Asymptotic Conditional Probabilities for First-Order Logic

Adam J. Grove
Stanford University
Stanford, CA 94305
email: grove@cs.stanford.edu

Joseph Y. Halpern
IBM Almaden Research Center
San Jose, CA 95120
email: halpern@almaden.ibm.com

Daphne Koller
Stanford University
Stanford, CA 94305
email: daphne@cs.stanford.edu

Abstract

Motivated by problems that arise in computing *degrees of belief*, we consider the problem of computing asymptotic conditional probabilities for first-order formulas. That is, given first-order formulas φ and θ, we consider the number of structures with domain $\{1, \ldots, N\}$ that satisfy θ, and compute the fraction of them in which φ is true. We then consider what happens to this probability as N gets large. This is closely connected to the work on 0-1 laws that considers the limiting probability of first-order formulas, except that now we are considering asymptotic *conditional* probabilities. Although work has been done on special cases of asymptotic conditional probabilities, no general theory has been developed. This is probably due in part to the fact that it has been known that, if there is a binary predicate symbol in the vocabulary, asymptotic conditional probabilities do not always exist. We show that in this general case, almost all the questions one might want to ask (such as deciding whether the asymptotic probability exists) are highly undecidable. On the other hand, we show that the situation with unary predicates only is much better. If the vocabulary consists only of unary predicate and constant symbols, it is decidable whether the limit exists, and if it does, there is an effective algorithm for computing it. The complexity depends on two parameters: whether there is a fixed finite vocabulary or an infinite one, and whether there is a bound on the depth of quantifier nesting.

Research sponsored in part by the Air Force Office of Scientific Research (AFSC), under Contract F49620-91-C-0080. The United States Government is authorized to reproduce and distribute reprints for governmental purposes. The first author was also supported by an IBM Graduate Fellowship.

1 Introduction

Consider an agent (or expert system) whose data consists of some facts θ, who would like to assign a *degree of belief* to a particular statement φ. For example, a doctor may want to assign a degree of belief to the hypothesis that a patient has a particular illness, based on the symptoms exhibited by the patient together with general information about symptoms and diseases. Since the actions the agent takes may depend crucially on this degree of belief, we would like techniques for computing degrees of belief in a principled manner. In this paper we investigate the properties of two formalisms for computing degrees of belief, based on the same general approach.

The approach is an old one, going back to Laplace [Lap20]. It is essentially what has been called *the principle of insufficient reason* [Kri86] or *the principle of indifference* [Key21]. The idea is to assign equal degree of belief to all basic "situations" consistent with the known facts. We consider two embodiments of this principle, which we call the *random-worlds method* and the *random-structures method*; they differ in how the idea of "situation" is interpreted.

In the random-worlds method, if we know that the domain has size N, the possible situations are all the worlds, or first-order models, with domain $\{1, \ldots, N\}$ that satisfy the constraints θ. We then compute the fraction of them in which the formula φ is true, and take this fraction to be our degree of belief in φ. The random-worlds method views each individual in $\{1, \ldots, N\}$ as having a distinct name (even though the name may not correspond to any constant in the vocabulary). Thus, two worlds that are isomorphic with respect to the symbols in the vocabulary are still treated as distinct situations. In contrast, the random-structures method only counts isomorphism classes of worlds (taking these as the basic situations). Intuitively, this says that structures that are indistinguishable by the language should only be counted once.[1]

[1] The random-worlds method considers what has been called in the literature *labeled* structures, while the random-structures method considers *unlabeled* structures [Com88]. We choose to

We often do not know the precise domain size N, but do know that it is large. Thus, we are particularly interested in knowing the asymptotic behavior of the degree of belief. Thus, we want to know (for both methods) whether the fraction of situations where φ is true converges to some limit as the domain size N gets large and, if it does, whether we can we characterize what this asymptotic probability is. The answer to these questions depends in part on the language that φ and θ are expressed in. Ultimately, we are interested in a language that allows first-order assertions (like "all penguins are birds") and statistical assertions (like "90% of birds fly"). Since the issues that arise with statistical assertions are significantly different from those that arise with first-order assertions, in this paper we restrict attention to the pure first-order case; in a companion paper [GHK92] we consider the case of statistical assertions.

To readers familiar with the work on asymptotic probabilities for various logics, our questions will seem quite familiar. For example, in the context of first-order formulas, it is well known that if we consider a formula with no constant or function symbols, then (for both methods) its asymptotic probability is either 0 or 1 [GKLT69, Fag76]. Moreover, it is known that both methods give the same asymptotic probability [Fag77]; as we shall see, this is not the case once we move to conditional probabilities. After the initial results, 0-1 laws were proved for special classes of first-order structures (such as graphs, tournaments, partial orders, etc.; see the overview paper [Com88] for details and further references). In many cases, the classes considered could be defined in terms of first-order constraints. Thus, these results can be viewed as special cases of the problem that we are interested in: computing asymptotic *conditional* probabilities relative to structures satisfying the constraints of a database. Lynch [Lyn80] showed that, for the random-worlds method, asymptotic limits exist for first-order formulas involving unary functions, although there is no 0-1 law. (Recall that the original 0-1 result is specifically for first-order logic *without* function symbols.) This can also be viewed as a special case of a conditional asymptotic probability for first-order logic without functions, since we can replace the unary functions by binary predicates, and condition on the fact that they are functions.

Despite all this work on special cases, to the best of our knowledge there has been no work on the general problem of asymptotic conditional probabilities. Perhaps the main reason for this is that it is well known [Fag76] that we do *not* always get convergence of conditional probabilities. In this paper, motivated by our desire to apply these methods to computing degrees of belief, we consider the general question of when asymptotic probabilities exist for first-order logic, and how to compute them if they do.

use our own terminology for these methods, partly because we feel it is more descriptive, and partly because there are other variants of the approach that do not fit into the standard labeled/unlabeled structures dichotomy (see [BGHK92]) .

We begin by showing that once we have a binary predicate symbol in the vocabulary, things are essentially as bad as possible. Not only does there not exist an asymptotic probability in general, but almost all the questions one might want to ask (such as deciding whether the asymptotic probability exists) are highly undecidable. Things get better if we restrict attention to *unary vocabularies*, *i.e.,* vocabularies consisting only of unary predicate symbols, equality, and constant symbols. In this case, the asymptotic conditional probability of φ given θ exists iff θ is satisfied in sufficiently large models (we define what we mean by "sufficiently large" later in the paper). We show that it is decidable whether the asymptotic limit exists and, if it does, there is an effective algorithm for computing it. Unary predicates are of great interest because they express properties of individuals. (For example, symptoms and diseases can be represented as unary predicates.)

We then turn our attention to the complexity of computing the asymptotic probability in the unary case. Our results, which are the same for random worlds and random structures, are summarized in Table 1. As the table shows, the complexity depends on two parameters. One is whether we have a fixed and finite or an infinite vocabulary, and the other is whether we allow bounded or unbounded quantification depth. For example, with an infinite vocabulary and unbounded quantification, checking whether the limit exists is NEXPTIME-complete, while the problem of computing the probability if it is exists is essentially #EXP-complete (where #EXP is the exponential time analogue of #P [Val79a]). Somewhat surprisingly in light of our other results, these results hold even if we restrict to formulas of quantification depth two. For unquantified formulas or depth one quantification, things become an exponential factor easier.[2] If we consider any fixed finite vocabulary, the problem of checking whether the limit exists and computing it are PSPACE-complete with unbounded quantification. On the other hand, for any bound on the quantification depth, we can effectively find a linear time and logarithmic space algorithm that solves the problem (where the constants depend on the bound). In the cases where computing the asymptotic probability is hard, we can show that finding a nontrivial estimate of the probability (*i.e.,* deciding if it lies in a nontrivial interval) is almost as difficult.

These results are of more than purely technical interest. The random-worlds method is of considerable

[2] The #P upper-bound for computing the limits requires a nonstandard definition of quantifier depth. The precise definition is found in Section 5.2.

	fixed finite vocabulary		infinite vocabulary	
	bounded depth	unbounded depth	depth < 2	depth ≥ 2
do limits exist?	in linear time	PSPACE-complete	NP-complete	NEXPTIME-complete
compute limits	in linear time	PSPACE-complete	#P-complete	#EXP-complete
approx. limits	in linear time	PSPACE-complete	(co-)NP-hard	(co-)NEXPTIME-hard

Table 1: Complexity of asymptotic probabilities for unary vocabularies

theoretical and practical importance. We have already mentioned its relevance to computing degrees of belief. There are well-known results from physics that show the close connection between the random-worlds method and *maximum entropy* [Jay78]; some formalization of similar results, but in a framework that is close to that of the current paper, can be found in [PV89, GHK92]. Essentially, the results say that in certain cases the asymptotic probability can be computed using maximum entropy methods.[3]

Given the wide use of maximum entropy, and its justification in terms of the random-worlds method, our results showing that it is not as widely applicable as one might hope come as somewhat of a surprise. Indeed, the difficulties of using the method once we move to binary predicates seem not to have been fully appreciated. In retrospect, this is not that hard to explain; in almost all applications where maximum entropy has been used the database is described in terms of unary predicates (or, equivalently, unary functions with a finite range, which of course can be encoded using unary predicates). For example, in physics applications we are interested in such predicates as quantum state (see [DD85]). Similarly, AI applications and expert systems typically use only unary predicates ([Che83]) such as symptoms and diseases. In these examples, the vocabulary typically used is relatively small, and quantifiers are rarely deeply nested. This, of course, is precisely the case where our linear time result applies.

2 Limiting probabilities

Let Φ be a set of of predicate and function symbols, and let $\mathcal{L}(\Phi)$ (resp., $\mathcal{L}^-(\Phi)$) denote the set of first-order formulas over Φ with equality (resp., without equality). In order to simplify the presentation, we begin by assuming that Φ is finite; the case of an infinite vocabulary is deferred to Section 2.3.

[3] These results are of far more interest when there are statistical assertions in the language, so we do not discuss them here (see [PV89, GHK92] for more details). We remark that our computations of the asymptotic probability in the case of pure first-order logic over unary predicates can be understood in terms of maximum entropy, but this interpretation does not give much insight in this case.

2.1 The random-worlds method

We begin by defining the random-worlds, or labeled, method. Given a formula $\xi \in \mathcal{L}(\Phi)$, let $\#world_N^\Phi(\xi)$ be the number of worlds, or first-order models, of ξ over Φ with domain $\{1,\ldots,N\}$. Note that the assumption that Φ is finite is required for $\#world_N^\Phi(\xi)$ to be well-defined. Let $\Pr_N^{w,\Phi}(\varphi|\theta) = \frac{\#world_N^\Phi(\varphi\wedge\theta)}{\#world_N^\Phi(\theta)}$. At first glance, it seems that the value of $\Pr_N^{w,\Phi}(\varphi|\theta)$ depends on the choice of Φ. The following proposition shows that this is not the case.

Proposition 2.1: *Let Φ, Φ' be finite vocabularies, and let φ, θ be formulas in both $\mathcal{L}(\Phi)$ and $\mathcal{L}(\Phi')$. Then $\Pr_N^{w,\Phi}(\varphi|\theta) = \Pr_N^{w,\Phi'}(\varphi|\theta)$.*

Based on this proposition, we omit reference to Φ in $\Pr_N^{w,\Phi}(\varphi|\theta)$, writing $\Pr_N^w(\varphi|\theta)$ instead.

We would like to define $\Pr_\infty^w(\varphi|\theta)$ as the limit $\lim_{N\to\infty} \Pr_N^w(\varphi|\theta)$. There is a small technical problem we have to deal with in this definition: we must decide what to do if $\#world_N^\Phi(\theta) = 0$, so that $\Pr_N^w(\varphi|\theta)$ is not well-defined. It might seem reasonable to say that the asymptotic probability is not well-defined if $\#world_N^\Phi(\theta) = 0$ for infinitely many N. However, suppose that θ is a formula that is true only when N is even and, for even N, $\varphi \wedge \theta$ holds in one third of the models of θ. In this case, we might want to say that there is an asymptotic conditional probability of 1/3, even though $\#world_N^\Phi(\theta) = 0$ for infinitely many N. Thus, we actually consider two notions: the persistent limit, denoted $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$, and the intermittent limit, denoted $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$ (the prefixes stand for the temporal logic representation of the persistence and intermittence properties [MP92]). In either case, we say that the limiting probability is either not well-defined, does not exist, or is some fraction between 0 or 1. The only difference between the two notions lies in when the limiting probability is taken to be well-defined. This difference is made precise in the following definition.

Definition 2.2: The asymptotic probability $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ is *well-defined* if $\#world_N^\Phi(\theta) \neq 0$ for all but finitely many N; $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$ is *well-defined* if $\#world_N^\Phi(\theta) \neq 0$ for infinitely many N. If the asymptotic probability $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ (resp., $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$) is well-defined, then we take $\Diamond\Box\Pr^w_\infty(\varphi|\theta)$ (resp., $\Box\Diamond\Pr^w_\infty(\varphi|\theta)$) to denote $\lim_{N\to\infty, \#world^\Phi_N(\theta)\neq 0} \Pr^w_N(\varphi|\theta)$. ∎

It will follow from our results that the two notions of limiting probability coincide if we restrict to unary predicates or to languages with no equality. In general, it is clear that if $\Diamond\Box\Pr^w_\infty(\varphi|\theta)$ is well-defined, then so is $\Box\Diamond\Pr^w_\infty(\varphi|\theta)$; the converse is not necessarily true.

2.2 The random-structures method

Even if two different worlds are isomorphic (as first-order structures), they contribute seperately to the count $\#world^\Phi_N(\xi)$. In contrast, the random-structures method is based on the intuition that we should be counting distinct *structures*, or *isomorphism classes* of models instead.

Formally, we proceed as follows. Given a formula $\xi \in \mathcal{L}(\Phi)$, let $\#struct^\Phi_N(\xi)$ be the number of isomorphism classes of worlds with domain $\{1, \ldots, N\}$ over the vocabulary Φ satisfying φ. We can then proceed, as before, to define $\Pr^{s,\Phi}_N(\varphi|\theta)$ as $\frac{\#struct^\Phi_N(\varphi\wedge\theta)}{\#struct^\Phi_N(\theta)}$. We define the persistent limit, denoted $\Diamond\Box\Pr^{s,\Phi}_\infty(\varphi|\theta)$, and the intermittent limit, denoted $\Box\Diamond\Pr^{s,\Phi}_\infty(\varphi|\theta)$, in terms of $\Pr^{s,\Phi}_N(\varphi|\theta)$, in analogy to the earlier definitions for random-worlds. It is clear that $\#world^\Phi_N(\theta) = 0$ iff $\#struct^\Phi_N(\theta) = 0$, so that well-definedness (both persistent and intermittent) is equivalent for the two methods, for any φ, θ.

As the following example shows, for the random-structures method, the analogue to Proposition 2.1 does not hold; the value of $\Pr^{s,\Phi}_N(\varphi|\theta)$, and even the value of the limit, depends on the choice of Φ. This example, together with Proposition 2.1, also demonstrates that the values of conditional probabilities generally differ between the random-worlds method and the random-structures method. This is in contrast to the case for unconditional probabilities; Fagin [Fag76] showed that the random-worlds and random-structures methods give the same answers for unconditional probabilities, if we do not have constant or function symbols in the language.

Example 2.3: Consider $\Phi = \{P\}$ for a unary predicate P. Let θ be $\exists! x\, P(x) \vee \neg\exists x\, P(x)$ (where, as usual, "$\exists!$" means "exists a unique"), and let φ be $\exists x\, P(x)$. For any domain size N, $\#struct^\Phi_N(\theta) = 2$: In one structure, there is exactly one element satisfying P and $N-1$ satisfying $\neg P$; in the other, all elements satisfy $\neg P$. Therefore, $\Diamond\Box\Pr^{s,\Phi}_\infty(\varphi|\theta) = \frac{1}{2}$.

Now, consider $\Phi' = \{P, Q\}$, for a new unary predicate Q. There are $2N$ structures where there exists an element satisfying P: the element satisfying P may or may not satisfy Q, and of the $N-1$ elements satisfying $\neg P$, any number between 0 and $N-1$ may also satisfy Q. On the other hand, there are $N+1$ structures where all elements satisfy $\neg P$: any number of elements between 0 and N may satisfy Q. Therefore, $\Pr^{s,\Phi'}_N(\varphi|\theta) = \frac{2N}{3N+1}$, and $\Pr^{s,\Phi'}_\infty(\varphi|\theta) = \frac{2}{3}$.

We know that the asymptotic limit for the random-worlds method will be the same, whether we use Φ or Φ'. Using Φ, notice that the single structure where $\exists! x\, P(x)$ is true contains N worlds (corresponding to the choice of element satisfying P), whereas the other possible structure contains only one world. Therefore, $\Diamond\Box\Pr^w_\infty(\varphi|\theta) = 1$. ∎

2.3 Infinite vocabularies

Up to now, we have assumed that the vocabulary Φ is finite. As we observed, this assumption is crucial in our definitions of $\#world^\Phi_N(\xi)$ and $\#struct^\Phi_N(\xi)$. Nevertheless, in many standard complexity arguments, it is crucial that the vocabulary is infinite. For example, satisfiability for propositional logic formulas is decidable in linear time if we assume a fixed finite vocabulary. We need to assume an infinite vocabulary to get NP-completeness.

How can we modify the random-worlds and random-structures methods to deal with an infinite vocabulary Φ? The issue is surprisingly subtle. One plausible choice depends on the observation that even if Φ is infinite, the set of symbols appearing in a given formula is always finite. We can thus do our computations relative to this set. More formally, if $\Phi_{\varphi\wedge\theta}$ denotes the set of symbols appearing in $\varphi\wedge\theta$, we could define $\Pr^{w,\Phi}_N(\varphi|\theta) = \Pr^{w,\Phi_{\varphi\wedge\theta}}_N(\varphi|\theta)$. Similarly, for the random-structures method, we could define $\Pr^{s,\Phi}_N(\varphi|\theta) = \Pr^{s,\Phi_{\varphi\wedge\theta}}_N(\varphi|\theta)$. The problem with this approach is that the values given by the random-structures approach depend on the vocabulary, and it is easy to find two equivalent formulas φ and φ' such that $\Phi_\varphi \neq \Phi_{\varphi'}$ and $\Pr^{s,\Phi}_N(\varphi|\theta) \neq \Pr^{s,\Phi}_N(\varphi'|\theta)$. (A simple example of this phenomenon can be obtained by modifying Example 2.3 slightly.) Thus, under this approach, the value of asymptotic conditional probabilities can depend on the precise syntax of the formulas involved. We view this as undesirable, and so in the following we concentrate on two different interpretations of the idea of an infinite vocabulary.[4]

The first of these two alternative approaches treats an infinite vocabulary as a limit of finite sub-vocabularies. Assume for ease of exposition that Φ is countable. Let Φ_m consist of the first m symbols in Φ. We can then define $\Pr^{w,\Phi}_N(\varphi|\theta) = \lim_{m\to\infty} \Pr^{w,\Phi_m}_N(\varphi|\theta)$ (where we take $\Pr^{w,\Phi_m}_N(\varphi|\theta)$ to be undefined if $\varphi, \theta \notin \mathcal{L}(\Phi_m)$). Similarly, we can define $\Pr^{s,\Phi}_N(\varphi|\theta) = \lim_{m\to\infty} \Pr^{s,\Phi_m}_N(\varphi|\theta)$.

[4] We note, however, that all our later complexity results concerning infinite vocabularies can be easily shown to hold for the definition just discussed.

The second possibility is quite different. In it, although there may be an infinite vocabulary in the background, we assume that each problem instance comes along with a finite vocabulary as part of the input. Thus, in our infinite vocabulary, we may have predicates that are relevant to medical applications, scientific applications, automobile insurance applications, etc. When thinking about medical applications, we use that finite portion of the infinite vocabulary that is appropriate. In this approach, we always deal with finite vocabularies, but ones whose size is potentially unbounded because we do not fix the relevant vocabulary in advance.

In essence, the first approach can be viewed as saying that there really is an infinite vocabulary, while the second approach considers there to be an infinite collection of finite (but unbounded) vocabularies. The distinction between these possibilities is not usually examined as closely as we have done here. This is because the difference is rarely important. For example, propositional satisfiability is NP-complete over an infinite vocabulary, no matter how we interpret "infinite." In our context, the difference turns out to be moderately significant. For random worlds, an argument based on Proposition 2.1 shows the two approaches lead to the same answers (as does the approach that we rejected where, when computing $\Pr_N^{\Phi}(\varphi|\theta)$, we restrict the vocabulary to $\Phi_{\varphi\wedge\theta}$). However, the two approaches can lead to quite different answers in the case of the random-structures approach.

In fact, and somewhat surprisingly, the random-worlds and random-structures methods turn out to agree if we use the first, limit, interpretation of infinite vocabularies. That is, $\Box\Diamond\Pr_{\infty}^{w,\Phi}(\varphi|\theta)$ exists iff $\Box\Diamond\Pr_{\infty}^{s,\Phi}(\varphi|\theta)$ does, and then $\Box\Diamond\Pr_{\infty}^{w,\Phi}(\varphi|\theta) = \Box\Diamond\Pr_{\infty}^{s,\Phi}(\varphi|\theta)$ (and similarly for the persistent limit). This is a consequence of the following result. If $F(N)$ and $G(N)$ are two functions of N, we write $F(N) \sim G(N)$ if $\lim_{N\to\infty} F(N)/G(N) = 1$.

Theorem 2.4: *If Φ is an infinite vocabulary, and $\varphi,\theta \in \mathcal{L}(\Phi)$, then $\Pr_N^{w,\Phi}(\varphi|\theta) \sim \Pr_N^{s,\Phi}(\varphi|\theta)$.*

3 Languages with binary predicates

In this section, we discuss asymptotic conditional probabilities over vocabularies that include at least one binary predicate symbol. These probabilities do not always exist, and almost all interesting questions about well-definedness, existence, and the values of limits are undecidable in general. Finally, we indicate why it appears to be hard to find interesting restrictions on the language—other than avoiding binary predicates completely—that give effectively computable limits. Proofs of all results can be found in the full paper.

3.1 Nonexistence

As we mentioned in the introduction, the fact that asymptotic conditional probabilities do not always exist is well known.

Theorem 3.1: *[Fag76] Let S be a binary predicate symbol. There exist formulas $\varphi,\theta \in \mathcal{L}(S)$ such that neither $\Box\Diamond\Pr_{\infty}^{w}(\varphi|\theta)$ nor $\Diamond\Box\Pr_{\infty}^{w}(\varphi|\theta)$ exists, although both are well-defined.*

The proof of this theorem is quite straightforward. Using a binary prediate (and equality), it is not hard to construct formulas *even* and *odd* such that *even* is true only in domains of even size, and *odd* is true only in domains of odd size. We then take φ to be *odd* and θ to be *even* $\vee$ *odd*. Clearly, this probability alternates between 0 and 1 as N increases, and does not approach an asymptotic limit.

Although this shows that the asymptotic limit does not exist in general, a good argument can be made that in this case there is a reasonable degree of belief that one can hold. In the absence of any information about domain size, 1/2 seems the natural answer. Perhaps if we modified our definition of degree of belief slightly, we could increase the applicability of our techniques.

There is indeed a reasonable modification that will let us assign a degree of belief of 1/2 in this case: we can use Cesàro limits instead of regular limits.[5] The Cesàro limit of a sequence $f(n)$ is the limit of the average value of $f(1),\ldots,f(n)$, as n goes to infinity. It is well known that if the regular limit exists, then so does the Cesàro limit, and they are equal. However, there are times when the Cesàro limit exists and the regular limit does not. For example, for a sequence of the form $1,0,1,0,\ldots$ (which, of course, is precisely the sequence that arises in the proof of Theorem 3.1), the regular limit does not exist, but the Cesàro limit does, and is 1/2.

Unfortunately, we can show that for any definition of limit satisfying some very basic restrictions (see [PS72] for some definitions), the limit of the conditional probabilities may not exist. That is, for any notion of limit in a large class, there exist sentences $\varphi,\theta \in \mathcal{L}(\{S\})$ (for a binary predicate S), such that the limit of (the well-defined subsequence of) $\Pr_N^{w}(\varphi|\theta)$ (resp., $\Pr_N^{s,\{S\}}(\varphi|\theta)$) does not exist, under that definition of limit. In particular, the Cesàro limit satisfies these restrictions; therefore, even for Cesàro limits, the non-existence problem still arises. This result is stated formally, and proved, in the full paper.

[5] We remark that Cesàro limits have been used before in the context of 0-1 laws; see Compton's overview [Com88] for details and further references.

We remark that, with a little more work, we can also show that equality is not necessary to obtain the non-existence of limits. We can avoid the use of equality at the cost of adding two binary predicates.

3.2 Complexity analysis

Many important decision problems associated with asymptotic conditional probabilities are undecidable. We analyze the complexity of these problems in terms of the *arithmetic hierarchy*. This is a hierarchy that extends the notions of r.e. (recursively enumerable) and co-r.e. sets. We briefly review the relevant definitions here, referring the reader to [Rog67] for further details. Consider a formula ξ in the language of arithmetic (*i.e.*, using $0, 1, +, \times$) having k free variables. The formula ξ, interpreted over the natural numbers, is said to define a *recursive set* if the set of k-tuples satisfying the formula is a recursive set. We can define more complex sets using quantification. We define a Σ_k^0 *prefix* as a block of quantifiers of the form $\exists x_1 \ldots x_h \forall y_1 \ldots y_m \ldots$, where there are k alternations of quantifiers. A Π_k^0 prefix is defined similarly, except that the quantifier block starts with a universal quantifier. A set A of natural numbers is in Σ_k^0 if there is a first-order formula $\xi(x) = Q\xi'$ in the language of arithmetic with one free variable x, where Q is a Σ_k^0 quantifier block and ξ' defines a recursive set, such that $n \in A$ iff $\xi(n)$ is true. We can similarly define what it means for a set to be in Π_k^0. A set is in Σ_1^0 iff it is r.e., and it is in Π_1^0 iff it is co-r.e. The hierarchy is known to be strict; higher levels of the hierarchy correspond problems which are "more undecidable."

In the full paper we prove the following results:

- Deciding whether $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$ is well-defined is Π_2^0-complete.
- Deciding whether $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ is well-defined is Σ_2^0-complete.
- Deciding whether $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ (resp., $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$) exists is Π_3^0-complete, even given an oracle that decides if the limit is well-defined.
- Deciding whether the limit $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ (resp., $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$) is in any fixed closed interval $[r_1, r_2] \subset [0, 1]$, where r_1 and r_2 are rational, is Π_2^0-complete, given an oracle that decides if the limit exists (note that $r_1 = r_2$ is allowed).

The lower bounds all use a fixed finite vocabulary, consisting of equality and a single binary predicate. In order to prove these results for a language without equality, or for the case of random structures, one binary predicate is not sufficient, but we never need more than four. For a language without equality, the first two problems drop in complexity to r.e.-completeness; the rest of the results continue to hold unchanged. In addition, we can show that all these results, both with and without equality, also hold for the random-structures method.

3.3 Is there any hope?

These results show that most interesting problems regarding asymptotic probabilities are badly undecidable in general. Are there restricted sublanguages for which these questions become tractable, or at least decidable? We focus on one important case in most of the remainder of the paper, namely, that of unary vocabularies. Here, we consider one other case.

The hardness results presented in the previous section hold for a language consisting of a single binary predicate symbol, but although the proofs of these results involve formulas with bounded alternation depth, they nevertheless require deeply nested quantifiers. It is possible to modify the same proofs to work for sentences with small, fixed, quantifier depth but with no bound on the size of vocabulary needed. What happens if there is a simultaneous bound on quantifer depth and vocabulary size?

Definition 3.2: Define $d(\xi)$ to be the *depth of quantifier nesting* in the formula ξ:

- $d(\xi) = 0$ for any atomic formula ξ,
- $d(\neg\xi) = d(\xi)$,
- $d(\xi_1 \land \xi_2) = \max(d(\xi_1), d(\xi_2))$,
- $d(\exists y\xi) = d(\xi) + 1$. ∎

Theorem 3.3: *Let Φ be a fixed, finite vocabulary, and let d be a constant. Then there exists a Turing machine* $\mathbf{M}$ *such that for any $\varphi, \theta \in \mathcal{L}(\Phi)$ of quantification depth $\leq d$,* $\mathbf{M}$ *decides in linear time whether $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ (resp., $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$, $\Diamond\Box\Pr_\infty^{s,\Phi}(\varphi|\theta)$, $\Box\Diamond\Pr_\infty^{s,\Phi}(\varphi|\theta)$) is well-defined, if so whether it exists, and if it exists computes an arbitrarily good rational approximation to its value.*

The proof of this theorem is based on the fact that there are only finitely many non-equivalent formulas of depth $\leq d$ over a fixed vocabulary. From the proof, it is easy to show that, given d and Φ, we can construct a finite family of linear-time Turing machines, one of which correctly computes the asymptotic probability. Unfortunately, this result is of little practical use because we cannot effectively tell which of the machines is the right one. If we could effectively describe how to construct the "correct" $\mathbf{M}$ from Φ and d, this would contradict all our earlier undecidability results. Thus, even for this extremely restrictive sublanguage we cannot effectively construct algorithms for computing degrees of belief.

4 Computing the limit for unary vocabularies

All our negative results so far depend on having at least one binary predicate symbol in the vocabulary; as we indicated in the introduction, this binary predicate symbol is essential. If we restrict attention to unary predicate and constant symbols, then the conditional asymptotic probability of φ given θ exists iff θ is satisfiable in arbitrarily large models. The question of whether θ is satisfiable in arbitrarily large models is decidable. Moreover, if the conditional asymptotic probability exists, then it is effectively computable. This gap between unary predicates and binary predicates is somewhat reminiscent of the fact that first-order logic over a vocabulary with only unary predicates (and constant symbols) is decidable, while if we allow even a single binary predicate symbol, then it becomes undecidable [DG79, Lew79]. This similarity is not coincidental; some of the techniques used to show that first-order logic over a vocabulary with unary predicate symbols is decidable are used here to show that asymptotic limits exist.

We begin by showing how the fact that the vocabulary is unary allows us to simplify formulas considerably. We then give precise expressions for the value of the asymptotic conditional probability for both the random-worlds and the random-structures methods, and discuss an algorithm for computing them.

4.1 Model descriptions

Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ denote a set of unary predicate symbols and let $\mathcal{C}$ denote a finite set of constant symbols. In this section, we always take $\Phi = \mathcal{P} \cup \mathcal{C}$ unless we explicitly state otherwise. We want to compute the asymptotic conditional probability of formulas in the language $\mathcal{L}(\Phi)$.

Our first step is to characterize the models of these formulas. It is well known [DG79], that if a formula in $\mathcal{L}(\Phi)$ is satisfiable at all, then it is satisfiable in a "small" model (one of size at most exponential in the size of the formula). We actually prove a somewhat stronger result from which this fact follows.

We start with some definitions. We define an *atom* over $\mathcal{P}$ to be a conjunction of the form $Q_1 \wedge \ldots \wedge Q_k$, where each Q_i is either P_i or $\neg P_i$. Note that there are $2^k = 2^{|\mathcal{P}|}$ atoms over $\mathcal{P}$, and that they are mutually exclusive and exhaustive. We use $A_1, \ldots, A_{2^{|\mathcal{P}|}}$ to denote the atoms over $\mathcal{P}$, listed in some fixed order.

A *model description over* Φ *of size* M is a syntactic characterization of a model. It describes exactly how many elements in the domain satisfy each atom (except that if there are $\geq M$ elements satisfying the atom, we say just that, rather than giving the exact count); for each element of $\mathcal{C}$, it describes which atom that element satisfies; and for each pair of elements in $\mathcal{C}$, it describes whether they are equal.

More formally, we proceed as follows. Given a formula $\xi(x)$ with one free variable x, we take $\exists^m x\, \xi(x)$ to be the formula that says there are precisely m domain elements satisfying ξ. Similarly, we define $\exists^{\geq m} x\, \xi(x)$ to be the formula that says that there are at least m domain elements satisfying ξ.

Definition 4.1: A *size* M *model description (over* Φ*)* consists of a conjunction of three types of formulas:

- $2^{|\mathcal{P}|}$ formulas of the form $\exists^m x\, A_i(x)$ for $m < M$ or $\exists^{\geq M} x\, A_i(x)$, one for each atom over $\mathcal{P}$,
- one formula of the form $A_i(c)$ for each element $c \in \mathcal{C}$, describing which atom c satisfies,
- a formula of the form $c = c'$ or $c \neq c'$ for each pair of elements $c, c' \in \mathcal{C}$ that satisfy the same atom (*i.e.,* such that $A_i(c)$ and $A_i(c')$, for some atom A_i, are both formulas in the model description).

Our interest in model descriptions is motivated by the following result, which is easily proved by induction on the quantifier depth of the sentence.

Theorem 4.2: *If* ξ *is a sentence in* $\mathcal{L}(\Phi)$*, then* ξ *is equivalent to a disjunction of consistent model descriptions of size* M *over* Φ*.*

Definition 4.3: For $\xi \in \mathcal{L}(\Phi)$, we define $\mathcal{M}_\xi^\Phi$ to be the set of consistent model descriptions of size $d(\xi) + |\mathcal{C}|$ over Φ such that ξ is equivalent to the disjunction of the model descriptions in $\mathcal{M}_\xi^\Phi$. ∎

4.2 Counting worlds

In our analysis of the random-worlds method, we first want to compute $\#world_N^\Phi(\xi)$ for a sentence ξ. As the following lemma shows, it suffices to compute $\#world_N^\Phi(\psi)$ for a model description ψ.

Lemma 4.4: *Let* ξ *be a sentence in* $\mathcal{L}(\Phi)$*. Then* $\#world_N^\Phi(\xi) = \sum_{\psi \in \mathcal{M}_\xi^\Phi} \#world_N^\Phi(\psi)$*.*

Thus, we now focus on the problem of computing $\#world_N^\Phi(\psi)$ for a model description ψ.

Definition 4.5: The *characteristic* of ψ is a tuple of the form $\langle (d_1, e_1), \ldots, (d_{2^{|\mathcal{P}|}}, e_{2^{|\mathcal{P}|}}) \rangle$, where

- $d_i = m$ if exactly $m < M$ domain elements satisfy A_i according to ψ,
- $d_i = *$ if at least M domain elements satisfy A_i according to ψ,
- e_i is the number of distinct domain elements which are interpretations of elements in $\mathcal{C}$ that satisfy A_i according to ψ. ∎

Note that we can compute e_i immediately from ψ.

Definition 4.6 :
Suppose $C = \langle (d_1, e_1), \ldots, (d_{2^{|\mathcal{P}|}}, e_{2^{|\mathcal{P}|}}) \rangle$ is the characteristic of ψ. We say that a component i in C is *active* if $d_i = *$; otherwise i is *passive*. Let $A(C)$ be the set of active components of C. We define

$$\begin{aligned} \Delta_1(\psi) &= |A(C)| \\ \Delta_2(\psi) &= \sum_{i \in A(C)} e_i + \sum_{i \notin A(C)} d_i \\ H(\psi) &= \prod_{i \notin A(C)} (d_i - e_i)! \end{aligned}$$ ∎

Lemma 4.7: *Let ψ be a consistent model description over Φ of size $M \geq |\mathcal{C}|$. If If $\Delta_1(\psi) = 0$ and $N > KM$ then $\#world_N^{\Phi}(\psi) = 0$. If $\Delta_1(\psi) > 0$, then $\#world_N^{\Phi}(\psi) \sim \frac{N^{\Delta_2(\psi)}(\Delta_1(\psi))^N}{H(\psi)\Delta_1(\psi)^{\Delta_2(\psi)}}$.*

Definition 4.8: Given a model description ψ over Φ, let the *degree* of ψ, written $\Delta(\psi)$, be the pair $(\Delta_1(\psi), \Delta_2(\psi))$. We order degrees lexicographically. We extend this definition to formulas. For $\theta \in \mathcal{L}(\Phi)$, we define the *degree of θ over* Φ, written $\Delta^{\Phi}(\theta)$, to be $\max_{\psi \in \mathcal{M}_\theta^\Phi} \Delta(\psi)$, and the *activity count* of θ, written $\Delta_1^{\Phi}(\theta)$, to be $\max_{\psi \in \mathcal{M}_\theta^\Phi} \Delta_1(\psi)$. Finally, for any degree $\delta = (\delta_1, \delta_2)$, let $\mathcal{M}_\theta^{\Phi,\delta}$ be the set of model descriptions $\psi \in \mathcal{M}_\theta^\Phi$ such that $\Delta(\psi) = \delta$. ∎

As Lemma 4.7 shows, asymptotically the number of models of ψ is completely determined by its degree. Moreover, if ψ_1 and ψ_2 are two model descriptions over Φ such that $\Delta(\psi_1) > \Delta(\psi_2)$, we have $\lim_{N\to\infty} \#world_N^{\Phi}(\psi_2)/\#world_N^{\Phi}(\psi_1) \to 0$. Thus, model descriptions whose degree is higher will dominate. These observations can be used in order to give us a precise formula for the asymptotic conditional probability. Using Lemmas 4.4 and 4.7, it is not hard to show:

Theorem 4.9: *Suppose $\varphi, \theta \in \mathcal{L}(\Phi)$. If $\Delta_1^{\Phi}(\theta) = 0$, then neither $\Box\Diamond\Pr_\infty^w(\varphi|\theta)$ nor $\Diamond\Box\Pr_\infty^w(\varphi|\theta)$ is well-defined. If $\Delta_1^{\Phi}(\theta) > 0$, then, for $\delta = \Delta^{\Phi}(\theta)$,*

$$\Box\Diamond\Pr_\infty^w(\varphi|\theta) = \Diamond\Box\Pr_\infty^w(\varphi|\theta) = \frac{\sum_{\psi \in \mathcal{M}_{\varphi\wedge\theta}^{\Phi,\delta}} 1/H(\psi)}{\sum_{\psi \in \mathcal{M}_{\theta}^{\Phi,\delta}} 1/H(\psi)} .$$

Thus, Theorem 4.9 gives us an explicit formula for calculating conditional asymptotic limits for the random-worlds method, when the limits exist.

Since the intermittent and persistent limits agree if we have only unary predicate symbols in the vocabulary, in the sequel we drop the prefixes $\Box\Diamond$ and $\Diamond\Box$. Note that this also holds for the random-structures method, since the limit is well-defined in the random-worlds method iff it is well-defined in the random-structures method.

4.3 Counting structures

We can carry out a similar analysis for the random-structures method in order to compute $\#struct_N^{\Phi}(\xi)$. The key result is the following analogue of Lemma 4.7, which shows that, for the random-structures method, model descriptions of higher activity count $\Delta_1(\psi)$ (rather than model descriptions of higher degree $\Delta(\psi)$) dominate.

Lemma 4.10: *Let ψ be a consistent model description over Φ of size $M \geq |\mathcal{C}|$. If $\Delta_1(\psi) = 0$ and $N > KM$, then $\#struct_N^{\Phi}(\psi) = 0$. If $\Delta_1(\psi) > 0$ then $\#struct_N^{\Phi}(\psi) \sim \frac{N^{\Delta_1(\psi)-1}}{(\Delta_1(\psi)-1)!}$.*

In analogy to Definition 4.8, we define $\mathcal{M}_\xi^{\Phi,\delta_1}$ to be the set of $\psi \in \mathcal{M}_\xi^\Phi$ for which $\Delta_1(\psi) = \delta_1$. We now get the following analogue to Theorem 4.9.

Theorem 4.11: *Consider sentences $\varphi, \theta \in \mathcal{L}(\Phi)$. If $\delta_1 = \Delta_1^{\Phi}(\theta) = 0$, then neither $\Box\Diamond\Pr_\infty^{s,\Phi}(\varphi|\theta)$ nor $\Diamond\Box\Pr_\infty^{s,\Phi}(\varphi|\theta)$ is well-defined. If $\delta_1 > 0$, then*

$$\Box\Diamond\Pr_\infty^{s,\Phi}(\varphi|\theta) = \Diamond\Box\Pr_\infty^{s,\Phi}(\varphi|\theta) = \frac{|\mathcal{M}_{\varphi\wedge\theta}^{\Phi,\delta_1}|}{|\mathcal{M}_{\theta}^{\Phi,\delta_1}|} .$$

Theorem 4.11 gives us an explicit formula for calculating conditional asymptotic limits for the random-structures method, when the limits exist.

4.4 The basic algorithms

The previous results suggest a straightforward algorithm for computing asymptotic conditional probabilities. The algorithm is essentially the same for both random worlds and random structures. In essence, it generates model descriptions one by one, and checks whether they are consistent with φ and with θ. (The detailed algorithm is described in the full paper.) The following propositions give bounds on the algorithm's complexity.

Let $|\xi|$ be the length of the formula ξ.

Proposition 4.12: *Given a model description ψ of size M over Φ and a sentence $\xi \in \mathcal{L}(\Phi)$, where $M \geq d(\xi) + |\mathcal{C}|$, we can decide if ψ is consistent with ξ in time $O(T_\xi)$ and space $O(\log T_\xi)$, for $T_\xi = (M2^{|\mathcal{P}|})^{d(\xi)}|\xi|$.*

Proposition 4.13: *Given Φ and $\varphi, \theta \in \mathcal{L}(\Phi)$, the problem of deciding whether $\Pr_\infty^w(\varphi|\theta)$ (resp., $\Pr_\infty^{s,\Phi}(\varphi|\theta)$) is well-defined can be decided in time $O(T'_{\varphi\wedge\theta})$ and space $O(\log T'_{\varphi\wedge\theta})$, where $T'_{\varphi\wedge\theta} = T_{\varphi\wedge\theta} \cdot M^{2^{|\mathcal{P}|}} 2^{|\mathcal{P}||\mathcal{C}|} 2^{|\mathcal{C}|^2}$, and $M = d(\varphi \wedge \theta) + |\mathcal{C}|$. Moreover, if the limit is well-defined, its value can be computed in the same time bounds.*

We conclude this section with an example illustrating how we can use the tools developed in this section to compute the limits for the two methods.

Example 4.14: Consider $\mathcal{P} = \{P\}$, $\mathcal{C} = \{c, d\}$. Let $\varphi =_{\rm def} P(c) \wedge P(d)$, and $\theta =_{\rm def}$ *true*. Both θ and φ are equivalent to a disjunction of model descriptions of size 2. The model descriptions of maximal activity count consistent with θ all have activity count 2, and conjuncts stating that $\exists^{\geq 2} x\, P(x)$ and $\exists^{\geq 2} x\, \neg P(x)$. This table shows the model descriptions consistent with θ satisfying this property (the two conjuncts common to all of them are not shown).

ψ	$\Delta_2(\psi)$	$P(c) \wedge P(d)$
$P(c) \wedge P(d) \wedge c = d$	1	$\surd$
$P(c) \wedge P(d) \wedge c \neq d$	2	$\surd$
$P(c) \wedge \neg P(d)$	2	X
$\neg P(c) \wedge P(d)$	2	X
$\neg P(c) \wedge \neg P(d) \wedge c = d$	1	X
$\neg P(c) \wedge \neg P(d) \wedge c \neq d$	2	X

The number of model descriptions of maximal degree $(\Delta_1(\psi), \Delta_2(\psi))$ is four, and of these one is consistent with φ. Therefore, $\Pr^w_\infty(\varphi|\theta) = 1/4$. However, all six model descriptions have the same activity count, and count equally for the random-structures method. Of these, two model descriptions are consistent with φ. Therefore $\Pr^{s,\mathcal{P}\cup\mathcal{C}}_\infty(\varphi|\theta) = 1/3$. ∎

5 Complexity analysis for the unary case

In this section we consider the complexity of various problems relating to asymptotic limits, when the vocabulary uses unary predicates and constant symbols only. The results we give depend on two factors: whether the vocabulary is fixed and finite or whether it is infinite, and whether we can place a bound on the depth of quantifier nesting.

5.1 Fixed finite vocabularies

We first consider complexity of various problems when Φ is fixed and finite. By "fixed" we mean that Φ is not regarded as an input parameter in the various problems we consider, but is fixed in advance.

Theorem 5.1: *For fixed vocabulary Φ and φ, θ in $\mathcal{L}(\Phi)$, the problem of deciding whether $\Pr^w_\infty(\varphi|\theta)$ (resp., $\Pr^{s,\Phi}_\infty(\varphi|\theta)$) is well-defined is complete for PSPACE. The PSPACE lower bound holds even for sentences in $\mathcal{L}^-(\{P\})$. If the asymptotic probability is well-defined, there is an algorithm that computes it in polynomial space.*

Approximating the asymptotic probability is no easier than computing it, even given an oracle that tells us whether it is well-defined.

Theorem 5.2: *Fix ϵ with $0 < \epsilon < 1$, and let P be a unary predicate. For formulas $\varphi, \theta \in \mathcal{L}^-(\{P\})$, the problem of deciding whether $\Pr^w_\infty(\varphi|\theta) \in [0, 1-\epsilon]$ (resp., $\Pr^{s,\{P\}}_\infty(\varphi|\theta) \in [0, 1-\epsilon]$) is PSPACE-complete, even given an oracle that tells us whether the limit is well-defined.*

We cannot strengthen the lower bound in this result to hold for an arbitrary nontrivial interval in $[0, 1]$ bounded by rationals (which would make the form of this theorem closer all our other approximation results). The reason is that there are only finitely many nonequivalent formulas in $\mathcal{L}^-(\{P\})$, and so finitely many possible values for $\Pr^w_\infty(\varphi|\theta)$. Some closed intervals will contain none of these values, and the associated approximation problems are trivial.

The theorems just given show that our assumption in this section (that the vocabulary is fixed and finite) is not enough by itself to lead to computationally easy problems. Nevertheless, there is some good news. Examining Proposition 4.13 shows that much of the computational difficulty is due to quantifier nesting. In particular, we can show that for a fixed finite vocabulary and bounded quantifier depth, we can effectively construct a linear-time algorithm for deciding whether the asymptotic probability is well-defined, and computing it if it is. Of course, the constant factor in these algorithms' running times will be exponential in d and $2^{|\mathcal{P}|}$. Note that in the binary case, although we can show that a linear-time algorithm exists, we cannot construct it.

5.2 Infinite vocabularies

In this section we assume that we have a vocabulary consisting of a countably infinite collection of unary predicate symbols, $\mathcal{Q}$, and a countably infinite collection of constants, $\mathcal{D}$. As discussed in Section 2.3, there are at least two distinct ways of interpreting this assumption (which do turn out to be equivalent in the case of random worlds, but are truly different for random structures). Fortunately, the complexity results we give hold under both interpretations. In the case of the random-worlds method, this is easy to show since, as shown in Proposition 2.1, the limiting probability is essentially independent of the vocabulary. From Theorem 2.4, it follows that, asymptotically, the random-worlds and random-structures method give the same answers if we have an infinite vocabulary, so again, the complexity is the same in this case. It is not quite as obvious that the complexity results should be the same with our other interpretation of "infinite vocabulary", where we view the vocabulary as unbounded, and part of the input. To make it easier to see that this is indeed the case, we state our subsequent results in terms of this, somewhat more cumbersome, interpretation.

In the rest of this section, $\Phi = \mathcal{P} \cup \mathcal{C}$ always denotes some finite subset of $\mathcal{Q} \cup \mathcal{D}$. We remind the reader that we now view Φ as an input, along with the formulas φ and θ.

Some of the later results in this section are best stated in terms of a nonstandard definition of quantifier depth. This definition treats the presence of constants as a form of quantification.

Definition 5.3: A formula ξ is said to have *nonstandard quantifier depth* $d'(\xi) = \ell$ if it is of the form $\xi_1 \wedge \ldots \wedge \xi_r$, where for each i, $d(\xi_i) \leq \ell - 1$, or $d(\xi_i) \leq \ell$ and ξ_i contains no constant symbols. For any vocabulary Φ, $\mathcal{L}_\ell(\Phi)$ (resp., $\mathcal{L}_\ell^-(\Phi)$) denotes the set of sentences $\xi \in \mathcal{L}(\Phi)$ (resp., $\xi \in \mathcal{L}^-(\Phi)$) such that $d'(\xi) \leq \ell$. ∎

Subsequent complexity results will be presented in terms of nonstandard depth. Since $d(\xi) \leq d'(\xi)$ for any ξ, the fact that the lower bounds hold for this definition only strengthens the results. We remark that the only upper bound that depends on nonstandard depth is the one stating that computing asymptotic probabilities for formulas of nonstandard depth 1 is in #P.

Given this definition, we proceed to analyze the complexity for the case of an infinite vocabulary.

Theorem 5.4: *Given a finite vocabulary $\Phi \subset \mathcal{Q} \cup \mathcal{D}$ and $\varphi, \theta \in \mathcal{L}(\Phi)$, the problem of deciding if $\Pr_\infty^w(\varphi|\theta)$ (resp., $\Pr_\infty^{s,\Phi}(\varphi|\theta)$) is well-defined is complete for NEXPTIME. The NEXPTIME lower bound holds even if $\Phi \subset \mathcal{Q}$ and for sentences in $\mathcal{L}_2^-(\Phi)$.*

We next consider the problem of computing the asymptotic probability $\Pr_\infty^w(\varphi|\theta)$, given that it is well-defined. We show that this problem is #EXP-complete. Recall that #P (see [Val79a]) is the class of integer functions computable as the number of accepting computations of a nondeterministic polynomial-time Turing machine. The class #EXP is the exponential time analogue. That is, #EXP is the class of integer functions computable as the number of accepting computations of a nondeterministic exponential time Turing machine. The function we are interested in is $\Pr_\infty^w(\varphi|\theta)$, which is not integer valued. Nevertheless, we want to show that it is in #EXP. In the spirit of similar definitions for #P (see, for example, [Val79b, PB83]) and NP (*e.g.*, [GJ79]) we extend the definition of #EXP to apply also to non-integer valued functions.

Definition 5.5: An arbitrary function f is said to be *#EXP-easy* if there exists an integer function f' in #EXP and a polynomial-time Turing machine $\mathbf{M}$ such that for all x, $f(x) = \mathbf{M}(f'(x))$.

A function f is *#EXP-hard* if, for #EXP-easy function g, there are polynomial-time functions h_1 and h_2 such that, for all x, $g(x) = h_2(f(h_1(x)))$.[6]

A function f is *#EXP-complete* if it is #EXP-easy and #EXP-hard.[7] ∎

We can similarly define analogues of these definitions for the class #P.

We now show that for an unbounded vocabulary, the problem of computing the asymptotic limit is #EXP-complete.[8]

Theorem 5.6: *Given Φ and $\varphi, \theta \in \mathcal{L}(\Phi)$, the function $\Pr_\infty^w(\varphi|\theta)$ (resp., $\Pr_\infty^{s,\Phi}(\varphi|\theta)$) is #EXP-complete. The #EXP lower bound holds even if $\Phi \subset \mathcal{Q}$ and for sentences in $\mathcal{L}_2^-(\Phi)$.*

Proof: We sketch the proof for the random-worlds method; the proof for the random-structures method is quite similar. For the upper bound, given φ and θ, observe that it is not difficult to construct a nondeterministic exponential time Turing machine which has one accepting path for each model description consistent with θ (resp., $\varphi \wedge \theta$). Unfortunately, in order to compute the asymptotic conditional probability, we do not want to count *all* model descriptions. Rather, as the formula in Theorem 4.9 shows, we want to count the model descriptions consistent with θ (resp., $\varphi \wedge \theta$) that have maximum degree, without knowing in advance what that degree is. In order to deal with this difficulty, we show how to construct a nondeterministic time Turing machine $\mathbf{M}$ such that the number of accepting paths of $\mathbf{M}$ encodes, for each degree δ, the sum of $1/H(\psi)$ for those model descriptions ψ of degree δ which are consistent with each of θ and $\varphi \wedge \theta$. From this number we can easily calculate the asymptotic conditional probability, using the formula in Theorem 4.9. We leave details to the full paper.

For the lower bound, we use a result due to Lewis [Lew80]; he shows that the satisfiability problem for $\mathcal{L}^-(\mathcal{Q})$ is complete for NEXPTIME. Lewis' proof involves a straightforward reduction of NEXPTIME to $\mathcal{L}^-(\mathcal{Q})$: given a nondeterministic Turing machine $\mathbf{M}$ that runs in exponential time, and an input w, Lewis constructs a sentence $\xi \in \mathcal{L}^-(\mathcal{Q})$, whose length is polynomial in the size of $\mathbf{M}$ and w, such that ξ is satisfiable iff there is an accepting computation of $\mathbf{M}$ on w. Moreover, ξ has quantifier depth 2. By slightly modifying Lewis' proof, we can show that, given w and $\mathbf{M}$, we

[6] Notice that we need the function h_2 as well as h_1. For example, if g is an integer-valued function and f always returns a rational value between 0 and 1, as is the case for us, then there is no function h_1 such that $g(x) = f(h_1(x))$.

[7] Using the term #EXP-complete for a non-integer valued function is a slight abuse of notation. We use it for the sake of simplicity.

[8] We allow the function computing the asymptotic limit to have a special value, say 2, if the asymptotic limit is not well-defined.

can construct a formula ξ and a finite set $\mathcal{P}$ such that $\xi \in \mathcal{L}^-(\mathcal{P})$ and the number of model descriptions over $\mathcal{P}$ consistent with ξ is exactly the number of accepting computations of $\mathbf{M}$ on w. Lewis' result allows us to prove that given a formula $\varphi \in \mathcal{L}^-(\mathcal{Q})$ and a finite set $\mathcal{P} \subset \mathcal{Q}$ such that $\varphi \in \mathcal{L}^-(\mathcal{P})$, counting the number of model descriptions over $\mathcal{P}$ consistent with φ is #EXP-complete. We then show how to reduce the problem of counting the number of model descriptions consistent with φ to the problem of computing an appropriate asymptotic conditional probability.

The major difficulty we need to overcome in showing this is the converse of the difficulty that arose in the upper bound: we now want to count *all* consistent model descriptions consistent with φ, while using the asymptotic conditional probability in the most obvious way would only let us count the number of model descriptions consistent with φ whose degree is maximal. In the full paper, we show how to construct a formula θ and how to effectively transform φ to a formula φ^Q (using an extra predicate symbol Q not in φ) such that the number of model descriptions of maximal degree consistent with $\varphi^Q \wedge \theta$ is precisely the number of model descriptions consistent with φ. Using this transformation, the rest of the proof is straightforward. ∎

As in Theorem 5.2, we can also show that checking nontrivial estimates of the asymptotic probability is hard, even if we restrict to formulas of nonstandard depth 2.

Theorem 5.7: *Fix rationals r_1 and r_2 such that $0 \leq r_1 \leq r_2 \leq 1$ and $[r_1, r_2] \neq [0,1]$. Given $\mathcal{P} \subset \mathcal{Q}$ and $\varphi, \theta \in \mathcal{L}^-(\mathcal{P})$, the problem of deciding whether $\Pr^w_\infty(\varphi|\theta) \in [r_1, r_2]$ (resp., $\Pr^{s,\mathcal{P}}_\infty(\varphi|\theta) \in [r_1, r_2]$) is both NEXPTIME- and co-NEXPTIME-hard, even given an oracle for deciding whether the limit exists.*

The above discussion analyzes the complexity of asymptotic conditional probabilities for formulas whose nonstandard quantification depth is at least 2. It remains only to consider the case of formulas whose nonstandard depth is at most 1. As we show in the full paper, analogous results hold, but the complexities drop by an exponential factor in this case.

References

[BGHK92] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller. From statistics to belief. In *Proc. National Conference on Artificial Intelligence (AAAI '92)*, pages 602–608, 1992.

[Che83] P. C. Cheeseman. A method of computing generalized Bayesian probability values for expert systems. In *Proc. Eighth International Joint Conference on Artificial Intelligence (IJCAI '83)*, pages 198–202, 1983.

[Com88] K. Compton. 0-1 laws in logic and combinatorics. In I. Rival, editor, *Proc. 1987 NATO Adv. Study Inst. on algorithms and order*, pages 353–383. Reidel, Dordrecht, Netherlands, 1988.

[DD85] K. G. Denbigh and J. S. Denbigh. *Entropy in Relation to Incomplete Knowledge.* Cambridge University Press, Cambridge, UK, 1985.

[DG79] B. Dreben and W. D. Goldfarb. *The Decision Problem: Solvable Classes of Quantificational Formulas.* Addison-Wesley, Reading, MA, 1979.

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

[Fag77] R. Fagin. The number of finite relational structures. *Discrete Mathematics*, 19:17–21, 1977.

[GHK92] A. J. Grove, J. Y. Halpern, and D. Koller. Random worlds and maximum entropy. In *Proc. 7th IEEE Symp. on Logic in Computer Science*, pages 22–33, 1992.

[GJ79] M. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness.* W. Freeman and Co., San Francisco, CA, 1979.

[GKLT69] Y. V. Glebskiĭ, D. I. Kogan, M. I. Liogon'kiĭ, and V. A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Kibernetika*, 2:17–28, 1969.

[Jay78] E. T. Jaynes. Where do we stand on maximum entropy? In R. D. Levine and M. Tribus, editors, *The Maximum Entropy Formalism*, pages 15–118. MIT Press, Cambridge, MA, 1978.

[Key21] J. M. Keynes. *A Treatise on Probability.* Macmillan, London, 1921.

[Kri86] J. von Kries. *Die Principien der Wahrscheinlichkeitsrechnung und Rational Expectation.* Freiburg, 1886.

[Lap20] P. S. de Laplace. *Essai Philosophique sur les Probabilités.* 1820. English translation is *Philosophical Essay on Probabilities*, Dover Publications, New York, 1951.

[Lew79] H. R. Lewis. *Unsolvable Classes of Quantificational Formulas.* Addison-Wesley, New York, 1979.

[Lew80] H. R. Lewis. Complexity results for classes of quantificational formulas. *Journal of Computer and System Sciences*, 21:317–353, 1980.

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

[MP92] Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems*, volume 1. Springer-Verlag, Berlin/New York, 1992.

[PB83] S. J. Provan and M. O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM Journal on Computing*, 12:777–788, 1983.

[PS72] R. E. Powell and S. M. Shah. *Summability Theory and Applications.* Van Nostrand Reinhold, 1972.

[PV89] J. B. Paris and A. Vencovska. On the applicability of maximum entropy to inexact reasoning. *International Journal of Approximate Reasoning*, 3:1–34, 1989.

[Rog67] H. Rogers, Jr. *Theory of Recursive Functions and Effective Computability.* McGraw-Hill, New York, 1967.

[Val79a] L. G. Valiant. The complexity of computing the permanent. *Theoretical Computer Science*, 8:189–201, 1979.

[Val79b] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM Journal on Computing*, 8:410–421, 1979.