THE LOGIC IN COMPUTER SCIENCE COLUMN

BY

YURI GUREVICH

Microsoft Research
One Microsoft Way, Redmond WA 98052, USA
gurevich@microsoft.com

ZERO-ONE LAWS: THESAURI AND PARAMETRIC CONDITIONS*

Andreas Blass† Yuri Gurevich

Abstract

The 0-1 law for first-order properties of finite structures and its proof via extension axioms were first obtained in the context of arbitrary finite structures for a fixed finite vocabulary. But it was soon observed that the result and the proof continue to work for structures subject to certain restrictions. Examples include undirected graphs, tournaments, and pure simplicial complexes. We discuss two ways of formalizing these extensions, Oberschelp's (1982) parametric conditions and our (2003) thesauri. We show that, if we restrict thesauri by requiring their probability distributions to be uniform, then they and parametric conditions are equivalent. Nevertheless, some situations admit more natural descriptions in terms of thesauri, and the thesaurus point of view suggests some possible extensions of the theory.

*Bulletin of European Association for Theoretical Computer Science, #91, February 2007.

†Mathematics Department, University of Michigan, Ann Arbor, MI 48109–1043, U.S.A., ablass@umich.edu

Quisani: I've been thinking about zero-one laws for first-order logic. I know it's a rather old topic, but I noticed something in the literature that I'd like to understand better. The first proof [5] established that, for any first-order sentence σ in a finite relational vocabulary Υ, the proportion of models of σ among all Υ-structures with base set $\{1, 2, \ldots, n\}$ approaches 0 or 1 as n tends to infinity. Fagin [4] rediscovered the result (with a simpler proof) and added, near the end of his paper, some remarks about what happens if, instead of considering all Υ-structures, we consider only those satisfying some specified sentence τ. He pointed out that for some but not all choices of τ, there is still a 0-1 law: The proportion of models of σ among models of τ with base set $\{1, 2, \ldots, n\}$ approaches 0 or 1 as n tends to infinity. He gave two examples of such τ, both in the language with just a single binary relation symbol E, the language of digraphs. One example was the sentence saying that τ is symmetric and irreflexive, so the models are undirected loopless graphs. The other example defined the class of tournaments. The case of undirected graphs was rediscovered in [2], where another example was added, pure d-dimensional simplicial complexes, formulated using a completely symmetric and completely irreflexive $(d + 1)$-ary relation.

I'd think there should be some general result explaining these variants of the 0-1 law.

Authors: There is such a result in Oberschelp's paper [6], but it seems he never published the proof. His result, expressed in terms of what he calls parametric conditions, covers the variants that you mentioned as well as others, for example involving graphs with several colors of edges. It is based on the same approach, via extension axioms, as the work in [4] and many later works. So it doesn't cover the 0-1 laws obtained by other methods, for example by Compton (see [3] and the references given there) for slowly-growing classes of structures.

Later, not knowing of Oberschelp's work, we introduced in [1] the notion of a thesaurus as a suitable context for Shelah's proof of the 0-1 law for choiceless polynomial time. It also provides a suitable context for the 0-1 law in the more restrictive context of first-order logic, once that logic is appropriately defined for thesauri.

Q: Does the thesaurus approach also depend on the extension axioms? And should it be combined with parametric conditions to produce a common generalization?

A: Both of your questions are answered — the first affirmatively and the second negatively — by the fact that the two approaches, parametric conditions and thesauri, are essentially equivalent, at least when applied to the case of uniform probability distributions on structures of any given size.

Q: That leaves me with a lot of questions: What are parametric conditions? What are thesauri? What exactly does "essentially equivalent" mean in this context? And what happens when the probability distributions aren't uniform?

A: Let's start with Oberschelp's parametric conditions. These are conjunctions of first-order universal formulas, for a relational vocabulary, having the special form

$$\forall x_1 \dots \forall x_k \, (D(\vec{x}) \to C(\vec{x})),$$

where $\vec{x}$ stands for the n-tuple of variables $x_1, \dots, x_k$, where $D(\vec{x})$ is the formula $\bigwedge_{1 \le i < j \le k} x_i \neq x_j$ saying that the values of these variables are distinct, and where $C(\vec{x})$ is a propositional combination of atomic formulas such that, in each atomic subformula of $C(\vec{x})$, all k of the variables occur.

Q: I assume that k is allowed to vary from one conjunct to another in a parametric condition, and that when $k \le 1$ the empty conjunction D is interpreted as true. So, for example, irreflexivity of a binary relation is expressed by the parametric condition $\forall x(\texttt{true} \to \neg R(x, x))$.

A: That's right, and it's easy to express the other conditions you mentioned earlier — symmetry for undirected graphs, asymmetry for tournaments, and complete irreflexivity and symmetry for simplicial complexes — as parametric conditions.

Q: I see that, but I don't yet see the significance of the requirement that all atomic subformulas of C use all the variables.

A: The simplest explanation is that if you drop this requirement then the extension axioms need not have asymptotic probability 1. For example, for almost all finite partially ordered sets, the longest chain has length three (see the proof of [3, Theorem 5.4]). So there are configurations, like a four-element chain, that can arise in partial orders but are absent with asymptotic probability 1. Traditional extension axioms, in contrast, imply that any configuration permitted by the underlying assumption τ must occur. The trouble comes from the transitivity clause in the definition of partial orders; it involves three variables but each atomic subformula uses only two of them.

Q: The example shows that some requirement is needed to eliminate the case of partial orders, but how does the "use all the variables" requirement connect with extension axioms? I guess what I'm really asking for is a sketch of Oberschelp's proof.

A: The crucial contribution of parametricity is that it permits a reformulation of the uniform probability measure on structures of a fixed size n in terms of independent choices of the truth values of instances of the relations.

Recall that, when we consider the class of all structures (of a given relational vocabulary) with universe $\{1, 2, \ldots, n\}$, the uniform probability measure on these structures can be described by saying that each instance $R(a_1, \ldots, a_r)$ (where R ranges over the relations of the structure and $\vec{a}$ over tuples of appropriate length from $\{1, 2, \ldots, n\}$) is independently assigned truth value true or false, with equal probability. When we deal with, say, loopless undirected graphs, this description must be modified, since $R(a, a)$ must be false and since $R(a_1, a_2)$ must have the same truth value as $R(a_2, a_1)$. Nevertheless, the uniform distribution can still be described in terms of independent flips of a fair coin: flip a coin for each 2-element subset $\{a_1, a_2\}$ of $\{1, 2, \ldots, n\}$ to determine both $R(a_1, a_2)$ and $R(a_2, a_1)$. Similarly in the case of tournaments, a single flip of a fair coin decides which one of $R(a_1, a_2)$ and $R(a_2, a_1)$ shall hold. And similarly in the other examples.

Something similar happens for arbitrary parametric conditions τ. To describe it, we need the notion of a k-type relative to τ. Temporarily fix a positive integer k, less than or equal to the maximum arity of the relation symbols in the vocabulary of τ. Consider all the atomic formulas that use exactly the variables $x_1, \ldots, x_k$, possibly more than once. A *k-type* is an assignment of truth values to these atomic formulas that makes $C(\vec{x})$ true whenever $\forall \vec{x}(D(\vec{x}) \to C(\vec{x}))$ is a conjunct of τ (up to renaming bound variables, so that $\vec{x}$ is $x_1, \ldots, x_k$). In other words, a k-type is an assignment of truth values that can be realized by a k-tuple of distinct elements in a model of τ.

Now the uniform distribution on models of τ with base set $\{1, 2, \ldots, n\}$ admits the following equivalent description: For each k and each k-element subset $\{a_1 < a_2 < \cdots < a_k\} \subseteq \{1, 2, \ldots, n\}$, choose, uniformly at random, a k-type to be realized by the k-tuple $(a_1, \ldots, a_k)$. This works because each of these types is realized by $(a_1, \ldots, a_k)$ in equally many models of τ with base set $\{1, 2, \ldots, n\}$ and because different increasing tuples $(a_1, \ldots, a_k)$ behave independently. Furthermore, once the k-types of increasing tuples $\vec{a}$ are chosen, they determine all the relations of the structure.

Q: What about instances of the relations where the arguments are not in increasing order?

A: They're included, because the atomic formulas to which a k-type assigns truth values include those in which the variables $\vec{x}$ occur out of order.

Once one has this alternative description of the uniform distribution on models, one can easily imitate the traditional proof of 0-1 laws. There is an extension axiom for each k-type with $k > 0$; it says that, for any distinct $x_1, \ldots, x_{k-1}$, there is an x_k, distinct from all of them, such that the tuple $(x_1, \ldots, x_{k-1}, x_k)$ realizes the given k-type. It is easy to check that each extension axiom has asymptotic probability 1 and that the theory axiomatized by the extension axioms is complete. (To prove completeness, one can proceed as in [4] because the theory is

$\aleph_0$-categorical, or one can eliminate quantifiers as in [2].)

The role of parametricity in this argument is to ensure that all the information about any k-tuple of distinct elements can be isolated in its k-type and the k'-types of its subtuples, a finite amount of information, whose size is independent of the size n of the base set. That allows us to formulate extension axioms and verify their asymptotic validity. Contrast this with the situation for, say, partial orders. Here the requirement of transitivity imposes correlations between a truth value $R(a,b)$ and many other truth values $R(a,c)$ and $R(b,c)$, for all c in the base set. The number of relation instances correlated with a single $R(a,b)$ thus grows with the structure and the proof described above breaks down. Oberschelp [6] summarizes this (in the case of a vocabulary with only one relation symbol) by saying "A parametric property defines a class of relations which can be determined by the independent choice of values (parameters) in fixed regions of the adjacency array."

Q: OK. I see that parametricity seems to be just what's needed to carry out the traditional proof of the 0-1 law via extension axioms. Now what are thesauri?

A: A *thesaurus* is a finite set of signa, so of course we have to say what a signum is, but let's first deal with a simplified notion of signum, which turns out to have the same generality as parametric conditions. A *signum* in this simplified sense consists of

- a symbol R (assumed to be different for the different signa in a thesaurus),
- a natural number j called the arity,
- a finite set V called the value set[1],
- a group G of permutations of $\{1, 2, \ldots, j\}$, and
- a homomorphism from G to the group of permutations of V.

For notational convenience, one often writes simply the symbol R when one really means the whole signum.

Q: That sounds pretty complicated; what's really going on here?

A: The symbol R and the arity j are analogous to what you have in a relational vocabulary of ordinary first-order logic — a symbol and the number of its argument places. Our R's, however, will not necessarily be 2-valued as in first-order logic, but v-valued, where v is the cardinality of the value set V. So we could, for

[1] In [1], the value set was always of the form $\{1, 2, \ldots, v\}$ for some positive integer v. Allowing arbitrary finite sets of values makes no essential difference but is technically convenient.

example, treat a graph with colored edges by having a single binary signum where V is the set of colors plus one additional value to indicate the absence of an edge.

The other two constituents of the signum, the group G and the homomorphism h, describe the symmetry properties that we intend R to satisfy. The idea is that permuting the j arguments of R by a permutation π in G results in a change of the value given by $h(\pi)$. More precisely, a structure $\mathfrak{A}$ for a thesaurus consists of a base set A together with, for each signum $\langle R, j, v, G, h\rangle$ (often abbreviated as just R), an interpretation $R^{\mathfrak{A}}$ assigning to each j-tuple of distinct elements $a_1, \dots, a_j$ in A a value $R^{\mathfrak{A}}(\vec{a})$, subject to the symmetry requirement

$$R^{\mathfrak{A}}(a_1, \dots, a_j) = h(\pi)(R^{\mathfrak{A}}(a_{\pi(1)}, \dots, a_{\pi(j)})).$$

Q: The following variation seems more natural to me:

$$R^{\mathfrak{A}}(a_{\pi(1)}, \dots, a_{\pi(j)}) = h(\pi)(R^{\mathfrak{A}}(a_1, \dots, a_j))$$

It explains how to obtain $R^{\mathfrak{A}}(a_{\pi(1)}, \dots, a_{\pi(j)})$ from $R^{\mathfrak{A}}(a_1, \dots, a_j)$.

A: This doesn't work unless you either put π^{-1} on one side of the equation or make h an anti-homomorphism. Here's the calculation, using your proposed variation. Let π and σ be two permutations in the group, let $\vec{a}$ be a j-tuple of elements of A, and let $\vec{b}$ be the j-tuple defined by $b_i = a_{\sigma i}$.

$$\begin{aligned} h(\pi)h(\sigma)R^{\mathfrak{A}}(a_1, \dots, a_j) &= h(\pi)R^{\mathfrak{A}}(a_{\sigma 1}, \dots, a_{\sigma j}) \\ &= h(\pi)R^{\mathfrak{A}}(b_1, \dots, b_j) \\ &= R^{\mathfrak{A}}(b_{\pi 1}, \dots, b_{\pi j}) \\ &= R^{\mathfrak{A}}(a_{\sigma\pi 1}, \dots, a_{\sigma\pi j}) \\ &= h(\sigma\pi)R^{\mathfrak{A}}(a_1, \dots, a_j), \end{aligned}$$

where we've applied your variation three times, once with σ, once with π, and once with $\sigma\pi$. So for this to work, we'd need $h(\sigma\pi) = h(\pi)h(\sigma)$, i.e., h should be an anti-homomorphism.

Q: I suppose using an anti-homomorphism wouldn't be a disaster, but it would defeat the purpose of my suggestion, increased naturality.

Now why does $R^{\mathfrak{A}}$ apply only to j-tuples of distinct elements?

A: Distinctness is technically convenient. For example, in tournaments, one wants the truth value of $R(a, b)$ to be negated if a and b are interchanged, except when $a = b$. So we think of a binary relation R as being given by two signa, one binary signum for distinct arguments and one unary signum for equal arguments. Similarly, a relation of higher arity would be represented by several signa, one for each way of partitioning the argument places into blocks with equal arguments.

Q: I see that, just as with parametric conditions, you can represent the uniform probability distributions on structures with base set $\{1, 2, \dots, n\}$ in terms of independent random choices for some instances $R(\vec{a})$. For each signum R, choose a representative from each G-orbit of j-tuples of distinct elements, and assign R random values at these representatives. Then propagate these assignments through the whole orbits by means of the symmetry requirement.

A: That's right. To be precise about these G-orbits, one should say that G acts naturally on the set of j-tuples of elements from any set by

$$\pi(a_1, \dots, a_j) = (a_{\pi^{-1}(1)}, \dots, a_{\pi^{-1}(j)}).$$

Q: With this formulation in terms of independent random choices, it should be possible to prove something analogous to extension axioms for the thesaurus context. I'd expect almost all Υ-structures to have the following property, for each n: Given any n distinct points $a_1, \dots, a_n$, there is a point b, distinct from all the a_i, and giving prescribed values for all signum instances $R^{\mathfrak{A}}(c_1, \dots, c_j)$ where one of the c_i is b and the others are distinct elements of $\{a_1, \dots, a_n\}$.

A: That's right, provided the prescribed values obey the symmetry requirement for thesaurus models. We're pleased that you remembered that the arguments of a signum are supposed to be distinct, so that b should occur only once in $R^{\mathfrak{A}}(c_1, \dots, c_j)$ and each a_i should occur at most once.

Q: This result should yield 0-1 laws for thesauri, except that you haven't yet defined first-order logic in the context of thesauri.

A: Indeed, we have not introduced a syntax to go with these semantical notions in [1], but it is not difficult to do so. Take atomic formulas to be $R(\vec{x}) = c$ where R is a signum (or the symbol part of it), $\vec{x}$ is a sequence of variables of length equal to the arity of R, and $c \in V$. Also allow equality as usual in first-order logic. Then form compound formulas using propositional connectives and quantifiers, just as in ordinary first-order logic. The semantics is obvious. (If the values of the variables in $\vec{x}$ are not all distinct, then $R(\vec{x}) = c$ is naturally taken to be false.)

If one is willing to stretch the notion of syntax a bit, then it would be appropriate to identify the atomic formulas $R(x_1, \dots, x_j) = c$ and $R(x_{\pi^{-1}(1)}, \dots, x_{\pi^{-1}(j)}) = h(\pi)(c)$ for any π in the group of the signum R, since the symmetry requirement for structures implies that these will always have the same truth value.

Once these definitions are in place, it is, as you said, not difficult to show, via extension axioms, that first-order sentences have asymptotic probabilities 0 or 1 over the class of all structures of a thesaurus.

Q: This should also be clear for another reason, once you explain how thesauri and parametric conditions are essentially equivalent. Having the 0-1 law for parametric conditions, we should be able to use the essential equivalence to deduce the 0-1 law for thesauri. But what exactly did you mean by essential equivalence?

A: Essential equivalence has several components. First, for each thesaurus Υ, there is a parametric condition τ (in some vocabulary) such that the Υ-structures with any particular base set (for example $\{1, 2, \ldots, n\}$) are in (natural) one-to-one correspondence with models of τ on the same base set. Second, for each first-order sentence of the thesaurus, there is a first-order sentence of the vocabulary of τ such that the models of these sentences match up under the correspondence above. Third, conversely, for each parametric condition τ there is a thesaurus Υ with a one-to-one correspondence as before. And fourth, for every first-order sentence of the vocabulary of τ, there is a first-order sentence of Υ with the corresponding models.

Q: That seems to be exactly what's needed in order to convert 0-1 laws from either of the two contexts to the other. So how do these correspondences work?

A: One direction is implicit in the syntax for thesauri described above. Given a thesaurus Υ, form a first-order vocabulary Υ' with the same atomic formulas. That is, for each j-ary signum R of Υ and each value c, let R_c be a j-ary relation symbol in Υ'. The intended interpretation is that $R_c(\vec{a})$ should mean $R(\vec{a}) = c$. Every Υ-structure gives, in this way, a structure (in the ordinary, first-order sense) for Υ'. The converse is in general false, but the collection of Υ'-structures arising from Υ-structures in this way can be described by a parametric condition τ. The conjuncts in τ express the symmetry requirements of the thesaurus, i.e.,

$$\forall \vec{x}(D(\vec{x}) \to (R_c(\vec{x}) \to R_{h(\pi)(c)}(x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(j)})))$$

for each signum R and each π in its group. There are also conjuncts saying that every j-tuple of distinct elements satisfies R_c for exactly one $c \in V$ and that R_c is false whenever two of its arguments are equal. We trust this makes the first two parts of "essentially equivalent" clear.

Q: Yes; "essentially equivalent" is now half clear. But I suspect that this was the easier half. How do you handle the reverse direction?

A: Here we have to convert a parametric condition τ into a thesaurus Υ. Let Υ consist of one j-ary signum T_j for each j up to the maximum arity of the relation symbols in the vocabulary of τ.

Q: Just one signum per arity, no matter how rich the vocabulary of τ is?

A: That's right. We compensate by using a rich set of values. Take the values of the j-ary signum T_j to be the j-types relative to τ. (This is one place where it's convenient to allow a signum to have any finite set of values, rather than only an initial segment of the positive integers as in [1].)

The group associated to the j-ary signum T_j is the symmetric group of all permutations of $\{1, 2, \dots, j\}$. To describe its action on the set of values, i.e., on the set of j-types, just let it act on the j variables occurring in the types. That is, the truth value assigned to an atomic formula θ by the type $h(\pi)(c)$ is the same as the truth value assigned by c to the formula obtained from θ by substituting $x_{\pi^{-1}(i)}$ for x_i for all i.

A structure for this thesaurus provides, for each j-tuple $\vec{a}$ of elements of the base set, a j-type to be realized by this j-tuple. This specifies which atomic formulas are to be true of this $\vec{a}$ and of all permutations of $\vec{a}$. The symmetry requirement on thesaurus structures is exactly what is needed to ensure that these specified truth values for the various permutations of $\vec{a}$ are consistent and thus describe a structure for the vocabulary of τ. Furthermore, since we use only j-types relative to τ, the resulting structures will be models of τ. And it is easy to check that every model of τ arises from exactly one Υ-structure.

Q: That takes care of the third part of essential equivalence. For the fourth part, you have to translate formulas in the vocabulary of τ into Υ-formulas. Since the syntax and semantics of thesauri treats connectives and quantifiers the same way as first-order logic does, it suffices to consider atomic formulas.

A: Right, and handling these is mainly just bookkeeping. Given an atomic formula θ in the vocabulary of τ, let $\{x_1, \dots, x_k\}$ be the set of distinct free variables in it. For each equivalence relation E on this set of variables, we can write a quantifier-free formula θ'_E, in the syntax associated to the thesaurus Υ, with variables $x_1, \dots, x_k$, saying that

- the equality pattern of the x_i is given by E, i.e.,

$$\bigwedge_{(i.j)\in E} (x_i = x_j) \wedge \bigwedge_{(i,j)\notin E} \neg(x_i = x_j),$$

  and

- the type realized by distinct values of x_i's gives θ the value true, i.e.,

$$\bigvee_c (T_r(x_{i_1}, \dots, x_{i_r}) = c),$$

  where $x_{i_1}, \dots, x_{i_r}$ are chosen representatives of the equivalence classes of E and where c ranges over those r-types that assign true to the formula obtained from θ by replacing each variable by the chosen representative of its equivalence class.

Then the disjunction of these formulas θ'_E, over all equivalence relations E, is satisfied by a tuple of elements $\vec{a}$ in exactly those Υ-structures that correspond to models of τ in which θ is satisfied by $\vec{a}$.

Q: This bookkeeping sounds complicated, but I think I get it. Your Υ-translation of θ just says that the values of the x_i's satisfy some pattern of equations and negated equations (an equality-type) and that the distinct ones among those values give T_r a value that corresponds to θ being true.

A: Right. So this finishes the explanation of how parametric conditions and thesauri are essentially equivalent.

Q: Yes, except you said earlier that you were using a simplified notion of signum. What's the full-scale notion?

A: In [1], our definition of a signum included, in addition to R, j, V, G, h as above[2], a probability distribution p on the set of V values, subject to the requirements that each value has non-zero probability and that the distribution is invariant under the group $h(G)$.

The purpose of p is to modify the probability distribution on the structures with a fixed base set $\{1, 2, \ldots, n\}$. Previously, we chose a representative j-tuple from each G-orbit in $\{1, 2, \ldots, n\}^j$, and we chose the values of R at these representatives uniformly at random. Now, we choose these values according to the probability distribution p. And then, as before, we propagate the chosen values to all the other j-tuples by means of the symmetry requirement for thesaurus-structures.

Q: You require that p is invariant under $h(G)$ to ensure that the probability distribution on structures is independent of the choice of representative tuples.

A: Exactly. And of course the requirement that each value have non-zero probability is just a normalization. Any value whose probability is zero could be omitted, since it is unused in almost all structures.

Q: Is this more general notion of thesaurus equivalent to anything in the parametric condition world?

A: As far as we know, parametric conditions have been used only in connection with the uniform probability distribution on the structures that satisfy the conditions. But there is nothing to prevent one from introducing non-uniform distributions in this context, in a way that is equivalent to general thesauri, via the essential equivalence described above.

---

[2] Actually, as noted in an earlier footnote, it had a number v rather than a set V, but the difference is irrelevant.

Q: Apart from non-uniform probabilities, are the other advantages to thinking in terms of thesauri rather than in terms of parametric conditions?

A: Yes, we see a few.

First, certain structures are naturally thought of as having multi-valued relations. For example, colored graphs, where the values of the edge relation would be the colors and "false" (the latter for where there is no edge). Similarly, a tournament in which the outcome of a game can be a tie or a win for either player seems to be naturally viewed as a three-valued binary relation on the set of players.

Second, thesauri suggest a generalization that may be worth exploring. Instead of a fixed set of values for each signum, we could let the number of values grow slowly with the size of the structure.

Q: Won't that mess up the extension axioms?

A: Not if "slowly" is taken seriously. The usual computation showing that extension axioms have asymptotic probability 1 still works if the number of values of any relation grows more slowly than n^ε for each positive real number ε, as n, the number of elements in the structure, tends to infinity. So for example, $\log n$ values would be OK.

Q: What does it mean that the computation still works?

A: It means that, using the same ideas as in the proof of the usual first-order extension axioms, one finds that almost all finite models for a generalized thesaurus of this sort have the following property for each fixed natural number k. Take k variables, say $x_1, \dots, x_k$ and specify possible values for all the (finitely many) atomic formulas that use only these variables, at most once each, and really do use x_k. Here "possible values" means that

- if an atomic formula begins with the signum R then the value must be in the value set of that signum, and
- if two atomic formulas begin with R, and their variables differ only by a permutation π of argument places, and π is in the group G of the signum R, then the values must correspond via $h(\pi)$, as required in the definition of thesaurus structures.

Then, if one interprets $x_1, \dots, x_{k-1}$ as any $k-1$ distinct elements of the structure, there will be an interpretation for x_k, distinct from these, and realizing all the given values for atomic formulas.

Q: OK, so it's really analogous to traditional extension axioms. How does the "usual computation" work in this situation.

A: Well, fix k and fix an assignment of values to atomic formulas as above. We want to show that, with asymptotic probability 1, given any distinct $x_1, \dots, x_{k-1}$, there is some x_k realizing the given values.

Consider a structure of size n, and let v be the largest of the cardinalities of the value sets, in this structure, for all the signa. Notice that the number of atomic formulas to which values are assigned is a constant f, because k and the thesaurus are fixed. So there are at most v^f possible assignments of values to these atomic formulas.

Temporarily, fix the interpretations of $x_1, \dots, x_{k-1}$. There are $n-k+1$ possible interpretations, distinct from these, for x_k. Each of these will, with the given interpretations of $x_1, \dots, x_{k-1}$, realize one of the $\le v^f$ possible assignments of values to atomic formulas, so it has probability $\ge 1/v^f$ of realizing the assignment we want. Therefore, the probability that no interpretation of x_k realizes our desired assignment is at most

$$\left(1 - \frac{1}{v^f}\right)^{n-k+1}.$$

Since $1 - t \le e^{-t}$ for all t, this is at most $e^{-n/(2v^f)}$, where the factor 2 (more than) compensates for omitting $-k + 1$ once n is larger than $2k$. Our assumption that v grows slowly compared with n means, in particular, that $2v^f < \sqrt{n}$ once n is large enough. For such large n, we conclude that the probability that no x_k realizes the desired values is $< e^{-\sqrt{n}}$.

Now un-fix the interpretation of $x_1, \dots, x_{k-1}$. There are at most n^{k-1} such interpretations, so the probability that at least one of them has no suitable x_k is $< n^k e^{-\sqrt{n}}$. That's the probability that our analog of the k^{th} extension axiom fails, and the upper bound $n^k e^{-\sqrt{n}}$ approaches 0 as $n \to \infty$.

We have not studied possible applications of this idea, or even the appropriate extensions of the syntax and semantics of first-order logic. Once the number of values isn't constant, relations seem intuitively to behave more like functions than like relations in the traditional first-order world. One could also equip the set V of values with some relations or functions so that it becomes a structure in its own right.

Q: You'll need some structure on V, at least implicitly, in order to formulate first-order sentences over such a generalized thesaurus. For ordinary thesauri, your first-order language had, in effect, names for all the values $c \in V$. But with V now allowed to grow, that would make the language vary with the structure — not a good idea if you want to talk about the asymptotic probability of fixed sentences as the structure grows.

A: The idea could still work if, as the structure grows, the language also grows, so that any fixed sentence would make sense in all sufficiently large structures. But in fact, in at least one situation, V is naturally a relational structure.

Q: What situation is that?

A: Consider a tournament in which the result of each game (i.e., the relationship between a pair of players) is not merely a win for one or the other (as in traditional tournaments) or a tie (as in a generalization mentioned above) but can be any one of several "degrees of victory", where these degrees come from a small, linearly ordered set V. As before, "small" should mean of cardinality $< n^{\varepsilon}$ for each fixed $\varepsilon > 0$ and all large n. The outcomes of the games are modeled by a V-valued function R subject to the (anti-)symmetry requirement that $R(x, y)$ and $R(y, x)$ are symmetrically located in V, i.e., each is the image of the other under the unique order-reversing permutation of V.

In this situation, it is natural to admit the linear ordering of V as part of the structure, so that there are atomic formulas like $R(x, y) < R(u, v)$.

Q: Do you also want to allow names for the elements of V? If so, then it seems that your suggestion of a growing language depends on making some arbitrary conventions here. How shall the interpretation of a particular name vary while n and therefore V grow? There would be no problem with names for the first element, the second, the last, etc., but what about names for the element one third of the way up, or the element in position $\lfloor \sqrt{|V|} \rfloor$?

A: For the purposes of the present discussion, we'd want a language for which the 0-1 law holds. Apart from that, the choice of language would be guided by potential applications.

Notice, though, that as long as we have the ordering relation on V available, names for the first element, the second, the last, etc. are not really needed, with asymptotic probability 1, because these elements are definable in terms of the ordering.

Q: Wait a minute. Those definitions use quantifiers. Do you intend to allow quantification over V in your language?

A: With asymptotic probability 1, we have quantification over V automatically, since the elements of V are almost surely the same as the values of R. So, for example, we can almost surely express "$R(x, y)$ is the first element of V" by the formula

$$\forall u\, \forall v\, \neg(R(u, v) < R(x, y)).$$

Q: I get it. This formula doesn't express the desired property in all structures, but it does in those structures where, as u and v range over the players, $R(u, v)$ ranges over *all* elements of V — and that's almost all structures.

A: Such expressive power requires some caution in the definition of structures, specifically in the choice of how V is allowed to grow. For example, we had better insist that the cardinality of V has the same parity in almost all structures, because that parity is definable by the truth value of the sentence

$$\exists x\, \exists y\, (R(x, y) = R(y, x)).$$

Fortunately, this parity issue turns out to be the only such problem, in this particular example. That is, if we restrict $|V|$ to have a fixed parity and to grow slowly with n, then there will be a 0-1 law for the first-order properties of tournaments of this sort. (Here the atomic sentences are of the forms $x = u$ and $R(x, y) < R(u, v)$, where x and y are distinct variables and u and v are distinct variables, but other pairs of variables might coincide. It would make no difference if we also allow $R(x, y) = R(u, v)$, since it can be defined in terms of $<$.) We'll put a sketch of the argument into an appendix of this paper.

There is another possibility suggested by thesauri, namely the possibility of playing with the groups in the signa, and perhaps bringing some group theory to bear on these topics.

Q: That reminds me of something that occurred to me while you were proving the equivalence between thesauri and parametric conditions. Although thesauri allow arbitrary groups of permutations of $\{1, 2, \ldots, j\}$ for a j-ary signum, you used only the full symmetric group in the thesauri that simulate given parametric conditions. Combining this with the simulation in the other direction, you've shown, in effect, that any thesaurus is equivalent to one in which all the groups involved in the signa are full symmetric groups. I'd think this equivalence, involving only thesauri, should have a proof that involves only thesauri, rather than going from thesauri to parametric conditions and back.

A: Yes, one can also obtain the same result by working only with thesauri. Let $\langle R, j, V, G, h\rangle$ be a signum, and let S be the full symmetric group of all permutations of $\{1, 2, \ldots, j\}$. Then the signum $\langle R, j, V, G, h\rangle$ is essentially equivalent to a signum $\langle R_+, j, V_+, S, h_+\rangle$, where of course we have to define V_+ and h_+ (not R_+ because it's just a label).

Q: Before you start defining V_+ and h_+, please tell me what "essentially equivalent" means here.

A: It means that there is, for each set A, a canonical bijection between the possible interpretations in A for the two signa.

Q: OK. Now what are V_+ and H_+.

A: V_+ is the set of functions $f : S \to V$ that are G-equivariant in the sense that, for each $\pi \in G$ and each $\sigma \in S$,

$$f(\pi\sigma) = h(\pi)(f(\sigma)).$$

The action h_+ of S on V_+ is given by

$$(h_+(\sigma)(f))(\tau) = f(\tau\sigma).$$

It is straightforward to calculate that h_+ is a homomorphism[3] from S into the group of permutations of V.

Q: On the basis of my recent experience, I suppose that, if you had written $\sigma\tau$ instead of $\tau\sigma$, then h_+ would have been an anti-homomorphism.

A: That's right.

The essential equivalence between the two signa is obtained as follows. Given an interpretation $R^{\mathfrak{A}}$ of the original signum, i.e., a G-equivariant map $A^j \to V$, we obtain an interpretation $R_+^{\mathfrak{A}_+}$ of the new signum, an S-equivariant map $A^j \to V_+$ by letting $R_+^{\mathfrak{A}_+}(\vec{a})$ be the element of V_+ defined by

$$R_+^{\mathfrak{A}_+}(\vec{a})(\sigma) = R^{\mathfrak{A}}(a_{\sigma^{-1}(1)}, \dots, a_{\sigma^{-1}(j)}).$$

The transformation in the other direction takes any S-equivariant map $R_+^{\mathfrak{A}_+} : A^j \to V_+$ to the G-equivariant map $R^{\mathfrak{A}} : A^j \to V$ defined by

$$R^{\mathfrak{A}}(\vec{a}) = (R_+^{\mathfrak{A}_+}(\vec{a}))(1),$$

where 1 is the identity permutation. Of course, there's a lot to be checked here: that the two transformations are well-defined and that they're inverse to each other. That should be a good exercise for you.

Q: OK, I'll check it when I have some spare time. But, in view of this equivalence, why did you define thesauri in terms of arbitrary permutation groups G rather than just the full symmetric group?

A: We weren't aware of the equivalence until recently. But it seems worthwhile to retain the generality of arbitrary groups, since some forms of symmetry are more naturally expressed using groups smaller than the full symmetric group. Notice also that the conversion from a thesaurus using arbitrary groups to one using full symmetric groups makes the value sets V considerably larger and more complicated, especially if the original groups were small.

[3] Category theorists would describe the operation sending any G-set (V, h) to the S-set (V_+, h_+) as the right adjoint of the forgetful functor from S-sets to G-sets.

Appendix

We outline here the proof of the 0-1 law for the generalized tournaments described above. In fact, we prove somewhat more, namely the 0-1 law for first-order sentences in certain two-sorted structures. Here is the precise formulation.

Define (for the purposes of this appendix only) a *generalized tournament* to be a finite two-sorted structure $\mathfrak{A} = (A, V, <, \pi, R)$, where A and V interpret the two sorts, where $<$ is a linear ordering of V, where π is the unique order-reversing permutation of V, and where R is a function $A^2 \to V$, subject to the requirements that

- $R(a, b) = \pi(R(b, a))$ for all $a, b \in A$ and
- $|V|$ is odd.

The main reason for the second of these requirements is that, as we saw above, the parity of $|V|$ is definable and must therefore be fixed (at least almost surely) if we are to have a 0-1 law. We chose "odd" rather than "even" so that the first requirement makes sense even when $a = b$. For even $|V|$, a similar argument would apply, but we would have to either make R a partial function defined only when the arguments are distinct, or modify the first requirement to accommodate some convention for the case of equal arguments.

We say that a generalized tournament $\mathfrak{A}$ as above has *size* (n, v) if $|A| = n$ and $|V| = v$. For a first-order sentence θ in the language of generalized tournaments, we define the (n, v)-probability of θ as the proportion of models of θ among generalized tournaments of size (n, v). We say that θ has *asymptotic probability* 1 if, for each $\varepsilon > 0$, there exist $M \in \mathbb{N}$ and $\delta > 0$ such that, whenever $M < v < n^\delta$, then the (n, v)-probability of θ is at least $1 - \varepsilon$. Note that our definition of asymptotic probability incorporates the requirement that v is small compared to n, namely $v < n^\delta$; just how small this is (i.e., the choice of δ) depends on θ and ε. This definition of asymptotic probability is intended specifically for use with generalized tournaments; we do not propose it for more general contexts, though we expect that something similar would be appropriate in greater generality.

Theorem 1. *Let θ be any sentence in the first-order language of generalized tournaments. Then one of θ and $\neg\theta$ has asymptotic probability 1.*

Proof. Consider the first-order theory $\mathcal{RGT}$ (which stands for "random generalized tournaments"), in the language of generalized tournaments, given by the following axioms.

- R is a binary function from A to V.
- $<$ is a linear ordering of V.

- V has a first element and a last element.
- Each element of V but the first (resp. last) has an immediate predecessor (resp. successor).
- V is infinite (i.e., infinitely many axioms, saying $|V| > k$ for each $k \in \mathbb{N}$).
- π is an order-reversing permutation of V.
- $R(x, y) = \pi(R(y, x))$.
- π has a fixed point.
- The extension axioms

$$(\forall \vec{x})(\forall \vec{u})\,[D(\vec{x}) \rightarrow (\exists y) \bigwedge_{i=1}^{k} (y \neq x_i \wedge R(x_i, y) = u_i)]$$

  where the variables $\vec{x} = x_1, \ldots, x_k$ and y range over the first sort and $\vec{u} = u_1, \ldots, u_k$ over the second. As before, $D(\vec{x})$ is the formula saying that all the components x_i are interpreted as distinct elements.

We computed above that each extension axiom has asymptotic probability 1. Each of the axioms saying $|V| > k$ has asymptotic probability 1; just take $M > k$ in the definition of asymptotic probability. The remaining axioms of $\mathcal{RGT}$ have asymptotic probability 1 for the trivial reason that they are true in all generalized tournaments.

Just as with the usual (one-sorted) notion of asymptotic probability, one sees that any conjunction of finitely many sentences of asymptotic probability 1 also has asymptotic probability 1 (if there are k conjuncts and if $\varepsilon > 0$ is given, then just take the largest of the relevant M's and the smallest of the δ's for the conjuncts, using ε/k in place of ε) and that any logical consequence of a sentence of asymptotic probability 1 also has asymptotic probability 1. Therefore, the theorem will follow if we can show that the theory $\mathcal{RGT}$ is complete.

We prove completeness by presenting a winning strategy for Duplicator in an Ehrenfeucht-Fraïssé game of an arbitrary but specified length (i.e., number of rounds) r, between two models of $\mathcal{RGT}$. The essential part of Duplicator's work takes place in V. Here, Duplicator uses a familiar strategy for discrete linear orders with endpoints. The involution π and the A of the models are handled by suitable bookkeeping devices, described below in terms of imaginary pebbles.

For convenient reference, we first summarize the usual strategy for the Duplicator in the r-round Ehrenfeucht-Fraïssé game on two discrete linear orders *without* endpoints. In order to win, Duplicator must (by definition of the game) ensure that corresponding pebbles are ordered the same way in both models. His winning strategy is to ensure that, in addition, the distances between corresponding pebbles are not too different. Specifically, if, after m moves, two pebbles are at a distance $\leq 2^{r-m}$ on one model, then the corresponding pebbles on the other model are the same distance apart. Distances greater than 2^{r-m} need not be matched exactly, though of course if the distance between two pebbles is $> 2^{r-m}$ on one model then the distance between the corresponding pebbles on the other model will also be $> 2^{r-m}$. The key to the proof is that, because the distance 2^{r-m}, below which matching is required, decreases by a factor 2 from each move to the next, Duplicator can always move so as to maintain the required matching. By doing so, he ensures that he wins the game.

For infinite discrete linear orders *with* endpoints, Duplicator's strategy is the same except that he pretends that, already at the start of the game, there is a pair of corresponding pebbles on the first elements of the models and another pair of corresponding pebbles on the last elements. (In fact, the strategy doesn't really need that the models are infinite; it suffices that they have large enough cardinalities so that the Duplicator's matching requirements are satisfied by the initial imaginary pebbles.)

With these preliminaries, we now present Duplicator's strategy for models of $\mathcal{RGT}$. It involves a more elaborate scheme of imaginary pebbles. At the start of the game, Duplicator should imagine pebbles already placed on the first, last, and middle elements of the V sort in both models. Here "middle element" means the unique element fixed by π. In addition, during the course of the game, whenever a pebble is on an element $q \in V$, Duplicator should imagine an associated pebble at $\pi(q)$. Furthermore, whenever pebbles are on two elements a and b of A, Duplicator should imagine an associated pebble at $R(a, b)$.

As in the proof for plain linear orderings, Duplicator must ensure that corresponding pebbles on the V sorts of the two models are ordered the same way, and he will, in addition, voluntarily ensure that sufficiently small distances between corresponding pebbles match exactly. But now, "sufficiently small" does not mean $\leq 2^{r-m}$ (after m rounds of an r-round game) but rather $\leq 2^{r^2-m^2}$. The reason for this change will become clear in a moment, but for now notice that, like 2^{r-m}, this new boundary of smallness, $2^{r^2-m^2}$, decreases by at least a factor 2 at every move.

It is clear that, if Duplicator can follow this strategy, then doing so guarantees that he wins. It remains to show that Duplicator always has a move that maintains the required matching of distances. The proof of this is in two cases, depending on which sort Spoiler adds a pebble to.

Suppose first that Spoiler puts a new pebble on the V sort of one of the two models. Then, exactly as in the proof for plain linear orders, Duplicator can find a suitable spot for the corresponding pebble on the other model, because the distance below which matching is required has decreased by at least a factor 2. The new pebbles played in this round give rise to a new pair of imaginary pebbles, located at π of the real pebbles. But the required matching of distances between these imaginary pebbles and the played pebbles at earlier moves is automatic because π preserves distances. We must also consider the distance between the new real pebble and the new imaginary pebble, but since these locations are related by π they lie on opposite sides of the middle of V. Their distance is twice the distance from either of them to the (initially imagined) pebble at the middle. Since the distance to the middle pebble is matched (if small), so is the distance between the real and imaginary new pebbles. This completes the proof for the case where Spoiler's new pebble is on the V part.

Now suppose that Spoiler puts a new pebble on the A sort of one of the models. This results in many new imaginary pebbles on the V part. If there were already m moves before the current one, then there could be as many as m pebbles already in the A part, say at elements $a_1, \dots, a_m$. The new pebble, say at b, gives rise to as many as $2m$ new imaginary pebbles, at the elements $R(a_i, b)$ and (their π-images) $R(b, a_i)$. In this situation, Duplicator should proceed as follows. Pretend that, instead of a single move of Spoiler producing all these imaginary pebbles, there were $m + 1$ moves, that during the first m of these moves Spoiler put pebbles on the elements $R(a_i, b)$ one at a time (resulting in imaginary pebbles at $R(b, a_i)$), and that at the last of the $m+1$ moves Spoiler put the (real) pebble at b. Let Duplicator pretend to respond to the first of these m moves as described in the preceding paragraph, i.e., let him put imaginary new pebbles in the appropriate places. At each step, the maximum distance below which he can maintain exact matching decreases by a factor 2, so in these m moves, it has decreased by 2^m. Fortunately, this is still large enough, because $2^{r^2-m^2}$ (from before these moves) decreased by a factor 2^m is still larger than $2^{r^2-(m+1)^2}$, the required distance for matching after these moves. (This is of course why we used $2^{r^2-m^2}$ rather than 2^{r-m}. Our choice is not the smallest, since it could accommodate a decrease by a factor 2^{2m+1}, but it seems to be the simplest.) Finally, after all the imaginary pebbles have been placed in V, Duplicator must find a place for his real pebble in A. It together with previous pebbles must produce, as values of R, the elements of V bearing the imaginary pebbles just placed. Fortunately, an extension axiom guarantees the existence of an appropriate element to receive this pebble. □

Bibliographic Remark: Two useful survey papers on 0-1 laws, Oberschelp's [6] and Compton's [3], were not treated kindly by the printing process. The title of [3], though given correctly in the table of contents of the book in which it appears, had "0-1" deleted on the first page of the paper itself. As a result, "0-1" is also missing in the MathSciNet entry and perhaps elsewhere. In [6], some of the pages were printed out of order, and numbered in the printed order rather than the intended order. Fortunately, all the pages are present and can be found in correct order by a local search.

References

[1] Andreas Blass and Yuri Gurevich, "Strong extension axioms and Shelah's zero-one law for choiceless polynomial time," *J. Symbolic Logic* 68 (2003) 65–131.

[2] Andreas Blass and Frank Harary, "Properties of almost all graphs and complexes," *J. Graph Theory* 3 (1979) 225–240.

[3] Kevin Compton, "0–1 laws in logic and combinatorics," *Algorithms and Order* (Proc. NATO Advanced Study Institute, Ottawa, 1987) ed. I. Rival, Kluwer (1989) 353–383.

[4] Ronald Fagin, "Probabilities on finite models," *J. Symbolic Logic* 41 (1976) 50–58.

[5] Y. V. Glebskii, D. I. Kogan, M. I. Liogon'kii, and V. A. Talanov, "Range and degree of realizability of formulas in the restricted predicate calculus," *Kibernetika* (Kiev) 1969, no. 2, 17–28; English translation *Cybernetics* 5 (1969) 142–154.

[6] Walter Oberschelp, "Asymptotic 0-1 laws in combinatorics," *Combinatorial Theory* (Proc. Conf. Schloss Rauischholzhausen, 1982) ed. D. Jungnickel and K. Vedder, Springer-Verlag, Lecture Notes in Mathematics 969 (1982) 276–292.