Second Order Inductive Logic and Wilmers' Principle

M.S.Kließ* and J.B.Paris†
School of Mathematics
The University of Manchester
Manchester M13 9PL
tel. 0161-275-5880

malte.kliess@postgrad.manchester.ac.uk
jeff.paris@manchester.ac.uk

March 10, 2014

We extend the framework of Inductive Logic to Second Order languages and introduce Wilmers' Principle, a rational principle for probability functions on Second Order languages. We derive a representation theorem for functions satisfying this principle and investigate its relationship to the first order principles of Regularity and Super Regularity.

Key words: Universal Certainty, Second Order Logic, Inductive Logic, Probability Logic, Uncertain Reasoning.

1 Introduction

In the framework of Pure Inductive Logic, a rational agent's belief function is usually regarded as a probability function on the set of first order sentences of a certain, fixed language L. This language contains constants representing the objects of the universe

*Supported by a University of Manchester School of Mathematics Research Studentship.
†Corresponding Author.

and predicates representing the properties of these objects. This allows an agent to express statements about the universe.

So far such statements have, to our knowledge, been limited to first order expressions, allowing the agent to make existential or universal statements about the objects. As the famous Geach-Kaplan statement shows (see e.g. [1]), an agent could increase her expressive power if she were to extend the set of expressions available to include second order statements.

As we identify an agent's belief in a statement with the value the agent's belief function assigns to it, allowing the agent to use second order expressions will require her to extend the domain of the her belief function to include second order sentences.

Unsurprisingly, this leads to a number of complications. Since Second Order logic is inherently incomplete (see e.g. [10]), we will have to be careful picking a suitable framework for Second Order Inductive Logic. At the same time, we would want to have a suitable interpretation of universal and existential quantification over the predicates in L.

In this paper we intend to provide such a framework, allowing an agent to extend her expressive power to second order logic. Once such a framework is given, we can study rational principles involving second order logic. We will give an example of one such principle, called Wilmers' Principle, and provide a representation theorem for second order belief functions that satisfy this principle. We will then consider the consequences of this principle for the thorny question of Universal Truth for first order statements.

2 Second Order probability functions

As with traditional Inductive Logic[1] for example [2], we will work in a unary language L with predicate symbols P_i and constants a_i for $i \in \{1, 2, 3, \ldots\} = \mathbb{N}^+$ but without function symbols or equality. Let $F_1 L, S_1 L, QFS_1 L$ respectively denote the first order formulae, sentences and quantifier free sentences of L.

Let $\mathcal{T}L$ denote the set of structures M for L in which the constants a_i are interpreted as themselves and $|M| = \{a_i \,|\, i \in \mathbb{N}^+\}$, so every element of the universe of M is denoted by a constant symbol. Similarly we shall use P_j to denote $\{a_i \,|\, M \models P_j(a_i)\}$, leaving the M implicit whenever this cannot cause confusion.

We say that $w : S_1 L \to [0, 1]$ is a *probability function on* $S_1 L$, if for any $\vartheta, \varphi \in S_1 L$, $\psi(x) \in F_1 L$,

[1] Actually Inductive Logic is more commonly presented with only finitely many predicate symbols but as we would in any case advocate the rationality of Unary Language Invariance in this context, see for example [8, Chapter 16], this would ultimately lead to the same situation.

(P1) If $\models \vartheta$, then $w(\vartheta) = 1$.

(P2) If $\vartheta \models \neg\varphi$, then $w(\vartheta \vee \varphi) = w(\vartheta) + w(\varphi)$.

(P3) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} \psi(a_i)\right)$.

To our mind the central problem of (Pure) Inductive Logic can be picturesquely captured as follows: Imagine an agent inhabiting a structure $M \in \mathcal{T}L$ but having no further knowledge, so in particular the agent has no particular interpretation in mind for the constant and predicate symbols. In that case what probability $w(\vartheta)$ should the agent rationally, or logically, give to $\vartheta \in S_1 L$? Or more precisely, since we obviously intend for these probability values to be coherent, what probability function w should the agent rationally or logically adopt?

In the absence of any clear definition of what is meant here by 'rationally' (which for the purpose of this paper we identify with 'logically') the usual method of tackling this question is by imposing certain ostensibly rational, or at least not irrational, requirements on w and seeing where that leads. For example the symmetry between the constants a_i, and between the predicates P_j, from the agent's point of view surely requires that w should satisfy:

Constant Exchangeability, Ex
w satisfies Ex, if for all $\vartheta \in S_1 L$ and all permutations σ of $\mathbb{N}^+$,

$$w(\vartheta(a_1, \dots, a_n)) = w(\vartheta(a_{\sigma(1)}, \dots, a_{\sigma(n)})).$$

Predicate Exchangeability, Px
w satisfies Px, if for all $\vartheta \in S_1 L$, and all permutations σ of $\mathbb{N}^+$,

$$w\left(\vartheta(P_1, \dots, P_m, a_1, \dots, a_n)\right) = w\left(\vartheta(P_{\sigma(1)}, \dots, P_{\sigma(m)}, a_1, \dots, a_n)\right)$$

whenever $\vartheta(P_{\sigma(1)}, \dots, P_{\sigma(m)}, a_1, \dots, a_n)$ is the result of replacing each predicate symbol, P_j, occurring in ϑ with $P_{\sigma(j)}$.

Given their standing we shall assume throughout that unless specifically indicated all the probability functions we consider satisfy Ex and Px.

Two further principles which we might impose on w are based on the idea that in the presence of no prior knowledge about the universe it would might seem unreasonable to summarily assign zero probability to sentences which could be true.

Regularity, Reg
w satisfies Reg, if for all $\vartheta \in QFS_1 L$ such that ϑ is satisfiable, $w(\vartheta) > 0$.

Super Regularity, SReg
w satisfies SReg, if for all $\vartheta \in S_1 L$ such that ϑ is satisfiable, $w(\vartheta) > 0$.

Note that we have separate principles for quantifier-free sentences and sentences that may contain quantifiers. The reason for this is that functions satisfying Regularity in general do not need to also satisfy Super Regularity.

The principle SReg[2] has been something of a thorn in the side of traditional Inductive Logic. Whilst seemingly quite reasonable it is inconsistent with several other rationally attractive principles such as Johnson's Sufficientness Postulate, see [6], [8], and in consequence is not satisfied by the probability functions comprising Carnap's Continuum of Inductive Methods, though all of them except the pathological c_0 do satisfy Reg. Indeed SReg already fails there for consistent but non-tautologous Π_1 sentences where the failure is sometimes referred to as the problem of *Universal Certainty.* In spite of some attempts, see for example [5], to tweak Johnson's Sufficientness Postulate and allow SReg to hold the issue still seems to us problematic. In this paper we shall be investigating a somewhat different approach which under certain circumstance has SReg as a consequence.

Returning again to our agent it is clear that for this unary language L there is a certain parallel between the constant symbols and the predicate symbols, as indeed is reflected in the principles Ex and Px. This suggests that from the agent's point of view we could just as well have quantifiers ranging over the predicates as over the constants, in other words *second order* quantifiers $\forall X$ and $\exists X$ where the second order variable X is intended to range over the predicates P_i.

To this end let F_2L, S_2L respectively denote the second order formulae and sentences of L and say that $w : S_2L \to [0,1]$ is a *probability function on* S_2L, if for any $\vartheta, \varphi \in S_2L$, $\psi(x), \eta(X) \in F_2L$,

(P1) If $\models \vartheta$, then $w(\vartheta) = 1$.

(P2) If $\vartheta \models \neg\varphi$, then $w(\vartheta \vee \varphi) = w(\vartheta) + w(\varphi)$.

(P3) $w(\exists x\, \psi(x)) = \lim_{n\to\infty} w\left(\bigvee_{i=1}^n \psi(a_i)\right)$.

(P4) $w(\exists X\, \eta(X)) = \lim_{n\to\infty} w\left(\bigvee_{i=1}^n \eta(P_i)\right)$.

So here we have added (P4) to the original requirements, reflecting the idea that the P_i exhaust the second order universe just as (P3) reflected the idea that the a_i exhaust the first order universe.

This extension from first to second order in this way is in fact hardly more than notational (though it will enable us to state and investigate new principles). For example if $M \in \mathcal{T}L$ then M automatically becomes a second order structure by taking the second order

[2]Also referred to as *Cournot's Principle*, see [3].

universe of M to simply be the set of interpretations of the relation symbols in M, i.e.

$$\{\{a_i \,|\, M \models P_j(a_i)\} \,|\, j \in \mathbb{N}^+\}.$$

For this reason we shall not distinguish between M as a first or second order structure and similarly we shall use $\mathcal{T}L$ for 'both' versions.

Notice then that the condition $\vartheta \models \varphi$ in (P1-2) is actually equivalent to

$$\forall M \in \mathcal{T}L, \text{ if } M \models \vartheta \text{ then } M \models \varphi$$

since if a sentence $\psi \in S_2L$ has a model then it has a model with denumerably many elements and subsets in its universe and hence by suitable naming a model in $\mathcal{T}L$.

With this extended notion of probability function all the standard properties can be proved just as before (e.g. see [7, Proposition 2.1], [4] or [8, Lemma 3.8]):

Lemma 1. *Let w be a probability function on S_2L. Then for $\vartheta, \varphi \in S_2L$,*

(a) $w(\neg\vartheta) = 1 - w(\vartheta)$.

(b) If $\models \neg\vartheta$*, then* $w(\vartheta) = 0$.

(c) If $\vartheta \models \varphi$*, then* $w(\vartheta) \leq w(\varphi)$.

(d) If $\vartheta \equiv \varphi$*, then* $w(\vartheta) = w(\varphi)$.

(e) $w(\vartheta \vee \varphi) = w(\vartheta) + w(\varphi) - w(\vartheta \wedge \varphi)$.

Lemma 2. *Let w be a probability function on S_2L and $\exists x_1 \ldots \exists x_k\, \vartheta(x_1, \ldots, x_k, \vec{P}, \vec{a})$, $\exists X_1 \ldots \exists X_k\, \psi(X_1, \ldots, X_k, \vec{P}, \vec{a}) \in S_2L$. Then*

$$\begin{aligned}
w(\exists x_1 \ldots \exists x_k\, \vartheta(x_1, \ldots, x_k, \vec{P}, \vec{a})) &= \lim_{n\to\infty} w\left(\bigvee_{i_1,\ldots,i_k \leq n} \vartheta(a_{i_1}, \ldots, a_{i_k}, \vec{P}, \vec{a})\right) \\
w(\forall x_1 \ldots \forall x_k\, \vartheta(x_1, \ldots, x_k, \vec{P}, \vec{a})) &= \lim_{n\to\infty} w\left(\bigwedge_{i_1,\ldots,i_k \leq n} \vartheta(a_{i_1}, \ldots, a_{i_k}, \vec{P}, \vec{a})\right) \\
w(\exists X_1 \ldots \exists X_k\, \psi(X_1, \ldots, X_k, \vec{P}, \vec{a})) &= \lim_{n\to\infty} w\left(\bigvee_{i_1,\ldots,i_k \leq n} \psi(P_{i_1}, \ldots, P_{i_k}, \vec{P}, \vec{a})\right) \\
w(\forall X_1 \ldots \forall X_k\, \psi(X_1, \ldots, X_k, \vec{P}, \vec{a})) &= \lim_{n\to\infty} w\left(\bigwedge_{i_1,\ldots,i_k \leq n} \psi(P_{i_1}, \ldots, P_{i_k}, \vec{P}, \vec{a})\right) \quad (1)
\end{aligned}$$

Just as structures in $\mathcal{T}L$ extend naturally from S_1L to S_2L so do probability functions on S_1L extend uniquely to probability functions on S_2L. Precisely:

Theorem 3. *Let w be a probability function on S_1L. Then w extends uniquely to a probability function on S_2L. Furthermore if w satisfies Ex and Px on S_1L then they are preserved in this extension to S_2L.*

Proof Let w be a probability function on S_1L, let $\mathcal{A}$ be the subsets of $\mathcal{T}L$ of the form

$$\{M \in \mathcal{T}L \mid M \models \vartheta\}$$

for $\vartheta \in QFS_1L$. Define a finitely additive measure μ on $\mathcal{A}$ by

$$\mu\{M \in \mathcal{T}L \mid M \models \vartheta\} = w(\vartheta).$$

By a compactness argument if φ and ϑ_n are in QFS_1L for $n \in \mathbb{N}$ and

$$\{M \in \mathcal{T}L \mid M \models \varphi\} = \bigcap_{n \in \mathbb{N}} \{M \in \mathcal{T}L \mid M \models \vartheta_n\}$$

then for some k

$$\{M \in \mathcal{T}L \mid M \models \varphi\} = \bigcap_{n \leq k} \{M \in \mathcal{T}L \mid M \models \vartheta_n\}$$

and from this it follows that μ preserves all infs in $\mathcal{A}$.

By Carathéodory's Extension Theorem then μ extends uniquely to a countably additive measure on the σ-algebra generated by $\mathcal{A}$. Since

$$\{M \in \mathcal{T}L \mid M \models \exists x\, \vartheta(x)\} = \bigcup_{n \in \mathbb{N}^+} \{M \in \mathcal{T}L \mid M \models \vartheta(a_n)\},$$

$$\{M \in \mathcal{T}L \mid M \models \exists X\, \vartheta(X)\} = \bigcup_{n \in \mathbb{N}^+} \{M \in \mathcal{T}L \mid M \models \vartheta(P_n)\},$$

all the sets

$$\{M \in \mathcal{T}L \mid M \models \vartheta\}$$

for $\vartheta \in S_2L$ will be in this σ-algebra and if we define w^+ on S_2L by

$$w^+(\vartheta) = \mu\{M \in \mathcal{T}L \mid M \models \vartheta\}$$

then w^+ will satisfy (P1-4). Since w^+ agrees with w on QFS_1L, by a result of Gaifman, [4], (essentially by induction on the length of formulae using (1)) w^+ agrees with w on S_1L and hence provides the required extension of w to S_2L.

Furthermore it is now easy to see by induction on quantifier complexity that this is the only possible extension of w to S_2L and if w satisfies Ex and Px on QFS_1L then it also does so on S_2L ■

Given this result we shall in future not particularly distinguish between a probability function defined on S_1L and its extension to S_2L.

3 Wilmers' Principle

So far we have established some initial technical results, allowing us to work with second order sentences in the framework of Pure Inductive Logic. In this section we suggest and discuss a rational principle for Second Order expressions that rational agents may want to accept as defining their beliefs.

The motivation for the principle is the following idea: Suppose we have a First Order formula $\vartheta(x)$ with just one free variable. Then in the agent's ambient structure M $\vartheta(x)$ *defines the subset* of the universe

$$\{a_i \mid M \models \vartheta(a_i)\}.$$

A rational agent then might feel that there should be a *name* for this set in the language, and thus $\vartheta(x)$ defines not only a subset of the universe, but also a predicate of the language, i.e.

$$\{a_i \mid M \models \vartheta(a_i)\} = \{a_i \mid M \models P_j(a_i)\}$$

for some (unary) P_j in the agent's language.

The formal definition of this principle in terms of probability functions is given by:

Wilmers' Principle, WP

Let w be a probability function on S_2L. Then w satisfies Wilmers' Principle, WP, *if*

$$w(\exists X\, \forall x\, (\vartheta(x) \leftrightarrow X(x))) = 1$$

whenever $\vartheta(x) \in F_1L$.[3]

This principle is based on the original suggestion by George Wilmers that it would be rational that w gave probability 1 to all the tautologies of second order monadic logic. The current version then is but a fragment of what was originally intended.

The main aim of this section is to provide a certain representation theorem for the probability functions satisfying Wilmers' Principle. This will turn out to be useful in the next section when we consider also Reg and SReg. We first need some notation and apparatus.

For each $n \in \mathbb{N}^+$, define $\wp(n)^\infty \times n^\infty$ and structures $M_{f,g}$ as follows:

- Let $\wp(n)$ denote the power set of $\{1, 2, \ldots, n\}$, let $\wp(n)^\infty$ denote the set of functions $f : \mathbb{N}^+ \to \wp(n)$ and let n^∞ denote the set of function $g : \mathbb{N}^+ \to \{1, 2, \ldots, n\}$. So $\wp(n)^\infty \times n^\infty$ denotes the set of all pairs $\langle f, g\rangle$ with $f \in \wp(n)^\infty$ and $g \in n^\infty$.

[3]As far as the actual statements of results in this paper are concerned it would make no difference if we took instead the version of Wilmers' Principle with $\vartheta(x) \in F_2L$.

- For each pair $\langle f, g\rangle \in \wp(n)^\infty \times n^\infty$, define the structure $M_{f,g}$ for L with finite universe $\{e_1, e_2, \ldots, e_n\}$ by $a_i^{M_{f,g}} = e_{g(i)}$ and $P_j^{M_{f,g}} = \{e_k \mid k \in f(j)\}$. So

$$M_{f,g} \models P_j(a_i) \iff M_{f,g} \models P_j(e_{g(i)}) \iff g(i) \in f(j).$$

[Here $a_i^{M_{f,g}}$ and $P_j^{M_{f,g}}$ are the interpretations of a_i and P_j in $M_{f,g}$.]

Let $n \in \mathbb{N}^+$ and let μ_n be a normalized, σ-additive measure on $\wp(n)^\infty \times n^\infty$. We say that

- μ_n is *invariant under Ex* if for any $\vartheta \in S_2L$ and any permutation $\sigma : \mathbb{N}^+ \to \mathbb{N}^+$,

$$\mu_n\{\langle f, g\rangle \mid M_{f,g} \models \vartheta\}) = \mu_n(\{\langle f, g\sigma\rangle \mid M_{f,g} \models \vartheta\},$$

- μ_n is *invariant under Px* if for any $\vartheta \in S_2L$ and any permutation $\sigma : \mathbb{N}^+ \to \mathbb{N}^+$,

$$\mu_n\{\langle f, g\rangle \mid M_{f,g} \models \vartheta)\}) = \mu_n(\{\langle f\sigma, g\rangle \mid M_{f,g} \models \vartheta\}.$$

We will use the measures μ_n to construct probability functions that only give weight to those structures $M_{f,g}$ that instantiate precisely n distinguishable constants. For this, it will be convenient to have a single second order sentence expressing this. Let ξ_n be the sentence

$$\exists x_1 \ldots \exists x_n \left[\bigwedge_{1\le i<j\le n} \exists X \neg(X(x_i) \leftrightarrow X(x_j)) \wedge \forall z \bigvee_{i=1}^{n} \forall X \, (X(x_i) \leftrightarrow X(z)) \right] \tag{2}$$

for $n \in \mathbb{N}^+$.

Theorem 4. *Let μ be a normalized, σ-additive measure on $\wp(n)^\infty \times n^\infty$ invariant under Ex and Px and suppose that*

$$\mu\{\langle f, g\rangle \mid \mathrm{ran}(f) = \wp(n) \text{ and } \mathrm{ran}(g) = n\} = 1. \tag{3}$$

Let w_μ be the function on S_2L defined by

$$w_\mu(\vartheta) = \mu\{\langle f, g\rangle \mid M_{f,g} \models \vartheta\}$$

for $\vartheta \in S_2L$ Then

(i) w_μ is a probability function on S_2L satisfying Px + Ex.

(ii) w_μ satisfies Wilmers' Principle, WP, and $w_\mu(\xi_n) = 1$.

Conversely let v be a probability function on S_2L satisfying Px + Ex + Wilmers' Principle and suppose that $v(\xi_n) = 1$. Then $v = w_\mu$ for some normalized, σ-additive measure μ on $\wp(n)^\infty \times n^\infty$ invariant under Ex + Px and satisfying (3).

Proof For (i) it is straightforward to check that w_μ satisfies (P1-2). For (P3) let $\psi(x) \in F_2L$ with x the only free variable. Then

$$\begin{aligned}
w_\mu(\exists x\, \psi(x)) &= \mu\{\langle f, g\rangle \mid M_{f,g} \models \exists x\, \psi(x)\} \\
&= \mu\left\{\bigcup_{k=1}^{n} \{\langle f, g\rangle \mid M_{f,g} \models \psi(e_k)\}\right\} \\
&= \mu\left\{\bigcup_{i=1}^{\infty} \{\langle f, g\rangle \mid M_{f,g} \models \psi(a_i)\}\right\}, \text{ since } g \text{ is onto for almost all } \langle f, g\rangle, \\
&= \lim_{m\to\infty} \mu\left\{\langle f, g\rangle \,\middle|\, M_{f,g} \models \bigvee_{i=1}^{m} \psi(a_i)\right\}, \text{ by } \sigma\text{-additivity of } \mu, \\
&= \lim_{m\to\infty} w_\mu\left(\bigvee_{i=1}^{m} \psi(a_i)\right),
\end{aligned}$$

and thus (P3) holds for w_μ. (P4) now follows using an analogous argument, but without the requirement of f being onto, for a sentence $\exists X\, \eta(X) \in S_2L$.

Since μ_n is invariant under Ex and Px, it follows that w_{μ_n} satisfies Ex + Px.

For (ii), we will first show that w_μ satisfies Wilmers' Principle. Let $\vartheta(x) \in F_2L$ with x the only free variable and let $\langle f, g\rangle \in \wp^\infty \times n^\infty$ with f onto. Then there is $j \in \mathbb{N}^+$ such that

$$\{e_k \mid M_{f,g} \models \vartheta(e_k)\} = \{e_k \mid k \in f(j)\} = \{e_k \mid M_{f,g} \models P_j(e_k)\}.$$

Hence,

$$M_{f,g} \models \forall x\, (P_j(x) \leftrightarrow \vartheta(x)),$$

and in turn,

$$M_{f,g} \models \exists X\, \forall x\, (X(x) \leftrightarrow \vartheta(x)).$$

By (3) then,

$$\mu\{\langle f, g\rangle \mid M_{f,g} \models \exists X\, \forall x\, (X(x) \leftrightarrow \vartheta(x))\} = 1$$

and hence

$$w_\mu(\exists X\, \forall x\, (X(x) \leftrightarrow \vartheta(x))) = 1.$$

An analogous proof shows that $w_\mu(\xi_n) = 1$.

For the converse let v be a probability function on S_2L with the properties as given. For $\varphi \in QFS_1L$ set

$$\mu\{\langle f, g\rangle \in \wp(n)^\infty \times n^\infty \mid M_{f,g} \models \varphi\} = v(\varphi). \tag{4}$$

Then μ is a finitely additive normalized measure on the algebra $\mathcal{A}$ of these subsets of $\wp(n)^\infty \times n^\infty$. Furthermore μ preserves infs, i.e. is a pre-measure. To see this suppose that $\varphi, \vartheta_n \in \mathit{QFS_1}L$ for $n \in \mathbb{N}$ and

$$\bigcap_{n\in\mathbb{N}} \{\langle f, g\rangle \,|\, M_{f,g} \models \vartheta_n\} = \{\langle f, g\rangle \,|\, M_{f,g} \models \varphi\}. \tag{5}$$

Suppose that there is no $k \in \mathbb{N}$ such that

$$\bigcap_{n\leq k} \{\langle f, g\rangle \,|\, M_{f,g} \models \vartheta_n\} = \{\langle f, g\rangle \,|\, M_{f,g} \models \varphi\}. \tag{6}$$

Then for each k we can find a structure M_{f_k,g_k} to satisfy

$$\{\vartheta_n \,|\, n \leq k\} \cup \{\neg\varphi\}.$$

Taking an ultraproduct of these structures with respect to a non-principle ultrafilter yields a structure M satisfying the set of sentences

$$\{\vartheta_n \,|\, n \in \mathbb{N}\} \cup \{\neg\varphi\} \tag{7}$$

whose universe is still the set of $e_k, k = 1, 2, \ldots, n$ and, when restricted to the original language, is of the form $M_{f,g}$ for some $\langle f, g\rangle \in \wp(n)^\infty \times n^\infty$. Furthermore since the $\vartheta_n, \varphi \in \mathit{QFS_1}L$ this $M_{f,g}$ also satisfies (7). But that contradicts (5), so such a k must exist and the preservation of this inf follows.

By Carathéodory's Extension Theorem we can now uniquely extend μ to the σ-algebra generated by $\mathcal{A}$. Since v satisfies Ex and Px then μ as defined by (4) will be invariant under Ex and Px and this property will be retained in this extension of μ.

We now need to show that

$$\mu\{\langle f, g\rangle \,|\, g \text{ is onto}\} = 1. \tag{8}$$

Since $v(\xi_n) = 1$ for some $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$

$$v\left(\bigwedge_{1\leq k<j\leq n} \exists X\, \neg(X(a_{i_k}) \leftrightarrow X(a_{i_j}))\right) > 0.$$

and for some $P_{q_{k,j}}$, $1 \leq k < j \leq n$,

$$v\left(\bigwedge_{1\leq k<j\leq n} \neg(P_{q_{k,j}}(a_{i_k}) \leftrightarrow P_{q_{k,j}}(a_{i_j}))\right) > 0.$$

By definition this left hand side is equal to

$$\mu\left\{\langle f, g\rangle \,\middle|\, M_{\langle f,g\rangle} \models \bigwedge_{1\leq k<j\leq n} \neg(P_{q_{k,j}}(a_{i_k}) \leftrightarrow P_{q_{k,j}}(a_{i_j}))\right\},$$

and similarly for disjunctions of such formulae $\bigwedge_{1\leq k<j\leq n} \neg(P_{q_{k,j}}(a_{i_k}) \leftrightarrow P_{q_{k,j}}(a_{i_j}))$. But since $v(\xi_n) = 1$, by (1), the limit probability of these disjunctions is 1, giving (8).

Hence by the first part w_μ is a probability function and by definition it agrees with v on QFS_1L. So by induction on the quantifier complexity of $\varphi \in SL2$,

$$v(\varphi) = w_\mu(\varphi) = \mu\{\langle f, g\rangle \mid M_{f,g} \models \varphi\}$$

It remains to show that μ satisfies (3). Notice that from (8) we are already half way there. Since $v(\mathrm{WP} \wedge \xi_n) = 1$,

$$\mu\{\langle f, g\rangle \mid M_{f,g} \models WP \wedge \xi_n\} = 1.$$

Let $\langle f, g\rangle$ be in this set with g onto. Then since $M_{\langle f,g\rangle} \models WP$ we have that for any P_i, P_j there are P_k, P_r, P_m such that

$$M_{\langle f,g\rangle} \models \forall x\, (P_k(x) \leftrightarrow \top)$$

$$M_{\langle f,g\rangle} \models \forall x\, (P_r(x) \leftrightarrow \neg P_i(x))$$

$$M_{\langle f,g\rangle} \models \forall x\, (P_m(x) \leftrightarrow (P_i(x) \wedge P_j(x)))$$

Hence the $P_j^{M_{\langle f,g\rangle}}$ form a Boolean Algebra. Also since $M_{f,g} \models \xi_n$ every pair $e_m \neq e_r$ are separated by some $P_j^{M_{\langle f,g\rangle}}$ so this Boolean Algebra must be all subsets of $\{e_1, e_2, \ldots, e_n\}$. I.e. f must be onto and (3) follows.

■

In the case $n = 1$ there is only one choice for $\langle f, g\rangle$ and similarly μ and in this case w_μ is c_0^L from Carnap's Continuum of Inductive Methods for L. For $n > 1$ there are infinitely many choices for μ invariant under Ex and Px. For example for $n = 2$ and $0 < c < 1$ we can define μ_c on the standard basis subsets of $\wp(2)^\infty \times 2^\infty$ by

$$\mu_c \left\{ \langle f, g\rangle \;\Big|\; \bigwedge_{k=1}^{n} f(j_k) = s_k \wedge \bigwedge_{r=1}^{m} g(i_r) = q_r \right\} = 2^{-m} \prod_{k=1}^{m} c^{|s_k|}(1-c)^{2-|s_k|}.$$

If we now extend μ_c to a normalized σ-additive, Ex and Px invariant measure measure on $\wp(2)^\infty \times 2^\infty$ then all the w_{μ_c} will be different.

We now give a version of a 'Ladder Theorem' which provides a classification of the probability functions satisfying Wilmers' Principle and will be useful in the next section. In this theorem the ξ_n are as given in (2).

Theorem 5 (Ladder Theorem). *Let w be a probability function on S_2L satisfying Wilmers' Principle. Then w can be represented as*

$$w = \lambda_0 w_0 + \sum_{n\in\mathbb{N}^+} \lambda_n w_n \tag{9}$$

with the w_n *satisfying Wilmers' Principle,* $w_n(\xi_n) = 1$ *and* $w_0(\xi_n) = 0$ *for* $n > 0$, $\lambda_n \geq 0$ *and* $\sum_n \lambda_n = 1$.

Proof Let w be a function on S_2L satisfying Wilmers' Principle. For $n > 0$, if $w(\xi_n) = 0$ let $\lambda_n = 0$ and take w_n to be any probability function satisfying Wilmers' Principle and $w_n(\xi_n) = 1$. On the other hand if $w(\xi_n) = \lambda_n > 0$ set $w_n = w(\cdot \mid \xi_n)$. Since ξ_n does not contain any constant or predicate symbols w_n satisfies Ex and Px and similarly w_n inherits Wilmers' Principle from w since w gives value 1 to any instance of this principle. Clearly $w_n(\xi_n) = 1$.

Now consider

$$\hat{w} = w - \sum_{n \in \mathbb{N}^+} \lambda_n \cdot w_{\mu_n}$$

and let $\lambda_0 = \hat{w}(\top)$. If $\lambda_0 = 0$, we are done. Otherwise $w_0 = \lambda_0^{-1}\hat{w}$, is a probability function satisfying Ex, Px and Wilmers' Principle. Also for $n > 0$, $w_0(\xi_n) = 0$ since $w_m(\xi_n) = 0$ for $0 < m \neq n$ and

$$w(\xi_n) = \lambda_n = \lambda_n w_n(\xi_n).$$

■

We have already seen that there are probability functions w satisfying Wilmers' Principle in whose Ladder Representation $\lambda_n > 0$. The same is true for λ_0:

We are now in a position to show that:

Proposition 6. *There are probability functions satisfying Wilmers' Principle with* $\lambda_0 > 0$ *in the Ladder Representation.*

Proof Let ξ_n be as above. Since the set $\{\neg\xi_n \mid n \in \mathbb{N}^+\}$ together with all instances of Wilmers' Principle is consistent it has a model $M \in \mathcal{T}L$. Define the probability function $V_M : S_2L \to \{0, 1\}$ by

$$V_M(\vartheta) = \begin{cases} 1 & \text{if } M \models \vartheta, \\ 0 & \text{if } M \models \neg\vartheta. \end{cases} \tag{10}$$

As given V_M does not satisfy Ex or Px. However by a method introduced into this field by Gaifman, see [4] or [7, Theorem 12.3], we can use V_M to construct a probability function w which does satisfy Ex and Px and which gives non-zero probability to the same sentences. Indeed in this case it will continue to give all the instances of Wilmers' Principle and the $\neg\xi_n$ probability 1 so for this w we must have that $\lambda_0 = 1$. ■

4 Regularity and Super Regularity

Concerning the regularity and super regularity[4] of probability functions satisfying Wilmers' Principle the first point to notice is that for $n > 0$ the w_n in (9) do not satisfy Reg since, for example, for $m > n$

$$\bigwedge_{j=1}^{m} \left(P_j(a_j) \wedge \bigwedge_{\substack{1 \leq i \leq m \\ i \neq j}} \neg P_j(a_i) \right)$$

is inconsistent with ξ_n and hence gets probability 0 according to w_n. Hence if $\lambda_0 = 0$ and only finitely many of the λ_n are non-zero in the representation (9) then w will not satisfy Reg.

However we do have:

Theorem 7. *If w satisfies Wilmers' Principle and either $\lambda_0 > 0$ or infinitely many λ_n are non-zero in the Ladder Representation of w then w satisfies Reg.*

Proof We start by considering the case when infinitely many of the λ_n are non-zero.

By the Disjunctive Normal Form Theorem it is enough to show that w gives non-zero probability to quantifier free sentences of the form

$$\bigwedge_{i=1}^{m} \bigwedge_{j=1}^{k} P_j^{\varepsilon_{i,j}}(a_i)$$

where the $\varepsilon_{i,j} \in \{0, 1\}$ and $P^1 = P,\ P^0 = \neg P$. Pick n large (i.e. compared with m, k) with $\lambda_n > 0$ and let $\langle f, g \rangle \in \wp(n)^\infty \times n^\infty$ be such that ξ_n and Wilmers' Principle hold in $M_{f,g}$.

For $j = 1, \ldots, n$ pick distinct $\vec{\delta}_j = \langle \delta_{1,j}, \delta_{2,j}, \ldots, \delta_{n,j} \rangle$ agreeing on the first m coordinates with $\langle \varepsilon_{1,j}, \varepsilon_{2,j}, \ldots, \varepsilon_{m,j} \rangle$ for $j \leq k$ and such that the vectors $\langle \delta_{i,1}, \delta_{i,2}, \ldots, \delta_{i,n} \rangle$ are also all distinct (possible since n is large). From the proof of the converse in Theorem 4 $\mathrm{ran}(f) = \wp(n)$ so for each $\vec{\delta} = \delta_1, \delta_2, \ldots, \delta_n \in \{0, 1\}^n$ there is a $\kappa(\vec{\delta}) \in \mathbb{N}^+$ such that

$$M_{f,g} \models \bigwedge_{i=1}^{n} P_{\kappa(\vec{\delta})}^{\delta_i}(e_i).$$

So

$$M_{f,g} \models \bigwedge_{j=1}^{n} \bigwedge_{i=1}^{n} P_{\kappa(\vec{\delta}_j)}^{\delta_{i,j}}(e_i),$$

[4] Which continue to apply only to first order sentences.

and

$$M_{f,g} \models \exists X_1, \ldots, X_n \, \exists x_1, \ldots, x_n \bigwedge_{j=1}^{n} \bigwedge_{i=1}^{n} X_j^{\delta_{i,j}}(x_i).$$

Hence from Theorem 4,

$$w_n \left(\exists X_1, \ldots, X_n \, \exists x_1, \ldots, x_n \bigwedge_{j=1}^{n} \bigwedge_{i=1}^{n} X_j^{\delta_{i,j}}(x_i) \right) = 1.$$

Noticing that by the choice of $\delta_{i,j}$ these X_j and x_i must be distinct we now obtain by Lemma 2, Ex and Px that

$$w_n \left(\bigwedge_{j=1}^{n} \bigwedge_{i=1}^{n} P_j^{\delta_{i,j}}(a_i) \right) > 0, \tag{11}$$

and in turn, as required,

$$w_n \left(\bigwedge_{j=1}^{k} \bigwedge_{i=1}^{m} P_j^{\varepsilon_{i,j}}(a_i) \right) > 0.$$

Turning now to the case when $\lambda_0 > 0$. Let μ_0 be the measure for w_0 as given in the proof of Theorem 3, so

$$\mu_0 \left\{ M \in \mathcal{T}L \,\middle|\, M \models WP \wedge \exists x_1 \ldots \exists x_n \bigwedge_{1 \leq i < j \leq n} \exists X \, \neg(X(x_i) \leftrightarrow X(x_j)) \right\} = 1.$$

Let $M \in \mathcal{T}L$ be in this measure 1 set, without loss of generality say

$$M \models \bigwedge_{1 \leq i < j \leq n} \exists X \, \neg(X(a_i) \leftrightarrow X(a_j)).$$

Then since $M \models WP$, just as before the sets

$$\{a_i \,|\, 1 \leq i \leq n, \; M \models P_j(a_i)\}$$

form a Boolean Algebra which must be all subsets of $\{a_1, a_2, \ldots, a_n\}$.

The demonstration that Reg holds now proceeds just as in the case above.

■

We now improve this Theorem to conclude SReg.

Theorem 8. *If w satisfies Wilmers' Principle and either $\lambda_0 > 0$ or infinitely many λ_n are non-zero in the Ladder Representation of w then w satisfies SReg.*

Proof We first give the proof for the case when infinitely many λ_n are non-zero in the Ladder Representation of w.

Since for this unary language L any consistent sentence in $S_1 L$ is logically equivalent to a disjunction of sentences of the form [5]

$$\bigwedge_{i=1}^{m} P^{\vec{\delta}_i}(a_i) \wedge \bigwedge_{i=1}^{s} \exists x\, P^{\vec{\tau}_i}(x) \wedge \bigwedge_{i=1}^{q} \forall x\, \neg P^{\vec{\gamma}_i}(x),$$

where $P^{\vec{\delta}_i}(a_i)$ is short for $\bigwedge_{j=1}^{r} P_j^{\delta_{i,j}}(a_i)$ etc., to show that w satisfies SReg it is enough, by Px, to show that for n large compared m and r and $\lambda_n > 0$ there is some choice of distinct predicate symbols, $R_1, \ldots, R_r$ say, such that

$$w_n\left(\bigwedge_{i=1}^{m} R^{\vec{\delta}_i}(a_i) \wedge \bigwedge_{\vec{\delta} \neq \vec{\delta}_1, \ldots, \vec{\delta}_m} \forall x\, \neg R^{\vec{\delta}}(x)\right) > 0.$$

Assume that $\langle 0, 0, \ldots, 0 \rangle$ is amongst these $\vec{\delta}_k$. [Changing $\langle 0, 0, \ldots, 0 \rangle$ to some other vector $\vec{\delta}_k$ only requires a minor modification of the proof which follows. Notice that there must be *some* $\vec{\delta}_k$ otherwise the sentence is actually a contradiction!]

By Theorem 7

$$w_n\left(\bigwedge_{i=1}^{m} P^{\vec{\delta}_i}(a_i)\right) > 0. \tag{12}$$

Let

$$\psi = \bigwedge_{i=1}^{m} P^{\vec{\delta}_i}(a_i).$$

From Wilmer's Principle for w_n

$$w_n\left(\exists X\, \forall x \left(X(x) \leftrightarrow \bigvee_{i=1}^{m} P^{\vec{\delta}_i}(x)\right)\right) = 1$$

so with (12) there is some predicate symbol R_{r+1} such that

$$w_n\left(\psi \wedge \forall x \left(R_{r+1}(x) \leftrightarrow \bigvee_{i=1}^{m} P^{\vec{\delta}_i}(x)\right)\right) > 0. \tag{13}$$

We now claim that for $t = 1, \ldots, r$ there are distinct predicate symbols $R_1, \ldots, R_t$ such that

$$w_n(\zeta_t(R_1, R_2, \ldots, R_t)) > 0 \tag{14}$$

[5] See for example [8, Chapter 10].

where $\zeta_t(R_1, R_2, \ldots, R_t)$ is

$$\psi \wedge \forall x \left(R_{r+1}(x) \leftrightarrow \bigvee_{i=1}^{r} P^{\vec{\delta_i}}(x) \right) \wedge \bigwedge_{j=1}^{t} \forall x \left(R_j(x) \leftrightarrow (R_{r+1}(x) \wedge P_j(x)) \right).$$

To see this suppose that we have such a $\zeta_t(R_1, R_2, \ldots, R_t)$ for some $t < r$ and distinct $R_1, \ldots, R_t$. We wish to show it for $t+1$. Since by Wilmers' Principle

$$w_n(\exists X \forall x (X(x) \leftrightarrow (R_{r+1}(x) \wedge P_{t+1}(x)))) = 1$$

there is a predicate symbol R_{t+1} such that

$$w_n(\zeta_{t+1}(R_1, R_2, \ldots, R_t, R_{t+1})) > 0 \tag{15}$$

where $\zeta_{t+1}(R_1, R_2, \ldots, R_t, R_{t+1})$ is

$$\zeta_t(R_1, R_2, \ldots, R_t) \wedge \forall x \left(R_{t+1}(x) \leftrightarrow (R_{r+1}(x) \wedge P_{t+1}(x)) \right).$$

Unfortunately we seem to have no immediate guarantee that R_{t+1} differs from $R_1, \ldots, R_t$.

So suppose, without loss of generality, that in fact $R_{t+1} = R_1 \neq R_2, R_3, R_4, \ldots, R_t$. By Px for distinct relation symbols $S_1, S_2, \ldots, S_t$ different from $P_1, P_2, \ldots, P_r, R_1, R_2, \ldots, R_t, R_{r+1}$ we have that

$$w_n(\zeta_{t+1}(R_1, R_2, \ldots, R_t, R_{t+1})) = w_n(\zeta_{t+1}(S_1, S_2, \ldots, S_t, S_1)) > 0.$$

We can take infinitely many such $S_1, S_2, \ldots, S_t$ so a pair of these must overlap in the sense that their conjunction has non-zero w_n probability. Without loss of generality we may suppose then that $w_n(\rho) > 0$ where ρ is

$$\zeta_{t+1}(R_1, R_2, \ldots, R_t, R_{t+1}) \wedge \zeta_{t+1}(S_1, S_2, \ldots, S_t, S_{t+1})$$

and $S_{t+1} = S_1$ (and $R_{t+1} = R_1$). But

$$\rho \models \forall x \left(R_{t+1}(x) \leftrightarrow (R_{r+1}(x) \wedge P_{t+1}(x)) \right) \wedge \forall x \left(S_{t+1}(x) \leftrightarrow (R_{r+1}(x) \wedge P_{t+1}(x)) \right)$$

so

$$\rho \models \forall x \left(R_{t+1}(x) \leftrightarrow S_{t+1}(x) \right)$$

and in turn

$$\rho \models \zeta_{t+1}(R_1, R_2, \ldots, R_t, S_{t+1}).$$

Hence

$$w_n(\zeta_{t+1}(R_1, R_2, \ldots, R_t, S_{t+1})) \geq w_n(\rho) > 0$$

and the $R_1, R_2, \ldots, R_t, S_{t+1}$ now are all different.

Having now found such distinct $R_1, R_2, \ldots, R_r$ let $\zeta = \zeta_r(R_1, R_2, \ldots, R_r)$, so $w_n(\zeta) > 0$.

It remains to show that

$$\zeta \models \bigwedge_{i=1}^{m} R^{\vec{\delta}_i}(a_i) \wedge \bigwedge_{\vec{\delta} \neq \vec{\delta}_1, \ldots, \vec{\delta}_m} \forall x \, \neg R^{\vec{\delta}}(x).$$

Since for $i = 1, \ldots, m$

$$\zeta \models \zeta \wedge \psi \models R_{r+1}(a_i) \wedge P^{\vec{\delta}_i}(a_i),$$

it follows that

$$\zeta \models \bigwedge_{i=1}^{m} R^{\vec{\delta}_i}(a_i),$$

which gets us halfway there.

Now suppose that $\vec{\delta}$ is not amongst the $\vec{\delta}_1, \ldots, \vec{\delta}_m$. Notice that by our initial assumption at least one of the coordinates of $\vec{\delta}$ is 1. In this case

$$\zeta \wedge R^{\vec{\delta}}(x) \models R_{r+1}(x)$$

and since

$$\zeta \wedge R_{r+1}(x) \models P_i(x) \leftrightarrow R_i(x),$$

$$\zeta \wedge R^{\vec{\delta}}(x) \models \zeta \wedge R_{r+1}(x) \wedge R^{\vec{\delta}}(x) \models P^{\vec{\delta}}(x) \wedge \bigvee_{i=1}^{m} P^{\vec{\delta}_i}(x) \models \bot,$$

which gives that

$$\zeta \models \forall x \, \neg R^{\vec{\delta}}(x).$$

We now have, as required, that

$$w_n \left(\bigwedge_{i=1}^{m} R^{\vec{\delta}_i}(a_i) \wedge \bigwedge_{\vec{\delta} \neq \vec{\delta}_1, \ldots, \vec{\delta}_m} \forall x \, \neg R^{\vec{\delta}}(x) \right) \geq w_n(\zeta) > 0$$

with the $R_1, R_2, \ldots, R_r$ are distinct.

The corresponding result to Theorem 8 when $\lambda_0 > 0$ follows similarly.

■

A straightforward corollary here then is:

Corollary 9. *If the probability function w satisfies Wilmers' Principle and Reg then it satisfies SReg.*

It is worth noting that strengthening Wilmers' Principle to allow the formula $\vartheta(x)$ to be second order does not provide any correspondingly stronger results than we already have. Firstly Reg is perforce first order and the second order version of SReg trivially cannot be proved because it would require the satisfiable negations of instances of Wilmers' Principle to have non-zero probability in the very presence of Wilmers' Principle!

In this note we have considered one possible second order principle in relation to just one first order consequence i.e. regularity. There are many more such second order principles, both monadic and polyadic, that one might similarly consider, the only limit being their perceived degree of rationality within the context of Inductive Logic and their interesting first order consequences.

References

[1] Boolos, G., To Be is to be a Value of a Variable (or Some Values of Some Variables), *The Journal of Philosophy*, 1984, **81**:436-449

[2] Carnap, R., A Basic System of Inductive Logic, in *Studies in Inductive Logic and Probability, Volume I*, Eds. R.Carnap & R.C.Jeffrey, University of California Press, 1971, pp33-165.

[3] Cournot, A.A., *Exposition de la théorie des chances et des probabilités,* L.Hachette, Paris, 1843.

[4] Gaifman, H., Concerning measures on first order calculi, *Israel Journal of Mathematics*, 1964, **2**:1-18.

[5] Hintikka, J. & Niiniluoto, I., An Axiomatic Foundation for the Logic of Inductive Generalization, in *Studies in Inductive Logic and Probability, Volume II*, ed. R. C. Jeffrey, University of California Press, 1980, pp157-192

[6] Johnson, W.E., Probability: The Deductive and Inductive Problems, *Mind*, 1932, **41**:409-423.

[7] Paris, J.B. *The Uncertain Reasoner's Companion,* Cambridge University Press, 1994.

[8] Paris, J.B. and Vencovská, A., Pure Inductive Logic, Monograph, to appear

[9] Paris, J.B. and Vencovská, A., A Note on Irrelevance in Inductive Logic, *Journal of Philosophical Logic*, 2011, **40**:357-370

[10] Shapiro, S., Foundations without Foundationalism, Oxford University Press, 1991