentences, can we learn the true rules of grammar from this? His paper [**222**] proposes several models for learning, including models that would distinguish the case of having only positive examples from those that would involve some kind of correction of errors.

To model the situation of language learning, Gold identified a language with its set of grammatically correct sentences. Up to routine issues of representation, this set could be understood as a set of natural numbers. It is a routine assumption, at least since Chomsky, that the grammar should be computably enumerable.

The problem, then, is, given an unknown computably enumerable set, to determine its index. The exact conditions under which an unknown set is "given" and the exact standard for "determining" its index vary from model to model. In all of these models, an algorithm, called the *learner*, is presented with a surjective enumeration $\eta : \mathbb{N} \to S$. The learner computes a function $L : \mathbb{N}^{<\omega} \to \mathbb{N}$. We will give various definitions of success for the learner, expressed as conditions on the convergence of the sequence $(L(\eta \restriction_t) : t \in \mathbb{N})$ and on the form of η. Harizanov's survey [**245**] contains a useful collection of models and recent results, and [**278**] is a standard reference.

Notice, also, that if we make a "promise" about S, that may change things.

EXAMPLE 7.2.1. Suppose that the learner knows at the outset that S is $\{n\}$ for some $n \in \mathbb{N}$. In this case, as soon as we see an element enumerated into S, we can be confident that we know S.

The "promise" is given as a "concept class." That is, we have a set $\mathcal{C} \subseteq \mathcal{P}(\mathbb{N})$ such that the learner knows at the outset that S must be an element of $\mathcal{C}$.

Gold proposed several variant models of learning, which vary, at least, in the conditions of training, learner power, and conditions of correctness. In the training, we can consider the following conditions:

Arbitrary Text: η may be an arbitrary surjective function
Computable Text: η may be any computable function
"Nice" Computable Text: η may be any computable function with property P, where P ranges over any complexity or regularity properties one might care to specify.

We also consider the power of the learner. If the learner is restricted to a computable function L, then Gold called it "effective." He also considered the broader class of unrestricted ("ineffective") learners.

Considering conditions of correctness, Gold identified the following variations:

Tester: $\varphi_{g_t} = \chi_S$
Generator: $S = W_{g_t}$.

Even at this level of generality, some results are possible on conditions of learning.

DEFINITION 7.2.2 (Gold). The *Collapsing Uncertainty Condition* is satisfied if and only if the following holds: Let $\mathcal{C}_t$ be the set of all elements of $\mathcal{C}$ which are consistent with $\eta \restriction_t$. Then $\mathcal{C}_t$ must be a decreasing sequence, with a singleton limit.

THEOREM 7.2.3 (Gold). *For ineffective identifiability in the limit, the collapsing uncertainty condition is sufficient.*

PROOF. Recall that, since all target sets under consideration are computably enumerable, we can enumerate $\mathcal{C} = \{c_n : n \in \mathbb{N}\}$. Then we define $L(\eta \restriction_t)$ to be the least element of $\mathcal{C}_t$. Note that if the enumeration of $\mathcal{C}$ is computable, then L is computable.

To show that this works, note that $S \in \mathcal{C}$. Suppose that $S = c_n$. Then there are at most $n - 1$ different objects before this position, and the collapsing uncertainty condition implies that there is some finite time at which each of these is eliminated. □

There have been many sets of notation proposed for the learning theories arising out of Gold's work. There seems, in the more recent literature, to be a convergence on the notation found in [**278**], which draws on some of the earlier notation traditions, and which is used here. The first notion we will consider is called *explanatory learning*. The idea is that the learner's sequence of guesses should converge to a single index, representing a single algorithm to generate S.

Definition 7.2.4. A learner $L : \mathbb{N}^{<\mathbb{N}} \to \mathbb{N}$ *Ex-learns a c.e. set S from text* (we sometimes write that L *TxtEx learns* S) if and only if for any enumeration $\eta : \mathbb{N} \twoheadrightarrow S$, the sequence $(L(\eta \restriction_t) : t \in \mathbb{N})$ converges to an index for S.

Again, context matters. It is not hard to learn a set in isolation, with no credible distractors.

Proposition 7.2.5. *Every c.e. set S is EX-learnable from text.*

Proof. Let e be the index for S. Then define L to be the constant function with value e. □

To provide the necessary context, we generally speak of a concept class, or a class of c.e. sets.

Definition 7.2.6. A learner L learns a class of c.e. sets $\mathcal{C}$ if and only if L learns every element of $\mathcal{C}$.

The following result, due to Lenore and Manuel Blum [**73**], demonstrates a sense in which TxtEx learning is really as hard as it intuitively feels like it should be. It is interesting, in our present discussion, to note that the Blums identified the "philosophical basis" of their work as arising from Popper's *Logic of Scientific Discovery* [**410**], which contained some ideas on algorithmic randomness that we discussed in Section 3.1.1.

Theorem 7.2.7 (L. Blum and M. Blum). *If L can TxtEx learn a c.e. set S, then there is a finite sequence σ of elements of S, called a* locking sequence*, such that if the learner outputs e after seeing σ, then $S = W_e$ and $L(\tau) = e$ for any $\tau \supseteq \sigma$.*

Proof. Suppose that L converges on all sequences, and suppose that there is no σ as described in the statement of the theorem. Then for each finite $\sigma \subseteq S$ with $W_{L(\sigma)} = S$, there is some $\tau \subseteq S$ extending σ such that $L(\tau) \neq L(\sigma)$. We now produce some enumeration η of S on which L fails to converge to an index for S.

Let $\eta_0 : \mathbb{N} \twoheadrightarrow S$, and set $\sigma_0 = \emptyset$. At stage $t+1$, we check whether $W_{L(\sigma_t)} = S$. If so, then we find some $\tau \supseteq \sigma_t$ within S such that $L(\tau) \neq L(\sigma_t)$. Then we set $\sigma_{t+1} = \tau\eta(t)$. On the other hand, if $W_{L(\sigma_t)} \neq S$, no attention is needed, so we set $\sigma_{t+1} = \sigma_t\eta(t)$. Let $\xi = \bigcup_{t\in\mathbb{N}} \sigma_t$. Note that $\xi : \mathbb{N} \twoheadrightarrow S$. However, for each t where $L(\xi \restriction_t) = S$, we have some $\hat{t} > t$ where $L(\xi \restriction_{\hat{t}}) \neq S$. □

The theorem on locking sequences is frequently useful in negative results on TxtEx learning. The following result lifts a similar idea to learning a class of sets.

Proposition 7.2.8 (Angluin 1980). *Let $\mathcal{C}$ be a class of c.e. sets. Then $\mathcal{C}$ is EX-learnable from text iff for every $S \in \mathcal{C}$ there is a finite $D_S \subseteq S$ such that for no $U \in \mathcal{C}$ do we have $D \subseteq U \subseteq S$*

PROOF. Suppose that L TxtEx-learns $\mathcal{C}$. Then for each $S \in \mathcal{C}$, choose a locking sequence D_S. Now if $D_S \subseteq U \subseteq S$, pick an enumeration η of U such that $ran\,(\eta \restriction_t) = D_S$ for some t, and note that the sequence $\{L\,(\eta \restriction_t) : t \in \mathbb{N}\}$ converges to an index for S, since D_S is a locking sequence. Consequently, L does not converge to an index for U on every enumeration of U. □

From this result, we can prove the following negative result.

EXAMPLE 7.2.9. Let $\mathcal{C}$ be the class consisting of $\mathbb{N}$ and all of its finite subsets. Then $\mathcal{C}$ is not Ex-learnable from text.

We might (and Gold did) consider an even more strict standard of convergence. Ex-learnability allows an arbitrary finite number of incorrect hypotheses before the learner settles down on the correct one.

DEFINITION 7.2.10. We say that L TxtFin-learns $\mathcal{C}$ if and only if for any enumeration $\eta : \mathbb{N} \twoheadrightarrow S$, the computation $L(\eta \restriction_t)$ returns a blank for some initial segment $t < \hat{t}$, and for $t \geq \hat{t}$, we have $L(\eta \restriction_t) = e$, where $S = W_e$.

This notion does occur in the literature, for instance in [**107**], but we will not explore its properties in detail here. It will become important in that the convergence criterion of PAC learning in Section 7.3 matches this one better than any of the others of this section.

TxtEx learning is restrictive in many ways. We first consider the "Text" restriction — that is, the restriction that we receive only positive information. This was a major dividing line for Gold, who believed that "most children are rarely informed when they make grammatical errors, and those that are informed take little heed." He found this observation discrepant in light of the restrictive nature of text learning, and left open the linguistic problem of how something as complex as natural language appears to be could be learned in this fashion. Gold's only cited source for this observation was a preprint of [**376**], whose published version certainly had detailed descriptions of studies including subtle correction of children's grammar — for instance, when an adult rephrases the child's sentence into a nearby correct sentence — and children's response to it. In any case, Gold did recognize the importance of receiving both positive and negative information.

DEFINITION 7.2.11. We say that $L : \left(\mathbb{N}^{<\mathbb{N}}\right)^2 \to \mathbb{N}$ Ex-learns S from an informant (that is, L *InfTxtEx-learns* S) iff for any $\eta_+ : \mathbb{N} \twoheadrightarrow S$ and $\eta_- : \mathbb{N} \twoheadrightarrow S^c$ then sequence $L(\eta \restriction_t, \xi \restriction_t)$ is eventually constant on some index for S.

As we might expect, InfTxtEx learnability is a somewhat broader concept.

EXAMPLE 7.2.12. Let $\mathcal{C}$ be the class consisting of $\mathbb{N}$ and all of its finite subsets. Then $\mathcal{C}$ is Ex-learnable from an informant. Indeed, a learner can start outputting an index for $\mathbb{N}$. If $\mathbb{N}$ is not the target, then at some stage, L will see a nonmember enumerated by η_-. At that point, L will switch to the assumption that the target is the finite set of members already enumerated by η_+, and output indices accordingly.

Another respect in which TxtEx learning is restrictive is that it requires that the learner produce *exactly* the target. That is, it most converge to an index for S exactly. This is generally more precision than is needed in applications. Indeed, we generally expect and accept some small amount of error. Of course, in a later section, the amount of error tolerated will be expressed in terms of probability, but the context of learning subsets of $\mathbb{N}$ will initially hold us back from a full use of probability. However, we can still approximate it by asking that the target set be learned up to finite difference.

In apparently independent papers in 1982, Osherson and Weinstein [**394**] and Case and Lynes [**108**] identified variant learning standards that allowed for errors on finitely many instances. Case and Lynes also included variants allowing a fixed finite bound on the error instances.

Definition 7.2.13. Let $a \in \mathbb{N} \cup \{\omega\}$, and let $S \subseteq \mathbb{N}$.

(1) If $A, B \subseteq \mathbb{N}$ and $a \in \mathbb{N}$, then we say that $A =^a B$ if and only if $|A \triangle B| \leq a$.
(2) If $A, B \subseteq \mathbb{N}$, then we say that $A =^* B$ if and only if $A \triangle B$ is finite.
(3) We say that L TxtEx$^\omega$-learns S if and only if for any enumeration $\eta : \mathbb{N} \twoheadrightarrow S$, the sequence $(L(\eta \upharpoonright_t) : t \in \mathbb{N})$ converges to an index for a set $\tilde{S} =^* S$.

Learning a class of sets is defined, as before, by learning every set in the class. Case and Lynes proved that for every $a \in \mathbb{N} \cup \{\omega\}$ we have a strict implication: TxtEx^a learnability implies TxtEx^b learnability for every $b < a$. A modern proof of this can be found in [**278**].

The importance of these classes from the point of view of probability is that the finite sets form a hard core of "small" sets. They have natural density zero, and any function $\mu : \mathcal{P}(\mathbb{N}) \to [0, 1]$ which is finitely additive and takes value zero on singletons must make all of these zero.

Building out from this core of small sets, Royer [**426**] described learning up to small density of errors.

Definition 7.2.14 ([**426**]). Let $r \in [0, 1]$. A learner L TxtAEx^r-learns S if and only if for any enumeration $\eta : \mathbb{M} \twoheadrightarrow S$, the sequence $(L(\eta \upharpoonright_t) : t \in \mathbb{N})$ converges to an index for a set $\tilde{S}$ such that $\tilde{S} \triangle S$ has density at most r.

Royer shows that for any $r_1 < r_2$, we have TxtAEx$^{r_1} \subsetneq$ TxtAeX^{r_2}, and that TxtEx$^\omega \subsetneq$TxtAEx^0.

We can relax the convergence requirements in a meaningful way, too. Every computably enumerable set has infinitely many indices. We might refrain from requiring the learner to converge on an index, and merely require that the learner converge on a *set*. Such a learner is said to be *behaviorally correct*, in that while the index it outputs (the explanation) may vary, it is consistent and correct in its identification of the extension of the hypothesis set. In particular, we state the following definition.

Definition 7.2.15. Let $\mathcal{C}$ be a family of computably enumerable sets.

(1) We say that a learner L TxtBc-learns $S \in \mathcal{C}$ if and only if for any enumeration $\eta : \mathbb{N} \twoheadrightarrow S$, there is some N such that for $t \geq n$ we have $W_{L(\eta \upharpoonright_t)} = S$.
(2) We say that a learner L TxtBc-learns $\mathcal{C}$ if and only if L TxtBc-learns every $S \in \mathcal{C}$.
(3) Let $a \in \mathbb{N} \cup \{\omega\}$. Then we say that a learner L TxtBc^a-learns $S \in \mathcal{C}$ if and only if for any enumeration $\eta : \mathbb{N} \twoheadrightarrow S$, there is some N such that for $t \geq n$ we have $W_{L(\eta \upharpoonright_t)} =^a S$.
(4) We say that a learner L TxtBc^a-learns $\mathcal{C}$ if and only if L TxtBc^a-learns every $S \in \mathcal{C}$.

As before, TxtBc^a learnability strictly implies TxtBc^b learnability for all $b < a$. Moreover, the relaxation of the convergence requirements from Ex to Bc learning is incomparable with the relaxation of the correctness requirements from Ex to Ex^ω learning. The following result is well-known, and an outline of the proof can be found in [**278**].

Theorem 7.2.16. *Let TxtBc be the set of classes which are TxtBc learnable, and* TxtEx^ω *be the set of classes which are* TxtEx^ω *learnable.*

(1) $\mathit{TxtBc} \nsubseteq \mathit{TxtEx}^\omega$
(2) $\mathit{TxtEx}^\omega \nsubseteq \mathit{TxtBc}$.

More recently, Karn has developed density criteria for behaviorally correct learning, and has shown that, again, the classes learnable up to given density form a strict hierarchy, as in Royer's results [**291**].

While the work on these notions of learning that is most relevant for the topic at hand is mostly older, it should not be supposed that this topic has ceased to be of interest. Indeed, there has been more recent work in learning countable algebraic structures [**470, 246**], in classifying the Turing degree of learnability [**67**], and in using these learning models to explore the effective aspects of Kleene's Recursion Theorem [**107**], to cite only a few examples.

It would be routine to define, for each $r \in [0, 1]$, learnability notions TxtEx^r and TxtBc^r, where the hypothesis is required to be correct except possibly on a set of natural density r. These would correspond, in the environment of learning, to the work in Section 4.5 on sets generically or coarsely computable at density r. Probably these notions are more permissive than the finite versions, but no work on these notions is known to the present author.

7.3. Probably Approximately Correct Learning and Vapnik–Chervonenkis Dimension

7.3.1. Allowing Randomization and Errors. In a sense, the models considered so far lack some essential features of realism. So far, we have always assumed that the algorithm can look at unbounded training sets, and need only converge (in some sense) in the limit, which may, of course, be quite far out. This is clearly not the situation for most of modern machine learning. More frequently, the algorithm trains on some finite set of data. Since the set of potential targets consistent with that finite set of data is still quite large, we need to weaken the success criterion to compensate.

In doing so, we capture two critical features of real situations. The first is familiar. We allow the hypothesis generated by the learning algorithm to differ from the true target by a set of positive probability (which we call the "A" probability). In most buinsess, industrial, and scientific climates, we can still be quite comfortable with this, as long as the A probability is, or can be made, sufficiently small.

The second new feature reflects an inherent issue of partial data. In earlier models, if the first few bits of training data we got weren't helpful, there was no problem; helpful data would emerge eventually. In a situation with a finite (generally small) amount of training data, we will, with positive probability (which we will call the "P" probability) encounter a set of training data that is all unhelpful. Again, this is often practically acceptable, as long as the P probability is, or can be made, sufficiently small.

The following model, and its setting, were first described in [**494**], and a standard reference is [**300**]. It is convenient, at this step, to move to a much broader collection of examples.

Definition 7.3.1 (Valiant).

1. Let X be a set that admits a probability measure. We call X the *instance space.*
2. Let $\mathcal{C}$ be a subset of $\mathcal{P}(X)$, called a *concept class.*
3. The elements of $\mathcal{C}$ are called *concepts.*

It is important for what follows that we note that at this point we do not commit to a particular measure on X; indeed, the diversity of possible measures has an important role to play. For the present, we only make the (very liberal) assumption that X admits some probability measure. We make no assumptions at all yet about the structure of $\mathcal{C}$ or its members. We now proceed to Valiant's important definition of a notion of learnability.

Definition 7.3.2. Let $\mathcal{C}$ be a concept class over X. We say that $\mathcal{C}$ is *PAC Learnable* (where PAC stands for *Probably Approximately Correct*) if and only if there is an algorithm L such that for every $c \in \mathcal{C}$, every $\epsilon, \delta \in (0, \frac{1}{2})$ and every probability measure $\mathcal{D}$ on X, the algorithm L behaves as follows: On input (ϵ, δ), the algorithm L will ask for some number $n = n(\epsilon, \delta)$ of examples, and will be given $\{(x_1, i_1), \ldots, (x_n, i_n)\}$ where x_j are independently randomly drawn according to $\mathcal{D}$, and $i_j = \chi_c(x_j)$. The algorithm will then output some $h \in \mathcal{C}$ so that with probability at least $1 - \delta$ in $\mathcal{D}^n$, the symmetric difference of h and c has probability at most ϵ in $\mathcal{D}$.

In this definition, ϵ bounds what we have called the A probability, and δ bounds what we have called the P probability. The wide range of possibilities for the measure $\mathcal{D}$ reflects externally imposed rules according to which the random training data may be given; we only require that the algorithm be judged only by the measure that produced its training set. Without some form of this restriction, adversarial learning would become absolutely impossible, since a new measure could give probability zero to the set from which the training data were drawn, or concentrate exactly on the error of a particular algorithm on a particular input.

A few examples will illuminate the definition.

Example 7.3.3. Let X be the real line, and let $\mathcal{C}$ be the set of positive half-lines (i.e. intervals of the form (a, ∞)). Given $\epsilon, \delta \in (0, \frac{1}{2})$ we find m so large that

$$(1 - \epsilon)^m < \delta,$$

and ask for m labeled examples, which we will call $S = \{(x_1, i_1), \ldots, (x_n, i_n)\}$. Set $P = \{(x_j, i_k) \in S : i_k = 1\}$. We now define $h = \inf P$, and return the hypothesis $H = (h, \infty)$.

To see that this algorithm succeeds, suppose that the target is (t, ∞). We know that $t \geq h$, because otherwise some of our training data was incorrect. Consequently, $(h, \infty) \triangle (t, \infty)$ is exactly (h, t). We define b to be the greatest such that $\mathcal{D}(t, b) = \epsilon$. Then what we must show is that the probability in $\mathcal{D}^m$ that $h < b$ is less than δ. This is bounded by the probability that none of our training elements is in (t, b); that is, but $(1 - \epsilon)^m < \delta$.

Example 7.3.4. Let X be the real plane, and let $\mathcal{C}$ be the set of filled axis-aligned rectangles. Again, it suffices for the algorithm to take a large enough sample, and take the tightest rectangle H consistent with the training set. We call the target T, and note that, again, by the honesty of our training set, we need only be concerned with $T - H$. Exactly as in the one-dimensional case, we can make this error sufficiently small with high probability.

Example 7.3.5. Let $X = \mathbb{R}^d$, and let $\mathcal{C}$ be the class of linear half-spaces (i.e. of subsets of $\mathbb{R}^d$, each defined by a single linear inequality). Again, a tightest-consistent-fit algorithm succeeds.

Example 7.3.6. Let $X = 2^k$, interpreted as assignments of truth values to Boolean variables. Let $\mathcal{C}$ be the class of k-CNF expressions — that is, the class of propositional formulas on k Boolean variables, in conjunctive normal form (where each expression $c \in \mathcal{C}$ is interpreted as the set of truth assignments that satisfy it). This class is PAC learnable.

The contrast of PAC learning with the earlier models suggests another line of questions which, to the knowledge of this author, has not yet been explored. In the language of the previous section, we might call the PAC learning condition defined here PACFin learning. That is, the learner is allowed only one guess at the hypothesis. We might, by analogy, seek to define PAC-Ex and PAC-Bc learnability to allow a convergent sequence of guesses, but it is not immediately clear how to do this in a way that still respects the randomness of the examples.

In another direction, work of Xiao relates PAC learning to zero-knowledge proofs and the complexity class **BPP**.

Theorem 7.3.7 ([**508**]). *Let* **ZK** *be the set of languages with zero-knowledge proofs as defined in section 4.3, and* **BPP** *be as defined in Definition 4.2.5. There is an oracle A such that PAC-learning the concept class of sets defined by Boolean circuits of size n^2 is PAC-complete among learning problems relative to A, but* $\mathbf{ZK}^A = \mathbf{BPP}^A$.

7.3.2. Classification by Regression.

While the closest-fit algorithms give results that satisfy the definition, it is not hard to believe that better methods are generally available. Suppose that the instance space is $\mathbb{R}^n$, and that the concept class consists of linear half-spaces. Then the learning problem is to find a *separating hyperplane* with all positive instances on one side and all negative instances on the other.

We consider, in particular, the case $n = 2$, so that a linear half-space will be defined by a linear inequality $w_1 x + w_0 y \geq b$. In a typical case, we will be given labeled examples, and will want to determine hypotheses for w_1 and w_2. Of course, the case of linear regression has a similar form in that we are given several points and want to find a line that "fits" those points. The critical difference is in our notion of fit. In the least-squares linear regression, we aim to minimize the sum of the squared errors, which are given by $\left(y_i - \left(-\frac{w_1}{w_0}x_i + \frac{b}{w_0}\right)\right)^2$. In classification, our goal is not to have the line fall as close to our training points as possible; that would often lead to drawing the hard classification boundary in the most problematic place possible, not to mention ignoring the labels on the training set.

Instead, we aim to minimize *classification* errors. For each vector $\vec{w} = (w_1, w_0)$, we have a set $M_{\vec{w}}$ of indices for points in the training set which are misclassified by the linear model arising from $\vec{w}$. We denote the target set by T. For algebraic convenience we formulate the training set as $\{(\vec{x}_i, t_i) : i \in I\}$, where

$$t_i = \begin{cases} 1 & \text{if } \vec{x}_i \in T \\ -1 & \text{otherwise} \end{cases},$$

and $\vec{x}_i = (x_i, y_i)$. Now the overall error function to be minimized is

$$E(\vec{w}) = -\sum_{i \in M_{\vec{w}}} (w_1 x_i + w_0 y_i - b) t_i,$$

so that a point correctly classified by $\vec{w}$ makes a negative contribution to E, and a stronger one as it is further inside the model-predicted region, and likewise an incorrectly classified point makes a positive contribution. We can minimize this error by the numerical technique of *gradient descent* (also called *steepest descent*; see [**136**] for details). We set a real parameter η and update $\vec{w}$ by

$$\vec{w}_{s+1} = \vec{w}_s - \eta \nabla E(\vec{w}),$$

which, with relatively minor calculation, can be rewritten as

$$\vec{w}_{s+1} = \vec{w}_s - \eta \sum_{i \in M_{\vec{w}_s}} \vec{x}_s t_s.$$

It is not at all obvious that this must converge. However, a theorem of Rosenblatt [**423**] shows that if T is indeed a linear half-space as assumed, this process will converge to a linear separator of the training set arising from $\mathbf{w}$. We call the resulting half-space $H_{\vec{w}}$. We note that it is at least a hypothesis *consistent* with the training set. We will see in the next section that this is enough to achieve our goal in PAC learning, by Lemma 7.3.13.

This approach of finding a sharp discriminator is, in certain practical respects, inconvenient. Most importantly, its performance is not graceful when the training data is noisy, as practical training data often is. A single noisy training point could result in a training set which is not linearly separable, and in an algorithm that never converges (although this possibility can be got around, using an approach of [**496**]). Moreover, the output hypothesis sharpens the situation more than is realistic; a point near the boundary probably should be classified as somewhat uncertain. Meanwhile the discontinuity of the threshold function

$$t(\vec{x}) = \begin{cases} 1 & \text{if } x \in H_{\vec{w}} \\ 0 & \text{otherwise} \end{cases}$$

creates serious concerns about the stability and convergence of gradient descent. Indeed, from the perspective of continuity, Rosenblatt's theorem is something of a miracle.

Part of the way forward from these discouraging observations is reflection on our goals. The function $t(\vec{x})$ can be viewed as a statement (based on our hypothesis $H_{\vec{w}}$) about whether $\vec{x}$ is in T or not (since T is unknown, we instead literally assign probability values to $\vec{x} \in H_{\vec{w}}$. We could read this more broadly as crisply assigning probabilities to the statements $\vec{x} \in T$. We might also imagine allowing ourselves to assign probabilities between the extreme values. If we took this approach, Bayes' rule would give us

$$P(H_{\vec{w}}|\vec{x}) = \frac{P(\vec{x}|H_{\vec{w}})P(H_{\vec{w}})}{P(\vec{x}|H_{\vec{w}})P(H_{\vec{w}}) + P(\vec{x}|\neg H_{\vec{w}})P(\neg H_{\vec{w}})}.$$

Now by a change of coordinates

$$a = \ln \frac{P(\vec{x}|H_{\vec{w}})P(H_{\vec{w}})}{P(\vec{x}|\neg H_{\vec{w}})P(\neg H_{\vec{w}})}$$

this formula achieves the familiar form of a logistic function,

$$P(H_{\vec{w}}|\vec{x}) = \frac{1}{1+e^{-a}}.$$

We might attempt, then, to fit our data with a hypothesis with linear level sets, and a logistic cross section in the direction orthogonal to the level sets. This technique is known as *logistic regression.*

To achieve this, we consider a logistic model given by parameters $\vec{w}$, and compute, for each point $(\vec{x}_i, t_i)$ in the training set S, the probability this model assigns to $\vec{x}_i$ having classificaiton value t_i. This method is actually that of maximum likelihood estimation, although for consistency we are actually minimizing an inverted likelihood function. We then regard the points of the training sets as Bernoulli random variables with $P(t_i = 1) = \pi_i$. At this point, we compute the joint probability of the full training set,

$$P(S|\vec{w}) = \prod_{i=1}^{N} \pi_i^{t_i}(1-\pi_i)^{1-t_i}.$$

Now to achieve a maximum likelihood estimate of the parameters, we will minimize the negative log of $P(S|\vec{w})$, given by

$$E(\vec{w}) = -\sum_{i=1}^{n} \ln \pi_i^{t_i}(1-\pi_i)^{1-t_i} = -\sum_{i=1}^{n}\left(t_i \ln \frac{\pi_i}{1-\pi_i} + \ln(1-\pi_i)\right).$$

This function is sometimes called the *cross-entropy error function.* Again using gradient descent, we can numerically estimate parameters $\vec{w}$ minimizing $E(\vec{w})$.

Logistic regression is effective for linearly separable classification problems. There are, however, techniques to bootstrap the idea to more complicated problems. One important body of these ideas is based on models — perhaps naive models — of biological neuronal networks. Some of these models trace back to earlier work, but found their first canonical exposition, complete with learning algorithms, in a 1962 book of Rosenblatt [**423**].

In this body of theory, a single "neuron" is a function of several input variables, usually conceptualized as a composition of two functions: first, a linear combination of the inputs is taken; then a step function (perhaps a logistic function) is evaluated on the result. The weights in the linear combination stage are computed by minimizing the classification error on a training set by gradient descent. The similarity of a single "neuron" (node) to logistic regression should be evident.

The great power of the theory comes, however, when a network of many neurons is created. Typically these networks are graded, so that the "neurons" in level 0 take inputs from the environment and those in level n take inputs from the nodes of level $n-1$. Now there are many weights — a set of weights from the linear combination stage of each node. Again, we iteratively modify the weights to minimize the error on a training set. Layer zero is called the "input" layer, and the final layer is called the output layer. Intermediate layers are called "hidden" layers. The term "deep learning," currently very common in analytics circles, refers (at least primarily) to deep neural networks, that is, those with many hidden layers. The relation of these artifical neural networks, as they are called, to literal biological neuronal networks seems to be the subject of some debate in neuroscience circles. A canonical reference on these biological networks, including a discussion of the perceptron and neural network models, is [**288**].

An interesting property of neural networks is their capacity to converge even on a random labeling. Indeed, in a recent paper [**510**] researchers compared the results of training a neural network both on true data and on a copy of the data in which labels were chosen at random. The optimization process converged a bit more slowly, but still had no trouble converging to a minimum loss assignment of weights. This is sensible enough from the perspective that these networks have sufficient semantic power to express arbitrary or near-arbitrary functions on the sample space (a result of this kind is proved in the same paper). However, it is difficult to reconcile with the relatively strong resistance of some of these models to overfitting.

Indeed, explaining the empirical success of deep neural networks seems to present a significant theoretical challenge. One approach, introduced by Tishby in 1999, is to understand the learning process as an information process: in learning, we try to find a compression of the input data that carries maximal information. Each layer represents an intermediate representation of the data, with a Markov chain property: it depends only on the immediately preceding layer. The theory arising out of this analysis attempts to understand the semantics of neural networks — the expressive power, resistance to overtraining, and efficiency of representation — in terms of the mutual information of various layers. There is an ongoing debate about the effectiveness of this Information Bottleneck theory in accounting for these features. Representatives of the competing perspectives can be found in [**488**] and [**433**].

Another perspective on understanding what neural networks do arises from understanding each layer as a transformation on the data, and considering the dynamics of this system. Sussillo and Barak [**474**] take up this question on recurrent neural networks, which follow a more liberal definition that allows feedback loops. In this case, analysis of linearization around fixed points and similar structures to recover the mechanisms implemented by a trained network. Later work shows that mean field theory and stability of the resulting system can predict success or failure of the training of the system [**420**]. In both the training phase of standard neural networks and the evaluation phase of recurrent neural networks, there appears to be considerable opportunity for analysis from the dynamical perspectives we will see in Chapter 8.

More extensive treatment of logistic regression can be found in [**326**]. Bishop has a book surveying several machine learning techniques, including logistic regression and neural networks [**72**], and the books [**279, 251**] are standard references. Deep neural networks and their semantics are an extremely active area of research. Recent surveys on deep learning can be found in [**334, 435**], and [**114**] has a brief survey with an extensive bibliography on the semantic issues.

7.3.3. Vapnik-Chervonenkis Dimension. In the late 1960s, Vapnik and Chervonenkis considered a problem arising in learning algorithms: whether it is possible "to draw conclusions about the probabilities of the events of an entire class S from one and the same sample." They understood this as a problem of convergence in probability in Bernoulli's law of large numbers. They knew that such uniformity could fail, and gave sufficient conditions for uniformity. The definition that follows is, in updated language, theirs [**497**].

Definition 7.3.8. Let S be a set, and let $\mathcal{C}$ be a concept class on instance set S.

(1) We define $\Pi_{\mathcal{C}}(S) = |\{S \cap c : c \in \mathcal{C}\}|$.
(2) The VC dimension of $\mathcal{C}$, denoted $\dim_{VC}(\mathcal{C})$ is the greatest integer d such that there is some set S of cardinality d such that $\Pi_{\mathcal{C}}(S) = 2^d$, provided that such a d exists. Otherwise, we say that $\dim_{VC}(\mathcal{C}) = \infty$.

Vapnik and Chervonenkis also identified the *growth function* $m_{\mathcal{C}}(r)$ to be the maximum value of $\Pi_{\mathcal{C}}(S)$ where S ranges over all sets of size r. They proved the following result.

Theorem 7.3.9. *For any concept class $\mathcal{C}$, the growth function $m_{\mathcal{C}}(r)$ is either equal to 2^r for all values of r, or it is bounded above by r^n, where n is the first value of r such that $m_{\mathcal{C}}(r) \neq 2^r$.*

They also proved that the condition that $\mathcal{C}$ has finite VC dimension is equivalent to the condition that as sample size grows, the maximum deviation of a sample frequency from the population probability approaches zero with probability one.

The two layers of probability — the probability of convergence and the deviation of sample from population probability — suggest something like PAC learning. It is not quite true that finite VC dimension is equivalent to PAC learnability. However, this equivalence does hold under some very mild conditions.

While the conditions are mild in the sense that most, if not all, examples that arise in the literature meet them, they are technical, and are frequently unstated. The following definition establishes the necessary vocabulary for stating the conditions under which finite VC dimension is equivalent to PAC learnability.

Definition 7.3.10. Let $R \subseteq \mathcal{P}(X)$, and let $\mathcal{D}$ be a probability distribution on X, and $\epsilon > 0$.

(1) We say that $N \subseteq X$ is an ϵ-transversal for R with respect to $\mathcal{D}$ if and only if for any $c \in R$ with $P_{\mathcal{D}}(c) > \epsilon$ we have $N \cap c \neq \emptyset$.
(2) For each $m \geq 1$, we denote by $Q^m_\epsilon(R)$ the set of $\vec{x} \in X^m$ such that the set of distinct elements of $\vec{x}$ does not form an ϵ-transversal for R with respect to $\mathcal{D}$.
(3) For each $m \geq 1$, we denote by $J^{2m}_\epsilon(R)$ the set of all $\vec{x}\vec{y} \in X^{2m}$ with $\vec{x}$ and $\vec{y}$ each of length m such that there is $c \in R$ with $P_{\mathcal{D}}(c) > \epsilon$ such that no element of c occurs in $\vec{x}$, but elements of c have density at least $\frac{\epsilon m}{2}$ in $\vec{y}$.
(4) We say that a concept class $\mathcal{C}$ is *well-behaved* if for every Borel set b, the sets $Q^m_\epsilon(R)$ and $J^{2m}_\epsilon(R)$ are measurable where $R = \{c \triangle b : c \in \mathcal{C}\}$.

This condition of "well-behavedness" is enough, as Blumer, Ehrenfeucht, Haussler, and Warmuth showed in the following theorem.

Theorem 7.3.11 ([**74**]). *Let $\mathcal{C}$ be a nontrivial well-behaved concept class. Then $\mathcal{C}$ is PAC learnable if and only if $\mathcal{C}$ has finite VC dimension.*

We will give a proof of this theorem from the following two lemmas.

LEMMA 7.3.12. *Let d be the VC dimension of $\mathcal{C}$. For $0 < \epsilon < 1/2$ and sample size less than* $\max\{\frac{1-\epsilon}{\epsilon}\ln\frac{1}{\delta}, d(1-2(\epsilon(1-\delta)+\delta))\}$ *no function is a learning function for $\mathcal{C}$.*

LEMMA 7.3.13. *If the VC dimension of $\mathcal{C}$ is $d < \infty$, then for any $0 < \epsilon < 1$ and sample size at least* $\max\{\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{13}{\epsilon}\}$ *any consistent learner learns $\mathcal{C}$.*

PROOF OF THEOREM FROM LEMMAS. Suppose that $\mathcal{C}$ has finite VC dimension. Then $\mathcal{C}$ is learnable by Lemma 7.3.13. On the other hand, the second lower bound of Lemma 7.3.12 grows arbitrarily large with d for appropriate choice of ϵ and δ. Then if d is infinite, the sample size must also be infinite. □

We now proceed with the proof of the two lemmas. The first is the more straightforward of the two.

PROOF OF LEMMA 7.3.12. Let $\mathcal{C}$ be a concept class of finite VC dimension, let $\epsilon \in (0, \frac{1}{2})$, and let m be the bound given.

Suppose $\mathcal{C}$ contains $S_1 \neq S_2$ with a nonempty intersection, and let L be a learning function. Let $a \in S_1 \cap S_2$, and $b \in S_2 - S_1$. Let P be the probability distribution concentrated on a and b with $P(b) = \epsilon$. Replace $X, \mathcal{C}$ with this reduced case, where $S_1 \cap S_2$ is represented by a and $S_2 - S_1$ by b.

We can show that if $m < \frac{\ln\frac{1}{\delta}}{-\ln(1-\epsilon)}$ then the probability of drawing a is at least δ. This also holds for

$$m \leq \frac{(1-\epsilon)\ln\frac{1}{\delta}}{\epsilon}$$

. If there is a sample of m elements in which each point is a, there are four things the learning function could do. If it guesses $\{a\}$, then it has eror ϵ for the target concept $\{a, b\}$. If it gives any other answer, it has error at least ϵ if the target concept is $\{a\}$. In any case, there is error at least ϵ with probability greater than δ.

Suppose $\mathcal{C}$ contains S_1, S_2 such that $S_1 \cup S_2 \neq X$. Then let $a \in X - (S_1 \cup S_2)$ and $b \in S_1$. Given any learning function L, we let P be as above. Then a similar analysis verifies the first lower bound. For the second term, note that the VC dimension must be witnessed by a set of $d \geq 1$ points in X that is shattered by $\mathcal{C}$. Let P be uniform on these points and 0 elsewhere. Replace X with this set.

Suppose we draw a sequence with ℓ different points. For each of the 2^ℓ labelings, there are 2^{d-1} concepts consistent with it. Whatever the hypothesis of the learning function, for every point of X not observed, it will be correct for half. Thus, the average error is at least

$$(d-\ell)/(2d) \geq (d-m)/(2d)$$

This implies that the average error is at least $(d-m)/(2d)$. Hence there must be a concept with at least this error. □

The proof of Lemma 7.3.13 is more technical, and requires some lemmas in its own right. Proofs of the following three results can be found in [**74**], but are omitted here.

LEMMA 7.3.14. *For any $\epsilon > 0$ and and $m > 2/\epsilon$, we have $P(Q_\epsilon^m) < 2P(J_\epsilon^{2m})$.*

We define $\Pi_{\mathcal{C}}(m)$ to be the $\max \Pi_{\mathcal{C}}(S)$ where $|S| = m$.

LEMMA 7.3.15. $P(J_\epsilon^{2m}) < \Pi_R(2m)2^{-\epsilon m/2}$ *for all $m \geq 1$ and $\epsilon > 0$*

Lemma 7.3.16. *For any $m \geq 1$, any $S \subseteq X$, and any $H \subseteq \mathcal{P}(X)$, we have $\Pi_H(m) = \Pi_R(m)$, where $R = \{h \triangle S : h \in \mathcal{C}\}$.*

We now proceed with a proposition that brings us a good deal closer to Lemma 7.3.13.

Proposition 7.3.17. *Let $\mathcal{C}$ be a nonempty well-behaved concept class on X, let P be a probability measure on X, and let S be Borel. Then for any $\epsilon > 0$, $m \geq \frac{2}{\epsilon}$, given m independent random examples of S drawn according to P, the probability that there is a hypothesis in $\mathcal{C}$ consistent with the examples and having error greater than ϵ is at most $2\Pi_H(2m)2^{-\epsilon m/2}$.*

Proof. Let R be as in Lemma 7.3.16. A hypothesis $h \in \mathcal{C}$ has error greater than ϵ only if its symmetric difference with S has probability greater than ϵ. Hence, if the example points are drawn from an ϵ-transversal for R, then every hypothesis with error greater than ϵ will have an example drawn from its symmetric difference with S. No such hypothesis will be consistent. So the probability that there exists a hypothesis consistent with the examples and having error greater than ϵ is the probability that the examples contain no ϵ-transversal.

Since H is well-behaved and S is Borel, we apply Lemma 7.3.14 to find that the probability that the examples contain no ϵ-transversal is bounded by $2P(J^{2m}_{\epsilon})$, and apply Lemma 7.3.15 to see that this is, in turn, bounded by $2\Pi_R(2m)2^{-\epsilon m/2}$. □

Continuing with the proof of Lemma 7.3.13, we define $\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}$ if $m \geq d$ and 2^d otherwise. We can now prove by a routine induction on d that if the VC dimension of H is d, then $\Pi_H(m) \leq \Phi_d(m)$. This completes the proof of Lemma 7.3.13, and of Theorem 7.3.11.

When read in their full detail, the results of Blumer, Ehrenfeucht, Haussler, and Warmuth establish not only the learnability of classes with finite VC dimension, but also the size of the training set necessary for the algorithm to succeed. Naturally, it is of interest to improve this bound as much as possible. In more recent years, Li, Tromp, and Vitanyi have used Kolmogorov complexity to work in this direction [**338**]. Their concern was first to optimize the bounds, and second that the number of examples needed (the so-called "sample complexity") be independent of representation.

Definition 7.3.18. A *representation system* is a tuple (R, Γ, c, Σ), were $R \subseteq \Gamma^*$ is the set of representations, and $c : R \to 2^{\Sigma^*}$.

Now given a set of representations R, a representation system determines a concept class, the range of c.

Definition 7.3.19. An Occam algorithm for a representation system (R, Γ, c, Σ) is a randomized algorithm that, for every $n \geq 1$ and every $\gamma > 0$, when given as input a set S of m examples of concept $c(r)$, with probability at least $1 - \gamma$ outputs $t \in R$ such that $c(t)$ is consistent with S, and such that for some function $f(m, n, \gamma)$ increasing in m, we have $K(t|r, n) < \frac{m}{f(m,n,\gamma)}$.

Theorem 7.3.20 (Li-Tromp-Vitanyi). *Suppose that there is an Occam algorithm for (R, Γ, c, Σ). Then there is an algorithm that PAC-learns the concept class* *ran*(c) *using a training set of size*

$$m = \max\left\{\frac{2}{\epsilon}\ln\frac{2}{\delta}, f^{-1}\left(\frac{2\ln 2}{\epsilon}, n, \frac{\delta}{2}\right)\right\}.$$

PROOF. On input ϵ, δ, the learning algorithm will take a training set of the size given from the oracle. It will then run the Occam algorithm with $\gamma = \delta/2$. The output is then a hypothesis in the concept class. Now the probability that this hypothesis has all m examples bad (outside its putatively large symmetric difference from the target) is bounded by the number of concepts of Kolmogorov complexity less than $\frac{m}{f(m,n,\gamma)}$, times $(1-\epsilon)^m$. We now compute this quantity.

As in the previous proof, we note that since $m \geq f^{-1}\left(\frac{2\ln 2}{\epsilon}, n, \frac{\delta}{2}\right)$ we have

$$\epsilon - \frac{\ln 2}{f(m,n,\gamma)} \geq \frac{\epsilon}{2},$$

so that since $m > \frac{2}{\epsilon}\ln\frac{2}{\delta}$ we also have

$$m\left(\epsilon - \frac{\ln 2}{f(m,n,\gamma)}\right) \geq \ln\frac{2}{\delta}.$$

By exponentiating both sides, we get

$$2^{m/f(m,n,\gamma)}(1-\epsilon)^m \leq \frac{\delta}{2}.$$

Consequently, the learning algorithm succeeds with probability δ. □

The same paper also establishes a partial converse: In the following sense, a learning algorithm really does require compression.

THEOREM 7.3.21 (Li-Tromp-Vitanyi). *Let* (R, Γ, c, Σ) *be a representation system, and let* L *be a deterministic algorithm that PAC-learns* ran(c) *using* m *examples. Then there is an Occam algorithm for* (R, Γ, c, Σ) *achieving compression* $f(m,n,\gamma) = \frac{1}{2\epsilon n}$.

It is difficult to overstate the liberality of the condition of well-behavedness in contrast to the actual requirements of most practical learning situations. One way to see this is to notice a special class of concept classes which itself includes nearly every imaginable learning problem, and which is itself a very small subclass of the well-behaved classes. This class will be formulated in terms of Π^0_1 classes. This formulation is introduced in [**95**], and the remainder of this section describes concepts and results from that paper. The following result is well-known, but a proof is given in [**110**], which is also a good general reference on Π^0_1 classes.

THEOREM 7.3.22. *Let* $c \subseteq 2^\omega$. *Then the following are equivalent:*

(1) *c is the set of all infinite paths through a computable subtree of* 2^ω
(2) *c is the set of all infinite paths through a* Π^0_1 *subtree of* 2^ω *(i.e. a co-c.e. subtree)*
(3) $c = \{x \in 2^\omega : \forall n\ R(n,x)\}$ *for some computable relation R, i.e. a relation R for which there is a Turing functional Φ such that $R(n,x)$ is defined by* $\Phi^x(n)$.

This equivalence gives rise to the following definition:

DEFINITION 7.3.23. Let $c \subseteq 2^\omega$. We say that c is a Π^0_1 class if and only if it satisfies one of the equivalent conditions in Theorem 7.3.22.

Most instance spaces with which one could want to work can, by standard techniques, be encoded in 2^ω, and for the purposes of this discussion we will take $X = 2^\omega$ for the remainder of this section. Similarly, a large collection of concepts can be understood as Π^0_1 classes. Unless otherwise noted, we use the standard product topology on 2^ω. It remains to describe the concept classes to be used. We make the following preliminary definitions:

Definition 7.3.24. (1) Let $f, g \in 2^{\leq\omega}$. Then $d(f, g)$ is defined to be 2^{-n} where n is the least natural number such that $f(n) \neq g(n)$.

(2) Let $f \in 2^{\leq\omega}$ and $r \in \mathbb{R}$. We denote by $B_r(f)$ the set $\{g \in 2^{\leq\omega} : d(f, g) < r\}$.

(3) Let $S \subseteq 2^\omega$. We say that S is *computable* if and only if there is a computable function $f_S : 2^{<\omega} \times \mathbb{Q} \to \{0, 1\}$ such that

$$f_S(\sigma, r) = \begin{cases} 1 & \text{if } B_r(\sigma) \cap S \neq \emptyset \\ 0 & \text{if } B_{2r}(\sigma) \cap S = \emptyset \\ 0 \text{ or } 1 & \text{otherwise} \end{cases}$$

There is an unfortunate clash of terminology in that the concept *classes* will have, for their members, Π^0_1 *classes.* We will never use the term ambiguously, but because both terms are so well-established it will be necessary to use both of them.

Definition 7.3.25. A weakly effective concept class is a computable enumeration $\varphi_e : \mathbb{N} \to \mathbb{N}$ such that $\varphi_e(n)$ is a Π^0_1 index for a Π^0_1 tree $T_{e,n}$.

We would like one additional property: that a finite part of an effective concept class $\mathcal{C}$ should not be able to distinguish a non-computable point of 2^ω from all computable points, in the sense that if $y \in 2^\omega$ is noncomputable, then any finite Boolean combination of members of $\mathcal{C}$ containing y should also contain a computable point. This is reasonable: it would strain our notion of an "effective" concept class if it should fail. And yet it can fail with a weakly effective concept class: our classes may have no computable members at all, for instance. For that reason, we define an effective concept class as follows.

Definition 7.3.26. An effective concept class is a weakly effective concept class φ_e such that for each n, the set c_n of paths through $T_{e,n}$ is computable as a subset of 2^ω.

Note that for the set c_n to be computable, it is not necessary that all of its elements are computable. We note that all standard examples of learning problems are effective concept classes. For instance, using standard representations, any example in [**72**], [**300**], or [**429**] is an effective concept class.

Example 7.3.27. The class of well-formed formulas of classical propositional calculus, and the class of k-CNF expressions (for any k) are effective concept classes, by the example above. Whether a given $y \in 2^\omega$ satisfies a particular formula can be determined by examining only finitely many terms of y.

Example 7.3.28. The class $\mathcal{C}$ of linear half-spaces in $\mathbb{R}^d$ bounded by hyperplanes with computable coefficients is an effective concept class. Recall that each linear half-space with computable coefficients is a computable set, since the distance of a point from the boundary can be computed.

Example 7.3.29. The class of convex d-gons in $\mathbb{R}^2$ with computable vertices is an effective concept class.

Note that the requirement of computable boundaries of these examples is not a practical restriction. The proof of the following result is straightforward, and can be found in [**95**]. (We denote a vector $\langle x_1, \ldots, x_d \rangle$ by $\bar{x}$.)

PROPOSITION 7.3.30. *For any probability measure μ on $\mathbb{R}^d$ absolutely continuous with respect to Lebesgue measure, any $\epsilon > 0$, and any hyperplane given by $f(\bar{x}) = 0$, there is a hyperplane given by $\bar{f}(\bar{x}) = 0$ where $\bar{f}$ has computable coefficients, and where the linear half-spaces defined by these hyperplanes are close in the following sense: If H_f is defined by $f(\bar{x}) \leq 0$, if H_f^0 is defined by $f(\bar{x}) < 0$, and $H_{\bar{f}}$ is defined by $\bar{f}(\bar{x}) \leq 0$, then $\mu\left(H_f \triangle H_{\bar{f}}\right) < \epsilon$ and $\mu\left(H_f^0 \triangle H_{\bar{f}}\right) < \epsilon$.*

It is an easy consequence of the definition that every effective concept class is well-behaved. Indeed, the sets required to be measurable are, in fact, low-level Borel.

7.3.4. The Index Set for PAC Learnable Classes. One straightforward decision problem arising out of learning theory is whether a given concept class is learnable or not. Beros [**67**] took up this problem for the Gold-style learning of computably enumerable sets, proving the following.

THEOREM 7.3.31. *The set of Σ_1^0 indices for uniformly computably enumerable families learnable in each of the following models is m-complete in the corresponding class.*

(1) *TxtFin* — Σ_3^0
(2) *TxtEx* — Σ_4^0
(3) *TxtBC* — Σ_5^0
(4) *TxtEx** — Σ_5^0

In full generality, the problem would be difficult to express for PAC learning; literally any element of the double powerset of any set can be a concept class. However, as we have seen, this level of generality is not necessary to comfortably take in all practical cases. In the context of effective concept classes, there is even a natural notion of an index set. In [**95**], the following calculation is made.

THEOREM 7.3.32. *The set of indices for effective concept classes of infinite VC dimension is m-complete Π_3^0 within the set of indices for effective concept classes, and the set of indices for effective concept classes of finite VC dimension is m-complete Σ_3^0 within the set of indices for effective concept classes.*

PROOF. A preparatory result will be helpful in demonstrating the bound.

LEMMA 7.3.33. *An effective concept class $\mathcal{C}$ has infinite VC dimension if and only if for every d there are (not necessarily uniformly) computable elements*

$$(x_i : i < d)$$

such that $\Pi_{\mathcal{C}}(x_i : i < d) = 2^d$.

PROOF. Let $(y_i : i < d)$ witness that $\mathcal{C}$ has VC dimension at least d, and denote by $D_1, \ldots, D_{2^d}$ elements of $\mathcal{C}$ which distinguish distinct subsets of $(y_i : i < d)$. For each $i < d$, there is a computable element x_i such that for every $j \leq 2^d$ we have $x_i \in D_j$ if and only if $y_i \in D_j$. Thus $x_1, \ldots, x_d$ witness that $\mathcal{C}$ has VC dimension at least d. The converse is obvious. □

Now we can define the set of effective concept classes of infinite VC dimension with a Π_3 formula. We begin by noting that if f is a computable function and T is a Π^0_1 tree, then it is a Π^0_1 condition that f is a path of T, and a Σ^0_1 condition that it is not, uniformly in a Π^0_1 index for T and a computable index for f. Further, if $\mathcal{C} = \varphi_e$ is an effective concept class, then for any $k \in \omega$, the condition that $k \in ran(\varphi_e)$ is a Σ^0_1 condition, uniformly in e and k.

Let $(x_1, \ldots, x_n)$ be a sequence of computable functions, $S \subseteq \{1, \ldots, n\}$, and c a Π^0_1 class, represented by a Π^0_1 index for a tree in which it is the set of paths. We abbreviate by $[c\restriction_n = S]\,(\bar{x})$ the statement that for each $i \in \{1, \ldots n\}$, we have $x_i \in c$ if and only if $i \in S$. Now $[c\restriction_n = S]\,(\bar{x})$ is a d-Σ^0_1 condition, uniformly in the indices for the x_i and c.

We now note that $\mathcal{C} = \varphi_e$ has infinite VC dimension if and only if

$$\forall (n \in \mathbb{N}) \exists x_1, \ldots, x_n \bigwedge_{S \subseteq (n+1)} \exists k \, [\varphi_e(k) \restriction_n = S]\,(\bar{x}).$$

From the comments above, this definition is Π^0_3.

Toward completeness, for each Π^0_3 set S, we will construct a sequence of effective concept classes $(\mathcal{C}_n : n \in \mathbb{N})$ such that $\mathcal{C}_n$ has infinite VC dimension if and only if $n \in S$. In the following lemma, to simplify notation, we suppress the dependence of f on n.

Lemma 7.3.34. *There is a* Δ^0_2 *function* $f : \mathbb{N} \to 2$ *such that* $f(s) = 1$ *for infinitely many* s *if and only if* $n \in S$.

Proof. It suffices (see [**457**]) to consider S of the form $\exists^\infty x \forall y R(x, y, n)$, where R is computable. Now we set

$$f(x) = \begin{cases} 1 & \text{if } \forall y R(x, y, n) \\ 0 & \text{otherwise} \end{cases}.$$

This function is Δ^0_2-computable, and has the necessary properties. □

Now by the Limit Lemma, there is a uniformly computable sequence

$$(f_s : s \in \mathbb{N})$$

of functions such that for each x, for all sufficiently large s, we have $f_s(x) = f(x)$.

We now define a set of functions that will serve as the elements that may eventually witness high VC dimension. Let $\{\pi_{s,t,j} : s, t, j \in \mathbb{N}, j < s\}$ be a discrete uniformly computable set of distinct elements of 2^ω such that $\pi_{s,t,j}(q) = \pi_{s,t',j'}(q)$ whenever $q < \min\{t, t'\}$.

We also initialize $G_{s,0} = \emptyset$ for each s. Denote by P_t a bijection

$$P_t : \mathcal{P}\left(\{1, \ldots, t\}\right) \to \{1, \ldots, 2^t\}.$$

At stage s of the construction, we consider $f_s(t)$ for each $t \leq s$. If $f_s(t) = 0$, then no action is required.

If $f_s(t) = 1$, then we find the least k such that $k \notin G_{t,s-1}$. Let $\{e_{t,i} : i < 2^t\}$ be Π^0_1 indices for trees such that $T_{e_{t,i}}$ consists exactly of the initial segments τ of $\pi_{t,k,j}$ where $j = P_t(S)$ for some $S \subseteq \{1, \ldots, t\}$ and $|\tau|$ is less than the first $z > s$ such that $f_z(t) = 0$. We let i_s be the least such that $\mathcal{C}_n(i_s)$ is undefined, and take $\mathcal{C}_n(i_s + \ell) = e_{t,\ell}$ for each $\ell < 2^t$. We also set $G_{t,s} = G_{t,s-1} \cup k$.

Now for each t with $f(t) = 1$, there will be some s such that $f_{s'}(t) = f_s(t) = 1$ for all $s' > s$. Thus at stage s we have added to $\mathcal{C}_n$ the Π^0_1 indices $\{e_{t,i} : i < 2^t\}$ guaranteeing that $\{\pi_{t,k,j} : j < t\}$ is shattered for some k.

For each t such that $f(t) = 0$ and each s such that $f_s(t) = 1$, there is some later stage s' such that $f_{s'}(t) = 0$, so any indices added at stage s will be indices for a tree with no paths.

Note that if the same t receives attention infinitely often, this does not inflate the VC dimension beyond t. Indeed, the sets of witnesses will be pairwise disjoint, so no concept in $\mathcal{C}_n$ will include any mixture of witnesses from different treatments; the resulting sets will not be shattered.

We further note that all the Π^0_1 classes in $\mathcal{C}_n$ are computable. Indeed, each $c \in \mathcal{C}_n$ is a finite set of computable paths. Thus, $\mathcal{C}_n$ is an effective concept class.

Now if $n \notin S$, then $f(s) = 1$ for at most finitely many s, so that the VC dimension of $\mathcal{C}_n$ is finite. If $n \in S$, then $f(s) = 1$ for infinitely many s, so that the VC dimension of $\mathcal{C}_n$ is infinite (since sets of arbitrarily large size will be shattered). □

7.3.5. PAC reducibility. As is usual in computation, some positive learnability results are obvious, and others can be proved by giving an algorithm (or computing a dimension). Negative results, those that show non-learnability, are, as usual, at a premium. This is part of the appeal, not just in learning, but in other areas of computation, for reducibilities.

Of course, reducibilities do much more than help us take maximal advantage of known negative results. They often help us pose interesting computational questions. Montalban has suggested the philosophical approach that in Turing computability, considering results that hold on a Turing "cone" — that is, a set of the form $\{X : A \leq_T X\}$ for some A — captures the "typical" behavior, in the sense that such results avoid a significant class of ad hoc counter examples: those that are built by diagonalizing against all computable functions [**380**]. This is an interesting approach, and even a debatable one, but the point is that it is something that can be expressed in terms of reducibility.

It was not long after the definition of PAC learning that the difficulty of negative results became apparent. It is natural enough to those who have studied computability or complexity: one negative result is an epoch-making breakthrough (if it ever comes). Beyond that, we depend on showing that a solution to one problem would imply a solution to another — that is, we depend on appropriate notions of reducibility. In 1990, Pitt and Warmuth introduced such a reducibility notion for PAC learning, that allowed the transfer of negative results (even conditional negative results) from one problem to another [**409**].

Definition 7.3.35. Let $\mathcal{C}_0$ be a concept class over the instance space X_0 and $\mathcal{C}_1$ a concept class over the instance space X_1. We say that $\mathcal{C}_0 \leq_{PAC} \mathcal{C}_1$ if and only if

(1) There is a polynomial-time computable function $g : X_0 \to X_1$, and a polynomial p_g such that if $x \in X_0$ is of size n, then $g(x)$ is of size at most $p_g(n)$,
(2) There is a function $h : \mathcal{C}_0 \to \mathcal{C}_1$, and a polynomial p_h such that if $c \in \mathcal{C}_0$ of size n, then $h(x)$ is of size at most $p_h(n)$, and
(3) For all $x \in X_0$ and all $c \in \mathcal{C}_0$, we have $x \in c$ if and only if $g(x) \in h(c)$.

The definition, as stated here, gives no clear meaning to the term "size." In Pitt and Warmuth's (and Valiant's) original examples, where each instance space is a set of finite objects and each concept class has at least an obvious finite representation, the issue of size poses no serious challenge. However, the broader class of examples offered by effective concept classes in general (not to say concept classes in general) invite closer scrutiny to this point. Recently, Senadheera [**442**] has argued that Kolmogorov complexity would be an appropriate notion of size to put all effective concept classes on an equal footing.

It is also worth noting that there is no *ipso facto* requirement that h be computable. While this clearly is not an absolute obstruction to the study of the partial ordering of concept classes by PAC reducibility, an intermediate reducibility may also be worthy of study, in addition to being more tractable.

Definition 7.3.36. Let $\mathcal{C}_0$ be a countable concept class over the instance space $X_0 \subseteq \omega^\omega$ and $\mathcal{C}_1$ a concept class over the instance space X_1. We say that $\mathcal{C}_0 \leq_{PACi} \mathcal{C}_1$ if and only if there are functions $g : X_0 \to X_1$ and $h : \mathcal{C}_0 \to \mathcal{C}_1$ such that g is represented by a Turing functional, such that h is a computable function, and such that for all $x \in X_0$ and all $c \in \mathcal{C}_0$, we have $x \in c$ if and only if $g(x) \in h(c)$.

Iin much of what has been done with PAC reducibility, the "computability" and "complexity" versions (that is, PACi and PAC reducibilities), the parallel between the two is very close. Consider, by way of example, the following foundational result.

Proposition 7.3.37. *Let $\mathcal{C}_0 \leq_{PACi} \mathcal{C}_1$, and suppose that $\mathcal{C}_1$ is PAC learnable. Then $\mathcal{C}_0$ is also PAC learnable. Moreover, if $\mathcal{C}_0 \leq_{PAC} \mathcal{C}_1$ and $\mathcal{C}_1$ is PAC learnable, then $\mathcal{C}_0$ is also PAC-learnable.*

Proof. Suppose that L_1 PAC learns $\mathcal{C}_1$. Given ϵ, δ, we receive from L_1 the number of training examples it needs, and request that training set $\{x_1, \ldots, x_n\} \subseteq X_0$. We then compute $g(x_i)$ for each i, and return that set as the training set for L_1. We take the hypothesis h_1 returned by L_1, and we, in our effort to learn $\mathcal{C}_0$, use hypothesis $h_0 := h_1 \circ g$. Note that this hypothesis satisfies the conditions of probable correctness required of PAC learning. Also note that this algorithm is efficient. □

The structure of this reducibility does not seem well-studied. We observe, of course, that the empty concept class on the empty instance space is reducible to any other concept class, and that any class is reducible to itself (by the identity functions). It seems clear on cardinality grounds to expect that both $\leq_{PAC}$ and $\leq_{PACi}$ admint incomparable concept classes, but this is less obvious, for instance, among effective concept classes. Of course, to a computability theorist, these degree structures suggest the Turing degrees and their well-studied structure. It seems reasonable also to consider whether the effective concept classes play a role analagous to the computably enumerable degrees. Senadheera has proved some important preliminary results in such a program.

Senadheera has made some beginning explorations of the degree structure, including the following results.

Theorem 7.3.38 ([**442**]). *Let $\preceq$ be either PAC reducibility or PACi reducibility. Then there exist $\preceq$-incomparable effective concept classes.*

Theorem 7.3.39 ([**442**]). *There is an effective concept class $\mathcal{K}$ such that for any effective concept class $\mathcal{C}$ we have both $\mathcal{C} \leq_{PAC} \mathcal{K}$ and $\mathcal{C} \leq_{PACi} \mathcal{K}$.*

Further preliminary work by Senadheera suggests that there may be an embedding of the 1-degrees into the PACi degrees, although it is not at all clear at the time of this writing how to adapt this embedding into an embedding of 1-degrees (or any other known degree structure) into PAC degrees.

7.4. NIP Theories

7.4.1. The Independence Property. Aside from a concept that is called a dimension, the shift at this point in our study to neostability theory seems slightly jarring. On the other hand, the initial definitions of NIP formulas and theories bear immediate resemblance to their analogues in the learning problem. The acronym *NIP* indicates the negation of the independence property, which was historically the first condition used, although we give NIP as the basic definition, in keeping with the modern approach found, for instance, in [**456**]. Indeed, [**456**] is the current standard reference in this field, and the exposition of the present section owes much to it. It is conventional in discussions of this kind to make calculations in a large model $\mathcal{U}_T$ of the theory under consideration, which is saturated in some relatively large cardinal and homogeneous.

Definition 7.4.1. Let T be a first-order L-theory, and $\varphi(\bar{x};\bar{y})$ a first-order L-formula, and S a set of n-tuples. Then we say that S is *shattered by* $\varphi(\bar{x};\bar{y})$ if and only if there is a family $\left(\bar{b}_i : i \in I\right)$ such that for each $V \subseteq S$ we have some i such that $\varphi(\bar{a};\bar{b}_i)$ holds exactly of those $\bar{a}$ with $\bar{a} \in V$.

Provided that S is finite and of cardinality d, the condition that $\varphi(\bar{x};\bar{y})$ shatters S is exactly the condition that distinct choices of $\bar{y}$ cause φ to define exactly 2^d subsets of S, in parallel with the conditions of Definition 7.3.8.

Definition 7.4.2. Let T be a first-order L-theory.

(1) A formula $\varphi(\bar{x};\bar{y})$ is said to be NIP if no infinite set S is shattered by $\varphi(\bar{x};\bar{y})$.
(2) A theory T is said to be NIP if and only if all formulas $\varphi(x;y) \in L$ are NIP.

One *prima facia* difference between this situation and that of Definition 7.3.8 is that it is not obvious that an NIP theory must have some finite size at which it does not shatter sets.

Proposition 7.4.3. *Let T be a theory and $\varphi(\bar{x};\bar{y})$ be NIP. Then there is some natural number d such that if $|S| = d+1$, then $\varphi(\bar{x};\bar{y})$ does not shatter S.*

Proof. Suppose that $\varphi(\bar{x};\bar{y})$ shatters sets of arbitrarily large finite sizes $(n_i : i \in \mathbb{N})$. Then for each i we can write a first-order sentence saying there exist $z_1, \ldots, z_{n_i}$ such that $\varphi(\bar{x};\bar{y})$ shatters $\{z_1, \ldots, z_{n_1}\}$. Since every finite subset of these sentences is consistent with T, it follows by compactness that $\varphi(\bar{x};\bar{y})$ shatters an infinite set. □

This number d is called the VC dimension of φ, and bears transparent resemblance to the VC dimension of a concept class. Note that this does not, at once, settle the question of whether the theory T must have such a dimension

The so-called *independence property* (T is said to have the independence property if and only if it fails to be NIP) has known relationships with several other of the standard benchmarks in stability and neostability theory. We say that a formula $\varphi(\bar{x}, \bar{y})$ is *unstable* if and only if for every $n \in \mathbb{N}$ there are sequences $(\bar{a}_i : i < n)$ and $(\bar{b}_i : i < n)$ such that $\mathcal{U} \models \varphi(\bar{a}_i; \bar{b}_j)$ if and only if $i < j$. A theory is unstable if and only if it has an unstable formula. Otherwise, it is said to be stable.

Proposition 7.4.4. *Let T be a complete first-order theory. Then if T is stable then T is NIP.*

Proof. Suppose that T has an independent formula $\varphi(\bar{x}; \bar{y})$ witnessed by an infinite set $S = (s_i : i \in \mathbb{N})$ shattered by φ with parameters $(\bar{b}_i : i \in I)$. We set $c_i = b_{s_0, \ldots, s_{i-1}}$, and observe that φ defines a linear ordering on the c_i, so that φ is unstable and T is unstable. □

Another important boundary is the *strict order property*. Let T be a first-order theory. A formula $\varphi(\bar{x}, \bar{y})$ is said to have the strict order property if and only if there is a sequence of tuples $(\bar{b}_i : i \in \mathbb{N})$ such that for all $i \in \mathbb{N}$ and for all $\bar{a}$, then $\varphi(\bar{a}, \bar{b}_i)$ strictly implies $\varphi(\bar{a}, \bar{b}_{i+1})$. We say that T has the strict order property if and only if it has a formula with the strict order property. In the following results, we will need the concept of indiscernible sequences.

Definition 7.4.5. Let T be a theory and $A \subseteq \mathcal{U}_T$.

(1) Let $I = (b_i : i \in \kappa)$ be an infinite sequence. We say that I is *indiscernible over A* if and only if for any $k \in \mathbb{N}$ and any two finite increasing subsequences of I with the same length satisfy the same first-order formulas with parameters from A.
(2) Let $\mathcal{I} = (I_j : j \in \lambda)$ be a sequence of sequences. We say that $\mathcal{I}$ is *mutually indiscernible over A* if and only if for each $j \in \lambda$, the sequence I_j is indiscernible over $A \cup \left(\bigcup_{i \neq j} I_i \right)$.

Theorem 7.4.6 (Theorem II.4.7 of [**448**]). *A complete theory is unstable if and only if it has either the strict order property or the independence property.*

Proof. Certainly if a formula φ of T has either the strict order property or the independce property, then both φ and T are unstable. For the converse, suppose that $\varphi(\bar{x}; \bar{y})$ is an unstable NIP formula of T, with an indiscernible sequence $(\bar{a}_i : i \in \mathbb{N})$ and a sequence $(\bar{b}_j : j \in \mathbb{N})$ such that $\varphi(\bar{a}_i; \bar{b}_j)$ if and only if $i < j$. Let $n = dim_{VC}\varphi + 1$, and V a set not defined by φ witnessing that φ does not shatter a set S of size n. We define

$$\psi_i(\bar{x}; \bar{y}) = \begin{cases} \varphi(\bar{x}; \bar{y}) & \text{if } i \in V \\ \neg\varphi(\bar{x}; \bar{y}) & \text{otherwise} \end{cases},$$

and note that $\Phi_0 := \bigwedge_{i<n} \psi_i(\bar{a}_i; y)$ is inconsistent with T.

On the other hand, $\bar{b}_N$ satisfies $\Phi_1 := \left(\bigwedge_{i<N} \varphi(\bar{a}_i, \bar{y}) \right) \wedge \left(\bigwedge_{N \leq i \leq n} \varphi(\bar{a}_i, \bar{y}) \right)$, so Φ_1 is consistent. From an appropriate sequence of formulas beginning with Φ_0 and ending with Φ_1, we can derive a formula with the strict order property. □

This is the sharpest result possible using these three properties. Indeed, the theory of real closed fields is NIP but has the strict order property; the theory of the random graph does not have the strict order property, but has the independence property. Of course, there are theories (true arithmetic, for example) that have both.

Definition 7.4.7. Let p be a partial type over a set A, and κ a cardinal. We then define dp-rank as follows. We say that $\mathrm{dp}-\mathrm{rk}(p, A) < \kappa$ if and only if for every family $(I_t : t < \kappa)$ of mutually indiscernable sequences over A and every realization b of p, there is some t such that I_t is indiscernible over $A \cup \{b\}$.

Proposition 7.4.8. *The following are equivalent:*

(1) *T is NIP.*
(2) *For every type p and set A, there is some κ such that* $\mathrm{dp}-\mathrm{rk}(p, A) < \kappa$*.*

Proof. First suppose that $\varphi(x; y)$ has the independence property. We will show that $\mathrm{dp}-\mathrm{rk}(p, \emptyset)$ is unbounded. For any cardinal κ, we can find a tuple $\bar{b}$ and an indiscernible sequence $(\bar{a}_i : i \in \omega \times \kappa)$ where $\varphi(\bar{a}_i; \bar{b})$ defines the set $\{i : i = (0, \alpha)\}$. Thus, the sequences $I_t = (a_{(n,t)} : n \in \omega)$ are mutually indiscernible, but none are indiscernible over b, so that $\mathrm{dp}-\mathrm{rk}(p, \emptyset) \geq \kappa$.

On the other hand, suppose that T is NIP, and let $(I_t : t \in X)$ be mutually indiscernible sequences over a set A. Let $\bar{b} \in \mathcal{U}$. We may, with a little work, assume that these are sequences of singletons, with $I_t = (a_{t,i} : (t, i) \in X \times \kappa_t)$. We add new unary predicates P_1, P_2 and a new binary predicate R, with $P_1 = \{a_{t,i} : (t, i) \in X \times \kappa_t\} \cup A$, with $P_2 = A$, and $R = \{(a_{t,i}, a_{t,j}) : i < j, t \in X\}$. In a large saturated structure, the corresponding sequences are mutually indiscernible over the interpretation of P_2. Using the dependence of T, we can find a set A_0, of size at most $|T|$, such that if two tuples have the same type over A_0, then they have the same type over b. We exclude those relatively few elements $t \in X$ supporting an element in P_0, resulting in sequences which are mutually indiscernible over Ab. □

Another measure, closely related to VC dimension, is the VC density, on which it sometimes seems easier to produce uniform explicit bounds.

Definition 7.4.9. Let $\mathcal{C}$ be a family of sets, and

$$\pi_{\mathcal{C}}(n) = \max_{|A|=n} |\{c \cap A : c \in \mathcal{C}\}| .$$

Then the VC-density of $\mathcal{C}$, denoted $\delta_{VC}(\mathcal{C})$, is defined as

$$\limsup_{n\to\infty} \frac{\log(\pi_{\mathcal{C}}(n))}{\log(n)}.$$

The relation of the function $\pi_{\mathcal{C}}$ to $\Pi_{\mathcal{C}}(S)$ in definition 7.3.8 is transparent. We note that the VC-density of a family of sets is finite if and only if the VC dimension is finite. While explicit global bounds on VC dimensions of formulas in a theory are difficult to obtain, some bounds on density seem simpler. Karpinski and Macintyre prove the following result, which they attribute to earlier unpublished work of Wilkie [**293**].

Proposition 7.4.10. *Let $\varphi(\bar{x}; \bar{y})$ be a formula in an o-minimal theory on $\mathbb{R}$, and $\mathcal{C}$ the family of sets defined by instantiating different parameters for $\bar{y}$. Then $\delta_{VC}(\mathcal{C}) \leq |\bar{y}|$.*

A later series of papers by Aschenbrenner, Dolich, Haskell, Macperson, and Starchenko [**34, 33**] undertakes the systematic version of this: a uniform bound on the VC-density of formulas in an NIP theory. They weaken the hypotheses of this last result to an arbitrary weakly o-minimal theory, and make similar calculations for the p-adic fields in Macintyre's language, for the Spencer-Shelah random graph described in Section 5.1.1, for certain infinite Abelian groups, and for theories of finite U-rank and lacking the finite cover property.

An important bridge between first-order logic and probability is the Keisler measure, which is particularly well-behaved in NIP theories.

Definition 7.4.11. Let T be a theory, and A a set of parameters. Then $L_x(A)$ is the algebra of sets definable in variable x with parameters from A. Then a Keisler measure over A is a finitely additive probability measure on that algebra.

If p is a 1-type over A, then a Keisler measure naturally arises that gives measure 1 to each set defined by a formula implied by the type p, and measure 0 to every other set. In this sense, a Keisler measure is a generalization of a type from the set of truth values $\{0, 1\}$ to the set $[0, 1]$, in keeping with the spirit of Chapter 2.

Under favorable circumstances, we can approximate Keisler measures by averages of points.

Proposition 7.4.12. *Let T be an NIP theory, $\mathcal{M} \models T$, and μ be a Keisler measure over M such that for every $N \supseteq M$, the measure μ has a unique extension to N. Let X be a Borel subset of $S_x(M)$, and $\varphi(x; y)$ a formula. For any $\epsilon > 0$ there are points $a_1, \ldots, a_n \in \mathcal{U}$ such that for any finite $\bar{b} \in \mathcal{U}$, we have*

$$\left| \mu\left(X \cap \varphi(x; b)\right) - \frac{1}{n} \left|\{i : \varphi(a_i; b)\}\right| \right| < \epsilon$$

.

This approximation result recalls the original work of Vapnik and Chevornenkis on the convergence of sample probabilities.

One of the more important consequences of NIP in the context of Keisler measures is the existence of invariant measures. A Keisler measure μ on a group G is said to be invariant if and only if it is translation invariant in the group G.

Theorem 7.4.13 ([**273**]). *Suppose T is NIP and G is a $\emptyset$-definable group in $\mathcal{U}$ with the property that there is some global type $p(x)$ and some model $M \models T$ such that $p(x)$ is satisfied by exactly the elements of G and every left G-translate of p is finitely satisfiable in M. Then there is a left-invariant Keisler measure on G which is finitely satisfiable in some small model M.*

We will have much more to say about invariant measures, especially in Sections 6.1 and 8.3. For now it suffices to note that certain NIP groups have such measures, giving them a property called *definable amenability.*

7.4.2. Examples of NIP Theories. Many structures of algebraic and combinatorial interest are NIP. Of course, every stable theory is NIP, as we have said. In particular, strongly minimal theories, like the theory of vector spaces are NIP.

Because the proof that stable theories are NIP is technical, it is worthwhile to see a sketch of the proof for $\mathbb{Q}$-vector spaces in its own right. Let $\varphi(\bar{x}; \bar{y})$ be a formula in the language of $\mathbb{Q}$-vector spaces. For each $\bar{b}$, the set defined by $\varphi(\bar{x}; \bar{b})$ is either an affine set or the complement of an affine set. Any infinite set S will have subsets which are not of this form. Consequently, φ cannot shatter S.

Because of Shelah's dichotomy in Theorem 7.4.6, it is natural to look for tame examples among ordered structures. The simplest ordered structures are the o-minimal structures. An ordered structure $\mathcal{M}$ is o-minimal if the only one-dimensional definable subsets are finite unions of points and intervals. The real ordered field is the canonical example of an o-minimal structure, and many others are known. It is straightforward that every o-minimal theory is NIP — intuitively, the definable sets all come from the order, and that does not give sufficient flexibility for an independent formula.

In particular, let T be an o-minimal theory with a large saturated model $\mathcal{U}$. Let $\varphi(\bar{x}, \bar{y})$ be a formula, where $|\bar{x}| + |\bar{y}| = n$. It is known from the work of Knight, Pillay, and Steinhorn in the 1980s [**318**] that for each $\bar{b}$ there is a cell decomposition of M^n; that is, a partition of M^n into finitely many definable sets $C_1, \ldots, C_k$, where each C_i is either the graph of a definable continuous function or the region between two definable continuous functions, and such that for each i, either C_i is contained in the set defined by $\varphi(\bar{x}; \bar{b})$ or disjoint from it. Since an infinite set will have many subsets that are not of this kind, φ cannot shatter an infinite set and T is NIP.

Many standard weakenings of o-minimality remain NIP. Indeed, the expansion of the real field with a predicate for any of a large family of dense subgroups of $\mathbb{R}_{>0}$ remains NIP [**86, 236, 255**].

Proposition 7.4.14 ([**237**]). *Any ordered Abelian group is NIP.*

Of course, the theory of valued fields is far from a complete theory. However, the theory of algebraically closed non-trivially valued fields, with specified characteristic for both the base field and the residue, is a complete theory.

Proposition 7.4.15 ([**421, 116**]). *Any complete theory of algebraically closed value fields is NIP.*

Proposition 7.4.16 ([**155**]). *For any prime p, the field $\mathbb{Q}_p$ is NIP.*

In general, the characterization of all NIP completions of an algebraically interesting theory is difficult, and is an area of active research. Since we have considered pseudofinite structures already in Section 5.3, it seems germane to note that we have a partial characterization of the NIP pseudofinite groups.

Theorem 7.4.17 ([**352**]). *Let G be a pseudofinite group with NIP theory, and suppose that there is some uniform bound on the length of chains*

$$C_G(F_1) < C_G(F_2) < \cdots C_G(F_n)$$

where each F_i is a subset of G and where $C_G(F_i)$ is the centralizer of F_i. Then G has a solvable definable normal subgroup of finite index.

In the same paper, though, it is shown that there is a pseudofinite group with NIP theory which is not solvable by finite.

7.4.3. Learning in NIP Theories. In 1992, Chris Laskowski published the landmark paper pointing out the relationship of the independence property to the Vapnik-Chervonenkis dimension. At this time, the connection of VC dimension to the rather new concept of PAC learnability was not widely known.

Certain special cases had already been observed. For instance, Stengle and Yukich [**468**] applied known model theory of the real ordered field to prove that in the semialgebraic sets, as well as sets definable in certain extensions of real closed fields, some uniformly definable families with a single parameter had finite VC dimension.

Laskowski, hearing of this result and its limitation to families with a single parameter, thought of a result of Shelah's on IP theories, and wrote his paper to explain the single parameter phenomenon by Shelah's "Single Variable Suffices" result.

Important to this nexus of thought is a result, apparently discovered independently in quick succession by Vapnik and Chervonenkis (implicitly) [**497**], and by Richard Dudley [**162**], Norbert Sauer [**432**], and Shelah [**447**] on a certain dichotomy. In Sauer's formulation, the result is as follows.

Theorem 7.4.18. *Let $\mathcal{C}$ be a family of subsets of an infinite set X. Then one of the following is true:*

(1) *There is some number N and some constant c such that for any $S \subseteq X$ of cardinality at least N, we have $\Pi_{\mathcal{C}} S \leq |S|^c$*
(2) *For any n there is a set $S \subseteq X$ such that $\Pi_{\mathcal{C}}(S) = 2^n$.*

I am grateful to Laskowski for explaining the history of this section.

It is worth noting that the hypothesis given by logistic regression is uniformly continuous, and so the hypothesis is definable in continuous first-order logic. At a workshop at the American Institute of Mathematics in 2006, a question was posed asking for examples of continuous NIP theories. While no proof seems to have been published, it seems likely that NIP continuous theories might stand in relation to non-sharp classifications like logistic regression similar to the relation first order NIP theories have to sharp PAC classification.

An issue to which this theorem directly gives rise is whether neural networks have finite VC dimension. Given the common assumption that what they are doing constitutes something akin to PAC learning — at least, they give a classification, and the standards for their success (as described, for instance, in [**230**]) seem similar — it would be natural to hope for finite VC dimension. In some cases, this is known.

Consider a feed-forward neural network architecture, where each node has the same activation function σ, with k inputs, m nodes, and ℓ total weights to be assigned. In a suitable expansion of the real field, we can write a $(k+\ell)$-ary formula $\Phi(\bar{x}, \bar{y})$ describing the function computed by the network. In that sense, the network represents the family of definable sets $\mathcal{C}_\Phi = \left\{\Phi(\bar{x}, \bar{\beta}) : \bar{\beta} \in \mathbb{R}^\ell\right\}$, and the VC dimension of the network is that of $\mathcal{C}_\Phi$. Locally, the computation at each node is described by a formula $\tau(\bar{x}, \bar{y})$, which will describe whether the node is activated or not. Of course, the form of τ, and the "suitable expansion" of $(\mathbb{R}, +, \cdot, 0, 1, <)$ will be determined by the activation function σ. We may take Φ to be a quantifier-free formula which is a Boolean combination of formulas of the forms $\tau(\bar{x}, \bar{y}) > 0$ and $\tau(\bar{x}, \bar{y}) = 0$.

Theorem 7.4.19 ([**292**]). *Let $(\bar{\alpha}_i : i \leq V)$ be a sequence of elements of $\mathbb{R}^k$, and $(\tau_i : i \leq s)$ be a sequence of smooth functions from $\mathbb{R}^{k+\ell} \to \mathbb{R}$. Let $(\Theta_i : i \leq r)$ be some collection of ℓ-ary functions that result from substituting some $\bar{\alpha}_j$ in the first k arguments of a τ_k, and define $F : \mathbb{R}^k \to \mathbb{R}^r$ by $F(\bar{y}) = (\Theta_1(\bar{y}), \ldots, \Theta_r(\bar{y}))$.*

Finally assume that there is a constant bound B such that if $F^{-1}(\bar{z})$ is an $(\ell - r)$-dimensional smooth manifold, then it has at most B connected components.

Then a feed-forward neural network determined by $(\tau_i : i \leq s)$ in the way described in the previous paragraph has

$$\dim_{VC}(\mathcal{C}_\Phi) \leq 2\log_2 B + (16 + 2\log_2 s)\ell.$$

In particular, the VC dimension is finite.

Certainly the Karpinski-Macintre result covers many important cases, in particular the activation function $\sigma(y) = \frac{1}{1+e^{-y}}$. On the other hand, Sontag [**458**] showed that there are neural networks of infinite VC dimension.

7.4.4. Learning in Other Theories. Of course, combinatorial dichotomies are the standard stock-in-trade of model theorists, and it is reasonable to expect that other model-theoretic dividing lines would correspond to other learning models. One such instance is the proof by Chase and Freitag [**115**] that stability corresponds to an online variant of PAC learning.

The basic setting of online learning dates back to a paper of Littlestone [**340**], and has a more familiar exposition in [**58**]. In one form, we consider an instance space X and a concept class $\mathcal{C}$, as before. The learning consists of a fixed number n of rounds. In round i, the learner is given some $x_i \in X$, and the learner outputs $L(x_i) \in \{0,1\}$, with the goal of minimizing $E_L(\bar{x}) = |\{i : L(i) \neq \chi_X(i)\}|$, and in particular with the goal of minimizing $E_L(\bar{x})$ as $\bar{x}$ ranges over all possible sequences.

There is a dimension here which plays a role in online learning analogous to that played by VC dimension in PAC learning.

Definition 7.4.20. [**340**] Let $\mathcal{C}$ be a concept class on an instance space X.

(1) We define the *thicket shatter function* $p_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ by letting $p_{\mathcal{C}}(n)$ in the following way: let T be a binary tree whose non-leaf nodes are labeled by elements of X and whose leaves are labeled by elements of $\mathcal{C}$, in such a way that the path from the root to C codes membership in C of the elements labeling the nodes along the path. Then $p_{\mathcal{C}}(n)$ is the maximum number of leaves on a binary tree of height n constructed in this way.

(2) The *Littlestone dimension of* $\mathcal{C}$, denoted $\dim_L \mathcal{C}$ is the greatest integer n such that $p_{\mathcal{C}}(n) = 2^n$, if such an integer exists, and infinite otherwise.

Littlestone showed that online learnability is equivalent to this dimension being finite. Siddharth Bhaskar [**69**] noticed that this rank is equal to the Shelah 2-rank. In the context of a theory, the treatment is similar to the NIP case: we take a formul $\phi(\bar{x}, \bar{y})$, and let $\mathcal{C} = \{\varphi(\bar{x}, \bar{a}) : \bar{a} \in \mathcal{M}^n\}$.

Definition 7.4.21 (Definition II.1.1 in [**448**] with $\lambda = 2$). Let φ be a formula. The *2-rank* of φ over a set of formulas Δ is defined inductively as follows:

(1) $R(p, \Delta) \geq 0$ if and only if φ is consistent.

(2) $R(p, \Delta) \geq \alpha + 1$ if and only if there are m-ary formulas ψ_0, ψ_1 with the following properties:

   (a) Each Ψ_0 is either and element of Δ or the negation of an element of Δ,

   (b) ψ_0 is equivalent to $\neg\psi_1$, and

   (c) For each i, we have $R(\varphi \cup \psi_i, \Delta) \geq \alpha$.

(3) $R(\varphi, \Delta) \geq \delta$ for a limit ordinal δ if and only if $R(\varphi, \Delta) \geq \alpha$ for all $\alpha < \delta$.

A formula is stable if and only if its 2-rank is finite. Consequently, online learnability of the definable concept classes of a theory is equivalent to stability.

More recently, Alon, et al. have shown that a privacy-preserving notion of learnability is also equivalent to finite 2-rank, and so to online learning [**19**]. It is tempting to think that there may be other standard model-theoretic thresholds that correspond to notions of learning.

Bibliography

1. M. Abért, N. Bergeron, I. Biringer, T. Genander, N. Nikolov, J. Raimbault, and I. Samet, *On the grown tof L^2-invariants for sequences of lattices in Lie groups*, Annals of Mathematics **185** (2017), 711–790.
2. M. Abért, Y. Glasner, and B. Virág, *Kesten's theorem for invariant random subgroups*, Duke Mathematical Journal **163** (2014), 465–488.
3. N. Ackerman, C. Freer, A. Kwiatowska, and R. Patel, *A classification of orbits admitting a unique invariant measure*, Annals of Pure and Applied Logic **168** (2017), 19–36.
4. N. Ackerman, C. Freer, J. Nešetřil, and R. Patel, *Invariant measures via inverse limits of finite structures*, European Journal of Combinatorics **52** (2016), 248–289.
5. N. Ackerman, C. Freer, and R. Patel, *Invariant measures concentrated on countale structures*, Forum of Mathematics, Sigma **4** (2016), 1–59.
6. ———, *Countable infinitary theories admitting an invariant measure*, preprint, 2017.
7. ———, *The entropy function of an invariant measure*, Proceedings of the 14th and 15th Asian Logic Conferences, World Scientific Publishing, 2019, pp. 3–34.
8. N. Ackerman, C. Freer, and D. Roy, *On the computability of conditional probability*, Journal of the Association for Computing Machinery **66** (2019), 23:1–23:40.
9. ******E. W. Adams, *The logic of conditionals*, Reidel, 1975.
10. E. W. Adams, *A primer of probability logic*, CSLI Lecture Notes, no. 68, CSLI Publications, 1999.
11. S. Adams and A. S. Kechris, *Linear algebraic groups and countable Borel equivalence relations*, Journal of the American Mathematical Society **13** (2000), 909–943.
12. Scott Adams, *Dilbert*, (2001), October 25, `https://dilbert.com/strip/2001-20-25`.
13. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular biology of the cell*, 5th ed., Garland, 2008.
14. M. Aldana, S. Coppersmith, and L. P. Kadanoff, *Boolean dynamics with random couplings*, Perspectives and Problems in Nonlinear Science, Springer, 2003, pp. 23–89.
15. D. Aldous, *A conjectured compactification of some finite reversible Markov chains*, Lecture notes for a lecture at the Courant Institute, 2012.
16. D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, The American Mathematical Monthly **93** (1986), 333–348.
17. K. Allen, L. Bienvenu, and T. A. Slaman, *On zeros of Martin-Löf random Brownian motion*, Journal of Logic and Analysis **6** (2014), 1–36.
18. J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge, 2003.
19. N. Alon, M. Bun, R. Livni, M. Malliaris, and S. Moran, *Private and online learnability are equivalent*, Journal of the ACM **69** (2022), 28.
20. N. Alon, R. A. Duke, H. Lefmann, V. Rödel, and R. Yuster, *The algorithmic aspects of the regularity lemma*, Journal of Algorithms **16** (1994), 80–109.
21. N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, 2008.
22. R. Alvir, W. Calvert, G. Goodman, V. Harizanov, J. Knight, A. Morozov, R. Miller, A. Soskova, and R. Weisshaar, *Interpreting a field in its Heisenberg group*, Journal of Symbolic Logic **87** (2022), 1215–1230.
23. J. J. Andrews and M. L. Curtis, *Free groups and handlebodies*, Proceedings of the American Mathematical Society **16** (1965), 192–195.
24. U. Andrews, I. Goldbring, and H. J. Keisler, *Definable closure in randomizations*, Annals of Pure and Applied Logic **166** (2015), 325–341.
25. ———, *Independence in randomizations*, Journal of Mathematical Logic **19** (2019), 1950005.

26. U. Andrews and H. J. Keisler, *Separable models of randomizations*, Journal of Symbolic Logic **80** (2015), 1149–1181.
27. U. Andrews, S. Lempp, J. S. Miller, K. M. Ng, L. San Mauro, and A. Sorbi, *Universal computably enumerable equivalence relations*, The Journal of Symbolic Logic **79** (2014), 60–88.
28. U. Andrews and A. Sorbi, *The complexity of index sets of classes of computably enumerable equivalence relations*, The Journal of Symbolic Logic **81** (2016), 1375–1395.
29. A. Arana, *Logical and semantic purity*, Protosociology **25** (2008), 36–48.
30. A. Arnould and P. Nicole, *The Port Royal Logic*, Gordon, 1861.
31. S. Arora and B. Barak, *Computational complexity*, Cambridge, 2009.
32. Е. А. Асарин and А. В. Покровский, Применение колмогоровской сложности к анализу динамики упровляемых систем, Автоматика и Телемеханика **1** (1986), 25–33.
33. M. Aschenbrenner, A. Dolich, D. Haskell, D. Macpherson, and S. Starchenko, *Vapnik-Chervonenkis density in some theories without the independence property, II*, Notre Dame Journal of Formal Logic **54** (2013), 311–363.
34. ———, *Vapnik-Chervonenkis density in some theories without the independence property, I*, Transactions of the American Mathematical Society **368** (2016), 5889–5949.
35. C. J. Ash and J. F. Knight, *Computable structures and the hyperarithmetical hierarchy*, Studies in Logic and the Foundations of Mathematics, vol. 144, Elsevier, 2000.
36. K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo, *Effective strong dimension in algorithmic information and computational complexity*, SIAM Journal on Computing **37** (2007), 671–705.
37. J. Avigad, *Inverting the Furstenberg correspondence*, Discrete and Continuous Dynamical Systems **32** (2012), 3421–3431.
38. J. Avigad, J. Hözl, and L. Serafin, *A formally verified proof of the central limit theorem*, Journal of Automated Reasoning **59** (2017), 389–423.
39. J. Ax, *The elementary theory of finite fields*, Annals of Mathematics **85** (1968), 239–271.
40. L. Babai, *Trading group theory for randomness*, STOC '85: Proceedings of the seventeenth annual ACM symposium on Theory of Computing, 1985, pp. 421–429.
41. A. Baker, *Transcendental number theory*, Cambridge Mathematical Library, Cambridge, 1990.
42. J. T. Baldwin and S. Shelah, *Randomness and semigenericity*, Transactions of the American Mathematical Society **349** (1997), 1359–1376.
43. S. Banach, *Sur le problème de la measure*, Fundamenta Mathematicae **4** (1923), 7–33.
44. A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), 509–512.
45. G. Barmpalias, D. Cenzer, and C. P. Porter, *The probability of a computable output from a random oracle*, ACM Transactions on Computational Logic **18** (2017), 18.
46. ———, *Random numbers as probabilities of machine behavior*, Theoretical Computer Science **673** (2017), 1–18.
47. George Barmpalias and Andrew Lewis-Pye, *Differences of halting probabilities*, Journal of Computer and System Sciences **89** (2017), 349–360.
48. L. Barreira, *Dimension and recurrence in hyperbolic dynamics*, Progress in Mathematics, no. 272, Birkhäuser, 2008.
49. L. Bartholdi, *Counting paths in groups*, L'Enseignement Mathématique **45** (1999), 83–131.
50. N. A. Bazhenov and B. S. Kalmurzaev, *On dark computably enumerable equivalence relations*, Siberian Mathematical Journal **59** (2018), 22–30.
51. V. Becher, Y. Bugeaud, and T. A. Slaman, *On simply normal numbers to different bases*, Mathematische Annalen **364** (2016), 125–150.
52. V. Becher, O. Carton, and P. A. Heiber, *Normality and automata*, Journal of Computer and System Sciences **81** (2015), 1592–1613.
53. V. Becher and P. A. Heiber, *Normal numbers and finite automata*, Theoretical Computer Science **477** (2013), 109–116.
54. J. Beck, *An algorithmic approach to the Lovász local lemma I*, Random Structures and Algorithms **2** (1991), 343–365.
55. H. Becker and A. S. Kechris, *Borel actions of Polish groups*, Bulletin of the American Mathematical Society **28** (1993), 334–341.

56. ______, *The descriptive set theory of Polish group actions*, London Mathematical Society Lecture Note Series, no. 232, Cambridge, 1996.
57. O. Becker, A. Lubotzky, and A. Thom, *Stability and invariant random subgroups*, Duke Mathematical Journal **168** (2019), 2207–2234.
58. S. Ben-David, D. Pál, and S. Shalev-Shwartz, *Agnostic online learning*, Conference on Learning Theory (COLT), 2009.
59. I. Ben Yaacov, *Schrödinger's cat*, Israel Journal of Mathematics **153** (2006), 157–191.
60. ______, *Continuous and random Vapnik-Chervonenkis classes*, Israel Journal of Mathematics **173** (2009), 309–333.
61. ______, *On theories of random variables*, Israel Journal of Mathematics **194** (2013), 957–1012.
62. I. Ben Yaacov, A. Berenstein, C. W. Henson, and A. Usvyatsov, *Model theory for metric structures*, Model theory with applications to algebra and analysis, vol. 2, London Mathematical Socieity Lecture Note Series, no. 350, Cambridge, 2008, pp. 315–429.
63. I. Ben Yaacov and H. J. Keisler, *Randomizations of models as metric structures*, Confluentes Mathematici **1** (2009), 197–223.
64. I. Ben Yaacov and A. P. Pedersen, *A proof of completeness for continuous first-order logic*, Journal of Symbolic Logic (2010), 168–190.
65. I. Ben Yaacov and A. Usvyatsov, *Continuous first order logic and local stability*, Transactions of the American Mathematical Society **362** (2010), 5213–5259.
66. C. Bernardi and A. Sorbi, *Classifying positive equivalence relations*, The Journal of Symbolic Logic **48** (1983), 529–538.
67. A. Beros, *Learning theory in the arithmetic hierarchy*, Journal of Symbolic Logic **79** (2014), 908–927.
68. Ö. Beyarslan, *Random hypergraphs in pseudofinite fields*, Journal of the Institute of Mathematics of Jussieu **9** (2010), 29–47.
69. S. Bhaskar, *Thicket density*, Journal of Symbolic Logic **86** (2021), 110–127.
70. L. Bienvenu, *Game-theoretic approaches to randomness: unpredictability and stochasticity*, Ph.D. thesis, Université de Provence, 2008.
71. L. Bienvenu, A. Day, M. Hoyrup, I. Mezhirov, and A. Shen, *A constructive version of Birkhoff's ergodic theorem for Martin-Löf random points*, Information and Computation **210** (2012), 21–30.
72. C. M. Bishop, *Pattern recognition and machine learning*, Information Science and Statistics, Springer, 2006.
73. L. Blum and M. Blum, *Toward a mathematical theory of inductive inference*, Information and control **28** (1975), 125–155.
74. A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, *Learnability and the Vapnik-Chervonenkis dimension*, Journal of the ACM **36** (1989), 929–965.
75. B. Bollobás, *Random graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics, no. 73, Cambridge, 2001.
76. G. Boole, *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*, Macmillan, 1851.
77. W. W. Boone, *Certain simple, unsolvable problems of group theory V, VI*, Indagationes Mathematicae **60** (1957), 22–27, 227–232.
78. ______, *The word problem*, Proceedings of the National Academy of Sciences of the USA **44** (1958), 1061–1065.
79. M. Borda, *Fundamentals in information theory and coding*, Springer, 2011.
80. A. Borel, *Density properties for certain subgroups of semi-simple groups wihtout compact components*, Annals of Mathematics **72** (1960), 179–188.
81. M. E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Matematico di Palermo (1884–1940) **27** (1909), 247–271.
82. C. Borgs, J. T. Chayes, L. Lov'asz, V. T. Sos, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Advances in Mathematics **219** (2008), 1801–1851.
83. C. Borgs, J. T. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi, *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Advances in Mathematics **219** (2008), 1801–1851.

84. K. H. Borgwardt, *The simplex method: A probabilistic analysis*, Algorithms and Combinatorics, no. 1, Springer, 1987.
85. L. Bowen, *Invariant random subgroups of the free group*, Groups, Geometry, and Dynamics **9** (2015), 891–916.
86. G. Boxall, *NIP for some pair-like theories*, Archive for Mathematical Logic **50** (2011), 353–359.
87. M. Braverman, *Parabolic Julia sets are polynomial time computable*, Nonlinearity **19** (2006), 1383–1401.
88. M. Braverman and M. Yampolsky, *Computability of julia sets*, Algorithms and Computation in Mathematics, no. 23, Springer, 2009.
89. J. Brody and M. C. Laskowski, *Rational limits of Shelah-Spencer graphs*, The Journal of Symbolic Logic **77** (2012), 580–592.
90. T. A. Brown, T. H. McNicholl, and A. G. Melnikov, *On the complexity of classifying Lebesgue spaces*, Journal of Symoblic Logic **85** (2020), 1254–1288.
91. Y. Bugeaud, *Distribution modulo one and Diophantine approximation*, Cambridge Tracts in Mathematics, no. 193, Cambridge, 2012.
92. S. Buss and M. Minnes, *Probabilistic algorithmic randomness*, The Journal of Symbolic Logic **78** (2013), 579–601.
93. D. Cai, N. Ackerman, and C. Freer, *An iterative step-function estimator for graphons*, preprint, 2015.
94. W. Calvert, *Metric structures and probabilistic computation*, Theoretical Computer Science **412** (2011), 2766–2775.
95. ———, *PAC learning, VC dimension, and the arithmetic hierarchy*, Archive for Mathematical Logic **54** (2015), 871–883.
96. W. Calvert, D. Cenzer, D. Gonzalez, and V. Harizanov, *Generically computable linear orderings*, Preprint, 2024.
97. W. Calvert, D. Cenzer, and V. Harizanov, *Densely computable structures*, Journal of Logic and Computation **32** (2022), 581–607.
98. ———, *Generically and coarsely computable isomorphisms*, Computability **11** (2022), 223–239.
99. ———, *Generically computable Abelian groups*, Unconventional Computation and Natural Computation, Lecture Notes in Computer Science, no. 14003, Springer, 2023, pp. 32–45.
100. W. Calvert, E. Gruner, E. Mayordomo, D. Turetsky, and J. D. Villano, *On the computable dimension of real numbers: normality, relativization, and randomness*, preprint, 2025.
101. W. Calvert, V. Harizanov, and A. Shlapentokh, *Turing degrees of isomorphism types of algebraic objects*, The Journal of the London Mathematical Society **75** (2007), 273–286.
102. ———, *Computability in infinite Galois theory and algorithmically random algebraic fields*, Journal of the London Mathematical Society **110** (2024), 370017.
103. W. Calvert and J. F. Knight, *Classification from a computable viewpoint*, Bulletin of Symbolic Logic **12** (2006), 191–219.
104. P. J. Cameron, *Transitivity of permutation groups on unordered sets*, Mathematische Zeitschrift **148** (1976), 127–139.
105. C. Camrud, *Generalized effective completeness for continuous logic*, Journal of Logic & Analysis **15** (2023), 1–17.
106. C. Carathéodory, *Über das lineare Mass von Punktmengen — eine Verallgemeinerung des Längenbegriffs*, Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-phisikalisch Klasse **1914** (1914), 404–426.
107. J. Case, S. Jain, and F. Stephan, *Effectivity questions for Kleene's recursion theorem*, Theoretical Computer Science **733** (2018), 55–70.
108. J. Case and C. Lynes, *Machine inductive inference and language identification*, International Colloquium on Automata, Languages, and Programming, 1982, pp. 107–115.
109. T. Ceccherini-Silberstein and M. Coornaert, *Cellular automata and groups*, Springer Monographs in Mathematics, Springer, 2010.
110. D. Cenzer, Π^0_1 *classes in computability theory*, Handbook of Computability, Studies in Logic and the Foundations of Mathematics, no. 140, Elsevier, 1999, pp. 37–85.
111. C. Chabauty, *Limite d'ensembles et géométrie des nombres*, Bulletin de la S. M. F. **78** (1950), 143–151.

112. G. J. Chaitin, *A theory of program size formally identical to information theory*, Journal of the Association for Computing Machinery **22** (1975), 329–340.
113. ______, *Algorithmic information theory*, Cambridge Tracts in Theoretical Computer Science, no. 1, Cambridge, 1987.
114. A. S. Charles, *Interpreting deep learning: The machine learning Rorschach test?*, SIAM News **51** (2018), no. 6, 1.
115. H. Chase and J. Freitag, *Model theory and machine learning*, Bulletin of Symbolic Logic **25** (2019), 319–332.
116. Z. Chatzidakis, *Théorie des modèles des corps valués*, lecture notes, 2008.
117. Z. Chatzidakis, L. van den Dries, and A. Macintyre, *Definable sets over finite fields*, Journal für die reine und angewandte Mathematik **427** (1992), 107–135.
118. G.-Y. Chen and L. Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, Electronic Journal of Probability **13** (2008), 26–78.
119. ______, *The L^2-cutoff for reversible Markov processes*, Journal of Functional Analysis **258** (2010), 2246–2315.
120. R. Chen, *Borel functors, interpretations, and strong conceptual completeness for $L_{\omega_1\omega}$*, Transactions of the American Mathematical Society **372** (2019), 8955–8983.
121. G. Cherlin and E. Hrushovski, *Finite structures with few types*, Annals of Mathematics Studies, no. 152, Princeton University Press, 2003.
122. A. Chernikov and S. Starchenko, *Definable regularity lemmas for NIP hypergraphs*, Quarterly Journal of Mathematics **72** (2021), 1401–1433.
123. N. Chomsky, *Three models for the description of language*, IRE Transactions on Information Theory **2** (1956), 113–124.
124. ______, *Knowledge of language: Its nature, origin, and use*, Convergence, Praeger Scientific, 1986.
125. F. Chung, *On concentrators, superconcentrators, generalizers, and nonblocking networks*, The Bell System Technical Journal **58** (1978), 1765–1777.
126. F. Chung and L. Lu, *Complex graphs and networks*, CBMS Regional Conference Series in Mathematics, no. 107, American Mathematical Society, 2006.
127. F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, *Duplication models for biological networks*, Journal of Computational Biology **10** (2003), 677–687.
128. B. Cisma, D. D. Dzhafarov, D. R. Hirschfeldt, C. G. Jockusch, R. Solomon, and L. B. Westrick, *The reverse mathematics of Hindman's theorem for sums of exactly two elements*, Computability **8** (2019), 253–263.
129. K. J. Compton, *Laws in logic and combinatorics*, Algorithms and Order, NATO ASI Series C, no. 255, Kluwer, 1989, pp. 353–383.
130. G. Conant and A. Pillay, *Pseudofinite groups and VC-dimension*, Journal of Mathematical Logic **21** (2021), 2150009.
131. A. Condon, *The complexity of stochastic games*, Information and Computation **96** (1992), 203–224.
132. C. T. Conley, A. S. Kechris, and B. D. Miller, *Stationary probability measures and topological realizations*, Israel Journal of Mathematics **198** (2013), 333–345.
133. C. T. Conley and B. D. Miller, *Measure reducibility of countable Borel equivalence relations*, Annals of Mathematics **185** (2017), 347–402.
134. D. Conlon and J. Fox, *Bounds for graph regularity and removal lemmas*, Geometric and Functional Analysis **22** (2012), 1191–1256.
135. A. Connes and B. Weiss, *Property T and asymptotically invariant sequences*, Israel Journal of Mathematics **37** (1980), 209–210.
136. S. D. Conte and C. de Boor, *Elementary numerical analysis*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill, 1980.
137. O. Cooley, W. Fang, D. Del Giudice, and M. Kang, *Subcritical random hypergraphs, high-order components, and hypertrees*, 2019 Proceedings of the Sixteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO), 2019, pp. 111–118.
138. M. Coornaert, *Topoligcal dimension and dynamical systems*, Universitext, Springer, 2015.
139. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, 3rd ed., MIT Press, 2009.
140. R. T. Cox, *Probability, frequency, and reasonable expectation*, American Journal of Physics **14** (1946), 1–13.

141. B. F. Csima, V. S. Harizanov, R. Miller, and A. Montalban, *Computability of Fraïssé limits*, Journal of Symbolic Logic **76** (2011), 66–93.
142. R. W. R. Darling and J. R. Norris, *Structure of large random hypergraphs*, The Annals of Applied Probability **15** (2005), 125–152.
143. A. P. Dawid, *Probability, causality, and the empirical world: a Bayes-de Finetti-Popper-Borel synthesis*, Statistical Science **10** (2004), 44–57.
144. B. de Finetti, *Funzione caratteristica di un fenomeno aleatorio*, Memorie della R. Accademia dei Lincei **4** (1930), 86–133.
145. ______, *Foresight: its logical laws, its subjective sources*, Breakthroughs in Statisitcs, Springer, 1992, Translated by H. E. Kyberg, Jr.; Original published 1937, pp. 134–174.
146. ______, *A translation of 'The characteristic function of a random phenomenon' by Bruno de Finetti*, D. Alvarez-Melis and T. Broderick, translators. arXiv:1512.01229, 2015.
147. K. de Leeuw, E. F. Moore, C. E. Shannon, and N. Shapiro, *Computability by probabilistic machines*, Automata Studies, Annals of Mathematics Studies, no. 34, Princeton, 1956, pp. 183–212.
148. M. Dehn, *Über unendliche diskontinuierliche Gruppen*, Mathematische Annalen **71** (1911), 116–144.
149. ______, *Transformation der kurven auf zweiseitigen flächen*, Mathematische Annalen **72** (1912), 413–421.
150. A. P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics **38** (1967), 325–339.
151. O. Demuth, *On constructive pseudonumbers*, Commentationes Mathematicae Universitatis Carolinae **16** (1975), 315–331.
152. M. Detlefsen and A. Arana, *Purity of methods*, Philosophers' Imprint **11** (2011), 1–20.
153. P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs*, Rendiconti di Matematica, Serie VII **28** (2008), 33–61.
154. A. Ditzen, *Definable equivalence relations on Polish spaces*, Ph.D. thesis, California Institute of Technology, 1992.
155. A. Dolich, D. Lippel, and J. Goodrick, *dp-minimal theories: basic facts and examples*, Notre Dame Journal of Formal Logic **52** (2011), 267–288.
156. M. D. Donsker, *An invariance principle for certain probability limit theorems*, Memoirs of the American Mathematical Society **6** (1951), 1–12.
157. R. Dougherty, S. Jackson, and A. S. Kechris, *The structure of hyperfinite Borel equivalence relations*, Transactions of the American Mathematical Society **341** (1994), 193–225.
158. R. G. Downey and E. J. Griffiths, *Schnorr randomness*, The Journal of Symbolic Logic **69** (2004), 533–554.
159. R. G. Downey and D. R. Hirschfeldt, *Algorithmic randomness and complexity*, Theory and Applications of Computability, Springer, 2010.
160. R. G. Downey, C. G. Jockusch Jr., and P. E. Schupp, *Asymptotic density and computably enumerable sets*, Journal of Mathematical Logic **13** (2013), 1350005.
161. A. Dudko and M. Yampolsky, *On computational complexity of Cremer Julia sets*, Fundamenta Mathematicae **252** (2021), 343–353.
162. R. M. Dudley, *Central limit theorems for empirical measures*, The annals of probability **6** (1978), 899–929.
163. J.-L. Duret, *Les corps faiblement algebriquement clos non separablement clos ont la properiete d'independance*, Model Theory of Algebra and Arithmetic, Lecture Notes in Mathematics, no. 834, Springer, 1980, pp. 136–162.
164. M. Džamonja and I. Tomašić, *Graphons arising from graphs definable over finite fields*, Colloquium Mathematicum **169** (2022), 269–305.
165. P. D. Eastman, *Are you my mother?*, Random House, 1960.
166. H.-D. Ebbinghaus and J. Flum, *Finite model theory*, 2nd ed., Springer Monographs in Mathematics, Springer, 2006.
167. G. Edgar, *Measure, topology, and fractal geometry*, second ed., Undergraduate Texts in Mathematics, Springer, 2008.
168. H. G. Eggleston, *Sets of fractional dimensions which occur in some problems of number theory*, Proceedings of the London Mathematical Society **54** (1952), 42–93.
169. K. Eickmeyer and M. Grohe, *Randomisation and derandomisation in descriptive complexity theory*, Logical Methods in Computer Science **7** (2011), 1–24.

170. G. Elek and B. Szegedy, *A measure-theoretica approach to the theory of dense hypergraphs*, Advances in Mathematics **231** (2012), 1731–1772.
171. R. Elwes, *Asymptotic classes of finite structures*, Journal of Symbolic Logic **72** (2007), 418–438.
172. H. B. Enderton, *A mathematical introduction to logic*, Academic Press, 1972.
173. I. Epstein, *Orbit inequivalent actions of non-amenable groups*, preprint, 2008.
174. P. Erdős, D. J. Kleitman, and B. L. Rothschild, *Asymptotic enumeration of K_n-free graphs*, Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), vol. 2, Acad. Naz. Lincei, 1976, pp. 19–27.
175. P. Erdős and L. Lovász, *Problems and results on 3-chromatic hypergraphs and some related questions*, Infinite and Finite Sets, vol. II, Colloq. Math. Soc. János Bolyai, no. 10, North-Holland, 1975, pp. 609–627.
176. P. Erdős and A. Rényi, *On the evolution of random graphs*, Matematikai Kutató Intézet Közelményei **A** (1960), 17–60.
177. P. Erdős and A. Rényi, *On random graphs I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297.
178. Ю. Л. Ершов, Позитивные Еквивалентности, Алгебра и Логика **10** (1971), 620–650.
179. R. Fagin, *Generalized first-order spectra and polynomial-time recognizable sets*, SIAM–AMS Proceedings, vol. 7, 1974.
180. ———, *Probabilities on finite models*, The Journal of Symbolic Logic **41** (1976), 50–58.
181. S. Fajardo and H. J. Keisler, *Model theory of stochastic processes*, Lecture Notes in Logic, no. 14, A K Peters, 2002.
182. K. Falconer, *Fractal geometry: Mathematical foundations and applications*, 2nd ed., Wiley, 2003.
183. U. Felgner, *Pseudo-endliche gruppen*, Jahrbuch der Kurt-Gödel-Gesellschaft **3** (1990), 94–108.
184. E. Fokina, V. Harizanov, and D. Turetsky, *Computability-theoretic categoricity and Scott families*, Annals of Pure and Applied Logic **170** (2019), 669–717.
185. G. B. Folland, *Real analysis*, 2nd ed., Pure and Applied Mathematics, Wiley, 1999.
186. M. Foreman, D. J. Rudolph, and B. Weiss, *The conjugacy problem in ergodic theory*, Annals of Mathematics **173** (2011), 1529–1586.
187. M. Foreman and B. Weiss, *An anti-classification theorem for ergodic measure presrving transormations*, Journal of the European Mathematical Society **6** (2004), 277–292.
188. W. Fouché, *Arithmetical representations of Brownian motion I*, The Journal of Symbolic Logic **65** (2000), 421–442.
189. ———, *Martin-Löf randomness, invariant measures and countable homogeneous structures*, Theory of Computing Systems **52** (2013), 65–79.
190. W. L. Fouché, *Algorithmic randomness and Ramsey properties of countable homogeneous structures*, Logic, language, information, and computation, Lecture Notes in Computer Science, no. 7456, Springer, 2012, pp. 246–256.
191. R. Fraïssé, *Sur l'extension aux relations de quelques propriétés des ordres*, Annales scientifiques de l'É.N.S. **71** (1954), 363–388.
192. J. N. Y. Franklin, N. Greenberg, J. S. Miller, and K. M. Ng, *Martin-Löf random points satisfy Birkhoff's ergodic theorem for effectively closed sets*, Proceedings of the American Mathematical Society **140** (2012), 3623–3628.
193. J. N. Y. Franklin, M.-C. Ho, and J. Knight, *Free structures and limiting density*, preprint, 2022.
194. J. N. Y. Franklin and T. H. McNicholl, *Degrees of and lowness for isometric isomorphism*, Journal of Logic & Analysis **12** (2020), 1–23.
195. J. N. Y. Franklin and C. P. Porter (eds.), *Algorithmic randomness*, Lecture Notes in Logic, no. 50, Cambridge University Press, 2020.
196. J. N. Y. Franklin and C. P. Porter (eds.), *Algorithmic randomness: Progress and prospects*, Lecture Notes in Logic, no. 50, Cambridge, 2020.
197. C. Freer, *Computable de Finetti measures*, Annals of Pure and Applied Logic **163** (2012), 530–546.
198. M. D. Fried and M. Jarden, *Field arithmetic*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 11, Springer, 2005.

199. H. Friedman and L. Stanley, *A Borel reducibility theory for classes of countable structures*, Journal of Symbolic Logic **54** (1989), 894–914.
200. A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–220.
201. T. Fritz, *A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics*, Advances in Mathematics **370** (2020), 107239.
202. T. Fritz and E. Fjeldgren Rischel, *Infinite products and zero-one laws in categorical probability*, Compositionality **2** (2020).
203. A. Furman, *What is a stationary measure?*, Notices of the American Mathematical Soceity **58** (2011), 1276–1277.
204. H. Furstenberg, *A Poisson formula for semi-simple Lie groups*, Annals of Mathematics **77** (1963), 335–386.
205. ______, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Mathematical Systems Theory **1** (1967), 1–49.
206. ______, *A note on Borel's density theorem*, Proceedings of the American Mathematical Society **55** (1976), 209–212.
207. ______, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, Journal D'Analyse Mathématique **31** (1977), 204–256.
208. S. Gaal and L. Gál, *The discrepancy of the sequence* $\{(2^n x)\}$, Indagationes Mathematicae **26** (1964), 129–143.
209. H. Gaifman, *Concerning measures in first order calculi*, Israel Journal of Mathematics **2** (1964), 1–18.
210. S. Gao, *Invariant descriptive set theory*, Pure and Applied Mathematics, CRC Press, 2009.
211. S. Gao and P. Gerdes, *Computably enumerable equivalence relations*, Studia Logica **67** (2001), 27–59.
212. W. I. Gasarch, *The P=?NP poll*, ACM SIGACT News **33** (2002), 34–47.
213. H. Geffner and J. Pearl, *A framework for reasoning with defaults*, Tech. Report R-94, Cognitive Systems Laboratory, UCLA, 1987.
214. D. Geiger and J. Pearl, *Logical and algorithmic properties of conditional independence and graphical models*, Annals of Statistics **21** (1993), 2001–2021.
215. T. Gelander, *Lecture notes on invariant random subgroups and lattices in rank one and higher rank*, preprint, 2015.
216. ______, *Kazhdan-Margulis theorem for invariant random subgroups*, Advances in Mathematics **327** (2018), 47–51.
217. G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, *Random hypergraphs and their applications*, Physical Review E **79** (2009), 066118–1–066118–10.
218. J. Gill, *Computational complexity of probabilistic Turing machines*, SIAM Journal on Computing **6** (1977), 675–695.
219. E. Glasner and B. Wiess, *Minimal actions of the group* $\mathbb{S}(\mathbb{Z})$ *of permutations of the integers*, Geometric and Functional Analysis **12** (2002), 964–988.
220. Yu. V. Glebskii, D. I. Kogan, M. I. Kogonkii, and V. A. Talanov, *Volume and fraction of satisfiability of formulas of the lower predicate calculus*, Kibernetika (Kiev) (1969), 17–27.
221. B. Goertzel, M. Iklé, I. F. Goertzel, and A. Heljakka, *Probabilistic logic networks*, Springer, 2009.
222. E. Mark Gold, *Language identification in the limit*, Information and Control **10** (1967), 447–474.
223. I. Goldbring and B. Hart, *Computability and the Connes Embedding Problem*, Bulletin of Symbolic Logic **22** (2016), 238–248.
224. I. Goldbring and V. C. Lopes, *Pseudofinite and pseudocompact metric structures*, Notre Dame Journal of Formal Logic **56** (2015), 493–510.
225. I. Goldbring and H. Towsner, *An approximate logic for measures*, Israel Journal of Mathematics **199** (2014), 867–913.
226. O. Goldreich, *A primer on pseudorandom generators*, University Lecture Series, no. 55, American Mathematical Society, 2010.
227. O. Goldreich, S. Micali, and A. Wigderson, *Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems*, Journal of the Association for Computing Machinery **38** (1991), 691–729.

228. S. Goldwasser, S. Micali, and C. Rackoff, *The knowledge complexity of interactive proof systems*, STOC '85: Proceedings of the seventeenth annual ACM symposium on Theory of Computing, 1985, pp. 291–304.
229. С. С. Гончаров and Ю. Л. Ершов, Конструктивные Модели, Сибирская Школа Алгебры и Логики, Научная Книга, 1999.
230. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, Adaptive Computation and Machine Learning, MIT Press, 2016.
231. D. Gorenstein, *Classifying the finite simple groups*, Bulletin of the American Mathematical Society **14** (1986), 1–98.
232. E. Grädel, P. G. Kolaitis, L. Libkin, M. Marx, J. Spencer, M. Y. Vardi, Y. Venema, and S. Weinstein, *Finite model theory and its applications*, Texts in Theoretical Computer Science, Springer, 2007.
233. N. Greenberg, J. S. Miller, A. Shen, and L. Brown Westrick, *Dimension 1 sequences are close to randoms*, Theoretical Computer Science **705** (2018), 99–112.
234. M. Gromov, *Random walk in random groups*, Geometric and Functional Analysis **13** (2003), 73–146.
235. Y. Guivarc'h, *Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire*, Astérisque **74** (1980), 47–98.
236. A. Günaydin and P. Hieronymi, *Dependent pairs*, Journal of Symbolic Logic **76** (2011), 377–390.
237. Y. Gurevich and P. H. Schmitt, *The theory of ordered Abelian groups does not have the independence property*, Transactions of the American Mathematical Society **284** (1984), 171–182.
238. I. Hacking, *The emergence of probability*, 2nd ed., Cambridge, 2006.
239. R. Haenni and N. Lehmann, *Probabilistic artumentation systems: a new perspecive on the Dempster-Shafer theory*, International Journal of Intelligent Systems **18** (2003), 93–106.
240. R. Haenni, J.-W. Romeijn, G. Wheeler, and J. Williamson, *Probabilistic logics and probabilistic networks*, Synthese Library, vol. 350, Springer, 2011.
241. T. Hailperin, *Sentential probability logic*, Lehigh University Press, 1996.
242. J. Y. Halpern, *A counterexample to theorems of Cox and Fine*, Journal of Artificial Intelligence Research **10** (1999), 67–85.
243. ______, *Cox's theorem revisited*, Journal of Artifical Intelligence Research **11** (1999), 429–435.
244. ______, *Reasoning about uncertainty*, MIT Press, 2003.
245. V. S. Harizanov, *Inductive inference systems for learning classes of algorithmically generated sets and structures*, Induction, Algorithmic Learning Theory, and Philosophy (M. Friend, N. B. Goethe, and V. S. Harizanov, eds.), Logic, Epistemology, and the Unity of Science, no. 9, Springer, 2007, pp. 27–54.
246. V. S. Harizanov and F. Stephan, *On the learnability of vector spaces*, Journal of Computer and System Sciences **73** (2007), 109–122.
247. L. A. Harrington, A. S. Kechris, and A. Louveau, *A Glimm-Effros dichotomy for Borel equivalence relations*, Journal of the American Mathematical Society **3** (1990), 903–928.
248. M. Harrison-Trainor, B. Khoussainov, and D. Turetsky, *Effective aspects of algorithmically random structures*, Computability **8** (2019), 359–375.
249. M. Harrison-Trainor, A. Melnikov, R. Miller, and A. Montalbán, *Computable functors and effective interpretability*, Journal of Symbolic Logic **82** (2017), 77–97.
250. M. Harrison-Trainor, R. Miller, and A. Montalbán, *Borel functors and infinitary interpretations*, Journal of Symbolic Logic **83** (2018), 1434–1456.
251. T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, 2nd ed., Springer Series in Statistics, Springer, 2011.
252. H. Hatami and S. Norine, *The entropy of random-free graphons and properties*, Combinatorics, Probability, and Computing **22** (2013), 517–526.
253. F. Hausdorff, *Dimension und äußeres Maß*, Mathematische Annalen **79** (1918), 157–179.
254. C. W. Henson, *A family of countable homogeneous graphs*, Pacific Journal of Mathematics **38** (1971), 69–83.
255. P. Hieronymi and T. Nell, *Distal and non-distal pairs*, Journal of Symbolic Logic **82** (2017), 375–383.

256. D. Hilbert, *Lectures on the foundations of geometry, 1891–1902*, vol. 1, Springer, 2004, translation due to V. Pambuccian, Fragments of Euclidean and hyperbolic geometry, *Scientia mathematicae Japonicae* 53 (2001) pp. 361–400.
257. J. Hintikka and G. Sandu, *Game-theoretical semantics*, Handbook of Logic and Language, Amsterdam, 1997, pp. 361–410.
258. D. R. Hirschfeldt, C. G. Jockusch Jr., T. H. McNicholl, and P. E. Schupp, *Asymptotic density and the coarse computability bound*, Computability **5** (2016), 13–27.
259. D. R. Hirschfeldt, B. Khoussainov, R. A. Shore, and A. M. Slinko, *Degree spectra and computable dimensions in algebraic structures*, Annals of Pure and Applied Logic **115** (2002), 71–113.
260. J. M. Hitchcock, *Correspondence principles for effective dimensions*, Theory of Computing Systems **38** (2005), 559–571.
261. J. M. Hitchcock and J. H. Lutz, *Why computational complexity requires stricter martingales*, Theory of Computing Systems **39** (2006), 277–296.
262. G. Hjorth, *Classification and orbit equivalence relations*, Mathematical Surveys and Monographs, vol. 75, American Mathematical Society, 2000.
263. ———, *On invariants for measure preserving transformations*, Fundamenta Mathematicae **169** (2001), 51–84.
264. ———, *A converse to Dye's Theorem*, Transactions of the American Mathematical Society **357** (2005), 3083–3103.
265. ———, *Glimm-Effros for coanalytic equivalence relations*, Journal of Symbolic Logic **74** (2009), 402–422.
266. G. Hjorth and A. S. Kechris, *Analytic equivalence relations and Ulm-type classifications*, Journal of Symbolic Logic **60** (1995), 1273–1300.
267. W. Hodges, *What is a structure theory?*, Bulletin of the London Mathematical Society **19** (1987), 209–237.
268. ———, *Model theory*, Encyclopedia of Mathematics and its Applications, vol. 42, Cambridge, 1993.
269. ———, *Groups in pseudofinite fields*, Model theory of groups and automorphism groups, London Mathematical Society Lecture Note Series, no. 244, Cambridge University Press, 1997, pp. 90–109.
270. B. Host and B. Kra, *Nilpotent structures in ergodic theory*, Mathematical Surveys and Monographs, vol. 236, American Mathematical Society, 2018.
271. E. Hrushovski, *Pseudo-finite fields and related structures*, Model Theory and Applications, Quaderni di Matematica, no. 11, Aracne, 2002, pp. 151–212.
272. ———, *Stable group theory and approximate subgroups*, Journal of the American Mathematical Society **25** (2012), 189–243.
273. E. Hrushovski, Y. Peterzil, and A. Pillay, *Groups, measures, and the NIP*, Journal of the American Mathematical Society **21** (2008), 563–596.
274. E. Hrushovski and A. Pillay, *Groups definable in local fields and pseudo-finite fields*, Israel Journal of Mathematics **85** (1994), 203–262.
275. ———, *Definable subgroups of algebraic groups over finite fields*, Journal für die reine und angewandte Mathematik **462** (1995), 69–91.
276. T. W. Hungerford, *Algebra*, Graduate Texts in Mathematics, no. 73, Springer, 1974.
277. N. Immerman, *Descriptive complexity*, Graduate Texts in Computer Science, Springer, 1999.
278. S. Jain, D. Osherson, J. Royer, and A. Sharma, *Systems that learn: An introduction to learning theory*, 2nd ed., Learning, Development, and Conceptual Change, MIT Press, 1999.
279. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, 2nd ed., Springer Texts in Statistics, Springer, 2021.
280. S. Janson, *Graphons, cut norm, and distance, couplings and rearrangements*, NYJM Monographs, no. 4, New York Journal of Mathematics, 2013.
281. R. C. Jeffrey, *The logic of decision*, 2nd ed., University of Chicago Press, 1983.
282. Z. Ji, A. Natarajan, T. Vidick, J. Wright, and H. Yuen, $\mathbf{MIP}^* = \mathbf{RE}$, Communications of the ACM **64** (2021), 131–138.
283. C. G. Jockusch Jr and P. E. Schupp, *Generic computability, turing degrees, and asymptotic density*, Journalof the London Mathematical Society, 2nd Series **85** (2012), 472–490.

284. V. Kaimanovich, I. Kapovich, and P. Schupp, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel Journal of Mathematics **157** (2007), 1–46.
285. ———, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel Journal of Mathematics **157** (2007), 1–46.
286. O. Kallenberg, *Foundations of modern probability*, Probability and its Applications, Springer, 1997.
287. ———, *Probabilistic symmetries and invariance principles*, Probability and its Applications, Springer, 2005.
288. E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. A. Siegelbaum, and A. J. Hudspeth, *Principles of neural science*, fifth ed., McGraw-Hill, 2013.
289. I. Kaplansky, *Infinite Abelian groups*, revised ed., University of Michigan Press, 1969.
290. I. Kapovich, A. Myasnikov, P. Schupp, and V. Shpilrain, *Generic-case complexity, decision problems in group theory, and random walks*, Journal of Algebra **264** (2003), 665–694.
291. S. Karn, *Behaviorally correct language identification with anomalies*, preprint, 2025.
292. M. Karpinski and A. Macintyre, *Polynomial bounds for VC dimension of sigmoidal and general Pfaffian neural networks*, Journal of Computer and System Sciences **54** (1997), 169–176.
293. ———, *Approximating volumes and integrals in o-minimal and p-minimal theories*, Connections between model theory and algeraic and analytic geometry, Quaderni di Matematica, no. 6, Arcane, 2000, pp. 149–177.
294. S. Kauffman, *Homeostasis and differentiation in random genetic control networks*, Nature **224** (1969), 177–178.
295. ———, *Metabolic stability and epigenesis in randomly constructed genetic nets*, Journal of Theoretical Biology **22** (1969), 437–467.
296. ———, *The large scale structure and dynamics of gene control circuits: an ensemble approach*, Journal of Theoretical Biology **44** (1974), 167–190.
297. ———, *The origins of order*, Oxford University Press, 1993.
298. S. M. Kautz, *Degrees of random sets*, Ph.D. thesis, Cornell University, 1991.
299. D. Kazhdan and G. Margulis, *A proof of Selberg's hypothesis*, Matematicheskii Sbornik **75** (1968), 163–168.
300. M. J. Kearns and U. V. Vazirani, *An introduction to computational learning theory*, MIT Press, 1994.
301. A. Kechris and B. D. Miller, *Topics in orbit equivalence*, Lecture Notes in Mathematics, no. 1852, Springer, 2004.
302. A. S. Kechris and B. D. Miller, *Topics in orbit equivalence*, Lecture Notes in Mathematics, no. 1852, Springer, 2004.
303. A. S. Kechris, V. G. Pestov, and S. Todorcevic, *Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups*, Geometric and Functional Analysis **15** (2005), 106–189.
304. ———, *Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups*, Geometric and Functional Analysis **15** (2005), 106–189.
305. A. S. Kechris and R. D. Tucker-Drob, *The complexity of classification results in ergodic theory*, Appalachian Set Theory 2006–2012, London Mathematical Society Lecture Note Series, no. 406, Cambridge, 2013, pp. 265–299.
306. Alexander S. Kechris, *Classical descriptive set theory*, Graduate Texts in Mathematics, no. 156, Springer, 1995.
307. H. J. Keisler, *Randomizing a model*, Advances in Mathematics **143** (1999), 124–158.
308. ———, *Probability quantifiers*, Model-Theoretic Logics, Perspectives in Logic, Cambridge University Press, 2016, Originally published 1985, pp. 509–556.
309. J. M. Keynes, *A treatise on probability*, MacMillan, 1921.
310. A. I. Khinchine, *Mathematical foundations of information theory*, Dover, 1957.
311. B. Khoussainov, *A quest for algorithmically random infinite structures*, Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), ACM, 2014, Article No. 56.
312. ———, *A quest for algorithmically random infinite structures, II*, Logical foundations of computer science, Lecture Notes in Computer Science, no. 9537, Springer, 2016, pp. 159–173.

313. H. Ki and T. Linton, *Normal numbers and subsets of* $\mathbb{N}$ *with given densities*, Fundamenta Mathematicae **144** (1994), 163–179.

314. S. Kiefer, R. Mayr, M. Shirmohammadi, and D. Wojtczak, *Strong determinacy of countable stochastic games*, Proceedings of the 32nd Annual ACM/IEEE Symposium on Logic in Computer Science, 2017.

315. J. H. Kim, O. Pikhurko, J. Spencer, and O. Verbitsky, *How complex are random graphs in first order logic?*, Random Structures & Algorithms **26** (2005), 119–145.

316. J. F. C. Kingman, *The ergodic theory of subadditive stochastic processes*, Journal of the Royal Statistical Soceity, Series B (Methodological **30** (1968), 499–510.

317. V. Klee and G. J. Minty, *How good is the simplex algorithm?*, Inequalities, III, Academic Press, 1972, pp. 159–175.

318. J. F. Knight, A. Pillay, and C. Steinhorn, *Definable sets in ordered structures II*, Transactions of the American Mathematical Society **295** (1986), 593–605.

319. D. E. Knuth, *The art of computer programming: Seminumerical algorithms*, 3rd ed., vol. 2, Pearson, 1998.

320. P. G. Kolaitis and M. Y. Vardi, *0–1 laws and decision problems for fragments of second-order logic*, Information and Computation **87** (1990), 302–338.

321. ———, *Infinitary logics and 0–1 laws*, Information and Computation **98** (1992), 258–294.

322. A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, 1950.

323. А. Н. Колмогоров, Три подхода к определению понятия ⟨⟨ количество информатсии ⟩⟩, Проблемы передачи информации **1** (1965), 3–11.

324. B. Kra, *Commentary on 'Ergodic theory of amenable group actions': old and new*, Bulletin of the American Mathematical Society **55** (2018), 343–345.

325. P. N. Kryloff and N. Bogoliouboff, *La théorie générale de la mesure dans son application à l'étude des systémes dynamiques de la mécanique non linéaire*, Annals of Mathematics **38** (1937), 65–113.

326. M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li, *Applied linear statistical models*, fifth ed., McGraw-Hill Irwin Series: Operations and Decision Sciences, McGraw-Hill, 2005.

327. A. Kučera and T. Slaman, *Randomnuess and recursive enumerability*, SIAM Journal on Computing **31** (2001), 199–211.

328. H. Lädesmäki, S. Hautaniemi, I. Shmulevich, and O. Yli-Harja, *Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks*, Signal Processing **86** (2006), 814–834.

329. H. Lädesmäki, I. Shmulevich, and O. Yli-Harja, *On learning gene regulatory networks under the Boolean network model*, Machine Learning **52** (2003), 147–167.

330. J. C. Lagarias, *The ultimate challenge: The* $3x+1$ *problem*, American Mathematical Society, 2010.

331. S. Lang and A. Weil, *Number of points of varieties in finite fields*, American Journal of Mthematics **76** (1954), 819–827.

332. M. C. Laskowski, *A simpler axiomatization of the Shelah-Spencer almost sure theories*, Israel Journal of Mathematics **161** (2007), 157–186.

333. S. L. Lauritzen and N. Wermuth, *Graphical models for associations between variables, some of which are qualitative and some quantitative*, The Annals of Statistics **17** (1989), 31–57.

334. Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature **521** (2015), 436–444.

335. M. Ledoux, *The concentration of measure*, Mathematical Surveys and Monographs, no. 89, American Mathematical Society, 2001.

336. Л. А. Левин, Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятностей, Проблемы передачи информации **10** (1974), 30–35.

337. M. B. Levin, *On the discrepancy estimate of normal numbers*, Acta Arithmetica **88** (1999), 99–111.

338. M. Li, J. Tromp, and P. Vitányi, *Sharpening Occam's razor*, Information Processing Letters **85** (2003), 267–274.

339. Ming Li and Paul Vitányi, *An introduction to Kolmogorov complexity and its applications*, third ed., Texts in Computer Science, Springer, 2008.

340. N. Littlestone, *Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm*, Machine Learning **2** (1988), 285–318.

341. L. Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, 2012.

342. L. Lovász and B. Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B **96** (2006), 933–957.

343. D. Loveland, *A new interpretation of the von Mises' concept of random sequence*, Zeitschrift fur mathematische Logik und Grundlagen dr Mathematik **12** (1966), 279–294.

344. A. Lubotzky and B. Weiss, *Groups and expanders*, Expanding Graphs, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 10, American Mathematical Society, 1993, pp. 95–109.

345. T. Luczak and J. Spencer, *When does the zero-one law hold?*, Journal of the American Mathematical Society **4** (1991), 451–468.

346. C. Lund, L. Fortnow, H. Karloff, and N. Nisan, *Algebraic methods for interactive proof systems*, Journal of the Association for Computing Machinery **39** (1992), 859–868.

347. J. H. Lutz, *Dimension in complexity classes*, SIAM Journal on Computing **32** (2003), 1236–1259.

348. J. H. Lutz and N. Lutz, *Algorithmic information, plane Kakeya sets, and conditional dimension*, ACM Transactions on Computation Theory **10** (2018), 7:1–7:22.

349. R. Lyons and Y. Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, no. 42, Cambridge, 2016.

350. D. Macpherson and C. Steinhorn, *One-dimensional asymptotic classes of finite structures*, Transactions of the American Mathematical Society **380** (2008), 411–448.

351. ______, *Definability in classes of finite structures*, Finite and algorithmic model theory, London Mathematical Society Lecture Note Series, no. 379, Cambridge, 2011, pp. 140–176.

352. D. Macpherson and K. Tent, *Pseudofinite groups with NIP theory and definability in finite simple groups*, Groups and Model Theory, Contemporary Mathematics, no. 576, American Mathematical Society, 2012, pp. 255–267.

353. M. Makkai and R. Paré, *Accessible categories: The foundations of categorical model theory*, Contemporary Mathematics, no. 104, American Mathematical Society, 1989.

354. M. Malliaris and A. Pillay, *The stable regularity lemma revisited*, Proceedings of the American Mathematical Society **144** (2016), 1761–1765.

355. M. Malliaris and S. Shelah, *Regularity lemmas for stable graphs*, Transactions of the American Mathematical Society **366** (2014), 1551–1585.

356. ______, *Notes on the stable regularity lemma*, Bulletin of Symbolic Logic **27** (2021), 415–425.

357. V. W. Marek and M. Truszczyński, *Nonmonotonic logic*, Artificial Intelligence, Springer, 1993.

358. G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik un ihrer Grenzgebiete, 3. Folge, no. 17, Springer, 1991.

359. D. Marker, *Model theory*, Graduate Texts in Mathematics, no. 217, Springer, 2002.

360. A. Marks and S. Unger, *Baire measurable paradoxical decompositions via matchings*, Advances in Mathematics **289** (2016), 397–410.

361. R. Marshall, *Robust classes of finite structures*, Ph.D. thesis, University of Leeds, 2008.

362. P. Martin-Löf, *The definition of random sequences*, Information and Control **9** (1966), 602–619.

363. K. R. Matthews, *Generalized $3x + 1$ mappings: Markov cxhains and ergodic theory*, The Ultimate Challenge: The $3x + 1$ Problem (J. C. Lagarias, ed.), American Mathematical Society, 2010, pp. 79–.

364. K. R. Matthews and A. M. Watts, *A generalization of Hasse's generalization of the Syracuse algorithm*, Acta Arithmetica **43** (1984), 167–175.

365. P. Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, no. 44, Cambridge University Press, 1995.

366. A. D. Matushkin, *Zero-one law for random uniform hypergraphs*, preprint, 2016.

367. A. D. Matushkin and M. E. Zhukovskii, *First order sentences about random graphs: small number of alternations*, Discrete Applied Mathematics **236** (2018), 329–346.

368. J. E. Maxfield, *Normal k-tuples*, Pacific Journal of Mathematics **3** (1953), 189–196.

369. E. Mayordomo, *A Kolmogorov complexity characterization of constructive Hausdorff dimension*, Information Processing Letters **84** (2002), 1–3.

370. W. Merkle, N. Mihalović, and T. A. Slaman, *Some results on effective randomness*, Theory of Computing Systems **39** (2006), 707–721.

371. J.-F. Mertens, *Stochastic games*, Handbook of Game Theory, vol. 3, Elsevier, 2002, pp. 1809–1832.

372. J.-F. Mertens and A. Neyman, *Stochastic games*, International Journal of Game Theory **10** (1981), 53–66.
373. P. Michel, *Busy beaver competition and Collatz-like problems*, Archive for Mathematical Logic **32** (1993), 351–367.
374. P. Michel and M. Margenstern, *Generalized* $3x+1$ *functions and the theory of computation*, The Ultimate Challenge: The $3x+1$ Problem (J. C. Lagarias, ed.), American Mathematical Society, 2010, pp. 105–.
375. C. F. Miller, III, *On group-theoretic decision problems and their classification*, Annals of Mathematics Studies, no. 68, Princeton University Press, 1971.
376. G. A. Miller and D. McNeill, *Psycholinguistics*, The Handbook of Social Psychology (G. Lindzey and E. Aronson, eds.), vol. 3, Addison-Wesley, 2nd ed., 1968, pp. 666–794.
377. G. L. Miller, *Riemann's hypothesis and tests for primality*, Journal of computer and system science **13** (1976), 300–317.
378. R. Miller, *Isomorphism and classification for countable structures*, Computability **8** (2019), 99–117.
379. J. Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, no. 160, Princeton University Press, 2006.
380. A. Montalbán, *Computable structure theory: Within the arithmetic*, Perspectives in Logic, Cambridge University Press, 2021.
381. A. Montalban, *Computable structure theory, part ii*, preprint, 2025.
382. A. Montalbán and A. Nies, *Borel structures: a brief survey*, Effective Mathematics of the Uncountable, Lecture Notes in Logic, no. 41, Cambridge, 2013, pp. 124–134.
383. C. C. Moore, *Ergodicity of flows on homogeneous spaces*, American Journal of Mathematics **88** (1966), 154–178.
384. A. De Morgan, *Formal logic: or, the calculus of inference, necessary and probable*, Taylor and Walton, 1847.
385. R. A. Moser and G. Tardos, *A constructive proof of the general Lovász local lemma*, Journal of the Association for Computing Machinery **57** (2010), 1–15.
386. A. A. Muchnik, A. L. Semenov, and V. A. Uspensky, *Mathematical metaphysics of randomness*, Theoretical Computer Science **207** (1998), 263–317.
387. K. P. Murphy, *Dynamic Bayesian networks*, Ph.D. thesis, University of California, Berkeley, 2002.
388. T. Neary, *Small polynomial time universal Turing machines*, Proceedings of the 4th Irish Conference on the Mathematical Foundations of Computer Science and Information Technology (T. Hurley, et al., ed.), 2006, pp. 325–329.
389. E. Nelson, *Internal set theory: a new approach to nonstandard analysis*, Bulletin of the American Mathematical Society **83** (1977), 1165–1198.
390. A. Nies, *Computability and randomness*, Oxford Logic Guides, no. 51, Oxford, 2009.
391. N. Nisan and A. Wigderson, *Hardness vs. randomness*, Journal of Computer Systems and Sciences **49** (1994), 149–167.
392. P. S. Novikov, *On algorithmic unsolvability of the word problem in group theory*, Trudy Matimaticheskogo Instituta imeni V.A. Steklova, no. 44, Izdat. Akad. Nauk SSSR, 1955.
393. D. S. Ornstein and B. Weiss, *Ergodic theory of amenable group actions. I: The Rohlin Lemma*, Bulletin of the American Mathematical Society **2** (1980), 161–164.
394. D. N. Osherson and S. Weinstein, *Criteria of language learning*, Information and Control **52** (1982), 123–138.
395. D. Osin, *A topological zero-one law and elementary equivalence of finitely generated groups*, Annals of Pure and Applied Logic **172** (2021), 102915.
396. R. Pal, I. Ivanov, A. Datta, M. L. Bittner, and E. R. Dougherty, *Generating Boolean networks with a prescribed attractor structure*, Bioinformatics **21** (2005), 4021–4025.
397. J. Paris and A. Venkovská, *Pure inductive logic*, Perspectives in Logic, Cambridge, 2015.
398. E. Patterson, *The algebra and machine representation of statistical models*, Ph.D. thesis, Stanford University, 2020.
399. J. Pearl, *Deciding consistency in inheritance networks*, Tech. Report 870053, Cognitive Systems Laboratory, UCLA, 1987.
400. ———, *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann, 1988.
401. ———, *Causality*, Cambridge, 2000.

402. J. Pearl and A. Paz, *Graphoids: A graph-based logic for reasoning about relevance relations*, Tech. Report 850038, Cognitive Systems Laboratory, UCLA, 1985.
403. Y. B. Pesin, *Dimension theory in dynamical systems*, University of Chicago Press, 1997.
404. P. Petersen, *Riemannian geometry*, 3rd ed., Graduate Texts in Mathematics, no. 171, Springer, 2016.
405. F. Petrov and A. Vershik, *Uncountable graphs and invariant measures on the set of universal countable graphs*, Random Structures and Algorithms **37** (2010), 389–406.
406. R. R. Phelps, *Lectures on choquet's theorem*, Lecture Notes in Mathematics, no. 1757, Springer, 2001.
407. A. Pillay and C. Steinhorn, *Discrete o-minimal structures*, Annals of Pure and Applied Logic **34** (1987), 275–289.
408. M. S. Pinsker, *On the complexity of a concentrator*, 7th International Teletraffic Conference, 1973.
409. L. Pitt and M. K. Warmuth, *Prediction-preserving reducibility*, Journal of Computer and System Sciences **41** (1990), 430–467.
410. K. Popper, *The logic of scientific discovery*, Routledge, 2002, Original edition published in 1935; this edition dates from text of 1959.
411. E. L. Post, *Recursive unsolvability of a problem of Thue*, The Journal of Symbolic Logic **12** (1947), 1–11.
412. P. Potgieter, *Algorithmically random series and Brownian motion*, Annals of Pure and Applied Logic **169** (2018), 1210–1226.
413. M. O. Rabin, *Probabilistic algorithm for testing primality*, Journal of Number Theory **12** (1980), 128–138.
414. C. Radin and L. Sadun, *Phase transitions in a complex network*, Journal of Physics A **46** (2013), 305002.
415. F. P. Ramsey, *Truth and probability*, The Foundations of Mathematics and other Logical Essays (R. B. Braithwaite, ed.), Harcourt Brace, 1931, First published 1926, pp. 156–198.
416. J. Reimann, *Computability and fractal dimension*, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 2004.
417. R. Reiter, *A logic for default reasoning*, Artificial Intelligence **13** (1980), 81–132.
418. ———, *Nonmonotonic reasoning*, Annual Review of Computer Science **1987** (1987), 147–186.
419. L. J. Richter, *Degrees of structures*, The Journal of Symbolic Logic **46** (1981), 723–731.
420. A. Rivkind and O. Barak, *Local dynamics in trained recurrent neural networks*, Physical Review Letters **118** (2017), no. 258101, 1–5.
421. A. Robinson, *Complete theories*, Studies in logic and the foundations of mathematics, North-Holland, 1956.
422. P. Roeper and H. Leblanc, *Probability theory and probability semantics*, Uversity of Toronto Press, 1999.
423. F. Rosenblatt, *Principles of neurodynamics; perceptrons and the theory of brain mechansims*, Spartan, 1962.
424. S. Ross, *Introduction to probability models*, 12th ed., Academic Press, 2019.
425. J. J. Rotman, *The theory of groups*, Allyn and Bacon, 1965.
426. J. S. Royer, *Inductive inference of approximations*, Information and Control **70** (1986), 156–178.
427. A. Rumyantsev, *Infinite computable version of Lovász local lemma*, preprint, 2010.
428. A. Rumyantsev and A. Shen, *Probabilistic constructions of computable objects and a computable version of Lovász local lemma*, Fundamental Informaticae **132** (2014), 1–14.
429. S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*, third ed., Prentice Hall Series in Artificial Intelligence, Prentice Hall, 2010.
430. J. Rute, *Algorithmic randomness and constructive/computable measure theory*, Algorithmic Randomness: Progress and Prospects (J. N. Y. Franklin and C. P. Porter, eds.), Lecture Notes in Logic, no. 50, Cambridge, 2020, pp. 58–114.
431. M. J. Ryten, *Model theory of finite difference fields and simple groups*, Ph.D. thesis, The University of Leeds, 2007.
432. N. Sauer, *On the density of families of sets*, Journal of Combinatorial Theory (A) **13** (1972), 145–147.

433. A. Saxe, Y. Bansal, J. Dapello, M. Advani, A. Kolchinsky, B. Tracey, and D. Cox, *On the information bottleneck theory of deep learning*, ICLR, 2018.
434. A.-M. Scheerer, *Computable absolutely normal numbers and discrepancies*, Mathematics of Computation (2017), 2911–2926.
435. J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Networks **61** (2015), 85–117.
436. K. Schmidt, *Asymptotically invariant sequences and an action of $SL(2, Z)$ on the 2-sphere*, Israel Journal of Mathematics **37** (1980), 193–208.
437. W. M. Schmidt, *Über die Normalität von Zahlen zu verschiedenen Basen*, Acta Arithmetica **VII** (1962), 299–309.
438. ______, *Irregularities of distribution, VII*, Acta Arithmetica **21** (1972), 45–50.
439. C. P. Schnorr, *A unified approach to the definition of random sequences*, Mathematical systems theory **5** (1971), 246–258.
440. ______, *Zufälligkeit und Wahrscheinlichkeit*, Lecture Notes in Mathematics, no. 218, Springer, 1971.
441. C. P. Schnorr and H. Stimm, *Endliche automaten und zufallsfolgen*, Acta Informatica **1** (1972), 345–359.
442. D. G. Senadheera, *Effective concept classes of PAC and PACi incomparable degrees, joins and embedding of degrees*, Ph.D. thesis, Southern Illinois University, 2022.
443. G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
444. A. Shamir, **IP = PSPACE**, Journal of the Association for Computing Machinery **39** (1992), 869–877.
445. C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379–423.
446. L. S. Shapley, *Stochastic games*, Proceedings of the National Academy of Sciences **39** (1953), 1095–1100.
447. S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific Journal of Mathematics **41** (1972), 247–261.
448. ______, *Classification theory and the number of non-isomorphic models*, revised ed., Studies in Logic and the Foundations of Mathematics, no. 92, North-Holland, 1990.
449. S. Shelah and J. Spencer, *Zero-one laws for sparse random graphs*, Journal of the American Mathematical Society **1** (1988), 97–115.
450. A. Shen, *IP = PSPACE: Simplified proof*, Journal of the Association for Computing Machinery **39** (1992), 878–880.
451. A. K. Shen, *On relations between different algorithmic definitions of randomness*, Sovient Mathematics Doklady **38** (1989), 316–319.
452. A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, no. 95, Springer, 1996.
453. I. Shmulevich and E. R. Dougherty, *Genomic signal processing*, Princeton Series in Applied Mathematics, Princeton University Press, 2007.
454. ______, *Probabilisitc boolean networks*, Society for Industrial and Applied Mathematics, 2010.
455. H. A. Simon, *On a class of skew distribution functions*, Biometrika **42** (1955), 425–440.
456. P. Simon, *A guide to NIP theories*, Lecture Notes in Logic, no. 44, Cambridge, 2015.
457. R. I. Soare, *Recursively enumerable sets and degrees*, Springer-Verlag, 1987.
458. E. D. Sontag, *Feedforward nets for interpolation and classification*, Journal of Computer and System Sciences **45** (1992), 20–48.
459. J. Spencer, *Asymptotic lower bounds for Ramsey functions*, Discrete Mathematics **20** (1977), 69–76.
460. ______, *Threshold functions for extension statements*, Journal of Combinatorial Theory A **53** (1990), 286–305.
461. ______, *Threshold spectra via the Ehrenfeucht game*, Discrete Applied Mathematics **30** (1991), 235–252.
462. ______, *The strange logic of random graphs*, Algorithms and Combinatorics, no. 22, Springer, 2001.
463. J. Spencer and K. St. John, *The tenacity of zero-one laws*, The Electronic Journal of Combinatorics **8** (2001), R17.1–R17.14.

464. J. Spencer and M. E. Zhukovskii, *Bounded quantifier depth spectra for random graphs*, Discrete Mathematics **339** (2016), 1651–1664.

465. L. Staiger, *Kolmogorov complexity and Hausdorff dimension*, Information and Computation **103** (1993), 159–194.

466. ______, *A tight upper bound on Kolmogorov complexity and uniformly optimal prediction*, Theory of Computing Systems **31** (1998), 215–229.

467. C. I. Steinhorn, *Borel structures and measure and category logics*, Model Theoretic Logics, Perspectives in Logic, no. 8, Springer, 1985, pp. 579–596.

468. G. Stengle and J. E. Yukich, *Some new Vapnik-Chervonenkis classes*, The Annals of Statistics **17** (1989), 1441–1446.

469. V. E. Stepanov, *Phase transitions in random graphs*, Theory of Probability and its Applications **15** (1970), 187–203.

470. F. Stephan and Yu. Ventsov, *Learning algebraic structures from text*, Theoretical Computer Science **268** (2001), 221–273.

471. L. J. Stockmeyer, *The polynomial-time hierarchy*, Theoretical Computer Science **3** (1976), 1–22.

472. G. Stuck and R. Zimmer, *Stabilizers for ergodic actions of higher rank semisimple groups*, Annals of Mathematics **139** (1994), 723–747.

473. M. Studený, *Conditional independence relations have no finite complete characterization*, preprint, 1992.

474. D. Sussillo and O. Barak, *Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks*, Neural Computation **25** (2013), 626–649.

475. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica **27** (1975), 199–245.

476. ______, *Regular partitions of graphs*, preprint, 9 pp., 1975.

477. T. Tao, *Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets*, Contributions to Discrete Mathematics **10** (2015), 22–98.

478. ______, *Szemerédi's proof of Szemerédi's theorem*, Acta Mathematica Hungarica **161** (2020), 443–487.

479. A. Tarski, *Algebraische Fassung des Maßproblems*, Fundamenta Mathematicae **31** (1938), 47–66.

480. ______, *Cardinal algebras*, Oxford University Press, 1949.

481. A. Taveneaux, *Randomness zoo*, preprint, 2012.

482. E. Thoma, *Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe*, Mathematische Zeitschfrift **85** (1964), 40–61.

483. ______, *Über unitäre Darstellungen abzählbarer, diskreter Gruppen*, Mathematische Annalen **153** (1964), 111–138.

484. S. Thomas, *The classification problem for torsion-free Abelian groups of finite rank*, Journal of the American Mathematical Society **16** (2003), 233–258.

485. S. Thomas and R. Tucker-Drob, *Invariant random subgroups of strictly diagonal limits of finite symmetric groups*, Bulletin of the London Mathematical Society **46** (2014), 1007–1020.

486. ______, *Invariant random subgroups of inductive limits of finite alternating groups*, Journal of Algebra **503** (2018), 474–533.

487. A. Thue, *Probleme über Veränderungen von Zeichenreihen nach gegeben regeln*, Skrifter utgit av Videnskapsselskapet i Kristiania **1** (1914), no. 10, 1–34.

488. N. Tishby and N. Zaslavsky, *Deep learning and the information bottleneck principle*, IEEE Information Theory Workshop, 2015.

489. S. Toda, *PP is as hard as the polynomial-time hierarchy*, SIAM Journal of Computing **20** (1991), 878–880.

490. H. Towsner, *Limits of sequences of Markov chains*, Electronic Journal of Probability **20** (2015), 1–23.

491. ______, *Algorithmic randomness in ergodic theory*, Algorithmic Randomness: Progress and Prospects (J. N. Y. Franklin and C. P. Porter, eds.), Lecture Notes in Logic, no. 50, Cambridge, 2020, pp. 40–57.

492. S. Vadhan, *Pseudorandomness*, Foundations and Trends in Theoretical Computer Science, vol. 7, Now, 2012.

493. L. G. Valiant, *The complexity of computing the permanent*, Theoretical Computer Science **8** (1979), 189–201.

494. ———, *A theory of the learnable*, Communications of the ACM **27** (1984), 1134–1142.
495. P. van der Hoorn, G. Lippner, and D. Krioukov, *Sparse maximum-entropy random graphs with a given power-law degree distribution*, Journal of Statistical Physics **173** (2018), 803–844.
496. V. N. Vapnik, *The nature of statistical learning theory*, 2nd ed., Statistics for Engineering and Information Science, Springer, 2000.
497. V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory of probability and its applications **16** (1971), 264–280.
498. A. M. Vershik, *Totally nonfree actions and the infinite symmetric group*, Moscow Mathematical Journal **12** (2012), 193–212.
499. N. Vieille, *Stochastic games: Recent results*, Handbook of Game Theory, vol. 3, Elsevier, 2002, pp. 1833–1850.
500. Susan Vineberg, *Dutch Book Arguments*, The Stanford Encyclopedia of Philosophy (Edward N. Zalta, ed.), Metaphysics Research Lab, Stanford University, spring 2016 ed., 2016.
501. A. Visser, *Numerations, λ-calculus, & arithmetic*, To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism, Academic Press, 1980, pp. 259–284.
502. J. von Neumann, *Zur allgemeinen Theorie des Masses*, Fundamenta Mathematicae **13** (1929), 73–116.
503. P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, no. 79, Springer, 1982.
504. R. Weber, *Computability theory*, Student Mathematical Library, no. 62, American Mathematical Society, 2012.
505. K. Weihrauch, *Computable analysis*, Texts in Theoretical Computer Science, Springer, 2000.
506. J. Williamson, *Probability logic*, Handbook of the Logic of Argument and Inference, Studies in Logic and Practical Reasoning, vol. 1, Elsevier, 2002, pp. 397–424.
507. J. S. Wilson, *On simple pseudofinite groups*, Journal of the London Mathematical Society **51** (1995), 471–490.
508. D. Xiao, *On basing* $\mathbf{ZK} \neq \mathbf{BPP}$ *on the hardness of PAC learning*, 24th Annual IEEE Conference on Computational Complexity, 2009, pp. 304–315.
509. A. C. Yao, *Theory and applications of trapdoor functions*, 23rd Annual Symposium on Foundations of Computer Science, IEEE, 1982, pp. 80–91.
510. C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning requires rethinking generalization*, ICLR, 2017.
511. X. Zheng and R. Rettinger, *On the extnesions of Solovay reducibility*, Computing and Combinatorics: 10th Annual International Conference, COCOON 2004, Lecture Notes in Computer Science, no. 3106, Springer, 2004, pp. 360–369.
512. M. E. Zhukovskii, *On infinite spectra of first order properties of random graphs*, Moscow Journal of Combinatorics and Number Theory **4** (2016), 73–102.
513. R. J. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Mathematics, no. 81, Birkhäuser, 1984.

Index