**RESEARCH STATEMENT**

ALEX KRUCKMAN

Introduction. My research lies at the intersection of logic, combinatorics, and probability. I am interested in the relationships between properties of random finite structures and the model theory of their infinitary limits. What do I mean by “properties” and “infinitary limits”? It turns out that there are a number of useful ways of associating limit objects to classes of finite structures, which I will describe below, and I am also interested in the relationships between these limits. When I say “properties”, I mean something more precise: those properties which are expressible by sentences of first-order logic.

For those unfamiliar with logical terminology, formulas of first-order logic are built up from a vocabulary of basic symbols. My vocabulary will always consist of at most countably many relation symbols (e.g. the edge relation of a graph). These basic relations, applied to variables, can be combined by Boolean operations ($\neg$, $\land$, $\lor$) and quantification ($\forall x$, $\exists x$) to form formulas. A sentence is a formula in which all the variables have been quantified. A structure $A$ is a set, given with interpretations of the symbols in the vocabulary (e.g. a graph). The notation $A \models \varphi$, read “$A$ is a model of $\varphi$”, means that the sentence $\varphi$ is true of the structure $A$. A (complete) theory is a maximally consistent set of sentences.

Let me begin with two examples, illustrating the motivating phenomena. It is well-known that the class $\mathcal{G}$ of finite graphs has a first-order zero-one law [GKLT69,Fag76]. That is, letting $\mathcal{G}(n)$ be the set of graphs with domain $[n] = \{1, \ldots, n\}$ and $\mu_n$ the uniform measure on $\mathcal{G}(n)$, we have that for every sentence $\varphi$ of first-order logic,

$$\lim_{n \to \infty} \mu_n(\{G \in \mathcal{G}(n) | G \models \varphi\}) = 0 \text{ or } 1.$$  

The set of all sentences $\varphi$ for which the value of the above limit is 1 is thus a complete theory, the almost-sure theory of $\mathcal{G}$. This theory is countably categorical, meaning that it has a unique countable model up to isomorphism. This model is the Rado graph $\mathcal{R}$, a countably infinite graph which reflects all of the properties which are almost-surely true of large finite graphs.

$\mathcal{R}$ can be constructed as the Fraïssé limit of $\mathcal{G}$: it is universal, in the sense that it contains an embedded copy of every finite graph, and homogeneous, in the sense that if two finite subgraphs $A$ and $B$ of $\mathcal{R}$ are isomorphic, then there is an automorphism of $\mathcal{R}$ moving $A$ to $B$. Indeed, any class of finite structures which satisfies the hereditary property, joint embedding property, and amalgamation property has a universal homogeneous Fraïssé limit (see [Hod93, Ch. 7]). We call such a class $\mathcal{K}$ a Fraïssé class and call the theory its Fraïssé limit the generic theory of $\mathcal{K}$.

Alternatively, $\mathcal{R}$ has a probabilistic construction: If we define a graph with underlying set $\mathbb{N}$ by assigning each possible edge independently with probability $1/2$, then the resulting graph is isomorphic to $\mathcal{R}$ with probability 1. Note that this construction is highly symmetric: the probability of seeing a given induced subgraph on a set of vertices $A \subseteq \mathbb{N}$ does not depend on the choice of $A$.

Now let’s consider $\mathcal{G}_\triangle$, the class of triangle-free graphs. $\mathcal{G}_\triangle$ is a Fraïssé class, with Fraïssé limit $\mathcal{H}$, the Henson graph [Hen71]. $\mathcal{G}_\triangle$ also has a first-order zero-one law, but its almost-sure theory diverges from its generic theory. Indeed, almost all large finite triangle-free graphs are bipartite [EKR76,KPR87] and hence do not contain any odd-length cycles, while $\mathcal{H}$ contains embedded copies of all triangle-free graphs. The first symmetric probabilistic construction of the Henson graph was given only recently by Petrov and Vershik [PV10].

A theory $T$ is pseudofinite if every sentence $\varphi \in T$ has a finite model. The fact that the theory $T_\mathcal{R}$ of $\mathcal{R}$ occurs as the almost-sure theory of $\mathcal{G}$ shows that $T_\mathcal{R}$ is pseudofinite. In fact, for each sentence $\varphi \in T_\mathcal{R}$, most large finite graphs satisfy $\varphi$. But this probabilistic argument fails for triangle-free graphs, and the pseudofiniteness of the generic theory of $\mathcal{G}_\triangle$ is a notorious open problem (see [Che92,Che11]).

My research follows two main threads. The first is the question of pseudofiniteness of theories of Fraïssé limits, and the coincidence or divergence of generic theories and almost-sure theories. The second is a model-theoretic analysis of symmetric probabilistic constructions. There are analogies between these two settings, which can be viewed as instances of a general measure/category analogy, and in future work (described below) I plan to connect the two threads.
Pseudofinite Fra"issé limits. There are, essentially, two ways to show that a theory $T$ is pseudofinite. One can explicitly construct finite structures which satisfy the sentences in $T$, or one can employ a probabilistic argument by specifying a sequence of probability measures $\mu_n$ on classes $K(n)$ of finite structures, such that each $\varphi \in T$ is satisfied with positive probability for large enough $n$. The first method has the advantage of being more explicit, and the constructions may be of combinatorial interest. But the second method tells us something more: the sentences of $T$ actually have many finite models as measured by the $\mu_n$. I have identified a combinatorial sufficient condition, disjoint $n$-amalgamation, for a countably categorical theory to be pseudofinite, witnessed by a natural sequence of probability measures $\mu_n$ (see [Krua]).

A type $p(\pi)$ in the variables $\pi$ over a set of parameters $A$ relative to a theory $T$ is the set of all first-order formulas with parameters from $A$ satisfied by some tuple of elements from a model of $T$. If $T$ has quantifier elimination (as in the case that it is the generic theory of a Fra"issé class), then a type is determined by its quantifier-free formulas, i.e. the isomorphism type of the induced substructure on the tuple. A type is nonredundant if it does not contain the formulas $x_i = x_j$ when $i \neq j$ or $x_i = a$ when $a \in A$.

A theory has disjoint $n$-amalgamation if whenever we have tuples of variables $\pi_1, \ldots, \pi_n$ and a system of types $\{p_S : S \subseteq [n]\}$ over $A$ such that $p_S$ is a non-redundant type in the variables $\{\pi_i : i \in S\}$, and $p_S \subseteq p_T$ when $S \subseteq T$, then there is some non-redundant type $p_{\{\}}$ in the variables $\{\pi_i : i \in [n]\}$ extending the $p_S$.

Disjoint 2-amalgamation is a property enjoyed by many combinatorial examples of countably categorical theories. A countably categorical theory $T$ has disjoint 2-amalgamation if and only if it has trivial algebraic closure: For all finite sets $A \subseteq M \models T$, $acl(A) = A$, in the sense that no point $b \in M \setminus A$ has a finite orbit under the group of automorphisms fixing $A$.

**Theorem 1** ([Krua]). If a countably categorical theory $T$ has disjoint $n$-amalgamation for all $n$, then $T$ is pseudofinite.

The measures $\mu_n$ witnessing pseudofiniteness describe random structures on with domain $[n]$ “built from below”: For each $i \in [n]$, we choose a 1-type uniformly at random, then for every pair $\{i, j\}$, we choose a 2-type uniformly at random extending the given 1-types, etc. Disjoint $n$-amalgamation ensures that such an extension always exists and that we are able to make all random choices as independently as possible.

Variations on the theme of $n$-amalgamation are important in a variety of model-theoretic contexts, including simple theories [CH99,Kol05,KKT08], definable groupoids [Hru12,GKK15], and abstract elementary classes [Bal09], and I believe that they are an important piece of the puzzle of pseudofinite Fra"issé limits. Indeed, almost all examples of countably categorical theories with trivial algebraic closure which are known to be pseudofinite either have disjoint $n$-amalgamation for all $n$, or are reducts of countably categorical theories with disjoint $n$-amalgamation for all $n$ (a reduct of a structure or a theory is obtained by forgetting some of the symbols in the vocabulary). The only exceptions (known at the time of this writing) are built from equivalence relations.

For example, I have considered two case studies of complicated generic theories of equivalence relations. $T_{eq}$, introduced by Shelah [She93], is the generic theory of the class of parameterized equivalence relations: the vocabulary consists of a sort of objects, a sort of parameters, and a ternary relation $E_a(y, z)$, where for each parameter $a$, $E_a$ is an equivalence relation on the objects. $T_{CPZ}$, introduced by Casanovas, Peláez, and Ziegler [CPZ11] is the generic theory of structures in the vocabulary with infinitely many relations $E_n$, where $E_n$ is an equivalence relation on the $n$-tuples, and all redundant tuples are in the same class. Neither of the theories is a reduct of a theory with disjoint $n$-amalgamation for all $n$, but both can be expressed as limits of such reducts in an appropriate sense. This implies that they are both pseudofinite, and, moreover, the pseudofiniteness can be explained via disjoint $n$-amalgamation.

Much of model theory since the advent of Shelah’s classification theory [She90] is concerned with the combinatorics of definable sets (subsets of models definable by formulas), and many dividing lines (stability, simplicity, NIP, etc.) in the complexity of theories have been established on this basis. Let me take a moment to discuss the relationship between the ideas above and these dividing lines.

**Theorem 2** ([Krua]). If a theory $T$ is a reduct of a countably categorical theory with disjoint 2- and 3-amalgamation, then $T$ is simple.

Kim and Pillay [KP98] conjectured that all countably categorical pseudofinite theories are simple. The theories $T_{eq}$, $T_{CPZ}$ are counterexamples to this conjecture. These theories are not simple, but they are NSOP$_1$ (they do Not have the 1-Strong Order Property). The property SOP$_1$, introduced in [DS04], is the first in a hierarchy of dividing lines SOP$_n$, lying between simplicity and the strict order property SOP.
Theorem 3 (Folklore). No countably categorical pseudofinite theory has the strict order property.

However, very little is known about the pseudofiniteness of countably categorical theories in the region between SOP$_1$ and SOP. As an update to the conjecture of Kim and Pillay, and a vast strengthening of Theorem 3, I suggest the following conjecture:

**Conjecture 4.** Every pseudofinite countably categorical theory is NSOP$_1$.

Chernikov and Ramsey have given an independence relation criterion for NSOP$_1$ [CR15], which, together with refinements of Hrushovski’s pseudofinite dimensions [Hru13] may be a promising lead toward Conjecture 4.

Conjecture 4 embraces the equivalence relation examples, but another option is to throw them out, as in the following definition. The motivation is to try to isolate a class of theories for which pseudofiniteness can always be explained by Theorem 1.

**Definition 5.** A primitive combinatorial theory (PCT) is a countably categorical theory with trivial algebraic closure (disjoint 2-amalgamation), such that for every finite set $A$ and every complete type $p$ over $A$, every $A$-definable equivalence relation on realizations of $p$ is $\emptyset$-definable in the empty vocabulary.

**Conjecture 6.** A PCT is pseudofinite if and only if it is a reduct of a countably categorical theory with $n$-amalgamation for all $n$.

The triviality of definable equivalence relations only needs to be checked on 1-types [Krub]. This makes it easy to verify that many natural generic theories of Fraïssé classes are PCTs.

**Theorem 7 ([Krub]).** The theory $T_H$ of the Henson graph is a PCT. Further, $T_H$ is not the reduct of any countably categorical theory with $n$-amalgamation for all $n$.

The primitive combinatorial theories are of independent model-theoretic interest, since the definition is designed to capture a general notion of “purely unstable” theory. Stable theories are the most well-behaved from a model-theoretic point of view, but examples tend to be algebraic (think of vector spaces or algebraically closed fields) rather than combinatorial. Indeed, the theory of an infinite dimensional vector space over a finite field is pseudofinite, but not for a good probabilistic reason: the finite models of sentences in the theory only occur in certain cardinalities, and must have a very rigid structure. But PCTs have no non-trivial stable behavior.

**Theorem 8 ([Krub]).** In a PCT, any stable formula $\varphi(x, \overline{y})$ is equivalent to a boolean combination of formulas isolating complete 1-types over $\emptyset$ and instances of equality. Any stable PCT is interdefinable with a theory of finitely many infinite partitioning unary predicates.

In [Krub], I investigate the interaction of PCTs with model-theoretic dividing lines. For example, I characterize when PCTs are simple with trivial forking and when they are distal. Some questions remain in this direction. The following conjecture is the most important - replacing “SOP$_3$” with “not simple” is equivalent to the stable forking conjecture for PCTs.

**Conjecture 9.** If a PCT fails to have disjoint 3-amalgamation, then it is SOP$_3$.

**Symmetric probabilistic constructions.** I will now describe a new kind of infinitary limit of finite structures. Given any finite structure $A$, we can assign a value $P(\varphi; A) \in [0, 1]$ to every quantifier-free formula $\varphi(x_1, \ldots, x_n)$, namely the probability that $n$ elements sampled uniformly and independently from $A$ satisfy $\varphi$. We say that a sequence of finite structures $\langle A_n \rangle$ converges if the sequence $\langle P(\varphi; A_n) \rangle$ converges for all quantifier-free formulas $\varphi$.

This is very different from convergence in the sense of zero-one laws. Instead of evaluating the truth of sentences with quantifiers globally on structures, we are evaluating quantifier-free formulas locally on elements of structures. The information captured by a convergent sequence amounts to a finitely additive probability measure on the Boolean algebra of quantifier-free formulas in countably many variables $\{x_i \mid i \in \mathbb{N}\}$. This measure extends in a natural and unique way to a Borel probability measure on the space $\text{Str}_L$ of L-structures with domain $\mathbb{N}$. Moreover, the resulting measure is invariant under the “logic action” of $S_\infty$, the symmetric group on $\mathbb{N}$. From now on I will simply write invariant measure to refer to this sort of probability measure.

That there are other useful and equivalent ways to package this information: the exchangeable arrays of Aldous Hoover, and Kallenberg [Ald81, Hoo79, Kal05], the flag algebras of Razborov [Raz07], the graphons
of Lovasz and Szegedy in the case of graphs [Lov12] and their generalizations by Austin [Aus08], Nesetril and Ossana de Mendez [NoM12], and others. But the invariant measures framework allows us to think of a sequence of finite structures as converging to an infinite random structure with domain \( \mathbb{N} \), i.e. the sort of symmetric probabilistic construction alluded to in the introduction.

We say that an invariant measure concentrates on a structure \( M \) if the isomorphism class of \( M \) has measure 1. The method due to Petrov and Vershik [PV10] of giving a symmetric probabilistic construction of the Henson graph was extended dramatically by Ackerman, Freer, and Patel to any structure with trivial algebraic closure.

**Theorem 10** ([AFP12]). There is an invariant measure concentrating on an \( L \)-structure \( M \) if and only if \( M \) has trivial algebraic closure.

In fact, the measures constructed in the above theorem are all ergodic, meaning that any almost-surely invariant subset of \( \text{Str}_L \) has measure 0 or 1. In particular, an ergodic measure gives the set of models of any sentence, even of the infinitary logic \( L_{\omega_1,\omega} \), measure 0 or 1, and it makes sense to talk about the complete theory of the invariant measure, consisting of those sentences which receive measure 1.

While the familiar examples of invariant measures concentrate on countable structures (e.g. the Rado graph), there are also simple examples of ergodic invariant measures which fail to concentrate on any particular countable structure [AFNP13]. We call such measures properly ergodic. In joint work with Ackerman, Freer, and Patel, we characterized the behavior of properly ergodic invariant measures.

**Theorem 11** ([AFKP]). Let \( \varphi \) be a sentence of \( L_{\omega_1,\omega} \). The following are equivalent:

1. For all countable fragments \( F \) of \( L_{\omega_1,\omega} \) and all complete \( F \)-theories \( T \) containing \( \varphi \) which have trivial definable closure, there are only countably many \( F \)-types consistent with \( T \).
2. Every ergodic measure \( \mu \) which concentrates on \( \varphi \) concentrates on a single model of \( \varphi \), i.e., is not properly ergodic.

The proof goes through a measured version of Morley’s analysis of sentences of \( L_{\omega_1,\omega} \) [Mor70], and it also yields the following theorem.

**Theorem 12** ([AFKP]). If \( \mu \) is a properly ergodic measure, then \( \text{Th}_{L_{\omega_1,\omega}}(\mu) \) has no models of any cardinality. However, if \( F \) is any countable fragment of \( L_{\omega_1,\omega} \), then \( \text{Th}_F(\mu) \) has continuum-many countable models.

There is a general measure/category analogy between the theory of invariant measures on \( \text{Str}_L \) on one hand and Fraïssé theory on the other. A class \( K \) of finite structures with the joint embedding property is the analogue of an ergodic invariant measure: every isomorphism-invariant Borel subset of the \( K \)-direct limits is either meager or comeager. A generalization of the usual amalgamation property, called weak amalgamation, is equivalent to the existence of a comeager isomorphism class, and hence corresponds to a measure which concentrates on a single structure up to isomorphism. I explore this analogy in notes on generalized Fraïssé theory, where I prove the following analogue of Theorem 12.

**Theorem 13** ([Kruc]). If \( K \) is class of finite structure which has the joint embedding property but does not have the weak amalgamation property, then the generic \( L_{\omega_1,\omega} \)-theory \( \text{Th}_{L_{\omega_1,\omega}}(K) \) has no models of any cardinality. However, if \( F \) is any countable fragment of \( L_{\omega_1,\omega} \), then the generic \( F \)-theory \( \text{Th}_F(K) \) has continuum-many countable models.

As the theory of symmetric probabilistic constructions develops, this analogy deserves to be studied further, and I plan to pursue this direction in future work. I will end with a connection to the pseudofiniteness questions in the previous section.

Any invariant measure \( \mu \) induces a sequence of measures \( \mu_n \) on the spaces \( \text{Str}_{L,n} \) of \( L \)-structures with domain \( [n] = \{1, \ldots, n\} \), simply by taking the random substructure on \( [n] \) induced by the random structure on \( \mathbb{N} \). Conversely, these measure \( \mu_n \) cohere and induce a unique invariant measure on \( \text{Str}_L \). Coherence in this sense is a reasonable “naturality” condition on the sequence of measures \( \mu_n \). For example, the “build from below” measures in the proof of Theorem 1 cohere.

**Question 14.** Characterize those invariant measures \( \mu \) concentrating on the Fraïssé limit of a class \( K \) which induce a sequence \( \mu_n \) of measures which witness the pseudofiniteness of the generic theory \( T_K \), in the sense that for all \( \varphi \in T_K \), \( \lim_{n \to \infty} \mu_n(\{ A \in \text{Str}_{L,n} \mid A \models \varphi \}) = 1 \). Which pseudofinite Fraïssé limits admit such an invariant measure?