CHAPTER 2

# Formulating Probability

## 2.1. Axioms for Probability

**2.1.1. The Prehistory of Probability.** While there have certainly been individual utterances related to probability from quite early on in recorded thought, as well as hints of at least potential organized knowledge of the field, it is difficult to avoid the impression that logic came first. Certainly mathematical logic came before mathematical probability.

In its medieval usage, "probability" was not a mathematical or quantitative concept at all. Some things, collectively called *scientia*, could be known certainly by "demonstration," a precursor of modern concepts of proof. Other kinds of propositions, collectively called *opino* — and it was thought a genuinely different category — were the subject of belief, including belief based on sense perception. To say that an opinion was *probable* meant exactly that some trustworthy person believed it. There was no paradox, then, when Gibbon said of a certain opinion, "Such a fact is probable but undoubtedly fals."

This led to the common accusation, not unfamiliar in our times, that one's opponent had first settled on an opinion, and then found some learned worthy whose words could be taken as acceptance. Such an approach, sometimes called "probabilism," was one that conscientious scholars would presumably take pains to avoid.

Although various problems on the fair division of the stakes in an uncompleted game of chance circulated in the years before it, the first accounts of probability in anything like the modern sense appear to have arisen in the context of the now-obscure Jansenist controversey in theological circles. Pascal was a collaborator and correspondent with several on the Jansenist side, including Antoine Arnauld and Pierre Nicole, authors of the 1662 *La Logique ou l'ard de penser*, better known as the *Port Royal Logic*.

The *Port Royal Logic* started in a relatively conventional way, treating in its first part the problems around forming concepts; in the second part, deductive reasoning generally; in the third, syllogistic reasoning specifically. The fourth part begins as an exposition of analysis and synthesis, the two traditional parts of geometric reasoning, but the final four chapters take up a very different kind of reasoning. The matter under examination there is the application of reason to events "which may or may not be, when we inquire about the *future*, but which cannot be otherwise when we inquire about the *past*" [**27**]. It would be foolish to think, in probabilist fashion, that something is false in general merely because it fails in a particular case, or to think it true merely because it is possible. These uncertain beliefs, though, are explicitly compared to games of channce, and the frequency of the various outcomes of these games are used to reckon the degree of probability of the proposition — for the first time a quantity. This analysis ends with the first published form of Pascal's wager.

The history of probability, or the course by which Pascal, Fermat, Leibniz, Huygens, and finally Bernoulli extended these observations into what we now recognize as mathematical probability is not the subject of the present book. However, it is important to note two features regarding the first emergence of mathematical probability. First, it is important to note that it emerged as a branch of logic intended to deal with uncertain events. Second, we note that the genesis of modern probability happens exactly with the merger of what are later called the Bayesian and frequentist interpretations of probability.

In the frequentist interpretation, probability represents the proportion of one outcome in a (perhaps large) class of indistinguishable trials. In the Bayesian interpretation, the probability of a hypothesis reflects our degree of beiief in a hypothesis. The authors of the *Port Royal Logic* used the frequency of truth to address the problem of assigning a degree of belief. Of course, the philosophy of probability, including the Bayesian/frequentist debate, has a large literature of its own. The explanation here of the early history of the subject was heavily influenced by [**213**].

The *Port Royal Logic* was widely circulated, but by 1847 Augustus De Morgan still felt a need to explain the inclusion of probability in his *Formal Logic*:

> The old doctrine of modals is made to give place to the numerical theory of probability. Many will object to this theory as extralogical. But I cannot see on what definition, founded on real distinction, the exclusion of it can be maintained. When I am told that logic considers the validity of the inference, indpendently of the truth or falsehood of the matter, or supplies the conditions under which the hypothetical truth of the matter of the premises gives hypothetical truth to the matter of the conclusion, I see a real definitionm, which propounds for consideration the forms and laws of inferential thought. But when it is further added that the only hypothetical truth shall be absolute truth, certain knowledge, I begin to see arbitrary distinction, wanting the reality of that which preceded.... Not however to dispute upon names, I mean that I should maintain, against those who would exclude the theory of probability from logic, that, call it by what name they like, it should accompany logic as a study.[**347**]

In any case, De Morgan described probability in the context of logic, including arithmetical calculations and the assignment of a scale of probability, offering the Farenheit temperature scale as an example, but exploring in more detail the interval $[-1, 1]$.

When Boole undertook a general mathematicization of logic, he, too, included probability in its scope. Significantly for our investigation of algorithmic randomness in Chapter 3 and of machine learning in Chapter 7, one of his first motivating examples is the use of the fact that "the same characters and successions of characters recur with determinate frequency" to interpret archaeological texts from Ireland and the Near East . Boole sets out an axiomatic calculus of probabilities that would be familiar today either to students of elementary probability or to the axiomatizers of the modern probability logics discussed in Section 2.2.2, as well as a "general method for the solution of problems in the theory of probabilities" [**71**].

Keynes described probability in much more strictly logical terms, axiomatizing an algebraic structure capable of serving as the possible values of probability [**274**]. For Keynes, too, probability was a kind of logic.

The tradition of interpreting probability as a generalization of logic continued for some time afterward in philosophical circles in the hands of Popper, Carnap, and others, and some of this work (in particular that of Popper) will be relevant for algorithmic randomness. However, it is not primarily this philosophical tradition that conerns us, but the communication of modern *mathematical* logic with modern probability.

**2.1.2. The Kolmogorov Axiomatization.** I have broken some with the standard scholarly delineation of the history. Most would place everything after Pascal — and certainly everything after Bernoulli — in the *history* of probability, not the prehistory. There is certainly fairness in this. However, the point I want to emphasize is that all of the probability discussed in the previous section belongs firmly to an age of the field at least as distant from the practice of modern probability as De Morgan's classification of syllogistic forms is from that of modern logic.

A fair point of delineation in logic might be the Zermelo-Frankel formulation of set theory, or, more likely, the quick succession in the 1930s of Gődel's completeness and incompleteness theorems, Tarski's undefinability theorem, and Turing's negative solution to the *Entscheidungsproblem*. That which came before is certainly not modern logic. That which came afterward will generally be accepted by modern logicians as being in continuity with modern research in the field. The analogous demarcation point in probability is Kolmogorov's measure-theoretic formulation of probability.

It is true that Kolmogorov's measure-theoretic approach opened up significant new vistas for the field, especially toward non-finite probability spaces, but his motivation, it seems germane to state, was

> ... to give an axiomatic foundation for the theory of probability. The author set himself the task of putting in their natural place, among the general notions of modern mathematics, the basic concepts of probability theory — concepts which until recently were considered to be quite peculiar.[**287**]

Kolmogorov developed probability as a special case of Lebesgue's measure theory.

DEFINITION 2.1.1. A probability space is a triple $(\Omega, \mathcal{B}, P)$ with the following properties:

(1) $\Omega$ is a set.
(2) $\mathcal{B}$ is a $\sigma$-algebra of subsets of $\Omega$.
(3) $\Omega \in \mathcal{B}$
(4) $P : \mathcal{B} \to \mathbb{R}^{\geq 0}$
(5) $P(\Omega) = 1$
(6) If $A, B \in \mathcal{B}$ are disjoint, then $P(A \cup B) = P(A) + P(B)$
(7) If $A_1 \supseteq A_2 \supseteq \cdots$ is a decreasing sequence of elements of $\mathcal{B}$ with empty intersection, we have $\lim_{n \in \mathbb{N}} P(A_n) = 0$.

It is routine to see that this last condition is equivalent to the assertion that $P$ is countably additive.

For any set $S$, an $S$-valued random variable is a function $X : \Omega \to S$. Typically the set $S$ has some measure of its own and random variables are restricted to be the measurable functions. From this perspective, for a measurable subset $T \subseteq S$, we write $P(X \in T) = P\left(X^{-1}(T)\right)$.

Using this formalism, it is possible to generalize the probability known to Bernoulli and Boole to contacts with what might be called a "completed infinity" of possible outcomes, those involving infinite-dimensional distributions. We express the expectation $E(X) = \int X \; dP$.

In this world, a large range of applications opens up. For instance, let $\mathcal{G}$ be a graph, and equip $\mathcal{G}$ with a sequence of functions $\eta_t : V \to \{0, 1\}$ on its vertices, as $t$ ranges over $\mathbb{N}$.

EXAMPLE 2.1.2. Each vertex represents an individual in a population susceptible in principle to an infectious disease. Those vertices $v$ with $\eta_t(v) = 1$ are said to be *infected* at time $t$ and those with $\eta_t(v) = 0$ are said to be healthy. Of course, many models could describe the dynamics of this system over time, perhaps prescribing $\eta_{t+1}(v)$ in terms of the values of $\eta_t$ on the neighborhood of $v$. This is called a *contact process*. Under specific models of the evolution of $\eta_t$, we could ask questions about the probability that, for instance, there is some $\hat{t}$ such that for $t > \hat{t}$ there is some $v$ such that $\eta_t(v) = 1$.

There arises, of course, the question of the necessesity of this approach. In a sense, this necessity is established by a theorem of de Finetti [**130, 132, 131**]. Given a measurable space $(\Omega, \mathcal{B})$ *betting system* is a function $b : \mathcal{B}' \times \Omega \to \mathbb{R}$, where $\mathcal{B}' \subseteq \mathcal{B}$ is countable, such that for each $S \in \mathcal{B}'$ there is a $q_S \in \mathbb{R}$ such that

$$b(S, x) := \begin{cases} 1 - q_S & \text{if } x \in S \\ -q_S & \text{otherwise} \end{cases}$$

In this sense, we can interpret any function $\mu : \mathcal{B} \to \mathbb{R}$ as a betting system $b_\mu$, by setting $q_S = \mu(S)$. A rational actor who, for each $S \in \mathcal{B}'$, believes with strength $\mu(S)$ that $x \in S$ will expect value 0 from $b_\mu(S, \cdot)$, and will accept any bet giving $\mu$-expectation larger than that.

If $b$ is a betting system, a Dutch book on $b$ is a set $\mathcal{C} \subseteq \mathcal{B}'$ such that for any $x \in \Omega$, we have

$$\sum_{S \in \mathcal{C}} b(S, x) < 0.$$

A Dutch book, then, is a combination bet guaranteeing a loss for the agent who adopts betting system $b$. The following result, variously attributed to [**374**] and [**131**], shows that the axioms of probability are necessary, in the sense that a Dutch book can be made on any betting system that violates them.

THEOREM 2.1.3. *Let $(\Omega, \mathcal{B})$ be a measurable space, and let $\mu$ be a function from $\mathcal{B}$ to $\mathbb{R}^{\geq 0}$. If $(\Omega, \mathcal{B}, \mu)$ is not a probability space, then there is a Dutch book on $b_\mu$.*

The proof consists of an elementary truth table construction, and is described in detail in [**453**].

A second result, which seems unequivocally attributed to de Finetti, suggests a kind of uniqueness for the probability measure on particular events. It seems difficult to locate in de Finetti's works an explicit statement of the theorem attributed to him — or, indeed, any proofs or many theorems — but the sentiments expressed by the theorem certainly arise in his work. A standard modern expression of the theorem can be found in [**256**]. For the statement of this theorem, we define a sequence $(X_i : i \in \mathbb{N})$ of random variables to be *excxhangeable* if and only if for any permutation $\sigma \in S_\infty$ of the natural numbers and for any $j \in \mathbb{N}$, the joint distribution of $(X_{\sigma(1)}, \ldots, X_{\sigma(j)})$ is equal to that of $(X_1, \ldots, X_j)$.

THEOREM 2.1.4 (de Finetti's Theorem). *Let $(X_i : i \in \mathbb{N})$ be a sequence of random variables over a Borel measure space, taking values in $S$, with probability measure $P$. Then the following are equivalent:*

(1) *$(X_i : i \in \mathbb{N})$ is exchangeable.*
(2) *There is some random variable $\nu$ taking values in the set of probability measures on $S$ so that $P$ is the product measure $\nu^\infty$ almost surely.*

This "uniqueness" aspect of this theorem is explained, for instance, in [**251**]: de Finetti did not believe in objective values of probability. However, he was able to give meaning to probability apart from any objective value by proving that any two observers would eventually agree upon a subjective probability measure. Dawid explains the impact of this theorem in some detail, including the observation that a person believing only in the exchangeability of a sequence of coin tosses would converge on a subjective probability distribution exactly matching that given by the usual (objective) axiomatization of Bernoulli trials [**129**].

If we consider the sequence $(X_i : i \in \mathbb{N})$ to be exchangeable, then as we observe $X_i$ for more and more $i$, the subjective probability measure we would all eventually agree on is $\nu^\infty$. In particular, if we write $P(T) = \nu^\infty(T)$ and take expected values, then we obtain

$$P(T) = E\nu^\infty = \int m^\infty P(\nu \in dm),$$

where $m$ in the integral ranges over all probability measures on $S$.

PROOF. The implication from 2 to 1 is obvious from the definitions. To show the opposite direction, let $\mathbb{X} = (X_i : i \in \mathbb{N})$ be an exchangeable sequence. Let $\mathcal{I}$ be the shift-invariant $\sigma$-algebra in the natural probability space of sequences on which $\mathbb{X}$ is a random variable, and let $\mathcal{I}_\mathbb{X}$ be the pre-image of this $\sigma$-algebra under $\mathbb{X}$.

For a probability measure $P$, a random variable $A$ and a $\sigma$-algebra $\mathcal{C}$, we define the conditional probability $P(A|\mathcal{C})$ in the standard way [**255**]. That is, we first define the conditional *expectation* first for the case where $A \in L^2$, so that $E(A|\mathcal{C})$ is the orthogonal projection of $A$ onto the subspace of $\mathcal{C}$-measurable funcitons. This can be extended, by continuity, to $A \in L^1$. We then define $P(A|\mathcal{C}) := E(\chi_A|\mathcal{C})$. Thus, $P(A|\mathcal{C})$ is not a constant, but a random element of the space of probability measures. This specializes to the elementary conditioning on a single event when $\mathcal{C}$ is the smallest $\sigma$-algebra containing that event. We now define $\nu = P(X_1|\mathcal{I}_\mathbb{X})$.

To verify that this $\nu$ has the properties required, we take $I \in \mathcal{I}$ and let $f_1, \ldots, f_m$ be bounded measurable functions, and with the dominated convergence theorem (recalling that we have assumed $\mathbb{X}$ is exchangeable) we let $n \in \mathbb{N}$ and calculate

$$E\left(\chi_I(\mathbb{X})\right)\prod_{i=1}^{m} f_i(X_i) = \lim_{n\to\infty}\frac{1}{n^m}\sum_{j_1,\ldots,j_m\leq n} E\left(\chi_I(\mathbb{X})\right)\prod_{k=1}^{m} f_k\left(X_{kn+j_k}\right)$$
$$= \lim_{n\to\infty} E\left(\chi_I(\mathbb{X})\right)\prod_{k=1}^{m}\frac{\sum_{j=1}^{n} f_k\left(X_{kn+j}\right)}{n}.$$

We now use the definition of conditional probability, combined with the mean ergodic theorem, to find that this quantity converges to $E\left(\chi_I(\mathbb{X})\right)\prod_{k=1}^{m}\nu f_k$. Now for any product-measurable set $B$, we have $P\left(\mathbb{X}\in B|\mathcal{I}_{\mathbb{X}}\right) = \nu^\infty(B)$ almost surely, as required. □

It is natural to ask about the effectiveness of the preceding proof: given a sequence of random variables that is, in some appropriate sense, effective, can we find a computable measure $\nu$? This question has added significance in light of Jeffrey's interpretation. Following the reasoning of algorithmic game theory, if the agents can't effectively find the Nash equilibrium, then its power to predict their behavior is limited. Similarly, if the so-called *de Finetti measure* is not computable, we could not expect computable agents to agree on it.

Theorem 2.1.5 ([**179**]). *Let $F_{\mathbb{X}}$ be the distribution of a real-valued exchangeable sequence $\mathbb{X}$ and $F_\nu$ be the distribution of the measure whose existence is guaranteed by Theorem 2.1.4. Then $F_{\mathbb{X}} \equiv_T F_\nu$ in the sense of type-two functions.*

Proof. Let $\mathbb{X}$ be an exchangeable sequence of real variables, and $B$ a Borel set. We hope to define a random variable $Y_B$ by $Y_B = F_\nu(B)$, and it suffices to define it where $B$ is an interval $(a,b)$ with rational endpoints.

Now if $\beta : \mathbb{N}\to\mathcal{B}$ is an enumeration of some Borel sets, by de Finetti's theorem, we have $P\left(\bigcap_{i=1}^{k}\{X_i\in\beta(i)\}\right) = E\left(\prod_{i=1}^{k} Y_{\beta(i)}\right)$.

It can be shown that if $R:\mathbb{N}\to\mathcal{B}$ is an enumeration of finite unions of open rectangles in $\mathbb{Q}^k$, then from a degree computing $F_{\mathbb{X}}$, we can enumerate all rational lower bounds on $P\left(\bigcap_{i=1}^{k}\{X_i\in R(i)\}\right)$. By noting that almost surely $F_\nu$ places no mass on boundaries, an enumeration of upper bounds can at least be simulated.

Lemma 2.1.6. *Let $\mathbb{Y}$ be a sequence of $[0,1]$ random variables with distribution $F_\eta$. Then $F_\eta$ is Turing equivalent to the sequence $E\left(\prod_{i=1}^{k} Y_{j(i)}\right)$ as $k$ ranges over $\mathbb{N}$ and $j$ ranges over $\mathbb{N}^k$.*

Proof. That the mixed moments $E\left(\prod_{i=1}^{k} Y_{j(i)}\right)$ should be computable relative to $F_\eta$ is straightforward from the computability of polynomials and of integrals.

On the other hand, let $B$ be a $k$-tuple of open intervals with rational endpoints, and consider that $F_\eta = E\left(\chi_B(\bar{Y})\right)$. We can find a sequence of polynomials converging pointwise to $\chi_B$ from below, completing the proof of the Lemma and the Theorem. □

□

In considering Kolmogorov's axiomatization of probability, it is also interesting to point out the recent formalization of a proof of the Central Limit Theorem in [**35**] in the Isabelle proof assistant. Much of the work in this formalization included formalization of the appropriate objects in measure theory.

## 2.2. Reasoning With Probability

**2.2.1. The Primacy of Probability.** If you want to compute a conjunction, you could think of

$$P(A \wedge B) = P(A|B) \cdot P(B).$$

Even under some additional hypotheses, you could reckon

$$P(A \wedge B|H) = P(A|B \wedge H) \cdot P(B|H).$$

Similarly, you could compute a disjunction by

$$P(A \vee B|H) = P(A|H) + P(B|H) - P(A \wedge B|H),$$

and negation by $P(\neg A|H) = 1 - P(A|H)$, and from there you're off to all propositional logic.

Now probability is a natural enough way for us, as mathematicians to think about this sort of thing, and in fact it's hard to think of a reason we would do anything else. In fact, there's a proof, due to Cox [**125**] that probability is enough. There's a bit of setup for it.

Let's use $L(A|B)$ to denote some "likelihood," some credibility that we assign to $A$ under the assumption of $B$. Probability is certainly the key example, but you could certainly think of others. For cheap ones, just take some function $f$ and take $L(A|B) = f(P(A|B))$. You could get something not unreasonable, for instance, by taking $f(x) = 100x$, or by thinking of "rareness" and mapping

$$f(x) = \begin{cases} \frac{1}{x} & \text{if } x > 10^{-100} \\ 10^{100} & \text{otherwise.} \end{cases}$$

And those are just the cheap ones.

A good start is to restrict what sorts of factors can be considered. One reasonable restriction arises from an accounting of the things that "should be" relevant, and restricting $L$ to only considering those. Take the upcoming election. You can certainly figure out the likelihood that Mitt Romney will win the electoral votes of both Ohio and Florida if you know his likelihood of winning Florida and his likelihood of winning Ohio given that he wins Florida. There may be other ways to calculate it, but if you know those two, then everything else has either been accounted for or is irrelevant. So we have the first axiom: There is some function $F$ such that

$$L(C \wedge B|A) \quad = \quad F\left(L(C|B \wedge A), L(B|A)\right). \tag{1}$$

If $L$ is probability, then we even know what this function is: it's multiplication.

If we also believe that $\wedge$ is associative, we get some restriction on $F$. If we start with $L(D \wedge C \wedge B|A)$, then by our axiom, it must equal

$$F\left(F\left(L(D|C \wedge B \wedge A), L(C|B \wedge A)\right), L(B|A)\right).$$

On the other hand, it must also equal

$$F\left(L(D|C \wedge B \wedge A), F\left(L(C|B \wedge A), L(B|A)\right)\right).$$

Consequently, we have

$$F\left(F(x,y),z\right) = F\left(x,F(y,z)\right).$$

LEMMA 2.2.1. *For any solution $F$ to the functional equation above, having second derivatives, there must exist a unary function $f$ and a constant $C$ such that $Cf(x,y) = f(x)f(y)$.*

PROOF. We start with the given equation,

$$F[F(x,y),z] = F[x,F(y,z)],$$

and let $F(x,y) = u$ and $F(y,z) = v$. Then our equation is

$$F(u,z) = F(x,v).$$

We differentiate with respect to $x$ and get

$$F_1(u,z)\frac{\partial u}{\partial x} = F_1(x,v).$$

Differentiating this with respect to $x$ gives

$$F_{11}(u,z)\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + F_1(u,z)\frac{\partial^2 u}{\partial x \partial y} = F_{12}(x,v)\frac{\partial v}{\partial y}.$$

If we differentiate each time with each of the three variables, we get a big mess, which simplifies to

$$\frac{\partial^2 u/\partial x\partial y}{\partial u/\partial x} - \frac{\partial^2 u/\partial y^2}{\partial u/\partial y} = \frac{\partial^2 v/\partial y\partial z}{\partial v}\partial z - \frac{\partial^2 v/\partial y^2}{\partial v/\partial y}.$$

Now we remember our logarithmic differentiation, and get

$$\frac{\partial}{\partial y}\ln\left(\frac{\partial u/\partial x}{\partial u/\partial y}\right) = -\frac{\partial}{\partial y}\ln\left(\frac{\partial v/\partial y}{\partial v/\partial z}\right).$$

By the definitions of $u$ and $v$, we have

$$\frac{\partial}{\partial y}\ln\left(\frac{F_1(x,y)}{F_2(x,y)}\right) = -\frac{\partial}{\partial y}\ln\left(\frac{F_1(y,z)}{F_2(y,z)}\right).$$

Since the $x$'s only appear on one side and the $z$'s on the other, each side of the equation must be determined entirely by $y$, so we set the left side equal to

$$\frac{d}{dy}\ln\Phi(y),$$

and the right hand side equal to its negative. By renaming variables in the equation for the right hand side, we can have

$$\frac{\partial}{\partial x}\ln\left(\frac{F_1(x,y)}{F_2(x,y)}\right) = -\frac{d}{dx}\ln\Phi(x).$$

We can now combine these to get

$$\frac{\partial}{\partial x}\ln\left(\frac{F_1(x,y)}{F_2(x,y)}\right)dx + \frac{\partial}{\partial y}\ln\left(\frac{F_1(x,y)}{F_2(x,y)}\right)dy = -d\ln\Phi(x) + d\ln\Phi(y).$$

We can integrate this, and get

$$\frac{F_1(x,y)}{F_2(x,y)} = C_0\frac{\Phi(y)}{\Phi(x)}.$$

When we combine this with all the first derivatives, we get

$$\frac{\partial F(x,y)}{\partial \cdot} = \frac{\Phi(F(x,y))}{\Phi(\cdot)}.$$

Now the equation comes up to be

$$\frac{dF}{\Phi(F)} = \frac{dx}{\Phi(x)} + \frac{dy}{\Phi(y)},$$

and we can integrate that to get $C_1 f(F) = f(x)f(y)$, as was to be shown. $\square$

As we already said, any $f$ will do, so we'll choose the identity. Then we have

$$C_1 F(p,q) = pq,$$

so that

$$F(p,q) = \frac{1}{C_1} pq,$$

and, in particular,

$$L(C \wedge B|A) = \frac{1}{C_1} \left(L(C|B \wedge A)\right) \cdot \left(L(B|A)\right).$$

This looks familiar.

We can also specify $C_1$ based on convention and first principles. Let $C = B$. Then this last equation is equivalent to

$$L(B|A) = \frac{1}{C_1} \left(L(B|B \wedge A)\right) \cdot \left(L(B|A)\right).$$

If we clear the denominator and divide both sides by $L(B|A)$, we get

$$C = L(B|B \wedge A).$$

So we have to pick what the likelihood of certainty is.

We're ready for the second assumption: There is some function $S$ such that

$$L(\neg B|A) \quad = \quad S(L(B|A)). \tag{2}$$

Intuitively, the likelihood of $B$ failing is determined by that of $B$ holding. Since we probably want $L(\neg\neg B|A) = L(B|A)$, we have $S(S(x)) = x$. Then by De Morgan's law (which we presumably also want),

$$S(L(C \vee B|A)) = L(\neg C \wedge \neg B|A).$$

But then using the previous part, we have

$$S(L(C \vee B|A)) = S(C|\neg B \wedge A) S(B|A).$$

With another application of $S$ and a little more work, we get

$$\frac{S(L(B|C \wedge A))L(C|A)}{S(L(B|A))} = S\left(\frac{S(L(C \vee B|A))}{S(L(B|A))}\right).$$

We can rewrite this to have the same given in all cases.

$$S\left(\frac{L(C \wedge B|A)}{L(C|A)}\right) L(C|A) = S\left(\frac{S(L(C \vee B|A))}{S(L(B|A))}\right) S(L(B|A)).$$

Now let $B = C \wedge D$, so that

$$S\left(\frac{L(C \wedge D|A}{L(C|A)}\right) L(C|A) = S\left(\frac{S(L(C|A))}{S(L(C \wedge D|A))}\right) S(L(C \wedge D|A)).$$

Now we have another functional equation, this time for S:

$$xS\left(\frac{S(y)}{x}\right) = yS\left(\frac{S(x)}{y}\right).$$

LEMMA 2.2.2. *Let $S$ be twice differentiable, and let it satisfy the previous equation. Then $S(p) = (1 - p^m)^{1/m}$.*

PROOF. Write $u := \frac{S(y)}{x}$ and $v := \frac{S(x)}{y}$. Then the equation is

$$xS(u) = yS(v).$$

If we differentiate by $x$ and then $y$, we get

$$\frac{uS''(u)S'(y)}{x} = \frac{vS''(v)S'(x)}{y}.$$

Multiplying by the previous version, we get

$$uS''(u)S(u)S'(y) = vS''(v)S(v)S'(x).$$

With the first derivatives, we can get

$$\frac{uS''(u)S(u)}{(uS'(u) - S(u))S'(u)} = \frac{vS''(v)S(v)}{(vS'(v) - S(v))S'(v)}.$$

Again, we have separated variables, so that the right hand side is equal to a constant $k$. Then

$$uS''(u)S(u) = k(uS'(u) - S(u))S'(u),$$

so that

$$\frac{dS'}{S'} = k\left(\frac{dS}{S} - \frac{du}{u}\right).$$

We can integrate to get

$$S' = A_0\left(\frac{S}{u})^k\right).$$

We can integrate again to get

$$S^{1-k} = A_0 u^{1-k} + A_1.$$

For this to actually satisfy the equation, $A_0 = -1$ and $A_1 = 1$. □

Now, in particular, $L(B|A)^m + L(\neg B|A)^m = 1$. We make another conventional assumption and assume $m = 1$. If $B = A$, then we find that the likelihood of impossibility is 0.

Now we also can get the usual inclusion-exclusion formula.

LEMMA 2.2.3. *The following hold:*

(1) $L(C \wedge B|A) + L(\neg C \wedge B|A) = L(B|A)$
(2) $L(C \vee B|A) = L(C|A) + L(B|A) - L(C \wedge B|A)$.

DEFINITION 2.2.4 ([**219**]). A conditional probability measure is a binary function $\mu : \mathcal{E} \times \mathcal{E} \to [0, 1]$ such that

(1) $\mu(U|U) = 1$ for any $U$
(2) If $V_1$ and $V_2$ are disjoint, then $\mu(V_1 \cup V_2) = \mu(V_1) + \mu(V_2)$
(3) $\mu(V|U) = \mu(V \cap U|U)$
(4) $\mu(U_1|U_3) = \mu(U_1|U_2) \times \mu(U_2|U_3)$ whenever $U_i \subseteq U_{i+1}$.

The function $L$ given above is a conditional probability measure.

THEOREM 2.2.5 (Cox, 1946). *Let $L(\cdot|\cdot)$ be a binary function on events such that*

(1) $\wedge$ *is associative in the second argument,*
(2) $\neg$ *is a reflection in the first argument,*
(3) *De Morgan's law* $\neg(C \vee B) \Leftrightarrow \neg C \wedge \neg B$ *holds in the first argument.*
(4) *There is a twice-differentiable function* $F$ *such that*

$$L(C \wedge B|A) = F(L(C|B \wedge A), L(B|A)),$$

(5) *There is a twice-differentiable function* $S$ *such that*

$$L(\neg B|A) = S(L(B|A)).$$

*Then* $L$ *is a conditional probability measure.*

Halpern noted some issues with this proof. The big one is that the functional equations are assumed to hold for all possible values. We could use continuity and deal with only a dense set of values. This means, among other things, that for any $U$ and any $x \in [0,1]$, there is some $V$ such that $L(V|U) = x$.

Also, it assumes that $L$ is real-valued. This means that every event is comparable in likelihood to every other event — even under any (different) conditions on either of them.

THEOREM 2.2.6 (Halpern 1999 [**217**]). *There is a function* $L_0$*, a finite domain* $W$*, and functions* $S$*,* $F$*, and* $G$ *such that*

(1) $L_0(\neg V|U) = S(L_0(V|U))$
(2) $L_0(V_1 \wedge V_2|U) = F(L_0(V_2|V_1 \wedge U), L_0(V_1|U)$
(3) $L_0(V|U) \in [0,1]$
(4) $S(x) = 1 - x$
(5) $G(x,y) = x + y$
(6) $F$ *is infinitely differentiable, strictly increasing in each argument, commutative,* $F(x,0) = 0$ *and* $F(x,1) = x$.
(7) $L_0(V_1 \vee V_2|U) = G(L_0(V_1|U), L_0(V_2|U))$.
(8) *There is no function* $f$ *making* $L$ *a conditional probability measure.*

PROOF. Let $W = \{w_1, \ldots, w_{12}\}$, and weight each point as follows:

$$\begin{aligned}
f(w_1) &= 3 \\
f(w_2) &= 2 \\
f(w_3) &= 6 \\
f(w_4) &= 5 \times 10^4 \\
f(w_5) &= 6 \times 10^4 \\
f(w_6) &= 8 \times 10^4 \\
f(w_7) &= 3 \times 10^8 \\
f(w_8) &= 8 \times 10^8 \\
f(w_9) &= 8 \times 10^8 \\
f(w_{10}) &= 3 \times 10^{18} \\
f(w_{11}) &= 2 \times 10^{18} \\
f(w_{12}) &= 14 \times 10^{18}
\end{aligned}$$

Define $f(U) = \sum_{w \in U} f(w)$, and define $P(U) = f(U)/f(W)$.

We will pick a certain $\delta$, and define

$$f'(w) = \left\{ \begin{array}{ll} (3-\delta) \times 10^{18} & \text{if } w = w_{10} \\ (2+\delta) \times 10^{18} & \text{if } w = w_{11} \; ll \\ f(w) & \text{otherwise} \end{array} \right. ,$$

and set $W' = \{w_{10}, w_{11}, w_{12}\}$. For any nonempty $U$, define

$$L_0(V|U) = \left\{ \begin{array}{ll} f'(V \cap U)/f(U) & \text{if } W' \subseteq U \\ f(V \cap U)/f(U) & \text{otherwise} \end{array} \right. .$$

Now pick $\Delta > 0$ so that whenever $P(V_1|U_1) > P(V_2|U_2)$ we have $L_0(V_1|U_1) > L_0(V_2|U|2)$

The remainder of the proof of the counterexample can be found in [**217**]. Further discussion of the conditions under which Cox's theorem does hold is found in [**218**]. □

Note that given an unconditional real-valued likelihood measure $L$, we can always define a conditional version with

$$L(A|B) = \frac{L(A \wedge \neg B) - L(\neg B)}{1 - L(\neg B)}.$$

Similarly, given a conditional likelihood measure, we can always define an unconditional version by conditioning on the entire sample space.

**2.2.2. Probability Logic.** In the most straightforward connection of probability to logic, and the one suggested by the preceding sections, we understand classical logic to have truth values at the two ends of the spectrum: certainty and impossibility. We then extend the possible truth values by assigning probabilities to sentences. There are several systems of syntax and semantics for this task in the literature. Although this interpretation goes back, as we have seen, deep in the history of probability, the root of current work in the area seems to be in the 1960's. At that time, Gaifman proposed assigning a measure to sentences.

Throughout, we let $L$ be a first-order signature, and $L(U)$ be an expansion of $L$ by constants $U$.

Definition 2.2.7. If $G$ is a set of sentences of $L(U)$, then we say that $\mu : G \to \mathbb{R}^{\geq 0}$, not everywhere zero, is a syntactic measure if and only if its domain is closed under boolean combinations and for all $\varphi, \psi \in G$, we have

(1) If $\vdash \varphi$ and $\vdash \psi$ then $\mu(\varphi) = \mu(\psi)$, and
(2) If $\vdash \neg(\varphi \wedge \psi)$ then $\mu(\varphi \vee \psi) = \mu(\varphi) + \mu(\psi)$.

We say that $\mu$ is a syntactic probability measure if and only if $\mu(\varphi) = 1$ whenever $\vdash \varphi$.

Gaifmain notes that a first-order theory gives a syntactic measure mapping the true sentences to 1 and the false ones to 0. It is natural to ask about the connection to Kolmogorov's continuity axiom (axiom 7 of the definition).

Lemma 2.2.8. *Let $\mu$ be a syntactic measure on the quantifier-free sentences of $L(U)$. Let $(\varphi_i : i \in \mathbb{N})$ be a sequence of quantifier-free sentences such that $\varphi_{i+1} \to \varphi_i$ for every $i$, with $\{\varphi_i : i \in \mathbb{N}\}$ inconsistent. Then* $\lim_{i \to \infty} \mu(\varphi_i) = 0$.

PROOF. By compactness, there must either be an $n$ for which $\{\varphi_i : i \leq n\}$ is inconsistent, or else $\{\varphi_i : i \in \mathbb{N}\}$ is consistent. Consequently, there must be some $n$ such that for $m > n$ we have $\mu(\varphi_m) = 0$. □

For the semantics of this approach, Gaifman proposed *probability models* of the following form.

DEFINITION 2.2.9. A probability model for $L$ is a pair $(U, \mu)$ where $U$ is a set of constants and $\mu$ is a syntactic probability measure on the quantifier-free formulas of $L(U)$.

There is, of course the problem of extending $\mu$ to sentences which are not quantifier-free. Gaifman gave the following characterization

THEOREM 2.2.10 (Gaifman1964). *Let $(U, \mu)$ be a probability model. Then there is a unique syntactic probability measure $\mu^*$ on the full set of sentences of $L(U)$ which extends $\mu$ and satisfies*

$$\mu^*(\exists x\varphi(x)) = \sup_{\bar{a} \subseteq U} \left\{ \mu^* \left( \bigvee_{i=1}^{n} \varphi(a_i) \right) \right\}.$$

PROOF. We show how to construct $\mu^*$ in the countable case. Let $\mathcal{K}$ be the set of all $L$-structures with universe $U$. We take $\mathcal{K}_0$ to be the Boolean algebra of subsets $B_\varphi$ of $\mathcal{K}$ defined by their satisfaction of a quantifer-free formula $\varphi$. We use $\mu$ to define a measure $\mu_1$ on $\mathcal{K}_0$.

Since $\mu_1$ is continuous (as per Lemma 2.2.8), it can be extended to a countably additive probability measure on the $\sigma$-algebra generated by $\mathcal{K}_0$. Since, for any quantifier-free $\varphi$, the set of structures satisfying $\exists x\varphi(x)$ is a countable union of elements of $\mathcal{K}_0$, we can define $\mu^*$ by induction on the quantifier depth. □

An important class of probability measures that we will explore at length in Chapter 6 consists of the *exchangeable* measures — that is, those invariant under permutation of the domain.

THEOREM 2.2.11 (Gaifman1964). *Let $\mu$ be a syntactic probability measure on $L(U)$. Then there is a set $V \supseteq U$ and a probability model $(V, m)$ agreeing with $\mu$ which is invariant under permutations of $V - U$.*

PROOF. Let $V$ be a nontrivial superset of $U$, and $S_D$ the set of permutations of $V - U$. We denote by $S_{D,0}$ the finitely supported elements of $S_D$. For each $\sigma \in S_{D,0}$ we define $[\sigma]$ to be the set of functions $\tau \in S_D$ such that $\tau \restriction_{dom(\sigma)} = \sigma$.

We now construct a measure on $S_D$ under which the $\sigma$-algebra generated by the sets $[\sigma]$ is measurable. Let $\{a_i : i \in \mathbb{N}\}$ be the distinct elements of $V - U$, and let $\nu$ be a countably additive probability measure on $V - U$ that gives each element positive measure. We can now interpret the product measure on $(V - U)^\omega$ as a measure on $S_D$ by identifying $\sigma \in F$ by $\sigma(\vec{a})$.

Let $m$ be a syntactic probability measure on $L(V)$. We define a syntactic measure $m_\nu$ on a sentence $\varphi(\bar{b})$, with $\bar{b}$ representing all members of $V - U$ occuring in it, by

$$m_\nu(\varphi(\bar{b})) = \sum_{\sigma \in S_{D,0}} \left( m \left( \varphi \left( \sigma(\bar{b}) \right) \right) \nu[\sigma] \right).$$

This "averaging" method achieves the necessary invariance, and also represents a strategy that we will see several times in Chapter 6. □

Since Gaifman, many others have approached this topic. Indeed, Williamson complains, "The problem is that the literature contains a plethora of new axiomatisations of probability on sentences, few of which bear a close resemblance [to] the mathematical formulation of probability" [**459**]. One approach that does follow mathematical probability closely is that of Adams, who requires the following for all sentences:

(1) $0 \leq P(\varphi) \leq 1$
(2) If $\varphi$ is logically true, then $P(\varphi) = 1$.
(3) If $\varphi$ logically implies $\psi$, then $P(\varphi) \leq P(\psi)$.
(4) If $\varphi$ and $\psi$ are logically inconsistent, then $P(\varphi \vee \psi) = P(\varphi) + P(\psi)$.

For obvious reasons, although carefully distinguishing his instantiation of the axioms by sentences, instead of sets, Adams calls these the Kolmogorov axioms. Frequently, Adams treats the *uncertainty* of a sentence $\varphi$, defined as $1 - P(\varphi)$, as the more fundamental concept, and demonstrates the following

THEOREM 2.2.12 ([**10**]). *Let* $\{\varphi_1, \ldots, \varphi_n\} \vdash \psi$.

(1) $1 - P(\psi) \leq \sum_{i=1}^{n} (1 - P(\varphi_i))$
(2) *Let* $u_1, \ldots, u_n$ *be nonnegative numbers with* $\sum_{i=1}^{n} u_i \leq 1$. *Assume also*
    (a) $\{\varphi_1, \ldots, \varphi_n\}$ *is consistent, and*
    (b) *For any* $T \subsetneq \{\varphi_1, \ldots, \varphi_n\}$ *we have* $T \nvdash \psi$.
    *Then there is a probability measure* $P$ *such that for all* $i$ *we have* $1 - P(\varphi_i) = u_1$ *and* $1 - P(\psi) = \sum_{i=1}^{n} u_i$.

Adams develops the theory in some detail, including a theory to recognize a change in probability assigned in the face of new informaiton. Hailperin proposed another probability logic similar in some ways to Gaifman's, and others have given a definition of *probabilistic logic* sufficient to interpret many of these attempts. In the framework of [**215**], we consider premises and conclusions of the form $P(\varphi) \in X$, where $X \subseteq [0, 1]$. Then in the logic of Gaifman, for instance, or in the logic proposed in [**216**], if $S$ is a set of sentences and $\psi$ is a sentence, then the entailment relation $S \models P(\psi) \in Y$ holds if every probability measure $P$ satisfying $S$, we have $P(\psi) \in Y$. The key enhancement in Hailperin's treatment is that the probabilities need not have sharp values.

DEFINITION 2.2.13 ([**215**]). A *probabilistic argumentation system* is a triple $\mathcal{A} = (V, W, \Phi, P)$, where

(1) $V$ is a set of propositional atoms, with $W \subseteq V$
(2) $\Phi$ is a set of sentences in $V$
(3) $P$ is a probability measure defined on subsets of the space $2^W$ of joint values for the elements of $W$.

In this context, for each propositional formula $\psi$, we define the set $A_{\mathcal{A}}(\psi)$ to be the set of $\sigma \in 2^V$ such that if the values of $\sigma$ are instantiated for the atoms $V$, then the formula $\psi$ is satisfied. We further define the *degree of support* for $\psi$ by

$$d_{s,\mathcal{A}}(\psi) = \frac{P(A_{\mathcal{A}}(\psi)) - P(A_{\mathcal{A}}(\perp))}{1 - P(A_{\mathcal{A}}(\perp))},$$

and the *degree of possibility* for $\psi$ by

$$d_{p,\mathcal{A}}(\psi) = \frac{1 - P(A_{\mathcal{A}}(\neg\psi))}{1 - P(A_{\mathcal{A}}(\perp))}.$$

This approach is equivalent to the belief functions of Dempster and Shafer.

DEFINITION 2.2.14 ([**136, 398**]). Let $\Omega$ be a set. Consider a partial function $B : \mathcal{P}(\Omega) \to [0,1]$. We say that $B$ is a *belief function* if and only if the following conditions hold:

(1) $B(\emptyset) = 0$
(2) $B(\Omega) = 1$
(3) $B\left(\bigcup_{i=1}^{n}\right) \geq \sum_{i=1}^{n}(-1)^{i+1} \sum_{|I|=i} B\left(\bigcap_{j\in I} U_j\right)$.

The connection arises from the following result.

THEOREM 2.2.15 ([**214, 215**]). *For any probabilistic argumentation system $\mathcal{A}$ there exists an equivalent Dempster-Shafer belief function $B$ such that $d_{s,\mathcal{A}} = B$, and conversely.*

It is worth repeating: there are many probability logics in the literature, many of them with similar names and many with very similar (but seldom identical formulations. Much of the work around them appears to happen in the literature of philosophical logic and artificial intelligence, and somewhat less, to date, in the literature of mathematical logic. While these boundaries are, of course, imprecise, a notable exception in recent years has been Paris and Vencovská's recent volume merging the more recent probability logics with the work on first-order predicate logic by Gaifman, with which this section started [**359**].

## 2.3. Axiomatizing and Computing Conditionality

**2.3.1. Independence and Bayesian Networks.** Kolmogorov identified independence as the central feature to distinguish probability in particular from measure theory in general, pointing out that the major theorems of probability known at his time generally either assumed independence as a hypothesis or introduced some weaker condition similar to it. In practical terms, too, dependence and independence have a special place. Much of the applicability of probability turns on the question of predicting one event with greater certainty than would otherwise be possible, using information about another event. For this reason, there has been substantial work in the artificial intelligence community to formalize this notion.

A major direction of this work has been Pearl's concept of Bayesian networks, thich are intended to simplify conditional probability computations by representing the dependence relations efficiently [**361**]. A full table of the joint distribution even of $n$ binary variables would be rather alrge for computation, and even with that distribution represented, the computation of even a simple conditional probability would be slow.

The first thing to typically do, then, is to compute whether things are independent, and ignore irrelevant information. Since computing the full joint distribution, and then computing from it the conditional and unconditional probabilities, as in the traditional definitions of independence, is computationally unfeasible (for instance, the full joint distribution table for $n$ Boolean variables would have $2^n$ entries), there must, Pearl reasoned, be some other approach.

Definition 2.3.1. We say that a ternary relation $I(X, Y, Z)$ on sets is an independence relation if and only if the following conditions hold:

(1) $I(X, Z, Y)$ if and only if $I(Y, Z, X)$.
(2) If $I(X, Z, Y \cup W)$, then $I(X, Z, Y)$ and $I(X, Z, W)$.
(3) If $I(X, Z \cup W, Y)$ and $I(X, Z \cup Y, W)$, then $I(X, Z, Y \cup W)$.
(4) If $I(X, Z, Y \cup W)$, then $I(X, Z \cup W, Y)$.
(5) If $I(X, Z, Y)$ and $I(X, Z \cup Y, W)$, then $I(X, Z, Y \cup W)$.

It is routine to verify that for any probability measure $P$, the relation $I_P(X, Y, Z)$ defined by $P(X|Y, Z) = P(X|Z)$ is an idependence relation. It should be noted that this differs from the definition of independence relation in model theory; in addition to the concepts like invariance that would require additional context even to formulate, transitivity is lacking. The first critical issue is the question of whether this characterization precisely captures the notion of conditional independence.

The following "Completeness Conjecture" was initially put forward, and would settle the matter were it true.

Conjecture 2.3.2 ([**362**]). *Let $I$ be an independence relation. Then there is a probability space $(\Omega, \mathcal{B}, P)$ such that $I$ is a relation on $\mathcal{B}^3$ and coincides exactly with probabilistic independence on $(\Omega, \mathcal{B}, P)$.*

As we will see, this conjecture is not quite true, but some important variants of it are true. A basic Horn formula is a first-order formula of the form $\left(\bigwedge_{i=0}^{n} \theta_i\right) \to \psi$, where each $\theta_i$ is atomic and $\psi$ is atomic. A Horn formula is a basic Horn formula, perhaps bound by some global quantifiers. All of the clauses in Definition 2.3.1 are (or can be expressed by) universal Horn formulas in the language with one ternary relation $I$.

Theorem 2.3.3 ([**427**]). *There is no finite Horn axiomatization satisfied exactly by the ternary relations arising as the conditional independence relation of some probability space.*

Proof. Let $N \in \mathbb{N}$, and $3 \leq k < N$. Let

$$S(j) := \begin{cases} j+1 & \text{if } j \in \{1, \ldots, n-1\} \\ n & \text{if } j = n \end{cases} ,$$

and $I_\mu$ the conditional independence relation given by probability measure $\mu$ on $\{0, \ldots, N\}$. Then we can show that the following conditions are equivalent:

(1) For all $j \in \{1, \ldots, k\}$, we have $I_\mu(\{0\}, \{j\}, \{S(j)\})$.
(2) For all $j \in \{1, \ldots, k\}$, we have $I_\mu(\{0\}, \{S(j)\}, \{j\})$.

Now let $\varphi_1, \ldots, \varphi_\ell$ be Horn formulas. Then we will find some independence relation $I_*$ satisfying $\bar{\varphi}$ but not arising as $I_\mu$ for any $\mu$. Pick $n$ large enough to exceed the number of $\theta_i$ conjuncts in the antecedents of all $\varphi_j$. We define $I_*(A, B, C)$ to hold of a trible $(A, B, C)$ of subsets of $\{1, \ldots, n\}$ in exactly the following cases:

(1) $(A, B, C) = (\{0\}, \{S(j)\}, \{j\})$ for some $j \in \{1, \ldots, n\}$.
(2) $(A, B, C) = (\{j\}, \{S(j)\}, \{0\})$ for some $j \in \{1, \ldots, n\}$.
(3) $A, B$, and $C$ are pairwise disjoint and nonempty.

This $I_*$ does not satisfy the conditions above, but most satisfy $\varphi_1, \ldots, \varphi_\ell$. □

There are some partial recoveries froom this obstruction, however. The following two results constitute a sort of completeness theorem with respect to a slightly exotic logic based on graphs.

DEFINITION 2.3.4. Let $G = (V, E)$ be an undirected graph, and $I : \mathcal{B}^3 \to \{0, 1\}$ an independence relation on a measurable space $(\Omega, \mathcal{B})$.

(1) We say that $G$ is a dependency map of $I$ if and only if there is a bijection $f : \Omega \to V$ such that if $I(X, Z, Y)$ then all paths in $G$ from elements of $f(X)$ to elements of $f(Y)$ pass through elements of $f(Z)$.
(2) We say that $G$ is an independency map of $I$ if and only if there is a bijection $f : \Omega \to V$ such that if all paths in $G$ from elements of $f(X)$ to elements of $f(Y)$ pass through elements of $f(Z)$, then $I(X, Y, Z)$.
(3) We say that $G$ is a perfect map of $I$ if and only if it is both a dependency map and an independency map.

In this sense, a graph which is a perfect map of $I$ constitutes a syntactic representation of the independence relation $I$. The completeness result for this logic is as follows.

THEOREM 2.3.5 ([**362**]). *Let $(\Omega, \mathcal{B})$ be a measurable space, and let $I : \mathcal{B}^3 \to \{0, 1\}$. Then the following are equivalent:*

(1) *There is an undericted graph that is a perfect map of $I$.*
(2) *$I$ satisfies the following conditions:*
    (a) *$I(X, Z, Y)$ if and only if $I(Y, Z, X)$*
    (b) *If $I(X, Z, Y \cup W)$, then $I(X, Z, Y)$ and $I(X, Z, W)$.*
    (c) *If $I(X, Z \cup W, Y)$ and $I(X, Z \cup Y, W)$, then $I(X, Z, Y \cup W)$.*
    (d) *If $I(X, Z, Y)$, then $I(X, Z \cup W, Y)$.*
    (e) *If $I(X, Z, Y)$ then either $I(X, Z, W)$ or $I(W, Z, Y)$*

PROOF. Vertex separation in any graph satisfies the stated condition, so one direction of implication is clear. In the other direction, we construct a graph. To do this, we start with a complete graph with a vertex for each element of $\mathcal{B}$, and delete each edge $(X, Y)$ where there is some $Z \in \mathcal{B}$ with $I(X, Z, Y)$. □

This gives rise to a representation of independence relations as graphs, and even a notion of maximal parsimony — avoiding the computational complexity explosion inherent in working with a full joint distribution table.

DEFINITION 2.3.6. A Markov Network for $I$ is an independency map $G$ of $I$ with the property that deleting any edge would result in a graph which is not an independency map.

THEOREM 2.3.7 ([**362**]). *Every function $I$ satisfying the equivalent conditions of Theorm 2.3.5 has a unique Markov network.*

PROOF. For each element $x \in \Omega$, a Markov blanket of $x$ is a subset $S \subseteq \Omega$ such that $I(\{x\}, S, U - (S \cup \{x\}))$. A Markov boundary of $x$ is a minimal Markov blanket. We can show that the Markov boundary of an element is unique. We then set exactly the elements of the Markov boundary of $x$ adjacent to $x$. □

The difficulty with Markov netwoks is that the transitivity condition is too strong. Of course, Theorem 2.3.7 does not assume transitivity, but a form of it comes back in when a graph is a perfect map of $I$. If $A$ occurs exactly when unconditionally independent events $B_1$ and $B_2$ occur, then an independence map will require an edge between $B_1$ and $B_2$, introducing a spurious dependency. This issue can be addressed by replacing undirected graphs by directed acyclic graphs.

Definition 2.3.8. Let $G = (V, E)$ be a directed acyclic graph.

(1) We say that $G$ to be an independency map of an independence relation $I$ if and only if there is a bijection $f : \Omega \to V$ such that for any disjoint sets $X, Y, Z$, if all paths in $G$ from elements of $f(X)$ to elements of $f(Y)$ pass through elements of $f(Z)$, then $I(X, Y, Z)$.
(2) We say that $G$ to be a dependency map of an independence relation $I$ if and only if there is a bijection $f : \Omega \to V$ such that for any disjoint sets $X, Y, Z$, if $I(X, Y, Z)$ then all paths in $G$ from elements of $f(X)$ to elements of $f(Y)$ pass through elements of $f(Z)$.
(3) We say that $G$ is a perfect map of $I$, if and only if it is both a dependency map and in independency map.
(4) We say that $G$ is a minimal independency map of $I$ if and only if it is an independency map and no edge can be deleted without losing this property.
(5) We say that $G$ is a Bayesian network of a probability measure $P$ if and only if it is a minimal independency map of the conditional independence relation $I_P$.

Theorem 2.3.9. *Let $(\Omega, \mathcal{B}, P)$ be a probability space. Then the following conditions on a directed acyclic graph $G$ are equivalent:*

(1) *$G$ is a Bayesian network of $P$.*
(2) *Each measurable set $X \in \mathcal{B}$ is conditionally independent of its non-descendents given its parents, and no proper subset of the parent relation satisfies this property.*

We now arrive at the canonical completeness result for Bayesian networks.

Theorem 2.3.10 ([**194**]). *For any directed acyclic graph $G = (V, E)$ there exists a probability space $(\Omega, \mathcal{B}, P)$ such that $\mathcal{D}$ is a perfect map of $P$.*

Proof. For any particular sets $X, Y, Z$ of vertices of $G$, with some path from $X$ to $Y$ not passing through $Z$. Then there is a probability measure on the vertices of $G$ such that $I_P(X, Z, Y)$ fails. We describe how to do this in the case of a finite graph. We associate with each element of $V$ a binary random variable, and partition $V$ as $V_1 \cup V_2$ where $V_1$ contains exactly the $n_1$ vertices on the path from $X$ to $Y$ avoiding $Z$ and $V_2$ has cardinality $n_2$. Preliminary to describing the full joint distribution we construct, set

$$f(x, y) = \begin{cases} \frac{1}{2} & \text{if } x = y = 0 \\ \frac{3}{4} & \text{if } x = y = 1 \\ \frac{1}{4} & \text{if } x = 1 \text{ and } y = 0 \\ \frac{1}{2} & \text{if } x = 0 \text{ and } y = 1 \end{cases}.$$

We now set

$$P(\bar{x}) = \left(\frac{1}{2}\right)^{n_2+1} \prod_{i=1}^{n_1-1} f(x_i, x_{i+1}).$$

We take the product of all such probability spaces $(V, \mathcal{F}, P)$ to obtain a single probability space satisfying exactly the desired independence relation. □

A Bayesian network does not require computation or storage of the full joint distribution, but only of the conditional distribution of each vertex conditioned on its parents. There are standard algorithms for constructing Bayesian networks, but their results depend on the ordering of the variables. Informally, it appears that the most compact (i.e. sparse) representations are achieved when causes precede effects.

To generate a Bayesian network for a finite system of variables, we first order the variables $X_1, \ldots, X_n$. Then at each stage $s$, choose, from $X_1, \ldots, X_{s-1}$ a minimal set $\Pi_s$ of variables such that $X_s$ is conditionally independent of $X_1, \ldots, X_{s-1}$ given $\Pi_s$. For each $j$ with $X_j \in \Pi_s$, we make an edge from $X_j$ to $X_s$, and no other edges to $X_s$. While the traditional applications involve finite sets of variables [**386**], this algorithm would still produce a computable Bayesian network on a countably infinite set of variables, provided the necessary conditional probability computations could be made — which, as we will see, is not always the case, even for finite sets of variables.

Pearl observed that the method of Bayesian networks could be extended to systems involving continuous random variables. His system, treated only in outline in [**361**] does not seem to have attracted as much further study as the similar models arising from contemporary work on conditional Gaussian distributions [**298**]. Instead of Boolean-valued random variables with distribution conditioned on the parent vertices, each vertex has a continuous (in Pearl's early examples, normal) distribution conditioned on a linear combination of the parent vertices.

There is, in fact, a difficulty in computing conditional probabilities. To express this difficulty, we introduce some terminology. If $\Omega$ and $X$ are computable Polish spaces (in the sense that there is a computably enumerable dense subset, called the *ideal points*, on which the distance function produces uniformly computable real numbers), then $f : \Omega \to X$ is said to be computable on a set $S \subseteq \Omega$ exactly when there is a compuable sequence $(U_n : n \in \mathbb{N})$ of computably enumerable open sets in $X$ such that for any ideal point $q$ of $X$, any rational $r > 0$, and any $n \in \mathbb{N}$, we have $f^{-1}(B_r(q)) \cap S = U_n \cap S$. We say that a function is *almost computable* if and only if it is computable on a set of full measure in $\Omega$, and *almost continuous* if and only if it is continuous on such a set. In that sense, the following result demonstrates in a fairly strong sense that having all information we might wish about random variables does not allow us to compute the conditional distribution.

THEOREM 2.3.11 ([**8**]). *There are $P$-almost computable $[0,1]$-random variables $X$ and $Z$ such that the conditional distribution map $z \mapsto P(X|Z=z)$ is $P_Z$-almost continuous but not $P_Z$-almost computable.*

PROOF. Let $N, C, X, U$, and $V$ be independent computable random variables with the following distributions:

- $N$**:** geometric distribution with parameter $p = \frac{1}{5}$
- $C$**:** Bernoulli distribution with parameter $p = \frac{1}{3}$
- $X, U, V$**:** each with uniform distribution taking values in $[0,1]$.

We produce the random variable $Y$ as follows. Define

$$Y_k = \frac{2\lfloor 2^k V \rfloor + C + U}{2^{k+1}},$$

and note that $\lim\limits_{k\to\infty} Y_k = V$ almost surely. Call this limit $Y_\infty$. We compose the variable $Y$ as a random mixture (depending on the random variable $N$) of the variables $\{Y_i : i \in \mathbb{N} \cup \{\infty\}\}$ by setting $h_n$ to be the time by which the $n$th Turing machine halts on input 0, or $\infty$ if it does not halt, and setting $Y = Y_{h(N)}$. Now we can show that this $Y$ is almost continuous.

As usual, we can simulate a geometric random variable $M$ with parameter $\frac{1}{2}$ by

$$M := \left\lfloor \frac{\log X}{\log\left(1 - \frac{1}{2}\right)} \right\rfloor.$$

We define $Z$ by $P(Z|Y)(x) = P\left(X|N = M(x)\right)$.

Now $P\left(X \in (2^{-n-1}, 2^{-n})|Z = x\right) = P(N = n|Y = x)$. From computing this, even on a set of full measure, we could compute $\emptyset'$. □

Finally, we observe a connection of Bayesian networks with the probabilistic argumentation systems of Section 2.2.2. To determine $P(\psi)$ from assumptions on the values of $P(\varphi_i)$ for each $i$, we construct a Bayesian network whose vertices are the propositional atoms $V$ of the probabilistic argumentation system. It is customary to assume that the independence relation on the propositional atoms is known *a priori*.

To compute the conditional probabilities represented by the network from the assumptions on $P(\varphi_i)$, we rewrite each $\varphi_i$ in disjunctive normal form as $\varphi_i = \bigvee_m \bigwedge_n A^{\pm}_{s_i(n)}$, where $s_i(n)$ enumerates propositional atoms of increasing rank (taking, as usual, the vertices with no edges in to have rank zero, and each other vertex to have the least rank strictly greater than that of any of its parents).

Now we have

$$P(\varphi_i) = \sum_m \prod_n \gamma_{n,m},$$

where $\gamma_{n,m}$ is the probability of $A^{\pm}_{s_i(n)}$ given its parents in the networks. These equations, as $i$ ranges over the set of assumptions, give constraints on the variables $\gamma_{n,m}$. By reversing the procedure, these constraints can be used to give constraints on $P(\psi)$. The full computation can be laborious, and is explained in full detail in [**215**].

**2.3.2. Nonmonotonic Logic and Conditionality.** It is perhaps not surprising that conditional probability can interpret certain nonmonotonic reasoning, since the standard examples generally refer to probabilistic judgments. Most people are willing, in general, given that $x$ is a bird, to deduce that $x$ can fly. On the other hand, if afterward informed that $x$ is one of the critically endangered kakapo, they would (perhaps after looking the matter up) not make that deduction. Such reasoning is called "nonmonotonic" because an increase in the information available may cause a recision of deductions that were valid from a smaller set of information.

There are well-documented logics to describe such reasoning. The subject in general is beyond the scope of the present book, but is treated in some detail in [**320**]. What will concern us here is Pearl's development of default logic through a semantics of conditional probability, and some exposition of the background theory of nonmonotonic reasoning and default logic will be given in service of that goal.

The syntax of (propositional) default logic has two types. One type consists of the well-formed formulas of the propositional language. The other consists of defaults. A default is a triple $\langle \alpha, J, \gamma \rangle$, where $\alpha$ and $\gamma$ are formulas and $J$ is a finite set of formulas. We then define the appropriate proof system.

DEFINITION 2.3.12. Let $T$ be a classical propositional theory, $W$ a set of propositional sentences, and $D$ be a set of defaults. A proof of $\varphi$ from $W$ relative to $T$ consists of a finite sequence $\varphi_1, \ldots, \varphi_n$ such that $\varphi_n = \varphi$ and such that for every $i$, one of the following holds:

(1) $\varphi_i \in W$
(2) $\varphi_i$ follows from $\{\varphi_1, \ldots, \varphi_{i-1}\}$ by the usual rules of propositional logic
(3) There is a default $\langle \alpha, J, \varphi_i \rangle \in D$ such that $\alpha \in \{\varphi_1, \ldots, \varphi_{i-1}\}$ and $J \cap T = \emptyset$.

Of course, the semantics for such a system will be a good deal more subtle than for classical propositional logic itself. In the example of the bird, we might express the "default" by saying that "Typically, birds are able to fly, unless they are either of a flightless species or wounded." (The problems of expressing these assertions exactly in propositional, as opposed to predicate, logic are well-understood, and need not detain us here; in any case, the example is standard.) This expresses the matter well enough in the sense that as the contextual theory $T$ changes we might go from concluding that $x$ can fly to not drawing that conclusion if $T$ grows to include the statement "$x$ is a kakapo." But to put semantics on this deductive system, we have to put a more concrete meaning to "Typically."

Since the goal for Marek and Truszczyński was logic programming, semantics were not an essential feature of their treatment. In Reiter's seminal paper on default logic, he also devotes primary place to the syntactical realm [**376**], although in a later survey he does explicitly warn against an exclusively "statistical" reading of defaults [**377**]. Indeed, it is not fair to interpret "Typically $A$" unequivocally as a mere statement that the probability of $A$ is high.

Pearl nevertheless gives a semantics for defaults in terms of limiting conditional probabilities. We introduce formulas of the form $P(A|B) \geq r$, where $A, B$ are sentences and $r \in [0,1]$. For default set $D$ whose defaults are all of the form $\langle A, \emptyset, B \rangle$, we compose the set

$$D_\epsilon := \{P(B|A) \geq 1 - \epsilon : \langle A, \emptyset, B \rangle \in D\} .$$

The semantic notion of entailment is then given by writing $(T, D) \models \varphi$ if and only if for every probability measure $P$ satisfying all conditions in $D_\epsilon$ also satisfies $P(\varphi|T) \geq 1 - O(\epsilon)$. We likewise say that a probability measure $P$ is a model of $(T, D)$ if and only if $P(\varphi) = 1$ for all $\varphi \in T$ and $P$ satisfies $D_\epsilon$ for every $\epsilon \geq 0$.

The calculation including our kakapo now goes in the following way. We will make reference to a bird named "Lisa" by researchers, who has played a pivotal role in the conservation of her species. We set

$$D := \{\langle \text{bird}, \emptyset, \text{fly} \rangle, \langle \text{kakapo}, \emptyset, \neg\text{fly} \rangle\} ,$$

and

$$T := \{\text{kakapo} \to \text{bird}\} .$$

Now from the axiom "Lisa is a bird," we can certainly write a proof in default logic of "Lisa can fly." We will denote "Lisa can fly" by "fly," and similarly for other properties. For any model $P \models (T, D)$, we have $P(\text{fly}|\text{bird}) \geq 1 - \epsilon$ for each $\epsilon > 0$. On the other hand,

$$P(\text{fly}|\text{bird}, \text{kakapo}) = \frac{P(\text{fly}|\text{kakapo})}{P(\text{bird}|\text{kakapo})} \leq \epsilon.$$

The system of derivations proposed here by Geffner and Pearl is not transparently identical to that of Reiter, Marek, and Truszczyński. Nevertheless, he does state a derivation system which is sound and complete with respect to this semantics, and which has some similarities to theirs.

THEOREM 2.3.13 ([**193**]). *Let $(T, D)$ be a default theory, where every element of $D$ is of the form $\langle a, \emptyset, b, \rangle$. Then for any sentence $\varphi$, we have $(T, D) \models \varphi$ if and only if $\varphi$ is derivable from $(T, D)$ using the following rules of inference:*

(1) *If $p \to q$, then from $p$ we can derive $q$.*
(2) *For each default $\langle p, \emptyset, q\rangle$, from $p$, we can derive $q$.*
(3) *If from $p$ we can derive each of $q, r$, then from $p \wedge q$ we can derive $r$.*
(4) *If from $p$ we can derive $q$ and from $p \wedge q$ we can derive $r$, then from $p$ we can derive $r$.*
(5) *If from each of p,q we can derive $r$, then from $p \vee q$ we can derive $r$.*

The default rule of this system certainly matches the appropriate special case of that from Definition 2.3.12, and the remaining rules are all sound derivations of propositional calculus. Conversely, by clause 1, all propositional calculus derivations are valid here. Thus, the Geffner-Pearl system coincides with the special case of Definition 2.3.12 in which $D$ consists of elements of the form $\langle a, \emptyset, b, \rangle$. The connection to Pearl's graph semantics for conditional probability is made by the following result.

THEOREM 2.3.14 ([**9, 360**]). *Let $D$ be a set of defaults of the form $\langle a, \emptyset, b\rangle$ with $a$ and $b$ atomic or negation atomic propositions. Then the following are equivalent.*

(1) *For each $\epsilon$, the set of $P$ such that $P \models D$ is non-empty.*
(2) *For every nonempty subset $D_0 \subseteq D$, there is a truth assignment $\sigma$ such that for some $\langle a, \emptyset, b, \rangle \in D_0$ we have $\sigma \models a \to b$, and for no such element do we have $\sigma \models a \wedge \neg b$.*
(3) *For every pair $\langle a_1, \emptyset, b\rangle, \langle a_2, \emptyset, \neg b\rangle$, the following hold:*
    (a) $a_1 \neq a_2$
    (b) *There is no cycle of the form $\langle k_1, \emptyset, k_2\rangle, \langle k_2, \emptyset, k_3\rangle \cdots \langle k_n \emptyset k_1\rangle$ in which both $a_1$ and $a_2$ are among the $k_i$.*

The final clause of this last result justifies the representation of $D$ as a directed graph that forbids certain kinds of cycles.

## 2.4. Adapted Spaces and Distributions

We now begin to consider a different logical approach to probability. It represents a significant departure from classical syntax, in exchange for modeling random variables and stochastic processes in a way that it much more tightly connected to the modern view of probability. In addition to being interesting in its own right, it is also an important precurser to the continuous first-order logic of Seciton 2.6. Fajardo and Keisler considered stochastic processes, with a focus on results that can be established in nonstandard analysis. Their process considers stochastic processes in an appropriate structure, and then passes to a saturated structure. Much of this work is accumulated in [**166**].

Exactly what sort of structure is appropriate is perhaps a natural logical question, but depends, of course, on the objects under study.

DEFINITION 2.4.1. Let $(\Omega, \mathcal{F}, P)$ be a probability space.

(1) An *X-valued stochastic process* is a function $x : \Omega \times [0,1] \to X$, where for each $t$, the function $x(\cdot, t) : \Omega \to X$ is measurable.
(2) We say that a stochastic process $x$ is *measurable* if it is measurable with respect to the product $\sigma$-algebra (using the Borel $\sigma$-algebra on $[0,1]$).

It is also standard in modern investigations of stochastic processes to consider the probability space with a *sequence* of $\sigma$-algebras, $(\mathcal{F}_t : t \in [0,1])$ with $\mathcal{F}_{t_1} \subseteq \mathcal{F}_{t_2}$ whenever $t_1 < t_2$. This is often said to capture the notion of added information over time, in the sense that the conditional distribution $P(x|\mathcal{F}_{t_2})$ caries more information than $P(x|\mathcal{F}_{t_1})$.

DEFINITION 2.4.2. An *adapted space* is a triple $(\Omega, (\mathcal{F}_t : t \in [0,1]), P)$, where

(1) For each $t < 1$, we have $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$,
(2) $\mathcal{F}_0$ contains all subsets of measure-zero sets.

It is these structures that constitute the structures of the logic to be described. The next step is to describe the formulas of the appropriate logic. The definition arises from the observation that two random variables $x, y$ are equidistributed if and only if for every bounded continuous function $\varphi$, we have $E(\varphi(x)) = E(\varphi(y))$.

DEFINITION 2.4.3. Let $x$ be a stochastic process defined on an adapted space $(\Omega, (\mathcal{F}_t : t \in [0,1]), P)$. We define the set of *adapted functions* by induction as follows.

(1) Every bounded continuous function $\varphi$ from $\mathbb{R}^n$ to $\mathbb{R}$, in conjunction with an $n$-tuple of check-points $\bar{t} \in [0,1]^n$, gives rise to an adapted function $\varphi_{\bar{t}}$. We evaluate $\varphi_{\bar{t}}$ on $x$ by the random variable $\varphi_{\bar{t}}(x)$ which arises from evaluating $x$ at each of the check-points, and then evaluating $\varphi$ on that $n$-tuple.
(2) If $\psi$ is a bounded continuous function from $\mathbb{R}^n$ to $\mathbb{R}$, and $f_1, \ldots, f_n$ are adapted functions, then $\psi(\bar{f})$ is an adapted function, evaluated in the obvious way.
(3) If $f$ is an adapted function and $t \in [0,1]$, then the expression $E(f|t)$ is an adapted function, defined (up to almost everywhere equivalence) by

$$E(f|t)(x)(\omega) = E(f(x)|\mathcal{F}_t)(\omega)$$

.

Now we define a fundamental equivalence relation.

DEFINITION 2.4.4. Let $x_i$ be a stochastic process on an adpated space $\Omega_i$. We say that $(\Omega_1, x_1) \equiv (\Omega_2, x_2)$ if and only if for each adapted function $f$, we have $E(f(x_1)) = E(f(x_2))$. Where the context is obvious, we sometimes write $x_1 \equiv x_2$.

As the notation suggests, this relation will serve in the place of elementary equivalence. It does, indeed, preserve many important properties.

A *martingale*, which we will see in a few contexts over the course of the present volume, is a stochastic process $x$ with the following two additional properties:

(1) For any fixed $t \in [0,1]$, we have $E(x((t)) < \infty$, and
(2) For any $t \in [0,1]$ and any $s < t$, we have $E(x(t)|\mathcal{F}_s) = x(s)$.

PROPOSITION 2.4.5. *If $x$ is a martingale and $x \equiv y$, then $y$ is a martingale.*

PROOF. Note that the truncated absolute value function

$$|x|_n = \begin{cases} |x| & \text{if } |x| \leq n \\ n & \text{otherwise} \end{cases}$$

gives rise to an adapted function $|\cdot|_{n,t}$. Thus,

$$\lim_{n\to\infty} \int |x|_{n,t} dP = \lim_{n\to\infty} \int |y_{n,t}| dP$$

since $x \equiv y$, making the individual terms in the limit equal. By the dominated convergence theorem, $E(y(t)) < \infty$ for every $t$.

Using a similar truncation strategy, we can reason from

$$\int |E\left(x(t)|\mathcal{F}_s\right) - x(s)|\, dP = 0$$

to the corresponding equation for $y$. □

This shows that $\equiv$ preserves at least some of the important properties. Similarly, we have the following.

PROPOSITION 2.4.6. *If, for all $t$ and almost all $\omega$, we have $x(\omega, t) = y(\omega, t)$, then $x \equiv y$.*

The definition of equivalence can be adapted to consider pairs of stochastic processes. If $x_1, x_2$ both take values in $X$, then we can consider the pair $(x_1, x_2)$ as a stochastic process taking values in $X^2$. If $X$ is a Polish space, then $X^2$ is, as well, so that the formulation of adapted functions in terms of continuous bounded functions still makes sense.

DEFINITION 2.4.7. An adapted space $\Omega$ is said to be saturated if for any adapted space $\Lambda$ and any pair of stochastic processes $x_1, x_2$ on $\Lambda$, the following holds:

- For every $y_1$ on $\Omega$ with $y_1 \equiv x_1$, there is a process $y_2$ on $\Omega$ such that $(x_1, x_2) \equiv (y_1, y_2)$.

A notion of homogeneity can be defined in a similar way (with almost sure bijections which preserve measure and $\mathcal{F}_t$ for all $t$ taking the role of automorphisms), as well as a back-and-forth relation and an appropriate version of Ehrenfeucht-Fraïssé games. Saturated adapted spaces exist, but the most obvious adapted spaces are not saturated.

PROPOSITION 2.4.8. *Let $(\Omega, (\mathcal{F}_t : t \in [0, 1]), P)$ be an adapted space with $\Omega$ a Polish space and $P$ the completion of a probability measure on the family of Borel sets. Then $\Omega$ is not saturated.*

It will be noticed that the set of adapted functions is unusually large for the set of formulas of a logic. Fortunately, it is not always necessary to consider the full set of such functions.

DEFINITION 2.4.9. We say that a set $F$ of adapted functions is *dense* if for any stochastic processes $x$ and sequence $(x_n : n \in \mathbb{N})$ of stochastic processes, then if $E(f(x_n)) \to E(x)$ for every $f \in F$, then the convergence holds for every adapted function $f$.

We can always take a countable set of bounded continuous functions that generate a dense set of adapted functions.

In considerations of definability, the base notion in classical model is the definable set. It is probably less natural to think of definable subsets of an adapted space, and, in any event, farther from the motivating examples of stochastic processes. Instead, we say what it means for one stochastic process to be definable from another.

DEFINITION 2.4.10. Let $(\Omega, (\mathcal{F}_t : t \in [0,1]), P)$ be an adapted space and $x$ a stochastic process on $x$.

(1) The *intrinsic $\sigma$-algebra* of $x$ on $\Omega$, denoted $\mathcal{I}_x$ is the $\sigma$-algebra generated by the null sets and the sets of the form $f(x)$ where $f$ ranges over all adapted functions.
(2) We say that $y$ is definable from $x$ if and only if $\mathcal{I}_y \subseteq \mathcal{I}_x$.

An important property of definable stochastic processes is the following argument that facilitates some back-and-forth arguments.

THEOREM 2.4.11. *Let $x_i$ be a stochastic process on $\Omega_i$. If $x_1 \equiv x_2$ and $y_1$ is a stochastic process on $\Omega_1$ definable from $x_1$, then there is a stochastic process $y_2$ on $\Omega_2$, definable from $x_2$, such that $(x_1, y_1) \equiv (x_2, y_2)$.*

## 2.5. Approximate Measure Logic

Motivated by several disparate attempts to incorporate measure and probability into model theory, including the work of [**273, 243**], Goldbring and Towsner proposed a logic that extended first-order logic to accommodate measure as a logical primitive.

DEFINITION 2.5.1 ([**203**]).
AML—seeapproximate measure logic
approximate measure logic—textbfLet $L$ be a first-order signature. Then the AML $L$-formulas are defined by induction:

(1) Every atomic $L$-formula is an AML $L$-formula.
(2) If $\varphi, \psi$ are AML $L$-formulas, then the following are AML $L$-formulas:
  (a) $\neg\varphi$
  (b) $\varphi \wedge \psi$
  (c) $\varphi \vee \psi$
  (d) $\exists x \varphi$
  (e) $\forall x \varphi$
(3) If $\bar{x}$ is a sequence of distinct variables, $q$ is a non-negative rational, and $\varphi$ is an AML $L$-formula, then the following are AML $L$-formulas:
  (a) $m_{\bar{x}} \leq q.\varphi$
  (b) $m_{\bar{x}} < q.\varphi$

It is clear how to interpret the formulas generated by most clauses of this definition, but the final clause requires some comment. First, a variable of $\bar{x}$ that is free in $\varphi$ is bound in $m_{\bar{x}} \leq q.\varphi$ or $m_{\bar{x}} < q.\varphi$, making this clause something like a quantifier. We interpret $m_{\bar{x}} < q.\varphi$ (for instance) as the statement that the set defined by $\varphi$ has measure less than $q$.

Toward formalizing this semantic ideology, we inductively define a rank on the formulas. A classical $L$-formula has rank 0, and for any $\varphi$, we have

$$rk(\varphi) = rk(\neg\varphi) = rk(\forall\varphi) = rk(\exists\varphi).$$

Further, we set

$$rk(\varphi \wedge \psi) = rk(\varphi \vee \psi) = \max(rk(\varphi), rk(\psi)).$$

On the other hand, we let

$$rk\,(m_{\bar{x}} \leq q.\varphi) = rk\,(m_{\bar{x}} < q.\varphi) = rk(\varphi) + 1.$$

Definition 2.5.2. Let $L$ be a first-order signature.

(1) An AML $L$-quasistructure is a first-order $L$-structure $\mathcal{M}$, with the following additional data:
   (a) For each $n \geq 1$, an algebra $\mathcal{B}_n$ of subsets of $M^n$, and a finitely additive measure $\mu_n$ on $\mathcal{B}_n$, with the property that for any $n, m$, the measure space $(M^{n+m}, \mathcal{B}_{n+m}, \mu_{n+m})$ extends the product space $(M^n, \mathcal{B}_n, \mu_n) \times (M^m, \mathcal{B}_m, \mu_m)$.
   (b) For each $n \geq 1$, a function $v_n : \mathcal{B}_n \to \{\oplus, \ominus, \odot\}$ such that $v_n(X) = \odot$ whenever $\mu_n(X) \notin \mathbb{Q}$.

(2) An AML $L$-structure is an AML $L$-quasistructure satisfying, for each $n \geq 1$, the condition that for any formula $\varphi(\bar{x}, \bar{y})$ of rank $n$, and for all $\bar{y}$, we have $\varphi(M, \bar{b}) \in \mathcal{B}_{|\bar{x}|}$.

Again, most clauses are intuitive. The assignments $v_n$ will designate when the structure determines the measure of a set exactly. In particular, we will say that $\mathcal{M} \models m_{\bar{x}} < r.\varphi(\bar{x}, \bar{b})$ if and only if *either* $\mu(\varphi(M), \bar{b}) < r$ *or* $\mu(\varphi(M), \bar{b}) = r$ and $v(\varphi(M), \bar{b}) = \ominus$, and we say that $\mathcal{M} \models m_{\bar{x}} \leq r.\varphi(\bar{x}, \bar{b})$ if and only if *either* $\mu(\varphi(M), \bar{b}) \leq r$ *or* $\mu(\varphi(M), \bar{b}) = r$ and $v(\varphi(M), \bar{b}) = \oplus$.

We can define much of the usual model-theoretic apparatus (e.g. substructures, definable sets, types) in the usual way. Indeed, Goldbring and Towsner were able to prove the following compactness theorem for AML.

Theorem 2.5.3 ([**203**]). *Let $T$ be an AML theory. If every finite subset of $T$ has a model, then $T$ has a model.*

Proof. Let $\mathcal{I}$ be the set of all finite subsets of $T$, and for each $T_i \in \mathcal{I}$, let $\mathcal{M}_i \models T_i$. We will construct an ultrafilter $D$ over $\mathcal{I}$ such that $\prod_D \mathcal{M}_i \models T$.

To this end, for each $\varphi \in T$, we define $\mathcal{J}_\varphi \subseteq \mathcal{I}$ to be the set of all members of $\mathcal{I}$ containing $\varphi$. The set $\{\mathcal{J}_\varphi : \varphi \in T\}$ is closed under finite intersection. Thus, it can be extended to a set $D \notin \{\emptyset, \mathcal{I}\}$ which is closed upwards, closed under finite intersection, and cannot be properly extended to another set having these properties (that is, $D$ is an ultrafilter).

We now construct a new structure in the following way. We let $\mathcal{M}_0 = \prod_{i \in \mathcal{I}} \mathcal{M}_i$, and let $\mathcal{M}$ be the quotient of $\mathcal{M}_0$ by the relation identifying two functions $f, g \in \mathcal{M}_0$ if and only if the set on which they agree is an element of $D$, and define a measure $\mu_n^{\mathcal{M}}$ on $\mathcal{M}$ as the product measure of the measures on the $\mathcal{M}_i$, modulo $D$. We claim that $\mathcal{M} \models T$.

Observe that, by definition, for each $\varphi \in T$, the set $J_\varphi$ is in $D$. We will show that $\mathcal{M} \models \varphi$ if and only if $J_\varphi \in D$, which would complete the proof. This would be quite true in classical first-order logic by the so-called *Łoś Theorem*, a standard result, whose proof proceeds by induction on the formation of formulas. If the result holds for $\varphi(\bar{x},\bar{y})$, we consider $m_{\bar{x}} < q.\varphi(\bar{x},\bar{y})$.

If $\mu^{\mathcal{M}_i}(\varphi(M_i,\bar{y})) < r$ on an element of $D$, then it follows that $\mu^{\mathcal{M}}(\varphi(M_i,\bar{y})) < r$, so that $\mathcal{M} \models m_{\bar{x}} < q.\varphi(\bar{x},\bar{y})$. The other cases are similar. □

It is important to note that this result includes a proof of the Łoś theorem for AML: a sentence is true in the ultraproduct if it is true on an element of the corresponding ultrafilter. Significant results from classical first-order model theory are possible here; the Tarski-Vaught test and the downward Löwenheim-Skolem theorem both hold. In the context of approximate measure logic, a substructure should, in addition to preserving the first-order signature, also preserves the algebras $\mathcal{B}_n$. A substructure is said to be elementary in the usual case, except that satisfaction of all AML formulas must be preserved.

THEOREM 2.5.4 ([**203**]). *Suppose that $\mathcal{M}$ is an AML substructure of $\mathcal{N}$. Then $\mathcal{M}$ is an AML-elementary substructure if and only if the following conditions hold:*

(1) *For all formulas $\varphi(\bar{x},y)$ and all $\bar{a} \in \mathcal{M}$, if $\mathcal{N} \models \varphi(\bar{a},b)$, then there is $b' \in \mathcal{M}$ with $\mathcal{M} \models \varphi(\bar{a},b')$, and*
(2) *For all formulas $\varphi(x,\bar{y})$ and all $\bar{a} \in \mathcal{M}$, measure is preserved in the sense that $\mu(\varphi(N,\bar{a}) \cap M) = \mu(\varphi(N,\bar{a}))$ and $v(\varphi(N,\bar{a}) \cap M) = v(\varphi(N,\bar{a}))$*

PROOF. If $\mathcal{M}$ is an elemtnary substructure of $\mathcal{N}$. Then both conditions follow. The opposite implication follows by the usual induction. □

THEOREM 2.5.5 ([**203**]). *Let $\mathcal{M}$ be an $L$ structure, with $X \subseteq \mathcal{N}$. Then there is an elementary substructure of $\mathcal{M}$ of cardinality at most* $\max(|X|,|L|,\aleph_0)$.

PROOF. We start with $X$ and iteratively close under the functions of $L$, as usual. The set $X_\omega$ obtained in this way respects the cardinality bound. We define $\mathcal{M}_0$ on $X_\omega$ by defining the measurable sets according to intersection with $X_\omega$. By the Tarski-Vaught test (Theorem 2.5.4), this is an elementary substructure. □

Goldbring and Towsner develop a significant theory of types in approximate measure logic. One significant additional observation is the notion of *approximation by definable stable sets*. A set $S$ is said to be approximated by definable stable sets if for every $\epsilon > 0$ there is a definable stable set whose symmetric difference from $S$ is less than $\epsilon$. Here, as in the previous section on adapted spaces, approximation of formulas has its place.

## 2.6. Continuous First-Order Logic and Metric Structures

**2.6.1. Continuous First-Order Logic.** Perhaps the great counterexample in applications of stability theory is the Hilbert space $L^2(\mathbb{C})$. This structure has a well-behaved (and important) independence theory. On the other hand, the norm guarantees the interpretability of an infinite linear ordering, guaranteeing instability. While we could, of course, pursue the matter from the perspective of thorn forking, another approach has proved productive.

The idea of this approach is that for many structures, collectively called *metric structures*, there is a metric (expressed or implied) that could import instability, even when the combinatorics of the structure otherwise act like a stable structure.

A seminal 2008 paper of Ben Yaacov, Berenstein, Henson, and Usvyatsov merged several approaches current at the time, including compact abstract theories.

The approach here is to absorb the metric into the logic, so that the definable sets of stable theories work like those in first-order stable theories, without interference from the metric.

A *metric structure* is a many-sorted structure in which each sort is a complete metric space of finite diameter. Key examples include Hilbert spaces, Banach spaces, probability spaces, and probability spaces with a distinguished automorphism. Those which do not natively have finite diameter may be treated as increasing sequences of balls. Probability spaces lend themselves naturally to this treatment, as we will see shortly, as do several more involved random structures, as we will see in Chapter 6.

The logic for metric structures, called *continuous first-order logic*, an extension of Łukasiewicz propositoinal logic, takes truth values in $[0, 1]$. Indeed, it is sometimes more conventional to shy away from calling these "truth" values at all; many authors would interpret the values in a geometrical sense, and not a credal one — an interesting contrast with the humble Jansenist origins of probability.

To the extent that the values do represent truth, they follow the slightly unusual convention of using 0 as a numerical value for True (or acceptance) and 1 as a numerical value for False (or rejection). The authors of [**61**] chose this convention to emphasize the metric nature of their logic.

Continuous first-order logic is propositional logic, which builds on work of Keisler and Henson (see [**58**] for a more detailed history). The following definitions are from [**60**].

DEFINITION 2.6.1. A *continuous signature* is an object of the form $\mathcal{L} = (\mathcal{R}, \mathcal{F}, \mathcal{G}, n, d)$ where

1. $\mathcal{R}$ (which we will call the set of *predicate symbols*) and $\mathcal{F}$ (which we will call the *function symbols*) are disjoint and $\mathcal{R}$ is nonempty, and
2. $n$ is a function associating to each member of $\mathcal{R} \cup \mathcal{F}$ its arity
3. $\mathcal{G}$ has the form $\{\delta_{s,i} : (0, 1] \to (0, 1] : s \in \mathcal{R} \cup \mathcal{F} \text{ and } i < n_s\}$, and
4. $d$ is a special symbol.

Because of the context, we will use the notation $n_S = n(S)$ for the arity function. We now define the class of structures. The similarity to the AML structures of the previous section will be evident.

DEFINITION 2.6.2. Let $\mathcal{L} = (\mathcal{R}, \mathcal{F}, \mathcal{G}, n, d)$ be a continuous signature. A *continuous $\mathcal{L}$-pre-structure* is an ordered pair $\mathfrak{M} = (M, \rho)$, where $M$ is a non-empty set, and $\rho$ is a function on $\mathcal{R} \cup \mathcal{F} \cup \{d\}$ taking values in sets of maps defined on Cartesian powers of $M$, such that

1. To each function symbol $f$, the function $\rho$ assigns a mapping $f^{\mathfrak{M}} : M^{n_f} \to M$
2. To each predicate symbol $P$, the function $\rho$ assigns a mapping $P^{\mathfrak{M}} : M^{n_P} \to [0, 1]$.
3. The function $\rho$ assigns $d$ to a pseudo-metric $d^{\mathfrak{M}} : M \times M \to [0, 1]$. (Note: a pseudo-metric is a symmetric nonnegative function into the reals satisfying the triangle inequality.)
4. For each $f \in \mathcal{F}$ for each $i < n_f$, the element $\delta_{f,i} \in \mathcal{G}$ is a modulus of continuity for $f$ in the $i$th argument; that is, for each $\epsilon \in (0, 1]$, we have

$$\forall \bar{a}, \bar{b}, c, e \left[ d^{\mathfrak{M}}(c, e) < \delta_{f,i}(\epsilon) \Rightarrow d^{\mathfrak{M}} \left( f^{\mathfrak{M}}(\bar{a}, c, \bar{b}), f^{\mathfrak{M}}(\bar{a}, e, \bar{b}) \right) < \epsilon \right]$$

where $lh(\bar{a}) = i$ and $lh(\bar{a}) + lh(\bar{b}) = n_f - 1$.

(5) For each $P \in \mathcal{R}$ for each $i < n_P$, the element $\delta_{P,i} \in \mathcal{G}$ is a modulus of continuity for $P$ in the $i$th argument; that is, for each $\epsilon \in (0, 1]$, we have

$$\forall \bar{a}, \bar{b}, c, e \left[d^{\mathfrak{M}}(c, e) < \delta_{f,i}(\epsilon) \Rightarrow |P^{\mathfrak{M}}(\bar{a}, c, \bar{b}) - P^{\mathfrak{M}}(\bar{a}, e, \bar{b})| < \epsilon\right]$$

where $lh(\bar{a}) = i$ and $lh(\bar{a}) + lh(\bar{b}) = n_P - 1$.

Definition 2.6.3. A *continuous weak $\mathcal{L}$-structure* is a continuous $\mathcal{L}$-pre-structure such that $\rho$ assigns to $d$ a metric.

Since we are concerned here with countable structures, we will not use the stronger notion of a *continuous $\mathcal{L}$-structure* common in the literature, which requires that $d$ be assigned to a *complete* metric. However, it is possible, given a continuous weak structure (even a pre-structure), to pass to a completion [**60**].

We now proceed to define formulas and their evaluation. The following definition may be found in [**58**], and is, in some ways similar to the adapted functions of Section 2.4.

Definition 2.6.4. Let $\mathcal{L}$ be a continuous signature.

(1) The $\mathcal{L}$-terms are defined inductively as follows:
  (a) Every variable is a term.
  (b) If $t_1, \ldots, t_n$ are terms and $f$ is an $n$-ary function symbol, then $f(t_1, \ldots, t_n)$ is a term.
(2) The atomic $\mathcal{L}$-formulas are the expressions $P(t_1, \ldots, t_n)$, where $P$ is an $n$-ary predicate symbol, and the expressions $d(t_1, t_2)$, where $t_i$ are terms.
(3) The quantifier-free $\mathcal{L}$-formulas are the smallest class containing all atomic $\mathcal{L}$-formulas and satisfying the following closure condition:
  - If $u : [0, 1]^n \to [0, 1]$ is continuous and $\varphi_i$ are formulas, then $u(\varphi_1, \ldots, \varphi_n)$ is a formula.
(4) The $\mathcal{L}$-formulas are the smallest class containing all the atomic $\mathcal{L}$-formulas and closed under both the previous condition and the following one:
  - If $\varphi$ is an $\mathcal{L}$-formula and $x$ is a variable, then $\sup_x \varphi$ and $\inf_x \varphi$ are both $\mathcal{L}$-formulas.

In practice, a much smaller set of formulas is entirely adequate — exactly as we saw for adapted spaces and approximate measure logic. Any $\mathcal{L}$-formula may be approximated to arbitrary accuracy, for instance, by a formula where $u$ is taken from $\dot{-}$, $x \mapsto 1 - x$, and $x \mapsto \frac{1}{2}x$. This result, and many others like it, are documented in [**58**]. Throughout the rest of the present paper, then, we will work only with this reduced set of formulas.

Definition 2.6.5. Let $V$ denote the set of variables, and let $\sigma : V \to M$. Let $\varphi$ be a formula.

(1) The *interpretation under $\sigma$* of a term $t$ (written $t^{\mathfrak{M},\sigma}$) is defined by replacing each variable $x$ in $t$ by $\sigma(x)$.
(2) Let $\varphi$ be a formula. We then define the *value of $\varphi$ in $\mathfrak{M}$ under $\sigma$* (written $\mathfrak{M}(\varphi, \sigma)$) as follows:
  (a) $\mathfrak{M}(P(\bar{t}), \sigma) := P^{\mathfrak{M}}(\overline{t^{\mathfrak{M},\sigma}})$
  (b) $\mathfrak{M}(\alpha \dot{-} \beta, \sigma) := \max(\mathfrak{M}(\alpha, \sigma) - \mathfrak{M}(\beta, \sigma), 0)$
  (c) $\mathfrak{M}(\neg\alpha, \sigma) := 1 - \mathfrak{M}(\alpha, \sigma)$
  (d) $\mathfrak{M}(\frac{1}{2}\alpha, \sigma) := \frac{1}{2}\mathfrak{M}(\alpha, \sigma)$

(e) $\mathfrak{M}(\sup_x \alpha, \sigma) := \sup_{a \in M} \mathfrak{M}(\alpha, \sigma_x^a)$, where $\sigma_x^a$ is equal to $\sigma$ except that $\sigma_x^a(x) = a$.

(3) We write $(\mathfrak{M}, \sigma) \models \varphi$ exactly when $\mathfrak{M}(\varphi, \sigma) = 0$.

Of course, if $\varphi$ has no free variables, then the value of $\mathfrak{M}(\varphi, \sigma)$ is independent of $\sigma$.

A sound and complete proof system has been described for continous first-order logic. Of course, it is only complete up to approximation. The following axiom system was given in [**60**], and soundness and completeness results were proved there. The axioms are stated here for completeness, since they are not well-known.

(A1) $(\varphi \dot{-} \psi) \dot{-} \varphi$
(A2) $((\chi \dot{-} \varphi) \dot{-} (\chi \dot{-} \psi)) \dot{-} (\psi \dot{-} \varphi)$
(A3) $(\varphi \dot{-} (\varphi \dot{-} \psi)) \dot{-} (\psi \dot{-} (\psi \dot{-} \varphi))$
(A4) $(\varphi \dot{-} \psi) \dot{-} (\neg\psi \dot{-} \neg\varphi)$
(A5) $\frac{1}{2}\varphi \dot{-} (\varphi \dot{-} \frac{1}{2}\varphi)$
(A6) $(\varphi \dot{-} \frac{1}{2}\varphi) \dot{-} \frac{1}{2}\varphi$.
(A7) $(\sup_x \psi \dot{-} \sup_x \varphi) \dot{-} \sup_x (\psi \dot{-} \varphi)$
(A8) $\varphi[t/x] \dot{-} \sup_x \varphi$ where no variable in $t$ is bound by a quantifier in $\varphi$.
(A9) $\sup_x \varphi \dot{-} \varphi$, wherever $x$ is not free in $\varphi$.
(A10) $d(x, x)$
(A11) $d(x, y) \dot{-} d(y, x)$
(A12) $(d(x, z) \dot{-} d(x, y)) \dot{-} d(y, z)$
(A13) For each $f \in \mathcal{F}$, each $\epsilon \in (0, 1]$, and each $r, q \in \mathcal{D}$ with $r > \epsilon$ and $q < \delta_{f,i}(\epsilon)$, the axiom $(q \dot{-} d(z, w)) \wedge (d(f(\bar{x}, z, \bar{y}), f(\bar{x}, w, \bar{y})) \dot{-} r)$, where $lh(\bar{x}) + lh(\bar{y}) = n_f - 1$.
(A14) For each $P \in \mathcal{R}$, each $\epsilon \in (0, 1]$, and each $r, q \in \mathcal{D}$ with $r > \epsilon$ and $q < \delta_{P,i}(\epsilon)$, the axiom $(q \dot{-} d(z, w)) \wedge ((P(\bar{x}, z, \bar{y}) \dot{-} P(\bar{x}, w, \bar{y})) \dot{-} r)$, where $lh(\bar{x}) + lh(\bar{y}) = n_P - 1$.

Axioms A1–A4 are those of Łukasiewicz propositional logic, and axioms A5–A6 complete the propositional part of continuous logic (primarily prescribing the behavior of $\frac{1}{2}$. Axioms A7–A9 describe the role of the quantifiers. Axioms A10–A12 guarantee that $d$ is a pseudo-metric, and axioms A13–A14 guarantee uniform continuity of functions and predicates. We write $\Gamma \vdash_Q \varphi$ whenever $\varphi$ is provable from $\Gamma$ in continuous first-order logic, as axiomatized above. Where no confusion is likely, we will write $\Gamma \vdash \varphi$.

THEOREM 2.6.6 ([**60**]). *Let $L$ be a continuous signature, and $\Gamma$ a set of $L$-fumrulas.*

(1) *For every formula $\varphi$, if $\Gamma \vdash \varphi$ then $\Gamma \models \varphi$.*
(2) *If $\Gamma$ is satisfiable, then $\Gamma$ is consistent.*
(3) *If $\Gamma$ is consistent, then $\Gamma$ is satisfiable.*

Every classical first-order structure can be considered as a metric structure. We take the discrete metric on the elements of the structure, and modulus of continuity $D(\epsilon) = \epsilon$. As usual, it is very easy to be continuous in a discrete topology. As a technicality, we consider the natural continuous versions of first-order formulas.

DEFINITION 2.6.7. Let $L$ be a classical first-order signature, and $\varphi(\bar{x})$ an $L$-formula. We define a continuous first-order formula $\tilde{\varphi}$ by induction on the form of $\varphi$.

(1) If $P$ is a relation symbol of $L$ and $\varphi(\bar{x})$ is of the form $P\left(t_1(\bar{x}), t_2(\bar{x}), \ldots, t_n(\bar{x})\right)$, for $L$-terms $t_i$, then $\tilde{\varphi}(x)$ is $\varphi(x)$.
(2) If $\varphi(\bar{x})$ is of the form $t_1(\bar{x}) = t_2(\bar{x})$, then $\tilde{\varphi}(\bar{x})$ is $d\left(t_1(\bar{x}), t_2(\bar{x})\right)$.
(3) If $\varphi(\bar{x})$ is of the form $\neg\psi(\bar{x})$, then $\tilde{\varphi}(\bar{x})$ is $1 \dot{-} \tilde{\psi}(\bar{x})$.
(4) If $\varphi(\bar{x})$ is of the form $(\psi_1 \wedge \psi_2)(\bar{x})$, then $\tilde{\varphi}(\bar{x})$ is $\max\left\{\tilde{\psi}_1(\bar{x}), \tilde{\psi}_2(\bar{x})\right\}$.
(5) If $\varphi(\bar{x})$ is of the form $\exists y \psi(\bar{x}, y)$, then $\tilde{\varphi}(\bar{x})$ is $\inf_y \tilde{\psi}(\bar{x}, y)$.

This transformation preserves truth values in the following sense.

LEMMA 2.6.8. *Let $\mathcal{M}$ be a classical first-order $L$-structure, with $\bar{a} \in \mathcal{M}$, and $\varphi(\bar{x})$ an $L$-formula. Then if $\mathcal{M} \models \varphi(\bar{a})$, then $\tilde{\varphi}(\bar{a})^{\mathcal{M}} = 0$.*

The truth preservation also works at the scale of structures.

PROPOSITION 2.6.9 ([**202**]). *Let $\mathcal{M}, \mathcal{N}$ be a classical first-order structure. Then if $\mathcal{M} \equiv \mathcal{N}$ as classical first-order structures, then $\mathcal{M} \equiv \mathcal{N}$ as discrete metric structures.*

PROOF. We may assume without loss of generality that $\mathcal{N} \preceq \mathcal{M}$ as classical structures. Now we may show by induction on a CFO formula $\varphi$ that there is only a finite set $R \subseteq [0, 1]$ of values it may have on the classical structure $\mathcal{M}$. Moreover, for any $r \in R$, there is a classical formula $\psi_r$such that $\varphi(\bar{a})$ has value $r$ on a classical structure $\mathcal{A}$ if and only if $\mathcal{A} \models \psi(\bar{a})$. From this, the result follows. □

Definability of elements is understood in metric structures in the sense of approximation.

DEFINITION 2.6.10. Let $\mathcal{M}$ be a metric structure, $A \subseteq \mathcal{M}$, and $b \in \mathcal{M}$. We say that $b$ is *definable* over $A$ if and only if there is a sequence $(\bar{a}_n : n \in \mathbb{N})$ and a sequence of CFO formulas $\varphi_n$ such that the sequence $\left(\mathcal{M}\left(\varphi_n(x, \bar{a}_n)\right) : n \in \mathbb{N}\right)$ converges to $d(x, b)$.

**2.6.2. Some Metric Structures.** Now let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$. We treat $\mathcal{H}$ as a many-sorted structure where the sorts are balls of increasing radius, and where the signature includes function symbolsfor the inclusions among these balls, scaling, addition, subtraction, and the inner product.

In this signature, we can formulate continuous first-order formulas which have value zero if and only if the structure satisfies the axioms (other than completeness) of a Hilbert space. For instance, the axiom that

$$\forall x \forall y \ \langle x, y \rangle = \langle y, x \rangle$$

can be written as

$$\sup_x \sup_y \langle x, y \rangle \dot{-} \langle y, x \rangle$$
$$\sup_x \sup_y \langle y, x \rangle \dot{-} \langle x, y \rangle.$$

We interpret distance on this structure in the usual way, by $d(x, y) = \sqrt{\langle x - y, x - y \rangle}$. At this point, a CFO structure in which all of the relevant axioms evaluate to zero will also be a complete metric space, and so will be a Hilbert space. We could also axiomatize infinite dimension by requiring, for each $n$, the condition

$$\inf_{x_1, \ldots, x_n} \max_{1 \leq i, j \leq n} \left| \rangle x_i, x_j \rangle - \delta_{ij} \right|.$$

The CFO theory of infinite-dimensional Hilbert spaces is $\kappa$-categorical for every infinite cardinal $\kappa$.

Returning to the original quesiton, we sketch the stability theory on Hilbert spaces. A (complete) $n$-type $p$ over a set $A$ is a set of CFO formulas in free variables $x_1, \ldots, x_n$, with parameters from $A$, such that there is some structure $\mathcal{M}$ and some $n$-tuple $\bar{b}$ such that $p$ is the set of all formulas whose value in $\mathcal{M}$ on $\bar{b}$ is zero. The set of all such types we denote by $S_n(A)$.

For a theory $T$ and set $A$, let $\mathcal{M}_S$ be a model of $T$ realizing every type in $S_n(A)$ for every $n$. Now the metric $d$ on $\mathcal{M}_S$ induces a natural metric on $S_n(A)$ in the following way. For any $p, q \in S_n(A)$, and for tuples $\bar{b}, \bar{c}$ realizing $p$ and $q$ respectively, we set $d(p, q, \bar{b}, \bar{c})$ to be the maximum of $d(b_i, c_i)$ as $i$ ranges from 1 to $n$. We define $d(p, q) = \inf_{\bar{b}, \bar{c}} d(p, q, \bar{b}, \bar{c})$.

Definition 2.6.11. Let $T$ be a CFO theory. We say that $T$ is $\lambda$-stable if for any $\mathcal{M} \models T$ and any set $A$ with cardinality at most $\lambda$, there is a $d$-dense set $D \subseteq S_1(A)$ with cardinality at most $\lambda$.

From this perspective, the following result shows that the transition to continuous logic has resolved the difficulty with which the present section started.

Proposition 2.6.12 ([**58**]). *The CFO theory of infinite-dimensional Hilbert spaces is $\omega$-stable.*

One can also observe that the notion of stability here does generalize stability for classical first-order structures. As we have seen, any first-order theory can be viewed as a continuous first-order theory, and its models viewed as metric structures, by giving the structures the discrete metric.

Theorem 2.6.13. *If $T$ is a classical theory, then $T$ is stable iff $T$ is stable as a discrete metric theory.*

Up to this point in the exposition, the connection of continuous first-order logic and metric structures to probability may seem distant, or, at most, contrived. The next example, however, will begin to show the applicability. Let $(\Omega, \mathcal{B}, P)$ be a probability space. We say that $S \in \mathcal{B}$ is an *atom* if it has positive measure and no subset has properly smaller nonzero measure. Further assume that $(\Omega, \mathcal{B}, P)$ has no atoms (that is, it is atomless).

Let $\hat{\mathcal{B}}$ be the quotient of $\mathcal{B}$ by the equivalence relation that makes $A_1$ and $A_2$ equivalent if and only if they have symmetric difference of measure zero. We can then consider a metric structure with universe $\hat{\mathcal{B}}$, with metric given by $d(A_1, A_2) = P(A_1 \triangle A_2)$, and with functions $\cap, \cup$, complement, and a predicate for $P$. This structure satisfies the axioms of Boolean algebras, the Kolmogorov axioms for $P$, and the additional axiom

$$\sup_x \sup_y |d(x, y) - P(x \triangle y)| .$$

Moreover, because $\Omega$ is atomless, the structure also satisfies the axiom

$$\sup_x \inf_y |P(x \cap y) - P(x \cap y^c)| .$$

Conversely, any metric structure satisfying all of these axioms arises from an atomless probability space in this way.

Proposition 2.6.14 ([**58**]). *Let $\mathcal{M}_1, \mathcal{M}_2$ be separable models of the axioms of atomless probability algebras as above. Then $\mathcal{M}_1 \cong \mathcal{M}_2$.*

This result is proved by a routine back-and-forth argument. Ben Yaacov [**55**] proved that in a sufficiently saturated, sufficiently homogeneous atomless probability algebra, $tp(\bar{a}/C) = tp(\bar{b}/C)$ if and only if for every function $u : \hat{\mathcal{B}}^n \to \hat{\mathcal{B}}$ mapping $\bar{x}$ to an intersection of literals of $\bar{x}$, we have $P\left(u(\bar{a})|C\right) = P\left(u(\bar{b})|C\right)$.

With this in mind, we can show the following.

Theorem 2.6.15 ([**58**]). *The CFO theory of atomless probability algebras is $\omega$-stable.*

Proof. Let $\mathcal{U}$ be a large saturated model, and let $(X^{\mathcal{M}}, \mathcal{B}^{\mathcal{M}}, P^{\mathcal{M}})$ be the corresponding probability space. Let $C \subseteq \mathcal{M}$ be a countable set, closed under Boolean combinations, and let $\mathcal{C}$ be a countable Boolean algebra of subsets of $X^{\mathcal{M}}$ corresponding to $C$.

We describe a dense set of 1-types over $C$. Let $\mathcal{F}$ be the set of types of form $tp(a/C)$ where $P\left(a|\langle\mathcal{C}\rangle\right)$ is an integrable step function on $(X, \mathcal{C}, P)$. □

An important element in modern probability is the theory of measure-preserving functions on probability spaces. We formalize this in the following way. An automorphism of the probability space $(X, \mathcal{B}, P)$ consists of an invertible measure-preserving transformation $\Phi : C_1 \to C_2$, where $C_i \in \mathcal{B}$ with $P(C_i) = 1$. Such a map induces a map $\hat{\Phi} : \hat{\mathcal{B}} \to \hat{\mathcal{B}}$ preserving both the measure and the algebra structure.

For any atomless probability algebra $\mathcal{A}$ and any automorphism $\tau$ induced by an automorphism of the probability space, the following formulas will have value 0.

(1) $\sup_x |P(x) - P(\tau(x))|$
(2) $\sup_x \inf_y |P(x \triangle y)|$
(3) $\sup_x \sup_y |P\left(\tau(x \cup y) \triangle \left(\tau(x) \cup \tau(y)\right)\right)|$
(4) $\sup_x \sup_y |P\left(\tau(x \cap y) \triangle \left(\tau(x) \cap \tau(y)\right)\right)|$

We add the following condition: an automorphism $\tau$ of probability spaces is said to be *aperiodic* if and only if for every $n \in \mathbb{N}$ and every $y \in \mathcal{B}$, we have $P\left(y \cap \tau^n(y)\right) = 0$. Rokhlin's lemma allows this to be translated to a condition on automorphisms of a probability algebra. This condition, as a CFO formula, is written

$$\inf_y \max\left\{\left|\frac{1}{n} - P(y)\right|, P\left(y \cap \tau^i(y)\right) : i \leq n\right\}$$

for each $n \geq 1$. The theory of atomless probability algebras with an aperiodic automorphism is complete.

Theorem 2.6.16 ([**58**]). *The CFO theory of atomless probability algebras with a generic automorphism is stable, but not $\omega$-stable.*

## 2.7. CFO as a generalization

To a great extent, many of the systems described here can be expressed in continuous first-order logic. We have already seen that the $\omega$-stable theory of atomless probability algebras includes the Kolmogorov axioms as CFO formulas.

Toward formulating Gaifman's measure models, let $(U, \mu)$ be a measure model in signature $L$. We will construct a metric structure in a metric signature $L'$, using a technique that will be developed further in Section 6.2. The structure will have one sort, and $B$. For each first-order $L$-sentence $\varphi$, we will have an constant symbol $[\![\varphi]\!] \in B$. We will also have a unary predicate $P$, and function symbols for the classical Boolean connectives on $B$, and Boolean operations on $B$, and a metric on $B$.

We arrange that for all quantifier-free $\varphi$, we have $P\left([\![\varphi]\!]\right) = \mu(\varphi)$. We will also axiomatize (hopefully the formalism is familiar by now) that every element $x \in B$ has some $\varphi_x$ such that $P(x \triangle [\![\varphi_x]\!]) = 0$ and that the Boolean connectives in $B$ correspond to Boolean connectives in $L$. Further, we require that

$$P(\exists x \varphi(x)) = \sup_{\bar{a} \subseteq U} \left\{ P\left( \bigwedge_{i=1}^{n} \varphi(a_i) \right) \right\}.$$

We define a metric on $B$, as usual, by the $P$-measure of the symmetric difference.

Gaifman does not seem to have defined a notion of morphism of measure models, but it seems reasonable that a homomorphism $f : (U_1, \mu_1) \to (U_2, \mu_2)$ would be a function $f : U_1 \to U_2$ that respects the measure of quantifier-free formulas. Under this definition, the morphisms of measure models would exactly correspond to the morphisms of the metric structures representing them.

It seems difficulty to believe that the reverse interpretability could be true: that there is a uniform transformation from metric structures to measure models such that the morphisms of the metric structures correspond exactly to the morphisms of the measure models. Probabilistic argumentation systems could be encoded as metric structures in much the same way, with the prospects of a reverse encoding even more daunting.

The same technique used to create metric structures to encode measure models also give metric structures for structures in approximate measure logic. This pair is interesting, because a significant fragment of type theory and stability has been worked out in both approximate measure logic and continuous first order logic. It seems unknown at present to what extent these agree: Do the types of the CFO theory correspond to those of the AML theory? Are forking, dividing, and stability conserved?

The challenge to incorporate Bayesian networks into metric structures arises from their hybrid nature — the underlying probability space and the events on it are continuous, while the graph structure is discrete. To address this, let $(\Omega, \mathcal{B}, P)$ be a probability space, and let $G$ be a Bayesian network of $P$. Let $\mathcal{A}$ be the probability algebra (atomless or not) corresponding to $(\Omega, \mathcal{B}, P)$. We add another sort $V$ of vertices with a discrete metric, along with a function $\nu : \hat{\mathcal{B}} \to V$, which we will prescribe to be a bijection, and a symmetric irreflexive binary relation on $V$. In this signature, we can, excepting probability zero events, write CFO specifications for $P\left(\nu^{-1}(x) | \nu{-}1(y_1, \ldots, y_n)\right)$. In this way, all necessary data about a Bayesian network is encoded in a metric structure.

In the other direction, we recall the use of Bayesian networks to compute inference in probabilistic argumentation systems, so that Bayesian networks can encode at least any metric structure arising as a representation of a probabilistic argumentation system.

Despite the obvious similarities, continuous first-order logic does not appear to generalize the logic of adapted spaces in an immediately meaningful way. Certainly there are bounded continuous functions which are not uniformly continuous. Likely some interesting cases on bounded adapted spaces can be formalized in continous first-order logic, but this does not seem to have been explored in the literature.

# 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. 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.
16. J.-P. Allouche and J. Shallit, *Automatic sequences*, Cambridge, 2003.
17. 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.
18. N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley, 2008.
19. 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*, preprint, 2020.
20. J. J. Andrews and M. L. Curtis, *Free groups and handlebodies*, Proceedings of the American Mathematical Society **16** (1965), 192–195.
21. U. Andrews, I. Goldbring, and H. J. Keisler, *Definable closure in randomizations*, Annals of Pure and Applied Logic **166** (2015), 325–341.
22. ———, *Independence in randomizations*, Journal of Mathematical Logic **19** (2019), 1950005.
23. U. Andrews and H. J. Keisler, *Separable models of randomizations*, Journal of Symbolic Logic **80** (2015), 1149–1181.
24. 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.
25. 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.

26. A. Arana, *Logical and semantic purity*, Protosociology **25** (2008), 36–48.
27. A. Arnould and P. Nicole, *The Port Royal Logic*, Gordon, 1861.
28. S. Arora and B. Barak, *Computational complexity*, Cambridge, 2009.
29. Е. А. Асарин and А. В. Покровский, Применение колмогоровской сложности к анализу динамики упровляемых систем, Автоматика и Телемеханика **1** (1986), 25–33.
30. 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.
31. ———, *Vapnik-Chervonenkis density in some theories without the independence property, I*, Transactions of the American Mathematical Society **368** (2016), 5889–5949.
32. 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.
33. 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.
34. J. Avigad, *Inverting the Furstenberg correspondence*, Discrete and Continuous Dynamical Systems **32** (2012), 3421–3431.
35. J. Avigad, J. Hözl, and L. Serafin, *A formally verified proof of the central limit theorem*, Journal of Automated Reasoning **59** (2017), 389–423.
36. J. Ax, *The elementary theory of finite fields*, Annals of Mathematics **85** (1968), 239–271.
37. L. Babai, *Trading group theory for randomness*, STOC '85: Proceedings of the seventeenth annual ACM symposium on Theory of Computing, 1985, pp. 421–429.
38. A. Baker, *Transcendental number theory*, Cambridge Mathematical Library, Cambridge, 1990.
39. J. T. Baldwin and S. Shelah, *Randomness and semigenericity*, Transactions of the American Mathematical Society **349** (1997), 1359–1376.
40. S. Banach, *Sur le problème de la measure*, Fundamenta Mathematicae **4** (1923), 7–33.
41. A.-L. Barabási and R. Albert, *Emergence of scaling in random networks*, Science **286** (1999), 509–512.
42. G. Barmpalias, D. Cenzer, and C. P. Porter, *The probability of a computable output from a random oracle*, preprint, 2016.
43. ———, *Random numbers as probabilities of machine behavior*, Theoretical Computer Science **673** (2017), 1–18.
44. George Barmpalias and Andrew Lewis-Pye, *Differences of halting probabilities*, preprint, 2016.
45. L. Barreira, *Dimension and recurrence in hyperbolic dynamics*, Progress in Mathematics, no. 272, Birkhäuser, 2008.
46. L. Bartholdi, *Counting paths in groups*, L'Enseignement Mathématique **45** (1999), 83–131.
47. N. A. Bazhenov and B. S. Kalmurzaev, *On dark computably enumerable equivalence relations*, Siberian Mathematical Journal **59** (2018), 22–30.
48. V. Becher, Y. Bugeaud, and T. A. Slaman, *On simply normal numbers to different bases*, Mathematische Annalen **364** (2016), 125–150.
49. V. Becher, O. Carton, and P. A. Heiber, *Normality and automata*, Journal of Computer and System Sciences **81** (2015), 1592–1613.
50. V. Becher and P. A. Heiber, *Normal numbers and finite automata*, Theoretical Computer Science **477** (2013), 109–116.
51. J. Beck, *An algorithmic approach to the Lovász local lemma I*, Random Structures and Algorithms **2** (1991), 343–365.
52. H. Becker and A. S. Kechris, *Borel actions of Polish groups*, Bulletin of the American Mathematical Society **28** (1993), 334–341.
53. O. Becker, A. Lubotzky, and A. Thom, *Stability and invariant random subgroups*, Duke Mathematical Journal **168** (2019), 2207–2234.
54. S. Ben-David, D. Pál, and S. Shalev-Shwartz, *Agnostic online learning*, Conference on Learning Theory (COLT), 2009.
55. I. Ben Yaacov, *Schrödinger's cat*, Israel Journal of Mathematics **153** (2006), 157–191.
56. ———, *Continuous and random Vapnik-Chervonenkis classes*, Israel Journal of Mathematics **173** (2009), 309–333.

57. ———, *On theories of random variables*, Israel Journal of Mathematics **194** (2013), 957–1012.
58. 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.
59. I. Ben Yaacov and H. J. Keisler, *Randomizations of models as metric structures*, Confluentes Mathematici **1** (2009), 197–223.
60. I. Ben Yaacov and A. P. Pedersen, *A proof of completeness for continuous first-order logic*, Journal of Symbolic Logic (2010), 168–190.
61. I. Ben Yaacov and A. Usvyatsov, *Continuous first order logic and local stability*, Transactions of the American Mathematical Society **362** (2010), 5213–5259.
62. C. Bernardi and A. Sorbi, *Classifying positive equivalence relations*, The Journal of Symbolic Logic **48** (1983), 529–538.
63. A. Beros, *Learning theory in the arithmetic hierarchy*, Journal of Symbolic Logic **79** (2014), 908–927.
64. Ö. Beyarslan, *Random hypergraphs in pseudofinite fields*, Journal of the Institute of Mathematics of Jussieu **9** (2010), 29–47.
65. L. Bienvenu, *Game-theoretic approaches to randomness: unpredictability and stochasticity*, Ph.D. thesis, Université de Provence, 2008.
66. 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.
67. C. M. Bishop, *Pattern recognition and machine learning*, Information Science and Statistics, Springer, 2006.
68. L. Blum and M. Blum, *Toward a mathematical theory of inductive inference*, Information and control **28** (1975), 125–155.
69. A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, *Learnability and the Vapnik-Chervonenkis dimension*, Journal of the ACM **36** (1989), 929–965.
70. B. Bollobás, *Random graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics, no. 73, Cambridge, 2001.
71. G. Boole, *An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities*, Macmillan, 1851.
72. W. W. Boone, *Certain simple, unsolvable problems of group theory V, VI*, Indagationes Mathematicae **60** (1957), 22–27, 227–232.
73. ———, *The word problem*, Proceedings of the National Academy of Sciences of the USA **44** (1958), 1061–1065.
74. M. Borda, *Fundamentals in information theory and coding*, Springer, 2011.
75. A. Borel, *Density properties for certain subgroups of semi-simple groups wihtout compact components*, Annals of Mathematics **72** (1960), 179–188.
76. M. E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rendiconti del Circolo Matematico di Palermo (1884–1940) **27** (1909), 247–271.
77. 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.
78. 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.
79. K. H. Borgwardt, *The simplex method: A probabilistic analysis*, Algorithms and Combinatorics, no. 1, Springer, 1987.
80. L. Bowen, *Invariant random subgroups of the free group*, Groups, Geometry, and Dynamics **9** (2015), 891–916.
81. G. Boxall, *NIP for some pair-like theories*, Archive for Mathematical Logic **50** (2011), 353–359.
82. M. Braverman, *Parabolic Julia sets are polynomial time computable*, Nonlinearity **19** (2006), 1383–1401.
83. M. Braverman and M. Yampolsky, *Computability of julia sets*, Algorithms and Computation in Mathematics, no. 23, Springer, 2009.

84. J. Brody and M. C. Laskowski, *Rational limits of Shelah-Spencer graphs*, The Journal of Symbolic Logic **77** (2012), 580–592.
85. Y. Bugeaud, *Distribution modulo one and Diophantine approximation*, Cambridge Tracts in Mathematics, no. 193, Cambridge, 2012.
86. S. Buss and M. Minnes, *Probabilistic algorithmic randomness*, The Journal of Symbolic Logic **78** (2013), 579–601.
87. D. Cai, N. Ackerman, and C. Freer, *An iterative step-function estimator for graphons*, preprint, 2015.
88. W. Calvert, *Metric structures and probabilistic computation*, Theoretical Computer Science **412** (2011), 2766–2775.
89. ———, *PAC learning, VC dimension, and the arithmetic hierarchy*, Archive for Mathematical Logic **54** (2015), 871–883.
90. 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.
91. ———, *Random fields*, preprint, 2020.
92. W. Calvert and J. F. Knight, *Classification from a computable viewpoint*, Bulletin of Symbolic Logic **12** (2006), 191–219.
93. P. J. Cameron, *Transitivity of permutation groups on unordered sets*, Mathematische Zeitschrift **148** (1976), 127–139.
94. 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.
95. J. Case, S. Jain, and F. Stephan, *Effectivity questions for Kleene's recursion theorem*, Theoretical Computer Science **733** (2018), 55–70.
96. J. Case and C. Lynes, *Machine inductive inference and language identification*, International Colloquium on Automata, Languages, and Programming, 1982, pp. 107–115.
97. T. Ceccherini-Silberstein and M. Coornaert, *Cellular automata and groups*, Springer Monographs in Mathematics, Springer, 2010.
98. D. Cenzer, $\Pi^0_1$ *classes in computability theory*, Handbook of Computability, Studies in Logic and the Foundations of Mathematics, no. 140, Elsevier, 1999, pp. 37–85.
99. C. Chabauty, *Limite d'ensembles et géométrie des nombres*, Bulletin de la S. M. F. **78** (1950), 143–151.
100. G. J. Chaitin, *A theory of program size formally identical to information theory*, Journal of the Association for Computing Machinery **22** (1975), 329–340.
101. ———, *Algorithmic information theory*, Cambridge Tracts in Theoretical Computer Science, no. 1, Cambridge, 1987.
102. A. S. Charles, *Interpreting deep learning: The machine learning Rorschach test?*, preprint, 2018.
103. H. Chase and J. Freitag, *Model theory and machine learning*, preprint, 2018.
104. Z. Chatzidakis, *Théorie des modèles des corps valués*, lecture notes, 2008.
105. 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.
106. G. Cherlin and E. Hrushovski, *Finite structures with few types*, Annals of Mathematics Studies, no. 152, Princeton University Press, 2003.
107. A. Chernikov and S. Starchenko, *Definable regularity lemmas for NIP hypergraphs*, Quarterly Journal of Mathematics **72** (2021), 1401–1433.
108. N. Chomsky, *Three models for the description of language*, IRE Transactions on Information Theory **2** (1956), 113–124.
109. ———, *Knowledge of language: Its nature, origin, and use*, Convergence, Praeger Scientific, 1986.
110. F. Chung, *On concentrators, superconcentrators, generalizers, and nonblocking networks*, The Bell System Technical Journal **58** (1978), 1765–1777.
111. F. Chung and L. Lu, *Complex graphs and networks*, CBMS Regional Conference Series in Mathematics, no. 107, American Mathematical Society, 2006.
112. F. Chung, L. Lu, T. G. Dewey, and D. J. Galas, *Duplication models for biological networks*, Journal of Computational Biology **10** (2003), 677–687.

113. 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.
114. K. J. Compton, *Laws in logic and combinatorics*, Algorithms and Order, NATO ASI Series C, no. 255, Kluwer, 1989, pp. 353–383.
115. G. Conant and A. Pillay, *Pseudofinite groups and VC-dimension*, preprint, 2018.
116. A. Condon, *The complexity of stochastic games*, Information and Computation **96** (1992), 203–224.
117. C. T. Conley, A. S. Kechris, and B. D. Miller, *Stationary probability measures and topological realizations*, Israel Journal of Mathematics **198** (2013), 333–345.
118. C. T. Conley and B. D. Miller, *Measure reducibility of countable Borel equivalence relations*, Annals of Mathematics **185** (2017), 347–402.
119. D. Conlon and J. Fox, *Bounds for graph regularity and removal lemmas*, Geometric and Functional Analysis **22** (2012), 1191–1256.
120. A. Connes and B. Weiss, *Property T and asymptotically invariant sequences*, Israel Journal of Mathematics **37** (1980), 209–210.
121. S. D. Conte and C. de Boor, *Elementary numerical analysis*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill, 1980.
122. 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.
123. M. Coornaert, *Topoligcal dimension and dynamical systems*, Universitext, Springer, 2015.
124. T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, 3rd ed., MIT Press, 2009.
125. R. T. Cox, *Probability, frequency, and reasonable expectation*, American Journal of Physics **14** (1946), 1–13.
126. B. F. Csima, V. S. Harizanov, R. Miller, and A. Montalban, *Computability of Fraïssé limits*, Journal of Symbolic Logic **76** (2011), 66–93.
127. R. W. R. Darling and J. R. Norris, *Structure of large random hypergraphs*, The Annals of Applied Probability **15** (2005), 125–152.
128. S. Ben David, P. Hrubes, S. Moran, A. Shpilka, and A. Yehudayoff, *Learnability can be undecidable*, Nature Machine Intellignece **1** (2019), 44–48.
129. A. P. Dawid, *Probability, causality, and the empirical world: a Bayes-de Finetti-Popper-Borel synthesis*, Statistical Science **10** (2004), 44–57.
130. B. de Finetti, *Funzione caratteristica di un fenomeno aleatorio*, Memorie della R. Accademia dei Lincei **4** (1930), 86–133.
131. ———, *Foresight: its logical laws, its subjective sources*, Breakthroughs in Statisitcs, Springer, 1992, Translated by H. E. Kyberg, Jr.; Original published 1937, pp. 134–174.
132. ———, *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.
133. 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.
134. M. Dehn, *Über unendliche diskontinuierliche Gruppen*, Mathematische Annalen **71** (1911), 116–144.
135. ———, *Transformation der kurven auf zweiseitigen flächen*, Mathematische Annalen **72** (1912), 413–421.
136. A. P. Dempster, *Upper and lower probabilities induced by a multivalued mapping*, The Annals of Mathematical Statistics **38** (1967), 325–339.
137. O. Demuth, *On constructive pseudonumbers*, Commentationes Mathematicae Universitatis Carolinae **16** (1975), 315–331.
138. M. Detlefsen and A. Arana, *Purity of methods*, Philosophers' Imprint **11** (2011), 1–20.
139. P. Diaconis and S. Janson, *Graph limits and exchangeable random graphs*, Rendiconti di Matematica, Serie VII **28** (2008), 33–61.
140. A. Ditzen, *Definable equivalence relations on Polish spaces*, Ph.D. thesis, California Institute of Technology, 1992.
141. A. Dolich, D. Lippel, and J. Goodrick, *dp-minimal theories: basic facts and examples*, Notre Dame Journal of Formal Logic **52** (2011), 267–288.

142. M. D. Donsker, *An invariance principle for certain probability limit theorems*, Memoirs of the American Mathematical Society **6** (1951), 1–12.
143. 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.
144. R. G. Downey and E. J. Griffiths, *Schnorr randomness*, The Journal of Symbolic Logic **69** (2004), 533–554.
145. R. G. Downey and D. R. Hirschfeldt, *Algorithmic randomness and complexity*, Theory and Applications of Computability, Springer, 2010.
146. R. G. Downey, C. G. Jockusch Jr., and P. E. Schupp, *Asymptotic density and computably enumerable sets*, Journal of Mathematical Logic **13** (2013), 1350005.
147. A. Dudko and M. Yampolsky, *On computational complexity of Cremer Julia sets*, Fundamenta Mathematicae **252** (2021), 343–353.
148. R. M. Dudley, *Central limit theorems for empirical measures*, The annals of probability **6** (1978), 899–929.
149. 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.
150. M. Džamonja and I. Tomašić, *Graphons arising from graphs definable over finite fields*, preprint, 2017.
151. P. D. Eastman, *Are you my mother?*, Random House, 1960.
152. G. Edgar, *Measure, topology, and fractal geometry*, second ed., Undergraduate Texts in Mathematics, Springer, 2008.
153. 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.
154. K. Eickmeyer and M. Grohe, *Randomisation and derandomisation in descriptive complexity theory*, Logical Methods in Computer Science **7** (2011), 1–24.
155. G. Elek and B. Szegedy, *A measure-theoretica approach to the theory of dense hypergraphs*, Advances in Mathematics **231** (2012), 1731–1772.
156. R. Elwes, *Asymptotic classes of finite structures*, Journal of Symbolic Logic **72** (2007), 418–438.
157. H. B. Enderton, *A mathematical introduction to logic*, Academic Press, 1972.
158. I. Epstein, *Orbit inequivalent actions of non-amenable groups*, preprint, 2008.
159. 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.
160. 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.
161. P. Erdős and A. Rényi, *On the evolution of random graphs*, Matematikai Kutató Intézet Közelményei **A** (1960), 17–60.
162. P. Erdős and A. Rényi, *On random graphs I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297.
163. Ю. Л. Ершов, Позитивные Еквивалентности, Алгебра и Логика **10** (1971), 620–650.
164. R. Fagin, *Generalized first-order spectra and polynomial-time recognizable sets*, SIAM–AMS Proceedings, vol. 7, 1974.
165. ——, *Probabilities on finite models*, The Journal of Symbolic Logic **41** (1976), 50–58.
166. S. Fajardo and H. J. Keisler, *Model theory of stochastic processes*, Lecture Notes in Logic, no. 14, A K Peters, 2002.
167. K. Falconer, *Fractal geometry: Mathematical foundations and applications*, 2nd ed., Wiley, 2003.
168. U. Felgner, *Pseudo-endliche gruppen*, Jahrbuch der Kurt-Gödel-Gesellschaft **3** (1990), 94–108.
169. E. Fokina, V. Harizanov, and D. Turetsky, *Computability-theoretic categoricity and Scott families*, preprint, 2019.
170. G. B. Folland, *Real analysis*, 2nd ed., Pure and Applied Mathematics, Wiley, 1999.
171. M. Foreman, D. J. Rudolph, and B. Weiss, *The conjugacy problem in ergodic theory*, Annals of Mathematics **173** (2011), 1529–1586.

172. M. Foreman and B. Weiss, *An anti-classification theorem for ergodic measure presrving transormations*, Journal of the European Mathematical Society **6** (2004), 277–292.
173. W. Fouché, *Arithmetical representations of Brownian motion I*, The Journal of Symbolic Logic **65** (2000), 421–442.
174. ———, *Martin-Löf randomness, invariant measures and countable homogeneous structures*, Theory of Computing Systems **52** (2013), 65–79.
175. 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.
176. R. Fraïssé, *Sur l'extension aux relations de quelques propriétés des ordres*, Annales scientifiques de l'É.N.S. **71** (1954), 363–388.
177. 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.
178. J. N. Y. Franklin and C. P. Porter (eds.), *Algorithmic randomness*, Lecture Notes in Logic, no. 50, Cambridge University Press, 2020.
179. C. Freer, *Computable de Finetti measures*, Annals of Pure and Applied Logic **163** (2012), 530–546.
180. M. D. Fried and M. Jarden, *Field arithmetic*, 2nd ed., Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 11, Springer, 2005.
181. H. Friedman and L. Stanley, *A Borel reducibility theory for classes of countable structures*, Journal of Symbolic Logic **54** (1989), 894–914.
182. A. Frieze and R. Kannan, *Quick approximation to matrices and applications*, Combinatorica **19** (1999), 175–220.
183. A. Furman, *What is a stationary measure?*, Notices of the American Mathematical Soceity **58** (2011), 1276–1277.
184. H. Furstenberg, *A Poisson formula for semi-simple Lie groups*, Annals of Mathematics **77** (1963), 335–386.
185. ———, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Mathematical Systems Theory **1** (1967), 1–49.
186. ———, *A note on Borel's density theorem*, Proceedings of the American Mathematical Society **55** (1976), 209–212.
187. ———, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, Journal D'Analyse Mathématique **31** (1977), 204–256.
188. S. Gaal and L. Gál, *The discrepancy of the sequence* $\{(2^n x)\}$, Indagationes Mathematicae **26** (1964), 129–143.
189. H. Gaifman, *Concerning measures in first order calculi*, Israel Journal of Mathematics **2** (1964), 1–18.
190. S. Gao, *Invariant descriptive set theory*, Pure and Applied Mathematics, CRC Press, 2009.
191. S. Gao and P. Gerdes, *Computably enumerable equivalence relations*, Studia Logica **67** (2001), 27–59.
192. W. I. Gasarch, *The P=?NP poll*, ACM SIGACT News **33** (2002), 34–47.
193. H. Geffner and J. Pearl, *A framework for reasoning with defaults*, Tech. Report R-94, Cognitive Systems Laboratory, UCLA, 1987.
194. D. Geiger and J. Pearl, *Logical and algorithmic properties of conditional independence and graphical models*, Annals of Statistics **21** (1993), 2001–2021.
195. T. Gelander, *Lecture notes on invariant random subgroups and lattices in rank one and higher rank*, preprint, 2015.
196. ———, *Kazhdan-Margulis theorem for invariant random subgroups*, Advances in Mathematics **327** (2018), 47–51.
197. 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.
198. J. Gill, *Computational complexity of probabilistic Turing machines*, SIAM Journal on Computing **6** (1977), 675–695.
199. E. Glasner and B. Weiss, *Minimal actions of the group* $\mathbb{S}(\mathbb{Z})$ *of permutations of the integers*, Geometric and Functional Analysis **12** (2002), 964–988.
200. 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.

201. E. Mark Gold, *Language identification in the limit*, Information and Control **10** (1967), 447–474.
202. I. Goldbring and V. C. Lopes, *Pseudofinite and pseudocompact metric structures*, Notre Dame Journal of Formal Logic **56** (2015), 493–510.
203. I. Goldbring and H. Towsner, *An approximate logic for measures*, Israel Journal of Mathematics **199** (2014), 867–913.
204. O. Goldreich, *A primer on pseudorandom generators*, University Lecture Series, no. 55, American Mathematical Society, 2010.
205. 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.
206. 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.
207. D. Gorenstein, *Classifying the finite simple groups*, Bulletin of the American Mathematical Society **14** (1986), 1–98.
208. 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.
209. M. Gromov, *Random walk in random groups*, Geometric and Functional Analysis **13** (2003), 73–146.
210. Y. Guivarc'h, *Sur la loi des grands nombres et le rayon spectral d'une marche aléatoire*, Astérisque **74** (1980), 47–98.
211. A. Günaydin and P. Hieronymi, *Dependent pairs*, Journal of Symbolic Logic **76** (2011), 377–390.
212. 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.
213. I. Hacking, *The emergence of probability*, 2nd ed., Cambridge, 2006.
214. 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.
215. R. Haenni, J.-W. Romeijn, G. Wheeler, and J. Williamson, *Probabilistic logics and probabilistic networks*, Synthese Library, vol. 350, Springer, 2011.
216. T. Hailperin, *Sentential probability logic*, Lehigh University Press, 1996.
217. J. Y. Halpern, *A counterexample to theorems of Cox and Fine*, Journal of Artificial Intelligence Research **10** (1999), 67–85.
218. ———, *Cox's theorem revisited*, Journal of Artifical Intelligence Research **11** (1999), 429–435.
219. ———, *Reasoning about uncertainty*, MIT Press, 2003.
220. 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.
221. V. S. Harizanov and F. Stephan, *On the learnability of vector spaces*, Journal of Computer and System Sciences **73** (2007), 109–122.
222. 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.
223. M. Harrison-Trainor, B. Khoussainov, and D. Turetsky, *Effective aspects of algorithmically random structures*, Computability **8** (2019), 359–375.
224. H. Hatami and S. Norine, *The entropy of random-free graphons and properties*, Combinatorics, Probability, and Computing **22** (2013), 517–526.
225. F. Hausdorff, *Dimension und äußeres Maß*, Mathematische Annalen **79** (1918), 157–179.
226. C. W. Henson, *A family of countable homogeneous graphs*, Pacific Journal of Mathematics **38** (1971), 69–83.
227. P. Hieronymi and T. Nell, *Distal and non-distal pairs*, Journal of Symbolic Logic **82** (2017), 375–383.
228. 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.

229. J. Hintikka and G. Sandu, *Game-theoretical semantics*, Handbook of Logic and Language, Amsterdam, 1997, pp. 361–410.
230. 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.
231. 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.
232. J. M. Hitchcock, *Correspondence principles for effective dimensions*, Theory of Computing Systems **38** (2005), 559–571.
233. J. M. Hitchcock and J. H. Lutz, *Why computational complexity requires stricter martingales*, Theory of Computing Systems **39** (2006), 277–296.
234. G. Hjorth, *Classification and orbit equivalence relations*, Mathematical Surveys and Monographs, vol. 75, American Mathematical Society, 2000.
235. ______, *On invariants for measure preserving transformations*, Fundamenta Mathematicae **169** (2001), 51–84.
236. ______, *A converse to Dye's Theorem*, Transactions of the American Mathematical Society **357** (2005), 3083–3103.
237. ______, *Glimm-Effros for coanalytic equivalence relations*, Journal of Symbolic Logic **74** (2009), 402–422.
238. G. Hjorth and A. S. Kechris, *Analytic equivalence relations and Ulm-type classifications*, Journal of Symbolic Logic **60** (1995), 1273–1300.
239. W. Hodges, *What is a structure theory?*, Bulletin of the London Mathematical Society **19** (1987), 209–237.
240. ______, *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.
241. B. Host and B. Kra, *Nilpotent structures in ergodic theory*, Mathematical Surveys and Monographs, vol. 236, American Mathematical Society, 2018.
242. E. Hrushovski, *Pseudo-finite fields and related structures*, Model Theory and Applications, Quaderni di Matematica, no. 11, Aracne, 2002, pp. 151–212.
243. ______, *Stable group theory and approximate subgroups*, Journal of the American Mathematical Society **25** (2012), 189–243.
244. E. Hrushovski, Y. Peterzil, and A. Pillay, *Groups, measures, and the NIP*, Journal of the American Mathematical Society **21** (2008), 563–596.
245. E. Hrushovski and A. Pillay, *Groups definable in local fields and pseudo-finite fields*, Israel Journal of Mathematics **85** (1994), 203–262.
246. ______, *Definable subgroups of algebraic groups over finite fields*, Journal für die reine und angewandte Mathematik **462** (1995), 69–91.
247. T. W. Hungerford, *Algebra*, Graduate Texts in Mathematics, no. 73, Springer, 1974.
248. N. Immerman, *Descriptive complexity*, Graduate Texts in Computer Science, Springer, 1999.
249. 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.
250. S. Janson, *Graphons, cut norm, and distance, couplings and rearrangements*, NYJM Monographs, no. 4, New York Journal of Mathematics, 2013.
251. R. C. Jeffrey, *The logic of decision*, 2nd ed., University of Chicago Press, 1983.
252. 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.
253. 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.
254. ______, *The subadditive ergodic theorem and generic stretching factors for free group automorphisms*, Israel Journal of Mathematics **157** (2007), 1–46.
255. O. Kallenberg, *Foundations of modern probability*, Probability and its Applications, Springer, 1997.
256. ______, *Probabilistic symmetries and invariance principles*, Probability and its Applications, Springer, 2005.
257. 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.

258. I. Kaplansky, *Infinite Abelian groups*, revised ed., University of Michigan Press, 1969.
259. 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.
260. M. Karpinski and A. Macintyre, *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.
261. S. Kauffman, *Homeostasis and differentiation in random genetic control networks*, Nature **224** (1969), 177–178.
262. ______, *Metabolic stability and epigenesis in randomly constructed genetic nets*, Journal of Theoretical Biology **22** (1969), 437–467.
263. ______, *The large scale structure and dynamics of gene control circuits: an ensemble approach*, Journal of Theoretical Biology **44** (1974), 167–190.
264. ______, *The origins of order*, Oxford University Press, 1993.
265. S. M. Kautz, *Degrees of random sets*, Ph.D. thesis, Cornell University, 1991.
266. D. Kazhdan and G. Margulis, *A proof of Selberg's hypothesis*, Matematicheskii Sbornik **75** (1968), 163–168.
267. M. J. Kearns and U. V. Vazirani, *An introduction to computational learning theory*, MIT Press, 1994.
268. A. S. Kechris and B. D. Miller, *Topics in orbit equivalence*, Lecture Notes in Mathematics, no. 1852, Springer, 2004.
269. 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.
270. ______, *Fraïssé limits, Ramsey theory, and topological dynamics of automorphism groups*, Geometric and Functional Analysis **15** (2005), 106–189.
271. 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.
272. H. J. Keisler, *Randomizing a model*, Advances in Mathematics **143** (1999), 124–158.
273. ______, *Probability quantifiers*, Model-Theoretic Logics, Perspectives in Logic, Cambridge University Press, 2016, Originally published 1985, pp. 509–556.
274. J. M. Keynes, *A treatise on probability*, MacMillan, 1921.
275. A. I. Khinchine, *Mathematical foundations of information theory*, Dover, 1957.
276. 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.
277. ______, *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.
278. H. Ki and T. Linton, *Normal numbers and subsets of $\mathbb{N}$ with given densities*, Fundamenta Mathematicae **144** (1994), 163–179.
279. 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.
280. 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.
281. J. F. C. Kingman, *The ergodic theory of subadditive stochastic processes*, Journal of the Royal Statistical Soceity, Series B (Methodological **30** (1968), 499–510.
282. V. Klee and G. J. Minty, *How good is the simplex algorithm?*, Inequalities, III, Academic Press, 1972, pp. 159–175.
283. J. F. Knight, A. Pillay, and C. Steinhorn, *Definable sets in ordered structures II*, Transactions of the American Mathematical Society **295** (1986), 593–605.
284. D. E. Knuth, *The art of computer programming: Seminumerical algorithms*, 3rd ed., vol. 2, Pearson, 1998.
285. 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.
286. ______, *Infinitary logics and 0–1 laws*, Information and Computation **98** (1992), 258–294.
287. A. N. Kolmogorov, *Foundations of the theory of probability*, Chelsea, 1950.

288. А. Н. Колмогоров, Три подхода к определению понятия ⟨⟨ количество информатсии ⟩⟩, Проблемы передачи информации **1** (1965), 3–11.
289. B. Kra, *Commentary on 'Ergodic theory of amenable group actions': old and new*, Bulletin of the American Mathematical Society **55** (2018), 343–345.
290. 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.
291. 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.
292. A. Kučera and T. Slaman, *Randomnuess and recursive enumerability*, SIAM Journal on Computing **31** (2001), 199–211.
293. 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.
294. H. Lädesmäki, I. Shmulevich, and O. Yli-Harja, *On learning gene regulatory networks under the Boolean network model*, Machine Learning **52** (2003), 147–167.
295. J. C. Lagarias, *The ultimate challenge: The* $3x+1$ *problem*, American Mathematical Society, 2010.
296. S. Lang and A. Weil, *Number of points of varieties in finite fields*, American Journal of Mthematics **76** (1954), 819–827.
297. M. C. Laskowski, *A simpler axiomatization of the Shelah-Spencer almost sure theories*, Israel Journal of Mathematics **161** (2007), 157–186.
298. 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.
299. Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, Nature **521** (2015), 436–444.
300. M. Ledoux, *The concentration of measure*, Mathematical Surveys and Monographs, no. 89, American Mathematical Society, 2001.
301. Л. А. Левин, Законы сохранения (невозрастания) информации и вопросы обоснования теории вероятностей, Проблемы передачи информации **10** (1974), 30–35.
302. M. B. Levin, *On the discrepancy estimate of normal numbers*, Acta Arithmetica **88** (1999), 99–111.
303. M. Li, J. Tromp, and P. Vitányi, *Sharpening Occam's razor*, Information Processing Letters **85** (2003), 267–274.
304. Ming Li and Paul Vitányi, *An introduction to Kolmogorov complexity and its applications*, third ed., Texts in Computer Science, Springer, 2008.
305. N. Littlestone, *Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm*, Machine Learning **2** (1988), 285–318.
306. L. Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, 2012.
307. L. Lovász and B. Szegedy, *Limits of dense graph sequences*, Journal of Combinatorial Theory, Series B **96** (2006), 933–957.
308. 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.
309. 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.
310. T. Luczak and J. Spencer, *When does the zero-one law hold?*, Journal of the American Mathematical Society **4** (1991), 451–468.
311. 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.
312. J. H. Lutz, *Dimension in complexity classes*, SIAM Journal on Computing **32** (2003), 1236–1259.
313. R. Lyons and Y. Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, no. 42, Cambridge, 2016.
314. D. Macpherson and C. Steinhorn, *One-dimensional asymptotic classes of finite structures*, Transactions of the American Mathematical Society **380** (2008), 411–448.
315. ———, *Definability in classes of finite structures*, Finite and algorithmic model theory, London Mathematical Society Lecture Note Series, no. 379, Cambridge, 2011, pp. 140–176.

316. 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.
317. M. Malliaris and A. Pillay, *The stable regularity lemma revisited*, Proceedings of the American Mathematical Society **144** (2016), 1761–1765.
318. M. Malliaris and S. Shelah, *Regularity lemmas for stable graphs*, Transactions of the American Mathematical Society **366** (2014), 1551–1585.
319. ———, *Notes on the stable regularity lemma*, Bulletin of Symbolic Logic **27** (2021), 415–425.
320. V. W. Marek and M. Truszczyński, *Nonmonotonic logic*, Artificial Intelligence, Springer, 1993.
321. G. A. Margulis, *Discrete subgroups of semisimple Lie groups*, Ergebnisse der Mathematik un ihrer Grenzgebiete, 3. Folge, no. 17, Springer, 1991.
322. D. Marker, *Model theory*, Graduate Texts in Mathematics, no. 217, Springer, 2002.
323. A. Marks and S. Unger, *Baire measurable paradoxical decompositions via matchings*, Advances in Mathematics **289** (2016), 397–410.
324. R. Marshall, *Robust classes of finite structures*, Ph.D. thesis, University of Leeds, 2008.
325. P. Martin-Löf, *The definition of random sequences*, Information and Control **9** (1966), 602–619.
326. 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–.
327. K. R. Matthews and A. M. Watts, *A generalization of Hasse's generalization of the Syracuse algorithm*, Acta Arithmetica **43** (1984), 167–175.
328. P. Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, no. 44, Cambridge University Press, 1995.
329. A. D. Matushkin, *Zero-one law for random uniform hypergraphs*, preprint, 2016.
330. A. D. Matushkin and M. E. Zhukovskii, *First order sentences about random graphs: small number of alternations*, Discrete Applied Mathematics **236** (2018), 329–346.
331. J. E. Maxfield, *Normal k-tuples*, Pacific Journal of Mathematics **3** (1953), 189–196.
332. E. Mayordomo, *A Kolmogorov complexity characterization of constructive Hausdorff dimension*, Information Processing Letters **84** (2002), 1–3.
333. W. Merkle, N. Mihalović, and T. A. Slaman, *Some results on effective randomness*, Theory of Computing Systems **39** (2006), 707–721.
334. J.-F. Mertens, *Stochastic games*, Handbook of Game Theory, vol. 3, Elsevier, 2002, pp. 1809–1832.
335. J.-F. Mertens and A. Neyman, *Stochastic games*, International Journal of Game Theory **10** (1981), 53–66.
336. P. Michel, *Busy beaver competition and Collatz-like problems*, Archive for Mathematical Logic **32** (1993), 351–367.
337. 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–.
338. C. F. Miller, III, *On group-theoretic decision problems and their classification*, Annals of Mathematics Studies, no. 68, Princeton University Press, 1971.
339. 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.
340. G. L. Miller, *Riemann's hypothesis and tests for primality*, Journal of computer and system science **13** (1976), 300–317.
341. R. Miller, *Isomorphism and classification for countable structures*, Computability **8** (2019), 99–117.
342. J. Milnor, *Dynamics in one complex variable*, 3rd ed., Annals of Mathematics Studies, no. 160, Princeton University Press, 2006.
343. A. Montalbán, *Computable structure theory: Within the arithmetic*, preprint, 2019.
344. *****A. Montalban, *Computable structure theory, part ii*, not yet seen, 2021?
345. 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.
346. C. C. Moore, *Ergodicity of flows on homogeneous spaces*, American Journal of Mathematics **88** (1966), 154–178.

347. A. De Morgan, *Formal logic: or, the calculus of inference, necessary and probable*, Taylor and Walton, 1847.
348. 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.
349. A. A. Muchnik, A. L. Semenov, and V. A. Uspensky, *Mathematical metaphysics of randomness*, Theoretical Computer Science **207** (1998), 263–317.
350. K. P. Murphy, *Dynamic Bayesian networks*, Ph.D. thesis, University of California, Berkeley, 2002.
351. 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.
352. A. Nies, *Computability and randomness*, Oxford Logic Guides, no. 51, Oxford, 2009.
353. N. Nisan and A. Wigderson, *Hardness vs. randomness*, Journal of Computer Systems and Sciences **49** (1994), 149–167.
354. 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.
355. 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.
356. D. N. Osherson and S. Weinstein, *Criteria of language learning*, Information and Control **52** (1982), 123–138.
357. D. Osin, *A topological zero-one law and elementary equivalence of finitely generated groups*, Annals of Pure and Applied Logic **172** (2021), 102915.
358. 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.
359. J. Paris and A. Venkovská, *Pure inductive logic*, Perspectives in Logic, Cambridge, 2015.
360. J. Pearl, *Deciding consistency in inheritance networks*, Tech. Report 870053, Cognitive Systems Laboratory, UCLA, 1987.
361. ———, *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann, 1988.
362. J. Pearl and A. Paz, *Graphoids: A graph-based logic for reasoning about relevance relations*, Tech. Report 850038, Cognitive Systems Laboratory, UCLA, 1985.
363. Y. B. Pesin, *Dimension theory in dynamical systems*, University of Chicago Press, 1997.
364. P. Petersen, *Riemannian geometry*, 3rd ed., Graduate Texts in Mathematics, no. 171, Springer, 2016.
365. 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.
366. R. R. Phelps, *Lectures on choquet's theorem*, Lecture Notes in Mathematics, no. 1757, Springer, 2001.
367. A. Pillay and C. Steinhorn, *Discrete o-minimal structures*, Annals of Pure and Applied Logic **34** (1987), 275–289.
368. M. S. Pinsker, *On the complexity of a concentrator*, 7th International Teletraffic Conference, 1973.
369. K. Popper, *The logic of scientific discovery*, Routledge, 2002, Original edition published in 1935; this edition dates from text of 1959.
370. E. L. Post, *Recursive unsolvability of a problem of Thue*, The Journal of Symbolic Logic **12** (1947), 1–11.
371. P. Potgieter, *Algorithmically random series and Brownian motion*, Annals of Pure and Applied Logic **169** (2018), 1210–1226.
372. M. O. Rabin, *Probabilistic algorithm for testing primality*, Journal of Number Theory **12** (1980), 128–138.
373. C. Radin and L. Sadun, *Phase transitions in a complex network*, Journal of Physics A **46** (2013), 305002.
374. 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.
375. J. Reimann, *Computability and fractal dimension*, Ph.D. thesis, Ruprecht-Karls-Universität Heidelberg, 2004.
376. R. Reiter, *A logic for default reasoning*, Artificial Intelligence **13** (1980), 81–132.
377. ———, *Nonmonotonic reasoning*, Annual Review of Computer Science **1987** (1987), 147–186.

378. L. J. Richter, *Degrees of structures*, The Journal of Symbolic Logic **46** (1981), 723–731.
379. A. Rivkind and O. Barak, *Local dynamics in trained recurrent neural networks*, Physical Review Letters **118** (2017), no. 258101, 1–5.
380. A. Robinson, *Complete theories*, Studies in logic and the foundations of mathematics, North-Holland, 1956.
381. F. Rosenblatt, *Principles of neurodynamics; perceptrons and the theory of brain mechansims*, Spartan, 1962.
382. S. Ross, *Introduction to probability models*, 12th ed., Academic Press, 2019.
383. J. J. Rotman, *The theory of groups*, Allyn and Bacon, 1965.
384. A. Rumyantsev, *Infinite computable version of Lovász local lemma*, preprint, 2010.
385. A. Rumyantsev and A. Shen, *Probabilistic constructions of computable objects and a computable version of Lovász local lemma*, Fundamental Informaticae **132** (2014), 1–14.
386. S. J. Russell and P. Norvig, *Artificial intelligence: A modern approach*, third ed., Prentice Hall Series in Artificial Intelligence, Prentice Hall, 2010.
387. 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.
388. M. J. Ryten, *Model theory of finite difference fields and simple groups*, Ph.D. thesis, The University of Leeds, 2007.
389. N. Sauer, *On the density of families of sets*, Journal of Combinatorial Theory (A) **13** (1972), 145–147.
390. 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.
391. A.-M. Scheerer, *Computable absolutely normal numbers and discrepancies*, Mathematics of Computation (2017), 2911–2926.
392. J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Networks **61** (2015), 85–117.
393. K. Schmidt, *Asymptotically invariant sequences and an action of* $SL(2, Z)$ *on the* 2*-sphere*, Israel Journal of Mathematics **37** (1980), 193–208.
394. W. M. Schmidt, *Über die Normalität von Zahlen zu verschiedenen Basen*, Acta Arithmetica **VII** (1962), 299–309.
395. ———, *Irregularities of distribution, VII*, Acta Arithmetica **21** (1972), 45–50.
396. C. P. Schnorr, *A unified approach to the definition of random sequences*, Mathematical systems theory **5** (1971), 246–258.
397. ———, *Zufälligkeit und Wahrscheinlichkeit*, Lecture Notes in Mathematics, no. 218, Springer, 1971.
398. G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.
399. A. Shamir, **IP = PSPACE**, Journal of the Association for Computing Machinery **39** (1992), 869–877.
400. C. E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal **27** (1948), 379–423.
401. L. S. Shapley, *Stochastic games*, Proceedings of the National Academy of Sciences **39** (1953), 1095–1100.
402. S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*, Pacific Journal of Mathematics **41** (1972), 247–261.
403. ———, *Classification theory and the number of non-isomorphic models*, revised ed., Studies in Logic and the Foundations of Mathematics, no. 92, North-Holland, 1990.
404. S. Shelah and J. Spencer, *Zero-one laws for sparse random graphs*, Journal of the American Mathematical Society **1** (1988), 97–115.
405. A. Shen, *IP = PSPACE: Simplified proof*, Journal of the Association for Computing Machinery **39** (1992), 878–880.
406. A. K. Shen, *On relations between different algorithmic definitions of randomness*, Sovient Mathematics Doklady **38** (1989), 316–319.
407. A. N. Shiryaev, *Probability*, second ed., Graduate Texts in Mathematics, no. 95, Springer, 1996.
408. I. Shmulevich and E. R. Dougherty, *Genomic signal processing*, Princeton Series in Applied Mathematics, Princeton University Press, 2007.

409. ———, *Probabilisitc boolean networks*, Society for Industrial and Applied Mathematics, 2010.
410. H. A. Simon, *On a class of skew distribution functions*, Biometrika **42** (1955), 425–440.
411. P. Simon, *A guide to NIP theories*, Lecture Notes in Logic, no. 44, Cambridge, 2015.
412. R. I. Soare, *Recursively enumerable sets and degrees*, Springer-Verlag, 1987.
413. J. Spencer, *Asymptotic lower bounds for Ramsey functions*, Discrete Mathematics **20** (1977), 69–76.
414. ———, *Threshold functions for extension statements*, Journal of Combinatorial Theory A **53** (1990), 286–305.
415. ———, *Threshold spectra via the Ehrenfeucht game*, Discrete Applied Mathematics **30** (1991), 235–252.
416. ———, *The strange logic of random graphs*, Algorithms and Combinatorics, no. 22, Springer, 2001.
417. J. Spencer and K. St. John, *The tenacity of zero-one laws*, The Electronic Journal of Combinatorics **8** (2001), R17.1–R17.14.
418. J. Spencer and M. E. Zhukovskii, *Bounded quantifier depth spectra for random graphs*, Discrete Mathematics **339** (2016), 1651–1664.
419. L. Staiger, *Kolmogorov complexity and Hausdorff dimension*, Information and Computation **103** (1993), 159–194.
420. ———, *A tight upper bound on Kolmogorov complexity and uniformly optimal prediction*, Theory of Computing Systems **31** (1998), 215–229.
421. C. I. Steinhorn, *Borel structures and measure and category logics*, Model Theoretic Logics, Perspectives in Logic, no. 8, Springer, 1985, pp. 579–596.
422. G. Stengle and J. E. Yukich, *Some new Vapnik-Chervonenkis classes*, The Annals of Statistics **17** (1989), 1441–1446.
423. V. E. Stepanov, *Phase transitions in random graphs*, Theory of Probability and its Applications **15** (1970), 187–203.
424. F. Stephan and Yu. Ventsov, *Learning algebraic structures from text*, Theoretical Computer Science **268** (2001), 221–273.
425. L. J. Stockmeyer, *The polynomial-time hierarchy*, Theoretical Computer Science **3** (1976), 1–22.
426. G. Stuck and R. Zimmer, *Stabilizers for ergodic actions of higher rank semisimple groups*, Annals of Mathematics **139** (1994), 723–747.
427. M. Studený, *Conditional independence relations have no finite complete characterization*, preprint, 1992.
428. D. Sussillo and O. Barak, *Opening the black box: Low-dimensional dynamics in high-dimensional recurrent neural networks*, Neural Computation **25** (2013), 626–649.
429. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arithmetica **27** (1975), 199–245.
430. ———, *Regular partitions of graphs*, preprint, 9 pp., 1975.
431. 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.
432. ———, *Szemerédi's proof of Szemerédi's theorem*, Acta Mathematica Hungarica **161** (2020), 443–487.
433. A. Tarski, *Algebraische Fassung des Maßproblems*, Fundamenta Mathematicae **31** (1938), 47–66.
434. ———, *Cardinal algebras*, Oxford University Press, 1949.
435. A. Taveneaux, *Randomness zoo*, preprint, 2012.
436. E. Thoma, *Die unzerlegbaren, positiv-definiten Klassenfunktionen der abzählbar unendlichen, symmetrischen Gruppe*, Mathematische Zeitschfrift **85** (1964), 40–61.
437. ———, *Über unitäre Darstellungen abzählbarer, diskreter Gruppen*, Mathematische Annalen **153** (1964), 111–138.
438. S. Thomas, *The classification problem for torsion-free Abelian groups of finite rank*, Journal of the American Mathematical Society **16** (2003), 233–258.
439. 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.
440. ———, *Invariant random subgroups of inductive limits of finite alternating groups*, Journal of Algebra **503** (2018), 474–533.

441. A. Thue, *Probleme über Veränderungen von Zeichenreihen nach gegeben regeln*, Skrifter utgit av Videnskapsselskapet i Kristiania **1** (1914), no. 10, 1–34.
442. N. Tishby and N. Zaslavsky, *Deep learning and the information bottleneck principle*, IEEE Information Theory Workshop, 2015.
443. S. Toda, *PP is as hard as the polynomial-time hierarchy*, SIAM Journal of Computing **20** (1991), 878–880.
444. H. Towsner, *$\sigma$-algebras for quasirandom hypergraphs*, Random Structures and Algorithms **50** (2017), 114–139.
445. ———, *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.
446. S. Vadhan, *Pseudorandomness*, Foundations and Trends in Theoretical Computer Science, vol. 7, Now, 2012.
447. L. G. Valiant, *The complexity of computing the permanent*, Theoretical Computer Science **8** (1979), 189–201.
448. ———, *A theory of the learnable*, Communications of the ACM **27** (1984), 1134–1142.
449. 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.
450. 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.
451. A. M. Vershik, *Totally nonfree actions and the infinite symmetric group*, Moscow Mathematical Journal **12** (2012), 193–212.
452. N. Vieille, *Stochastic games: Recent results*, Handbook of Game Theory, vol. 3, Elsevier, 2002, pp. 1833–1850.
453. Susan Vineberg, *Dutch Book Arguments*, The Stanford Encyclopedia of Philosophy (Edward N. Zalta, ed.), Metaphysics Research Lab, Stanford University, spring 2016 ed., 2016.
454. A. Visser, *Numerations, $\lambda$-calculus, & arithmetic*, To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus, and Formalism, Academic Press, 1980, pp. 259–284.
455. J. von Neumann, *Zur allgemeinen Theorie des Masses*, Fundamenta Mathematicae **13** (1929), 73–116.
456. P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, no. 79, Springer, 1982.
457. R. Weber, *Computability theory*, Student Mathematical Library, no. 62, American Mathematical Society, 2012.
458. K. Weihrauch, *Computable analysis*, Texts in Theoretical Computer Science, Springer, 2000.
459. 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.
460. J. S. Wilson, *On simple pseudofinite groups*, Journal of the London Mathematical Society **51** (1995), 471–490.
461. D. Xiao, *New perspectives on the complexity of computational learning, and other problems in theoretical computer science*, Ph.D. thesis, Princeton University, 2009.
462. ———, *On basing $\mathbf{ZK} \neq \mathbf{BPP}$ on the hardness of PAC learning*, 24th Annual IEEE Conference on Computational Complexity, 2009, pp. 304–315.
463. A. C. Yao, *Theory and applications of trapdoor functions*, 23rd Annual Symposium on Foundations of Computer Science, IEEE, 1982, pp. 80–91.
464. C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, *Understanding deep learning requires rethinking generalization*, ICLR, 2017.
465. 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.
466. M. E. Zhukovskii, *On infinite spectra of first order properties of random graphs*, Moscow Journal of Combinatorics and Number Theory **4** (2016), 73–102.
467. R. J. Zimmer, *Ergodic theory and semisimple groups*, Monographs in Mathematics, no. 81, Birkhäuser, 1984.