Asymptotic probabilities of languages with generalized quantifiers

Guy Fayolle, Stéphane Grumbach, Christophe Tollu

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Asymptotic probabilities of languages with generalized quantifiers

Guy FAYOLLE
Stéphane GRUMBACH
Christophe TOLLU

N° 1916
Mai 1993

PROGRAMME 1

Architectures parallèles,
bases de données,
réseaux et systèmes distribués

Rapport de recherche

1993

Probabilités Asymptotiques de Langages avec des Quantificateurs Généralisés*

Guy Fayolle Stéphane Grumbach Christophe Tollu

I.N.R.I.A.
Rocquencourt BP 105
78143 Le Chesnay
France
guy.fayolle@inria.fr
stephane.grumbach@inria.fr

L.I.P.N.
URA 1507, Université Paris-Nord
93430 Villetaneuse
France
ct@lipn.univ-paris13.fr

Résumé

Nous étudions l'effet sur le comportement asymptotique des énoncés, de l'ajout de certaines familles de quantificateurs généralisés à la logique du premier ordre. Tous nos résultats sont établis pour des langages sans variable libre à l'intérieur des quantificateurs généralisés. Pour une classe $\mathcal{K}$ de structures finies fermée par isomorphisme, le quantificateur $Q_{\mathcal{K}}$ est dit fortement monotone si le problème d'appartenance à la classe est preservé pour une forme souple d'extensions. Notre premier théorème (loi 0/1 pour la logique du premier ordre avec un ensemble quelconque de quantificateurs fortement monotones) généralise un critère connu pour montrer que presque aucun graphe ne satisfait une propriété donnée. Nous montrons aussi une loi 0/1 pour la logique du premier ordre avec des quantificateurs de Härtig (quantificateurs d'équicardinalité) et une loi limite pour un fragment de la logique du premier ordre avec des quantificateurs de Rescher (exprimant des inégalités entre les cardinalités). Les preuves de ces deux derniers résultats font appel à l'énumération combinatoire standard et à des techniques plus sophistiquées d'analyse complexe. Nous montrons aussi que la loi 0/1 n'est pas vérifiée pour une extension de la logique du premier ordre avec des quantificateurs de Härtig si les restrictions syntaxiques mentionnées ci-dessus sont relachées. Nous obtenons donc la meilleure borne supérieure pour l'existence de loi 0/1 avec des quantificateurs de Härtig.

*Cet article est paru dans *Proc. of IEEE Logic in Computer Science*, Montreal 93.

Asymptotic Probabilities of Languages with Generalized Quantifiers

Guy Fayolle   Stéphane Grumbach

I.N.R.I.A.
Rocquencourt BP 105
78153 Le Chesnay
France
guy.fayolle.stephane.grumbach@inria.fr

Christophe Tollu

L.I.P.N.
URA 1507, Université Paris-Nord
93430 Villetaneuse
France
ct@lipn.univ-paris13.fr

May 3, 1993

Abstract

We study the impact of adding certain families of generalized quantifiers to first-order logic (FO) on the asymptotic behavior of sentences. All our results are stated and proved for languages disallowing free variables in the scope of generalized quantifiers. For a class $\mathcal{K}$ of finite structures closed under isomorphism, the quantifier $Q_{\mathcal{K}}$ is said to be strongly monotonic, sm, if membership in the class is preserved under a loose form of extensions. Our first theorem (0/1 law for FO with any set of sm quantifiers) subsumes a previous criterion for proving that almost no graphs satisfy a given property [BH79]. We also establish a 0/1 law for FO with Härtig quantifiers (equicardinality quantifiers) and a limit law for a fragment of FO with Rescher quantifiers (expressing inequalities of cardinalities). The proofs of these last two results combine standard combinatorial enumerations with more sophisticated techniques from complex analysis. We also prove that the 0/1 law fails for the extension of FO with Härtig quantifiers if the above syntactic restriction is relaxed. We therefore get the best upper bound for the existence of a 0/1 law for FO with Härtig quantifiers.

1 Introduction

The idea of extending first-order logic by means of generalized quantifiers dates back to the work of Mostowski [Mos57] on *cardinality quantifiers*, which was an attempt to remedy the fact that key notions of modern mathematics, such as the notion of a finite set or the notion of an uncountable set, were not first-order definable over the class of all (either finite or infinite) structures. In Mostowski's stride, miscellaneous quantifiers, inspired by probabilistic or topological concepts, came to light. A decade later, Lindström [Lin66] gave a very general definition of a quantifier, allowing practically any class $\mathcal{K}$ of structures to be used for defining a new quantifier $Q_{\mathcal{K}}$ that captures membership in that class. Since then, the study of languages with added quantifiers has been an important line of research of *abstract model theory* [BF85].

Meanwhile, *finite model theory* had emerged as an important research area [Fag90]. The steadily growing interest of logicians in finite structures was a consequence of the strengthened connections between logic and computer science. Researchers rapidly realized that first-order logic (FO) was not tuned properly for this new challenge. In particular, FO lacks any form of recursion mechanism that reveals necessary to define usual properties of finite structures. For the last two decades, a considerable amount of work has been achieved, in the context of finite model theory, on logics whose expressive power surpasses FO's: Gurevich and Shelah [GS86], among others, investigated and compared various fixed-point extensions of first-order logic, and Kolaitis and Vardi [KV90b, KV92b] undertook a careful examination of infinitary languages. Most of the work on extended logics over finite structures was related to important problems of descritive complexity. Very recently, generalized quantifiers have been studied in the realm of finite structures [KV92a, Hel92].

The restriction to finite structures also enabled the design and development of specific methods, among which 0/1 laws appear as central. This line of research was initiated by Fagin [Fag76] and Glebski *et alii* [GKLT69] who independently proved the following startling result: given any FO sentence φ, if all structures of size n are considered equiprobable, then the limit, as $n \to \infty$, of the probability that φ is satisfied by a random structure of size n, always exists and is equal to either 0 or 1. Languages enjoying such a property are said to have a 0/1 law. By now, the 0/1 law has been shown to hold for numerous extensions of first-order logic without functions or constants: fixed-point logics [BGK85, KV87], the infinitary logic with a finite number of variables $L^{\omega}_{\infty\omega}$ [KV90b] and some prenex classes of existential second-order logic [KV90a].

Our aim in the present paper is to study the asymptotic behavior of sentences of FO(**Q**), where **Q** is a set of generalized quantifiers. We first focus on *strongly monotonic* quantifiers. A quantifier Q, binding one formula, is strongly monotonic if it is defined by a class $\mathcal{K}$ of structures $\langle A, X\rangle$, with $X \subseteq A^k$ for some integer k, that is preserved under a loose form of extension: $\langle A, X\rangle \in \mathcal{K}$, $A \subseteq B$ and $X \subseteq Y$ entail $\langle B, Y\rangle \in \mathcal{K}$. Strong monotonicity is a neat framework for capturing numerous classes of graphs, among which planarity, n-colorability, and so on. We give a simple proof that the 0/1 law holds for the logics $\mathrm{FO}^*(\mathbf{Q})$ extending (without free variables in the scope of generalized quantifiers) FO by any family of strongly monotonic quantifiers. Hence, we yield a generalization of a criterion of Blass and Harary [BH79] for proving that almost no graphs satisfy some properties.

We then turn to extensions of FO with *counting*. To our knowledge, the only result on the asymptotic probabilities of extensions of first-order logic with counting can be found in [Kny90], where it is proved that, for a rational r such that $0 \leq r \leq 1$, if the asymptotic probability of a formula $\varphi(x)$ is different from r, sentences of the form: "there is at least a fraction r of the elements of the domain satisfying $\varphi(x)$" have a 0/1 law. The restriction on r is crucial. Indeed, the sentence expressing that "there is at least one half of the elements of the domain satisfying $P(x)$", where P is some unary predicate, has asymptotic probability $\frac{1}{2}$. Our results differ from Knyazev's in two respects: (i) we are concerned with families of quantifiers expressing equalities or inequalities of first-order definable relations of arbitrary arities, and (ii) our proofs rely upon radically different methods, namely the Laplace method and the saddle-point method for computing integrals. It should be noted that neither of our last two theorems can be deduced from Knyazev's main result.

Counting is an essential primitive in database query languages. Logical languages generally lack the ability to express counting, though it is very easy to count on any computational device [AV91]. 0/1 laws have been used in this context to get upper bounds on the expressive power of query languages. These results give a better understanding of the expressive power gained with counting primitives such as Härtig's and Rescher's quantifiers.

The paper is organized as follows. In the next section, we review the main definitions. Section 3 is devoted to strongly monotonic quantifiers. In Sections 4 and 5, two theorems are proved, which establish a 0/1 law (respectively a limit-law) for extensions of FO by means of the Härtig quantifiers (respectively the Rescher quantifiers).

2 Asymptotic Probabilities and Generalized Quantifiers

In this section, we introduce the basic concepts and tools that will be used throughout the paper.

Let $\langle R_1, \ldots, R_k\rangle$ be a purely relational schema and $\tau = \langle r_1, \ldots, r_k\rangle$ be its *similarity type*, i.e. the arity of R_i is r_i. $Struc_{<\omega}[\tau]$ and $Struc_n[\tau]$ respectively denote the class of finite τ-structures and the class of τ-structures of domain $n = \{0, \ldots, n-1\}$. If $\mathcal{L}[\tau]$ is a logic for $Struc_{<\omega}[\tau]$ and φ is a sentence of $\mathcal{L}[\tau]$, $\mu_n(\varphi)$ denotes the proportion of structures of $Struc_n[\tau]$ which satisfy φ:

$$\mu_n(\varphi) = \frac{|\{\mathcal{A} \in Struc_n[\tau] \mid \mathcal{A} \models \varphi\}|}{|Struc_n[\tau]|}.$$

The *asymptotic probability* $\mu(\varphi)$ of φ is the limit, if it exists, of $\mu_n(\varphi)$, as $n \to \infty$. A property is *almost surely (a.s.) true* (resp. *almost surely false*) if its asymptotic probability is 1 (resp. 0). If the asymptotic probability is defined for every sentence of $\mathcal{L}[\tau]$, $\mathcal{L}[\tau]$ is said to have a *limit law*. If, in addition, the asymptotic probability is either 0 or 1, $\mathcal{L}[\tau]$ is said to have a *(labelled) 0/1 law*.

In [Fag76], Fagin proved that there is a τ-structure over the set ω of all integers, called the *random countable τ-structure*, which plays a key role in the study of the asymptotic probabilities of FO[τ] sentences. In the sequel, $\Omega[\tau]$ will denote the random countable τ-structure. The following well-known property will turn out to be essential to the proofs of Theorems 4.1, 5.1.

Proposition 2.1 *[Fag76] For every $\ell \geq 1$, the number of equivalence classes of ℓ-tuples induced by the isomorphisms of $\Omega[\tau]$ is finite and equal to $N_\ell(\tau)$, which depends only upon ℓ and the similarity type τ. Furthermore,*

- *each equivalence class C is characterized on $\Omega[\tau]$ by a $FO[\tau]$ sentence without quantifier (a complete open description), $\Phi_C(x_1, \ldots, x_\ell)$, i.e. for each ℓ-tuple $\langle a_1, \ldots, a_\ell\rangle$ of integers,*

$$\langle a_1, \ldots, a_\ell\rangle \in C \text{ iff } \Omega[\tau] \models \Phi_C(a_1, \ldots, a_\ell);$$

- *each equivalence class C is* equiprobable, *i.e. for a random ℓ-tuple $\overline{a}$,*

$$\text{Prob}(\overline{a} \in C) = \frac{1}{N_\ell(\tau)};$$

- *each $FO[\tau]$ formula with ℓ free variables $\varphi(\overline{x})$ is* asymptotically equivalent *to a union of equivalence classes of ℓ-tuples, i.e. there exist classes $C_1, \ldots, C_p$ such that:*

$$\lim_{n\to\infty} \forall \overline{x}\,(\varphi(\overline{x}) \Leftrightarrow \bigvee_{1\leq i\leq p} \Phi_{C_i}(\overline{x})) = 1.$$

Let $\mathcal{K}$ be a class of τ-structures closed under isomorphism. We associate $\mathcal{K}$ with a *generalized quantifier* $Q_\mathcal{K}$, of type τ. FO($Q_\mathcal{K}$) is the logic extending FO by the following formation rule.

- If $\varphi_1, \ldots, \varphi_k$ are formulas and, for each $i \in \{1, \ldots, k\}$, $\overline{x}_i$ is a tuple of variables of arity r_i, then $Q_\mathcal{K}\,\overline{x}_1, \ldots, \overline{x}_k(\varphi_1(\overline{x}_1), \ldots, \varphi_k(\overline{x}_k))$ is a formula. Note that $Q_\mathcal{K}$ binds all occurrences of variables from $\overline{x}_i$ in φ_i.

The semantics of $Q_\mathcal{K}$ is defined by:

- $\mathcal{A} \models Q_\mathcal{K}\,\overline{x}_1, \ldots, \overline{x}_k(\varphi_1(\overline{x}_1), \ldots, \varphi_k(\overline{x}_k))$ iff $\langle A, \varphi_1^\mathcal{A}, \ldots, \varphi_k^\mathcal{A}\rangle \in \mathcal{K}$,
  where A is the domain of $\mathcal{A}$
  and $\varphi_i^\mathcal{A} = \{\overline{a} \in A^{r_i} \mid \mathcal{A} \models \varphi_i(\overline{a})\}$.

Obviously, for any set $\mathbf{Q}$ of generalized quantifiers, logics FO($\mathbf{Q}$) can be defined in the same way. In the following sections, we shall disallow the occurrence of free variables in generalized quantifiers and FO*($\mathbf{Q}$) will denote the sublanguage of FO($\mathbf{Q}$) obtained by replacing the above formation rule for $Q_\mathcal{K}$ with the following one.

- If $\varphi_1, \ldots, \varphi_k$ are formulas and, for each $i \in \{1, \ldots, k\}$, $\overline{x}_i$ is a tuple of variables of arity r_i and φ_i has exactly r_i free variables, then $Q_\mathcal{K}\,\overline{x}_1, \ldots, \overline{x}_k(\varphi_1(\overline{x}_1), \ldots, \varphi_k(\overline{x}_k))$ is a formula. The latter closed formula will be referred to as a *generalized atomic sentence*.

3 Strongly Monotonic Quantifiers

In this section, we introduce the notion of a strongly monotonic quantifier and prove that extensions of first-order logic with any family of strongly monotonic quantifiers have a 0/1 law.

Definition 3.1 *A quantifier $Q_\mathcal{K}$ is said to be* strongly monotonic *if for all τ-structures $\mathcal{A} = \langle A, R_1, \ldots, R_k\rangle$ and $\mathcal{B} = \langle B, R'_1, \ldots, R'_k\rangle$, if $\mathcal{A} \in \mathcal{K}$ and $\mathcal{A} \subseteq \mathcal{B}$ (i.e. $A \subseteq B$ and $R_i \subseteq R'_i$, for each $i \in \{1, \ldots, k\}$), then $\mathcal{B} \in \mathcal{K}$.*

There are plenty of natural examples of strongly monotonic quantifiers. The most obvious strongly monotonic quantifiers are the first-order existential quantifier $\exists$, and the counting quantifiers $\exists i\, x$, meaning "there are at least i distinct x's" [Imm86]. Consider now the class $\mathcal{K}_{\neg P}$ of non planar graphs. The quantifier $Q_{\mathcal{K}_{\neg P}}$ is obviously strongly monotonic; it is known ([Ber76], p. 21) that a graph is *planar* if it does not contain a subgraph—in the graph-theoretic sense—homeomorphic either to $K_{3,3}$ (the bi-partite 3,3-graph) or to K_5 (the 5-clique). The notion of *n-colorability* provides us with another example. If $\mathcal{K}_{\neg nC}$ is the class of graphs that are not n-colorable, then $Q_{\mathcal{K}_{\neg nC}}$ is a strongly monotonic quantifier. More generally, if a property of graphs has a forbidden subgraph characterization, its complement can be expressed by a strongly monotonic quantifier. It is now folklore in the theory of random graphs that such properties are satisfied by almost no graphs [BH79]: it is an easy consequence of the extension axioms that, for any graph H, almost every graph has an induced subgraph isomorphic to H. Theorem 3.1 below thus includes as subcases the results about the asymptotic probability of planarity, n-colorability, chordality, etc.

Let $\mathbf{Q}_{sm}$ be any set of strongly monotonic quantifiers.

Theorem 3.1 *FO*($\mathbf{Q}_{sm}$) has a (labelled) 0/1 law.*

Proof : Obviously, the set of a.s. true or a.s. false properties is closed under finite conjunctions and negations. So, since there is no free variable within the scope of generalized quantifiers, it is sufficient to prove that every generalized atomic sentence $\Phi : Q_\mathcal{K}\,\overline{x}_1, \ldots, \overline{x}_k(\varphi_1(\overline{x}_1), \ldots, \varphi_k(\overline{x}_k))$, with a strongly monotonic quantifier $Q_\mathcal{K}$, has asymptotic probability 0 or 1. If $\mathcal{K} = \emptyset$, then Φ is always false, and its probability is uniformly 0. Otherwise, let n be such that the probability that a τ-structure over n does not belong to $\mathcal{K}$ is p ($0 \leq p < 1$). Then, by a simple combinatorial argument and by the strong monotonicity, the probability that a τ-structure over m (with $m \geq 2n$) does not belong to $\mathcal{K}$, is at most p^2. Therefore, the ratio of τ-structures of $Struc_n[\tau]$ which do not belong to $\mathcal{K}$ tends to 0 as n tends to ∞, and the asymptotic probability of Φ is 1. □

The above proof actually gives a more accurate insight into the asymptotic probability than Theorem 3.1.

- The speed of convergence of the asymptotic probability of a sentence in $FO^*(\mathbf{Q}_{sm})$ is exponential.
- If $\mathcal{K}$ defines a non empty subclass, then the sentence Φ : $Q_{\mathcal{K}}\overline{x}_1, \dots, \overline{x}_k(\varphi_1(\overline{x}_1), \dots, \varphi_k(\overline{x}_k))$ has asymptotic probability 1.

It follows that the asymptotic probability of planarity and n-colorability of a graph is 0. Strong monotonicity thus gives a broader language for tackling with the problem of the asymptotic probability of important properties of relational structures.

4 Härtig Quantifiers

This section focuses on the asymptotic probabilities of sentences with Härtig quantifiers [Här65]. Härtig quantifiers state that two defined relations have the same cardinality. The Härtig quantifier $I_{r,s}$ is of type $\langle r, s\rangle$, i.e. it binds two formulas and r variables in the first one and s variables in the second one. The syntax of $FO^*(I_{r,s})$ is given by the new formation rule:

- if $\varphi(\overline{x})$ and $\psi(\overline{y})$ are two FO formulas such that the set of free variables of φ (resp. ψ) is $\overline{x}$, of length r (resp. $\overline{y}$, of length s), then $I_{r,s}\,\overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$ is a sentence of $FO^*(I_{r,s})$.

Its semantics is given by:

- $\mathcal{A} \models I_{r,s}\,\overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$ iff $|\varphi^{\mathcal{A}}| = |\psi^{\mathcal{A}}|$, where $\varphi^{\mathcal{A}} = \{\overline{a} \in A^r \mid \mathcal{A} \models \varphi(\overline{a})\}$ and $\psi^{\mathcal{A}} = \{\overline{a} \in A^s \mid \mathcal{A} \models \psi(\overline{a})\}$.

It is shown in [GT92] that, for any fixed relational schema, there is a sentence of $FO^*(I_{r,s})$ that is not expressible in terms of Härtig quantifiers of type $\langle r', s'\rangle$, with $\max\{r', s'\} < \max\{r, s\}$. Let $\mathbf{I}$ stand for the set of Härtig quantifiers of any type.

We next establish our main theorem on the asymptotic probabilities of sentences in $FO^*(\mathbf{I})$.

Theorem 4.1 *The asymptotic probability of every sentence of $FO^*(\mathbf{I})$ is defined and is either 0 or 1, i.e. $FO^*(\mathbf{I})$ has a labelled 0/1 law.*

Proof : Without loss of generality, we may restrict to generalized atomic sentences with quantifiers of the form $I_{r,r}$. Indeed, assume that $\overline{x} = \langle x_1, \dots, x_r\rangle$ and $\overline{y} = \langle y_1, \dots, y_s\rangle$ with $r < s$. The formula:

$$I\langle x_1, \dots, x_r\rangle, \langle y_1, \dots, y_s\rangle(\varphi(x_1, \dots, x_r), \psi(y_1, \dots, y_s))$$

is equivalent to:
$I\langle x_1, \dots, x_r, x_{r+1}, \dots, x_s\rangle, \langle y_1, \dots, y_s\rangle$

$$((\varphi(x_1, \dots, x_r) \wedge \bigwedge_{i=1}^{s-r} x_{r+i} = x_i), \psi(y_1, \dots, y_r)).$$

Now, Proposition 2.1 enables us to reformulate our problem in combinatorial terms. Indeed, for some finite sets I and J we have:

$$\lim_{n\to\infty} \forall \overline{x}\,(\varphi(\overline{x}) \Leftrightarrow \bigvee_{i \in I} \Phi_{C_i}(\overline{x})) = 1$$

and

$$\lim_{n\to\infty} \forall \overline{y}\,(\psi(\overline{y}) \Leftrightarrow \bigvee_{j \in J} \Phi_{C_j}(\overline{y})) = 1.$$

Therefore, the asymptotic behavior of

$$I\,\overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$$

is identical to the asymptotic behavior of

$$I\,\overline{x}, \overline{y}(\bigvee_{i \in I'} \Phi_{C_i}(\overline{x}), \bigvee_{j \in J'} \Phi_{C_j}(\overline{y})),$$

with $I' = I - J$ and $J' = J - I$. We set $\ell = N_r(\tau)$, $\ell_1 = card(I')$ and $\ell_2 = card(J')$. The equivalent combinatorial problem is the following:

> One randomly distributes n balls into ℓ equiprobable urns. For each pair $\langle \ell_1, \ell_2\rangle$ of integers such that $\ell_1 + \ell_2 \leq \ell$, let A_n denote the probability that the number of balls in the first ℓ_1 urns be equal to the number of balls in the next ℓ_2 ones. An easy computation yields:

$$A_n = \frac{1}{\ell^n} \sum_{\substack{p=0 \\ p \text{ even}}}^{n} \binom{n}{p} \binom{p}{p/2} (\ell_1 \ell_2)^{p/2} (\ell - \ell_1 - \ell_2)^{n-p}.$$

It is immediate that if the limit of A_n, as $n \to \infty$, exists and is equal to either 0 or 1, then $FO^*(\mathbf{I})$ has a labelled 0/1 law. Lemma 4.2 thereafter shows that this is indeed the case and gives also accurate information about the speed of convergence. □

Lemma 4.2 *As $n \to \infty$, $A_n \to 0$ at exponential speed, except when $\ell_1 = \ell_2$, in which case the speed of convergence is $O\left(\frac{1}{\sqrt{n}}\right)$.*

Proof : See the Appendix (section A.1). □

As an aside, the next proposition shows that the complexity of the decision problem for the asymptotic truth of a sentence of $FO^*(\mathbf{I})$ is as "low" as the complexity of the counterpart problem for FO [Gra83].

Proposition 4.3 *The decision problem for the value of the asymptotic probabilities of sentences of $FO^*(\mathbf{I})$ is PSPACE-complete.*

Proof : It relies on the fact that the value of the asymptotic probability of a sentence of $FO^*(\mathbf{I})$ depends only on the values of asymptotic probabilities of FO sentences. Indeed, a sentence of the form $I_{r,r}\,\overline{x},\overline{y}(\varphi(\overline{x}),\psi(\overline{y}))$ has asymptotic probability 0 (1) iff $\lim_{n\to\infty}\forall\overline{x}\,(\varphi(\overline{x})\Leftrightarrow\psi(\overline{x}))=0$ (1). □

If free variables can occur within the scope of a Härtig quantifier, the 0/1 law (and even the limit law) fails.

Theorem 4.4 *There is a sentence φ of $FO(\mathbf{I})$ whose probability $\mu_n(\varphi)$ does not have a limit.*

Proof : It is sufficient to show that the property of graphs, expressible in $FO(\mathbf{I})$, that says "there is (at least) one vertex adjacent to exactly the half of the other vertices" does not have an asymptotic probability. Indeed, while the sentence is obviously always false on graphs of even order, its probability does not tend to 0, as $n\to\infty$, on graphs of order $2n+1$. To see this last point, we shall rely on a general theorem given in [Bol85] (Chap. III, Theorem 1, p. 57), which asserts in particular the following

Proposition 4.5 *Let $\mathcal{G}(n,p)$ be the model consisting of all graphs with vertex set $\{1,\ldots,n\}$, in which the edges are chosen independently with probability p, which can be a function of n. Let $\epsilon>0$ be fixed and $\epsilon n^{\frac{-3}{2}}\le p\le 1-\epsilon n^{\frac{-3}{2}}$, let $k=k(n)$ be a natural number and set $\lambda_k(n)=nb(k;n-1,p)$, where*

$$b(j;n,p)=\binom{n}{j}p^j(1-p)^{n-j}.$$

Denote also by X_k the random variable representing the number of vertices of degree k in a random graph. One sees that $\lambda_k(n)$ is exactly the expectation of X_k. Then the following property holds :

$$\textit{If}\ \lim_{n\to\infty}\lambda_k(n)=\infty\ ,\ \textit{then}\ \lim_{n\to\infty}Prob(X_k\ge \ell)=1\ ,$$

for every fixed ℓ. □

Thus, we are in a position to apply Proposition 4.5, with $p=\frac{1}{2}$, for graphs of order $2n+1$. Using Stirling's formula, we have immediately

$$\lambda_n(2n+1)=(2n+1)\binom{2n}{n}4^{-n}\sim 2\sqrt{\frac{n}{\pi}}\to\infty\,.$$

The probability $\mu_n(\varphi)$ of the sentence φ of $FO(\mathbf{I})$, which we have considered, has no proper limit, but takes, as $n\to\infty$, (only) the two values 0 and 1. This concludes the proof of Theorem 4.4 □

The same technique can be applied to *similarity quantifiers* S_r, of type $\langle r,r\rangle$, which state that two defined relations are isomorphic, and not only of equal cardinalities (Härtig quantifiers). The syntax is similar to Härtig quantifiers with S in place of I. The semantics is defined as follows:

- $\mathcal{A}\models S_r\,\overline{x},\overline{y}(\varphi(\overline{x}),\psi(\overline{y}))$ iff $\langle A,\varphi^{\mathcal{A}}\rangle\cong\langle A,\psi^{\mathcal{A}}\rangle$.

Let $\mathbf{S}$ be the set of similarity quantifiers of all types. We define $FO^*(\mathbf{S})$ in the same way as $FO^*(\mathbf{I})$.

Corollary 4.6 *The asymptotic probability of every sentence of $FO^*(\mathbf{S})$ is defined and is either 0 or 1, i.e. $FO^*(\mathbf{S})$ has a labelled 0/1 law.*

The proof follows from Theorem 4.1. Indeed, if $\lim_{n\to\infty}\forall\overline{x}\,(\varphi(\overline{x})\Leftrightarrow\psi(\overline{x}))=0$, then the probability that $|\ \varphi^{\mathcal{A}}\ |=|\ \psi^{\mathcal{A}}\ |$ (which majorates the probability that $\langle A,\varphi^{\mathcal{A}}\rangle\cong\langle A,\psi^{\mathcal{A}}\rangle$) goes to 0 when $n\to\infty$. Otherwise, if $\lim_{n\to\infty}\forall\overline{x}\,(\varphi(\overline{x})\Leftrightarrow\psi(\overline{x}))=1$, then $\lim_{n\to\infty}S_r\,\overline{x},\overline{y}(\varphi(\overline{x}),\psi(\overline{y}))=1$.

5 Rescher Quantifiers

We now proceed to the study of a natural extension of $FO^*(\mathbf{I})$ including the possibility of testing inequalities of cardinalities of first-order definable relations. The corresponding quantifiers is referred to in the literature as the *Rescher quantifiers* $R_{r,s}$ [Res62], with $R_{r,s}$ of type $\langle r,s\rangle$. Here again, the syntax of $FO^*(R_{r,s})$ is given by the new formation rule:

- if $\varphi(\overline{x})$ and $\psi(\overline{y})$ are two FO formulas such that the set of free variables of φ (resp. ψ) is $\overline{x}$, of length r (resp. $\overline{y}$, of length s), then $R_{r,s}\,\overline{x},\overline{y}(\varphi(\overline{x}),\psi(\overline{y}))$ is a sentence of $FO^*(R_{r,s})$.

Its semantics is given by:

- $\mathcal{A}\models R_{r,s}\,\overline{x},\overline{y}(\varphi(\overline{x}),\psi(\overline{y}))$ iff $|\ \varphi^{\mathcal{A}}\ |\le|\ \psi^{\mathcal{A}}\ |$.

Obviously, the 0/1 law fails in this new context. If P is a unary predicate, the asymptotic probability of $R_{1,1}\, x, y(P(x), \neg P(y))$ is $\frac{1}{2}$. Nevertheless, the asymptotic behavior of a fragment of $\mathrm{FO}^{*}(\mathbf{R})$, where $\mathbf{R}$ stands for the set of Rescher quantifiers of all types, is fairly regular, as shown by the next theorem. Recall that a sentence of $\mathrm{FO}^{*}(\mathbf{R})$ of the form $R_{r,s}\, \overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$ is called a *generalized atomic* sentence.

Theorem 5.1 *The asymptotic probability of every* generalized atomic *sentence of $FO^{*}(\mathbf{R})$ is defined and its value ranges over* $\{0, \frac{1}{2}, 1\}$.

Proof : It goes along the same lines as the proof of Theorem 4.1. We consider generalized atomic sentences like $R_{r,r}\, \overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$. Once again, we reformulate our problem:

> One randomly distributes n balls into ℓ equiprobable urns. For each pair $\langle \ell_1, \ell_2 \rangle$ of integers such that $\ell_1 + \ell_2 \leq \ell$, let B_n denote the probability that the number of balls in the first ℓ_1 urns be lesser than the number of balls in the next ℓ_2 ones. An easy computation yields:

$$B_n = \frac{1}{\ell^n} \sum_{p=0}^{n} \binom{n}{p} (\ell - \ell_1 - \ell_2)^{n-p} \sum_{q=0}^{\lfloor \frac{p}{2} \rfloor} \binom{p}{q} \ell_1^{q}\, \ell_2^{p-q}.$$

Clearly, if $\lim_{n \to \infty} B_n$ always exists and belongs to a set $\mathbf{B}$, then the asymptotic probability of all sentences of the form $R_{r,r}\, \overline{x}, \overline{y}(\varphi(\overline{x}), \psi(\overline{y}))$ exists and belongs to $\mathbf{B}$. The next lemma explicitly describes $\mathbf{B}$, hence completing the proof of Theorem 5.1. □

Lemma 5.2 *As $n \to \infty$, $B_n \to 0$, $\frac{1}{2}$ or 1.*

Proof : See the Appendix (section A.2). □

The limit law of Theorem 5.1 can be easily generalized to sentences of $\mathrm{FO}^{*}(\mathbf{R})$ with one generalized atomic sub-sentence, or under some technical assumptions to sentences with several generalized atomic sub-sentences, with dyadic limit law. We conjecture that in the general case, $\mathrm{FO}^{*}(\mathbf{R})$ admits a dyadic limit law. Note that if the existence of the asymptotic probability for all sentences of $\mathrm{FO}^{*}(\mathbf{R})$ $[\tau]$ is established, the finiteness of the set of equivalence classes of ℓ-tuples over the random countable structure $\Omega[\tau]$ (Proposition 2.1) will ensure that the set of possible values is finite for each τ.

6 Conclusions

We undertook a study of the asymptotic truth of sentences for extensions of FO with generalized quantifiers, and came up with a characterization of the asymptotic behavior of two kinds of languages, corresponding on the one hand to strongly monotonic quantifiers and on the other hand to counting quantifiers. As for $\mathrm{FO}^{*}(\mathbf{I})$, the syntactic constraint (no free variables within the scope of Härtig quantifiers) we imposed on the formulas revealed essential: our counterexample to the 0/1 law for the full language $\mathrm{FO}(\mathbf{I})$ indicates the very place where the boundary of 0/1 laws is broken in the hierarchy of counting extensions of FO.

In most cases, the 0/1 law no longer holds for $\mathrm{FO}(\mathbf{Q})$ where $\mathbf{Q}$ is a (set of) generalized quantifier(s). This is the case for instance of partially ordered quantifiers, such as *Henkin quantifiers* [Hen61], which yield the same expressive power as existential second-order logic [BG86]. For example, the following sentence expresses the fact that the cardinality of the domain is even.

$$\begin{pmatrix} \forall x \exists v \\ \forall y \exists w \end{pmatrix} \begin{pmatrix} (x = y \Leftrightarrow v = w) \wedge \\ (x = w \Leftrightarrow y = v) \wedge \\ (v \neq x)) \end{pmatrix}.$$

Indeed, it states that there exist two unary functions f and g, which are equal and one-one; since the domain is finite, they define a bijection, which is involutive and has no fixpoint.

In the body of the paper, we did not consider languages of the form $\mathcal{L}^{*}(\mathbf{Q})$, $\mathcal{L}$ being a logic stronger than FO, or non uniform probability measures. It is clear that the asymptotic probabilities of sentences are similar if FO is replaced with FP, namely first-order logic plus a positive fixed-point operator. There are two basic reasons for that: (i) each FP formula with ℓ free variables is asymptotically equivalent to a union of equivalence classes of ℓ-tuples of the random countable structure Ω, thus ensuring that Proposition 2.1 applies to FP; (ii) all variables of an expression $Q_{r_1, \ldots, r_p} \overline{x}_1, \ldots, \overline{x}_p(\varphi_1, \ldots, \varphi_p)$ occurring in a formula of $\mathrm{FP}^{*}(\mathbf{Q}_{sm})$ or $\mathrm{FP}^{*}(\mathbf{I})$ are bound, so the fixed-point construct cannot "involve" a generalized quantifier. As for $\mathrm{L}^{\omega *}_{\infty\omega}(\mathbf{Q}_{sm})$, obviously, it does not have a 0/1 law, the parity of the domain is expressible in $\mathrm{L}^{\omega *}_{\infty\omega}(\mathbf{C})$, where $\mathbf{C}$ stands for the set all counting quantifiers $\exists i\, x$ [KV92a]. The asymptotic behaviour of $\mathrm{L}^{\omega *}_{\infty\omega}(\mathbf{I})$ is an open question.

Acknowledgements

The authors wish to thank Yuri Gurevich for fruitful discussions.

A Appendix

A.1 Proof of Lemma 4.2

We suppose that n is even. This restriction is harmless and the case n odd can be dealt with in an analogous way.

Define $\alpha = \ell_1/\ell$, $\beta = \ell_2/\ell$, $n = 2k$ and $p = 2u$. A_n can thus be rewritten as

$$A_n \equiv \tilde{A}_k = \sum_{u=0}^{k} \binom{2k}{2u} \binom{2u}{u} (\alpha\beta)^u \left[(1-\alpha-\beta)^2\right]^{k-u},$$

i.e.

$$\tilde{A}_k = (2k)! \sum_{u=0}^{k} \frac{(\alpha\beta)^u \left[(1-\alpha-\beta)^2\right]^{k-u}}{(u!)^2 \left[2(k-u)\right]!}.$$

Setting also

$$\tilde{A}_k \stackrel{\text{def}}{=} (2k)!\, B_k$$

$$B(z) \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} B_j z^{2j}, \text{ so that } \tilde{A}_k = \frac{d^{2k} B(z)}{dz^{2k}}\bigg|_{z=0},$$

and using the convolution character of $\tilde{A}_k$, one notices that $B(z)$ can be expressed in terms of the product of the two following series:

1. the classicaly so-called *modified* Bessel function

$$\mathbf{J}_0(ix) \equiv \mathbf{I}_0(x) = \sum_{j=0}^{\infty} \frac{1}{(j!)^2} \left(\frac{x}{2}\right)^{2j};$$

2. the series

$$F(z) \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} \frac{x^{2j}}{(2j)!} = \frac{1}{2}\left[e^x + e^{-x}\right].$$

This yields directly

$$B(z) = \mathbf{I}_0(2\sqrt{\alpha\beta}\, z) F[(1-\alpha-\beta)z].$$

We now use Sonine's integral representation [LS77]

$$\mathbf{I}_0(z) = \frac{1}{2i\pi} \int_{|\omega|=1} \frac{e^{\frac{iz}{2}(\omega - \frac{1}{\omega})}}{\omega} d\omega.$$

Then, by a symmetry argument with respect to z, it follows that

$$B(z) = \frac{1}{2i\pi} \int_{|\omega|=1} \frac{e^{z[i\sqrt{\alpha\beta}(\omega - \frac{1}{\omega}) + 1 - \alpha - \beta]}}{\omega} d\omega.$$

Finally, making the change of variable $\omega = e^{i\theta}$, we get

$$\tilde{A}_k = \frac{1}{\pi} \int_0^{\pi} \left[2\sqrt{\alpha\beta}\cos\theta + (1-\alpha-\beta)\right]^{2k} d\theta.$$

It is known that $\tilde{A}_k$ can be expressed directly in terms of the Legendre functions of first and second kind, according to the values of the parameters. It is nevertheless here more direct to apply the *saddle point* method, see e.g. [LS77], which produces an asymptotic expansion for integrals of the form

$$H(\lambda) = \int_a^b e^{\lambda f(x)} dx, \text{ as } \lambda \to \infty. \qquad (1)$$

We expedite first two borderline cases

i) $\alpha\beta = 0$, say for instance $\alpha = 0$. Then

$$\tilde{A}_k = (1-\beta)^{2k},$$

which tends to zero exponentially fast.

ii) $\alpha = \beta = 1/2$. Then

$$\tilde{A}_k = \frac{(2k)!}{4^k (k!)^2} \sim \frac{1}{\sqrt{k\pi}}\left[1 + O(k^{-1})\right],$$

by using Stirling's formula.

Assuming $\alpha\beta \neq 0$ and α, β not both equal to 1/2, we have

$$\tilde{A}_k = \frac{1}{\pi} \int_0^{\pi} e^{k\left[\log[2\sqrt{\alpha\beta}\cos\theta + (1-\alpha-\beta)]^2\right]} d\theta,$$

which is of the form 1. The behavior of $\tilde{A}_k$, as $k \to \infty$ depends on the maximum of the function

$$|2\sqrt{\alpha\beta}\cos\theta + (1-\alpha-\beta)|$$

and on the position of the point, say θ_0, $0 \leq \theta_0 \leq \pi$, where this maximum is attained. Here $\theta_0 = 0$ is an end of the interval of integration. Then we get approximately *one half* of a Gaussian integral, see [LS77], so that

$$\tilde{A}_k \sim \frac{\left[1 - (\sqrt{\alpha} - \sqrt{\beta})^2\right]^{2k}}{2} \sqrt{\frac{1 - (\sqrt{\alpha} - \sqrt{\beta})^2}{2k\pi\sqrt{\alpha\beta}}} \left[1 + O\left(\frac{1}{k}\right)\right],$$

showing again the convergence to 0 at exponential speed. The proof of the lemma is concluded. □

A.2 Proof of Lemma 5.2

We use the same notation as in the proof of lemma 4.2. It follows that

$$B_n = \sum_{p=0}^{n} \binom{n}{p} (1-\alpha-\beta)^{n-p}(\alpha+\beta)^p f_p\,,$$

where

$$f_p = \sum_{q=0}^{\lfloor \frac{p}{2} \rfloor} \binom{p}{q} \left(\frac{\alpha}{\alpha+\beta}\right)^q \left(\frac{\beta}{\alpha+\beta}\right)^{p-q}.$$

From Toeplitz lemma, see [Loe77], it follows that if f_p tends to a limit as $p \to \infty$, then B_p will tend to the same limit. It turns out that f_p can be expressed in an integral form, allowing to use the *saddle-point* method, see [LS77]. Introduce the cumulative distribution function of the binomial distribution

$$B(k;p,r) = \sum_{p=0}^{k} \binom{p}{q} r^q s^{p-q}\,,$$

where $0 \le k \le p$, $0 \le r, s$ and $r+s=1$. Then $f_p = B(\lfloor \frac{p}{2} \rfloor; p, r)$. It is sweet to remark that $B(k;p,r)$ admits the following integral representation, see [Fel70].

$$B(k;p,r) = (p-k)\binom{n}{p} \int_0^s t^{p-k-1}(1-t)^k\,dt\,.$$

Assume p is odd, for instance $p = 2u+1$, and set $s = \frac{\beta}{\alpha+\beta}$. Then

$$f_p \stackrel{\text{def}}{=} \tilde{f}_u = \frac{(2u+1)!}{(u!)^2} \int_0^{\frac{\beta}{\alpha+\beta}} t^u(1-t)^u\,dt\,.$$

As in the proof of lemma 4.2, we are in a position to apply the saddle point method to the above integral, which yields three different cases, according to the position of the point $t = 1/2$ (which is the point where the function $\log t(1-t)$ reaches its maximum) with respect to s :

1. $\beta > \alpha$. Denoting by $I(u)$ the integral coming in $\tilde{f}_u$, we obtain

   $$I(u) \sim \frac{4^{-u}}{2}\sqrt{\frac{\pi}{u}}\left[1 + O(u^{-1})\right]\,,$$

   so that, after applying Stirling's formula to evaluate the coefficient $\frac{\tilde{f}_u}{I(u)}$,

   $$\tilde{f}_u \sim 1 + O\left(\frac{1}{u}\right).$$

2. $\beta < \alpha$. Here, on $[0,s]$, the function $\log t(1-t)$ has its maximum at $t = s$, so that $\tilde{f}_u \sim$

   $$\left[\frac{4\alpha\beta}{(\alpha+\beta)^2}\right]^u \frac{2^{3/2}\alpha\beta}{\sqrt{\pi u(\alpha^2+\beta^2)}(\alpha+\beta)}\left[1 + O\left(\frac{1}{u}\right)\right],$$

   which tends to 0 exponentially fast.

3. $\beta = \alpha = 1/2$. Now we obtain a quantity equivalent to the area of a *semi-gaussian* and

   $$\tilde{f}_u \sim \frac{1}{2} + O\left(\frac{1}{u}\right).$$

By mimicking the above derivation, one can see that the same limits hold in the case $p = 2u$. Hence f_p, and consequently B_p, have, when $p \to \infty$, one of the limits 0, 1, or $\frac{1}{2}$, depending on the respective values of α and β. The lemma is proved. □

References

[AV91] S. Abiteboul and V. Vianu. Generic computation and its complexity. In *Proc. 23rd ACM Symp. on Theory of Computing* (1991) 209–219.

[Ber76] C. Berge. *Graphs and hypergraphs.* North-Holland mathematical library, Vol. 6 (North-Holland, Amsterdam, 2nd edition, 1976).

[BF85] J. Barwise and S. Feferman, editors. *Model theoretic logics.* Perspectives in Mathematical Logic (Springer Verlag, Berlin, 1985).

[BG86] A. Blass and Y. Gurevich. Henkin quantifiers and complete problems. *Annals of Pure and Applied Logic* **32** (1986) 1–16.

[BGK85] A. Blass, Y. Gurevich, and D. Kozen. A zero-one law for logic with a fixed-point operator. *Information and Control* **67** (1985) 70–90.

[BH79] A. Blass and F. Harary. Properties of almost all graphs and complexes. *Journal of Graph Theory* **3** (1979) 225–240.

[Bol85] B. Bollobas. *Random Graphs* (Academic Press, 1985).

[Fag76] R. Fagin. Probabilities on finite models. *Journal of Symbolic Logic* **41**(1) (1976) 50–58.

[Fag90] R. Fagin. Finite model theory—a personal perspective. In S. Abiteboul and P. Kanellakis, editors: *Proc. 3rd Int. Conf. on Database Theory.* Lecture Notes in Computer Science, Vol. 470 (Springer-Verlag, Berlin, 1990) 3–24.

[Fel70] W. Feller. *An introduction to Probability Theory and its Applications*, volume I. (Wiley, 3rd edition, 1970).

[GKLT69] Y. Glebskii, D. Kogan, M. Liogon'kiĭ, and V. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetica* **2** (Kiev, 1969) 17–28. English translation. *Cybernetics* **5** (1969) 142–154.

[Gra83] E. Grandjean. Complexity of the first order theory of almost all structures. *Information and Control* **52** (1983) 180–204.

[GS86] Y. Gurevich and S. Shelah. Fixed-point extensions of first order logic. *Annals of Pure and Applied Logic* **32** (1986) 265–280.

[GT92] S. Grumbach and C. Tollu. Query languages with counters. In J. Biskup and R. Hull, editors: *Proc. 4th Int. Conf. on Database Theory.* Lecture Notes in Computer Science. Vol. 646 (Springer-Verlag, Berlin, 1992) 124–139.

[Här65] H. Härtig. Über einen Quantifikator mit zwei Wirkungsbereichen. In L. Kalmar, editor: *Colloquium on Foundations of Mathematics, Mathematical Machines and their Applications* (Budapest, 1965) 31–36.

[Hel92] L. Hella. Logical hierarchies in PTIME. In *Proc. 7th IEEE Symp. on Logic in Computer Science* (1992) 360–368.

[Hen61] L. Henkin. Some remarks on infinitely long formulas. In *Infinitistic methods. Proc. Symp. Foundations of Mathematics, Warsaw* (Pergamon Press, London, 1961) 167–183.

[Imm86] N. Immerman. Relational queries computable in polynomial time. *Information and Control* **68** (1986) 86–104.

[Kny90] V. Knyazev. Zero-one law for an extension of first-order predicate language. *Kybernetika* **2** (Kiev, 1990) 110–113. English translation. *Cybernetics* **26** (1990) 292–296.

[KV87] P. Kolaitis and M.Y. Vardi. The decision problem for the probabilities of higher-order properties. In *Proc. 19th ACM Symp. on Theory of Computing* (1987) 425–435.

[KV90a] P. Kolaitis and M.Y. Vardi. 0/1 laws and decision problems for fragments of second-order logic. *Information and Computation* **87** (1990) 302–338.

[KV90b] P. Kolaitis and M.Y. Vardi. 0/1 laws for infinitary logic. In *Proc. 5th IEEE Symp. on Logic in Computer Science* (1990) 156–167.

[KV92a] P. Kolaitis and J. Väänänen. Generalized quantifiers and pebble games on finite structures. In *Proc. 7th IEEE Symp. on Logic in Computer Science* (1992) 348–359.

[KV92b] P. Kolaitis and M.Y. Vardi. Fixpoint logic vs. infinitary logic in finite-model theory. In *Proc. 7th IEEE Symp. on Logic in Computer Science* (1992) 46–57.

[Loe77] M Loeve. *Probability Theory I.* (Springer-Verlag, Berlin, 1977).

[LS77] M. Lavrentiev and B. Shabat. *Méthode de la Théorie d'une Fonction d'une Variable Complexe.* (MIR, Moscou, 1977).

[Lin66] P. Lindström. First order predicate logic with generalized quantifiers. *Theoria* **32** (1966) 186–195.

[Mos57] A. Mostowski. On a generalization of quantifiers. *Fundamenta Mathematicæ* **44** (1957) 12–36.

[Res62] N. Rescher. Plurality quantification. *Journal of Symbolic Logic* **27** (1962) 373–374.

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR.1916