MANCHESTER
1824

The University
of Manchester

The Finite Values Property

Howarth, E. and Paris, J.B.

2017

MIMS EPrint: **2017.2**

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

The Finite Values Property

E.Howarth* and J.B.Paris†
School of Mathematics
The University of Manchester
Manchester M13 9PL

lizhowarth@outlook.com, jeff.paris@manchester.ac.uk

November 24, 2015

Dedicated to Gabriele Kern-Isberner

Abstract

We argue that the simplicity condition on a probability function on sentences of a predicate language L that it takes only finitely many values on the sentences of any finite sub-language of L can be viewed as rational. We then go on to investigate consequences of this condition, linking it to the model theoretic notion of quantifier elimination.

Key words: Finite Values Property, Quantifier Elimination, Inductive Logic, Logical Probability, Rationality, Uncertain Reasoning.

*Supported by a UK Engineering and Physical Sciences Research Council (EPSRC) Studentship.

†Supported by a UK Engineering and Physical Sciences Research Council Research Grant.

Introduction

A much studied problem over the last 35 years in AI is how one should choose a particular probability function to satisfy a given set of (satisfiable) constraints. Obviously in all but the most trivial cases such a choice has to go beyond what is actually stated in the constraints per se and as a result various methodologies have been proposed on the basis of what it seems reasonable to additionally assume in the circumstances. For example if the constraints actually apply to an objective probability function then an approach based on some sort of averaging of all the possible candidate probability functions might seem appropriate whilst if one is seeking a subjective, common sense, probability function then one might opt for minimizing the information content beyond what is already there in the constraints, for example by maximizing entropy (relevant to the context of this paper and volume see [1],[12],[13],[14],[15],[21],[22],[23]).

In this paper we shall consider a property of the chosen probability function which, in the context of a predicate language, seems attractive on the basis of both pragmatic and rational considerations, namely that in a sense to be explained shortly the chosen probability function only takes finitely many values on the sentences of any particular finite sublanguage. 'Pragmatic' because such a property can simplify predictions and 'rational' in the sense of Occam's Razor, that it is rational to adopt as simple an explanation as possible.

Firstly however we need to introduce the particular context in which we shall be working, namely Pure Inductive Logic, see for example [17], [18], [19].

Context

Let L be the first order language with constant symbols a_n, $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$ and relation symbols $R_1, R_2, \ldots, R_q$ of arities $r_1, r_2, \ldots, r_q$ respectively but no function symbols nor equality.

Let $SL, QFSL$ denote the sentences and quantifier free sentences of L and let $SL^{(n)}$ denote those sentences of L which do not mention any constants a_r with $r > n$. In other words sentences of the finite sublanguage of L with just the relation symbols $R_1, R_2, \ldots, R_q$ and the constant symbols $a_1, a_2, \ldots, a_n$.

We say that a function $w : SL \to [0, 1]$ is a *probability function on SL* if for all $\theta, \phi, \exists x\, \psi(x) \in SL$

(P1) $\vDash \theta \;\Rightarrow\; w(\theta) = 1.$

(P2) $\theta \vDash \neg\phi \;\Rightarrow\; w(\theta \vee \phi) = w(\theta) + w(\phi).$

(P3) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n)).$

With this definition all the expected properties of probabilities hold, in particular because of their significance to this paper, if $\theta, \phi \in SL$ and $\theta \models \phi$ then $w(\theta) \leq w(\phi)$ and if $\models \theta \leftrightarrow \phi$ then $w(\theta) = w(\phi)$ (see for example [19, Chapter 3] for more details).

Let $\mathcal{T}L$ be the set of structures M for L with universe the interpretations of the a_n (also denoted a_n). So for $M \in \mathcal{T}L$ every element in the universe of M is named by a constant.

In our view the central question which Pure Inductive Logic aims to investigate is:

> Question: Given an agent $\mathcal{A}$ inhabiting an unknown structure $M \in \mathcal{T}L$ and $\theta \in SL$ what probability $w(\theta)$ should $\mathcal{A}$ *rationally*, or *logically*, assign to θ?

– or more generally given that we obviously expect $\mathcal{A}$'s answers to be mutually consistent:

> Question: Given an agent $\mathcal{A}$ inhabiting an unknown structure $M \in \mathcal{T}L$, *rationally* or *logically*, what probability function w should $\mathcal{A}$ adopt?

It is important to appreciate here that by 'probability' we mean subjective probability (i.e. degree of belief or willingness to bet) and that $\mathcal{A}$ should know nothing more about M, so have no intended interpretation in mind for the constant and relation symbols.

The key obstacle in answering this question is of course what we mean by 'rational'. In the absence of any precise definition of this term the main approach (since Carnap essentially founded the subject in the 1940's, see for example [2], [3], [4], [5], [6]) is to postulate possible properties of w which are 'rational' in some intuitive sense and then investigate what consequences these entail. Of these properties the most widely accepted is that w should satisfy:

Constant Exchangeability, Ex:

For $\phi(a_{i_1}, a_{i_2}, \ldots, a_{i_m}) \in SL$,[1]

$$w(\phi(a_1, a_2, \ldots, a_m)) = w(\phi(a_{i_1}, a_{i_2}, \ldots, a_{i_m})).$$

The rational justification here is that the agent has no knowledge about any of the a_i so it would be irrational to treat them differently when assigning probabilities.

Constant Exchangeability, Ex, is so widely accepted in this area that we will henceforth *assume it throughout for all the probability functions we consider.*

The purpose of this paper is to investigate a further ostensibly rational principle (the Finite Values Property, FVP) which is based on the putative idea that simplicity is a facet of rationality.

The Finite Values Property

Let w be a probability function on SL satisfying Ex. We say that w satisfies the *Finite Values Property*, FVP, if

$$\{w(\theta) \mid \theta \in SL^{(n)}\}$$

is finite for each $n \in \mathbb{N}$.

The Finite Values Property may seem rather surprising, since although at each 'level' n, the number of constants in each sentence of $SL^{(n)}$ is bounded, no such restriction is placed on the length or complexity of these sentences. FVP is the formal version of the pragmatically and rationally desirable 'finiteness property' alluded to in the introduction.[2]

We shall say that w satisfies FVP_n if

$$\{w(\theta) \mid \theta \in SL^{(n)}\}$$

is finite. So FVP amounts to $\forall n \in \mathbb{N}$, FVP_n. Clearly if w satisfies FVP_n then it also satisfies FVP_m for $m < n$.

[1]The convention is that when a sentence ϕ is written in this form it is assumed (unless otherwise stated) that the displayed constants are distinct and include all the constants actually occurring in ϕ.

[2]We would point out that in the context of Pure Inductive Logic our current favoured choice of probability functions on grounds of rationality are the so called *homogeneous* probability functions, see [19, Chapter 30], and these do indeed additionally satisfy this further pragmatic requirement of FVP, see [9].

The aim of this paper is to initiate an investigation into which probability functions satisfy this property and what further structure they must have. Our first result is that FVP always holds if the language L is unary (i.e. each $r_i = 1$), a standard assumption in fact in Inductive Logic up to the 21st century.

Theorem 1. *If L is unary and w is a probability function on SL then w satisfies FVP.*

Proof. Suppose L is unary and let $\alpha_i(x)$ for $i = 1, 2, \ldots, 2^q$ enumerate the *atoms* of L, that is the formulae of L of the form

$$R_1(x)^{\epsilon_1} \wedge R_2(x)^{\epsilon_2} \wedge \ldots \wedge R_q(x)^{\epsilon_q}$$

where the $\epsilon_i \in \{0, 1\}$ and for a formula ϕ, $\phi^1 = \phi$, $\phi^0 = \neg\phi$.

Let $\theta(a_1, \ldots, a_n) \in SL^{(n)}$. It is well known, see for example [16, Theorem 4], that θ is logically equivalent to a sentence θ' of the form

$$\bigvee_{k=1}^{l} \left(\bigwedge_{j=1}^{2^q} (\exists x\, \alpha_j(x))^{\epsilon_{k_j}} \wedge \bigwedge_{i=1}^{n} \alpha_{f_{k_i}}(a_i) \right),$$

where each $\vec{\epsilon_k} \in \{0, 1\}^{2^q}$ and the disjuncts are disjoint and satisfiable.

Let B be the set of satisfiable disjuncts (up to logical equivalence)

$$\bigwedge_{j=1}^{2^q} (\exists x\, \alpha_j(x))^{\epsilon_j} \wedge \bigwedge_{i=1}^{n} \alpha_{f_i}(a_i). \tag{1}$$

Since there are 2 choices for ϵ_j for each $j = 1, \ldots, 2^q$, and at most 2^q choices for α_{f_i} for each $i = 1, \ldots, n$, this gives

$$|B| \leq 2^{2^q + qn}.$$

Since $\theta(a_1, \ldots, a_n)$ is logically equivalent to the disjunction of some subset of B, the size of $SL^{(n)}$ up to logical equivalence is bounded by the number $2^{|B|}$ of distinct subsets of B. Since logically equivalent sentences must get the same probability FVP_n, and hence FVP, follows. □

It turns out that Theorem 1 is a special, though important, case of a much more general result. Before we can give that result however it will be useful to introduce the concept of *ions*.

Ions

The following characterization of FVP_n shows that the set of sentences B in Theorem 1 has a counterpart wherever FVP_n occurs.

Theorem 2. *A probability function w on SL satisfies FVP_n just if there is a finite set of sentences*

$$B = \{\phi_1, \ldots, \phi_g\} \subset SL^{(n)}$$

such that

- $w(\phi_i \wedge \phi_j) = 0$ *for any* $1 \leq i < j \leq g$,
- $\sum_{i=1}^{g} w(\phi_i) = 1$,
- *for any* $\theta \in SL^{(n)}$ *there is a subset* B_θ *of* B *such that*

$$w\big(\theta \leftrightarrow \bigvee_{\phi \in B_\theta} \phi\big) = 1.$$

Proof. From left to right, suppose that w satisfies FVP_n. Initially let $B' = \{\top\}$ where $\top \in SL^{(0)}$ (the set of sentences of L mentioning no constants) is a tautology. Now repeatedly 'split' sentences in B' as follows. If possible pick $\phi \in B'$ for which there exists $\theta \in SL^{(n)}$ such that

$$0 \ < \ w(\phi \wedge \theta),\ w(\phi \wedge \neg\theta) \ < \ w(\phi),$$

and replace ϕ in B' by $\phi \wedge \theta$ and $\phi \wedge \neg\theta$. Repeat this step until no such ϕ remains. Note that at each stage of this process

$$w\big(\bigvee_{\phi \in B'} \phi\big) = 1 \tag{2}$$

and for any distinct $\phi, \eta \in B'$, $w(\phi \wedge \eta) = 0$.

To show that this process must halt after a finite number of steps suppose on the contrary that it did not. Then the $w(\phi)$ for ϕ appearing in B' at some stage cannot be bounded away from 0 since otherwise by (2) the $|B'|$ would also have to be uniformly bounded at all stages, which clearly is not the case. But if the $w(\phi)$ for ϕ appearing in a B' at some stage are not bounded away from 0 then this contradicts FVP_n.

Let

$$B = \{\phi_1, \phi_2, \ldots, \phi_g\} \subset SL^{(n)}$$

be the halting B', so for any $\theta \in SL^{(n)}$ and any $\phi_j \in B$,

$$w(\theta \wedge \phi_j) \in \{0, w(\phi_j)\}.$$

For $\theta \in SL^{(n)}$ let $B_\theta = \{\phi_j \in B \mid w(\theta \wedge \phi_j) = w(\phi_j)\}$. Then from (2)

$$w(\theta) = w\big(\theta \wedge \bigvee_{\phi \in B} \phi\big) = \sum_{\phi \in B} w(\theta \wedge \phi) = \sum_{\phi \in B_\theta} w(\phi) \tag{3}$$

since $w(\theta \wedge \phi) = 0$ for $\phi \in B - B_\theta$.

Furthermore, since

$$w(\phi_j) = w(\phi_j \wedge \theta) + w(\phi_j \wedge \neg\theta)$$

we have

$$B_{\neg\theta} = B - B_\theta. \tag{4}$$

Hence from (3) for $\neg\theta$,

$$w(\neg\theta \wedge \bigvee_{\phi \in B_\theta} \phi) = w(\bigvee_{\psi \in B_{\neg\theta}} \psi \wedge \bigvee_{\phi \in B_\theta} \phi) = 0$$

since by (4) for $\phi \in B_\theta, \psi \in B_{\neg\theta}$, $\phi \neq \psi$ so $w(\phi \wedge \psi) = 0$. Together with (3) this forces that

$$w\big(\theta \leftrightarrow \bigvee_{\phi \in B_\theta} \phi\big) = 1.$$

In the other direction, it is clear that if $B = \{\phi_1, \ldots, \phi_g\} \subset SL^{(n)}$ is as described in the statement of the result, then for any $\theta \in SL^{(n)}$

$$w(\theta) = \sum_{\phi \in B_\theta} w(\phi),$$

and since the number of possible subsets B_θ of B is finite, then so is the range of $w \restriction SL^{(n)}$. $\square$

We will call such a set $B \subset SL^{(n)}$ with the properties given in Theorem 2 a set of *n-ions* for w. Note from the above result that when ϕ is an n-ion for w and $\theta \in SL^{(n)}$, $B_{\theta \wedge \phi}$ is either equal to $\{\phi\}$ or to $\emptyset$, so that

$$w(\theta \wedge \phi) \in \{0, w(\phi)\}. \tag{5}$$

Theorem 2 suggests a connection between FVP and the well known and very significant notion of *Quantifier Elimination* in Model Theory, a connection which we shall investigate in the next section.

Generalized Quantifier Elimination

Let T be a theory in $L^{(0)}$, meaning that $T \subset SL^{(0)}$ and T is closed under logical consequence. We shall say that T satisfies *Generalized Quantifier Elimination*, GQE, if there is a finite set of formulae of $L^{(0)}$,

$$\{\zeta_i(x_1, x_2, \ldots, x_k) \mid i = 1, 2, \ldots, m\}$$

such that for each formula $\theta(x_1, x_2, \ldots, x_n)$ of $L^{(0)}$ there is a Boolean combination $\psi(x_1, x_2, \ldots, x_n)$ of formulae $\zeta_i(y_1, y_2, \ldots, y_k)$ where $\{y_1, y_2, \ldots, y_k\} \subseteq \{x_1, x_2, \ldots, x_n\}$ and $i \in \{1, 2, \ldots, m\}$, such that $T \models \theta \leftrightarrow \psi$.

Note that this reduces to the standard notion of simply *Quantifier Elimination* in the case where the $\zeta_i(x_1, x_2, \ldots, x_k)$ are just the $R_i(x_1, x_2, \ldots, x_{r_i})$ for R_i a relation symbol of L and $k = \max\{r_i\}$.

Theorem 3. *Let w be a probability function on SL satisfying Ex and such that*

$$Th(w) = \{\phi \in SL^{(0)} \mid w(\phi) = 1\}$$

satisfies GQE. Then w satisfies FVP.

Proof. Let w be as given with the $\zeta_i(x_1, x_2, \ldots, x_k)$ for $i = 1, 2, \ldots, m$ as in the definition of GQE. Then for $\theta(x_1, x_2, \ldots, x_n)$ a formula of $L^{(0)}$ there is a Boolean combination $\psi(x_1, x_2, \ldots, x_n)$ of formulae $\zeta_i(y_1, y_2, \ldots, y_k)$ where $\{y_1, y_2, \ldots, y_k\} \subseteq \{x_1, x_2, \ldots, x_n\}$ and $i \in \{1, 2, \ldots, m\}$, such that $Th(w) \models \theta \leftrightarrow \psi$. Hence $\theta \leftrightarrow \psi$ is a logical consequence of some finite subset of $Th(w)$ and since the members of $Th(w)$ are all sentences it further follows that

$$\theta(a_1, \ldots, a_n) \leftrightarrow \psi(a_1, \ldots, a_n)$$

is also a logical consequence of this finite subset. Therefore, since $w(\phi) = 1$ for each $\phi \in Th(w)$,

$$w(\theta(a_1, \ldots, a_n) \leftrightarrow \psi(a_1, \ldots, a_n)) = 1$$

and consequently

$$w(\theta(a_1, \ldots, a_n)) = w(\psi(a_1, \ldots, a_n)).$$

But clearly here there are, up to logical equivalence, only finitely many choices for $\psi(a_1, \ldots, a_n)$ and hence only finitely many choices for $w(\theta)$ when $\theta \in SL^{(n)}$. $\square$

To date all the examples we have of probability functions satisfying FVP actually also have theories which satisfy GQE. For example in the unary case we see that we can take the ζ to be the possible disjuncts in (1). Other examples are the t-heterogeneous and homogeneous probability functions (see [9]) and the probability functions ${}^{\circ}w^{\Psi}$ (see [19, page 184]) when Ψ is standard.

For later use we observe that for L a not purely unary language there are probability functions on SL which fail even FVP_0. For example, in the notation of [19, page 217] let

$$\bar{p}_n = \langle 0, n^{-1}, n^{-1}, \ldots, n^{-1}, 0, 0, \ldots \rangle$$

for $n \in \mathbb{N}^+$ where there are n copies of n^{-1}. Then the probability function

$$w = \sum_{n=1}^{\infty} 2^{-n} v^{\bar{p}_n, L} \tag{6}$$

satisfies Ex. However for each $m \in \mathbb{N}^+$ there is a sentence $\theta_m \in SL^{(0)}$ such that

$$v^{\bar{p}_n, L}(\theta_m) = \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n, \end{cases}$$

so $w(\theta_m) = 2^{-m}$. (With apologies for the terseness of this example we refer the reader to [9].)

This observation may suggest that the converse to Theorem 3 holds, that if w satisfies FVP then $Th(w)$ satisfies GQE. We now show that this does indeed hold in cases where the n-ions of w for some large enough n have a certain property.

Theorem 4. *Let w satisfy FVP and suppose that there is some k such that if*

$$\{\zeta_1(a_1, a_2, \ldots, a_k), \ldots, \zeta_m(a_1, a_2, \ldots, a_k)\}$$

is a set of k-ions for w then for $n \geq k$ the set of Boolean combinations $\psi(a_1, a_2, \ldots, a_n)$ of sentences $\zeta_i(b_1, b_2, \ldots, b_k)$, where $\{b_1, b_2, \ldots, b_k\} \subseteq \{a_1, a_2, \ldots, a_n\}$ and $i \in \{1, 2, \ldots, m\}$, includes a set of n-ions for w. Then $Th(w)$ satisfies GQE.

Proof. Let $\rho(x_1, x_2)$ be the formula

$$\bigwedge_{i=1}^{q} \bigwedge_{f=1}^{r_i} \forall z_1, \ldots, z_{f-1}, z_{f+1}, \ldots, z_{r_i}$$
$$(R_i(z_1, \ldots, z_{f-1}, x_1, z_{f+1}, \ldots, z_{r_i}) \leftrightarrow R_i(z_1, \ldots, z_{f-1}, x_2, z_{f+1}, \ldots, z_{r_i}))$$

which expresses that x_1 and x_2 are indistinguishable from each other as far as the relations $R_1, \ldots, R_q$ of L are concerned.

The formula $\rho(x_1, x_2)$ clearly acts like equality in that it satisfies the axioms of equality (see for example [16]), in particular satisfying that for each $i = 1, 2, \ldots, q$,

$$\models \left(\bigwedge_{f=1}^{r_i} \rho(x_f, x_{r_i+f}) \right) \to (R_i(x_1, x_2, \ldots, x_{r_i}) \leftrightarrow R_i(x_{r_i+1}, x_{r_i+2}, \ldots, x_{2r_i})).$$

Consequently we also have that for any formula $\phi(x_1, x_2, \ldots, x_n)$ of $L^{(0)}$ (or even L) that

$$\models \left(\bigwedge_{f=1}^{n} \rho(x_f, x_{n+f}) \right) \to (\phi(x_1, x_2, \ldots, x_n) \leftrightarrow \phi(x_{n+1}, x_{n+2}, \ldots, x_{2n})). \tag{7}$$

Let w be as described in the statement of the theorem. Let $\theta(x_1, x_2, \ldots, x_n)$ be a formula of $L^{(0)}$ and (without loss of generality) let $n \geq k$. We claim that there is a Boolean combination $\psi(x_1, x_2, \ldots, x_n)$ of the $\rho(x_i, x_j)$ for $1 \leq i, j \leq n$ and the $\zeta_i(y_1, y_2, \ldots, y_k)$, where $\{y_1, \ldots, y_k\} \subseteq \{x_1, \ldots, x_n\}$, such that

$$w(\forall x_1, \ldots, x_n \, (\theta(x_1, \ldots, x_n) \leftrightarrow \psi(x_1, \ldots, x_n))) = 1,$$

which clearly gives the required result.

Let $\mathcal{H} = \{H_1, H_2, \ldots, H_e\}$ be a partition of $\{1, 2, \ldots, n\}$ and let $\eta_{\mathcal{H}}(x_1, \ldots, x_n)$ be the formula

$$\bigwedge_{i=1}^{e} \left(\bigwedge_{s,t \in H_i} \rho(x_s, x_t) \wedge \bigwedge_{1 \leq i < j \leq e} \bigwedge_{\substack{s \in H_i \\ t \in H_j}} \neg \rho(x_s, x_t) \right).$$

By the assumption of the theorem there is a Boolean combination $\psi_{\mathcal{H}}(x_1, x_2, \ldots, x_n)$ of the $\zeta_i(y_1, y_2, \ldots, y_k)$, where $\{y_1, y_2, \ldots, y_k\} \subseteq \{x_1, x_2, \ldots, x_n\}$ and $i \in \{1, 2, \ldots, m\}$, such that

$$w(\psi_{\mathcal{H}}(a_1, \ldots, a_n) \leftrightarrow (\theta(a_1, \ldots, a_n) \wedge \eta_{\mathcal{H}}(a_1, \ldots, a_n))) = 1$$

and hence

$$w((\psi_{\mathcal{H}}(a_1, \ldots, a_n) \wedge \eta_{\mathcal{H}}(a_1, \ldots, a_n)) \leftrightarrow (\theta(a_1, \ldots, a_n) \wedge \eta_{\mathcal{H}}(a_1, \ldots, a_n))) = 1. \tag{8}$$

We can now refine (8) to give[3]

$$w((\psi_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}))\leftrightarrow(\theta(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})))=1 \tag{9}$$

for *any* $i_1,i_2,\ldots,i_n$, not necessarily distinct. To see this let $\mathcal{G}=\{G_1,G_2,\ldots,G_c\}$ be the partition of $\{1,2,\ldots,n\}$ such that k,j are in the same class just if $i_k=i_j$. If $\mathcal{G}$ is not a refinement of $\mathcal{H}$ then both of

$$\psi_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}) \text{ and } \theta(a_{i_1},\ldots,a_{i_n}))\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})$$

are inconsistent so get probability 0 and the required conclusion (9) holds.

On the other hand if $\mathcal{G}$ is a refinement of $\mathcal{H}$ then from the fact that ρ satisfies the axioms of equality (7) we have that

$$(\psi_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}))\leftrightarrow(\psi_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n})\wedge\eta_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n}))$$

gets probability 1 according to w, where h_t is the least i_j such that t,j are in the same equivalence class according to $\mathcal{H}$. Similarly we have that

$$(\theta(a_{i_1},\ldots,a_{i_n}))\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}))\leftrightarrow(\theta(a_{h_1},\ldots,a_{h_n})\wedge\eta_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n}))$$

gets probability 1 according to w.

Hence the sentence in (9) gets the same probability as

$$(\psi_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n})\wedge\eta_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n}))\leftrightarrow(\theta(a_{h_1},\ldots,a_{h_n})\wedge\eta_{\mathcal{H}}(a_{h_1},\ldots,a_{h_n})). \tag{10}$$

---

[3]One might have supposed that we could get a much less torturous proof here by noting that, by the assumption of the theorem, there is such a Boolean combination $\psi(a_1,\ldots,a_n)$ for which

$$w(\psi(a_1,\ldots,a_n)\leftrightarrow\theta(a_1,\ldots,a_n))=1$$

and then concluding that for any $i_1,i_2,\ldots,i_n$ we must also have

$$w(\psi(a_{i_1},\ldots,a_{i_n})\leftrightarrow\theta(a_{i_1},\ldots,a_{i_n}))=1.$$

Unfortunately this need not be the case. For example if L has just a single binary relation symbol R and w is the probability function defined by

$$w(R(a_i,a_j))=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{if } i\neq j,\end{cases}$$

then w satisfies Ex and for $\perp$ a contradiction, $w(R(a_1,a_2)\leftrightarrow\perp)=1$ but $w(R(a_1,a_1)\leftrightarrow\perp)=0$.

Applying the corresponding argument in the case of

$$\psi_{\mathcal{H}}(a_1,\ldots,a_n)\wedge\eta_{\mathcal{H}}(a_1,\ldots,a_n) \text{ and } \theta(a_1,\ldots,a_n)\wedge\eta_{\mathcal{H}}(a_1,\ldots,a_n),$$

and noticing that the the non-refinement situation cannot occur in this case, we see that

$$(\psi_{\mathcal{H}}(a_1,\ldots,a_n)\wedge\eta_{\mathcal{H}}(a_1,\ldots,a_n))\leftrightarrow(\theta(a_1,\ldots,a_n)\wedge\eta_{\mathcal{H}}(a_1,\ldots,a_n)) \tag{11}$$

has the same probability, i.e. 1 by (8), as

$$(\psi_{\mathcal{H}}(a_{s_1},\ldots,a_{s_n})\wedge\eta_{\mathcal{H}}(a_{s_1},\ldots,a_{s_n}))\leftrightarrow(\theta(a_{s_1},\ldots,a_{s_n})\wedge\eta_{\mathcal{H}}(a_{s_1},\ldots,a_{s_n})) \tag{12}$$

where s_t is the least j such that t,j are in the same equivalence class according to $\mathcal{H}$. But clearly the sentences in (12) and (10), and consequently also (9), must get the same probability by Ex, to wit probability 1.

To complete the proof notice that $\theta(a_{i_1},\ldots,a_{i_n}))$ is logically equivalent to

$$\bigvee_{\mathcal{H}}(\theta(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}))$$

where the disjunction is over all partitions $\mathcal{H}$ of $\{1,2,\ldots,n\}$ and therefore by (9) to

$$\bigvee_{\mathcal{H}}(\psi_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})).$$

Hence we now have that

$$\bigwedge_{i_1,\ldots,i_n\leq r}\left(\theta(a_{i_1},\ldots,a_{i_n})\ \leftrightarrow\ \bigvee_{\mathcal{H}}(\psi_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n})\wedge\eta_{\mathcal{H}}(a_{i_1},\ldots,a_{i_n}))\right)$$

gets probability 1 according to w. Taking the limit $r\to\infty$ (and using the standard result [19, Lemma 3.8]) now gives as required that

$$w(\forall x_1,\ldots,x_n\,(\theta(x_1,\ldots,x_n)\leftrightarrow\bigvee_{\mathcal{H}}(\eta_{\mathcal{H}}(x_1,\ldots,x_n)\wedge\psi_{\mathcal{H}}(x_1,\ldots,x_n))))=1,$$

the disjunction over $\mathcal{H}$ being simply a Boolean combination of the $\rho(x_i,x_j)$ for $1\leq i,j\leq n$ and the $\zeta_i(y_1,y_2,\ldots,y_k)$.

□

At this time we know of no example of a probability function w satisfying FVP which does not satisfy the requirement of Theorem 4. This might lead one to conjecture that FVP and GQE are essentially the same thing, a point we shall revisit later in the concluding section.

The Strong Finite Values Property

Given the suggested rationality of FVP it seems natural to take a further step in this direction and consider probability functions satisfying Ex for which even

$$\{w(\theta) \,|\, \theta \in SL\}$$

is finite. We say that such a probability function satisfies the *Strong Finite Values Property*, SFVP.

The plan now is to give a characterization (in fact several) of the probability functions satisfying SFVP. Before doing so however we need to introduce some notation.
A sentence $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n}) \in QFSL$ is a *state description*[4] for $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$ if it is (up to the order of conjuncts) of the form

$$\bigwedge_{j=1}^{q} \bigwedge_{b_1,\ldots,b_{r_j}} \pm R_j(b_1, \ldots, b_{r_j})$$

where the $b_1, \ldots, b_{r_j} \in \{a_{i_1}, a_{i_2}, \ldots, a_{i_n}\}$ and $\pm R_j(\vec{b})$ stands for one of $R_j(\vec{b})$, $\neg R_j(\vec{b})$. We shall adopt the usual notation that state descriptions are designated by upper case letters Θ, Φ, Ψ etc..

Let I_n be the set of state descriptions of L for $a_1, \ldots, a_n$ which are invariant up to logical equivalence under any permutation of $a_1, \ldots, a_n$. That is:

$$I_n = \{\Phi(a_1, \ldots, a_n) \mid \Phi(a_1, \ldots, a_n) \equiv \Phi(a_{\sigma(1)}, \ldots, a_{\sigma(n)}) \quad \forall \sigma \in \mathsf{S}_n\}, \tag{13}$$

where S_n is the set of permutations of $\{1, 2, \ldots, n\}$ (and where logically equivalent members are identified). Let r be at least as large as the arity of any relation in L, i.e. $r \geq \max\{r_j \,|\, j = 1, 2, \ldots, q\}$.

For $\Theta(a_1, \ldots, a_r) \in I_r$ we define a unique structure $M_\Theta \in \mathcal{T}L$ as follows: For $j = 1, 2, \ldots, q$ and not necessarily distinct $i_1, i_2, \ldots, i_{r_j} \in \mathbb{N}^+$ set

$$M \models R_j(a_{i_1}, a_{i_2}, \ldots, a_{i_{r_j}}) \iff \Theta(a_1, \ldots, a_r) \models R_j(a_{\tau(i_1)}, a_{\tau(i_2)}, \ldots, a_{\tau(i_{r_j})})$$

where $\tau(i_t)$ is the least s such that $i_t = i_s$.

[4]State descriptions are important in this subject because by a result of Gaifman [7] every probability function on SL is already determined by its values on the state descriptions.

In this case, referring to $\langle \tau(i_1), \tau(i_2), \ldots, \tau(i_{r_j}) \rangle$ as the collapse of $\langle i_1, i_2, \ldots, i_{r_j} \rangle$, notice that if σ is a permutation of $\mathbb{N}$ and then $\langle \sigma(i_1), \sigma(i_2), \ldots, \sigma(i_{r_j}) \rangle$ has the identical collapse, and furthermore all its coordinates are in $\{1, 2, \ldots, r\}$. From this it follows that $M_\Theta \models \Theta$, for each $n \in \mathbb{N}^+$ M_Θ is a model of

$$\bigvee_{\Phi \in I_n} \Phi(a_1, \ldots, a_n) \tag{14}$$

for each $n \in \mathbb{N}^+$, and furthermore M_Θ is the unique structure in $\mathcal{T}L$ with these two properties. Hence

$$\{M \in \mathcal{T}L \,|\, M \models \bigvee_{\Phi \in I_n} \Phi(a_1, \ldots, a_n), \quad \forall n \in \mathbb{N}^+\} = \{M_\Theta \,|\, \Theta \in I_r\}. \tag{15}$$

Now define the probability function V_{M_Θ} on SL by:

$$V_{M_\Theta}(\phi) = \begin{cases} 1 & \text{if } M_\Theta \models \phi, \\ 0 & \text{if } M_\Theta \models \neg\phi. \end{cases}$$

Note that by the construction of M_Θ, V_{M_Θ} satisfies Ex.

The following theorem characterizes those probability functions on SL satisfying SFVP (and Ex).

Theorem 5. *If w is a probability function on SL then the following statements are equivalent:*

1. *w satisfies SFVP.*
2. *$w\left(\bigvee_{\Phi(a_1,\ldots,a_n) \in I_n} \Phi(a_1, \ldots, a_n)\right) = 1$ for each $n \in \mathbb{N}^+$.*
3. *w is a convex sum of the functions V_{M_Θ} for $\Theta \in I_r$.*
4. *For every $n \in \mathbb{N}$, $\theta \in SL^{(n)}$ and $\sigma \in \mathsf{S}_n$,*

$$w(\theta(a_1, \ldots, a_n) \longleftrightarrow \theta(a_{\sigma(1)}, \ldots, a_{\sigma(n)})) = 1.$$

Since the proof of this result (as we know it) requires a digression into Nonstandard Analysis we shall refer the curious reader to [8, page 112].

Theorem 5 clearly shows that SFVP puts very strong constraints on a probability function, too strong in our view to be considered as a desirable simplicity condition.

FVP and Super Regularity

As far as practical applications are concerned there is a major question we have essentially ignored up to now, namely given a satisfiable $\theta \in SL^{(0)}$ is there a probability function w satisfying Ex + FVP and such that $w(\theta) = 1$? Equivalently given a satisfiable finite set of linear constraints

$$\sum_{i=1}^{n} \beta_{i,j} w(\theta_i) = 0, \quad j = 1, 2, \ldots, m$$

with the $\theta_j \in SL^{(0)}$ is there necessarily a probability function w both satisfying these constraints, FVP and Ex? Unfortunately at present we do not know the answer to this question though we would conjecture that it is yes.

Of course we would obtain an immediate affirmative answer to this conjecture if we could find a probability function satisfying Ex + FVP + SReg, where SReg, standing for *Super Regularity*, is the requirement that $w(\theta) > 0$ for all satisfiable $\theta \in SL$. Unfortunately this is not possible when L is not purely unary as the following theorem shows:

Theorem 6. *If L is not purely unary and w is a probability function on SL satisfying FVP_n for some $n \in \mathbb{N}$, then w does not satisfy SReg.*

Proof. Suppose L and w are as described. Then by Theorem 2, there is some set of n-ions for w

$$B = \{\phi_1, \ldots, \phi_g\} \subset SL^{(n)}$$

such that $w(\phi_i \wedge \phi_j) = 0$ for $1 \leq i < j \leq g$,

$$\sum_{i=1}^{g} w(\phi_i) = 1,$$

and for every $\theta \in SL^{(n)}$ there is some $B_\theta \subseteq B$ such that

$$w\left(\theta \leftrightarrow \bigvee_{\phi_i \in B_\theta} \phi_i\right) = 1.$$

Suppose that

$$\models \bigvee_{i=1}^{g} \phi_i$$

and for each $\theta \in SL^{(n)}$

$$\models \theta \leftrightarrow \bigvee_{\phi_i \in B_\theta} \phi_i.$$

Then by Theorem 2, every probability function on SL would satisfy FVP_n with n-ions B, contradicting our earlier observation that there are probability functions on SL, for example (6), for which FVP_0 (and hence FVP_n) fails. Therefore, either $\neg \bigvee_{i=1}^{g} \phi_i$ is consistent, but assigned probability zero by w, or for some $\theta \in SL^{(n)}$, $\neg \left(\theta \leftrightarrow \bigvee_{\phi_i \in B_\theta} \phi_i\right)$ is consistent, but assigned probability zero by w. In either case, w fails to satisfy SReg. $\square$

FVP and FVP_n

Given our results so far a natural question to ask is whether FVP_n might actually imply FVP for sufficiently large n. The *simple* answer to this question is, perhaps unsurprisingly, no. We give here an outline to show that FVP_1 does not imply FVP (or even FVP_2). Similar examples can, with some effort, be constructed in general to show that FVP_n does not imply FVP_{n+1}, see [10].

Let L be the language with a single binary relation symbol R. Let M be the structure for L with universe $\mathbb{Z}$, R interpreted as immediate successor and for $n \in \mathbb{N}^+$ let a_n^M, the interpretation of a_n in M, be $n/2$ if n is even and $(1-n)/2$ if n is odd. Notice that $M \in \mathcal{T}L$.

For any $i, j \in \mathbb{Z}$ there is an isomorphism of M sending i to j so for $\theta(a_1) \in SL^{(1)}$,

$$M \models \theta(a_i) \iff M \models \theta(a_j). \tag{16}$$

Hence the a_i for which $\theta(a_i)$ holds in M is either all of them or none of them. This similarity however breaks down when we allow sentences from $SL^{(2)}$ since for $m \in \mathbb{N}$ we can clearly write down provably disjoint sentences $\psi_m(a_i, a_j)$ such that

$$M \models \psi_m(a_i, a_j) \iff |a_i^M - a_j^M| = m. \tag{17}$$

Now define a probability function V_M on SL by

$$V_M(\theta(a_1, a_2, \ldots, a_n)) = \begin{cases} 1 & \text{if } M \models \theta(a_1, a_2, \ldots, a_n), \\ 0 & \text{otherwise,} \end{cases}$$

and in turn a further function w on SL by

$$w(\theta(a_1, a_2, \ldots, a_n)) = \sum_{\tau} V_M(\theta(a_{\tau(1)}, a_{\tau(2)}, \ldots, a_{\tau(n)})) \cdot \prod_{i=1}^{n} 2^{-\tau(i)}$$

where τ runs over all maps from $\{1, 2, \ldots, n\}$ into $\mathbb{N}^+$. By a theorem of Gaifman, see [7] or [19, Chapter 26], w is a probability function on SL satisfying Ex.

By (16) all the $V_M(\theta(a_i))$ are 0 or all are 1 so $w(\theta(a_1))$ is either 0 or 1 and w satisfies FVP_1. However for $m \in \mathbb{N}$ the $w(\psi_m(a_1, a_2))$ are clearly non-zero and have sum at most 1 (because they are provably disjoint) so w must fail FVP_2.

We referred to this construction and the supplement at [10] as giving a 'simple answer' to the question of whether FVP_n alone implies FVP. The reason for this qualifier is that in these examples as the n increases so does the largest arity of the relation symbols used in the language. As far as we currently know it is possible that if w satisfies FVP_n on a language L with no relation symbols of arity greater than n then w must also satisfy FVP.

We would conjecture that this is indeed the case, and even more that once we have FVP_n at this largest arity level then the j-ions beyond that are as described earlier just Boolean combinations of the n-ions.

Conclusion

Theorem 2 shows that if a function w satisfies FVP_n, its n-ions correspond to various 'possible worlds', in each of which w is able to 'decide' every $\theta \in SL^{(n)}$, so that the probability it assigns to any such θ is the sum of the probabilities assigned to those worlds where θ is decided positively. This demonstrates that there is an underlying simplicity to those functions which satisfy FVP, beyond the superficial simplicity evident in its definition, in that it entails a rather 'neat', and arguably natural, way of assigning probabilities.

Simplicity, as a feature of probability functions used to model rational belief, was endorsed by Kemeny in [11], and considered by Paris & Vencovská in [20], but seems otherwise to have received little attention in Inductive Logic. Kemeny is not explicit about what constitutes simplicity, and the notion discussed by Paris & Vencovská is rather different from that considered here in relation to FVP. With these and likely other different ideas of simplicity available it would be reckless to claim without qualification that simplicity is always a desirable feature of probability functions, in fact we reach the opposite conclusion in the case of the Strong Finite Values Property. However, the particular simplicity

entailed by FVP and interpreted above in terms of systematic reasoning about 'possible worlds', seems to be an appealing and arguably a rational feature.

In the course of this paper we have made two somewhat rash conjectures:

- If $\theta \in SL$ is consistent then there is a probability function w satisfying FVP for which $w(\theta) = 1$.
- If w satisfies FVP_n when n is the largest arity of any relation symbol in the language L then w satisfies the FVP and furthermore the j-ions for $j > n$ are just Boolean combinations of the n-ions.

Clearly a positive answer to these conjectures would considerably strengthen the structural importance of FVP. In particular as shown in Theorem 4 a positive answer to this second bullet point would equate FVP with GQE.

References

[1] Beierle, C., Finthammer, M. & Kern-Isberner, G., Relational Probabilistic Conditionals and Their Instantiations under Maximum Entropy Semantics for First-Order Knowledge Bases, *Entropy*, 2015, **17**(2):852-865.

[2] Carnap, R., On Inductive Logic, *Philosophy of Science*, 1945, **12**(2):72-97.

[3] Carnap, R., On the Application of Inductive Logic, *Philosophy and Phenomenology Research*, 1947, **8**:133-147.

[4] Carnap, R., *Logical Foundations of Probability*, University of Chicago Press, Chicago, Routledge & Kegan Paul, London, 1950.

[5] Carnap, R., *The Continuum of Inductive Methods,* University of Chicago Press, 1952.

[6] Carnap, R., The Aim of Inductive Logic, in *Logic, Methodology and Philosophy of Science,* eds. E.Nagel, P.Suppes & A.Tarski, Stanford University Press, Stanford, California, 1962, pp303-318.

[7] Gaifman, H., Concerning Measures on First Order Calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.

[8] Howarth, E., *New Rationality Principles in Pure Inductive Logic*, Ph.D. Thesis, Manchester University, June 2015. Available at http://www.maths.manchester.ac.uk/∼jeff/theses/lwthesis.pdf

[9] Howarth, E. & Paris, J.B., The Theory of Spectrum Exchangeability, *Review of Symbolic Logic,* **8**(01):108-130, 2015.

[10] Howarth, E. & Paris, J.B., A proof that FVP_n does not imply FVP_{n+1}. Available at http://www.maths.manchester.ac.uk/∼jeff/papers/lw150729GKIsup.pdf

[11] Kemeny, J.G., Carnap's Theory of Probability and Induction, in *The Philosophy of Rudolf Carnap*, ed. P.A.Schilpp, La Salle, Illinois, Open Court, 1963, pp711-738.

[12] Kern-Isberner, G., Conditionals in Nonmonotonic Reasoning and Belief Revision - Considering Conditionals as Agents, in *Lecture Notes in Computer Science, 2087*, Springer, 2001, ISBN 3-540-42367-2

[13] Kern-Isberner, G. & Lukasiewicz, T., Combining probabilistic logic programming with the power of maximum entropy, *Artificial Intelligence*, 2004, **157**(1-2):139-202.

[14] Kern-Isberner, G. & Thimm, M., Novel Semantical Approaches to Relational Probabilistic Conditionals, in the *Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR'10)*, Eds. F.Lin, U.Sattler, M.Truszczynski, AAAI Press, Toronto, Canada, May 2010, pp382-392.

[15] Landes, J. & Williamson, J., Justifying Objective Bayesianism on Predicate Languages, *Entropy [Online]*, 2015, **17**:2459-2543.

[16] Paris, J.B., *A Short Course in Predicate Logic.* 2015, ISBN 978-87-403-0795-5. Available online at bookboon.com.

[17] Paris, J.B., Pure Inductive Logic, in *The Continuum Companion to Philosophical Logic*, eds. L.Horsten & R.Pettigrew, Continuum International Publishing Group, London, 2011, pp428-449.

[18] Paris, J.B., *Pure Inductive Logic: Workshop Notes for Progic 2015*. URL= <http://www.maths.manchester.ac.uk/∼jeff/lecture-notes/Progic15.pdf>

[19] Paris, J.B. & Vencovská, A., *Pure Inductive Logic*, in the Association of Symbolic Logic Perspectives in Mathematical Logic Series, Cambridge University Press, April 2015.

[20] Paris, J.B. & Vencovská, A., The Twin Continua of Inductive Methods, in *Logic Without Borders*, eds. A.Hirvonen, J.Kontinen, R.Kossak, & A.Villaveces, Ontos Mathematical Logic, Frankfurt-Heusenstamm, Germany, 2015, pp355-366.

[21] Paris, J.B. & Rad, S.R., Inference Processes for Quantified Predicate Knowledge, in *Logic, Language, Information and Computation,* WoLLIC, Edinburgh, 2008, eds. W.Hodges & R.deQueiroz, Springer LNAI 5110, pp249-259.

[22] Paris, J.B. & Rad, S.R., A note on the least informative model of a theory, in *Programs, Proofs, Processes, CiE 2010,* eds. F.Ferreira, B.Löwe, E.Mayordomo, & L.Mendes Gomes, Springer LNCS 6158, pp342-351.

[23] Williamson, J., *In Defence of Objective Bayesianism*, Oxford University Press, 2010.