On the Expressive Power of Counting

Stéphane Grumbach, Christophe Tollu

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the Expressive Power of Counting

Stéphane Grumbach , Christophe Tollu

N° 2330

Septembre 1994

PROGRAMME 1

Architectures parallèles,

bases de données,

réseaux et systèmes distribués

apport de recherche

1994

ISSN 0249-6399

On the Expressive Power of Counting

Stéphane Grumbach * , Christophe Tollu **

Programme 1 — Architectures parallèles, bases de données, réseaux et systèmes distribués
Projet Verso

Rapport de recherche n° 2330 — Septembre 1994 — 35 pages

Abstract: We investigate the expressive power of various extensions of first-order, inductive, and infinitary logic with counting quantifiers. We consider in particular a LOGSPACE extension of first-order logic, and a PTIME extension of fixpoint logic with counters. Counting is a fundamental tool of algorithms. It is essential in the case of unordered structures. Our aim is to understand the expressive power gained with a limited counting ability. We consider two problems: (i) unnested counters, and (ii) counters with no free variables. We prove a hierarchy result based on the arity of the counters under the first restriction. The proof is based on a game technique that is introduced in the paper. We also establish results on the asymptotic probabilities of sentences with counters under the second restriction. In particular, we show that first-order logic with equality of the cardinalities of relations has a 0/1 law.

Key-words: query languages, first-order logic, inductive logic, infinitary logic, counting, generalized quantifiers, finite model theory, descriptive complexity

(Résumé : tsvp)

*stephane.grumbach@inria.fr
**LIPN-Institut Galilée-URA 1507, Université Paris-Nord, 93430 Villetaneuse, France, ct@ura1507.univ-paris13.fr

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Le Pouvoir d'Expression des Compteurs

Résumé : Nous étudions le pouvoir d'expression d'extensions de la logique du premier ordre, de la logique inductive et de la logique infinitaire avec des quantificateurs de comptage. Nous considérons en particulier une extension LOGSPACE de la logique du premier ordre et une extension PTIME de la logique point-fixe. Compter est une opération fondamentale de l'algorithmique. C'est essentiel dans le cas de structures non ordonnées. Notre but est de comprendre le pouvoir d'expression résultant d'une capacité limitée de comptage. Nous considérons deux problèmes : (i) les compteurs non imbriqués, et (ii) les compteurs sans variable libre. Nous montrons que l'arité des compteurs induit une hiérarchie stricte dans le cas de la première restriction. La preuve repose sur des techniques de jeu, définies dans l'article. Nous montrons aussi des résultats sur les probabilités asymptotiques des énoncés avec compteurs sous la seconde restriction. Nous montrons en particulier que la logique du premier-ordre avec un quantificateur d'équicardinalité des relations admet une loi 0/1.

Mots-clé : langages de requêtes, logique du premier ordre, logique inductive, compteurs, quantificateurs généralisés, théorie des modèles finis, complexité descriptive

1 Introduction

Counting is a fundamental operation of numerous algorithms. Counters constitute also an essential primitive of query languages. In relational databases, practical query languages, such as SQL, provide counters as built-in functions of the languages. Counters map (database or defined) relations to integers. They are of great importance from a practical point of view. Moreover, counters raise challenging theoretical problems. Logical languages generally lack the ability to express counting, though it is very easy to count on any computational device [AV91].

Finite model theory offers an elegant paradigm to study the expressive power of counting primitives. It emerged as an important research area [Gur88, Fag93]. 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 [GS85], among others, investigated and compared various fixpoint extensions of first-order logic, and Kolaitis and Vardi [KV90b, KV92a] undertook a careful examination of infinitary languages. Most of the work on extended logics over finite structures was related to important problems of descriptive complexity.

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 al.* [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 $n \rightarrow \infty$, 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. Fagin's proof is particularly interesting. He showed the existence of a countable structure, Ω, called the random countable structure, which is the unique (up to isomorphism) model of an infinite set of axioms, the extension axioms, and such that there is a transfer property, that is for every first-order sentence ϕ, the asymptotic probability of ϕ is 1 iff $\Omega \models \phi$. By now, the 0/1 law and the transfer property have been shown to hold for numerous extensions of first-order logic without functions or constants: fixpoint logics [BGK85, KV87], the infinitary logic with a finite number of variables $L^{\omega}_{\infty,\omega}$ [KV92b] and some prenex classes of existential second-order logic [KV90a].

Counting mechanisms have been the focus of a great interest in classical logic in the past. 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]. Very recently, generalized quantifiers have been studied in the realm of finite structures [KV92c, Hel92, DH94].

Our aim in the present paper is to study the impact of restricted counting mechanisms on various logics, such as first-order logic, fixpoint logic, and infinitary logic with a finite number of variables. We focus on two-sorted logics, with a sort *Domain* unordered and a sort *Integer* with a linear order. Relations are defined over the first sort only. The counters map relations to integers (which are therefore never stored in the relations and only used as selection arguments). It is possible to check if the cardinality of a relation is equal (resp. less than or equal, greater than or equal) to a given integer, or to compare cardinalities of relations (with $=$ or $\leq$). All these expressions define generalized quantifiers, some of them being well-known in the literature such as Härtig's quantifiers [Har65] and Rescher's quantifiers [Res62].

We first consider an extension of first-order logic with counters, FO+C. It is shown that it has a rather limited expressive power and can be evaluated in LOGSPACE data complexity. The extension of fixpoint logic with counting, FO+IFP+C, enjoys rather nice properties. It can be restricted to unary counters (indeed, polyadic counters can be simulated with nested monadic counters and the fixpoint operator). Moreover, every PTIME property on the values of the counters can be expressed in FO+IFP+C. This is due to the fact that FO+IFP characterizes PTIME on ordered inputs [Var82, Imm86]. These two aspects differ strongly from the first-order extension. This language has been studied by other authors, and shown to be particularly robust by [Ott92]. Finally, we consider an extension of infinitary logic with a finite number of variables with counting, $\mathrm{L}^{\omega}_{\infty\omega}$+C.

We first consider the expressive power of the languages when restricted to unnested counters. We denote these restrictions by FO+$\mathrm{C_u}$, FO+IFP+$\mathrm{C_u}$, and $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C_u}$. We define extensions of Ehrenfeucht-Fraïssé games [Fra54, Ehr61], which characterize the languages FO+$\mathrm{C_u}$, and $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C_u}$. These games differ from the game presented in [IL90], since here the counters are polyadic and unnested, while in this previous game, the counters are monadic and nested. The game is used to prove a hierarchy result based on the arity of the counters. We prove that FO+$\mathrm{C}^k_{\mathrm{u}} \subset$ FO+$\mathrm{C}^{k+1}_{\mathrm{u}}$, where FO+$\mathrm{C}^k_{\mathrm{u}}$ is the sublogic of FO+$\mathrm{C_u}$ restricted to counters of arity at most k. The same holds for FO+IFP+$\mathrm{C}^k_{\mathrm{u}}$, and $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C}^k_{\mathrm{u}}$.

We then turn to the asymptotic probabilities of yet another restriction of the counting logics, where free variables are disallowed in counting expressions. To our knowledge, the only results, on the asymptotic probabilities of extensions of first-order logic with counting can be found in [Kny90, GT92, FGT93]. It is proved in [Kny90] 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 $\frac{1}{2}$.

We establish a 0/1 law for FO with Härtig quantifiers (equicardinality quantifiers) and a limit law for a fragment of FO with Rescher quantifiers (majority quantifiers). 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. The results carry over for fixpoint logic and infinitary logic with a finite number of variables.

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 introduce extensions of first-order, fixpoint, and infinitary logic with counters. Section 3 is devoted to the games characterizing the versions of the previous logics with unnested counters. The hierarchy result is proved in Section 4. The asymptotic probabilities are presented in Section 5. Finally, we mention some open problems in the last section.

2 Logics with Counters

In this section, we define languages extending first-order logic, inductive logic and infinitary logic with finitely many variables with counters. An expression of the form $count(\overline{x}, \varphi(\overline{x}))$ is interpreted as the integer giving the cardinality of the relation defined by the formula $\varphi(\overline{x})$. These languages will be introduced in as general a setting as possible and we shall investigate a few of their structural properties. The present section will serve as a reference for the thorough examination (in terms of expressive power and asymptotic behavior) of the sublogics we shall deal with in Sections 3 through 5.

2.1 First-Order Logic with Counters

Let τ be a fixed purely relational signature, i.e. a (finite) sequence $\langle R_1, \ldots, R_p\rangle$ of relation symbols (excluding constant or function symbols). From now on, we shall refer to first-order logic over τ as FO$[\tau]$. For a fixed integer $k \geq 1$, we define the language FO+C^k, extending FO with counters giving the cardinality of definable relations of arity $\leq k$.

Let $\mathcal{B} = \langle B, =, R_1^{\mathcal{B}}, \ldots, R_p^{\mathcal{B}}\rangle$ be a τ-structure of finite domain B, such that $|B| = n$. Let $\mathcal{B}_k = \mathcal{B} \sqcup \mathcal{N}_k$ denote the two-sorted structure which is the union of $\mathcal{B}$ and $\mathcal{N}_k$, where

$$\mathcal{N}_k = \langle \mathbf{n^k+1} \cup \{\infty\}, 0, \ldots, n^k, \infty, =, \leq\rangle.$$

The domain of $\mathcal{N}_k$ is the union of the set of the first n^k+1 natural numbers and ∞. The relations $=$ and $\leq$ have their standard meaning. The elements of B constitute the *Domain* sort, whereas the domain of $\mathcal{N}_k$ constitutes the *Integer* sort.

Let $\mathrm{Str}_{<\omega}[\tau]$ denote the class of all τ-structures with a finite domain and $\mathrm{Str}^k_{<\omega}[\tau]$ denote the class of two-sorted structures $\mathcal{B}_k$, for $\mathcal{B}$ ranging over $\mathrm{Str}_{<\omega}[\tau]$. We call FO$^{\#}[\tau]$ the expansion of the first-order language for $\mathrm{Str}^k_{<\omega}[\tau]$ with an *Integer* constant symbol $\underline{n}$ for every $n \in \omega$ (notice that $\mathrm{FO}^{\#}[\tau]$ does not depend upon k). In the remainder of the paper, in order to get rid of mentioning the type of a variable when writing a formula, we shall always use the letters $x, y, z, t, \ldots$ for the variables of *Domain* sort, and the letters $i, j, k, l, \ldots$ for the variables of *Integer* sort.

The following definitions of *counting terms* and $\mathrm{FO{+}C}^k[\tau]$-*formulas* are mutually recursive.

Definition 2.1 Let $\varphi(\overline{x}, \overline{y}, \overline{\imath})$ be a formula of $\mathrm{FO{+}C}^k[\tau]$, such that $\overline{x} = \langle x_1, \ldots, x_\ell \rangle$ is an ℓ-tuple ($\ell \leq k$) of *Domain* variables **free** in φ. We define an *Integer* term t by:

$$t \equiv count(\overline{x}, \varphi(\overline{x}, \overline{y}, \overline{\imath})).$$

t is called a *counting term* of arity ℓ. If Free(φ) is the set of free variables of φ, then Free(t)= Free(φ) $- \{\overline{x}\}$.

Definition 2.2 The set of *formulas* of $\mathrm{FO{+}C}^k[\tau]$ is the smallest set containing $\mathrm{FO}^{\#}[\tau]$, and closed under the following constructs:

- If φ and ψ are formulas, x a *Domain* variable and i an *Integer* variable, then $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\exists x \varphi$, $\forall x \varphi$, $\exists i \varphi$ and $\forall i \varphi$ are formulas.
- If C_1 and C_2 are two *Integer* terms (constants, variables or counting terms), then $C_1 = C_2$ and $C_1 \leq C_2$ are formulas.

Let us now show how to interpret a formula of $\mathrm{FO{+}C}^k[\tau]$ in a structure $\mathcal{B}_k \in \mathrm{Str}^k_{<\omega}[\tau]$. The interpretation of variables or constants of both types is straightforward (for $m \in \omega$, $\underline{m}$ is interpreted as m if $m \leq n^k$, and as ∞ otherwise), so we concentrate on the case of the counting terms:

- Let Σ be a sort-preserving assignment of the free variables of the counting term of arity ℓ $t = count(\overline{x}, \varphi(\overline{x}, \overline{y}, \overline{\imath}))$ in $\mathcal{B}_k$. The interpretation of t in $\mathcal{B}_k$ is the element of *Integer* sort defined by:

$$t[\mathcal{B}_k, \Sigma] = |\{\overline{a} \in B^\ell \mid (\mathcal{B}_k, \Sigma) \models \varphi(\overline{a})\}|.$$

The interpretation of a formula of $\mathrm{FO{+}C}^k[\tau]$ in a structure $\mathcal{B}_k$ is now standard. Note that the formulas of $\mathrm{FO{+}C}^k[\tau]$ without free variables of *Integer* sort (the ones we are chiefly interested in) are easily interpreted in $\mathcal{B} \in \mathrm{Str}_{<\omega}[\tau]$.

Remark : From now on, we shall feel free to drop the mention of the underlying signature whenever it is obvious from the context.

Example 2.1 Let $\tau = \{G\}$, where G is a binary relation symbol, and $k \geq 2$. The τ-structures are directed graphs. The following formula of FO+C$^k[\tau]$,

$$count(\langle x, y\rangle, G(x, y) \wedge G(y, x)) = \underline{\mathrm{m}}.$$

is true in a directed graph $\mathcal{G}$ iff the number of symmetric edges is m.

Example 2.2 [GO93] Consider two equivalence relations E_1 and E_2. The FO+C^1 sentence $\psi(E_1, E_2)$ expresses that E_1 and E_2 have the same number of equivalence classes of size i, for every $i \in \omega$, thus that E_1 and E_2 are isomorphic.

$$\psi(E_1, E_2) \ : \forall i(count\,(x, count(y, E_1(x, y)) = i) = count\,(x, count(y, E_2(x, y)) = i)).$$

So far, we have always worked with counters of bounded arity: we had fixed an integer $k \geq 1$. It is of course possible to consider counters of any arity.

Definition 2.3

$$\mathrm{FO} + \mathrm{C}[\tau] = \bigcup_{k=1}^{\infty}(\mathrm{FO} + \mathrm{C}^k[\tau]).$$

The only difficulty is to interpret the formulas of $\mathrm{FO} + \mathrm{C}[\tau]$ (possibly with free variables of *Integer* sort) uniformly in a structure of $\bigcup_{k=1}^{\infty} \mathrm{Str}^k_{<\omega}[\tau]$. Let $\psi \in \mathrm{FO} + \mathrm{C}[\tau]$, and k be the greatest arity of the counters occurring in ψ; then ψ is interpreted in $\mathrm{Str}^k_{<\omega}[\tau]$. As for the formulas of $\mathrm{FO} + \mathrm{C}[\tau]$ without free variables of *Integer* sort, they are interpreted in $\mathrm{Str}_{<\omega}[\tau]$.

2.2 First-Order Logic with Primitive Recursive Counting

The expressive power of FO+C^k is still very weak. We thus now enrich the arithmetic constructs over the sort *Integer* by means of all LOGSPACE-computable arithmetic functions of any arity and co-arity 1. That can be properly achieved thanks to the global functions introduced by Gurevich [Gur83].

Definition 2.4 A *global function* F of arity k and co-arity ℓ is a mapping assigning to every initial segment of the integers $\mathcal{I} = \langle\{0, \ldots, n\}, 0, End^{\mathcal{I}}, =, \leq\rangle$ a (local) arithmetic function $F^{\mathcal{I}} \ : \ \mathbf{n}^k \to \mathbf{n}^\ell$. *Primitive recursive* global functions are defined as the closure, under the usual composition and primitive recursion schemata, of a small set of initial global functions including the global constant functions with respective values 0 and $End^{\mathcal{I}}$, i.e. $|I| = n + 1$, the global projection functions and the global successor functions, whose corresponding local functions are defined only for tuples which are not greatest with respect to the lexicographic order.

The following result was proved by Gurevich.

Theorem 2.1 [Gur83] The global primitive recursive functions exactly characterize the LOGSPACE-computable ones. □

The initial segment of the natural numbers which appears in a structure $\mathcal{B}_k$ is slightly different from the structure $\mathcal{I}$ occurring in Definition 2.4. It is harmless to set $F^{\mathcal{I}}(\infty) = \infty$ and $F^{\mathcal{I}}(m) = \infty$ when $F^{\mathcal{I}}$ is undefined for m.

Definition 2.5 The set of *primitive recursive counting expressions* is the closure of the set of *Integer* terms (constants, variables or counting terms) under the following construct:

- If $C_1, \ldots, C_k$ are *primitive recursive counting expressions* and F is a global primitive recursive function of arity k and co-arity 1, then $F(C_1, \ldots, C_k)$ is a *primitive recursive counting expression.*

This definition yields a new counting extension of FO.

Definition 2.6 FO+C^k_{pr} is defined exactly in the same way as FO+C^k with the primitive recursive counting expressions (corresponding to local functions over $\langle \mathbf{n^k+1} \cup \{\infty\}, 0, \ldots, n^k, \infty, =, \leq \rangle$) in place of the counting terms.

$$\mathrm{FO} + \mathrm{C}_{pr} = \bigcup_{k=1}^{\infty} (\mathrm{FO} + \mathrm{C}^k_{pr}).$$

We are now able to express usual queries involving counting, such as *Even*, which answers "yes" iff the underlying structure is of even cardinality:

Example 2.3 *Even* $: \exists i(count(x, x = x) = i + i)$.

We now give lower and upper bounds on the expressivity and complexity of FO+C_{pr}.

Proposition 2.2 FO $\subset$ FO+C $\subset$ FO+C_{pr} $\subset$ LOGSPACE.

Proof : FO is obviously properly contained in FO+C since the latter enables to compare the respective cardinalities of two definable relations. FO+C is properly contained in FO+C_{pr}: indeed, as FO cannot express all LOGSPACE-computable relations over ordered inputs, there are properties of the integers which can be defined in FO+C_{pr} but not in FO+C (*Even* is an example of such a property). The proof that for instance *Even* is not definable in FO with ordered inputs can be done using Ehrenfeucht-Fraïssé games. The containment FO+C_{pr} $\subseteq$ LOGSPACE is trivial since the primitive recursive global functions exactly encodes the LOGSPACE-computable ones ([Gur83], see Proposition 2.1 above). It is a proper inclusion: indeed, FO+C_{pr} is a sublogic of the inductive logic with counting introduced by Immerman [Imm86] and which eventually turned out to be unable to express a LOGSPACE-computable property of graphs ([CFI89], see below Proposition 2.3). □

Remark : One could allow more powerful functions over the sort *Integer*. The only constraint is to stay within a reasonable complexity class. For instance, if we had chosen to define

FO+C_{rec} by replacing primitive recursive counting expressions with recursive ones, we would have been led to:

$$\mathrm{FO} \subset \mathrm{FO} + \mathrm{C}_{pr} \subseteq \mathrm{FO} + \mathrm{C}_{rec} \subset PTIME.$$

Indeed, the algebra of recursive global functions has been proved by Gurevich to capture all the PTIME-computable ones [Gur83].

2.3 Fixpoint Logic with Counters

We now explore the possibility to extend FO both with counters and an operator enabling to define new relations by monotone induction up to a fixpoint. Fixpoint extensions of FO with counting have been investigated in [GT92], but the following presentation owes much to [Ott92].

In sharp contrast to the case of the family $\{\mathrm{FO} + \mathrm{C}^k\}_{1 \leq k < \omega}$, we shall no longer have to introduce polyadic counters: indeed, thanks to the fixpoint operator and the possibility to use arbitrarily deep nestings of counters, counters of arity k will be definable in terms of monadic ones[1]. As a consequence, formulas will be interpreted in structures of $\mathrm{Str}^1_{<\omega}$.

First of all, we consider the following formation rule:

- Let $\varphi(\overline{x}, \overline{\imath}, X)$ be a formula whose free variables of sort *Domain* include $\overline{x} = \langle x_1, \ldots, x_r \rangle$ and whose free variables of sort *Integer* include $\overline{\imath} = \langle i_1, \ldots, i_s \rangle$. X is a (possibly mixed) relational variable (not belonging to τ) of arity $\langle r, s \rangle$. Then the expression:

$$\psi \equiv [\mathrm{IFP}_{\overline{x}, \overline{\imath}, X}\, \varphi(\overline{x}, \overline{\imath}, X)](\overline{x}, \overline{\imath})$$

  also is a formula. One has: Free(ψ)=Free(φ)$-\{X\}$.

The semantics of such a formula is similar to the one of formulas of FO+IFP. Let Σ be a sort-preserving interpretation in $\mathcal{B}_1 \in \mathrm{Str}^1_{<\omega}$, of all the free variables of φ except for $\overline{x}$, $\overline{\imath}$ and X. φ induces an operator F_φ on the powerset of $B^r \times \{0, \ldots, n\}^s$,

$$P \mapsto \{(\overline{a}, \overline{m}) \in B^r \times \{0, \ldots, n\}^s \mid (\mathcal{B}_1, \Sigma) \models \varphi(\overline{a}, \overline{m}, P)\}.$$

The interpretation of $\psi \;:\; [\mathrm{IFP}_{\overline{x}, \overline{\imath}, X}\, \varphi(\overline{x}, \overline{\imath}, X)](\overline{x}, \overline{\imath})$ in $(\mathcal{B}_1, \Sigma)$ is the least fixpoint of the monotone operator that assigns $\mathrm{G}(P) = P \cup \mathrm{F}_\varphi(P)$ to P. Thus

$$(\mathcal{B}_1, \Sigma) \models \psi(\overline{a}, \overline{m}) \quad \Leftrightarrow \quad (\overline{a}, \overline{m}) \in \bigcup_{j<\omega} G^j_\varphi(\emptyset).$$

The definition of the concept of (monadic) counting terms $t = count(x, \varphi(x, \overline{y}, \overline{\imath}))$ now involves fixpoint formulas and we get:

Definition 2.7 FO+IFP+C is the closure of FO$^\#$ under the two schemata of Definition 2.2 and the following one:

[1] We may nonetheless take the liberty to use polyadic counters to simplify the writing of formulas.

- If $\varphi(\overline{x}, \overline{\imath}, X)$ is a formula of FO+IFP+C, then so is $[\mathrm{IFP}_{\overline{x},\overline{\imath},X}\,\varphi(\overline{x}, \overline{\imath}, X)](\overline{x}, \overline{\imath})$.

Example 2.4 [IL90, GO93] The method of *stable colorings* of graphs provides a PTIME graph-canonization algorithm for almost all graphs. It is not difficult to show that the stable coloring of a graph is definable in FO+IFP+C (for details, we refer to [IL90, CFI89]).

Remark: FO+IFP+C already appeared in the literature ([CFI89, IL90, Ott92, GO93]: the above presentation is taken from [Ott92] and [GO93]). As already stressed at the beginning of the section, counters of arity ≥ 2 would not increase the expressive power of FO+IFP+C; for a concise justification of this fact, the reader is referred to [GO93], where the extension of Datalog with counting is also considered[2]. In Sections 3 through 5, we shall sometimes focus on restrictions of FO+IFP+C (unnested counters, restrictions on free variables within the scope of a counter) which will ruin this specific feature of FO+IFP+C.

Grädel and Otto [Ott92, GO93] have carried out an in-depth investigation of the expressive power of FO+IFP+C and shown its robustness with respect to alternative definitions of fixpoint logic with counting and to complexity theory. In particular, they observe that the above definition of FO+IFP+C is equivalent to the one enriching FO+IFP with all counting quantifiers $\exists^{\geq m} x$ $(m \geq 1)$ meaning "there exist at least m distinct x's such that..."

It is obvious that FO+IFP+C can only express PTIME-computable properties: indeed, in the presence of a binary predicate always interpreted as a linear order on the structures (which is the situation for the sort *Integer*), a property is definable in FO+IFP if and only if is computable in polynomial time on a deterministic Turing machine [Imm86, Var82]. It had been conjectured by Immerman [Imm86] that FO+IFP+C would capture all PTIME-computable properties of graphs. Unfortunately, as proven by Cai, Fürer and Immerman in [CFI89], this conjecture fails dramatically, since there is a LOGSPACE-computable property of graphs that is not definable in FO+IFP+C. Nonetheless, it is obviously more powerful a language than FO+IFP. We thus naturally get:

Proposition 2.3 FO+IFP $\subset$ FO+IFP+C $\subset$ PTIME. □

2.4 Infinitary logic with counters

In the sequel, we shall sometimes be concerned with infinite formulas with a finite number of distinct variables. From a computer science point of view, such languages lack an effective syntax. Nonetheless, they constitute an efficient means to analyze recursive extensions of FO and their relation to first-order logic on certain classes of structures, as proved for instance by Kolaitis and Vardi [KV92a] or Dawar, Lindell and Weinstein [DLW94]. We now give a short reminder of the main definitions concerning infinitary logic.

Definition 2.8 Let τ be a relational signature. The set $\mathrm{L}_{\infty\omega}[\tau]$ of infinitary formulas is the smallest set of expressions containing $\mathrm{FO}[\tau]$ and satisfying:

[2] Datalog is the language whose programs consist of sets of definite Horn clauses without function symbols. The relations between Datalog and FO+IFP are now folklore in the Database literature.

- If φ is an infinitary formula, then so is $\neg\varphi$;
- If φ is an infinitary formula and v a variable, then $\exists v\varphi$ and $\forall v\varphi$ are infinitary formulas;
- If Φ is a set of infinitary formulas, then $\bigwedge \Phi$ and $\bigvee \Phi$ are infinitary formulas.

Remarks: (i) It is worth noticing that infinite strings of quantifiers are not allowed. (ii) The semantics of infinite formulas is straightforwardly inspired from the semantics of first-order formulas: for example, $\bigwedge \Phi$ simply means the conjunction of all formulas in Φ. (iii) As far as finite structures are concerned $\mathrm{L}_{\infty\omega}$ obviously is much too powerful a language, since every class of finite structure is definable by a formula of $\mathrm{L}_{\infty\omega}$. That is why it is necessary to constrain the formulas of $\mathrm{L}_{\infty\omega}$ so as to keep its expressivity within more reasonable limits.

Definition 2.9 The formulas of $\mathrm{L}^k_{\infty\omega}$ are the formulas of $\mathrm{L}_{\infty\omega}$ with at most k distinct variables (either free or bound). The formulas of $\mathrm{L}^\omega_{\infty\omega}$ are the formulas of $\mathrm{L}_{\infty\omega}$ with a finite number of distinct variables (either free or bound):

$$\mathrm{L}^\omega_{\infty\omega} = \bigcup_{k<\omega} \mathrm{L}^k_{\infty\omega}.$$

Example 2.5 [KV92b] Connectivity of finite graphs is expressible in $\mathrm{L}^3_{\infty\omega}$: there is an FO formula $p_n(x,y)$ using at most three variables x, y and z asserting that there is a path of length $\leq n$ from x to y. Let G be the edge relation, we have:

$$p_1(x,y) \equiv G(x,y),$$

and

$$p_{n+1}(x,y) \equiv p_n(x,y) \vee \exists z[G(x,z) \wedge \exists x(x = z \wedge p_n(x,y))].$$

So we have:

$$Conn\ :\ \forall x \forall y (\bigvee_{n=1}^{\infty} p_n(x,y)).$$

Remark : $\mathrm{L}^\omega_{\infty\omega}$ is strictly more powerful than FO+IFP, since it can express non-recursive classes of structures. Just consider the formula:

$$\forall x \forall y (\bigvee_{n \in P} p_n(x,y)),$$

where P is a non-recursive set of integers.

We now turn to infinitary logic with counting. Instead of adding to the formation rules of $\mathrm{L}^\omega_{\infty\omega}$ the formation rule for the counting terms, we adopt an equivalent definition, which facilitates the statements of interesting propositions. In fact, the proof (sketched in [GO93]) of Proposition 2.4 below gives a canonical way of unwinding every sentence of FO+IFP+C (thus containing counting terms) into an infinite formula with counting quantifiers and a finite number of variables.

Definition 2.10 We call *infinitary logic with counting and k variables* (and we write $\mathrm{L}^k_{\infty\omega} + C$) the language obtained by extending $\mathrm{L}^k_{\infty\omega}$ with the set of counting quantifiers $\exists^{\geq m}$ for all $m \in \omega$.

We call *infinitary logic with counting* (and we write $\mathrm{L}^\omega_{\infty\omega} + C$) the language obtained by extending $\mathrm{L}^\omega_{\infty\omega}$ with the set of counting quantifiers $\exists^{\geq m}$ for all $m \in \omega$:

$$\mathrm{L}^\omega_{\infty\omega} + C = \bigcup_{k<\omega} \mathrm{L}^k_{\infty\omega} + C.$$

Obviously, FO+IFP+C $\subset L^\omega_{\infty\omega} + C$ (a strict containment, since all properties expressible in FO+IFP+C are recursive). Grädel and Otto [GO93] gave an elegant characterization of the formulas of $L^\omega_{\infty\omega} + C$ which are equivalent to a formula of FO+IFP+C.

Proposition 2.4 [GO93] FO+IFP+C is equivalent to the sublanguage of $L^\omega_{\infty\omega} + C$ consisting of all formulas of the form:

$$\bigvee_{n\in\omega} (\exists^{\geq m} x(x = x) \wedge \neg\exists^{\geq m+1} x(x = x) \wedge \varphi_n),$$

for sequences $(\varphi_n)_{n\in\omega}$ in some $L^k_{\omega\omega} + C$ (i.e. FO+$\{\exists^{\geq m}\}_{m\in\omega}$ with at most k distinct variables) such that the mapping $n \mapsto \varphi_n$ is PTIME constructible. □

The main interest of this result is that it is in sharp contrast to the case of FO+IFP itself: indeed, the language consisting of all formulas $\bigvee_{n\in\omega}(\exists^{\geq m} x(x = x) \wedge \neg\exists^{\geq m+1} x(x = x) \wedge \varphi_n)$, for PTIME constructible families $(\varphi_n)_{n\in\omega}$ in some $L^k_{\omega\omega}$ (i.e. FO with at most k variables) is strictly more powerful than FO+IFP [GO93].

3 Games for logics with counters

This section is devoted to a combinatorial characterization, in terms of two-player games with perfect information, of the elementary equivalence with respect to some sublanguages of FO+C_{pr} and $\mathrm{L}^\omega_{\infty\omega}$+C. Games have long revealed a key tool in the investigation of expressiveness of various languages either on finite or infinite structures. Games have been introduced to characterize elementary equivalence with respect to first-order logic [Ehr61, Fra54], infinitary logic with finitely many variables [Bar77], and extensions of the latters by means of generalized quantifiers [IL90, KV92c]. Our games are very much inspired from the games introduced in [IL90, CFI89].

3.1 First-Order Counting Games

In this and the following sections, we shall restrict ourselves to the study of the counting languages with *no nesting of counters*. Before we give precise definitions, we introduce a notion that is a mere generalization of the classical concept of quantifier-depth. It will play a key role in the main result of this section. The following two definitions are mutually recursive.

Definition 3.1 Let τ be a fixed signature. The *quantifier depth* $d(C)$ of a primitive recursive counting expression C is defined as follows:

- $d(C) = 0$ if C is an *Integer* variable or constant;
- $d(count(\overline{x}, \varphi(\overline{x}, \overline{y}, \overline{\imath})) = d(\varphi) + \mathrm{arity}(\overline{x})$;
- $d(F(C_1, \ldots, C_k)) = \mathrm{Max}\{d(C_i) \mid 1 \leq i \leq k\}$ if F is a primitive recursive global function of arity k and co-arity 1.

Definition 3.2 Let φ be a formula of FO+C. The *quantifier depth* $d(\varphi)$ of φ is defined by induction on the structure of φ:

- $d(\varphi) = 0$ if φ is an atomic formula of FO;
- $d(\exists i \varphi) = d(\forall i \varphi) = d(\varphi)$ if i is an *Integer* variable;
- $d(\neg\varphi) = d(\varphi)$;
- $d(\varphi \wedge \psi) = d(\varphi \vee \psi) = \mathrm{Max}\{d(\varphi), d(\psi)\}$;
- $d(\exists x \varphi) = d(\forall x \varphi) = d(\varphi) + 1$ if x is a *Domain* variable appearing free in φ;
- $d(C_1 = C_2) = d(C_1 \leq C_2) = \mathrm{Max}\{d(C_1), d(C_2)\}$ if C_1 and C_2 are primitive recursive counting expressions.

Definition 3.3 Let FO+$\mathrm{C}_{\mathrm{u}}^k$ denote the sublogic of FO+C_{pr}^k obtained by restricting the formulas to the ones with no nesting of counters, FO+$\mathrm{C}_{\mathrm{u},\ell}^k$ the sublogic of FO+$\mathrm{C}_{\mathrm{u}}^k$ obtained by restriction to the formulas with at most ℓ distinct variables (either free or bound) of sort *Domain*, and FO+$\mathrm{C}_{\mathrm{u},\ell,r}^k$ the sublogic of FO+$\mathrm{C}_{\mathrm{u},\ell}^k$ obtained by restriction to the formulas of quantifier depth less than or equal to r.

Although the constraints we impose on the formulas of FO+$\mathrm{C}_{\mathrm{u}}^k$ weaken its expressive power, there are numerous properties of graphs that are definable in the sublogics of FO+$\mathrm{C}_{\mathrm{u}}^k$, as shown by the next example.

Example 3.1 A graph G is *regular* if all the vertices have the same degree. $Reg(G) \in$ FO+$\mathrm{C}_{\mathrm{u},2,1}^1$ is a formula expressing that G is regular:

$$Reg(G) \ : \ \exists i \forall x \ count(y, G(x,y) \vee G(y,x)) = i.$$

A graph G is *Eulerian* if there exists a Euler cycle, i.e. a cycle passing through each edge exactly once. Finite connected undirected graphs are Eulerian if and only if all vertices have even degrees, which can be expressed by $Euler(G) \in$ FO+$\mathrm{C}_{\mathrm{u},2,1}^1$:

$$Euler(G) \ : \ \forall x \exists i \ count(y, G(x,y)) = 2i.$$

We now define an equivalence relation, associated to the formulas of FO+$\mathrm{C}_{\mathrm{u},\ell}^k$, over the class of finite relational structures.

Definition 3.4 Let $\mathcal{A}$ and $\mathcal{B}$ be two finite τ-structures. If both structures satisfy exactly the same sentences of FO+$\mathrm{C}_{\mathrm{u},\ell,r}^k$, they are said to be $(k;\ell,r)$-equivalent and we write $\mathcal{A} \equiv_{\ell,r}^k \mathcal{B}$.

If both structures satisfy exactly the same sentences of FO+$\mathrm{C}_{\mathrm{u},\ell}^k$, they are said to be $(k;\ell)$-equivalent and we write $\mathcal{A} \equiv_{\ell}^k \mathcal{B}$.

The game we present now will precisely characterize the above notion of equivalence.

Definition 3.5 We define the game $\mathcal{C}^k_{\ell,r}$ (also referred to as the game $\mathcal{C}^k_\ell$ of length r, i.e. with r moves) on structures $\mathcal{A}$ and $\mathcal{B}$. There are two players, **I** and **II**, and for each variable x_i, $i = 1, \ldots, \ell$, a pair of x_i-pebbles. We distinguish between two kinds of moves: the *classical moves*, identical to the ones occurring in an Ehrenfeucht-Fraïssé game for first-order logic [Ehr61, Fra54],

1. **I** picks up a pair of x_i-pebbles and puts one of the pebble on an element of the domain of one of the two structures, say A; **II** replies by putting the other pebble of the pair on an element of the domain of the other structure;

and the *counting moves*,

1. **I** picks up k' ($k' \leq k$) pairs of x_i-pebbles and chooses a set S_I of k'-tuples in one of the two structures, say $S_I \subseteq A^{k'}$. **II** replies by choosing a set S_{II} of k'-tuples in the domain of the other structure, say $S_{II} \subseteq B^{k'}$. S_I and S_{II} must satisfy $|S_I| = |S_{II}|$;

2. **I** puts one pebble of each pair he has taken, i.e. k' pebbles, on one of the k'-tuples of S_{II}. **II** replies by putting the remaining pebbles (one in each pair, i.e. k' pebbles as well) on a k'-tuple of S_I.

I can decide at any moment in the game to trigger a unique *counting move*, hence forcing **II** to reply by a move of the same type.

The game goes as follows. Suppose the players have already taken r' classical moves ($r' \leq r - k'$ for some $k' \leq k$). Player **I** triggers a counting move by seizing k' pairs of x_i-pebbles. After this **unique** counting move, the game restarts for another $r - (r' + k')$ classical moves.

Let $\langle u, v\rangle$ be a pair of partial functions,

$$u : \{x_1, \ldots, x_\ell\} \rightarrow A; \quad v : \{x_1, \ldots, x_\ell\} \rightarrow B$$

such that the domains of u and v are equal. Let $D_u \subseteq \{x_1, \ldots, x_\ell\}$ be the domain of u. An *ℓ-configuration* on $\mathcal{A}, \mathcal{B}$ is a valid position of the game $\mathcal{C}^k_{\ell,r}$ between $\mathcal{A}$ and $\mathcal{B}$. $u(x_i) = a$ means that a pebble of the x_i-pair is on the element $a \in A$. If $x_i \notin D_u$, this means that the x_i-pebbles are not currently placed on the structures.

Let $\langle u_s, v_s\rangle$ be the configuration of the game after move s. *Player* **I** *wins the* $\mathcal{C}^k_{\ell,r}$ *game after move* s if the mapping $u_s(x_i) \mapsto v_s(x_i)$, for each $x_i \in D_u$, is not an isomorphism between the induced substructures. Player **II** wins if **I** does not, i.e. if after each of the moves $1, \ldots, r$, the latter mapping is an isomorphism. Player **II** is said to have a winning strategy for the $\mathcal{C}^k_{\ell,r}$ game if, whatever **I** plays, **II** wins, and to have a winning strategy for $\mathcal{C}^k_\ell$ if, for every r, he has a winning strategy for the $\mathcal{C}^k_{\ell,r}$ game.

The $\mathcal{C}^k_{\ell,r}$ game defined above is quite close to the game ([IL90, CFI89]) for characterizing elementary equivalence with respect to the formulas of FO+$\{\exists^{\geq m}\}_{1 \leq m < \omega}$ (i.e. FO enriched with all counting quantifiers $\exists^{\geq m} x$ meaning "there are at least m x's such that...") with ℓ variables and of quantifier-depth r. Let us denote the latter game by $\mathcal{C}^{1,*}_{\ell,r}$. All moves are counting moves of the form:

1. **I** seizes a pair of x_i-pebbles and chooses a set S_I in one of the two structures, say $S_I \subseteq A$. **II** replies by choosing a set S_{II} of elements in the opposite structure, say B. S_I and S_{II} must have the same cardinality.

2. **I** puts one of the pebbles of the x_i-pair on one element $b \in S_{II}$. **II** replies by putting the remaining pebble of the x_i-pair on one element $a \in S_I$.

The following theorem establishes the connection between $\mathcal{C}^{1,*}_{\ell,r}$ and FO with all counting quantifiers.

Theorem 3.1 [IL90, CFI89] Let $\mathcal{A}$ and $\mathcal{B}$ be two finite structures. The following two statements are equivalent:

- $\mathcal{A}$ and $\mathcal{B}$ satisfy exactly the same formulas of FO+$\{\exists^{\geq m}\}_{1\leq m<\omega}$ with ℓ variables and of quantifier-depth r.

- Player **II** has a winning strategy for the $\mathcal{C}^{1,*}_{\ell,r}$ game between $\mathcal{A}$ and $\mathcal{B}$.

It is obvious that a counting move of $\mathcal{C}^{k}_{\ell,r}$ merely generalizes the moves of $\mathcal{C}^{k,*}_{\ell,r}$. The main difference between our game and theirs is that we have unnested polyadic counters while they have nested monadic counting quantifiers.

We now are in a position to generalize the well-known result of Ehrenfeucht and Fraïssé to these new games.

Theorem 3.2 Let $\mathcal{A}$ and $\mathcal{B}$ be two τ-structures for a given signature τ. Let k and ℓ be two integers. The following statements are equivalent:

- $\mathcal{A} \equiv^k_\ell \mathcal{B}$;

- Player **II** has a winning strategy for the $\mathcal{C}^k_\ell$ game between $\mathcal{A}$ and $\mathcal{B}$.

The proof relies in part upon the following lemma, stating that the distinguishability of two structures by a formula of FO+$\mathrm{C}^k_{\mathrm{u},\ell}$ amounts to the distinguishability by a formula containing only one component of the form "there exist exactly i k'-tuples satisfying a certain property", which accounts for the presence of a single *counting move* in the games we are dealing with.

Lemma 3.3 Let $\mathcal{A}$ and $\mathcal{B}$ be two τ-structures of a given signature τ. If $\mathcal{A}$ and $\mathcal{B}$ are separated by a sentence φ of FO+$\mathrm{C}^k_{\mathrm{u}}$ with ℓ variables, then they are also separated by a sentence of FO+$\mathrm{C}^k_{\mathrm{u}}$ of smaller or equal quantifier depth, containing at most ℓ distinct variables and a unique atom of the form $count(\overline{x}, \varphi(\overline{x})) = i$.

Proof : It is a simple proof by induction on the structure of φ. Almost all cases are classical and we omit them. Let us just show how to deal with subformulas involving *Integer* terms. If $\mathcal{A}$ and $\mathcal{B}$ are separated by a formula $\exists i \psi$, then they are separated by $\psi(\underline{\mathrm{n}})$, for some $n \in \omega$. If $\mathcal{A}$ and $\mathcal{B}$ are separated by a formula $F(C_1, \ldots, C_p) = F'(C'_1, \ldots, C'_q)$, where F and F' are primitive recursive global functions of respective arity p and q, then $\mathcal{A}$ and $\mathcal{B}$ are separated by $C = \underline{\mathrm{n}}$, for some primitive recursive expression $C \in \{C_1, \ldots, C_p, C'_1, \ldots, C'_q\}$ and some $n \in \omega$. $\square$

The proof of Theorem 3.2 follows immediately from Theorem 3.1 and Lemma 3.3.

3.2 Infinitary Counting Games

Games for infinitary logic already have a rather long history. They have been widely used in the literature on finite model theory, mainly because they provide a powerful tool for investigating the expressiveness of numerous languages below infinitary logic (FO+IFP for instance). A game for every $\mathrm{L}^{\ell}_{\infty\omega} + \{\exists^{\geq m}\}_{1 \leq m < \omega}$ is introduced in [IL90]. It is straightforwardly inspired from $\mathcal{C}^{1,*}_{\ell,r}$. We follow the same pattern and define a game $\mathcal{C}^{k}_{\ell,\omega}$. The subscript ω just indicates that the number of moves is unlimited.

The game goes exactly in the same way as $\mathcal{C}^{k}_{\ell,r}$ except for the restriction to r moves: the same goes on indefinitely.

Player **I** *wins the* $\mathcal{C}^{k}_{\ell,\omega}$ *game after move s* if the mapping $u_s(x_i) \mapsto v_s(x_i)$, for each $x_i \in D_u$, is not an isomorphism between the substructures induced by the respective positions of the pebbles. Player **II** wins if **I** never does, i.e. if after each of the moves $1, 2, \ldots, s, \ldots$, the latter mapping is an isomorphism. Player **II** is said to have a winning strategy for the $\mathcal{C}^{k}_{\ell,\omega}$ game if, whatever **I** plays, **II** can maintain indefinitely the isomorphism between the induced substructures. Obviously, $\mathcal{C}^{k}_{\ell,\omega}$ aims at capturing the notion of equivalence with respect to infinitary logic with bounded-arity counters.

Definition 3.6 Let $\mathrm{L}^{\ell}_{\infty\omega}+\mathrm{C}^{k}_{\mathrm{u}}$ be the sublanguage of $\mathrm{L}^{\ell}_{\infty\omega}+\mathrm{C}$ consisting of the formulas with unnested counters of arity $\leq k$.

Definition 3.7 Let $\mathcal{A}$ and $\mathcal{B}$ be two structures over the same signature. If $\mathcal{A}$ and $\mathcal{B}$ satisfy exactly the same sentences of $\mathrm{L}^{\ell}_{\infty\omega}+\mathrm{C}^{k}_{\mathrm{u}}$, we say that they are $(k; \ell, \omega)$-equivalent, and we write $\mathcal{A} \equiv^{k}_{\ell,\omega} \mathcal{B}$.

The only difference between $\mathcal{C}^{k}_{\ell,\omega}$ and $\mathcal{C}^{k}_{\ell}$ is that, in the case of $\mathcal{C}^{k}_{\ell}$, the strategy of player **II** can *a priori* depend upon the planned number of moves of the game. However, because of the finiteness of the structures, **II** has a winning strategy for $\mathcal{C}^{k}_{\ell,\omega}$ iff he has a winning for $\mathcal{C}^{k}_{\ell,r}$, for each $r < \omega$ (similar remarks have been made in [IL90, KV92b]). The following theorem extends Theorem 3.2.

Theorem 3.4 Let $\mathcal{A}$ and $\mathcal{B}$ be two finite structures over the same signature. The following three statements are equivalent:

- $\mathcal{A} \equiv^k_{\ell,\omega} \mathcal{B}$;
- $\mathcal{A} \equiv^k_{\ell} \mathcal{B}$;
- Player **II** has a winning strategy for the $\mathcal{C}^k_{\ell,\omega}$ game on $\mathcal{A}$ and $\mathcal{B}$.

We omit the proof, which would go along the same lines as the proof of Theorem 3.2.

4 An arity-based hierarchy

We now turn to the study of the impact of the arity of the counters on the expressive power of the logics in the family $\{\mathrm{FO{+}C^k_u}\}_{k\in\omega}$, (formulas with unnested counters). We show that the latter family constitute a strict hierarchy, i.e. the expressiveness is strictly monotone with respect to the arity of the counters. Our result has a flavor quite similar to the work of Dublish and Maheshwari [DM89] who proved a strict hierarchy theorem for fixpoint logics restricted to formulas with only one occurrence of the IFP operator.

Theorem 4.1 For every $k \in \omega$, one has the following proper inclusion:

$$\mathrm{FO} + \mathrm{C^k_u}[\tau] \subset \mathrm{FO} + \mathrm{C^{k+1}_u}[\tau],$$

where the underlying signature τ is binary (it does not depend upon k).

Remark : The hierarchy of Theorem 4.1 obviously collapses when the signature is restricted to unary relations.

In order to separate $\mathrm{FO{+}C^k_u}[\tau]$ from $\mathrm{FO{+}C^{k+1}_u}[\tau]$, we adopt the following strategy: for every ℓ, we construct two structures $\mathcal{A}^k_\ell$ and $\mathcal{B}^k_\ell$ such that $\mathcal{A}^k_\ell \equiv^k_\ell \mathcal{B}^k_\ell$ but $\mathcal{A}^k_\ell \models \phi$ and $\mathcal{B}^k_\ell \models \neg\phi$, where $\phi \in \mathrm{FO{+}C^{k+1}_{u,p}}$, with p a constant *not depending* on ℓ.

As the main argument of the proof of Theorem 4.1 involves heavy combinatorial constructions, we attempt at making it clearer by splitting it into two parts: (i) in a first step, we deal with the case $k = 2$ and then (ii) we generalize the method to any k. In the case $k = 2$, we shall directly construct $\mathcal{A}^2_\ell$ and $\mathcal{B}^2_\ell$ as directed graphs, whereas, in the general case, $\mathcal{A}^k_\ell$ and $\mathcal{B}^k_\ell$ will at first appear as k-hypergraphs, before they are encoded as binary structures.

We first outline the construction in the case $k = 2$ with $\tau = \{G\}$ where G is a binary relation symbol. We show that there is a sentence ϕ in $\mathrm{FO{+}C^3_u}[\tau]$ with 3 variables such that for every ℓ there exist two directed graphs $\mathcal{A}^2_\ell$ and $\mathcal{B}^2_\ell$ such that $\mathcal{A}^2_\ell \equiv^2_\ell \mathcal{B}^2_\ell$ and $\mathcal{A}^2_\ell \models \phi$ and $\mathcal{B}^2_\ell \models \neg\phi$. The two graphs are almost identical; $\mathcal{B}^2_\ell$ is obtained from $\mathcal{A}^2_\ell$ by shifting two edges.

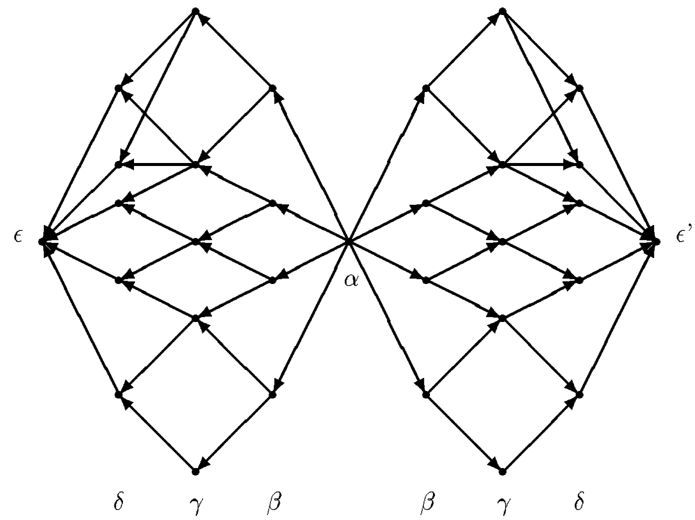


Figure 1: The butterfly

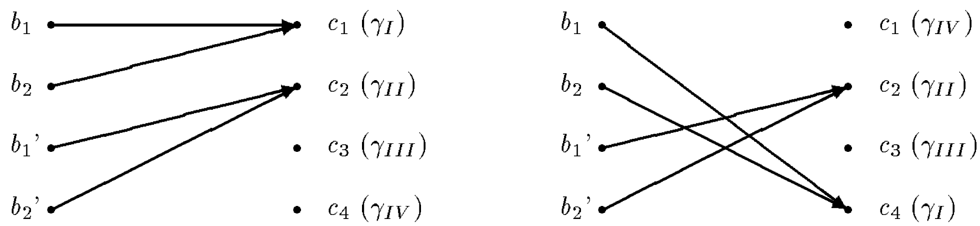
The graph $\mathcal{A}_\ell^2$ looks like a butterfly (see Figure 1), with two wings (W_1, W_2) symmetric with respect to a central vertex denoted α. Let us describe the wing W_1: we first construct a rooted dag (directed acyclic graph) of depth 4 and maximum width $\ell 2^{\ell+1}$. The root is α. We distinguish different types of vertices depending in particular from their distance from α and their degree. There are: 2^ℓ vertices of type β at distance 1 from α; $\ell 2^{\ell+1}$ vertices of type γ (divided into $\ell 2^{\ell-1}$ vertices of each of the types γ_I, γ_{II}, γ_{III}, and γ_{IV}) at distance 2 from α; 2^ℓ vertices of type δ at distance 3 from α. And finally, one vertex ϵ at distance 4 from α.

There are *directed edges* linking: α to all vertices of type β; each vertex of type β to $\ell(2^{\ell-1}-2)$ vertices of type γ (with similar number of vertices of each of the types γ_I, γ_{II}, γ_{III}, and γ_{IV}), each vertex of type γ to either $(2^{\ell-2}-2)$ or $2^{\ell-2}$ vertices of type δ; and each vertex of type δ to the vertex ϵ.

For every pair of vertices (x, y) at distance at least 2 from each other, there are many (at least $2^{\ell-2}$) (undirected) paths from x to y. The graph is *path-preserving*, that is erasing an edge $\langle a, b\rangle$ does not affect the existence of a path from x to y, for every pair $\langle x, y\rangle$ such that $x \neq a$ or $y \neq b$. This condition will be very helpful in the sequel.

One ensures that the dag contains the following pattern:

- There are four vertices b_1, b_2, b'_1 and b'_2, of type β, and four vertices c_1, c_2, c_3 and c_4, of respective type γ_I, γ_{II}, γ_{III} and γ_{IV}, such that: for every $i = 1, 2$, there is an edge from b_i to c_1, and from b'_i to c_2; there is no other edge inside the pattern. (See wing W_1 in Figure 2.) Outside the pattern, b_1, b_2, b'_1 and b'_2 have exactly the same connections,



(the wing W_1) (the wing W_2)

Figure 2: The different patterns in the two wings

and so have c_1 and c_3, and c_2 and c_4, respectively. It is the latter subgraph that will be concerned with the transformation of $\mathcal{A}_\ell^2$ into $\mathcal{B}_\ell^2$ (see two paragraphs below).

The rooted dag is now expanded by generating ℓ distinct *copies* of each vertex of type γ. This expansion is carried out as follows:

- For every $j \in \{I, II, III, IV\}$, each vertex $z \in \gamma_j$ is replaced with ℓ vertices $z_1, \ldots, z_\ell$ of the same type. There is an edge $\langle x, z_i \rangle$ (respectively $\langle z_i, y \rangle$) iff there was an edge $\langle x, z \rangle$ (respectively $\langle z, y \rangle$).

The result of the latter expansion is the wing W_1. The wing W_2 is constructed from W_1 by choosing one copy of the pattern involving $b_1, b_2, b_1', b_2', c_1, c_2, c_3, c_4$ and: (i) removing 2 edges, $\langle b_1, c_1 \rangle$ and $\langle b_2, c_1 \rangle$, of type $\langle \beta, \gamma_I \rangle$ and (ii) adding 2 edges, $\langle b_1, c_4 \rangle$ and $\langle b_2, c_4 \rangle$, of type $\langle \beta, \gamma_{II} \rangle$. The copy of c_4 affected by the transformation was of type γ_{IV} in W_1, but becomes of type γ_{II} in W_2, while the corresponding copy of c_1 moves from type γ_I to type $\gamma_I II$. (See Figure 2.)

The structure of the wings W_1 and W_2 is recalled in the following array. The attribute *Wing 1* (respectively *Wing 2*) denotes the number of nodes of the given type in W_1 (respectively W_2).

Type	in-degree	out-degree	Wing 1	Wing 2
α	0	2^ℓ	1	1
β	1	$\ell(2^{\ell-1}-2)$	2^ℓ	2^ℓ
γ_I	$2^{\ell-2}$	$2^{\ell-2}$	$\ell 2^{\ell-1}$	$\ell 2^{\ell-1}-1$
γ_{II}	$2^{\ell-2}$	$2^{\ell-2}-2$	$\ell 2^{\ell-1}$	$\ell 2^{\ell-1}+1$
γ_{III}	$2^{\ell-2}-2$	$2^{\ell-2}$	$\ell 2^{\ell-1}$	$\ell 2^{\ell-1}+1$
γ_{IV}	$2^{\ell-2}-2$	$2^{\ell-2}-2$	$\ell 2^{\ell-1}$	$\ell 2^{\ell-1}-1$
δ	$\ell(2^{\ell-1}-2)$	1	2^ℓ	2^ℓ
ϵ	2^ℓ	0	1	1

Let ρ be some renaming of the vertices of W_1, such that $\rho(\alpha) = \alpha$, $\rho(\epsilon) = \epsilon'$, and the range of ρ is disjoint from the domain of both W_1 and W_2 except for α. The two graphs $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$ are constructed as follows:

- $\mathcal{A}_\ell^2 = W_1 \cup \rho(W_1)$, and
- $\mathcal{B}_\ell^2 = W_2 \cup \rho(W_1)$.

The two graphs $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$ disagree on the following property ϕ expressible in $\mathrm{FO} + \mathrm{C}_\mathrm{u}^3$:

$$\phi \equiv count(\langle x, y, z\rangle, G(\alpha, x) \wedge G(x, y) \wedge G(y, z) \wedge G(z, \epsilon)) = count(\langle x, y, z\rangle, G(\alpha, x) \wedge G(x, y) \wedge G(y, z) \wedge G(z, \epsilon')).$$

$\mathcal{A}_\ell^2$ satisfies ϕ while $\mathcal{B}_\ell^2$ does not. Indeed, there exists an integer i such that:

$$\mathcal{A}_\ell^2 \models count(\langle x, y, z\rangle, G(\alpha, x) \wedge G(x, y) \wedge G(y, z) \wedge G(z, \epsilon)) = i,$$

$$\mathcal{B}_\ell^2 \models count(\langle x, y, z\rangle, G(\alpha, x) \wedge G(x, y) \wedge G(y, z) \wedge G(z, \epsilon)) \neq i.$$

The value of the counter in $\mathcal{A}_\ell^2$ is $i = \ell 2^{\ell+1}(2^{\ell-2} - 1)^2$, whereas in $\mathcal{B}_\ell^2$, it is $(i - 4)$.

We now prove the following lemma.

Lemma 4.2 Player **II** has a winning strategy for the game $\mathcal{C}_\ell^2$ on $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$, for every $\ell \in \omega$.

Proof : We prove that whatever the strategy of Player **I** is, Player **II** can keep the two substructures $\mathcal{A}_\ell^2/\{a_1, ..., a_\ell\}$ and $\mathcal{B}_\ell^2/\{b_1, ..., b_\ell\}$ isomorphic, where $a_1, ..., a_\ell$ ($b_1, ..., b_\ell$) are the ℓ-configurations on $\mathcal{A}_\ell^2$ ($\mathcal{B}_\ell^2$) (i.e. the positions of the pebbles). First of all, it is clear that if Player **I** does not trigger a counting move, Player **II** has a winning strategy consisting in "playing the identity". This follows from the fact that the degrees of the nodes are exponential in ℓ. We prove the result by induction on the number of classical moves preceding the counting move.

Basis : Assume that Player **I** starts the game by triggering a counting move, and selects a set of nodes (or a set of pairs of nodes) in one of the two graphs. Player **II** has to answer by a set of nodes (or a set of pairs of nodes) of the same cardinality in the other graph.

The only property of a vertex expressible in first-order logic with ℓ variables is its distance from the vertices α and ϵ. For an arbitrary pair of vertices a and b in the graphs, we can check the following properties: (i) existence of an edge from a to b (or b to a); (ii) existence of an undirected path of length n between a and b. Note that in the case there is such an undirected path, there are plenty of them. Moreover, as soon as two nodes are of types at distance[3] $n = 0$, or $n \geq 2$, then there are many paths of length at least $\mathrm{Max}\{2, n\}$. For types at distance 1, there are many undirected paths of length at least 3. It follows that the only interesting properties concern the types of the nodes and the existence of an edge between them.

The following strategy constitutes a winning strategy for Player **II**. After Player **I**'s choice of a set, S_I, of nodes (or pairs of nodes), Player **II** chooses a set S_{II} of nodes (resp.

[3] The distance between types is defined in an obvious way corresponding to their distance in the graph. For instance $d(\alpha, \alpha) = 0$ and $d(\alpha, \epsilon) = 4$.

pairs of nodes) such that S_{II} contains the same quantity of nodes (resp. pairs of nodes) of each type α, β, γ, δ, (resp. each type of pair of nodes, among $\{\alpha,\beta,\gamma,\delta\}^2$) as S_I, with the same number of edges between the nodes. The properties of the graphs ensure precisely that this can always be achieved for tuples with at most two arguments. The rest of the winning strategy goes as without counting move.

Induction : Assume that if Player **I** triggers a counting move after at most n classical moves, then Player **II** has a winning strategy. We prove that Player **II** has a winning strategy if Player **I** triggers a counting move after exactly $(n+1)$ classical moves.

If Player **I** puts a pebble on an $(n+1)^{th}$ node a on one of the two structures, such that its image on the other structure is of the same type, then Player **II** plays the identity (i.e. the vertex which has the same label in the other structure), otherwise, Player **II** plays as if Player **I** had played a free copy a' of a such that a''s image on the other structure is of the same type as a'. Since the last two nodes pebbled on the two structures before triggering a counting move have exactly the same edges to all other vertices, the winning strategy of Player **II** in the case where Player **I** triggers a counting move after at most n classical moves (induction hypothesis) still constitutes a winning strategy for Player **II** in the case where Player **I** triggers a counting move after $(n+1)$ classical moves. $\square$

We now turn to a generalization of the above method in order to construct structures separating first-order logic with $(k+1)$-ary counters, FO+$\mathrm{C}_{\mathrm{u}}^{k+1}$, from first-order logic with k-ary counters, FO+$\mathrm{C}_{\mathrm{u}}^{k}$. Instead of graphs, we first consider relations of arity k, i.e. k-hypergraphs. We then prove that the k-hypergraphs can be easily encoded in a fixed signature containing two binary relations. This seemingly awkward construction based on hypergraphs instead of (binary) graphs, gives more intuition about the undistinguishability of the two structures with counters of arity up to k.

Let $\tau = \{R\}$ where R is a k-ary relation. We show that there is a sentence ϕ in FO+$\mathrm{C}_{\mathrm{u}}^{k+1}$ with $(k+1)$ variables such that for every ℓ there exist two k-ary relations $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ such that $\mathcal{A}_\ell^k \equiv_\ell^k \mathcal{B}_\ell^k$ and $\mathcal{A}_\ell^k \models \phi$ and $\mathcal{B}_\ell^k \models \neg\phi$. The two k-ary relations are almost identical and they are in the spirit of the graphs of the case $k=2$, that is they have a butterfly shape with two symmetric wings. $\mathcal{B}_\ell^k$ is obtained from $\mathcal{A}_\ell^k$ by replacing two hyperedges.

The construction of the hypergraphs $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ is very similar to the construction of the graphs $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$. We first construct a (binary) graph, and then define the hypergraphs based on this initial binary graph.

We first construct a graph W as the union of the wings W_1 and W_2 introduced above. As a result, W contains as a subgraph the pattern involving the *eight* vertices b_i, b_i' and c_j, for $i=1,2$ and $j=1,\ldots,4$, and the *six* edges $\langle b_i, c_1\rangle$ (of type $\langle\beta,\gamma_I\rangle$) and $\langle b_i', c_2\rangle$ and $\langle b_i, c_4\rangle$ (of type $\langle\beta,\gamma_{II}\rangle$) $i=1,2$. (It is worth noticing that, in W, b_1 and b_2 have out-degree $\ell(2^{\ell-1}-2)+2$). We expand W by introducing directed paths of length k between α and the nodes of type β. To do so, we define $(k-1)$ new types of nodes $\beta_{k-1},\ldots,\beta_2,\beta_1$ (β_1 merely replaces β)[4]. There is an edge from α to all vertices of type β_{k-1}, and edges from vertices of type β_j to vertices of type β_{j-1}, for $j=2,\ldots,k-1$. The part of the graph composed of

[4] We use superscripts b^j to denote vertices of type β_j.

vertices of type β (i.e. β_1), γ, δ and ϵ remains unchanged. The characteristics of the nodes of new types are given by:

Type	in-degree	out-degree	W
β_{k-1}	1	$2^{\ell-1}$	2^ℓ
β_j	$2^{\ell-1}$	$2^{\ell-1}$	2^ℓ
β_1	$2^{\ell-1}$	$\ell(2^{\ell-1}-2)$	2^ℓ

Let E be the resulting graph. The relation R of arity k is now defined as follows:

$$\forall x_1, ..., x_k R(x_1, ..., x_k) \Leftrightarrow (E(x_1, x_2) \wedge E(x_2, x_3) \wedge ... \wedge E(x_{k-1}, x_k)).$$

The whole structures $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ also have a butterfly shape, but with wings longer than in the case of $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$.

We now construct the hypergraphs W_1^k and W_2^k from R: W_1^k is obtained by removing *two* hyperedges from R, namely:

- $\langle b^{k-1}, ..., b^2, b_1, c_4\rangle$ and $\langle b^{k-1}, ..., b^2, b_2, c_4\rangle$, of type $\langle \beta_{k-1}, ..., \beta_2, \beta_1, \gamma_{II}\rangle$,

and W_2^k is obtained by removing *two* hyperedges of R, namely:

- $\langle b^{k-1}, ..., b^2, b_1, c_1\rangle$ and $\langle b^{k-1}, ..., b^2, b_2, c_1\rangle$, of type $\langle \beta_{k-1}, ..., \beta_2, \beta_1, \gamma_I\rangle$.

The above construction of W_1^k and W_2^k ensures that every proper segment of a hyperedge is a proper segment of many other hyperedges. This is the fundamental property used in Lemma 4.3; it guarantees that the two wings have the same counting properties for counters of arity less than or equal to k.

Now, let ρ be a renaming of the constants of W_1^k, mapping α to itself and such that the range of ρ is disjoint from the domain of both W_1^k and W_2^k except for α. We finally define $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ as follows:

- $\mathcal{A}_\ell^k = W_1^k \cup \rho(W_1^k)$, and
- $\mathcal{B}_\ell^k = W_2^k \cup \rho(W_1^k)$.

Consider now the sentence ϕ in FO+$\mathrm{C}_{\mathrm{u}}^{k+1}$ defined by:

$$\begin{aligned}\phi \equiv\ & count(\langle x_1, ..., x_{k+1}\rangle, R(\alpha, x_1, ..., x_{k-1}) \wedge R(x_1, ..., x_k) \wedge R(x_2, ..., x_{k+1}) \wedge R(x_3, ..., x_{k+1}, \epsilon)) \\ & = count(\langle x_1, ..., x_{k+1}\rangle, R(\alpha, x_1, ..., x_{k-1}) \wedge R(x_1, ..., x_k) \wedge R(x_2, ..., x_{k+1}) \wedge R(x_3, ..., x_{k+1}, \epsilon')).\end{aligned}$$

It is clear that: $\mathcal{A}_\ell^k$ satisfies ϕ while $\mathcal{B}_\ell^k$ does not. Indeed, there exists an integer i such that:

$$\mathcal{A}_\ell^k \models count(\langle x_1, ..., x_{k+1}\rangle, R(\alpha, x_1, ..., x_{k-1}) \wedge R(x_1, ..., x_k) \wedge R(x_2, ..., x_{k+1}) \wedge R(x_3, ..., x_{k+1}, \epsilon)) = i,$$

$$\mathcal{B}_\ell^k \models count(\langle x_1, ..., x_{k+1}\rangle, R(\alpha, x_1, ..., x_{k-1}) \wedge R(x_1, ..., x_k) \wedge R(x_2, ..., x_{k+1}) \wedge R(x_3, ..., x_{k+1}, \epsilon)) \neq i.$$

Once again, the difference between the values on $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ of the above counter is 4.

We now prove the following lemma.

Lemma 4.3 Player **II** has a winning strategy for the game $\mathcal{C}_\ell^k$ on $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$, for every $\ell \in \omega$.

Proof : We show how the winning strategy of Player II on the graphs $\mathcal{A}_\ell^2$ and $\mathcal{B}_\ell^2$ can be extended to a winning strategy on the hypergraphs $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$. The proof goes along the same lines as the proof of Lemma 4.2. It is clear that if Player **I** doesn't trigger a counting move, the strategy, consisting for Player **II**, to play the identity, is a winning strategy for the game $\mathcal{C}_\ell^k$.

We establish the proof by induction on the number of classical moves preceding the counting move as before. It is easy to see that the induction is done exactly as for Lemma 4.2. We only have to verify the basis of the induction. This is done also in a very similar way. For a k-tuple, the only first-order expressible properties concern the types of the nodes and the existence of an edge between them. This follows from the construction of $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$. The proof then carries over as in Lemma 4.2. $\square$

Lemma 4.3 can be easily generalized to a fixed signature containing two binary relations.

Lemma 4.4 The structures $\mathcal{A}_\ell^k$ and $\mathcal{B}_\ell^k$ over a signature containing a relation of arity k can be encoded in $\mathcal{E}_\ell^2$ and $\mathcal{D}_\ell^2$ over a signature of arity 2 only, and such that Player **II** has a winning strategy for the game $\mathcal{C}_\ell^k$ on $\mathcal{E}_\ell^2$ and $\mathcal{D}_\ell^2$, for every $\ell \in \omega$.

Proof : Let $\tau = \{S, T\}$ be a binary signature. The encoding is based on the definition of the hypergraph starting from the graph W. We assume that the first relation S contains the graph $W \cup \rho(W)$. $\mathcal{A}_\ell^k$ is obtained from $W \cup \rho(W)$ by removing four hyperedges (two in each wing), namely:

- $(b^{k-1}, ..., b^2, b_1, c_4)$ and $(b^{k-1}, ..., b^2, b_2, c_4)$ in W, and
- $(\rho(b^{k-1}), ..., \rho(b^2), \rho(b_1), \rho(c_4))$ and $(\rho(b^{k-1}), ..., \rho(b^2), \rho(b_2), \rho(c_4))$ in $\rho(W)$.

$\mathcal{B}_\ell^k$ is obtained from $W \cup \rho(W)$ by removing four hyperedges (two in each wing), namely:

- $(b^{k-1}, ..., b^2, b_1, c_1)$ and $(b^{k-1}, ..., b^2, b_2, c_1)$ in W, and
- $(\rho(b^{k-1}), ..., \rho(b^2), \rho(b_1), \rho(c_4))$ and $(\rho(b^{k-1}), ..., \rho(b^2), \rho(b_2), \rho(c_4))$ in $\rho(W)$.

The second binary relation, T, suffices to store the four hyperedges removed to define each graph. Let $\mathcal{E}_\ell^2$ (respectively $\mathcal{D}_\ell^2$) be the τ-structure defined with S and T as above. It is easy to see that the structure $\mathcal{E}_\ell^2$ (respectively $\mathcal{D}_\ell^2$) is first-order reducible (and reciprocally) to $\mathcal{A}_\ell^k$ (resp. $\mathcal{B}_\ell^k$). Therefore, they satisfy both the same sentences of $\mathrm{FO} + \mathrm{C}_\mathrm{u}^k[\tau]$. $\square$

Theorem 4.1 follows easily from lemmas 4.2, 4.3 and 4.4. It admits the following generalization.

Corollary 4.5 There is a binary signature τ such that for every $k \in \omega$,

- $\mathrm{FO} + \mathrm{IFP} + \mathrm{C}^k_{\mathrm{u}}[\tau] \subset \mathrm{FO} + \mathrm{IFP} + \mathrm{C}^{k+1}_{\mathrm{u}}[\tau]$.
- $\mathrm{L}^{\omega}_{\infty\omega}{+}\mathrm{C}^k_{\mathrm{u}}[\tau] \subset \mathrm{L}^{\omega}_{\infty\omega}{+}\mathrm{C}^{k+1}_{\mathrm{u}}[\tau]$.

Proof : The proof is made on the same structures as in the proof of Theorem 4.1. It is easy to see that the winning strategy of Player II can be maintained in the same manner for an unbounded number of moves. □

5 Asymptotic probabilities

We study properties of the counting languages under stronger restrictions than in the previous section. The restrictions now are: (i) there is no free occurrence of a variable within the scope of a counting term, and (ii) no integer variables are allowed in the intermediate relations[5]. We see that under these restrictions, the counting extensions of FO, FO+IFP and $\mathrm{L}^{\omega}_{\infty\omega}$, respectively denoted by FO+$\mathrm{C}^*_=$, FO+IFP+$\mathrm{C}^*_=$ and $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C}^*_=$, all admit a 0/1 law.

We then introduce several extensions of FO+$\mathrm{C}^*_=$, respectively called FO+$\mathrm{C}^*_\leq$, FO+C^*_+ and FO+$\mathrm{C}^*_\times$, with restricted arithmetics on the side. While in FO+$\mathrm{C}^*_=$ one can only test the equality of the cardinalities of definable relations, FO+$\mathrm{C}^*_\leq$ extends FO+$\mathrm{C}^*_=$ by allowing the counting expressions to be compared modulo the usual order on the integers, and FO+C^*_+ and FO+$\mathrm{C}^*_\times$ respectively extend FO+$\mathrm{C}^*_\leq$ with the addition and both the addition and the multiplication .

We present a series of results and conjectures on the asymptotic probabilistic behavior of those languages. As a by-product of our results, we obtain a strict hierarchy of counting languages with respect to the expressive power. To our knowledge, the only results on the asymptotic probabilities of extensions of first-order logic with counters can be found in [Kny90, GT92, FGT93]. In [Kny90], 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: otherwise, 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 of the signature, would be almost surely true or almost surely false, which is trivially false. Unfortunately, because of the restriction on r, Knyazev's result and its proof technique (by induction on the structural complexity of the sentences) are of no use to us.

5.1 0/1 laws

Let $\langle R_1, \ldots, R_k\rangle$ be a purely relational signature and $\tau = \langle r_1, \ldots, r_k\rangle$ be its *similarity type*, i.e. the arity of R_i is r_i. $\mathrm{Str}_{<\omega}[\tau]$ and $\mathrm{Str}_n[\tau]$ respectively denote the class of finite τ-structures and the class of τ-structures with domain $\mathbf{n} = \{0, \ldots, n-1\}$. If $\mathcal{L}[\tau]$ is a logic

[5] This restriction is important in the case of the fixpoint.

for $\mathrm{Str}_{<\omega}[\tau]$ and φ is a sentence of $\mathcal{L}[\tau]$, $\mu_n(\varphi)$ denotes the proportion of structures of $\mathrm{Str}_n[\tau]$ which satisfy φ:

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

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 *(labeled) 0/1 law*.

First-order logic without constant or function symbols (FO) was the first logic to be proved to enjoy a 0/1 law [GKLT69, Fag76]. Moreover, Fagin considered the (infinite) set of all *extension axioms*, which constitute an ω-categorical and complete theory Θ, whose unique countable model, the *random countable structure*, denoted by Ω, satisfies: a sentence $\varphi \in FO$ has asymptotic probability 1 iff it is true in Ω (iff it is a theorem of Θ). This property, called the *Transfer Property* for FO, was later proved to carry over to other logics.

We next define formally, the languages FO+$\mathrm{C}^*_=$, FO+IFP+$\mathrm{C}^*_=$, and $\mathrm{L}^\omega_{\infty\omega}$+$\mathrm{C}^*_=$.

Definition 5.1 The *counting terms* of FO+$\mathrm{C}^*_=$, are all of the form $count(\overline{x}, \varphi(\overline{x}))$, where $\varphi(\overline{x})$ is a formula of FO whose only free variables are $\overline{x}$. The *counting expressions* of FO+$\mathrm{C}^*_=$ consist of the counting terms and of the variables and constants of *Integer* sort.

Definition 5.2 The *atomic formulas* of FO+$\mathrm{C}^*_=$ are either formulas of FO or of the form $C_1 = C_2$ where C_1 and C_2 are counting expressions. The *formulas* of FO+$\mathrm{C}^*_=$ are built from the atomic formulas by means of the usual first-order constructs.

Counting expressions and formulas of FO+IFP+$\mathrm{C}^*_=$ and $\mathrm{L}^\omega_{\infty\omega}$+$\mathrm{C}^*_=$ are defined similarly with FO+IFP (resp. $\mathrm{L}^\omega_{\infty\omega}$) in place of FO, and the assumption that the relations are of sort domain. Therefore, there are no integers in the recursively defined relations of FO+IFP.

We first show that FO+$\mathrm{C}^*_=$ has a 0/1 law. In sharp contrast with other 0/1 laws, our proof will not establish the Transfer Property for FO+$\mathrm{C}^*_=$. Indeed, for immediate reasons, there is no such result for sentences of FO+$\mathrm{C}^*_=$. Moreover, Theorem 5.2 is not a consequence of the 0/1 law for Knyazev's language. Yet, a 0/1 law for sentences of the form $count(x, \varphi(x)) = count(x, \psi(x))$ would derive from a 0/1 law for all expressions of the form: "there is *exactly* a fraction r of the elements of the domain satisfying $\varphi(x)$", where r is *any* rational such that $0 \leq r \leq 1$.

We first show how to reduce our problem to the study of the asymptotic behavior of a term A_n expressing a probability in a simplified world. This reduction relies on fundamental properties of the equivalence classes of k-tuples of elements of the countable random structure (two tuples are equivalent if there is an automorphism mapping one to the other). Let τ be a fixed relational signature and $\varphi(\overline{x})$ and $\psi(\overline{x})$ be two formulas (with $\overline{x}$ as their k-tuple of free variables) built over τ. Each equivalence class of k-tuples over the domain of the random countable structure over τ is determined by a *complete open description* with k variables. Each FO formula with k free variables corresponds almost surely to a disjoint union of such equivalence classes. So asking whether, asymptotically, $count(\overline{x}, \varphi(\overline{x})) = count(\overline{x}, \psi(\overline{x}))$, amounts to comparing the respective cardinalities of two different unions of equivalence classes. The next fundamental proposition follows from results in [Fag76, Gra83]. In particular, it states that FO-formulas enjoy quantifier-elimination in the random countable structure.

Proposition 5.1 For every $\ell \geq 1$, the number of equivalence classes of ℓ-tuples induced by the isomorphisms of $\Omega[\tau]$ is finite and depends only upon ℓ and the similarity type τ (this number is denoted by $N_\ell(\tau)$). Furthermore,

- each equivalence class C is characterized on $\Omega[\tau]$ by an FO$[\tau]$ formula without quantifier (a complete open description), $\Phi_C(x_1, \dots, x_\ell)$, i.e. for each ℓ-tuple $\langle a_1, \dots, a_\ell \rangle$ of integers,
$$\langle a_1, \dots, a_\ell \rangle \in C \text{ iff } \Omega[\tau] \models \Phi_C(a_1, \dots, a_\ell);$$
- each equivalence class C is *equiprobable*, i.e. for a random ℓ-tuple $\overline{a}$,
$$\mathrm{Prob}(\overline{a} \in C) = \frac{1}{N_\ell(\tau)};$$
- 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 $C_1, \dots, C_p$ such that:
$$\lim_{n \to \infty} \forall \overline{x} \, (\varphi(\overline{x}) \Leftrightarrow \bigvee_{1 \leq i \leq p} \Phi_{C_i}(\overline{x})) = 1.$$

We can now state the main result which shows that every sentence of FO+$\mathrm{C}^*_=$ is almost surely true or almost surely false.

Theorem 5.2 [FGT93] FO+$\mathrm{C}^*_=$ has a 0/1 law.

Proof : It is clear that if the asymptotic probabilities of properties φ and ψ are 0 or 1, so are the asymptotic probabilities of properties expressed by $\neg\varphi$, $\varphi \wedge \psi$, and $\varphi \vee \psi$ (i.e. almost sure truth or falsity is preserved by boolean constructs). We can therefore restrict ourselves to considering sentences of the form (i) $count(\overline{x}, \varphi(\overline{x})) = count(\overline{y}, \psi(\overline{y}))$, or (ii) $count(\overline{x}, \varphi(\overline{x})) = i$, where i is an integer. The asymptotic probability of sentences of the second form is obviously 0. Moreover, without loss of generality, we may further restrict to atomic sentences of the form: $count(\overline{x}, \varphi(\overline{x})) = count(\overline{x}, \psi(\overline{x}))$, with counters of the same arity. 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 sentence:

$$count(x_1, \dots, x_r, \varphi(x_1, \dots, x_r)) = count(y_1, \dots, y_s, \psi(y_1, \dots, y_s))$$

is equivalent to:

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

Now, Proposition 5.1 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

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

is identical to the asymptotic behavior of

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

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+$C^*_=$ has a labeled 0/1 law. It is shown in [FGT93], using classical methods from complex analysis, involving the Laplace method for computing integrals and the saddle-point method, that, as $n \to \infty$, $A_n \to 0$ or 1 at exponential speed, except when $\ell_1 = \ell_2$, in which case the speed of convergence is $O\left(\frac{1}{\sqrt{n}}\right)$. □

The 0/1 law of FO+$C^*_=$ has the immediate consequence that *Even*, the query giving the parity of the cardinality, is not definable in the language. It shows the importance of 0/1 laws in this context since this result cannot be proved using the $\mathcal{C}^k_\ell$ games defined in the previous section. Indeed, two structures with different cardinalities can be distinguished by a sentence of quantifier depth 1. For a fixed signature, the decision problem for the probabilities of FO+$C^*_=$ is PSPACE complete. This follows from Grandjean's result [Gra83].

Moreover, it has been shown in [FGT93], that the assumption of Definition 5.1 that counting expressions have no free variable is necessary to get a 0/1 law. Indeed, there is a sentence ϕ in FO+C whose probability $\mu_n(\phi)$ does not have a limit.

Theorem 5.2, admits the following generalization:

Theorem 5.3 The language $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C}^{*}_{=}$ admits a 0/1 law.

Proof : We prove that for each k, $\mathrm{L}^{k}_{\infty\omega}$+$\mathrm{C}^{*}_{=}$ admits a 0/1 law. It has been shown in [KV92b], that each $\mathrm{L}^{k}_{\infty\omega}[\tau]$ 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:

$$\Theta_k \models \forall \overline{x}\, (\varphi(\overline{x}) \Leftrightarrow \bigvee_{1\leq i\leq p} \Phi_{C_i}(\overline{x})),$$

where Θ_k is the finite set of extension axioms with at most k variables over signature τ. Θ_k has a finite model. Let $\models_{\mathrm{f}}$ denote validity in finite structures. It follows that:

$$\Theta_k \models_{\mathrm{f}} count(\overline{x}, \varphi(\overline{x})) = count(\overline{x}, \bigvee_{1\leq i\leq p} \Phi_{C_i}(\overline{x})).$$

Therefore atomic sentences of the form $count(\overline{x}, \varphi(\overline{x})) = t$, where t is some *Integer* term, and $\varphi(\overline{x})$ is a formula in $\mathrm{L}^{k}_{\infty\omega}[\tau]$, have asymptotic probability 0 or 1.

Let ψ be any sentence in $\mathrm{L}^{k}_{\infty\omega}$+$\mathrm{C}^{*}_{=}$. Then,

$$\Theta_k \models_{\mathrm{f}} \psi \Leftrightarrow \psi',$$

where ψ' is obtained from ψ by replacing each expression $count(\overline{x}, \varphi(\overline{x}))$ by the corresponding $count(\overline{x}, \bigvee_{1\leq i\leq p} \Phi_{C_i}(\overline{x}))$ such that:

$$\Theta_k \models_{\mathrm{f}} \forall \overline{x}\, (\varphi(\overline{x}) \Leftrightarrow \bigvee_{1\leq i\leq p} \Phi_{C_i}(\overline{x})).$$

ψ' is an infinitary sentence containing a finite number of distinct atomic sentences of the form:

$$count(\overline{x}, \varphi(\overline{x})) = t.$$

Therefore, since these atomic sentences have asymptotic probability 0 or 1, ψ' has asymptotic probability 0 or 1, and so does ψ. $\square$

It follows from the previous theorem that FO+IFP+$\mathrm{C}^{*}_{=}$ admits a 0/1 law. Note that the computational complexity of the decision problem for the value of the asymptotic probability of a sentence in FO+IFP+$\mathrm{C}^{*}_{=}$ is the same as for FO+IFP, that is EXPTIME complete [BGK85, KV87]. On the other hand, $\mathrm{L}^{\omega}_{\infty\omega}$+$\mathrm{C}^{*}_{=}$ does not admit an effective syntax, and the complexity of the decision problem is meaningless [KV92b].

5.2 The power of arithmetic constructs

In this section, we consider extensions of the previous languages with limited arithmetic constructs. We first define formally these extensions.

Definition 5.3 The *counting terms* of FO+C$^*_<$, FO+C$^*_+$ and FO+C$^*_\times$ are the counting terms of FO+C$^*_=$. The *counting expressions* of FO+C$^*_<$ are the counting expressions of FO+C$^*_=$. The set of *counting expressions* of FO+C$^*_+$ is the closure of the set of counting expressions of FO+C$^*_=$ under $(C_1, C_2) \mapsto C_1 + C_2$, and the set of *counting expressions* of FO+C$^*_\times$ is the closure of the set of counting expressions of FO+C$^*_+$ under $(C_1, C_2) \mapsto C_1 \times C_2$, and $(C_1, C_2) \mapsto C_1 + C_2$.

Definition 5.4 The *atomic formulas* of FO+C$^*_<$, FO+C$^*_+$ and FO+C$^*_\times$ are either atomic formulas of FO, or of the form $C_1 = C_2$, or $C_1 \leq C_2$ where C_1 and C_2 are counting expressions of respectively FO+C$^*_<$, FO+C$^*_+$ or FO+C$^*_\times$. The *formulas* are built from the atomic formulas by means of the usual first-order constructs.

We first consider the language FO+C$^*_<$ which extends FO+C$^*_=$ by allowing counting expressions to be compared modulo the usual order on the integers. It is clear that FO+C$^*_<$ does not enjoy a 0/1 law. If R is a monadic relation, the following sentence has probability $\frac{1}{2}$:

$$count(x, R(x)) \leq count(x, \neg R(x)).$$

Let us first consider the asymptotic probabilities of sentences under strong syntactic restrictions. We consider the restrictions of FO+C$^*_<$, FO+C$^*_+$ and FO+C$^*_\times$, to (i) universal relational signatures, i.e. containing a unique relation symbol R, and (ii) monadic counters only. We denote the corresponding sublanguages FO+C$^{*1}_<[R]$, FO+C$^{*1}_+[R]$ and FO+C$^{*1}_\times[R]$.

Theorem 5.4 The asymptotic probability of any sentence of FO+C$^{*1}_<[R]$ always exists and is equal to 0, $\frac{1}{2}$ or 1.

Proof : We first prove the result for atomic sentences. It is sufficient to consider atomic sentences of the form:

$$\Theta : count(x, \varphi(x)) \leq count(x, \psi(x)).$$

We prove that the asymptotic probability of Θ equals 0, $\frac{1}{2}$ or 1. There are exactly two equivalence classes of elements of ω induced by the automorphisms of $\Omega(R)$, namely $\{x \mid R(x, \ldots, x)\}$ and $\{x \mid \neg R(x, \ldots, x)\}$. The interpretations of $\varphi(x)$ and $\psi(x)$ on $\Omega(R)$ are either the empty set or unions of the equivalence classes determined by $R(x, \ldots, x)$ and $\neg R(x, \ldots, x)$. Thus one of the following sentence $\forall x(\varphi(x) \Leftrightarrow \top)$, $\forall x(\varphi(x) \Leftrightarrow \bot)$, $\forall x(\varphi(x) \Leftrightarrow R(x, \ldots, x))$ and $\forall x(\varphi(x) \Leftrightarrow \neg R(x, \ldots, x))$ is almost surely true (and the others are almost surely false), and the same holds for $\psi(x)$.

Let $\Phi_\theta \sim count(x, R(x, \ldots, x))\, \theta\, count(x, \neg R(x, \ldots, x))$. It is easy to verify that the asymptotic probability of Θ is 0 or 1 unless $\Omega(R) \models \forall x(\varphi(x) \Leftrightarrow R(x, \ldots, x))$ and $\Omega(R) \models \forall x(\psi(x) \Leftrightarrow \neg R(x, \ldots, x))$ (or reciprocally). The asymptotic behavior of Θ is then identical to the asymptotic behavior of $\Phi_\leq$. Let $\mu_n(\Phi)$ denote the probability that Φ be true on a random structure of domain $\mathbf{n} = \{0, \ldots, n-1\}$. A given structure $\mathcal{B}$ on $\mathbf{n}$ satisfies $\Phi_\leq$ iff the complement structure $\overline{\mathcal{B}}$ does not ($\overline{x} \in \mathcal{B}$ iff $\overline{x} \notin \overline{\mathcal{B}}$), unless both $\mathcal{B}$ and $\overline{\mathcal{B}}$ satisfy $\Phi_=$. It follows from Theorem 5.2, that this last case has asymptotic probability 0. It can also be verified easily as follows. For $n = 2k+1$, $\mu_n(\Phi_=)$ trivially equals 0. For $n = 2k$, $\mu_n(\Phi_=)$ is given by a straightforward enumeration:

$$\frac{1}{2^{n^v}} \binom{n}{n/2} 2^{n^v - n} = \frac{1}{2^n} \binom{n}{n/2},$$

where v is the arity of R. Stirling's formula immediately gives:

$$\mu_n(\Phi_=) = O(1/\sqrt{n}).$$

Therefore, the asymptotic probability of $\mu_n(\Phi_\leq)$ is $\frac{1}{2}$, and the speed of convergence is $O(1/\sqrt{n})$. In all other cases, the asymptotic probability of atomic sentences is either 0 or 1.

The proof generalizes easily to general sentences. Indeed, since the signature contains only one relation symbol, sentences of asymptotic probability $\frac{1}{2}$ are almost surely equivalent to one of the two sentences (i) $count(x, R(\overline{x}))\, \theta\, count(x, \neg R(\overline{x}))$ or (ii) $count(x, \neg R(\overline{x}))\, \theta\, count(x, R(\overline{x}))$. Boolean combinations of these two sentences have asymptotic probabilities 0, $\frac{1}{2}$, or 1. The set of asymptotic probabilities $\{0, \frac{1}{2}, 1\}$ is therefore preserved under first-order constructions. □

Again, for a fixed signature, computing the asymptotic probabilities of FO+$C^{*1}_{\leq}[R]$ is PSPACE complete and *Even* is not definable since it has no asymptotic probability.

A more general result was proved in [FGT93]. It was shown that the asymptotic probability of *atomic sentences* in FO+$C^*_{\leq}$ (without restrictions on the signature or the arity of the counters) is among 0, $\frac{1}{2}$ or 1. The technique is similar to the one used in Theorem 5.2, but it cannot be generalized to all sentences in an easy way. The same reformulation as for Theorem 5.2 can be used for giving our problem a combinatorial form. We just have to adapt it to our new purpose: what we now are interested in is the probability that the first ℓ_1 urns contain at least as many balls as the next ℓ_2 ones, which is given by the following expression:

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

The speed of convergence is $O\left(\frac{1}{\sqrt{n}}\right)$. The decision problem is again PSPACE complete.

Consider now the language FO+C^*_+ which extends FO+$C^*_{\leq}$ with addition. The convergence of the probabilities is no longer ensured in FO+C^*_+. There are indeed sentences which do not have an asymptotic probability. For example, the parity of the domain, whose probability switches from 0 to 1 is definable as follows.

$$\text{Even} \equiv \exists i\; count(x, x = x) = i + i.$$

We can prove the following restricted result.

Proposition 5.5 The asymptotic probability of every sentence φ in FO+$C_{+}^{*1}[R]$ *without Integer variable* exists and is equal to 0, $\frac{1}{2}$ or 1.

Proof : Consider a sentence of FO+$C_{+}^{*1}[R]$ with no integer variable. In this fragment, one can merely express boolean combinations of equations or inequations of the form:

$$\sum_{\varphi \in \Phi} count(x, \varphi(x))\, \theta \sum_{\psi \in \Psi} count(x, \psi(x)),$$

where θ is either $=$ or $\leq$, and Φ and Ψ are finite sets of formulas in FO$[R]$ with a unique free variable. As already mentioned, the $\varphi(x)$'s and the $\psi(x)$'s are asymptotically equivalent to $x = x$, $x \neq x$, $R(x, \ldots, x)$, or $\neg R(x, \ldots, x)$. So every atomic formula is asymptotically equivalent to a formula of the form:

$$\Theta \sim count(x, R(x, ..., x))\, \theta\; r\; count(x, x = x),$$

where r is a rational between 0 and 1. If θ is $=$, the asymptotic probability is 0. Assume indeed that $r = \frac{p}{q}$. For $n = qk + s$, where $0 < s < q$ and $k \in N$, $\mu_n(\Phi_=)$ uniformly equals 0. For $n = qk$, $\mu_n(\Phi_=)$ is given by a straightforward enumeration:

$$\frac{1}{2^{n^v}} \binom{n}{\frac{pn}{q}} 2^{n^v - n} = \frac{1}{2^n} \binom{n}{\frac{pn}{q}},$$

where v is the arity of R. Stirling's formula immediately gives:

$$\mu_n(\Phi_=) = O(1/\sqrt{n}).$$

If θ is $\leq$ and $r \neq 1/2$ then it follows from Knyazev's result that the asymptotic probability is 0 or 1. Now, if $r = 1/2$, it follows from Theorem 5.4, that the asymptotic probability is 1/2. □

If we turn to FO+$C_{\times}^{*}$, the expressive power still increases (otherwise, $\times$ would be definable from $+$ in first-order logic). It is indeed possible, as soon as the multiplication is available, to express that the cardinality of a relation is a perfect square or a prime number. As for their asymptotic behavior, it is clear that, in FO+$C_{\times}^{*}$, we lose any reasonable form of regularity.

If we consider the restrictions of FO+IFP+C and $L_{\infty\omega}^{\omega}$+C defined as above for FO+C, the results presented in this section carry over if FO+C is replaced by FO+IFP+C or $L_{\infty\omega}^{\omega}$+C. In fact, the asymptotic behavior is preserved whenever one adds a construct enabling to define new relations over the *Domain* sort, such that FO with this construct enjoys the transfer property.

The results are of interest by themselves. Depending on the arithmetic allowed on the counters, the asymptotic behavior of the sentences varies dramatically. The results can be compared with Lynch's theorems [Lyn80] on the asymptotic probabilities of sentences on the class of structures with a modular successor and a modular addition. In particular, FO sentences on the class of structures with a modular successor relation have a 0/1 law. FO sentences on the class of structures with a successor relation have a limit law. FO sentences on the class of structures with a successor relation and a modular addition have a periodic law. In contrast with ours, Lynch's results are obtained by developing specific game-theoretic techniques.

6 Conclusion

We focused in this work on extensions of first-order, inductive and infinitary logic in presence of counting mechanisms. We proved various results on the expressive power of rather restricted counting primitives in these languages. The proofs requires two kinds of tools: extensions of Ehrenfeucht-Fraïssé games to counters, and combinatorial techniques linked with classical analytical methods. There are important problems that we have been unable to solve yet, and seem to be of some difficulty. We present and discuss some conjectures and open problems below.

We proved that the languages $\mathrm{FO} + \mathrm{C}_u^k$ define a strict hierarchy. It is very likely that, without global functions, the expressive power of FO with counters strictly increases with the arity of the counters, even in the case of nested counters. The reason for that is that it is impossible to sum up integers in the absence of recursion mechanisms. Indeed, summing up a set of integers is a global phenomenon, whereas FO can only express local properties [Gai81].

We suggest that techniques similar to the ones used for establishing the 0/1 law for $\mathrm{FO+C}^*_=$ yield a "convergence law" for $\mathrm{FO+C}^*_\leq$, although the new term to be studied does not lend itself easily to a transformation into polynomials of Le Gendre. We conjecture that:

- $\mathrm{FO+C}^*_\leq$ has a limit law, and the limits are rational numbers.

Trivially, for a given signature τ, if there is a limit law, the set of possible asymptotic probabilities of sentences of $\mathrm{FO+C}^*_\leq[\tau]$ remains finite. The proof of the finiteness of the set of possible asymptotic probabilities is straightforward. Indeed, let l be the number of equivalence classes of k-tuples of elements of the domain of the random countable structure over a given signature. That number being fixed, we get a finite number of limits, given by the pairs of integers (l_1, l_2) satisfying $l_1 + l_2 \leq l$.

Although we did not come up with a definite proof for $\mathrm{FO+C}^*_+$, we most naturally think that it enjoys a "periodic law". We conjecture that:

- FO+C^*_+ has a periodic law, and the limits are rational numbers.

Or, in other words, for every formula φ in FO+C^*_+, there exists an integer a such that for each b, $b < a$, $lim_{n\to\infty}\mu_{b+an}(\varphi)$ exists.

Acknowledgements : We are greatly indebted to Guy Fayolle for the collaboration on the proof of Theorem 5.3. We also thank Jianwen Su for comments on an earlier draft.

References

[AV91] S. Abiteboul and V. Vianu. Generic computation and its complexity. In *Proc. ACM Symp. on Theory of Computing*, New Orleans, May 1991.

[Bar77] J. Barwise. On Moschovakis' closure ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.

[BF85] J. Barwise and S. Feferman. *Model Theoretic Logics.* Springer-Verlag, 1985.

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

[CFI89] J. Cai, M. Furer, and N. Immerman. An optimal lower bound on the number of variables for graph identification. In *Proc IEEE Foundations of Computer Science*, pages 612–617, 1989.

[DH94] A. Dawar and L. Hella. The expressive power of finitely many generalized quantifiers. In *Proc. IEEE Symp. of Logic in Computer Science*, pages 20–29, 1994.

[DLW94] A. Dawar, S. Lindell, and S. Weinstein. Infinitary logic and inductive definability over finite structures. *Information and Computation*, 1994.

[DM89] P. Dublish and S. Maheshwari. Expressibility of bounded-arity fixed-point query hierarchies. In *Proc. 8th ACM Symp. on Principles of Database Systems*, pages 324–335, Philadelphia, 1989.

[Ehr61] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fund. Math*, 49, 1961.

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

[Fag93] R. Fagin. Finite model theory - a personal perspective. *Theoretical Computer Science*, 116:3–31, 1993.

[FGT93] G. Fayolle, S. Grumbach, and C. Tollu. Asymptotic probabilities of languages with generalized quantifiers. In *Proc. IEEE Symp. of Logic in Computer Science*, pages 199–207, Montreal, June 1993.

[Fra54] R. Fraïssé. Sur les classifications des systèmes de relations. *Publ. Sci. Univ Alger*, I:1, 1954.

[Gai81] H. Gaifman. On local and non local properties. In J. Stern, editor, *Proc. Herbrand Symposium Logic Colloquium*, pages 105–135. North Holland, 1981.

[GKLT69] Y. Glebskii, D. Kogan, M. Liogon'kii, and V. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetica*, 1969.

[GO93] E. Grädel and M. Otto. Inductive definability with counting on finite structures. In *Proc. of Computer Science Logic 92*, pages 231–247. LNCS, n702, 1993.

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

[GS85] Y. Gurevich and S. Shelah. Fixed-point extensions of first order logic. In *Proc IEEE Foundations of Computer Science*, 1985.

[GT92] S. Grumbach and C. Tollu. Query languages with counters. In J. Biskup and R. Hull, editors, *Proc. Int. Conf. on Database Theory*, pages 124–139, Berlin, 1992. Springer Verlag. LNCS 646.

[Gur83] Y. Gurevich. Algebras of feasible functions. In *Proc IEEE Foundations of Computer Science*, pages 210–214, 1983.

[Gur88] Y. Gurevich. *Current Trends in Theoretical Computer Science, E. Borger Ed.*, chapter Logic and the Challenge of Computer Science, pages 1–57. Computer Science Press, 1988.

[Har65] H. Hartig. über einen quantifikator mit zwei wirkungsbereichen. In L. Kalmar, editor, *Colloquium on Foundations of Mathematics, Mathematical Machines and their Applications*, pages 31–36, Budapest, 1965.

[Hel92] L. Hella. Logical hierarchies in PTIME. In *Proc. 7th Symp. of Logic in Computer Science*, 1992.

[IL90] N. Immerman and E. Lander. *Complexity Theory Retrospective*, chapter Describing Graphs : A First-Order Approach to Graph Canonization, pages 59–81. Springer Verlag, Ed. A. Selman, 1990.

[Imm86] N. Immerman. Relational queries computable in polynomial time. *Inf. and Control*, 68:86–104, 1986.

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

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

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

[KV90b] P. Kolaitis and M.Y. Vardi. On the expressive power of datalog: Tools and a case study. In *Proc. 9th ACM Symp. on Principles of Database Systems*, pages 61–71, 1990.

[KV92a] P. Kolaitis and M.Y. Vardi. Fixpoint logic vs. infinitary logic in finite model theory. In *Proc. 7th Symp. of Logic in Computer Science*, pages 46–57, 1992.

[KV92b] P. Kolaitis and M.Y. Vardi. Infinitary logic and 0-1 laws. *Information and Computation*, 98(2):258–294, 1992.

[KV92c] P. Kolaitis and J.A. Väänänen. Generalized quantifiers and pebble games on finite structures. In *Proc. 7th Symp. of Logic in Computer Science*, 1992. to appear in the Annals of Pure and Applied Logic.

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

[Lyn80] J. Lynch. Almost sure theories. *Annals of Mathematical Logic*, 18:91–135, 1980.

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

[Ott92] M. Otto. The expressive power of fixed-point logic with counting. In *manuscript*, 1992.

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

[Var82] M. Vardi. The complexity of relational query languages. In *Proc. 14th ACM Symp. on Theory of Computing*, pages 137–146, 1982.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399