Probability and the Finite Model Property

John T. Baldwin *
Department of Mathematics, Statistics and Computer Science
University of Illinois at Chicago

May 2, 2000

Abstract

Theorem: Each of the $\aleph_0$-categorical stable theories constructed by Hrushovski can be viewed a countable model of an almost sure theory of finite structures. Thus, each theory has the finite model property.

1 Introduction

A complete first order theory T has the *finite model property* if every sentence $\phi \in T$ has a finite model. Clearly, a theory with the finite model property is not finitely axiomatizable. It is easy to see that in general the converse fails (take a disjoint union of a dense linear order with a non-finitely axiomatizable theory). Hrushovski (see the exposition in [5].) constructed a family T_α of stable $\aleph_0$-categorical theories parameterized by a suitable irrational α with $0 < \alpha < 1$. Since these theories are 'continuous' in α, it is straightforward to show they cannot be finitely axiomatizable. But such an argument does not suffice to show the T_α have the finite model property. However, if a theory T is achieved as an almost sure theory for a collection of finite models then T does have the finite model property. More precisely, suppose there is a class $\mathbf{K}_0$ of finite structures and a probability measure P_n on $\mathbf{K}_n$ (the members of $\mathbf{K}_0$ whose universe has cardinality n with $n > 0$). Suppose further that T is the collection of sentences ϕ such that $\lim_{n\to\infty} P_n(\phi) = 1$. Then T has the finite model property. For, if ϕ in T, $\lim_{n\to\infty} P_n(\phi) = 1$ so for sufficiently large n, $P_n(\phi) > 0$ and so ϕ has a finite model.

This paper requires familiarity with both the Hrushovski construction as described in [5] and the account of the Shelah-Spencer 0-1 law as presented in [2]. The question of whether there are continuum many $\aleph_0$-categorical theories with the finite model property was specifically raised in [4]; since the argument

*Partially supported by NSF grant DMS-9803496. primary classification: 03C45 secondary 03C30, 60C05

here is independent of α, we have a positive solution. I would like to acknowledge extremely helpful conversations on this paper with Jim Lynch, John Wood, and Alan Woods. This paper also provides a 'new' proof of the Shelah-Spencer theorem, by taking r_n below as 1. Our proof of Theorem 1.9 gives an alternative to the use of the second moment method in the original argument.

The theory T_α was given as the theory of a generic (or homogeneous universal in an appropriate sense) structure for a class $\mathbf{K}_0 = \mathbf{K}_\alpha$ of finite structures. Fix an irrational α. For any finite graph G, let $\delta(G) = \delta_\alpha(G) = |G| - \alpha e(G)$, (where $e(G)$ is number of edges of the graph G). For each α define ℓ_n/m_n to be the best rational approximation of α which is less than α with $m_n \leq n$. A sequence of ordered pairs $\mathbf{b}_i = \langle b_i^0, b_i^1 \rangle$ for $i < \omega$ is defined from these rational approximations. Then $\mathbf{K}_\alpha$ is the collection of finite graphs G such that for every $H \subset G$, $\delta(H)$ is above the polygonal line formed by connecting $\mathbf{b}_i$ to $\mathbf{b}_{i+1}$. If there is no bound on the size of the y-coordinates we say α has *infinite index.* The Baire category theorem guarantees that there are α with infinite index. The choice of the b_i guarantees that $\mathbf{K}_\alpha$ has the amalgamation property (with 'no-edges' amalgamation); the theory T_α of the resulting generic is $\aleph_0$-categorical if and only if α has infinite index. (The amalgamation argument is given in [5]; we do not rely on any details of the argument here.) The case where α does not have infinite index is treated in [3]. The new point there is the proof that even in this case the theory remains stable.

By representing T_α as an almost sure theory we will show

1.1 Theorem. *Each T_α has the finite model property.*

We give now the general argument which works for each irrational α. In fact, for the $\aleph_0$-categorical examples singled out by Hrushovsky, (i.e. when α has infinite index) the proof can be significantly simplified and we will sketch that simplification at the end.

We write $A \leq B$ if for every B' with $A \subseteq B' \subseteq B$,

$$\delta(B'/A) = \delta(B') - \delta(A) \geq 0.$$

Let $\mathbf{J}_n$ be the collection of isomorphism types of symmetric reflexive binary relational structures (graphs) with universe n. First consider the standard probability measure $\hat{P}_n$ on $\mathbf{J}_n$ with edge probability $n^{-\alpha}$:

$$\hat{P}_n(G) = n^{-\alpha e} \cdot (1 - n^{-\alpha})^{(\binom{n}{2} - e)}$$

for each $G \in \mathbf{J}_n$ where e is the number of edges of G. Now we relativize this measure to $\mathbf{K}_n$, the members of $\mathbf{K}_\alpha$ with cardinality n. Thinking of $\mathbf{K}_n$ as a subset of $\mathbf{J}_n$, set $r_n = \hat{P}_n(\mathbf{K}_n)$. Define a probability measure on $\mathbf{K}_n$ by setting

$$P_n(G) = \frac{\hat{P}_n(G)}{r_n}.$$

We will show that for this measure, for each ϕ, $\lim_{n\to\infty} P_n(\phi) = 1$ or $\lim_{n\to\infty} P_n(\phi) = 0$. For this, we show as in [2] that the semigeneric models for $\mathbf{K}_\alpha$ are elementarily equivalent and that T_α is finitely determined in the sense of [1]. Let $\mathbf{K}$ (which of course depends implicitly on α) denote the universal class of models whose finite substructures are $\mathbf{K}_0 = \mathbf{K}_\alpha$.

1.2 Definition. The countable model M is $(\mathbf{K}_\alpha, \leq)$-*semigeneric*, or just semigeneric, if

1. $M \in \mathbf{K}$
2. If $A \leq B \in \mathbf{K}_0$ and $g : A \mapsto M$, then for each finite m there exists an embedding $\hat{g}$ of B into M which extends g such that
   (a) $\mathrm{cl}^m_M(\hat{g}B) = \hat{g}B \cup \mathrm{cl}^m(gA)$
   (b) $M|\,\mathrm{cl}^m_M(gA)\hat{g}B$ is the free join over gA of $\mathrm{cl}^m_M(gA)$ and $\hat{g}B$.

Each theory T_α is axiomatized by a collection of axioms for semigenericity as defined in Notation 1.28 of [2]. That notation assigns a formula $\phi^m_{A,B,C}$ to each triple $\langle A, B, C \in \mathbf{K}_0\rangle$ and each integer m where $A \leq B$ and C is a possible candidate for $\mathrm{icl^m_M(A)}$. It is then straightforward to show:

1.3 Lemma. *The structure $N \in \mathbf{K}$ is semigeneric, if and only if for each appropriate triple $\langle A, B, C\rangle$ and each $m < \omega$, $N \models \phi^m_{A,B,C}$*

Note that there is a crucial dependence on the class $\mathbf{K}_0$ in determining these axioms. Although formally the axioms in the case appear the same as in [2] the actual axioms differ because the class $\mathbf{K}_0$ is different. Now the main result is:

1.4 Theorem. *For each appropriate triple $\langle A, B, C\rangle$ and each $m < \omega$,*

$$\lim_{n\to\infty} P_n(\phi^m_{A,B,C}) = 1.$$

Proof. Exactly as in the proof of Theorem 2.30 of [2], the argument is reduced to showing the probabilistic Theorem 1.9. We need the following definitions.

1.5 Remark. The essential aspect of the proof of Theorem 2.30 of [2] is that if $A \subseteq B \subseteq B'$ and $\delta(B'/B) < 0$ then the probability of extending an embedding of A to an embedding of B' is less than the probability of extending an embedding of A to an embedding of B exactly as described in Theorem 1.9.

1.6 Definition. Let $A \subseteq B \in \mathbf{K}_0$. Fix a graph G with $|G| = n$ and any $W \subseteq n$ and suppose f embeds A into G. Then $X_{f,W}$ is the number of sets (i.e. possible ranges) $\hat{g}(B - A) \subseteq W$ such that $\hat{g}$ is an extension of f embedding B into G over A.

Viewing $X_{f,W}$ as a function of G as G ranges through $\mathbf{K}_n$, makes $X_{f,W}$ a random variable. Similarly, we will shortly calculate the expectation $\mu_{f,W}$ of $X_{f,W}$ and we write X_f when $W = n$.

1.7 Observation. Fix graphs $A \subseteq B \in \mathbf{K}_0$ with $A \leq B$. Let f map A into n. Further let g_1, g_2 be extensions of f mapping B into n. The probability that g_i is a graph homomorphism depends only on B, i.e is the same for both g_1 and g_2. For, any graph G_1 on n which makes g_1 a homomorphism is transformed by any permutation of n which takes image of g_1 to the image of g_2 into a graph G_2 on n which makes g_2 a homomorphism. For this fixed f, A, B, with $e = e(A, B)$ denoting the number of edges between B and A, let p_n be the probability that an extension of f to B is a graph homomorphism: $\frac{n^{-e\alpha}}{r_n}$.

We write $f \approx g$ if $f = O(g)$ and $g = O(f)$.

1.8 Lemma. *Fix $A, B \in \mathbf{K}_\alpha$, with $e = e(A, B)$ denoting the number of edges between B and A, where $v = |B - A|$, and f is an embedding of A into n. For any $W \subset n$,*

$$E(X_{f,W}) = \mu_{f,W} \approx |W|^v \cdot p_n \approx |W|^v \frac{n^{-e\alpha}}{r_n}.$$

Proof. The probability of an extension g of f being a graph homomorphism is p_n. The number of such embeddings is approximately $|W|^v$. Since expectation is additive, we have the result.

Now, our main aim is to prove:

1.9 Theorem. *Fix graphs $A \subseteq B \in \mathbf{K}_0$ with $A \leq B$. Let V be the event (which depends on constants c, d): for every embedding $f : A \to n$,*

$$d\mu_f \cdot (\log n)^{-(v+1)} < X_f < c\mu_f. \tag{1}$$

Then, for some choice of c, d

$$\lim_{n \to \infty} P_n(V) = 1.$$

Proof. We prove the upper and lower bound separately, although there is some duplication of a main idea in the argument. We fix $A \leq B$ with $|B - A| = v$. We require the following definitions and fact from [2]. In particular, by induction, we can assume that B is primitive over A in the following sense.

1.10 Definition.

1. We say B is in the *intrinsic closure* of A and write $A \prec_i B$ if for every B' with $A \subseteq B' \subset B$, $\delta(B/B') < 0$.
2. We say B is *primitive* over A if $A \leq B$ but for every B' with $A \subset B' \subset B$, $B' \prec_i B$.

1.1 Fact. *[2] There is a binary function $t : \omega \times \omega \to \omega$ which is monotone increasing in both arguments such that if $A, B \in \mathbf{K}_0$, $A \prec_i B$ then for any $M \in \overline{\mathbf{K}}$ with $A \subseteq M$, $X_f \leq t(|A|, |B|)$.*

Upper Bound for Theorem 1.9: Since B is a primitive extension of A, $Ab \prec_i B$ for any $b \in B - A$. Thus, by Fact 1.1, for $M = t(|A|, |B|)$, there are fewer than M extensions g_i with any fixed image of b. The range of each extension g_i can intersect at most $M|B|^2$ extensions so if $X_f = s'$, there is a set of $s = s'/(M|B|^2)$ disjoint extensions. If $s < 3n^v p_n$, (p_n is defined in Observation 1.7.) we are finished. Let $p = |A|$. There are less than $n^p(n^{vs}/s!)$ pairs of an embedding f taking A into n and a set of s extensions (disjoint over rng f) $\langle g_1, \ldots g_s \rangle$ each taking B into W. The probability that each of the g_i is a graph homomorphism extending f is p_n so (via Markov) the probability of such a pair of an embedding and s homomorphisms is at most $n^p n^{vs} p_n^s / s!$. By Stirling's formula, this is less than

$$\frac{n^p (2.72 n^v)^s p_n^s}{(2\pi)^{1/2} s^{s+1/2} (2.71)^{1/(12s+1)}}$$

which tends to 0 if $s \geq 3n^v p_n$. There at most n^v possible choices for s, so the probability that for some f and some $s > 3n^v p_n$, $X_f = s$ is at most $n^v n^p (\frac{2.72}{3})^{3n^v p_n}$. As detailed, in a more complicated situation, at the end of the lower bound proof, this tends to 0 as n tends to infinity.) Thus, a.a. for each f, $s < 3n^v p_n$. In particular, a.a. for each f,

$$X_f \leq 3(M|B|^2) n^v p_n$$

which is the required upper bound with $c = 3(M|B|^2)$.

Lower Bound:

We have fixed A and B. We first calculate a lower bound on the number of extensions of f mapping into a subset W of M_n. For this we relativize the upper bound argument by considering only mappings into W. Choose M as in the proof of the upper bound argument so that each extension g_i of f can intersect at most $M|B|^2$ extensions so if $X_{f,W} = s'$, there is a set of $s = s'/(M|B|^2)$ disjoint extensions of f into W. Let $h = 3 \cdot (M|B|^2)$. The following analysis applies for any b with $0 < b < \mu_{f,W}$; for us, $b = \frac{\mu_{f,W}}{[1+\ln n]}$. For each n and each f mapping A into n, just expanding the definition of expectation,

$$\mu_{f,W} = \sum_{s > h\mu_{f,W}} s \cdot P(X_{f,W} = s) + \sum_{b \leq s \leq h\mu_{f,W}} s \cdot P(X_{f,W} = s) + \sum_{s<b} s \cdot P(X_{f,W} = s).$$

Thus, adding and subtracting $b \cdot P(X_{f,W} > b)$,

$$\begin{aligned} \mu_{f,W} \quad \leq & \quad \sum_{s > h\mu_{f,W}} (s-b) \cdot P(X_{f,W} = s) \\ + & \quad (h\mu_{f,W} - b) \cdot P(b \leq X_{f,W} \leq h\mu_{f,W}) \\ + & \quad b \cdot P(X_{f,W} \geq b) + b \cdot P(X_{f,W} < b). \end{aligned}$$

This implies

$$(\mu_{f,W} - b) \leq \sum_{s > h\mu_{f,W}} (s-b) \cdot P(X_{f,W} = s) + (h\mu_{f,W} - b) \cdot P(b \leq X_{f,W} \leq h\mu_{f,W}).$$

We conclude, (neglecting one division by $h\mu_{f,W} - b$ since it only improves the inequality, as $h\mu_{f,W} - b > 1$),

$$\begin{aligned} P(b \leq X_{f,W}) &\geq P(b \leq X_{f,W} \leq h\mu_{f,W}) \\ &\geq \frac{\mu_{f,W} - b}{h\mu_{f,W} - b} - \sum_{s > h\mu_{f,W}} (s-b) \cdot P(X_{f,W} = s). \end{aligned}$$

Now s varies through the values of $X_{f,W}$, which is always at most $|W|^v$. As in the upper bound argument, if there are s embeddings, there are $t = \frac{s}{M|B|^2}$ whose ranges intersect in the range of f. Thus, invoking Markov's inequality, for each s

$$P(X_{f,W} = s) \leq |W|^{vt} p_n^t / t!. \tag{2}$$

In contrast to the upper bound argument, we treat each f separately for the moment. By Stirling's formula, for each t, the right hand side of Equation 2 is less than

$$\frac{(2.72|W|^v)^t p_n^t}{(2\pi)^{1/2} t^{t+1/2} (2.71)^{1/(12t+1)}}.$$

Recall that $\mu_{f,W} = |W|^v \cdot p_n$. Thus, for $|W|^v \geq s > h\mu_{f,W}$, $\frac{s}{M|B|^2} > \frac{3M|B|^2|W|^v p_n}{M|B|^2}$ so the $t^{t+1/2}$ dominates $(3|W|^v p_n)^t$. These conditions on s yield, for an appropriate constant c_2,

$$\sum_{s > h\mu_{f,W}} (s-b) \cdot P(X_{f,W} = s) \leq c_2 |W|^v \cdot (2.72/3)^t \leq c_2 |W|^v \cdot (2.72/3)^{|W|^v p_n}$$

since $t = \frac{s}{M|B|^2}$ and $s > M|B|^2|W|^v p_n$. Using this inequality in our estimate for $P(X_{f,W} \leq b)$ yields:

$$P(X_{f,W} \leq b) \leq \left[1 - \left(\frac{\mu_{f,W} - b}{h\mu_{f,W} - b} - c_2|W|^v \cdot (2.72/3)^{|W|^v p_n}\right)\right].$$

Choose k so that $(1 - \frac{1}{4h})^k < \frac{1}{e}$. (This choice will be used in a couple of paragraphs; the necessity for this step was pointed out to me by Jim Lynch.) Let $R = [1 + \ln n] \cdot |A| \cdot k$. As n contains R disjoint subsets W_i of length n/R, and the X_{f,W_i} are independent, the probability that for all i, $X_{f,W_i} < \frac{\mu_{f,W_i}}{R}$ is less than

$$\left[1 - \left(\frac{\mu_{f,W_i} - b}{h\mu_{f,W_i} - b} - c_2|W_i|^v \cdot (2.72/3)^{|W_i|^v p_n}\right)\right]^R.$$

Note that the computation in Lemma 1.8 of $\mu_{f,W}$ depends only on the cardinality of W so it makes sense to write $\mu_{f,|W_i|}$. We have:

$$P(X_f < \mu_{f,|W_i|}/R) \leq \left[1 - \left(\frac{\mu_{f,|W_i|} - b}{h\mu_{f,|W_i|} - b} - c_2|W_i|^v \cdot (2.72/3)^{|W_i|^v p_n}\right)\right]^R.$$

Now, let V_ℓ (V lower) be the event that $X_f > \frac{\mu_{f,|W_i|}}{R}$ for all f. Thus $P(V_\ell)$ is bounded below by one minus the number of f's times the probability that an f has few extensions. That is,

$$P(V_\ell) \geq 1 - \left[n^p \cdot \left(1 - (\frac{\mu_{f,|W_i|} - b}{h\mu_{f,|W_i|} - b} - c_2|W_i|^v \cdot (2.72/3)^{|W_i|^v p_n}) \right)^R \right].$$

More explicitly, taking $b = \frac{\mu_{f,|W_i|}}{R}$ and $|W_i|$ as n/R,

$$P(V_\ell) \geq 1 - \left[n^p \cdot \left(1 - (\frac{1-\frac{1}{R}}{h-\frac{1}{R}} - c_2(\frac{n}{R})^v \cdot (2.72/3)^{(\frac{n}{R})^v p_n}) \right)^R \right].$$

The right hand side of this expression goes to one. First, for sufficiently large n,

$$c_2(\frac{n}{R})^v \cdot (2.72/3)^{(\frac{n}{R})^v p_n} < \frac{1}{4h}.$$

This follows from the fact that for $0 < g < 1$, $a > 0$ and positive integers b, c, $n^b g^{\frac{n^a}{(\ln n)^c}}$ tends to 0 as n tends to infinity. (To see this note that for every $\alpha > 0$, for suffciently large n, $(\ln n)^c < n^\alpha$ so $\frac{n^a}{(\ln n)^c} > n^{a-\alpha} = n^{b'}$ with $b' > 0$ and by L'Hopital's rule for any integer b and $b' > 0$, $n^b g^{n^{b'}}$ tends to zero as n tends to infinity.) More easily,

$$\frac{1-\frac{1}{R}}{h-\frac{1}{R}} > \frac{1}{2h}.$$

Thus

$$1 - (\frac{1-\frac{1}{R}}{h-\frac{1}{R}} - c_2(R)^v \cdot (2.72/3)^{(\frac{n}{R})^v p_n}) \leq 1 - \frac{1}{4h}.$$

Since for any integer p and any n and c with $0 < c < \frac{1}{e}$, $n^p c^{p \cdot \ln n}$ tends to 0 as n tends to infinity, we deduce from $R = [1 + \ln n] \cdot |A| \cdot k$ and $(1 - \frac{1}{4h})^k < \frac{1}{e}$ that the probability that for all f, $X_f > b$ tends to one. But since for each i, by Lemma 1.8

$$\mu_{f,|W_i|} = |W_i|^v \cdot \frac{n^{-e\alpha}}{r_n} = (n/R)^v \cdot \frac{n^{-e\alpha}}{r_n},$$

a.a. for every f, since $r_n \leq 1$,

$$X_f \geq (n/R)^v \cdot \frac{n^{-e\alpha}}{r_n} \cdot (R)^{-1} \geq n^v \cdot n^{-e\alpha} \cdot (R)^{-(v+1)}.$$

Since $R = k \cdot |A| \cdot \ln n$, incorporating $(k \cdot |A|)^{-(v+1)}$ in the constant d, we get the required result. $\square_{1.9}$

This completes the proof. The $\aleph_0$-categorical case is simpler in several respects. Since for any $M \in \mathbf{K}_0$, and any finite A, $\mathrm{icl_M(A)}$ is finite it is possible to replace the axioms defining semigenericity by a set of Π_2 sentences. The probability argument no longer requires the upper bound. (Under the argument given here, this is not much of a saving since the upper bound argument is essentially embedded in the lower bound proof. However, one could use the second moment argument of [6] instead.) Note that by Lindstrom's theorem, the $\aleph_0$-categorical T_α are model complete while in general T_α is not model complete.

References

[1] J.T. Baldwin. Random expansions of geometries. submitted.

[2] J.T. Baldwin and S. Shelah. Randomness and semigenericity. *Transactions of the American Mathematical Society*, 349:1359–1376, 1997.

[3] J.T. Baldwin and Niandong Shi. Stable generic structures. *Annals of Pure and Applied Logic*, 79:1–35, 1996.

[4] G.L. Cherlin. Combinatorial problems connected with finite homogeneity. *Contemporary Mathematics*, pages 3–30, 1992 (Part 3).

[5] B. Herwig. Weight ω in stable theories with few types. *J. Symbolic Logic*, 60:353–373, 1995.

[6] S. Shelah and J. Spencer. Zero-one laws for sparse random graphs. *Journal of A.M.S.*, 1:97–115, 1988.