0-1 LAWS IN LOGIC: AN OVERVIEW

ANAND ABRAHAM

Abstract. I give an overview of 0-1 laws in logic. I give a proof of the 0-1 law for first-order logic, and then guide the reader through a series of extensions of the logic in which the law fails or continues to hold.

Contents

1. Introduction

In this paper, I give an overview of 0-1 laws in logic. This is a fun topic about the limits of what a language can say. The intended audience of this paper is people who have some level of mathematical maturity, though no experience in logic is assumed. All logical concepts or results that are used in this paper are defined or stated beforehand. Because I try to go over a large amount of results, it is an unfortunate necessity that not all proofs are given. In many cases, the proofs are long enough to require a paper of their own, or at very least too complicated to easily give at the appropriate level. The purpose of this paper is above all to bring 0-1 laws to the attention of the reader, and to excite them. If sufficiently excited, the curious reader is encouraged to look into the sources which the omitted proofs come from.

The name "0-1 law" seems out of place in logic. In probability theory, a 0-1 law says that under certain conditions, a certain event has either probability 0 or 1. The field of logic is concerned with what can be said within a specific mathematical language–exactly what structures a sentence picks out. To bring 0-1 laws into logic, we have to introduce a notion of the probability of a sentence. It turns out that in first-order logic with only relation symbols, the probability of every sentence is either 0 or 1. This is an interesting result, because it says that no sentence of first-order logic can divide the space of structures into two decently-sized halves.

The first proof of this fact was found by Fagin in [1]. Although the actual proof itself is not terribly challenging, what is interesting about it is how it establishes a

Date: Sept. 26th, 2018.

correspondence between countable random structures and asymptotic probability. This correspondence is not obvious at all, and exciting to consider, since countable random objects are not just used for the proof of the 0-1 law, but are interesting in their own right.

2. Background

This section is meant to serve as a brief overview of first-order logic for those not familiar with it. Here we give the definitions of a language and a model, as well as the statements of some some classic and useful theorems. Those who are already familiar with first-order logic may skip this section.

To be able to talk about the formulas of a logic, we first need a vocabulary. A *language* $\mathcal{L}$ is a collection of relation, function, and constant symbols. Relation and function symbols also have specified *arities*: the number of arguments they take. A language with only relation symbols is called *relational*. Also featured in logical formulas are the standard propositional connectives, parentheses, variables, and quantifiers. Formula are formed as follows: a *term* is a variable, a constant, or a k-ary function applied to k terms. An *atomic proposition* is a k-ary relation applied to k terms. *Formulas* are defined inductively as follows: atomic propositions are formulas, boolean combinations of formulas are formulas, and quantifications over formulas are formulas. Any variable which is quantified over somewhere in a formula is *bound*, and otherwise it is *free*. A *sentence* is a formula with no free variables.

It is common to write $\varphi(x, y)$ to indicate that φ has x and y as free variables. A formula such as $\varphi(a, b)$ is to be be interpreted as the substitution of a and b for x and y, respectively. We will also often deal with arbitrary but finite lists of variables. In that case, it is convenient to use $\vec{x}$ to refer to them. So $\exists x_1 \dots \exists x_n \varphi(x_1, \dots, x_n)$ may be written more succinctly as $\exists \vec{x} \varphi(\vec{x})$. In general, we will use φ, ψ, χ, and θ to mean formulas, P to mean a unary relation, R to mean a general relation, a, b to mean elements, and x, y to mean variables. Another piece of notation worth noting is $\exists! x$, which is to be read "there exists a unique x such that". $\exists! x \varphi(x)$ is defined to be equivalent to $\exists x(\varphi(x) \wedge \forall y(\varphi(y) \rightarrow y = x))$. Finally, $\perp$ is used to denote a sentence which is always false (such as $\exists x, x \neq x$), and $\top$ is used to denote a sentence which is always true (such as $\forall x, x = x$).

A sentence is the kind of object which can be true or false depending on context. On the other hand, a formula with free variables is to be thought of as only being true or false once we know what the variables mean. To use an analogy from English, something like "all birds can fly" is like a sentence, while "he knows how to swim" is like a formula. The former has a truth value: false, because penguins cannot. The latter's truth value depends on what is meant by "he". One way of interpreting "he knows how to swim" might say that it is true if, no matter who "he" is, the sentence is true. We call this the *universal closure* of "he knows how to swim". The universal closure of a formula is the formula, with the universal quantification of each free variable appended to the front. This can be used as a notion of "truth" for a formula when it is convenient to have one.

A *model* is a setting in which sentences can be true or false. More specifically, a *model* for the language $\mathcal{L}$, also known as an $\mathcal{L}$-*structure*, consists of a base set A and a collection of interpretations for each symbol in the language: for each constant, an element of A is picked out, for each k-ary relation, a subset of A^k is picked out, and for each k-ary function, a function $A^k \to A$ is picked out. Formally, the model is all of these choices put together in a tuple. We call an element of the base set an element of the model, and often as an abuse of notation we use the same symbol for both the whole model and its base set. A sentence is true in a model if it is true when all relations, functions, and constants are interpreted as the model specifies. If φ is true in the model A, we write $A \models \varphi$. For technical definitions of all of the above, the reader can refer to [2]. The *size* of a model is the size of its domain.

Models are all around us–while they might not capture all the structure of a mathematical object, they give us a way to choose what part of it we are looking at and abstract away the structure. $\mathbb{N}$, the natural numbers, gives birth to a variety of models. We can look at its order, the addition operation, the multiplication operation, the exponentiation operation, or any combination thereof, to pick out different models.

Two important concepts are *validity* and *satisfiability*. A sentence is said to be *valid* if it is true in every model, and a sentence is said to be *satisfiable* if it is true in at least one model. These concepts also have finite analogues: a sentence is said to be *finitely valid* if it is true in all finite models, and it is said to be *finitely satisfiable* if it has a finite model. Finite validity and satisfiability are not the same as validity and satisfiability. For example, the sentence "if a function is surjective, it is injective" is expressible in first-order logic, and this is true in every finite model, since the function's domain and range are the same finite set. This is false in general models, though: for example, we can map $\mathbb{N}$ onto itself by mapping every n to $\lfloor \frac{n}{2} \rfloor$. On the other hand, "$R$ is a dense linear order without endpoints" is satisfiable, since $\mathbb{Q}$ is a model, but is not finitely satisfiable, since density necessitates infinitely many elements.

Two models A, B are said to be *isomorphic* if there exists an f such that for every atomic formula $\varphi(\vec{x})$ and every $\vec{a}$ consisting of elements of A, $A \models \varphi(\vec{a})$ if and only if $B \models \varphi(\vec{b})$, where $\vec{b}$ is the result of applying f to each element of $\vec{a}$. Essentially, this definition says that an isomorphism is a map that respects functions, relations, and constants.

Technically, there exists an empty model for every language. This model is very strange, and plenty of properties that we expect of most models do not work here–all existential statements are false, because nothing exists, and all universal statements are true, because they quantify over an empty set. Since the empty model is so strange, it is convenient to pretend it does not exist, rather than state it as an edge case whenever necessary. In this paper it is assumed that the empty model is nonexistent.

When considering formulas, we care less about their syntactic forms and more about their meaning. For example, no one would dispute that $\varphi \wedge (\psi \wedge \chi)$ and $(\varphi \wedge \psi) \wedge \chi$ mean essentially the same thing. We say that two sentences φ and ψ are *equivalent* if they are true in exactly the same models. We say that two formulas φ and ψ are equivalent if in every model exactly the same elements (or tuples) satisfy both formulas.

If we only care about formulas up to equivalence, we can pick nice representatives for each formula. A formula is said to be in *prenex normal form* if all of its quantifiers are at the front. The following proposition says that we can always put a formula into prenex normal form. Its proof is by induction, the main method of proving properties of all formulas.

Proposition 2.1. *Every formula is equivalent to one in prenex normal form.*

Proof. (Sketch) First, note that any atomic proposition is trivially in prenex normal form, because it has no quantifiers. Now, suppose φ is in prenex normal form. Then $\forall x\varphi$ and $\exists x\varphi$ are too, where x is an arbitrary variable. Furthermore, $\neg\varphi$ is, because $\neg\forall x\varphi$ is equivalent to $\exists x\neg\varphi$ and vice versa, so we can shuffle the negation past each quantifier. Now suppose that ψ is in prenex normal form too. We may assume that φ and ψ do not share any variables, because we can simply rename all of the variables in ψ and get an equivalent formula. Now, in $\varphi \wedge \psi$ and $\varphi \vee \psi$ we can simply put all of the quantifiers from φ in the front, and all the quantifiers from ψ after that. Since neither of them capture variables in the other, this does not cause a problem. So any boolean combination of formulas, each of which can be put in prenex normal form, can also be put in prenex normal form. Since any formula is either atomic, the quantification over a formula, or the boolean combination of some, it follows that all formulas may be put in prenex normal form.

There is a slight caveat in that $(\forall x\varphi) \vee \psi$ and $\forall x(\varphi \vee \psi)$ are not completely equivalent, because in the empty model all universal sentences are trivially true. But this problem disappears if we banish the empty model, as we were already planning to do. $\square$

One thing to note about this proof is that it provides an algorithm for turning formulas into prenex normal form, as well, so it is never mysterious to which prenex normal form formula a formula is equivalent. Essentially, this allows us to assume that all formulas have all of their quantification in the front, which lets us classify formulas based on their quantifiers.

First order logic also comes with its own notion of what is provable. I will not go into the details here, but some notation is in order. Let Γ be a set of sentences. We say $\Gamma \vdash \varphi$ if φ can be proven from Γ. Any φ such that $\Gamma \vdash \varphi$ is called a *consequence* of Γ. We say that a set of sentences Γ is *consistent* so long as it cannot prove a contradiction (like $\exists x(x \neq x)$). A crucial property of proofs is that they are finite in nature: if φ follows from Γ, then it follows from finitely many elements of Γ. One foundational result in logic is the soundness-completeness theorem.

Theorem 2.2 (Soundness-Completeness)**.** *For any language $\mathcal{L}$ and set of sentences Γ, $\Gamma \vDash \varphi$ if and only if $\Gamma \vdash \varphi$.*

The backwards direction is called soundness and the forwards is called completeness. Soundness can be proven by showing that if the premises of a proof hold in a model, the conclusion does. Completion requires a more elaborate construction first given by Gödel in [3]. An alternate phrasing of this theorem is "Γ is consistent if and only if it has a model." This is because being inconsistent is the same as saying $\Gamma \vdash \perp$, which by soundness-completeness is the same as $\Gamma \vDash \perp$, which means that the set of models satisfying Γ is empty. This statement of the theorem is useful in the proof of another important theorem, the compactness theorem.

Theorem 2.3 (Compactness)**.** *For any language $\mathcal{L}$ and set of sentences Γ, Γ has a model if and only if all of its finite subsets Γ' have a model.*

Proof. If Γ has a model, then that model is already a model of all of its finite subsets. On the other hand, if Γ does not have a model, then by completeness a contradiction can be deduced from Γ. As proofs are finite, this means that a contradiction can be deduced from a finite subset of Γ. So there is a finite subset of Γ without a model. $\square$

This theorem gives a profound limit on what can be defined by sets of sentences in first-order logic. One consequence is that if a set of sentence has infinite models, we can add constants to the language and axioms saying that they are unique to get models of higher cardinality. This is the "upwards" part of the following result.

Theorem 2.4 (Upwards-Downwards Löwenheim-Skolem-Tarski)**.** *If the set of sentences Γ has an infinite model, then it has models of every infinite cardinality.*

A proof of this fact can be found in [2]. True to its name, it was proven independently, at least in part, by Löwenheim, Skolem, and Tarski.

A set of sentences is also often called a *theory*. A theory is said to be *complete* if for every sentence in the language, either it or its negation is a consequence of the theory. The following is a useful criterion for proving if a theory is complete.

Proposition 2.5 (Łoś-Vaught test)**.** *Let T be a consistent theory with no finite models, and suppose that for some cardinal κ, T only has one model of size κ up to isomorphism. Then T is complete.*

Proof. Suppose T is not complete. Then there is some φ for which $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are both consistent. Thus both have models, and since both models are models of T, they are infinite. Thus by upwards-downwards Löwenheim-Skolem-Tarski, for every infinite cardinality there is both a model of $T \cup \{\varphi\}$ and one of $T \cup \{\neg\varphi\}$. Both are models of T, and they cannot be isomorphic, so T does not have exactly one model up to isomorphism for any infinite cardinality. By contraposition, if T does have exactly one model up to isomorphism for some infinite cardinality, then it is complete. $\square$

Some more concepts that are useful in logic are those of computability and decidability. A set is said to be *computable* if there exists an algorithm for deciding whether or not an element is in that set. There is a formal definition of computability, but for most purposes an intuitive notion suffices. A theory is said to be *decidable* if the set of consequences of the theory is computable. A theory is said to be *computably axiomatizable* if there is a computable set of sentences of which every sentence of the theory is a consequence. A complete and computably axiomatizable theory is decidable: to decide whether φ is a consequence, we can list all the proofs from the axioms of the theory, and since the theory is complete eventually either φ or $\neg\varphi$ will show up.

3. 0-1 Law for First Order Logic

In order to prove the 0-1 law for first-order logic, we first need a reasonable notion of the probability of a sentence. When considering a finite class of structures, this appears to be simple: just take the fraction of structures in which the sentence is true. Unfortunately, the space of models is impossibly large, so our approach will be to take the limit of a series of finite approximations instead.

For a sentence φ, let $l_n(\varphi)$ be the fraction of all models of size n where φ holds. Let $l(\varphi) = \lim_{n\to\infty} l_n(\varphi)$. In the case that this limit exists, $l(\varphi)$ is called the *(labeled) asymptotic probability* of φ. When $l(\varphi) = 1$, we say that φ is *almost surely true*, and when $l(\varphi) = 0$ we say it is *almost surely false*. The following lemma shows that the propositional connectives treat almost-sure truth and falsity the same as truth and falsity.

Lemma 3.1. *(a) $l(\varphi) = 1$ if and only if $l(\neg\varphi) = 0$.*
(b) $l(\varphi \wedge \psi) = 1$ if and only if $l(\varphi) = 1$ and $l(\psi) = 1$.

Proof. (a) Suppose $l(\varphi) = 1$. Since, in a model, φ is true if and only if $\neg\varphi$ is false, it follows that $l_n(\neg\varphi) = 1 - l_n(\varphi)$. So $\lim_{n\to\infty} l_n(\neg\varphi) = 1 - \lim_{n\to\infty} l_n(\varphi) = 0$. The other direction is the same.

(b) Suppose $l(\varphi \wedge \psi) = 1$. Since φ and ψ have to be true for $\varphi \wedge \psi$ to be, $l_n(\varphi) > l_n(\varphi \wedge \psi)$ and $l_n(\psi) > l_n(\varphi \wedge \psi)$ for all n. As $\lim_{n\to\infty} l_n(\varphi \wedge \psi) = 1$, it follows that $\lim_{n\to\infty} l_n(\varphi)$ and $\lim_{n\to\infty} l_n(\psi)$ are 1. Now suppose $l(\psi) = 1$ and $l(\varphi) = 1$. By inclusion-exclusion, $l(\varphi) + l_n(\psi) - l_n(\varphi \wedge \psi) \leq 1$. So we have:

$$l_n(\varphi) + l_n(\psi) - 1 \leq l_n(\varphi \wedge \psi) \leq 1$$

Taking the limit $n \to \infty$, we get that $1 \leq l(\varphi \wedge \psi) \leq 1$, so $l(\varphi \wedge \psi) = 1$. □

Since every Boolean combination of formulas is built up from $\neg$ and $\wedge$, this means that the propositional truth conditions of $\vee$ and $\to$ also carry over with almost-sure validity. In particular, this means that if $l(\varphi) = 1$ and $l(\varphi \to \psi) = 1$, then $l(\psi) = 1$. That is, the consequence of almost-surely true sentences is almost-surely true.

With this notion of the probability of a sentence, we can now state the 0-1 law for first order logic.

Theorem 3.2. *Let φ be a sentence of first-order logic over a finite relational vocabulary. Then φ is either almost surely true or almost surely false.*

To prove this, we construct a model known as the *countable random graph*, first constructed in [4]. Then we prove a transfer principle, showing that being true in the graph is equivalent to being almost surely true. Because the countable random graph is a model, every sentence is either true in it or false in it, so it will follow that every sentence is either almost surely true or almost surely false.

To construct the countable random graph, we take a countable infinite set and, between any pair of points, assign an edge with a probability of $\frac{1}{2}$. It turns out that if you do this, then with probability 1 the graph will have the following property, $(*)$:

- $(*)$ For any finite set of vertices S, for any partition of S into $U \sqcup V$, there is a vertex x such that x is adjacent to every element of U and not adjacent to any element of V.

This property means that we can build up any finite graph as a subgraph of the random graph, since when picking out points we can always pick out a point which is related to the other points in exactly the way that we want. In addition, we will show that all graphs which satisfy $(*)$ are isomorphic, which is why we call our graph *the* countable random graph.

Proposition 3.3. *There exists a graph satisfying $(*)$.*

Proof. Take the natural numbers as our vertex set. We construct a graph by, between any two points, assigning a $\frac{1}{2}$ chance that there is an edge between them. Now, given any finite $S \subset \mathbb{N}$, with $S = U \sqcup V$, $|S| = n$, we can calculate the chance that there is an element of $\mathbb{N}$ connected to all elements of U and none of V. Let x

be the first number not in S. The chance that x is connected to or not connected to any element of S is $\frac{1}{2}$, and since these events are independent, the chance that x is connected to all of them in the "right" way is $\frac{1}{2^n}$. Similarly, the chance that $x+1$ is connected to all of them in the "right" way is also $\frac{1}{2^n}$, and this is the same for any element above x. Since there are infinitely many of these, and the same finite chance that the event occurs for each of them, the event that there is a point which connects to U and V in exactly the right way has probability 1.

Now, there are countably many choices of S and finitely many ways of decomposing each S into $U \sqcup V$, so the chance that $(*)$ holds is the intersection of countably many events of probability 1. Thus, $(*)$ holds with probability one. □

This does not give an explicit construction of a graph satisfying property $(*)$, but it does show that such a graph exists–in fact, many constructions of this graph exist, since by flipping a coin to assign edges between points you will almost surely come across such a graph. The following proposition shows that the graph constructed is unique up to isomorphism.

Proposition 3.4. *Any two countable graphs satisfying $(*)$ are isomorphic.*

Proof. Let A and B be graphs satisfying $(*)$. We will use a "back-and-forth" construction, showing that we can always extend a partial isomorphism between finite subsets of A, B by adding either a point to the range or to the domain. Let A', B' be finite subsets of A, B with f an isomorphism between them. Let $x \in A, x \notin A'$. To extend f to include x as an input, we need to make sure that $f(x)$ is related to all of the other members of B' in the "right" way. By $(*)$, there is an element y of B which is connected to exactly the elements of B' that correspond to elements of A' which x is connected to. So let $f(x) = y$. Similarly, given a $y \in B, y \notin B'$ we can use $(*)$ in A to find a suitable x, and again set $f(x) = y$ and maintain f as an isomorphism. These are called the *back-and-forth properties.*

Now we can construct an isomorphism between A and B proper. We will define it in stages. Enumerate $A = \{x_1, x_2, \dots\}$ and $B = \{y_1, y_2, \dots\}$. Let f_0 be the map from $\emptyset$ to $\emptyset$. Given f_{2k}, we construct f_{2k+1} as follows: if x_k is already in the domain of f_{2k}, do nothing. If it is not, find a suitable y to map it to, and let $f_{2k+1} = f_{2k} \cup \{(x_k, y)\}$. Given f_{2k+1}, we construct f_{2k+2} the same way in reverse: if y_k is already in the range of f_{2k+1}, do nothing, but if not, find a suitable x and set $f_{2k+2} = f_{2k+1} \cup \{(x, y_k)\}$. So, at odd stages we ensure that every element of A is in the domain, and at even stages we ensure that every element of B is in the range.

Let $f = \bigcup_{n \in \mathbb{N}} f_n$. By construction f is a bijection between A and B, and for any $a, b \in A$, there is a partial isomorphism f_k which f extends that has both in the domain, which guarantees that a and b are adjacent if and only if $f(a)$ and $f(b)$ are. □

Unfortunately, $\mathcal{L}$-structures are not exactly the same as graphs. However, an analogue version of $(*)$ for $\mathcal{L}$-structures can be stated, and analogues of propositions 3.3 and 3.4 can be proved. For any $\mathcal{L}$, the analogue of $(*)$ for $\mathcal{L}$-structures, $(*')$, is given as follows:

- $(*')$ For any finite subset S of the domain, with for each k-ary $R \in \mathcal{L}$, $U_R \subset (S \cup \{x\})^k$ such that every tuple in U_R has an occurrence of x, there exists an $x \notin S$ such that for all R, $R\vec{a} \iff \vec{a} \in U_R$.

This says that for any finite subset of S, we can find an x which relates to all of its elements in exactly the way that we want. The analogue of proposition 3.3 can be proved by more or less the same argument.

Proposition 3.5. *There exists a countable $\mathcal{L}$-structure satisfying $(*')$.*

Proof. Again, let the domain be $\mathbb{N}$ and decide each relation with probability $\frac{1}{2}$. For any given choice of S and the U_Rs, there is a finite probability for any point not in S that it is a suitable x, so with probability 1 there is a suitable x. Since our vocabulary is finite there are a countable amount of choices total, and the intersection of countably many events of probability 1 has probability 1. □

Similarly, the analogue of proposition 3.4 can also be proved:

Proposition 3.6. *Any two countable $\mathcal{L}$-structures satisfying $(*')$ are isomorphic.*

Proof. Given two such $\mathcal{L}$-structures, it is not hard to see that they satisfy the back-and-forth properties, so that we can always extend a partial isomorphism with either an element from the domain or the range. Because of this, the exact same construction as in the proof of proposition 3.4 also constructs an isomorphism between these two $\mathcal{L}$-structures. □

So, in the same way that a countable random graph exists, there is also a countable random $\mathcal{L}$-structure for any $\mathcal{L}$.

In order to study this model with methods from logic, we need to give it an axiomatization. Luckily, the property $(*')$ fundamentally concerns finite objects, so we can give a set of axioms which capture it. These are called *extension axioms.* In a given $\mathcal{L} = \{R_1, \ldots, R_m\}$, for any n and $U_R \subset \{x_1, \ldots, x_n, x_{n+1}\}^k$ for every k-ary $R \in \mathcal{L}$, the extension axiom $\psi_{n,U_{R_1},\ldots,U_{R_m}}$ is defined as follows:

$$\forall x_1, \ldots x_n\Big(\bigwedge_{1 \leq i \neq j \leq n} x_i \neq x_j \to \Big(\exists x_{n+1}\Big(\bigwedge_{1 \leq i \leq n} x_i \neq x_{n+1} \wedge \bigwedge_{R \in \mathcal{L}} \Big(\bigwedge_{\vec{x} \in U_R} R\vec{x} \wedge \bigwedge_{\vec{x} \notin U_R} \neg R\vec{x}\Big)\Big)\Big)\Big)$$

In other words, this says that for any collection of n elements, and a specific way a new element might be related to them, there is such an element. Let T_{ext} be the set of all such extension axioms. Put together, all of the extension axioms have the same meaning as $(*')$. The property we will make use of for the proof of the 0-1 law is that T_{ext} is complete.

Proposition 3.7. *T_{ext} is complete.*

Proof. The countable random $\mathcal{L}$-structure is a model of T_{ext}, and since every model of T_{ext} satisfies $(*')$, it is the only countable model of T_{ext} up to isomorphism. T_{ext} must also not have any finite models, because for any finite set there is an extension axiom demanding that there exist an element outside that set. Thus, by the Łoś-Vaught test, T_{ext} is complete. □

With this in hand, we can finally connect this all back to asymptotic probability. The following lemma will do the majority of the work:

Lemma 3.8. *Every extension axiom is almost surely true.*

Proof. Consider the extension axiom $\psi_{n,U_{R_1},\ldots,U_{R_m}}$. Given any tuple of distinct elements $a_1, \ldots, a_n$, we add an additional element a, randomly deciding all of its atomic relations with the other elements. Let δ be the probability that $a_1, \ldots, a_n, a$

satisfy every element of every U_R and no atomic relations not in a U_R. In other words, this is the chance that a finite (although large) number of coin flips turn out the right way. So $\delta > 0$.

Now, let's calculate $l_k(\neg\psi_{n,U_{R_1},\dots,U_{R_m}})$. This is the chance that there exist distinct $a_1, \dots, a_n$ such that any new a we add will not be related in the right way. For any given choice of $a_1, \dots, a_n$, the chance that it is such a tuple is $(1-\delta)^{k-n}$, since this is the chance that every other element fails to be related to $a_1, \dots, a_n$ in exactly the right way. Taking the union bound over all choices of $a_1, \dots, a_n$, the total chance is bounded above by $\frac{k!}{n!}(1-\delta)^{k-n}$, which is bounded above by $k^n(1-\delta)^{k-n}$. Taking the limit as $k \to \infty$, this goes to 0, because the exponential $(1-\delta)^{k-n}$ decays much faster than the polynomial k^n grows. Therefore, $l(\neg\psi_{n,U_{R_1},\dots,U_{R_m}}) = 0$, so $l(\psi_{n,U_{R_1},\dots,U_{R_m}}) = 1$. □

Now we can finish the proof of the 0-1 law for first-order logic, restated below for convenience.

Theorem 3.2. *Let φ be a sentence of first-order logic over a finite relational vocabulary. Then φ is either almost surely true or almost surely false.*

Proof. Let φ be a sentence of first order logic with relations from $\mathcal{L}$. Since T_{ext} is complete, either $T_{ext} \models \varphi$ or $T_{ext} \models \neg\varphi$. Suppose the former is true. Then by completeness, $T_{ext} \vdash \varphi$. This means that φ is the consequence of some finite subset $T \subset T_{ext}$. So φ is true in every model in which $\bigwedge T$ is true. But $\bigwedge T$ is the conjunction of almost-surely true sentences, so it is almost surely true. Therefore, φ is almost surely true. Similarly, if $T_{ext} \models \neg\varphi$, then $\neg\varphi$ is almost surely true, so φ is almost surely false. □

An objection one might have to the 0-1 law, as presented, is that, in considering all finite models of a certain size, we are repeating a lot of the same structure. At some level, isomorphic models are essentially the same. By looking at all finite models, models which have isomorphic, differently labeled copies will have more say in deciding the asymptotic probability of a sentence. As is, l weights models which have more isomorphic copies as more important than models without them. In other words, we are weighting models based on how symmetric they are– how many automorphisms they have. To rectify this, we define $u_n(\varphi)$ as the fraction of isomorphism classes of models of size n which satisfy φ, and $u(\varphi)$ as $\lim_{n\to\infty} u_n(\varphi)$. This quantity is called the *unlabeled asymptotic probability* of φ.

It turns out that this coincides with the asymptotic probability of φ in first-order logic. Although when considering small graphs, we see a lot of graphs with plenty of automorphisms, which might be cause for worry, it turns out that on a larger scale, the vast majority of graphs have no nontrivial automorphisms. We call a graph with no nontrivial automorphisms *asymmetric.*

Theorem 3.9. *Let a_n be the number of asymmetric graphs of size n, and $\#_n$ be the number of graphs of size n. Then $\lim_{n\to\infty} \frac{a_n}{\#_n} = 1$.*

A proof of this fact can be found in [4]; it is combinatorial in nature. Suppose we have a relational language $\mathcal{L}$ with at least one non-unary relation. This guarantees that every $\mathcal{L}$-structure defines at least one graph: with k-ary R, define a, b adjacent if $Ra \dots ab$ holds. Any automorphism of our $\mathcal{L}$-structure must be an automorphism of this graph. It follows from this that as $n \to \infty$, the proportion of models with any nontrivial automorphisms goes to zero. So, when calculating (labeled) asymptotic probabilities, the share of the vote given to models with nontrivial automorphisms goes to zero, so the labeled and unlabeled asymptotic probabilities are the same. This is the unlabeled 0-1 law for first order logic.

Theorem 3.10. *(Unlabeled 0-1 law) Let $\mathcal{L}$ have at least one non-unary relation symbol. Then for any first order sentence φ, its unlabeled asymptotic probability is either 1 or 0.*

Proof. See [5] for the full proof. The idea is given in the above paragraph. □

There is an interesting divide between the finite and the infinite in this case. Most finite graphs are asymmetric, having no nontrivial automorphisms. But if you randomly pick a countable graph it will be extremely symmetric–proposition 3.4 constructs a symmetry of it. In fact, the proof of proposition 3.4 can be modified so that if you start with any two isomorphic finite graphs within the countable random graph, you can extend the isomorphism between them to an automorphism of the whole graph. Thus it is interesting that the proof of the unlabeled 0-1 law relies both on the asymmetry of most finite graphs and on the symmetry of the random graph.

It is worth comparing almost-sure validity to finite validity. The proof of the 0-1 law gives a complete and computable axiomatization for the set of almost-surely valid formulas. On the other hand, the set of formulas which are finitely valid is not decidable, a result known as Trakhtenbrot's theorem. For the proof of Trakhtenbrot's theorem, see [6]. All finitely-valid sentences are, of course, almost-surely valid, but by expanding the set of finitely-valid sentences to almost-surely valid ones, the set becomes computable.

4. The Frontier of 0-1 Laws

After seeing a result like the 0-1 law for first-order logic, a natural question arises: how far does this go? In any formal system with a notion of "finite model" and sentences which can be true or false in models, we can formulate the concept of the asymptotic probability of a sentence. A good path towards understanding whether 0-1 laws hold is seeing how fragile or durable they can be. If we want to keep our definition of models the same as in first-order-logic, there are two ways we can change a system to get new 0-1 law statements. First, we can restrict what kinds of models are allowed when we are calculating our asymptotic probability, making it into some sort of conditional asymptotic probability. Second, we can change what can be said by adding new sentences to our logic. Both ways have places where 0-1 laws still hold and where they do not. In this section, I give examples of where 0-1 laws remain and where they no longer hold.

4.1. **Restricting the class of finite models.** Most of the time, when we do abstract math, we do not care about arbitrary models, but rather models that have certain properties. So it is natural to ask whether 0-1 laws still hold when restricting the properties of the models under consideration.

We have already, in fact, mostly proven a result of this sort, barring a few details: the 0-1 law for graphs. It is important to note that (simple) graphs are not exactly the same thing as models–the adjacency relation is required to be symmetric and irreflexive. So taking a fraction of all graphs is different from taking a fraction of all arbitrary binary relations. For example, the asymptotic probability of the sentence $\forall x \neg Rxx$ is 1 when restricted to graphs, because this is always true in a graph.

To finish the proof of the 0-1 law for graphs, instead of dealing with the countable random $\mathcal{L}$-structure, we deal with the countable random graph, axiomatizing it via a set of (easier to state) extension axioms, and showing that truth in that model corresponds to almost sure truth or falsity. More or less the same proofs that are used in the case of arbitrary $\mathcal{L}$-structures apply to graphs. In general, for any type of object where we can construct a random version of it by independent coin flips, we can adapt the above arguments to prove a 0-1 law for it.

This is not true of more complicated objects, though. Consider the case of countable functions. It is still totally possible to construct a random function– just assign a probability distribution P to $\mathbb{N}$ and for each x, decide the value of $f(x)$ from there. Now we can ask a question like "Does f have a fixed point?" and calculate the probability of it being true. Well, the chance that any individual point n is fixed is $P(n)$, and since these flips are independent, the probability that any point is fixed is $\sum_{n=1}^{\infty} P(n)$, which is 1 since P is a probability distribution.

However, consider the finite case. What is the asymptotic probability that the sentence "f has a fixed point" is true in the language of functions? Well, for a given function on n elements, for each element, the probability that it is sent to itself is $\frac{1}{n}$, so the probability that f does not have any fixed points is is $1 - \frac{1}{n}$. So the total probability is $(1 - \frac{1}{n})^n$. Taking limit as $n \to \infty$, we find that this is $\frac{1}{e}$, which is decidedly not 0 or 1. So our notion of asymptotic probability does not coincide with truth for a random function.

In fact, a random function itself is not completely determined: even with a specific distribution, models created from it may not be isomorphic. Suppose our distribution P is geometric: $P(n) = \frac{1}{2^n}$. Consider the property "f has (at least) two fixed points." This is bounded above by the probability that 1 is a fixed point and something after is, plus the probability that 2 is a fixed point and something after is, and so forth, which is

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \left(\sum_{k=n+1}^{\infty} \frac{1}{2^k} \right) = \sum_{n=1}^{\infty} \left(\frac{1}{2^n} \right)^2 = \sum_{n=1}^{\infty} \frac{1}{4^n} = \frac{1}{3}$$

However, this probability is also bounded below by the chance that 1 is a fixed point and something after is. This is the first term in the series, which is $\frac{1}{4}$. In other words, the probability that a function chosen from this distribution has at least two fixed points is neither 0 nor 1; it is somewhere between $\frac{1}{4}$ and $\frac{1}{3}$. Since having at least two fixed points is a property preserved by isomorphism, there is no "random function" in the same way that there is a random graph, because when picking two random functions, there is a nonzero chance that they are non-isomorphic.

We can think of adding functions to our vocabulary as having a binary relation and restricting the models under consideration to those in which that relation is the graph of a function. Similarly, adding a constant to our vocabulary can be thought of as adding a unary relation and restricting the models under consideration to those in which that unary relation has a unique element that satisfies it. We have shown above that the introduction of functions kills the 0-1 law. It turns out that adding constants does too, and this is even easier to see.

Suppose we have a vocabulary with at least one unary relation symbol P and a constant symbol c. Consider the sentence Pc. P is a randomly chosen set, and c

is a randomly chosen element, so this will be true in exactly half of the models, no matter what the size of the models are. So this sentence has asymptotic probability $\frac{1}{2}$.

Still, it turns out that in both the case of functions and constants, a weakened form of the 0-1 law still holds. In the above cases, the asymptotic probability still did converge, although it did not converge to 0 or 1. It turns out that this happens so long as we are limited only to unary functions.

Theorem 4.1. *In a first-order language with only relations and unary functions, for every sentence φ, $\lim_{n\to\infty} l_n(\varphi)$ exists.*

The proof of this fact is too long and too challenging for the scope of this paper, although the interested reader should refer to [7] Unfortunately, the moment we allow even binary functions, the above result fails.

Theorem 4.2. *In a language with just a binary function symbol, there is a sentence φ whose asymptotic probability does not converge.*

The proof of this fact is also outside of the scope of this paper, though it can be found in [8]. So even in fairly simple extensions of relational first-order logic, the 0-1 law fails, and asymptotic probabilities can even fail to exist. So there is a jump in the power of first-order logic when we go from only allowing relations to allowing functions and constants: it gains the capability to say things that are not either almost surely true or almost surely false.

4.2. **Extending the language.** One common way to extend first-order logic is to move to second-order logic. In first-order logic, we can only quantify over individual elements. Because of this, it is impossible to define things such as a well-order, which requires the ability to quantify over sets. Second-order logic fixes this by allowing us to quantify over relations too. For example, to say "R is a well-order," the formula $\forall P\exists x\forall y(Px \wedge (Py \to (xRy)))$, conjoined with the order axioms, would suffice. The notion of a model of second-order logic stays the same, but the 0-1 law is affected because there are more sentences to consider.

In full second-order logic, the 0-1 law from first-order logic does not carry over. In fact, not all sentences need to have an asymptotic probability. This is because sentences in second-order logic can keep track of the size of the model in nontrivial ways.

Proposition 4.3. *The 0-1 law fails in second-order logic.*

Proof. Let φ be the sentence $\exists R\forall x\exists! y(y \neq x \wedge Rxy \wedge Ryx)$. This essentially says that there is a relation which matches each element of the domain with a unique other element. This can be done if and only if the size of the model is even. So $u_n(\varphi)$ alternates eternally between 0 and 1, and hence never converges. $\square$

Therefore, full second-order logic is too strong to maintain the 0-1 law. However, there are plenty of interesting fragments of second-order logic to consider. *Monadic second-order logic* is the fragment of second-order logic where only unary relations may be quantified over. *Existential second-order logic* only allows existential quantification over relations. And *existential monadic second-order logic* only has existential quantification over unary relations. Note that the above counterexample also shows that the 0-1 law fails for existential second order logic, since its only second-order quantifier is existential. The following counterexample is adapted from [9], and shows that the 0-1 law fails for existential monadic second-order logic, and thus also for monadic second-order logic.

Proposition 4.4. *The 0-1 law fails for existential monadic second-order logic.*

Proof. The majority of the work in this proof is done by the following lemma from [9].

Lemma 4.5. *There is a first-order formula $\varphi(x, y)$ in a vocabulary with a sequence of unary relation symbols $\vec{P}$ such that $\exists \vec{P}$ "$\varphi(x, y)$ defines a linear order" has asymptotic probability 1.*

With this lemma in hand, the proof becomes a lot easier. The key here is that we are going to make another sentence which captures the parity of the model, using this linear order. The sentence is as follows:

$$\begin{aligned} &\exists \vec{P} \exists Q \text{“}\varphi(x, y) \text{ is a linear order”} \wedge \\ &\text{“not both the first and last element are in } Q\text{”} \wedge \\ &\forall x \forall y \varphi(x, y) \to (Qx \leftrightarrow \neg Qy) \end{aligned} \tag{4.6}$$

This says that Q is a relation which targets every other element of our linear order, and does not contain both the first and last elements. Thus, this sentence is true precisely when our model has even size, so long as φ is actually a linear order. But asymptotically, φ does become a linear order, so the asymptotic probability of the sentence for even models approaches 1, and for odd models approaches 0.

□

A quick corollary of this is that the 0-1 law fails for universal second-order logic as well. This is because the negation of a purely existential sentence is a purely universal sentence–you can shuffle the negation past all of the quantifiers, flipping them. So the negation of the above counterexample also does not converge.

We have found that in a large number of restrictions of second-order logic, the 0-1 law fails. It turns out, however, that there are interesting classes of second-order logic for which the 0-1 law does hold. To state these classes, rather than just restricting the second-order quantifiers, we also restrict the first-order part. A *prefix class* is a set of first-order sentences, in prenex normal form, defined by their quantifiers. Two prefix classes of note are the *Bernays-Schönfinkel class* and the *Ackermann class*. The former consists of all sentences of the form $\exists^*\forall^*\varphi$, that is, with as many existential quantifiers as necessary followed by arbitrarily many universal quantifiers. The Ackermann class is all sentences of the form $\exists^*\forall\exists^*\varphi$, consisting of one universal quantifier sandwiched between existentials.

For a prefix class $\mathcal{C}$, $\Sigma^1_1(\mathcal{C})$ is the set of sentences consisting solely of elements of that prefix class with existential second-order quantifiers. Σ^1_1 on its own is just existential second-order logic. Similarly, $\Pi^1_1(\mathcal{C})$ and Π^1_1 represent universal relation quantifications over sentences from $\mathcal{C}$ or any formula, respectively. It turns out that for both of the above classes, with existential relation quantifiers appended, the 0-1 law still holds. This was first proven in [10] by Kolaitis and Vardi.

Proposition 4.7. *The 0-1 law holds for Σ^1_1(Bernays-Schönfinkel).*

Proof. Again, we will show that almost sure validity corresponds to truth or falsity in the countable random $\mathcal{L}$-structure. First, for one direction, we show the following:

Lemma 4.8. *Suppose φ is Π_1^1 and true in the countable random $\mathcal{L}$-structure. Then there is a ψ such that $l(\psi) = 1$ and $\psi \to \varphi$ is valid. In particular, $l(\varphi) = 1$.*

Proof. Let $\varphi = \forall \vec{R}\theta(\vec{R})$, where θ contains no second-order quantifiers. Let T_{ext} be the set of extension axioms for our language, $\mathcal{L}$, and let $\mathcal{L}'$ be $\mathcal{L} \cup \vec{R}$. Let $T = T_{ext} \cup \{\neg\theta\vec{R}\}$. Now, suppose that there is no ψ such that $l(\psi) = 1$ and $\psi \to \varphi$ is valid. Then for each sentence which is almost surely true, there is a model of it in which that sentence is true and φ is false.

In particular, this means that any finite subset of T has an $\mathcal{L}'$-model. By compactness, T has a model, A. Now, the reduct of A to $\mathcal{L}$ satisfies all the extension axioms, so it is isomorphic to the random graph. However, by construction there is a choice of $\vec{R}$ such that $\theta(\vec{R})$ is false, so φ is not true in A. Therefore, φ is false in the random $\mathcal{L}$-structure. By contraposition, if φ is true in the countable random $\mathcal{L}$-structure, then such a ψ exists. □

This directly implies that if there is a Σ_1^1 sentence which is false in the countable random graph, its asymptotic probability is 0 (because the asymptotic probability of its negation is 1). Note that for this direction it does not matter which prefix class we are in.

Now, suppose φ is a Σ_1^1(*Bernays-Schönfinkel*) sentence which is true in the countable random graph. We need to show that its asymptotic probability is 1. We can then write it as $\exists\vec{R}\exists\vec{x}\forall\vec{y}\theta(\vec{R},\vec{x},\vec{y})$. Now, let $\vec{a} = a_1, \ldots, a_n$ be elements of the countable random $\mathcal{L}$-structure such that $\exists\vec{R}\forall\vec{y}\theta(\vec{R},\vec{a},\vec{y})$ is true. Let A be the finite substructure of the countable random $\mathcal{L}$-structure with domain equal to these elements. Since A is finite, we can describe the property "this model has A as a substructure" via a sentence in first order logic, which states that there exist unique elements $x_1, \ldots, x_n$ which relate to each other in the same way that $a_1, \ldots, a_n$ do. This sentence is true in the countable random $\mathcal{L}$-structure, so it is a consequence of finitely many extension axioms.

Let ψ be the conjunction of these finitely many extension axioms. Since it is the conjunction of almost surely true sentences, it is almost surely true. Hence, there is a finite model B in which ψ is true. Note that B has an isomorphic copy of A as a substructure. Now, within the countable random $\mathcal{L}$-structure, we can use the extension property to extend A to an isomorphic copy of B, call it B'. Let $P_1, \ldots, P_n$ be the relations which witness the second-order existential quantifiers in φ. Now, $\forall\vec{y}\theta(\vec{P},\vec{a},\vec{y})$ is a universal statement that is true in the countable random $\mathcal{L}$-structure, so it is true in any substructure, so it is true when restricted to B'. Therefore, $\exists\vec{R}\exists\vec{x}\forall\vec{y}\theta(\vec{R},\vec{x},\vec{y})$ is true in B', and thus true in B.

Thus in all models of B, φ is true. So $\psi \to \varphi$. As ψ has asymptotic probability 1, it follows that φ does too. □

Proposition 4.9. *The 0-1 law holds for Σ_1^1(Ackermann).*

The proof of this fact is too involved for the purposes of this paper, though it can be found in [11]

A prefix class is called *solvable* if the task of determining whether one of its members is satisfiable is decidable. It is known that the only solvable prefix classes in first order logic with identity are the Bernay-Schönfinkel class and the Ackermann class. Kolaitis and Vardi conjectured that a 0-1 law holds for $\Sigma_1^1(\mathcal{C})$ precisely when the prefix class is solvable. The proof of this fact was finished by Pacholski and Szwast in [13].

Theorem 4.10. *The 0-1 law holds for $\Sigma_1^1(\mathcal{C})$ if and only if the prefix class $\mathcal{C}$ is solvable.*

Naturally, it is interesting to ask whether there might be a reason for this correspondence. To prove that both prefix classes are solvable, it was used that they are both *finitely controllable.* A class of sentences is *finitely controllable* if for each of its elements, it is either unsatisfiable or finitely satisfiable. From finite controllability, we can computably enumerate the satisfiable members of a class by listing all finite models, and we can computably enumerate the unsatisfiable members of a class by the completeness theorem. So satisfiability in a finitely controllable class is decidable.

The following observation, made by Kolaitis and Vardi in [12], connects finite controllability of a class to asymptotic probability.

Proposition 4.11. *Suppose that for every $\varphi \in \Sigma_1^1(\mathcal{C})$, φ is true in the countable random structure if and only if it is almost surely true. Then $\mathcal{C}$ is finitely-controllable.*

Proof. Let θ be a sentence in a language with $\vec{R}$ as the relations, and suppose it is satisfiable. We need to show that it is finitely satisfiable. Well, by turning all of the relation symbols in θ into variables, we get that $\exists\vec{R}\theta(\vec{R})$ is satisfiable. Let A be a model of it–if it is finite, we are good, and if it is infinite, we can assume it is countable by downwards-Löwenheim-Skolem-Tarski. Now, since A is a countable structure, it is embeddable in the countable random structure. This means that the sentence $\exists P\exists\vec{R}\theta(\vec{R})^P$ is true in the countable random structure, where $\theta(\vec{R})^P$ is $\theta(\vec{R})$ with all quantifiers relativized to P, i.e. $\forall x$ replaced with $\forall x, Px \to$ and the same with $\exists$. This is true because the copy of A in the random structure is a suitable P. Now, by hypothesis, this means that $\exists P\exists\vec{R}\theta(\vec{R})^P$ is almost surely true. Thus, $\exists P\exists\vec{R}\theta(\vec{R})^P$ has a finite model. Now, within this finite model, take as a submodel the elements picked out by P. This will be a finite model of $\exists\vec{R}\theta(\vec{R})$, and thus a finite model of θ when the relations are chosen the right way. ☐

This gives at least a partial explanation for the correspondence between the solvability of prefix classes and 0-1 laws. Fleshing out the connection in greater depth is still an interesting open question.

Acknowledgements. I would like to thank my mentor, Sarah Reitzes, for her help on this project throughout the summer. I would also like to thank Peter May for organizing the REU.

References

[1] R. Fagin. *Probabilities on Finite Models.* Journal of Symbolic Logic 41 (1), 1976.
[2] C. C. Chang and H. J. Keisler. *Model Theory.* North-Holland Publishing Co., Amsterdam, Third edition. 1990.
[3] K. Gödel. *Über die Vollständigkeit des Logikkalküls.* Doctoral dissertation, University Of Vienna. 1929.
[4] P. Erdös and A. Rényi. *Asymmetric Graphs.* Acta Mathematica Academiae Scientiarum Hungaricae 14. 1963.
[5] H. Ebbinghaus and J. Flum. *Finite Model Theory.* Springer, New York. 1995.

[6] B. Trakhtenbrot. *The Impossibility of an Algorithm for the Decidability Problem on Finite Classes*. Proceedings of the USSR Academy of Sciences 70 (4). 1950.

[7] J. Lynch. *Probabilities of first-order sentences about unary functions*. Transactions of the American Mathematical Society 287. 1985.

[8] K. J. Compton, C. W. Henson, and S. Shelah. *Nonconvergence, undecidability, and intractability in asymptotic problems*. Annals of Pure and Applied Logic 36. 1987.

[9] M. Kaufmann. *Counterexample to the 0-1 law for existential monadic second-order logic*. CLI Internal Note 31, Computational Logic Inc. 1987.

[10] P. G. Kolaitis and M. Y. Vardi. *The decision problem for the probabilities of higher order properties*. Proceedings of the 19th ACM Symposium on the Theory of Computing. 1987.

[11] P. G. Kolaitis and M. Y. Vardi. *0-1 laws and decision problems for fragments of second-order logic*. Proceedings of the 3rd IEEE Symposium on Logic in Computer Science. 1988.

[12] P. G. Kolaitis and M. Y. Vardi. *0-1 laws for fragments of second-order logic*. Research Report RJ 7508, IBM. 1990.

[13] L. Pacholski and W. Szwast. *The 0-1 law fails for the class of existential second-order Gödel sentences with equality*. Proceedings of the 30th IEEE Symposium on Foundations of Computer Science. 1989.