Inference Processes for Quantified Predicate Knowledge

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

School of Mathematics
University of Manchester
Manchester M13 9PL, UK
jeff.paris@manchester.ac.uk,
soroush.rafiee-rad@postgrad.manchester.ac.uk

Abstract. We describe a method for extending an inference process for propositional probability logic to predicate probability logic in the case where the language in purely unary and show that the method is well defined for the Minimum Distance and CM^∞ inference processes.

Keywords: Uncertain reasoning, probability logic, inference processes, Renyi Entropies.

1 Motivation

In this paper, which extends [3], we consider the following problem

Given that we know only that a probability function w on a predicate language L satisfies the finite set K of constraints

$$\sum_{i=1}^{q} c_{ij} w(\theta_i) \leq b_j, \quad j = 1, \ldots, m$$

where the θ_i are sentences of L and the $c_{ij}, b_j \in \mathbb{R}$ what value should be given to $w(\theta)$, for a sentence θ of L?

in the limited case where L has just finitely many unary predicate symbols $P_1, \ldots, P_n$ and countably many constant symbols $a_1, a_2, \ldots$ (which are intended to exhaust the universe) but no function symbols nor equality.

The relevance of this question for AI is that we imagine an agent whose knowledge consists of just K wishing to nevertheless assign probabilities to other sentences of the language. Indeed if, as seems quite reasonable, we require these assigned values to be consistent as a whole with K and w being a probability function then the question amounts to asking how one should best pick a probability function w on L, that is a function w on the set SL sentences of the language L satisfying that for all $\theta, \phi, \exists x \psi(x) \in SL$

* Supported by a MATHLOGAP Research Studentship - MEST-CT-2004-504029.

W. Hodges and R. de Queiroz (Eds.): WoLLIC 2008, LNAI 5110, pp. 249–259, 2008.

(P1) If $\models \theta$ then $w(\theta) = 1$

(P2) If $\models \neg(\theta \wedge \phi)$ then $w(\theta \vee \phi) = w(\theta) + w(\phi)$

(P3) $w(\exists x \psi(x)) = \lim_{r \to \infty} w(\bigvee_{j=1}^{r} \psi(a_j))$,

given that w must also satisfy K.

A number of possible answers to such a question have been proposed both for propositional and predicate languages, for example [1], [2], [3], [9], [10], [11], [17], [19], [20], [21], [22], based on various underlying assumptions about the form and origin of the knowledge and the probability function w, see [3] for a discussion. As in that paper we shall assume that w is a subjective probability function corresponding to the beliefs of some agent and that the assigning agent intends to act rationally or logically (though we shall not make any attempt to define these terms here, instead simply leaving it to the reader to decide to what extent our proposals fulfill that intention).

The method we shall describe in this note extends a well developed approach (see [17, Chapter 6]) for the analogous problem in the *propositional* case to the limited *predicate* situation when the language L is *purely unary*. This same path has already been trodden in [3] in a special case (viz. the Maximum Entropy Inference Process). The main novelty in this paper is in giving a general result which applies to a wide range of inference processes.

In the next section we explain, in the specific case of the Minimum Distance Inference Process (see [17, p76]), this method for picking a probability function satisfying K and the key limit result, which we prove in the subsequent section. In the final section we consider how this specific case generalizes.

2 The Method

The idea, as explained in [3], for assigning a probability to a sentence $\theta(a_1, a_2, \ldots, a_m)$ from SL is that this should be the limit as r tends to ∞ of the probability that one would assign to it being true in a finite structure with universe $\{ a_i \mid i \leq r \}$. In other words we wish to approximate our beliefs in what holds in a universe with denumerably many individuals $a_1, a_2, \ldots$ with our beliefs of what holds in its large finite substructures.

In more detail let L^k be the sublanguage of L with the same unary predicate symbols $P_1, \ldots, P_n$ but only the constant symbols $a_1, ..., a_k$ and let $Q_1, ..., Q_J$, $J = 2^n$ enumerate all formulas of the form

$$P_1^{\epsilon_1}(x) \wedge P_2^{\epsilon_2}(x) \wedge ... \wedge P_n^{\epsilon_n}(x) \tag{1}$$

where the $\epsilon_1, \epsilon_2, \ldots, \epsilon_n \in \{0, 1\}$ and $P^1 = P, P^0 = \neg P$. Let $\mathfrak{L}^r$ be the propositional language with the propositional variables $P_j(a_i)$, $i = 1, ..., r$ $j = 1, ..., n$. For $k < r$ define $()^{(r)} : SL^k \to S\mathfrak{L}^r$ inductively as follows:

$$P_j(a_i)^{(r)} = P_j(a_i),$$

$$(\neg\phi)^{(r)} = \neg\phi^{(r)},$$

$$(\phi \vee \theta)^{(r)} = \phi^{(r)} \vee \theta^{(r)},$$
$$(\phi \wedge \theta)^{(r)} = \phi^{(r)} \wedge \theta^{(r)},$$
$$(\exists x \psi(x))^{(r)} = \bigvee_{i=1}^{r} \psi(a_i)^{(r)}.$$

Let $K^{(r)}$ be the result of replacing every sentence θ_i in K by $\theta_i^{(r)}$. As indicated above we now wish to make a 'rational' choice $N(K^{(r)})$ of a probability function satisfying $K^{(r)}$ and thence define our 'rational' probability function w satisfying K by

$$w(\theta) = \lim_{r \to \infty} N(K^{(r)})(\theta^{(r)}).$$

Apart from the question of whether this limit even exists (which we will confront in the next section), and even then satisfies K, we need to justify our assignment of probabilities for these finite substructures satisfying the $K^{(r)}$. Fortunately however we are now essentially working in the propositional calculus and a number of such assignment processes, in this context called *Inference Processes,* N, for picking a probability function $N(K)$ (or set of functions, see [22]) satisfying a probabilistic propositional knowledge base K have been studied, and to some extent justified, see for example [17, Chapter 6].

Currently the generally most accepted inference process according to this criterion of rationality is the so called Maximum Entropy Inference Process, and indeed the programme we are advocating in this paper has already been carried out for maximum entropy in [3]. What we plan to do here is to retread this path using the Minimum Distance Inference Process, MD, and then point out how the necessary steps actually hold for a wide range of other inference processes.

Before proceeding with the proof we need to recall the definition of MD. Given a finite proposition language, say with propositional variables $p_1, p_2, \ldots, p_k$, a probability function v on the sentences of this language is determined by (and will be identified with) the vector

$$\langle v(\beta_1), v(\beta_2), \ldots, v(\beta_{2^k}) \rangle \in \mathbb{D}_{2^k} = \{ \langle x_1, x_2, \ldots, x_{2^k} \rangle \mid x_i \geq 0, \ \sum_i x_i = 1 \}$$

where the β_j run through the *atoms*

$$p_1^{\epsilon_1} \wedge p_2^{\epsilon_2} \wedge \ldots \wedge p_k^{\epsilon_k}.$$

Given a non-empty closed and convex subset C of $\mathbb{D}_{2^k}$ we define $MD(C)$ to be that (unique) probability function $\langle x_1, x_2, \ldots, x_{2^k} \rangle \in C$ for which $\sum_i x_i^2$ is minimal. Equivalently that point in C closest in Euclidean distance to the probability function $\langle 2^{-k}, 2^{-k}, \ldots, 2^{-k} \rangle$, which we can think of as representing complete ignorance.

3 The Existence of the Limit

The following results appear in [3] (also there as Lemma 1, Lemma 2 and Theorem 3) and we shall use them in what follows.

Lemma 1. *If $\theta, \phi \in SL^k$ and $k \leq r$ and $\theta \equiv \phi$ then $\theta^{(r)} \equiv \phi^{(r)}$.*

Let α_i for $i = 1, ..., J^k$ enumerate the exhaustive and exclusive set of sentences of the form

$$\bigwedge_{i=1}^{k} Q_{m_i}(a_i)$$

where the Q_i are as in (1).

Lemma 2. *Any sentence $\theta \in SL^k$ is equivalent to a disjunction of consistent sentences $\phi_{i,\boldsymbol{\epsilon}}$ of the form*

$$\alpha_i \wedge \bigwedge_{j=1}^{J} (\exists x Q_j(x))^{\epsilon_j}$$

where the $\epsilon_j \in \{0, 1\}$ and $\models \neg(\phi_{i,\boldsymbol{\epsilon}} \wedge \phi_{j,\boldsymbol{\delta}})$ whenever $\langle i, \boldsymbol{\epsilon} \rangle \neq \langle j, \boldsymbol{\delta} \rangle$.

Theorem 3. *$K^{(r)}$ is a satisfiable subset of $S\mathfrak{L}^r$ for large r.*

Theorem 4. *For $\theta \in SL$:*

$$w(\theta) = \lim_{r \to \infty} MD(K^{(r)})(\theta^{(r)})$$

exists and is a probability function on L satisfying K.

Proof. Assume throughout that r is large so that Theorem 3 applies. By Lemma 2 every sentence $\theta(a_1, ..., a_k) \in SL$ is equivalent to a disjunction of consistent sentences of the form

$$\phi_{i,\boldsymbol{\epsilon}} = \alpha_i \wedge \bigwedge_{j=1}^{J} (\exists x Q_j(x))^{\epsilon_j}.$$

If $\alpha_i = \bigwedge_{j=1}^{k} Q_{m_j}(a_j)$ then let

$$A_i = \{ m_j \,|\, j = 1, ..., k \}, \; P_{\boldsymbol{\epsilon}} = \{ j \,|\, \epsilon_j = 1 \}, \; P_{i,\boldsymbol{\epsilon}} = \{ j \,|\, j \in P_{\boldsymbol{\epsilon}} \text{ and } j \notin A_i \}$$

so

$$\phi_{i,\boldsymbol{\epsilon}}^{(r)} = \alpha_i \wedge \bigwedge_{j=1}^{J} \left(\bigvee_{i=1}^{r} Q_j(a_i) \right)^{\epsilon_j}$$

will be equivalent to

$$\bigvee_{\substack{m_j \in P_{\boldsymbol{\epsilon}} \text{ for } j=k+1,...,r \\ P_{i,\boldsymbol{\epsilon}} \subseteq \{ m_j \,|\, k+1 \leq j \leq r \}}} \left(\alpha_i \wedge \bigwedge_{j=k+1}^{r} Q_{m_j}(a_j) \right). \quad (2)$$

If we set $p_{\boldsymbol{\epsilon}} = |P_{\boldsymbol{\epsilon}}|$ and $p_{i,\boldsymbol{\epsilon}} = |P_{i,\boldsymbol{\epsilon}}|$ then the number of disjuncts (i.e. atoms of $\mathfrak{L}^r$) in (2) will be

$$\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}} (-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}.$$

Let $w^{(r)} = MD(K^{(r)})$. Since MD satisfies renaming (see [17, p95]), for every atom ζ in the disjunction in (2)

$$w^{(r)}(\zeta) = \frac{w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}})}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}}.$$

Hence if the ζ_j enumerate the atoms of $\mathfrak{L}^r$,

$$\begin{aligned}\sum_{j=1}^{J^r}(w^{(r)}(\zeta_j))^2 &= \sum_{i,\boldsymbol{\epsilon}} \left(\frac{w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}})}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}} \right)^2 \\ &= \sum_{i,\boldsymbol{\epsilon}} \frac{(w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}}))^2}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}}.\end{aligned}$$

From this it follows that w satisfying $K^{(r)}$ is equivalent to some set of linear inequalities

$$\sum_{i,\boldsymbol{\epsilon}} c_{i,\boldsymbol{\epsilon},j} w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}}) \leq b_j, \quad j = 1, \ldots, m \tag{3}$$

where the $c_{i,\boldsymbol{\epsilon},j}$ and b_j *do not depend on* r. Hence the vector of values $w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}})$ (as $i, \boldsymbol{\epsilon}$ vary) is that vector $x_{i,\boldsymbol{\epsilon}} \geq 0$ satisfying (3) for which

$$\sum_{i,\boldsymbol{\epsilon}} \frac{x^2_{i,\boldsymbol{\epsilon}}}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}} \tag{4}$$

is minimal.

Let $c_1 < c_2 < ... < c_k$ be the distinct values for $p_{\boldsymbol{\epsilon}}$ which occur here and define the sets

$$T_0 = \{ \boldsymbol{x} \mid \sum_{i,\boldsymbol{\epsilon}} c_{i,\boldsymbol{\epsilon},j} w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}}) \leq b_j, \quad j = 1, \ldots, m\}$$

and

$$T_{j+1} = \{ \boldsymbol{x} \in T_j \mid \sum_{i, p_{\boldsymbol{\epsilon}} = c_{j+1}} x^2_{i,\boldsymbol{\epsilon}} \text{ is minimal} \}$$

for $0 \leq j < k$. Since these T_j are closed and convex any two points in T_j agree on those coordinates $\langle i, \boldsymbol{\epsilon} \rangle$ with $p_{\boldsymbol{\epsilon}} \leq c_j$. Hence T_k consists of a single point, $\boldsymbol{X}$ say. Notice that this point does not depend on r.

Since $\boldsymbol{X} \in T_0$ by (4)

$$\sum_{i,\boldsymbol{\epsilon}} \frac{(w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}}))^2}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}} \leq \sum_{i,\boldsymbol{\epsilon}} \frac{X^2_{i,\boldsymbol{\epsilon}}}{\sum_{j=0}^{p_{i,\boldsymbol{\epsilon}}}(-1)^j \binom{p_{i,\boldsymbol{\epsilon}}}{j} (p_{\boldsymbol{\epsilon}} - j)^{r-k}}. \tag{5}$$

The $w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}})$ have a convergent subsequence (as $r \to \infty$), which for notational convenience we shall assume is the whole sequence, say

$$Y_{i,\epsilon} = \lim_{r \to \infty} w^{(r)}(\phi_{i,\epsilon}^{(r)}).$$

We shall show that $Y_{i,\epsilon} = X_{i,\epsilon}$ in turn for each of the cases $p_\epsilon = c_1, c_2, \ldots, c_k$.

Firstly for the case of c_1, multiplying (5) by c_1^{r-k} and taking the limit as $r \to \infty$ gives

$$\sum_{i, p_\epsilon = c_1} Y_{i,\epsilon}^2 \leq \sum_{i, p_\epsilon = c_1} X_{i,\epsilon}^2$$

and hence for such ϵ, $Y_{i,\epsilon} = X_{i,\epsilon}$ by definition of T_1.

To handle the case $p_\epsilon = c_2$ and beyond we will need the following lemma.

Lemma 5. *Let $B \subseteq R^m$ be a convex polyhedron with corners $\boldsymbol{a_1}, \boldsymbol{a_2}, ..., \boldsymbol{a_q}$. Let $\boldsymbol{c} \in B$ and let $f : R^m \to R^n$ be the projection function given by*

$$f \langle x_1, x_2, ..., x_m \rangle = \langle x_1, x_2, ..., x_n \rangle$$

Suppose that $\boldsymbol{y_j} \in R^n$ for $j \in \mathbb{N}$ are such that $f^{-1}(\boldsymbol{y_j}) \cap B \neq \emptyset$ for all j and

$$lim_{j \to \infty} \boldsymbol{y_j} = f(\boldsymbol{c}).$$

Then there is a subsequence $\boldsymbol{z_j} \in B$ converging to $\boldsymbol{c}$ such that the $f(\boldsymbol{z_j})$ form a subsequence of the $\boldsymbol{y_j}$.

Proof. Any point in B can be written as a linear combination

$$\boldsymbol{c} + \sum_{i=1}^{q} \lambda_i \boldsymbol{e_i},$$

where $\boldsymbol{e_i} = \boldsymbol{a_i} - \boldsymbol{c}$ and the $\lambda_i \geq 0$ with sum ≤ 1, so any $\boldsymbol{x} \in f(B)$ can be written as

$$f(\boldsymbol{c}) + \sum_{i=1}^{q} \lambda_i f(\boldsymbol{e}_i)$$

with $\lambda_i > 0$ with sum at most 1, where we drop any terms with $\lambda_i = 0$ (but to avoid messy notation assume there are none, and that this is true also for the $\boldsymbol{y_j}$, otherwise pick a suitable subsequence with the same zero terms throughout). Now for each $\boldsymbol{y_j}$ pick one such presentation:

$$\boldsymbol{y}_j = f(\boldsymbol{c}) + \sum_{i=1}^{q} \lambda_{ij} f(\boldsymbol{e}_i)$$

and set

$$\boldsymbol{z}_j = \boldsymbol{c} + \sum_{i=1}^{q} \lambda_{ij} \boldsymbol{e}_i.$$

It is obvious that the $f(\boldsymbol{z_j})$ form a subsequence of the $\boldsymbol{y_j}$. To show that $\lim_{j \to \infty} \boldsymbol{z_j} = \boldsymbol{c}$ it is enough to show that $\lim_{j \to \infty} \sum_{i=1}^{q} \lambda_{ij} \boldsymbol{e}_i = \boldsymbol{0}$. To this end we will show that $\lim_{j \to \infty} \lambda_{ij} = 0$. We know that $\lim_{j \to \infty} \boldsymbol{y_j} = f(c)$ and so we have

$$\lim_{j\to\infty}\sum_{i=1}^{q}\lambda_{ij}f(\boldsymbol{e}_i)=\boldsymbol{0}.$$

Let

$$\boldsymbol{t_j}=\sum_{i=1}^{q}\lambda_{ij}f(\boldsymbol{e}_i),$$

so

$$\lim_{j\to\infty}\boldsymbol{t_j}=\boldsymbol{0}$$

and $\boldsymbol{t_j}$ is in the convex polyhedron with corners $f(\boldsymbol{e_i})$. For each $\boldsymbol{t_j}$ pick a smallest set $f(\boldsymbol{e_{i_1}}),...,f(\boldsymbol{e_{i_h}})$ such that:

$$\boldsymbol{t_j}=\sum_{k=1}^{h}\lambda_{i_kj}f(\boldsymbol{e_{i_k}}) \tag{6}$$

with $\lambda_{i_kj}\geq 0$ and $\sum_{i_k}\lambda_{i_kj}\leq 1$. By taking a subsequence if necessary we can assume that the $\boldsymbol{t_j}$ all have the same smallest set and that $\lambda_{i_kj}\to\lambda_{i_k}$ as $j\to\infty$. For simplicity of notation we assume that these smallest sets are all the $\boldsymbol{e_i}$, so (6) will become

$$\boldsymbol{t_j}=\sum_{i=1}^{q}\lambda_{ij}f(\boldsymbol{e_i}) \tag{7}$$

and

$$\boldsymbol{0}=\sum_{i=1}^{q}\lambda_{i}f(\boldsymbol{e_i}). \tag{8}$$

Now if all the $\lambda_i=0$ we have the required result, otherwise suppose some of the $\lambda_i>0$. Then from (7) and (8) we will have:

$$\boldsymbol{t_j}=\sum_{i=1}^{q}(\lambda_{ij}-\nu\lambda_i)f(\boldsymbol{e_i}) \tag{9}$$

Now as we increase ν from 0 one of the coefficients in (9) will become zero while others are still non-negative and this contradicts the choice of smallest set, and so is a contradiction. Hence we must have that all the $\lambda_i=0$, as required. ■

To continue the proof of Theorem 4 suppose that $Y_{i,\boldsymbol{\epsilon}}\neq X_{i,\boldsymbol{\epsilon}}$ for some $p_{\boldsymbol{\epsilon}}=c_2$. We have already shown that we do have equality when $p_{\boldsymbol{\epsilon}}=c_1$ so by Lemma 5 there is a sequence of vectors $z^{(r)}_{i,\boldsymbol{\epsilon}}\in T_0$ such that for each $i,\boldsymbol{\epsilon}$

$$\lim_{r\to\infty}z^{(r)}_{i,\boldsymbol{\epsilon}}=X_{i,\boldsymbol{\epsilon}}$$

and the $z^{(r)}_{i,\boldsymbol{\epsilon}}$ form a subsequence of the $w^{(r)}(\phi^{(r)}_{i,\boldsymbol{\epsilon}})$ for $p_{\boldsymbol{\epsilon}}=c_1$. For simplicity of notation we will again assume that this subsequent is the whole sequence.

By definition of $w^{(r)}$,

$$\sum_{i,\epsilon} \frac{(w^{(r)}(\phi_{i,\epsilon}^{(r)}))^2}{\sum_{j=0}^{p_{i,\epsilon}}(-1)^j \binom{p_{i,\epsilon}}{j} (p_\epsilon - j)^{r-k}} \leq \sum_{i,\epsilon} \frac{z_{i,\epsilon}^2}{\sum_{j=0}^{p_{i,\epsilon}}(-1)^j \binom{p_{i,\epsilon}}{j} (p_\epsilon - j)^{r-k}}.$$

The parts of these sums for $p_\epsilon = c_1$ are equal, so can be cancelled out. Multiplying both sides of what remains by c_2^{r-k} and taking the limit as $r \to \infty$ we obtain, just as in the case of c_1 that

$$\sum_{i,p_\epsilon=c_2} Y_{i,\epsilon} \leq \sum_{i,p_\epsilon=c_2} X_{i,\epsilon}.$$

But since, as we have already shown, the vector $Y_{i,\epsilon}$ is in T_1 this inequality must also go the other way since the vector $X_{i,\epsilon}$ is in T_2. We conclude as required that $Y_{i,\epsilon} = X_{i,\epsilon}$ whenever $p_\epsilon = c_2$. Similar arguments give the same result for $c_3, c_4, \ldots, c_k$, as required.

Finally, since any linear identity satisfied by all the $w^{(r)}$ eventually will be satisfied by their limit it is clear that this limit is a probability function satisfying K. ∎

4 Some Generalizations

So far we have proved Theorem 4 for a specific inference process, Minimum Distance, but in fact analogous proofs give the the result too for the Maximum Entropy Inference Process (already proved in [3]), the Limiting Centre of Mass Inference Process (see [17, p73-74]) and the spectrum of other inference processes based on generalized Renyi Entropies. [For further results along these lines see [23].]

In our original question we imagined an agent wishing to assign probabilities to all sentences on the basis of qualified knowledge K. A special case of this is when K simply amounts to the assertion that some consistent, finite, set of axioms $\mathcal{T}$ hold categorically, i.e.

$$K = \{ w(\phi) = 1 \mid \phi \in \mathcal{T} \}.$$

In this case our question might be reformulated as

> *Given a finite (consistent) set $\mathcal{T}$ of first order axioms what should we take as the default or most normal model of $\mathcal{T}$? More precisely, if we know only that the structure M with universe $\{ a_i \mid i \in \mathbb{N} \}$ is a model of $\mathcal{T}$ what probability should we give to a sentence $\theta(a_1, a_2, \ldots, a_n)$ being true in M?*

There are various approaches one might take to this question depending on the interpretation of 'most normal'. For example within a model theory context one might consider a *prime model*, where such exists, to be the 'most normal' in the sense of being the smallest and the canonical example (see for example [5, p96], [8, p336]). On the other hand one might feel that if possible the default model should be existentially closed (see [24] for a precise definition) in the sense that any quantifier free formula which could be satisfied in a superstructure model of $\mathcal{T}$ was already satisfied in the default model. Alternatively we might consider arguing via the distribution of models, see for example [1], [2], [9], [10], [11], in order to make the default the 'average' model.

Furthermore, at first sight it would appear that there was already a rather well studied approach to this problem via Inductive Logic. In that subject, see for example [4], [7], [13], [16], this same problem with $\mathcal{T} = \emptyset$ is quite central. So it might seem that a solution to our problem here could be had by simply taking a rationally justified probability function w championed within Inductive Logic for the case of a completely empty knowledge base and then conditioning w on $\bigwedge \mathcal{T}$. The first problem with that approach however is that there is currently no clearly favored rational solution to the Inductive Logic problem. But more seriously, those solutions w which have been proposed generally give non-tautologous universal sentences probability 0, see for example [12], [14], [15], [16, p22-23], [17, p196-197], and once $w(\bigwedge \mathcal{T}) = 0$ such conditioning will not be possible.[1,2]

However if we assume that the sentences of $\mathcal{T}$ come from the *purely unary language* of the preceding sections then the method described in this paper, based on any of the above inference processes, indeed in this simple case of *categorical* knowledge, $K = \{\, w(\phi) = 1 \mid \phi \in \mathcal{T} \,\}$, based on *any* inference process just satisfying the Renaming Principle, can be applied, and in fact always yield the same answer. Namely that, in the notation of the proof of Theorem 4, if $\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^s$ are all the vectors $\boldsymbol{\epsilon}$ for which $\bigwedge_{j=1}^{J} (\exists x Q_j(x))^{\epsilon_j}$ is consistent with $\mathcal{T}$ and amongst which $p_{\boldsymbol{\epsilon}}$ takes its largest value then $w(\theta(a_1, \ldots, a_k)) = H/K$ where

$$K = |\{\, \phi_{i,\boldsymbol{\epsilon}^r} \mid \phi_{i,\boldsymbol{\epsilon}^r} \text{ is consistent with } \bigwedge \mathcal{T},\, 1 \leq i \leq J^k,\, 1 \leq r \leq s \,\}|,$$

$$H = |\{\, \phi_{i,\boldsymbol{\epsilon}^r} \mid \phi_{i,\boldsymbol{\epsilon}^r} \text{ is consistent with } \theta(a_1, \ldots, a_k) \wedge \bigwedge \mathcal{T}, 1 \leq i \leq J^k,\, 1 \leq r \leq s \,\}|.$$

In particular then w gives probability 1 to

$$\bigvee_{i=1}^{s} \bigwedge_{j=1}^{J} (\exists x Q_j(x))^{\epsilon_j^i},$$

---

[1] It is true that proposals have been made for solutions to the Inductive Logic problem which give some non-tautologous universal sentences non-zero probability, see for example [6], [12], [14], [15], [18]. However they seem (to us) too ad hoc to be seriously considered 'logical'.

[2] This apparent discontinuity between the cases when $\mathcal{T} =/\neq \emptyset$ is intriguing – the method we shall apply in this paper still works when $\mathcal{T} = \emptyset$ but gives an unsatisfactory solution to the inductive logic problem, unsatisfactory in that it corresponds to the so called completely independent solution which entertains no induction i.e. learning by example, see for example [17, p172].

(and probability $1/s$ to each of the disjuncts), thus exclusively favoring those models M of $\mathcal{T}$ in which as many of the Q_j are satisfied as possible, that is the existentially closed models of $\mathcal{T}$.

It would of course be nice to extend this approach (or develop an alternative) to more than just these rather trivial unary languages. For example to the theory saying that the relation $<$ is transitive. In this case what is a 'sensible' probability to even give to $a_1 < a_2$? Certainly the simple method suggested here fails, but whether it can be suitable adapted to make it more applicable, whilst at the same time retaining credibility in relation to the original philosophical question apparently remains to be investigated.

Acknowledgement

We would like to thank the referees for their useful comments.

References

1. Bacchus, F., Grove, A.J., Halpern, J.Y., Koller, D.: Generating new beliefs from old. In: Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence (UAI 1994), pp. 37–45 (1994)
2. Bacchus, F., Grove, A.J., Halpern, J.Y., Koller, D.: From statistical knowledge to degrees of belief. Artificial Intelligence 87, 75–143 (1996)
3. Barnett, O.W., Paris, J.B.: Maximum Entropy inference with qualified knowledge. Logic Journal of the IGPL 16(1), 85–98 (2008)
4. Carnap, R.: A basic system of inductive logic. In: Jeffrey, R.C. (ed.) Studies in Inductive Logic and Probability, vol. II, pp. 7–155. University of California Press (1980)
5. Chang, C.C., Keisler, H.J.: Model Theory. Studies in Logic and the Foundations of Mathematics, vol. 73. North Holland, Amsterdam (1973)
6. Dimitracopoulos, C., Paris, J.B., Vencovská, A., Wilmers, G.M.: A multivariate probability distribution based on the propositional calculus, Manchester Centre for Pure Mathematics, University of Manchester, UK, preprint number 1999/6, `http://www.maths.manchester.ac.uk/~jeff/`
7. Fitelson, B.: Inductive Logic, `http://fitelson.org/il.pdf`
8. Hodges, W.: Model Theory. Cambridge University Press, Cambridge (1993)
9. Grove, A.J., Halpern, J.Y., Koller, D.: Random Worlds and Maximum Entropy. Journal of Artificial Intelligence Research 2, 33–88 (1994)
10. Grove, A.J., Halpern, J.Y., Koller, D.: Asymptotic conditional probabilities: the unary case. SIAM J. of Computing 25(1), 1–51 (1996)
11. Grove, A.J., Halpern, J.Y., Koller, D.: Asymptotic conditional probabilities: the non-unary case. J. Symbolic Logic 61(1), 250–276 (1996)
12. Hintikka, J., Niiniluoto, I.: An axiomatic foundation for the logic of inductive generalization. In: Jeffrey, R.C. (ed.) Studies in Inductive Logic and Probability, vol. II, pp. 158–181. University of California Press, Berkeley, Los Angeles (1980)
13. Johnson, W.E.: Probability: The deductive and inductive problems. Mind 41(164), 409–423 (1932)

14. Kuipers, T.A.F.: A survey of inductive systems. In: Jeffrey, R.C. (ed.) Studies in Inductive Logic and Probability, vol. II, pp. 183–192. University of California Press, Berkeley, Los Angeles (1980)
15. Kuipers, T.A.F.: On the generalization of the continuum of inductive methods to universal hypotheses. Synthese 37, 255–284 (1978)
16. Paris, J.B.: A short course on Inductive Logic. In: JAIST 2007 (2007), http://www.maths.manchester.ac.uk/~jeff
17. Paris, J.B.: The Uncertain Reasoner's Companion. Cambridge University Press, Cambridge (1994)
18. Paris, J.B.: On the distribution of probability functions in the natural world. In: Hendricks, V.F., Pedersen, S.A., Jøgensen, K.F. (eds.) Probability Theory: Philosophy, Recent History and Relations to Science. Synthese Library, vol. 297, pp. 125–145 (2001)
19. Paris, J.B.: Vencovská, On the applicability of maximum entropy to inexact reasoning. International Journal of Approximate Reasoning 3(1), 1–34 (1989)
20. Paris, J.B., Vencovská: A note on the inevitability of maximum entropy. International Journal of Approximate Reasoning 4(3), 183–224 (1990)
21. Paris, J.B., Vencovská: In defense of the maximum entropy inference process. International Journal of Approximate Reasoning 17(1), 77–103 (1997)
22. Paris, J.B., Vencovská, A.: Common sense and stochastic independence. In: Corfield, D., Williamson, J. (eds.) Foundations of Bayesianism, pp. 203–240. Kluwer Academic Press, Dordrecht (2001)
23. Rad, S.R.: PhD Thesis, Manchester University, Manchester, UK (to appear)
24. http://en.wikipedia.org/wiki/Existentially_closed_model