

*A note on the least informative model of a theory*

Paris, J. B. and Rad, S. R.

2009

MIMS EPrint: **2009.54**

Manchester Institute for Mathematical Sciences  
School of Mathematics

The University of Manchester

Reports available from: <http://eprints.maths.manchester.ac.uk/>

And by contacting: The MIMS Secretary  
School of Mathematics  
The University of Manchester  
Manchester, M13 9PL, UK

ISSN 1749-9097

# A note on the least informative model of a theory

J. B. Paris and S. R. Rad\*  
School of Mathematics  
The University of Manchester  
Manchester M13 9PL

jeff.paris@manchester.ac.uk    soroush.rafaee-rad@postgrad.manchester.ac.uk

September 2, 2009

## Abstract

We consider one possible interpretation of the ‘least informative model’ of a relational and finite theory and show that it is well defined for a particular class of  $\Pi_1$  theories. We conjecture that it is always defined for  $\Pi_1$  theories.

Key words: Uncertain reasoning, probability logic, inference processes, Polyadic Inductive Logic.

## Introduction

Let  $\Psi$  be a consistent sentence of a first order language  $L$  and let  $\theta(x_1, x_2, \dots, x_n)$  be a formula of  $L$ . Then given a structure  $M$  for  $L$  with universe  $\{a_i \mid i \in \mathbb{N}^+\}$  about which we know only that  $M$  is a model of  $\Psi$  a natural question one might ask is how likely, or probable, is it that  $M \models \theta(a_1, a_2, \dots, a_n)$ ?

In this short note we consider one (limited) approach to answering this question which was originally (see [6]) based on adapting methods of uncertain reasoning developed for propositional probabilistic knowledge bases, in particular the Maximum Entropy Inference Process, though for the purposes of this paper we will not need to recall that history.

## Notation

From now on we assume of  $L$  that it is a finite relational first order language without functions or constants. We shall also assume for the present that  $L$  does not contain equality. Let  $\mathcal{T}$  be the set of structures for  $L$  with universe

---

\*Supported by a MATHLOGAPS Research Studentship, MEST-CT-2004-504029.

$\{a_i \mid i \in \mathbb{N}^+\}$  and let  $L_a$  be  $L$  augmented with constants  $a_i$  for  $i \in \mathbb{N}^+$ , which of course are interpreted in the structures  $M \in \mathcal{T}$  by the elements  $a_i$  of  $M$ . Let  $\mathcal{T}^{(n)}$  be the (finite) set of structures for  $L$  with universe  $\{a_i \mid 1 \leq i \leq n\}$  and let  $L_a^{(n)}$  be  $L$  augmented with the constants  $a_i$  for  $1 \leq i \leq n$ .

As above let  $\Psi$  be a consistent sentence in  $L$ . Then provided

$$\{M \in \mathcal{T}^{(n)} \mid M \models \Psi\} \neq \emptyset$$

we can define a probability function<sup>1</sup>  $w_n$  on the quantifier free sentences of  $SL_a^{(n)}$  by

$$w_n(\theta) = \frac{|\{M \in \mathcal{T}^{(n)} \mid M \models \theta\}|}{|\mathcal{T}^{(n)}|}.$$

Provided the limit is well defined we now set  $w$  to be the probability function on the quantifier free sentences of  $SL_a$  which is the limit of the  $w_n$ , so in this case  $w$  satisfies

$$\begin{aligned} \text{(P1)} \quad & \models \theta \Rightarrow w(\theta) = 1, \\ \text{(P2)} \quad & \models \neg(\theta \wedge \phi) \Rightarrow w(\theta \vee \phi) = w(\theta) + w(\phi), \end{aligned}$$

on the quantifier free  $\theta, \phi$  of  $SL_a$ . By a theorem of Gaifman, see [11],  $w$  has a unique extension to a probability function on  $SL_a$ , meaning that it satisfies (P1),(P2) for any sentences  $\theta, \phi$  and for  $\exists x \psi(x) \in SL_a$ ,

$$\text{(P3)} \quad w(\exists x \psi(x)) = \lim_{n \rightarrow \infty} w(\psi(a_1) \vee \psi(a_2) \vee \dots \vee \psi(a_n)).$$

Then provided  $w$  is well defined and  $w(\Psi) = 1$  we putatively propose  $w(\theta)$ , for  $\theta \in SL_a$ , as the probability that  $M \models \theta$  for  $M \in \mathcal{T}$  an otherwise unknown model of  $\Psi$ .

The first thing to say about this proposal is that the  $w_n$  may not be well defined simply because  $\Psi$  has no finite models. So suppose from now on that  $\Psi$  does have a finite model, say of cardinality  $n$ . Then by just making ‘clones’ of some element in that model we can construct models of any finite cardinality greater than  $n$  so all the subsequent  $w_n$  will be defined. Still however the limit need not be well defined.

As an obvious example here let  $L$  have relations  $G$ ,  $R$  and  $P$  of arities 3,2,1 respectively, let  $x =_G y$  abbreviate

$$\forall u, t (G(x, u, t) \leftrightarrow G(y, u, t))$$

and let  $\Psi_1$  be the conjunction of:

---

<sup>1</sup>Meaning that  $w_n$  is a map from the set  $SL_a^{(n)}$  of sentences of  $L_a^{(n)}$  to  $[0, 1]$  such that for  $\theta, \phi, \exists x \psi(x) \in SL_a^{(n)}$ ,  
(P1)  $\models \theta \Rightarrow w_n(\theta) = 1$ ,  
(P2)  $\models \neg(\theta \wedge \phi) \Rightarrow w_n(\theta \vee \phi) = w_n(\theta) + w_n(\phi)$ ,  
(P3)  $w_n(\exists x \psi(x)) = w_n(\psi(a_1) \vee \psi(a_2) \vee \dots \vee \psi(a_n))$ .

$$\begin{aligned}
& \forall x, y, z (x =_G y \rightarrow (R(x, z) \rightarrow R(y, z))) \\
& \quad \forall x, y (R(x, y) \leftrightarrow R(y, x)) \\
& \forall x, y, z ((R(x, y) \wedge R(x, z)) \rightarrow (x =_G y \vee x =_G z \vee y =_G z)) \\
& \quad \forall x \exists y (x \neq_G y \wedge R(x, y)) \\
& \quad \forall x \neg R(x, x)
\end{aligned}$$

and  $\Psi_2$  be the conjunction of:

$$\begin{aligned}
& \forall x, y, z (x =_G y \rightarrow (R(x, z) \rightarrow R(y, z))) \\
& \quad \forall x, y (R(x, y) \leftrightarrow R(y, x)) \\
& \quad \forall x, y, z ((R(x, y) \wedge R(x, z)) \rightarrow (y =_G z)) \\
& \forall x, y, z, t ((R(x, y) \wedge R(z, t) \wedge (x =_G y) \wedge (z =_G t)) \rightarrow (x =_G z)) \\
& \quad \forall x \exists y R(x, y) \\
& \quad \exists x R(x, x)
\end{aligned}$$

Then  $\models \neg(\Psi_1 \wedge \Psi_2)$  and for  $n$  even the proportion of  $M \in \mathcal{T}^{(n)}$  satisfying  $M \models \Psi_1 \vee \Psi_2$  which also satisfy  $\Psi_1$  tends to zero as even  $n \rightarrow \infty$  whereas for  $n$  odd it is the proportion which satisfy  $\Psi_2$  which tend to zero. Hence if we take  $\Psi = \Psi_1 \vee \Psi_2$  then  $w_n(\Psi_1)$  bobs backwards and forwards between being close to 1 and close to 0 as  $n \rightarrow \infty$  and no overall limit exists. [For fuller details see [19].]

A second problem with this approach is that whilst the  $w_n$  may have a limit  $w$  this probability function, once extended to all sentences of  $L_a$  via Gaifman's Theorem, may no longer give  $w(\Psi)$  probability 1. This happens for example in the case when  $L$  has a single binary relation  $R$  and  $\Psi$  is  $\exists x \forall y R(x, y)$ . The informal explanation of why this happens is that the overriding majority of structures in  $\mathcal{T}^{(n)}$  which model this  $\Psi$  will have just one  $a_i$  satisfying  $\forall y R(x, y)$ . Thus as  $n \rightarrow \infty$  the probability that any one  $a_i$  satisfies this tends to zero with the result that in the limit none of them have non-zero probability of satisfying it. Again details can be found in [19].

On a more optimistic note however the limit  $w$  does exist for (consistent)  $\Psi$  when  $L$  is purely unary, see for example [1], [2], [6], (and also the developments in [13], [14], [15]). So from now on we shall assume that  $L$  has at least one non-unary relation.

## The Conjecture

A feature of these two failures relevant to this paper is that in the first  $\Psi$  is of quantifier complexity  $\Pi_2$  whilst in the second it is  $\Sigma_2$ .<sup>2</sup> On the other hand it is not difficult to check that if  $\Psi$  is  $\Sigma_1$  then the limit  $w$  exists and  $w(\Psi) = 1$ , see [19]. This leaves open the question of what happens when  $\Psi$  is  $\Pi_1$ .

We would conjecture that in the case  $\Psi$  is  $\Pi_1$  and consistent then the limit  $w$  always exists and furthermore satisfies  $w(\Psi) = 1$ .

The main contribution of this paper towards confirming this conjecture is to show that it is true for such  $\Psi$  when

$$|\{M \in \mathcal{T}^{(n)} \mid M \models \Psi\}|$$

is polynomially bounded.

There is some evidence that confirming the conjecture in general when we also allow in equality may not be very easy. This comes from results in Graph Theory concerned with what is there called the ‘speed of hereditary classes’. A class of graphs  $\mathcal{F}$  is said to be *hereditary* if it is closed under isomorphisms and subgraphs, in other words whenever a graph  $G$  is in  $\mathcal{F}$  and  $H$  is the subgraph of  $G$  formed by restricting the edges to some subset of the vertices of  $G$  then  $H \in \mathcal{F}$ . Given a hereditary class  $\mathcal{F}$  of graphs let  $f_n$  be the number of graphs in  $\mathcal{F}$  with vertices  $\{a_i \mid 1 \leq i \leq n\}$ . A prominent question in Graph Theory over the last 15 years (see for example [3], [4], [5], [20]) is what  $f_n$  may look like as a function of  $n$ . From results obtained to date it would appear that  $f_n$  must fall into one of 4 bands. The lowest of these, which is well understood, is when  $f_n$  is bounded by a polynomial. After that however the bands are much wider and within them there *seems*, as far as is currently known, to be scope for  $f_n$  to behave uncommonly badly.

The reason these results, or lack of, are relevant to our conjecture here is that if  $L$  has a single binary relation  $R$  then the models in  $\mathcal{T}^{(n)}$  of a  $\Pi_1$  sentence  $\Psi$  which implies that  $R$  is symmetric are just the graphs with vertices  $\{a_i \mid 1 \leq i \leq n\}$  in some hereditary class  $\mathcal{F}$ . From this then it would seem that understanding the behavior of the  $w_n$  may not come so easily.

The reader familiar with the 0-1 laws of Fagin [10] and Glebskiĭ et al [12] and subsequent developments (see for example [7], [8], [9], [16], [23]) concerning the asymptotic frequency of models of  $\Psi$  in  $\mathcal{T}^{(n)}$  may wonder if everything that we are conjecturing has not already been answered in the course of that body research. As far as we have been able to discover it seems that it has not. The problem here is that unless  $\Psi$  is a tautology (assuming as we are that it does not mention any constants) then in the limit the frequency of its models in  $\mathcal{T}^{(n)}$  will drop to zero. Thus the obvious idea that the limit of the  $w_n(\theta)$  will be the ration of the limiting frequencies of  $\Psi \wedge \theta$  and  $\Psi$  gets us nowhere because this amounts to just 0/0 !

---

<sup>2</sup>Note however that in the case that  $\Psi$  is the (necessarily)  $\Pi_2$  sentence asserting that  $R(x, y)$  defines a linear order without end points the limit does exist and give  $\Psi$  probability 1.

## Slow sentences

In this section we shall confirm the above conjecture in the case that  $\Psi$  is *slow*, meaning that

$$|\{M \in \mathcal{T}^{(n)} \mid M \models \Psi\}|$$

is polynomially bounded. First however we introduce some notation and give a syntactic characterization of ‘slowness’.

Suppose  $L$  has relations  $R_1, \dots, R_q$  of arities  $h_1, \dots, h_q$ . Let  $b_1, b_2, \dots, b_m$  stand for some distinct choices from the constants  $a_i$  (a convention we adopt throughout). Then a *state description* for  $b_1, \dots, b_m$  is a sentence of the form

$$\Theta(b_1, b_2, \dots, b_m) = \bigwedge_{s=1}^q \bigwedge_{i_1, i_2, \dots, i_{h_s} \in \{1, \dots, m\}} \pm R_s(b_{i_1}, b_{i_2}, \dots, b_{i_{h_s}}), \quad (1)$$

where  $\pm R$  stands for  $R$  or  $\neg R$  respectively. In other words a state description for  $b_1, \dots, b_m$  determines a structure for  $L$  whose universe is the set  $\{b_1, \dots, b_m\}$ .

Given such a state description we say that  $b_i$  and  $b_j$  are *indistinguishable* with respect to  $\Theta(b_1, \dots, b_m)$  if

$$\Theta(b_1, \dots, b_m) \wedge b_i = b_j$$

is consistent (with the axioms of equality) and in this case write  $b_i \sim_{\Theta} b_j$ . Clearly  $\sim_{\Theta}$  is an equivalence relation. Let  $\|\Theta(b_1, \dots, b_m)\|$  be the number of equivalence classes with respect to  $\sim_{\Theta}$ .

For  $\Theta(b_1, \dots, b_m)$  a state description and  $i_1, \dots, i_k \in \{1, 2, \dots, m\}$ , set  $\Theta[b_{i_1}, \dots, b_{i_k}]$  to be the (unique) state description for  $b_{i_1}, \dots, b_{i_k}$  consistent with  $\Theta(b_1, \dots, b_m)$ .

The following rather technical lemma will be useful in what follows.

**Lemma 1** *Let  $r$  be at least the largest arity of any relation in  $L$ , let  $p \geq k \geq r$  and let  $\Phi(a_1, \dots, a_m)$  be a state description (of  $L$ ) with  $\|\Phi(a_1, \dots, a_m)\| = p$ . Then there is some  $k \leq s \leq k+r$  and  $1 \leq i_1 < i_2 < \dots < i_s \leq m$  such that  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_s}]\| = s$ .*

**Proof** If  $p \leq k+r$  taking  $a_{i_1}, a_{i_2}, \dots, a_{i_p}$  to be representatives from the equivalence classes of  $\sim_{\Phi}$  will give  $\|\Phi(a_{i_1}, a_{i_2}, \dots, a_{i_p})\| = p$  so assume  $k+r < p$ . Suppose we have picked  $a_{i_1}, a_{i_2}, \dots, a_{i_t}$  with  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}]\| = t$ . It is enough to show that we can find some  $j_1, j_2, \dots, j_s \leq m$  distinct from the  $a_{i_1}, \dots, a_{i_t}$  such that  $1 \leq s \leq r$  and  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}, a_{j_1}, \dots, a_{j_s}]\| = t+s$ .

To this end let  $a_{j_1}$  be inequivalent to each of the  $a_{i_1}, a_{i_2}, \dots, a_{i_t}$  modulo  $\sim_{\Phi}$ . If  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}, a_{j_1}]\| = t+1$  we are done. Otherwise, according to this state description  $\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}, a_{j_1}]$   $a_{j_1}$  and  $a_{i_q}$  are indistinguishable for some  $1 \leq q \leq t$ . Then since they are distinguishable modulo  $\sim_{\Phi}$  and  $r$  is the largest arity of any relation in  $L$  we can find  $j_2, \dots, j_q$  with  $q \leq r$  such that

$a_{j_1}$  and  $a_{i_g}$  are distinguishable according to  $\Phi[a_{i_g}, a_{j_1}, a_{j_2}, a_{j_3}, \dots, a_{j_q}]$ . Indeed we may further assume that none of these  $a_{j_2}, a_{j_3}, \dots, a_{j_q}$  are indistinguishable in  $\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}, a_{j_1}, a_{j_2}, \dots, a_{j_q}]$  from any of these other constants mentioned there, otherwise we could simply remove them. It follows then that  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_t}, a_{j_1}, a_{j_2}, \dots, a_{j_q}]\| = t + q$ , as required. ■

In the case  $r = 2$  it can be shown, see [21], that we can take  $s = k$  in this Lemma. However we cannot hope to have this result for  $s = k$  when  $r > 2$ . For example let  $L$  have just the ternary relation  $R$  and let  $\Phi(a_1, a_2, a_3, a_4, a_5, a_6)$  imply  $R(a_i, a_j, a_k)$  just if  $\{i, j, k\}$  is one of  $\{1, 2, 3\}$  or  $\{4, 5, 6\}$  (so it implies  $\neg R(a_i, a_j, a_k)$  otherwise). In this case

$$\|\Phi(a_1, a_2, a_3, a_4, a_5, a_6)\| = 6$$

but we cannot find  $a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4}, a_{i_5}$  such that  $\|\Phi[a_{i_1}, a_{i_2}, a_{i_3}, a_{i_4}, a_{i_5}]\| = 5$ .

The next theorem gives a characterization of the slow  $\Pi_1$  sentences.

**Theorem 2** *Let  $r$  be at least the largest arity of any relation in  $L$  and  $\Psi$  a consistent  $\Pi_1$  sentence of  $L$ . Then*

$$|\{M \in \mathcal{T}^{(n)} \mid M \models \Psi\}| = o(n^k)$$

*iff for some state descriptions  $\Phi_i(a_1, a_2, \dots, a_{k+r})$ ,  $i = 1, \dots, h$ , with  $\|\Phi_i(a_1, a_2, \dots, a_{k+r})\| < k$ ,*

$$\Psi \equiv \forall x_1, \dots, x_{k+r} \bigvee_{j=1}^h \Phi_j(x_1, x_2, \dots, x_{k+r}). \quad (2)$$

**Proof** Let the  $\Phi_i(a_1, \dots, a_{k+r})$  for  $i = 1, \dots, h$  list all state descriptions with  $\|\Phi_i(a_1, a_2, \dots, a_{k+r})\| < k$  which are consistent with  $\Psi$ . If

$$\Psi \wedge \neg \forall x_1, \dots, x_{k+r} \bigvee_{j=1}^h \Phi_j(x_1, x_2, \dots, x_{k+r}) \quad (3)$$

was consistent it would have a model  $M$  in  $\mathcal{T}$  and hence there would be a state description  $\Theta(a_{i_1}, a_{i_2}, \dots, a_{i_{k+r}})$  true in  $M$  with  $\|\Theta(a_{i_1}, a_{i_2}, \dots, a_{i_{k+r}})\| \geq k$ . Clearly by permuting these constants we may assume  $a_{i_j}$  is just  $a_j$  for  $j = 1, \dots, k+r$ . In this case take  $n$  large and let  $\Xi(a_1, a_2, \dots, a_{k+r}, a_{k+r+1}, \dots, a_n)$  be a state description which implies (equivalently extends)  $\Theta(a_1, a_2, \dots, a_{k+r})$  such that each of the  $a_{k+r+1}, \dots, a_n$  is equivalent according to  $\sim_\Xi$  to some  $a_i$  with  $i \leq k+r$ . In other words no new equivalence classes are created in going from  $\sim_\Theta$  to  $\sim_\Xi$ , they just enlarge. Then just as the structure  $M_\Theta \in \mathcal{T}^{(k+r)}$  determined by  $\Theta$  is a model of  $\Psi$  so also is  $M_\Xi \in \mathcal{T}^{(n)}$ . However since  $\sim_\Theta$  has at least  $k$  equivalence classes such a  $\Xi$  can be formed in at least  $k^{n-k-r}$  ways, contradicting the given bound.

From (3) then

$$\vdash \Psi \rightarrow \forall x_1, \dots, x_{k+r} \bigvee_{j=1}^h \Phi_j(x_1, x_2, \dots, x_{k+r}). \quad (4)$$

For the provability of the other direction here suppose on the contrary that

$$\neg \Psi \wedge \forall x_1, \dots, x_{k+r} \bigvee_{j=1}^h \Phi_j(x_1, x_2, \dots, x_{k+r})$$

was consistent, so had a model  $M \in \mathcal{T}$ . Then since  $\Psi$  is  $\Pi_1$  there is some large  $m$  such that the state description  $\Phi(a_1, \dots, a_m)$  determined by  $M$  is inconsistent with  $\Psi$ . If  $\|\Phi(a_1, \dots, a_m)\| \geq k$  then by Lemma 1 we could find some  $i_1, i_2, \dots, i_s \leq m$  with  $s \leq k+r$  such that  $\|\Phi[a_{i_1}, \dots, a_{i_s}]\| \geq k$ . Since we can permute the elements of  $M$  we may suppose that  $i_j = j$  for  $j = 1, \dots, s$ . But in that case since  $M$  is a model of the second conjunct of (4),  $\Phi[a_1, \dots, a_s, a_{s+1}, \dots, a_{k+r}]$  would have to be one of the  $\Phi_j(a_1, \dots, a_{k+r})$  which is a contradiction because it has too many equivalence classes.

Hence  $\|\Phi(a_1, \dots, a_m)\| < k$  and without loss of generality we may assume that  $a_1, a_2, \dots, a_{k+r}$  contains representatives of all the equivalence classes of  $\sim_\Phi$ . In that case  $\Phi[a_1, a_2, \dots, a_{k+r}]$  must again, as above, be one of the  $\Phi_j(a_1, \dots, a_{k+r})$ . But then  $\Phi[a_1, a_2, \dots, a_{k+r}]$  must be consistent with  $\Psi$ , indeed it determines a model of  $\Psi$  in  $\mathcal{T}^{(k+r)}$ , so  $\Phi(a_1, a_2, \dots, a_m)$  will also be consistent with  $\Psi$  since it is formed by simply duplicating  $a_i$  in  $\Phi[a_1, a_2, \dots, a_{k+r}]$ .

Turning now to the other direction of the equivalence stated in the Theorem assume that (2) holds and let the state description  $\Phi(a_1, a_2, \dots, a_m)$  determine a model of  $\Psi$  in  $\mathcal{T}^{(n)}$ . Then as above if  $\|\Phi(a_1, a_2, \dots, a_m)\| \geq k$  we could cut this down to a  $\Phi[a_{i_1}, \dots, a_{i_{k+r}}]$  satisfying  $\|\Phi[a_{i_1}, \dots, a_{i_{k+r}}]\| \geq k$  and this would still be consistent with  $\Psi$ . But clearly this is not consistent with the right hand side of (2), contradiction.

We conclude that  $\|\Phi(a_1, a_2, \dots, a_m)\| < k$  and in turn that if  $a_{i_1}, \dots, a_{i_{k+r}}$  contain representative from all the equivalence classes of  $\sim_\Phi$  then  $\Phi[a_{i_1}, \dots, a_{i_{k+r}}]$  is one of the  $\Phi_j(a_{i_1}, \dots, a_{i_{k+r}})$ . Hence  $\Phi(a_1, a_2, \dots, a_m)$  is determined by this  $j \in \{1, 2, \dots, h\}$ , the choice of  $i_1, \dots, i_{k+r}$  and the choice of which of the (at most)  $k-1$  equivalence classes contain the remaining  $a_i$  for  $1 \leq i \leq m$ ,  $i \neq i_1, \dots, i_{k+r}$ , which overall amounts to just  $o(n^k)$  choices. ■

We are now in a position to prove the main result of this paper.

**Theorem 3** *Let  $\Psi$  be a consistent  $\Pi_1$  slow sentence of  $L$ . Then the limit  $w$  of the  $w_n$  exists and satisfies  $w(\Psi) = 1$ .*

**Proof** It is clear that if the limit  $w$  exists then it satisfies  $w(\Psi) = 1$  so we only need to show that the limit exists. Let  $m$  be large and let  $\Theta(a_1, \dots, a_m)$  be a



state description consistent with  $\Psi$ , so this sentence defines a model in  $\mathcal{T}^{(m)}$ . We want to count the number of state descriptions  $\Phi(a_1, a_2, \dots, a_n)$  consistent with  $\Psi$  and extending  $\Theta(a_1, \dots, a_m)$ . Since  $\Psi$  is slow let the  $\Phi_j(a_1, \dots, a_{k+r})$  etc. be as in Theorem 2. Let  $a_{i_1}, a_{i_2}, \dots, a_{i_s}$  be the first (viz-a-viz the indices) elements of the distinct equivalence classes of  $\sim_\Phi$ , so  $\|\Phi[a_{i_1}, a_{i_2}, \dots, a_{i_s}]\| = s$ , which is less than  $k$  as in the arguments above. Let  $r$  be minimal such that  $i_r > m$ . Then  $\Phi(a_1, a_2, \dots, a_n)$  is determined by  $a_{i_r}, a_{i_{r+1}}, \dots, a_{i_s}$ ,  $\Phi(a_1, a_2, \dots, a_m, a_{i_r}, a_{i_{r+1}}, \dots, a_{i_s})$  (which determines  $a_{i_1}, \dots, a_{i_{r-1}}$ ) and the assignment of the remaining  $a_i$  for  $1 \leq i \leq n$  not amongst the  $a_1, a_2, \dots, a_m, a_{i_r}, a_{i_{r+1}}, \dots, a_{i_s}$  to the equivalence classes determined by  $a_{i_1}, a_{i_2}, \dots, a_{i_s}$ .

In order to count these  $\Phi(a_1, a_2, \dots, a_n)$  notice that if we choose  $a_{i_r}, a_{i_{r+1}}, \dots, a_{i_s}$  ( $s < k$ ) first then there are only a fixed (independent of  $n$ ) finite number of possible  $\Phi(a_1, a_2, \dots, a_m, a_{i_r}, a_{i_{r+1}}, \dots, a_{i_s})$  (this determines  $s$ ) and for each of these there are  $s^{n-m-s+r+1}$  ways to assign the remaining  $a_i$  to classes. Altogether then this gives us  $dc^n(1+o)$  possible  $\Phi(a_1, a_2, \dots, a_n)$  for some constant  $d$  and  $c < k$ .

The Theorem now follows by noticing that  $|\mathcal{T}^{(n)}|$  is just the sum of these  $dc^n(1+o)$  over the finite number of possible  $\Theta(a_1, \dots, a_m)$ . ■

## Discussion

Jon Williamson in [22] has suggested an alternative approach to the problem of what probability  $v(\theta)$  to give to a sentence  $\theta$  of  $L_a$  being true in  $M \in \mathcal{T}$  when all we know about  $M$  is that it is a model of  $\Psi$ . The idea is that given a probability function on  $SL_a$  we let  $v_n$  be the restriction of  $v$  to  $SL_a^{(n)}$  and define a partial ordering on such  $v$  by

$$v \prec v' \quad \text{iff for all } n \text{ eventually } E(v_n) < E(v'_n)$$

where  $E(v_n)$ , the negative entropy of  $v_n$ , is the sum over the state descriptions  $\Theta$  of  $SL_a^{(n)}$  of  $v_n(\Theta) \log(v_n(\Theta))$ . Since negative entropy is generally accepted as a measure of 'lack of information', one might argue that in our current context of knowing only  $\Psi$  one should assign  $\theta$  probability  $v(\theta)$  for that probability function  $v$  which satisfies  $v(\Psi) = 1$  and is  $\prec$ -minimal amongst all such functions.

As with the approach discussed in this paper this minimal choice exists (and gives the same answers) in the cases of unary languages and in the cases of  $\Psi$  being  $\Sigma_1$  but can fail to be defined for  $\Sigma_2$  and  $\Pi_2$   $\Psi$  (see [19]). We would conjecture that the method also succeeds for slow  $\Pi_1$  sentences, and indeed more generally that both approaches are defined and give the same answers for all  $\Pi_1$   $\Psi$ . However that is not always the case for higher quantifier complexity, Williamson's method can work even when  $\Psi$  has no finite models, for example when  $\Psi$  defines a dense linear ordering.

The results given in this paper suffer the obvious weakness that the language and structures do not include equality. Of course we could include in the  $\Pi_1$

sentence  $\Psi$  the (finite)  $\Pi_1$  axiomatization of the equality axioms appropriate to  $L$  and carry on as before. However in that case our structures would be what Mendelson refers to in [17] as ‘non-normal’ structures rather than structures in which  $=$  is interpreted as real equality, so this is not much help. Given the results from Graph Theory (which do allow equality) we would conjecture that Theorem ?? also holds when we properly allow equality into the language.

We finally mention that these method of this paper and Williamson’s were originally introduced to address the more general problem of inference from predicate probabilistic knowledge bases as in [6] and [18]. However it would seem that the special case problem considered here is really the obstacle to be overcome in this endeavor.

## References

- [1] Bacchus, F., Grove, A. J., Halpern, J. Y. and D. Koller, D., Generating new beliefs from old, *Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence*, 1994, pp37-45.
- [2] Bacchus, F., Grove, A. J., Halpern, J. Y. and Koller, D., From statistical knowledge to degrees of belief, *Artificial Intelligence*, **87**, 1996, pp75-143.
- [3] Balogh, J., Bollobas, B. and Weinreich, The speed of hereditary properties of graphs, *Journal of Combinatorial Theory*, series B **79**, 2000, pp131-156.
- [4] Balogh, J., Bollobas, B. and Weinreich, D., The penultimate range of growth for graph properties, *European Journal of Combinatorics*, **22**(3), 2001, pp277-289.
- [5] Ballobas, B. and Thomason, A. G., Projections of bodies and hereditary properties of hyper graphs, *Bulletin of the London Mathematical Society*, **27**(5), 1995, pp417-424.
- [6] Barnett, O. W. and Paris, J. B., Maximum Entropy inference with qualified knowledge, *Logic Journal of the IGPL*, **16**(1), 2008, pp85-98.
- [7] Compton, K. J., A Logical Approach to Asymptotic Combinatorics I. First Order Properties, *Advances in Mathematics*, **65**, 1987, pp65-96.
- [8] Compton, K. J., 0-1 Laws in Logic and Combinatorics, in *Proc. 1987 NATO Adv. Study Inst. on Algorithms and Order*, Ed. I.Rival, Reidel, Dordrecht, 1988.
- [9] Compton, K. J., A Logical Approach to Asymptotic Combinatorics II: Monadic Second Order Properties, *Journal of Combinatorial Theory*, Series A **50**, 1989, pp110-131.
- [10] Fagin, R., Probabilities on Finite Models, *Journal of Symbolic Logic*, bf 41, 1976, pp50-58.

- [11] H. Gaifman, Concerning measures in first order calculi, *Israel J. of Mathematics* **24**, 1964, pp1-18.
- [12] Glebskiĭ, Y. V., Kogan, D. I., Liogon'kiĭ, M. I. and Talanov, V. A., Range and Degree of Formulas in the Restricted Predicate Calculus, *Cybernetics*, **5**, 1972, pp142-154.
- [13] A.J. Grove, J.Y. Halpern, D. Koller, Random Worlds and Maximum Entropy, *Journal of Artificial Intelligence Research*, **2**,1994, pp33-88.
- [14] A.J. Grove, J.Y. Halpern, D. Koller, Asymptotic conditional probabilities: the unary case. *SIAM Journal of Computing*, **25**(1), 1996, pp1-51.
- [15] A.J. Grove, J.Y. Halpern, D. Koller, Asymptotic conditional probabilities: the non-unary case. *Journal of Symbolic Logic*, **61**(1), 1996, pp250-276.
- [16] Lynch, F. J., An Extension of 0-1 Laws, Random Structures and Algorithms, **5**(1), 1994, pp155-172.
- [17] Mendelson, E., *Introduction to Mathematical Logic*, 4th Edition, International Thompson Publishing, 1997.
- [18] Paris, J. B. and Rad, S. R., Inference Processes for Quantified Predicate Knowledge, in *Logic, Language, Information and Computation*, WoLLIC, Edinburgh, 2008, Eds. W.Hodges & R. de Queiroz, Springer LNAI 5110, pp249-259.
- [19] Rad, S. R., *Inference Processes For Probabilistic First Order Languages* , Ph.D. Thesis, University of Manchester, 2009. Available at <http://www.maths.manchester.ac.uk/~jeff/>
- [20] Scheinerman, E. R. and Zito, J., On the size of hereditary classes of graphs, *Journal Combinatorial Theory Ser. B* **61**, 1994, pp16-39.
- [21] Vencovská, A., Binary Induction and Carnap's Continuum, *Proceedings of the 7th Workshop on Uncertainty Processing (WUPES)*, Mikulov, 2006. See <http://mtr.utia.cas.cz/wupes06/articles/data/vencovska.pdf>
- [22] Williamson, J., Objective Bayesian Probabilistic Logic, *Journal of Algorithms in Cognition, Informatics and Logic*, **63**, 2008, pp167-183.
- [23] Woods, A. R., Counting Finite Models, *Journal of Symbolic Logic*, 1995, **62**(1), 925-949.