Pure Inductive Logic
Praha, 2017

,

1. Basics Notation. Definition of probability functions on SL and their properties (Pa)-(Pe). Examples of probability functions V_M and c_∞. Conditional probability. Central question. Principles Ex, Px and SN.

2. Justifications of Probability, Specifying prob. functions on QFSL Dutch Book argument. Conditional Dutch Book argument. State descriptions. Probability defined via state descriptions. Ex for state descriptions means full Ex.

3. Extending Probability Functions from QFSL to SL, Unary PIL Gaifman's theorem. Atoms. Unary Ex. Functions $w_{\vec{c}}$. Principle IP.

4. de Finetti's Representation Theorem, Further Unary Princiles de Finetti's representation theorem. Principle of Regularity Reg. Principle of Instantial Relevance PIR. Ex implies PIR. Atom Exchangeability. Reichenbach Axiom RA. Johnson Sufficientness Principle JSP.

5. Carnap's Continuum and another continuum Carnap's Continuum. Unary Language Invariance ULi. Generalised Principle of Instantial Relevance GPIR. Functions w_δ^L. NP-continuum. Recovery. Weak Irrelevance Principle WIP.

6. Polyadic Pure Inductive Logic Functions w^D. Representation theorem for Ex and L containing just one, binary, relation symbol. General representation theorem for Ex.

7. Analogy Strong Analogy Principle SAP. General Analogy principle GAP. Counterparts Principles CP and SCP. Language Invariance Li. Li implies CP and SCP.

8. Spectrum Exchangeability Spectrum Exchangeability Principle Sx. Variable Exchangeability Principle Vx. Functions $u^{\bar{p},L}$ and their properties.

9. Principle of Induction Carnap's Continuum and NP continuum for polyadic languages. Principle of Induction PI.

Introduction

Following R.Carnap, we may define *inductive logic* as a theory of logical probability providing rules for inductive thinking. *Probability* in this context represents something like *rational degree of belief*, as opposed to *frequency.*

Recall that *probability* has long been understood in two ways, either as *physical, objective* in which case it is meant to stand for *frequency,* or *propensity*, that is, a physical quantity 'out there', or as *Bayesian, inductive* in which case it stands for a *reasonable expectation representing a state of knowledge*, a degree of belief. Inductive logic is concerned with the latter.

An example (due to Patrick Maher) which perhaps best illustrates the difference between these two main concepts of probability is to imagine a situation in which somebody is about to flip a coin of which we know that rather than having head and tail (panna, orel) it has either two heads or two tails. The physical, frequency-based probability of a toss yielding head is 0 or 1 although we do not know which. However - as long as no other information about the coin is available - the logical probability, due to the symmetry of the situation, has to be 1/2. It would seem irrational and incoherent to make it anything else. If we know other things, the logical probability might change, or not. For example, the fact that the person tossing the coin has black hair should be judged irrelevant but the fact that the person throwing the coin is known to be in possession of a coin with two heads appears relevant and the (logical) probability should arguably increase. (Note that the physical probability would remain 0 or 1).

Pure inductive logic is the part of inductive logic which is concerned purely with the theoretical and mathematical foundations of the subject. Rather than attempting to argue what degrees of belief should be adopted in concrete situations, as above, on the grounds of symmetry, relevance, analogy etc, it investigates in an abstract setting what happens when various (combinations) of principles motivated by such desiderata are adopted. That is, PIL investigates what properties the process of probability assignments must have when we commit ourselves to some basic tenets; what can be proved to hold about probability assignments on the basis of given axioms. The relation of pure and applied inductive logic resembles that of pure and applied mathematics - the problem of interpreting the theoretical framework and results being left to the applied IL. We shall see that rather like set theory, inductive logic offers a number of basic axioms (principles) that formalise intuitively rational and desirable properties of the system, some incompatible with others. However, there is far less consensus on any generally acceptable collection of these (unlike in set theory, where ZF has the central stage), and there might yet be surprising developments of the subject.

The first substantial steps in the development of PIL were taken by W.E.Johnson and R.Carnap. They worked with unary properties only, and independently arrived at proposing what came to be known as Carnap's continuum of inductive methods. This 'Continuum' does in a certain sense determine, up to a real parameter, how we should practice inductive inference in the unary context, provided of course we are willing to accept the principles Carnap and Johnson start with. We shall explain their results during the course. Their work was done in the earlier parts of the 20th century and it was followed by much controversy, since for Carnap this was a part of an ambitious program to account for all human rational reasoning; nowadays there is a broad agreement that it cannot be done.

Carnap himself distinguished pure and applied inductive logic, and it could be said that it is the applied one which causes the trouble. The question how to assign probabilities to sentences (or propositions ....) rationally in an abstract, uninterpreted setting on the other hand appears, to me at least, entirely legitimate and in spite of a lot of recent effort going into it, as challenging as ever. The formal framework developed and results reached over the last twenty years make it possible to work with relations rather than just unary predicates and this more general perspective has thrown new light even on the unary problems. However, the 'holy grail' of inductive logic (as Jon Williamson puts it in his book on Inductive Logic), a best, generally acceptable collection of principles determining how to assign probabilities, remains elusive. In the light of its potential interest to artificial intelligence, it is well worth further effort. This course is about some of what is known to date.

1 Basics

We work in first order predicate logic with quantifiers $\forall, \exists$, connectives $\wedge, \vee, \to$ $\neg$, variables $(x, y, \ldots, x_1, x_2, \ldots)$, parentheses (,).

Language L:

- finitely many relation (predicate) symbols $R_1, \ldots, R_q$ of arities $r_1, \ldots, r_q$,
- countably many constant symbols $a_1, a_2, a_3, \ldots$,
- no equality nor function symbols.

Written as $L = \{R_1, \ldots, R_q\}$.

For formulae and sentences of L we use lower case Greek letters, and when useful, we list the constants and free variables appearing in them in brackets. For example, for L containing 3 unary relation symbols (predicates) R_1, R_2, R_3,

$$\phi(a_2, a_7) = (R_1(a_2) \wedge R_2(a_2)) \to R_3(a_7), \quad \theta(a_1) = R_2(a_1) \vee \exists y R_2(y)$$

$$\psi(x, a_1) = R_2(a_1) \vee R_3(x).$$

are sentences/formula of L.

SL ... the set of all sentences of the language L
$QFSL$ ... the set of all quantifier free sentences of the language L.

To reduce subscripts:

Other letters can stand for the a_i.
For example $b_1, b_2, \ldots, b_m$ (sometimes written as $\vec{b}$) for $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$.
R, Q, P (also) stand for relation symbols.

$\mathcal{T}L$ ... the set of structures for L, each with universe $\{a_1, \ldots, a_n, \ldots\}$ and with every a_i interptreted as itself.

Note: $\theta \in SL$ is consistent just if there is $M \in \mathcal{T}L$ such that $M \models \theta$ (because θ is consistent just when it has a countable model, and such a model can be modified so that all individuals in it are (interpretations of) some of the a_i).

Definition A function $w : SL \to [0, 1]$ is a probability function on SL if for all θ, ϕ and $\exists x\, \psi(x) \in SL$

(P1) If θ is logically valid $[\models \theta]$ then $w(\theta) = 1$.

(P2) If θ and ϕ are mutually exclusive $[\models \neg(\theta \wedge \phi)]$ then

$$w(\theta \vee \phi) = w(\theta) + w(\phi)\,.$$

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

Notes: (P3) is called the Gaifman's condition and it is intended to capture the idea that all individuals in the universe are named as $a_1, a_2, \ldots$. A function $w : QFSL \to [0,1]$ satisfying (P1) and (P2) is referred to as a probability function on quantifier-free formulae. Later we will see that any such function has a unique extension to a probability function on SL (Gaifman's Theorem).

Example (i) Let $M \in \mathcal{T}L$. Define $V_M : SL \to \{0,1\}$ by

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

Then V_M is a probability function on SL.

(ii) Define $c_\infty : SL \to [0,1]$ by setting

$$c_\infty(R_i(b_1, \ldots, b_{r_i})) = \frac{1}{2}$$

for any R_i and any $b_1, \ldots, b_{r_i}$, and by requiring all the distinct instantiations of the predicates or their negations to be stochastically independent. That is,

$$c_\infty\left(\bigwedge_{j=1}^{k} \pm R_{i_j}(b^j_1, \ldots, b^j_{r_{i_j}})\right) = \frac{1}{2^k}$$

(where $\pm R$ stands for R or $\neg R$).

As we shall see using the DNF theorem and then the promised Gaifman's theorem, c_∞ extends uniquely to a probability function on SL, 'the fairest one' (also called the *completely independent probability function*). When the language is not clear from the context, we write c^L_∞.

Properties of probability functions

Let $w : SL \to [0,1]$ satisfy (P1) and (P2). Then for $\theta, \phi \in SL$,

(Pa) $\quad w(\neg\theta) = 1 - w(\theta)$.

Proof. We have that $\models \theta \vee \neg\theta$ and $\models \neg(\theta \wedge \neg\theta)$ so by (P1) and (P2),

$$1 = w(\theta \vee \neg\theta) = w(\theta) + w(\neg\theta).$$

(Pb) $\quad \models \neg\theta \;\Rightarrow\; w(\theta) = 0$.

Proof. From $\models \neg\theta$ we have $w(\neg\theta) = 1$ by (P1) so from (Pa), $w(\theta) = 0$.

(Pc) $\quad \theta \models \phi \;\Rightarrow\; w(\theta) \leq w(\phi)$.

Proof. If $\theta \models \phi$ then $\models \neg(\neg\phi \wedge \theta)$ so from (P2), (Pa) and the fact that w takes values in $[0,1]$,

$$1 \geq w(\neg\phi \vee \theta) = w(\neg\phi) + w(\theta) = 1 - w(\phi) + w(\theta)$$

from which the required inequality follows.

(Pd) $\quad \theta \equiv \phi \;\Rightarrow\; w(\theta) = w(\phi)$.

Proof. If $\theta \equiv \phi$ then $\theta \models \phi$ and $\phi \models \theta$. By (Pc), $w(\theta) \leq w(\phi)$ and $w(\phi) \leq w(\theta)$ so $w(\theta) = w(\phi)$.

(Pe) $\quad w(\theta \vee \phi) = w(\theta) + w(\phi) - w(\theta \wedge \phi)$.

Proof. Since

$$\theta \vee \phi \equiv ((\theta \wedge \neg\phi) \vee (\neg\theta \wedge \phi)) \vee (\theta \wedge \phi)$$

and the disjuncts on the RHS are mutually exclusive, by (Pd) and repeated application of (P2), we have

$$w(\theta \vee \phi) = w(\theta \wedge \neg\phi) + w(\neg\theta \wedge \phi) + w(\theta \wedge \phi) \tag{2}$$

Using (Pd) and (P2) with

$$\theta \equiv (\theta \wedge \neg\phi) \vee (\theta \wedge \phi), \quad \phi \equiv (\phi \wedge \neg\theta) \vee (\phi \wedge \theta)$$

yields

$$w(\theta) = w(\theta \wedge \neg\phi) + w(\theta \wedge \phi), \quad w(\phi) = w(\phi \wedge \neg\theta) + w(\phi \wedge \theta)$$

so adding and subtracting $w(\theta \wedge \phi)$ to the RHS of (2) yields the result.

1.0.1 Conditional probability

Definition Given a probability function w on SL and $\phi \in SL$ with $w(\phi) > 0$ we define the *conditional probability function* $w(.\,|\,\phi) : SL \to [0,1]$ by

$$w(\theta\,|\,\phi) = \frac{w(\theta \wedge \phi)}{w(\phi)}. \tag{3}$$

Proposition 1 *Let w be a probability function on SL, $\phi \in SL$ and $w(\phi) > 0$. Then $w(.\,|\,\phi)$ is a probability function and $w(\theta\,|\,\phi) = 1$ whenever $\phi \vDash \theta$.*

Proof. To show (P1) suppose that $\vDash \theta$. Then $\phi \equiv \theta \wedge \phi$ so $w(\theta \wedge \phi) = w(\phi)$ by property (Pd) and in turn $w(\theta\,|\,\phi) = 1$.

For (P2) suppose that $\vDash \neg(\eta \wedge \theta)$. Then $\models \neg((\eta \wedge \phi) \wedge (\theta \wedge \phi))$ so since

$$(\theta \vee \eta) \wedge \phi \equiv (\theta \wedge \phi) \vee (\eta \wedge \phi),$$

$$\begin{aligned} w((\theta \vee \eta) \wedge \phi) &= w((\theta \wedge \phi) \vee (\eta \wedge \phi)), \quad \text{by property (Pd)},\\ &= w(\theta \wedge \phi) + w(\eta \wedge \phi), \quad \text{by (P2) for } w, \end{aligned}$$

and dividing by $w(\phi)$ gives the result.

For (P3), note that

$$\exists x\, \psi(x) \wedge \phi \equiv \exists x\, (\psi(x) \wedge \phi),$$

$$\left(\bigvee_{i=1}^{n} \psi(a_i)\right) \wedge \phi \equiv \bigvee_{i=1}^{n} (\psi(a_i) \wedge \phi),$$

so using property (Pd) and (P3) for w,

$$\begin{aligned} w(\exists x\, \psi(x) \wedge \phi) &= w(\exists x\, (\psi(x) \wedge \phi))\\ &= \lim_{n\to\infty} w\left(\bigvee_{i=1}^{n} (\psi(a_i) \wedge \phi)\right)\\ &= \lim_{n\to\infty} w\left(\left(\bigvee_{i=1}^{n} \psi(a_i)\right) \wedge \phi\right) \end{aligned}$$

and the result follows after dividing both sides by $w(\phi)$.

Finally, if $\phi \models \theta$ then $\phi \equiv \theta \wedge \phi$ so $w(\theta \wedge \phi) = w(\phi)$ by property (Pd) and in turn $w(\theta \,|\, \phi) = 1$. $\square$

Along with Carnap and others, we will assume that conditional probability models updating, that is degrees of belief on the basis of ϕ correspond to $w(. \mid \phi)$.

Example (continued). V_M corresponds to an agent who is already sure about everything, who can only learn what he believes with probability 1 anyway, and whose probability function does not change upon learning it. That is, $V_M(.|\phi)$ is defined only when $M \vDash \phi$ and in that case $V_M(\theta|\phi) = V_M(\theta)$ for all $\theta \in SL$. c_∞ on the other hand is extremely open-minded: for example, an agent using a language with just one unary predicate R and employing c_∞ would continue giving $R(a_n)$ belief $\frac{1}{2}$ after being told that $R(a_1) \ldots, R(a_{n-1})$, regardless of n.

We are now in a position to formulate the main problem of the subject.

Question: *Assuming a rational agent inhabits a structure $M \in \mathcal{T}L$ but knows nothing about which one it is (about what is true in M). What probability function $w : SL \to [0,1]$ should he adopt when $w(\theta)$ is to represent his degree of belief that a sentence $\theta \in SL$ is true in this ambient structure M?*

This can be seen as the central question, since after learning some facts expressed e.g. by sentences $\phi_1, \phi_2, \ldots, \phi_n$, provided that $w\left(\bigwedge_{i=1}^n \phi_i\right) \neq 0$, the agent could/should adopt the conditional probability

$$w\left(. \,\middle|\, \bigwedge_{i=1}^n \phi_i\right) : SL \to [0,1]$$

as his probability function for the new context when $\phi_1, \phi_2, \ldots, \phi_n$ are known to be true.

The question has puzzled generations of logicians. The two examples of probability functions which we have considered so far, V_M and c_∞ clearly are not particularly suitable.

There are many other probability functions, and to judge them, various principles have been proposed as desirable (rational) for a probability function to be adopted on the basis of zero knowledge. Next we will introduce some of these principles.

Some Basic Principles

The most basic principles are based on symmetry, and justified by arguing that if there is a symmetry in the situation then it would be irrational of the agent to break that symmetry when assigning probabilities.

One obvious such symmetry relates to the constants $a_1, a_2, a_3, \ldots$. In the situation of zero knowledge, the agent has no reason to treat these any differently - the subscripts on the a's are simply to allow us to list them easily, the agent is not supposed to 'know' that a_1 comes before a_2 which comes before . . . in our list. This consideration leads to:

The Constant Exchangeability Principle, Ex

For $\theta(a_1, a_2, \ldots, a_m) \in SL$ and any other m-tuple of distinct constants $b_1, b_2, \ldots, b_m$,

$$w(\theta(a_1, a_2, \ldots, a_m)) = w(\theta(b_1, b_2, \ldots, b_m)). \tag{4}$$

Equivalently, for any permutation σ of $\mathbb{N}^+$ and any $\theta(a_1, a_2, \ldots, a_m) \in SL$

$$w(\theta(a_1, a_2, \ldots, a_m)) = w(\theta(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)})). \tag{5}$$

We will be assuming Ex of almost all probability function which we will be considering.

Similarly, since in the situation of zero knowledge there is no reason to distinguish between relations of the same arity:

The Principle of Predicate Exchangeability, Px

If R, R' are relation symbols of L with the same arity then for $\theta \in SL$,

$$w(\theta) = w(\theta')$$

where θ' is the result of simultaneously replacing R by R' and R' by R throughout θ.

The following, somewhat more contentious principle, is based on the claim that in the situation of zero knowledge there is a symmetry between any relation symbol R of L and its negation $\neg R$ and so our agent has no reason to treat R any differently than $\neg R$. Since $\neg\neg R$ is logically equivalent to R this leads to:

The Strong Negation Principle, SN

For $\theta \in SL$,

$$w(\theta) = w(\theta')$$

where θ' is the result of replacing each occurrence of R in θ by $\neg R$.

Example (continued). c_∞ satisfies Ex,Px and SN. There are some special M such that V_M satisfies Ex and/or Px, but no V_M can satisfy SN.

Problems

Problem 1 *(a) Show that if w_1 and w_2 are probability functions on SL then $\frac{1}{2}(w_1+w_2)$ is also a probability function on SL.*

(b) Let $\langle D, \mathcal{B}, \mu\rangle$ be a measure space, $\mu(D) = 1$, and let $d \mapsto w_d$ be an assignment of probability functions on SL to the elements of D such that for each $\theta \in SL$, the function $d \mapsto w_d(\theta)$ is (Lebesgue) measurable. Show that w defined by*

$$w(\theta) = \int_D w_d(\theta) d\mu$$

is also a probability function on SL.

Problem 2 *Show that for $\theta, \phi \in SL$, the following are equivalent:*

(i) $w(\theta) \le w(\phi)$ *for all probability functions w on SL.*
(ii) $\theta \models \phi$.

Problem 3 *Let $w : SL \to [0,1]$ satisfy (P1), (P2). Then condition (P3) is equivalent to:*

$$(P3') \qquad w(\exists x\, \psi(x)) = \sum_{n=1}^{\infty} w\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right)$$

for $\exists x\, \psi(x) \in SL$.

Problem 4 *Show that there are only finitely many structures $M \in \mathcal{T}L$ such that V_M satisfies Ex, and that no V_M satisfies SN.*

Solutions

1 (a) This follows directly by checking $P1 - P3$.

(b) ($P(1)$ and ($P2$) are straightforward. To check ($P3$), recall Lebesgue Dominated convergence theorem:

Let f_n, $n \geq 1$, be a sequence of measurable functions such that f_n converges to f almost everywhere. Suppose there exists an integrable function $g \geq 0$ such that, for all $n \geq 0$, $|f_n| \leq g$ almost everywhere. Then f_n, f are integrable and

$$\lim_{n\to\infty} \int_D f_n d\mu = \int_D f d\mu.$$

Let $\exists x \psi(x) \in SL$ and for $d \in D$ and $n \in \mathbb{N}$ define

$$f_n(d) = w_d(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n)),$$

$$f(d) = w_d(\exists x \psi(x)),$$

$$g(d) = 1.$$

By (P3) which holds for each w_d, $\lim_{n\to\infty} f_n(d) = f(d)$ for each d. The f_n, f are measurable by the assumption made in the question and $0 \leq f_n \leq 1$ since tha values of w_d are all in $[0,1]$. Hence

$$\lim_{n\to\infty} w(\psi(a_1)\vee\psi(a_2)\vee\ldots\vee\psi(a_n)) = \lim_{n\to\infty} \int_D w_d(\psi(a_1)\vee\psi(a_2)\vee\ldots\vee\psi(a_n)) =$$

$$= \lim_{n\to\infty} \int_D f_n(d) d\mu = \int_D f(d) d\mu = \int_D w_d(\exists x \psi(x)) d\mu = w(\exists x \psi(x))$$

so (P3) for w follows.

2 In view of property (Pc), we only need to show that if $\theta \nvDash \phi$ then there is a probability function w such that $w(\phi) < w(\theta)$. But in this case $\{\theta, \neg\phi\}$ is consistent, so has a a model $M \in \mathcal{T}L$. Then $V_M(\theta) = V_M(\neg\phi) = 1$ but $V_M(\phi) = 1 - V_M(\neg\phi) = 0$, as required.

3 Recall that the properties (Pa)-(Pe) follow just from (P1) and (P2), so we can use them of w. Since

$$\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n) \;\equiv\; \bigvee_{j=1}^{n} \left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i) \right)$$

we have

$$\begin{aligned} w(\psi(a_1) \vee \psi(a_2) \vee \ldots \vee \psi(a_n)) &= w\left(\bigvee_{j=1}^{n} \left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i) \right) \right) \\ &= \sum_{j=1}^{n} w\left(\psi(a_j) \wedge \neg \bigvee_{i=1}^{j-1} \psi(a_i) \right) \end{aligned}$$

by repeated use of (P2) since the disjuncts here are all disjoint. The result follows.

4 (i) No V_M can satisfy SN because that would require, for example,

$$V_M(R_1(a_1, \ldots, a_{r_1})) = V_M(\neg R_1(a_1, \ldots, a_{r_1}))$$

but by property (Pa), we have

$$V_M(R_1(a_1, \ldots, a_{r_1})) = 1 - V_M(\neg R_1(a_1, \ldots, a_{r_1}))$$

amd V_M takes only 0,1 values.

(ii) Let r be the highest arity of a relation symbol in L. Any structure $M \in \mathcal{T}L$ such that V_M satisfies Ex is completely specified by the values $V_M(R_i(b_1, \ldots, b_{r_i}))$ for $i \in \{1, \ldots, q\}$ and $b_1, \ldots, b_{r_i}$ (not necessarily distinct) from $\{a_1, \ldots, a_r\}$. There are only finitely many combinations of such values, so the result follows.

Conventions:

• Throughout the course, $=$ between sentences is sometimes used in place of $\equiv$ (logical equivalence).

• If $w(\phi) = 0$ then

$$w(\theta \,|\, \phi) = c, \quad \text{or} \quad w(\theta \,|\, \phi) = w(\eta \,|\, \zeta),$$

stand for

$$w(\theta \wedge \phi) = c\, w(\phi), \quad \text{or} \quad w(\theta \wedge \phi)\, w(\zeta) = w(\eta \wedge \zeta)\, w(\phi)$$

respectively.

2 Dutch Book argument for using probability

Why use probability? Some other approaches, like *truth-functional belief functions*, would be much simpler to work with, since belief values of any sentence can be worked out from belief values of atomic sentences.

The most frequently quoted justification for belief as probability is the Dutch Book argument (Ramsey, de Finetti). It is based on identifying belief with willingness to bet.

Imagine an agent situated in a structure M (which is unknown to him) and required to choose, *for any* $\theta \in SL$ *and* $0 \leq p \leq 1$, one of two wagers - on or against θ (where $s > 0$ is a stake):

(Bet1_p) Get $s(1-p)$ if θ is true in M, pay sp if θ is false in iM.

(Bet2_p) Pay $s(1-p)$ if θ is true in M, get sp if θ is false in M.

Note that:

• The two bets are complementary so that when the agent chooses one of them, his opponent (bookie) is allocated the other one. Hence, with each p, we assume that the agent is able to choose (at least) one of them. For if the agent is not happy to accept Bet1_p then presumably he thinks that the bookie would be getting a better deal, and Bet2_p allows him to swap roles.

• Clearly, Bet1_0 Bet2_1 are acceptable to the agent - greatest possible gain, no risk of loss.

• If $\mathrm{Bet1}_p$ is acceptable to the agent and $0 \leq q < p$ then $\mathrm{Bet1}_q$ is acceptable to him (with $\mathrm{Bet1}_q$: larger gain if θ is true in M and smaller loss if θ is false).

• Similarly if $\mathrm{Bet2}_p$ is acceptable to the agent and $p < q \leq 1$ then $\mathrm{Bet2}_q$ is acceptable.

Consequently, there is some $P \in [0,1]$ such that for all $p < P$, $\mathrm{Bet1}_p$ is acceptable to the agent and for all $p > P$, $\mathrm{Bet2}_p$ is acceptable. Define $Bel(\theta)$ to be that P:

$$Bel(\theta) = \begin{array}{l} \text{the supremum of those } p \in [0,1] \\ \text{for which } \mathrm{Bet1}_p \text{ is acceptable to the agent.} \end{array}$$

$Bel(\theta)$ is a measure of the agent's willingness to bet on θ and in a sense it quantifies the agent's belief that θ is true.

Clearly, this function Bel should be such that the agent cannot be *Dutch-booked*. That is, such that there is no set of (simultaneous) bets each of which is acceptable to the agent but whose combined effect is to cause the agent certain loss no matter what his ambient structure $M \in \mathcal{T}L$ turns out to be.

To analyse the situation, notice that if the agent accepts $\mathrm{Bet1}_p$ for θ (that is, he bets on θ) he will in the event of the ambient structure being M "gain"

$$s(1-p)V_M(\theta) - sp(1 - V_M(\theta)) = s(V_M(\theta) - p)$$

(referring to loss as negative gain). Clearly in $\mathrm{Bet2}_p$ the gain is minus this, i.e. $-s(V_M(\theta) - p)$.

In making the notion of Dutch Book precise, we also include a condition that guarantees that the sets of bets under consideration are such that there is a limit on the loss that the agent or the bookmaker can be exposed to in the worst possible case.

We say that $Bel : SL \to [0,1]$ *can be Dutch-Booked* if there are

• sets A, B (finite or infinitely countable),
• $K > 0$,
• sentences $\theta_i \in SL$, $p_i \in [0, Bel(\theta_i))$ and stakes s_i for $i \in A$,
• sentences ϕ_j, $q_j \in (Bel(\phi_j), 1]$ and stakes $t_j > 0$ for $j \in B$,

such that for all $M \in \mathcal{T}L$

$$\sum_{i \in A} |s_i(V_M(\theta_i) - p_i)|\,, \quad \sum_{j \in B} |t_j(V_M(\phi_j) - q_j)| < K \tag{6}$$

and

$$\sum_{i \in A} s_i(V_M(\theta_i) - p_i) + \sum_{j \in B} (-t_j)(V_M(\phi_j) - q_j) < 0. \tag{7}$$

Theorem 2 *Suppose that* $Bel : SL \to [0,1]$ *is such it cannot be Dutch-Booked. Then Bel satisfies (P1-3).*

Proof For (P1) suppose that $\theta \in SL$ and $\models \theta$ but $Bel(\theta) < 1$. Then for $Bel(\theta) < q < 1$ the agent accepts Bet2_q. But since $V_M(\theta) = 1$ for all $M \in \mathcal{T}$ we have that with stake 1,

$$(-1)(V_M(\theta) - q) = q - 1 < 0$$

which gives an instance of (7), contradiction.

Suppose that (P2) fails, say $\theta, \phi \in SL$ are such that $\models \neg(\theta \wedge \phi)$ but

$$Bel(\theta) + Bel(\phi) < Bel(\theta \vee \phi).$$

At most one of θ, ϕ can be true in any $M \in \mathcal{T}$ so

$$V_M(\theta \vee \phi) = V_M(\theta) + V_M(\phi).$$

Pick $p > Bel(\theta)$, $q > Bel(\phi)$, $r < Bel(\theta \vee \phi)$ such that $p + q < r$. Then with stakes 1,1,1,

$$(-1)(V_M(\theta) - p) + (-1)(V_M(\phi) - q) + (V_M(\theta \vee \phi) - r) = (p + q) - r < 0$$

giving an instance of (7) and contradicting our assumption. A similar argument when

$$Bel(\theta) + Bel(\phi) > Bel(\theta \vee \phi)$$

shows that this cannot hold either so we must have equality here.

Finally suppose that $\exists x\, \psi(x) \in SL$. By Problem I.3 and the fact that we have already proved that (P1), (P2) hold for Bel, it is enough to derive a contradiction from the assumption that

$$\sum_{n=1}^{\infty} Bel\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right) \neq Bel(\exists x\, \psi(x)).$$

Notice that since the sentences on the left hand side here are disjoint both sides are bounded by 1.

We cannot have $>$ here since then that would hold for the sum of a finite number of terms on the left hand side, contradicting (Pc). So we may suppose that we have $<$ here. In this case we can pick

$$p_n > Bel\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right) \qquad \text{for } n = 1, 2, \ldots$$

and $r < Bel(\exists x\, \psi(x))$ with $\sum_{n=1}^{\infty} p_n < r$. Since for $M \in \mathcal{T}$,

$$V_M(\exists x\, \psi(x)) \;=\; \sum_{n=1}^{\infty} V_M\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right)$$

we get, as with the argument above for (P2), that for all stakes 1,

$$\begin{aligned}(V_M(\exists x\, \psi(x)) - r) \;+\; \sum_{n=1}^{\infty}(-1)\left(V_M\left(\psi(a_n) \wedge \neg \bigvee_{i=1}^{n-1} \psi(a_i)\right) - p_n\right) \\ \;=\; -r + \sum_{n=1}^{\infty} p_n < 0,\end{aligned}$$

giving an instance of (7) in contradiction to our assumption.

The Dutch Book argument can also be extended to conditional bets to justify the standard definition of the derived conditional probability given by (3). The idea is that not only is the agent offered unconditional bets as above but also bets about $\theta \in SL$ being true in his ambient structure M *given* that $\phi \in SL$ has turned out to be true in it. Similarly to the above unconditional case then for $\theta, \phi \in SL$, $0 \leq p \leq 1$ and for a stake $s > 0$ the agent is offered a choice of one of two wagers:

(CBet1$_p$): Get $s(1-p)$ if θ is true in M, pay sp if θ is false in it;

(CBet2$_p$): Pay $s(1-p)$ if θ is true in M, get sp if θ is false in it;

with all bets null and void if $M \nvDash \phi$.

Defining $Bel(\theta \mid \phi)$ to be the supremum of those $p \in [0, 1]$ for which CBet1$_p$ is acceptable to the agent, and modifying the notion a Dutch Book for the conditional context, we can show (see Problem 6 for details) that the requirement of no Dutch book against the agent still forces Bel to satisfy (P1),(P2),(P3) and moreover that for all θ, ϕ we have $Bel(\theta \mid \phi) \cdot Bel(\phi) = Bel(\theta \wedge \phi)$.

Consequently, a belief function that avoids all Dutch Books must be a probability function with conditional probability satisfying (3).

Conversely,

Theorem 3 *Suppose that $Bel : SL \to [0,1]$ is a probability function. Then Bel cannot be (conditionally) Dutch Booked.*

We shall prove this soon when we have the means to do it.

Specifying Probability Functions on $QFSL$

Let $L = \{R_1, R_2, \ldots, R_q\}$ where R_i has arity r_i. For distinct constants $b_1, b_2, \ldots, b_m$, a *state description* for $b_1, b_2, \ldots, b_m$ is a sentence of L of the form

$$\Phi(b_1, b_2, \ldots, b_m) = \bigwedge_{i=1}^{q} \bigwedge_{c_1, c_2, \ldots, c_{r_i}} \pm R_i(c_1, c_2, \ldots, c_{r_i})$$

where the $c_1, c_2, \ldots, c_{r_i}$ range over all (not necessarily distinct) choices from $b_1, b_2, \ldots, b_m$ and $\pm R_i$ stands for either R_i or $\neg R_i$. Note that:

• We shall identify two state descriptions if they are the same *up to the ordering of their conjuncts*.

• A state description tells us which of the $R_i(c_1, c_2, \ldots, c_{r_i})$ hold and which do not hold for R_i a relation symbol from our language and any arguments from $b_1, b_2, \ldots, b_m$.

• Any two distinct (inequivalent) state descriptions for $b_1, b_2, \ldots, , b_m$ are exclusive in the sense that their conjunction is inconsistent.

• The state descriptions for $b_1, b_2, \ldots, , b_m$ are exhaustive in the sense that the disjunction of all of them is a tautology.

• For $m = 0$ the sole state description is taken to be a tautology (denoted $\top$).

• Upper case Θ, Φ, Ψ always denote state descriptions.

Example If $L = \{R, P\}$, where P is unary binary and R is binary then

$$P(a_1) \wedge \neg P(a_2) \wedge R(a_1, a_1) \wedge \neg R(a_1, a_2) \wedge \neg R(a_2, a_1) \wedge R(a_2, a_2)$$

is a state description for a_1, a_2.

By the Disjunctive Normal Form Theorem any $\theta(b_1, b_2, \dots, b_m) \in QFSL$ is logically equivalent to a disjunction of state descriptions for $b_1, b_2, \dots, b_m$

$$\theta(\vec{b}) \equiv \bigvee_{\Theta \in S} \Theta(\vec{b})$$

where S is some subset of the set of all state descriptions for m constants. Hence for any w satisfying (P2),

$$w(\theta(\vec{b})) = \sum_{\Theta \in S} w(\Theta(\vec{b})). \tag{8}$$

The values of any probability function on quantifier free sentences are thus determined by its values on state descriptions. Also (adding constants on the right hand side if necessary), just by its values on state descriptions for $a_1, a_2, \dots, a_n$ $(n \in \mathbb{N})$.

Note that if w satisfies Ex, Px or SN respectively on state descriptions then it satisfies them on all $\theta \in QFSL$.

Now assume a function w is defined on the state descriptions $\Theta(a_1, a_2, \dots, a_m)$, $m \in \mathbb{N}$ only, and it satisfies:

(i) $w(\Theta(a_1, a_2, \dots, a_m)) \geq 0$,

(ii) $w(\top) = 1$,

(iii) $$w(\Theta(a_1, a_2, \dots, a_m)) = \sum_{\Phi(a_1, \dots, a_{m+1}) \models \Theta(a_1, \dots, a_m)} w(\Phi(a_1, a_2, \dots, a_{m+1})). \tag{9}$$

Then w extends to a function on $QFSL$ satisfying (P1) and (P2) by setting (unambiguously by (iii))

$$w(\theta(b_1, b_2, \dots, b_m)) = \sum_{\Theta(a_1, \dots, a_k) \models \theta(b_1, \dots, b_m)} w(\Theta(a_1, a_2, \dots, a_k)) \tag{10}$$

where k is sufficiently large that all of the b_i are amongst $a_1, a_2, \dots, a_k$.

Furthermore, in view of the following lemma, there is an easy-to-check condition for this extension to satisfy Ex.

Lemma 4 *Let w satisfy (P1) and (P2) and assume that for any state description $\Phi(a_1, \ldots, a_n)$ and τ a permutation of $\{1, 2, \ldots, n\}$,*

$$w(\Phi(a_1, \ldots, a_n)) = w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)}))\,. \tag{11}$$

Then w satisfies Ex on quantifier-free formulas.[1]

Proof If $\Theta(a_1, \ldots, a_m)$ is a state description and $b_1, b_2, \ldots, b_m$ is any other m tuple of distinct constants, $b_j = a_{i_j}$, then then there is a permutation τ of $\{1, 2, \ldots, n\}$, where

$$n = \max\{i_1, \ldots, i_m)\},$$

such that $\tau(j) = i_j$ for $j = 1, 2, \ldots, m$. So

$$\begin{aligned}
w(\Theta(a_1, \ldots, a_m)) &= \sum_{\Phi(a_1,\ldots,a_n) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_1, \ldots, a_n)) \\
&= \sum_{\Phi(a_1,\ldots,a_n) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)})), \quad \text{by (11)}, \\
&= \sum_{\Phi(a_{\tau(1)},\ldots,a_{\tau(n)}) \models \Theta(a_{\tau(1)},\ldots,a_{\tau(m)})} w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)})) \\
&= \sum_{\Psi(a_1,\ldots,a_n) \models \Theta(a_{\tau(1)},\ldots,a_{\tau(m)})} w(\Psi(a_1, \ldots, a_n)) \\
&= w(\Theta(a_{\tau(1)}, \ldots, a_{\tau(m)})) \\
&= w(\Theta(a_{i_1}, \ldots, a_{i_m})) \\
&= w(\Theta(b_1, \ldots, b_m))\,.
\end{aligned}$$

It follows that w satisfies (4) on state descriptions and hence by virtue of (8) on $QFSL$. □

[1] Note that the condition in this lemma differs from the requirement that w satisfies Ex for state descriptions because the latter would employ permutations of $\mathbb{N}$ rather than (just) $\{1, 2, \ldots, n\}$.

Problems

Problem 5 *Let L contain two unary predicates P and Q, and let*

$$\theta = \forall x(P(x) \vee Q(x)), \quad \phi = P(a_1) \vee P(a_2), \quad \psi = Q(a_1) \wedge Q(a_2).$$

Assume that $Bel : SL \to [0,1]$ satisfies

$$(a) \;\; Bel(\theta) = 0.8, \quad Bel(\phi) = 0.3 \quad \text{and} \quad Bel(\psi) = 0.3,$$

or

$$(b) \;\; Bel(\theta) = 0.6, \quad Bel(\phi) = 0.3 \quad \text{and} \quad Bel(\psi) = 0.3.$$

In each case decide weather or not Bel can be Dutch-booked and if so, find a corresponding Dutch book.

Problem 6 *(i) Write down what the gain/loss of the agent is after accepting $CBet1_p$ or $CBet2_p$ respectively for a stake $s > 0$ when the ambient structure is M.*

(ii) Let $Bel : SL \to [0,1]$, $Bel(.\,|\,.) : SL \times SL \to [0,1]$. Suggest what it means to say that Bel could be Dutch Booked.

(iii) Show that if Bel as above cannot be Dutch Booked than Bel satisfies (P1),(P2),(P3) and for all $\theta, \phi \in SL$, $Bel(\theta\,|\,\phi) \cdot Bel(\phi) = Bel(\theta \wedge \phi)$.

Problem 7 *Let $L = \{R\}$, where R is binary. A state description for m constants, $\Theta(b_1, b_2, \ldots, b_m)$, can be represented by an $m \times m$ $\{0,1\}$ matrix $D_\Theta = (d_{i,j})$ with*

$$d_{i,j} = \begin{cases} 1 & \text{if } \Theta \models R(b_i, b_j) \\ 0 & \text{if } \Theta \models \neg R(b_i, b_j) \end{cases}$$

Express Ex and SN in terms of conditions on values w gives to state descriptions as represented by these matrices.

Problem 8 *Let L_1, L_2 be the languages $\{P\}, \{R\}$ where P is unary and R binary. Let w_2 be a probability function on SL_2. Show that there is a unique probability function w_1 on SL_1 such that*

$$w_1\big(\bigwedge_{i=1}^{n} P^{\epsilon_i}(a_i)\big) = w_2\big(\bigwedge_{i=1}^{n} R^{\epsilon_i}(a_{2i+1}, a_{2i+2})\big),$$

where $\epsilon_i \in \{0,1\}$ and P^1, P^0 stand for P, $\neg P$ respectively (and similarly for R^ϵ).

Solutions to Problems

5 (a) If $Bel : SL \to [0,1]$ satisfied (a) and (P1)-(P3)[1] then, by (Pe), (Pa) we would have

$$Bel(\neg\phi \wedge \neg\psi) = Bel(\neg\phi) + Bel(\neg\psi) - Bel(\neg\phi \vee \neg\psi) \geq 0.7 + 0.7 - 1 = 0.4$$

but since

$$\neg\phi \wedge \neg\psi \models (\neg P(a_1) \wedge \neg Q(a_1)) \vee (\neg P(a_2) \wedge \neg Q(a_2)),$$

it follows that

$$\neg\phi \wedge \neg\psi \models \neg\theta$$

and hence we would also have by (Pc) and (Pa)

$$Bel(\neg\phi \wedge \neg\psi) \leq Bel(\neg\theta) = 1 - Bel(\theta) = 0.2.$$

This is a contradiction, so Bel cannot satisfy both the conditions (a) and (P1)-(P3) and hence by Theorem 2 it can be Dutch booked.

An example of a Dutch book are the following bets with the same stake $s > 0$: $\text{Bet1}_{0.75}$ on θ and $\text{Bet2}_{0.35}$ against both ψ and ϕ/ Then the total gain is

$$\begin{aligned} s(V_M(\theta) - 0.75) - s(V_M(\phi) - 0.35) - s(V_M(\psi) - 0.35) = \\ -0.05s + (V_M(\theta) - V_M(\phi) - V_M(\psi))s \end{aligned}$$

and since $\theta \models \phi \vee \psi$, for any M it must hold that $V_M(\theta) \leq V_M(\phi) + V_M(\psi)$ so the result is always negative as required.

To show that no Dutch book can be found for (b), by Theorem 3 it suffices to find a probability function which agrees with Bel on θ, ϕ, ψ. Using the obvious notation, let M_1, M_2, M_3 be the following structures:

M_1 :

	a_1	a_2	a_3	a_4	a_5	a_6	...
P	0	0	0	0	0	0	...
Q	1	0	1	0	1	0	...

M_2 :

	a_1	a_2	a_3	a_4	a_5	a_6	...
P	1	1	1	1	1	1	...
Q	1	0	1	0	1	0	...

[1] And hence properties (Pa)-(Pe).

$$M_3: \qquad \begin{array}{c|ccccccc} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \dots \\ \hline P & 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ Q & 1 & 1 & 1 & 1 & 1 & 1 & \dots \end{array}$$

Then

$$M_1 \models \neg\theta \wedge \neg\phi \wedge \neg\psi, \quad M_2 \models \theta \wedge \phi \wedge \neg\psi, \quad M_3 \models \theta \wedge \neg\phi \wedge \psi,$$

so

$$w = 0.4\, V_{M_1} + 0.3\, V_{M_2} + 0.3\, V_{M_3}$$

has the required properties.

6 (i) Accepting CBet1$_p$ means gaining

$$sV_M(\phi)(V_M(\theta) - p) = s(V_M(\phi \wedge \theta) - pV_M(\phi))$$

whilst accepting CBet2$_p$ means gaining minus this.

(ii) There are sets (finite or countably infinite) A, B, C, D, sentences θ_i, stakes $s_i > 0$, $p_i \in [0, Bel(\theta_i))$ for $i \in A$, sentences ϕ_i, stakes $t_i > 0$, $q_i \in (Bel(\phi_i), 1]$ for $i \in B$, sentences η_i, ψ_i, stakes $u_i > 0$, $r_i \in [0, Bel(\eta_i \,|\, \psi_i))$ for $i \in C$, sentences ζ_i, ξ_i, stakes $v_i > 0$, $m_i \in (Bel(\zeta_i \,|\, \xi_i), 1]$ for $i \in D$ such that for all $M \in \mathcal{T}L$ we have

$$\textstyle\sum_{i\in A} s_i(V_M(\theta_i) - p_i) + \sum_{i\in B}(-t_i)(V_M(\phi_i) - q_i) +$$
$$\sum_{i\in C} u_i V_M(\psi_i)(V_M(\eta_i) - r_i) + \sum_{i\in D}(-v_i)V_M(\xi_i)(V_M(\zeta_i) - m_i) \;<\; 0 \qquad (12)$$

(and in case of A, B, C, D infinite there is $K > 0$ such that for all $M \in \mathcal{T}L$ the series above converge absolutely with sums less than K).

(iii) By Theorem 2 we can already assume that Bel is a probability function. Suppose first that

$$Bel(\theta \,|\, \phi) \cdot Bel(\phi) < Bel(\theta \wedge \phi). \qquad (13)$$

If $Bel(\theta \,|\, \phi) < Bel(\theta \wedge \phi)$ then picking $Bel(\theta \,|\, \phi) < r < p < Bel(\theta \wedge \phi)$ gives

$$-V_M(\phi)(V_M(\theta) - r) + (V_M(\theta \wedge \phi) - p) = rV_M(\phi) - p \leq r - p < 0$$

for any M, since $V_M(\phi)V_M(\theta) = V_M(\theta\wedge\phi)$, contradicting the given no Dutch Book condition. Hence with (13), $Bel(\phi) < 1$. We also have $Bel(\theta\,|\,\phi) < 1$ since otherwise $Bel(\phi) < Bel(\theta\wedge\phi)$, contradicting Bel being a probability function (property (Pc)). Hence we can pick $Bel(\theta\,|\,\phi) < r$, $Bel(\phi) < q$, $p < Bel(\theta\wedge\phi)$ with $qr < p$. But then considering the corresponding wagers with stakes $1, r, 1$ gives

$$-V_M(\phi)(V_M(\theta) - r) - r(V_M(\phi) - q) + (V_M(\theta\wedge\phi) - p)$$

and furnishes a Dutch Book since it is straightforward to check that its value is $rq - p < 0$ regardless of M.

We have shown that (13) cannot hold. So if the required equality fails it must be because

$$Bel(\theta\,|\,\phi)\cdot Bel(\phi) > Bel(\theta\wedge\phi). \tag{14}$$

But in this case pick $Bel(\theta\,|\,\phi) > r$, $Bel(\phi) > q$, $p > Bel(\theta\wedge\phi)$ with $qr > p$ and obtain a Dutch Book via

$$V_M(\phi)(V_M(\theta) - r) + r(V_M(\phi) - q) - (V_M(\theta\wedge\phi) - p).$$

7 We shall write $w(D_\Theta)$ for $w(\Theta)$ etc.

Ex: Ex is the condition that if D is an $m\times m$ $\{0,1\}$ matrix, σ is a permutation of $\{1,2,\ldots,m\}$ and σD obtains from D by simultaneously permuting rows and columns according to σ (that is $\sigma D = e_{i,j}$ where $e_{i,j} = d_{\sigma^{-1}(i),\sigma^{-1}(j)}$) then $w(D) = w(\sigma D)$.

SN: if D is an $m\times m$ $\{0,1\}$ matrix and $\overline{D}$ obtains from D upon replacing every 1 by 0 and every 0 by 1, then $w(D) = w(\overline{D})$.

8 w_1 satisfies conditions (9): (i) and (ii) are obvious, and (iii) holds since for a state description

$$\Theta(a_1,a_2,\ldots,a_m)) = \bigwedge_{i=1}^{m} P^{\epsilon_i}(a_i)$$

of L_1,

$$\sum_{\Phi(a_1,\ldots,a_{m+1})\models\Theta(a_1,\ldots,a_m)} w_1(\Phi(a_1,a_2,\ldots,a_{m+1})) \;=$$

$$w_1\left(\bigwedge_{i=1}^{m} P^{\epsilon_i}(a_i) \wedge P(a_{m+1})\right) + w_1\left(\bigwedge_{i=1}^{m} P^{\epsilon_i}(a_i) \wedge \neg P(a_{m+1})\right) \quad =$$

$$w_2\left(\bigwedge_{i=1}^{m} R^{\epsilon_i}(a_{2i+1}, a_{2i+2}) \wedge (R(a_{2m+1}, a_{2m+2}) \vee \neg R(a_{2m+1}, a_{2m+2}))\right) =$$

$$w_2\left(\bigwedge_{i=1}^{m} R^{\epsilon_i}(a_{2i+1}, a_{2i+2})\right) = w_1(\Theta(a_1, a_2, \ldots, a_m)).$$

3 Extending Probability Functions from QFSL to all sentences

Theorem 5 (Gaifman's Theorem) *Suppose that $w^- : QFSL \to [0,1]$ satisfies (P1) and (P2) for $\theta, \phi \in QFSL$. Then w^- has a unique extension to a probability function w on SL satisfying (P1-3) for any $\theta, \phi, \exists x\, \psi(x) \in SL$. Furthermore if w^- satisfies Ex, Px, SN (respectively) on $QFSL$ then so will its extension w to SL.*

Proof Let w^- be as in the statement of the theorem. For $\theta \in QFSL$ the subsets

$$[\theta] = \{\, M \in \mathcal{T}L \mid M \vDash \theta \,\}$$

of $\mathcal{T}L$ form an algebra, $\mathcal{A}$ say, of sets and μ_{w^-} defined by

$$\mu_{w^-}([\theta]) = w^-(\theta) \qquad \text{for } \theta \in QFSL$$

is easily seen to be a finitely additive measure on this algebra.

Indeed μ_{w^-} is (trivially) a pre-measure. For suppose $\theta, \phi_i \in QFSL$ for $i \in \mathbb{N}$ with the $[\phi_i]$ disjoint and

$$\bigcup_{i \in \mathbb{N}} [\phi_i] = [\theta]. \tag{15}$$

Then it must be the case that for some finite n

$$\bigcup_{i \leq n} [\phi_i] = [\theta],$$

otherwise

$$\{\, \neg\phi_i \mid i \in \mathbb{N} \,\} \cup \{\theta\}$$

would be finitely satisfiable and hence, by the Compactness Theorem for the Predicate Calculus, would be satisfiable in some structure for L. Although this particular structure need not be in $\mathcal{T}L$ its substructure with universe the $\{a_1, a_2, a_3, \ldots\}$ will be, and will satisfy the same quantifier free sentences, thus contradicting (15). So from the disjointness of the $[\phi_i]$ we must have that $[\phi_i] = \emptyset$ for $i > n$ (so $\mu_{w^-}([\phi_i]) = 0$), giving

$$\mu_{w^-}([\theta]) = \sum_{i \leq n} \mu_{w^-}([\phi_i]) = \sum_{i \in \mathbb{N}} \mu_{w^-}([\phi_i]),$$

and confirming the requirement to be a pre-measure.

Hence by Carathéodory's Extension Theorem (see for example [2]) there is a unique extension μ_w of μ_{w^-} defined on the σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$. Notice that for $\exists x\, \psi(x) \in SL$ (where there may be some constants appearing in $\psi(x)$)

$$\begin{aligned}
[\exists x\, \psi(x)] &= \{\, M \in \mathcal{T}L \mid M \vDash \exists x\, \psi(x)\,\} \\
&= \{\, M \in \mathcal{T}L \mid M \vDash \psi(a_i),\ \text{some } i \in \mathbb{N}^+\,\} \\
&= \bigcup_{i \in \mathbb{N}^+} \{\, M \in \mathcal{T}L \mid M \vDash \psi(a_i)\,\} \\
&= \bigcup_{i \in \mathbb{N}^+} [\psi(a_i)] \qquad (16)
\end{aligned}$$

so since $\mathcal{B}$ is closed under complements and countable unions $\mathcal{B}$ contains all the sets $[\theta]$ for $\theta \in SL$.

Now define a function w on SL by setting

$$w(\theta) = \mu_w([\theta]).$$

Notice that w extends w^- as μ_w extends μ_{w^-}. Since μ_w is a measure w satisfies (P1-2) and also (P3) from (16) and the fact that μ_w is countably additive.

This probability function must be the unique extension of w^- to SL satisfying (P1-3). For suppose that there was another such probability function, u say. By property (Pd) it is enough to show that u and w agree on sentences θ in Prenex Normal Form. This can be done by induction on the quantifier complexity of θ, see [1] for some technical details.[1]

[1] The induction is straightforward on the basis of the following result (Lemma 3.8 from [1]):

For $\exists x_1, \ldots, x_k\, \theta(x_1, \ldots, x_k, \vec{a}) \in SL$ and w a probability function on SL,

$$w(\exists x_1, \ldots, x_k\, \theta(x_1, \ldots, x_k, \vec{a})) = lim_{n\to\infty}\, w\left(\bigvee_{i_1, i_2, \ldots, i_k \leq n} \theta(a_{i_1}, a_{i_2}, \ldots, a_{i_k}, \vec{a})\right). \qquad (17)$$

.

The last part for Ex can also be shown by this method but alternatively we can argue as follows: Assume that w satisfies Ex on QFSL. Let $\theta(a_1, \dots, a_m) \in SL$ and let $b_1, \dots, b_m$ be distinct constants: $b_j = a_{k_j}$. Let σ be a permutation of $\mathbb{N}^+$ such that $\sigma(j) = k_j$ for $j = 1, \dots m$, so $b_j = a_{\sigma(j)}$ (such a permutaton clearly exists). The function $v : SL \to [0, 1]$ defined by

$$v(\phi(a_{i_1}, a_{i_2}, \dots, a_{i_n})) = w(\phi(a_{\sigma(i_1)}, a_{\sigma(i_2)}, \dots, a_{\sigma(i_n)}))$$

is also a probability function which agrees with w on QFSL. Since the extension is unique, $v = w$ on SL and hence in particular

$$w(\theta(a_1, \dots, a_n)) = v(\theta(a_1, \dots, a_n)) = w(\theta(a_{\sigma(1)}, \dots, a_{\sigma(n)})) = w(\theta(b_1, \dots, b_n))$$

showing that w satisfies Ex on the whole of SL.

The cases of Px and SN are similar. □

Corollary 6 *Suppose that w is a probability function on SL. Then for some countably additive measure μ_w on the algebra $\mathcal{B}$ of subsets of $\mathcal{T}L$,*

$$w = \int_{\mathcal{T}L} V_M \, d\mu_w.$$

Proof Let $u^- = w \restriction QFSL$. From the proof of the above theorem, there is a countably additive measure μ_u on $\mathcal{B}$ such that for $\theta \in SL$,

$$u(\theta) = \mu_u([\theta]) = \int_{\mathcal{T}L} V_M(\theta) \, d\mu_u.$$

Since u and w coincide on QFSL, by uniqueness u and w must be the same on SL and we can take $\mu_w = \mu_u$. □

This is the result we need to prove the promised converse to the Dutch Book Theorem (Theorem 3), see Problem 9.

Unary Pure Inductive Logic

Pure Inductive Logic was first developed for unary languages (Johnson, Carnap). We will now survey the most significant results within this context; all relation symbols $R_1, R_2, \dots, R_q$ in L are assumed to be unary. (As such they are referred to more often as predicate rather than relation symbols.)

By $\alpha_1(x), \alpha_2(x), \dots, \alpha_{2^q}(x)$ we denote the 2^q *atoms* of L, that is, the formulae of the form

$$\pm R_1(x) \wedge \pm R_2(x) \wedge \dots \wedge \pm R_q(x).$$

These atoms are pairwise disjoint (exclusive) and exhaustive, that is,

$$\text{For } i \neq k, \;\; \vDash \forall x \, \neg(\alpha_i(x) \wedge \alpha_k(x)) \qquad \text{and} \qquad \vDash \forall x \bigvee_{j=1}^{2^q} \alpha_j(x).$$

We list them in the lexicographic order with + before - so for example when $L = \{R_1, R_2, R_3\}$, we have

$$\begin{array}{ll} \alpha_1(x) = R_1(x) \wedge R_2(x) \wedge R_3(x), & \alpha_2(x) = R_1(x) \wedge R_2(x) \wedge \neg R_3(x), \\ \alpha_3(x) = R_1(x) \wedge \neg R_2(x) \wedge R_3(x), & \alpha_4(x), = R_1(x) \wedge \neg R_2(x) \wedge \neg R_3(x), \\ \alpha_5(x) = \neg R_1(x) \wedge R_2(x) \wedge R_3(x), & \alpha_6(x) = \neg R_1(x) \wedge R_2(x) \wedge \neg R_3(x), \\ \alpha_7(x) = \neg R_1(x) \wedge \neg R_2(x) \wedge R_3(x), & \alpha_8(x) = \neg R_1(x) \wedge \neg R_2(x) \wedge \neg R_3(x). \end{array}$$

A state description $\Theta(b_1, b_2, \dots, b_m)$ has the form

$$\bigwedge_{j=1}^{q} \bigwedge_{i=1}^{m} \pm R_j(b_i) \;\equiv\; \bigwedge_{i=1}^{m} \alpha_{h_i}(b_i).$$

In the unary context Ex can be expressed in a particularly simple way. For a state description $\bigwedge_{i=1}^m \alpha_{h_i}(b_i)$, define its *signature* to be the vector $\langle m_1, m_2, \dots, m_{2^q} \rangle$, owhere $m_j = |\{i \mid h_i = j\}|$.

Constant Exchangeability (Unary Version): $w\left(\bigwedge_{i=1}^m \alpha_{h_i}(b_i)\right)$ *depends only on the signature of* $\bigwedge_{i=1}^m \alpha_{h_i}(b_i)$.

Informally, this is because Ex says that it does not matter which $b_1, \dots, b_m$ figure in $\bigwedge_{i=1}^m \alpha_{h_i}(b_i)$ and the order of conjuncts does not matter by (Pd).

[Formally, assume that the value of w on $\bigwedge_{i=1}^n \alpha_{h_i}(b_i)$ depends only on the signature of $\bigwedge_{i=1}^n \alpha_{h_i}(b_i)$. Then the condition from Lemma 4 holds since for a permutation τ of $\{1, 2, \dots, m\}$, the state descriptions $\bigwedge_{i=1}^m \alpha_{h_i}(a_i)$ and $\bigwedge_{i=1}^m \alpha_{h_i}(a_{\tau(i)})$ have the same signature. Hence by that Lemma w satisfies Ex on $QFSL$ and by Gaifman's theorem also on the whole of SL.

Conversely, if w satisfies Ex and $a_{k_1}, \dots, a_{k_m}$, $a_{j_1}, \dots, a_{j_m}$ are distinct constants and $\Phi = \bigwedge_{i=1}^m \alpha_{h_i}(a_{k_i})$, $\Theta = \bigwedge_{i=1}^m \alpha_{g_i}(a_{j_i})$ are state descriptions with the same signature then there is a bijection $\sigma : \{k_1, \ldots, k_m\} \to \{j_1, \ldots, j_m\}$ extendable to a permutation σ of $\mathbb{N}^+$ such that $\bigwedge_{i=1}^m \alpha_{g_i}(a_{j_i}) = \bigwedge_{i=1}^m \alpha_{h_i}(a_{\sigma(k_i)})$. By Ex, w gives the same values to Φ, Θ as required.]

Functions $w_{\vec{c}}$

Let

$$\mathbb{D}_{2^q} = \{ \langle x_1, x_2, \ldots, x_{2^q} \rangle \in \mathbb{R}^{2^q} \mid x_1, \ldots, x_{2^q} \geq 0, \ \sum_{i=1}^{2^q} x_i = 1 \}$$

and

$$\vec{c} = \langle c_1, c_2, \ldots, c_{2^q} \rangle \in \mathbb{D}_{2^q}.$$

Define $w_{\vec{c}}$ by setting

$$w_{\vec{c}} \left(\bigwedge_{i=1}^m \alpha_{h_i}(a_i) \right) = \prod_{i=1}^m w_{\vec{c}}(\alpha_{h_i}(a_i)) = \prod_{i=1}^m c_{h_i} = \prod_{j=1}^{2^q} c_j^{m_j} \tag{18}$$

where $m_j = |\{ i \mid h_i = j \}|$ for $j = 1, 2, \ldots, 2^q$. Then conditions (9) are satisfied, so $w_{\vec{c}}$ extends uniquely to a probability function on $QFSL$ satisfying (P1-2) - and hence to a probability function on SL - via

$$w_{\vec{c}} \left(\bigwedge_{i=1}^m \alpha_{h_i}(b_i) \right) = \sum_{\Phi(a_1,\ldots,a_n) \models \bigwedge_{i=1}^m \alpha_{h_i}(b_i)} w_{\vec{c}}(\Phi(a_1, \ldots, a_n)),$$

where the Φ are state descriptions as usual. This gives again

$$w_{\vec{c}} \left(\bigwedge_{i=1}^m \alpha_{h_i}(b_i) \right) = \prod_{i=1}^m c_{h_i}$$

as apparent from the following example:

$$w_{\vec{c}}\,(\alpha_1(a_2) \wedge \alpha_3(a_4)) = \sum_{\Phi(a_1,\ldots,a_4) \models \alpha_1(a_2) \wedge \alpha_3(a_4)} w_{\vec{c}}(\Phi(a_1, \ldots, a_4))$$

$$= \sum_{k,j=1}^{2^q} w_{\vec{c}}(\alpha_k(a_1) \wedge \alpha_1(a_2) \wedge \alpha_j(a_3) \wedge \alpha_3(a_4)) = \sum_{k,j=1}^{2^q} c_k c_1 c_j c_3 =$$

$$= \left(\sum_{k=1}^{2^q} c_k \right) c_1 \left(\sum_{j=1}^{2^q} c_k \right) c_3 = c_1 c_3.$$

Clearly, $w_{\vec{c}}$ satisfies Ex. However, Px and SN hold only for special choices of $\vec{c}$, see Problem 11. They do satisfy the following strong independence condition:

The Constant Irrelevance Principle, IP

If $\theta, \phi \in QFSL$ have no constant symbols in common then[1]

$$w(\theta \wedge \phi) = w(\theta) \cdot w(\phi)$$

Proposition 7 *Let w be a probability function on SL satisfying Ex. Then w satisfies IP just if $w = w_{\vec{c}}$ for some $\vec{c} \in \mathbb{D}_{2^q}$.*

Proof First notice that for $\vec{c} \in \mathbb{D}_{2^q}$ and state descriptions $\bigwedge_{i=1}^n \alpha_{h_i}(a_{j_i})$, $\bigwedge_{i=1}^m \alpha_{g_i}(a_{k_i})$ with no constant symbols in common,

$$\begin{aligned} w_{\vec{c}}\left(\bigwedge_{i=1}^n \alpha_{h_i}(a_{j_i}) \wedge \bigwedge_{i=1}^m \alpha_{g_i}(a_{k_i})\right) &= \prod_{i=1}^n c_{h_i} \cdot \prod_{i=1}^m c_{g_i} \\ &= w_{\vec{c}}\left(\bigwedge_{i=1}^n \alpha_{h_i}(a_{j_i})\right) \cdot w_{\vec{c}}\left(\bigwedge_{i=1}^m \alpha_{g_i}(a_{k_i})\right). \end{aligned}$$

Hence if $\theta, \phi \in QFSL$ have no constant symbols in common and $\theta \equiv \bigvee_{\Theta \in S} \Theta$, $\phi \equiv \bigvee_{\Phi \in T} \Phi$ with the Θ, Φ state descriptions (for the a_i in θ, ϕ respectively) then

$$\begin{aligned} w_{\vec{c}}(\theta \wedge \phi) &= w_{\vec{c}}\left(\bigvee_{\Theta \in S} \Theta \wedge \bigvee_{\Phi \in T} \Phi\right) = w_{\vec{c}}\left(\bigvee_{\Theta \in S} \bigvee_{\Phi \in T} \Theta \wedge \Phi\right) \\ &= \sum_{\Theta \in S} \sum_{\Phi \in T} w_{\vec{c}}(\Theta \wedge \Phi) = \sum_{\Theta \in S} \sum_{\Phi \in T} w_{\vec{c}}(\Theta) \cdot w_{\vec{c}}(\Phi) \\ &= \sum_{\Theta \in S} w_{\vec{c}}(\Theta) \cdot \sum_{\Phi \in T} w_{\vec{c}}(\Phi) = w_{\vec{c}}(\theta) \cdot w_{\vec{c}}(\phi). \end{aligned} \tag{19}$$

Conversely if w satisfies Ex and IP then by repeated application

[1] We remark that in the presence of Ex, changing QFSL to SL produces an equivalent principle, see [1, Lemma 6.2].

$$w\left(\bigwedge_{i=1}^{n}\alpha_{h_i}(a_{j_i})\right)=\prod_{i=1}^{n}w(\alpha_{h_i}(a_{j_i}))=\prod_{i=1}^{n}w(\alpha_{h_i}(a_1))=\prod_{i=1}^{n}c_{h_i}$$

where $c_i = w(\alpha_i(a_1))$ for $i = 1, 2, \ldots, 2^q$. Since w is determined by its values on state descriptions this forces $w = w_{\vec{c}}$, as required. $\square$

Problems

Problem 9 *Using Corollary 6, prove Theorem 3 .*

Problem 10 *Let* $\lambda > 0$*. Show that there is a unique probability function* w *such that for any* $0 \leq j \leq 2^q$ *and any state description* $\bigwedge_{i=1}^m \alpha_{h_i}(a_i)$

$$w\left(\alpha_j(a_{m+1}) \mid \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right) = \frac{m_j + \lambda 2^{-q}}{m + \lambda} \tag{20}$$

where $m_j = |\{i \mid h_i = j\}|$*. Show that this* w *satisfies Ex.*

Problem 11 *Let* L *contain just two unary predicates.*

(a) Write down conditions under which $w_{\vec{c}}$ *satisfy Px and SN respectively.*

(b) Find $\vec{c}, \vec{d}$ *such that* $w_{\vec{c}}$ *does not satisfy Px but* $\frac{1}{2}(w_{\vec{c}} + w_{\vec{d}})$ *does.*

Problem 12 *Let* L *be a language with a single, unary, predicate* Q*. Let* $N \in \mathbb{N}$ *and let* $\vec{d} = \langle d_1, d_2, \ldots, d_N \rangle$ *be a* $\{0,1\}$*-vector. Define* $w^{\vec{d}}$ *for state descriptions by setting* $w^{\vec{d}}(\top) = 1$ *and*

$$w^{\vec{d}}\left(\bigwedge_{i \leq m} Q^{t_i}(a_i)\right)$$

to be the probability of (uniformly) randomly picking, with replacement, $h_1, h_2, \ldots, h_m$ *from* $\{1, 2, \ldots, N\}$ *such that for each* $i \leq m$*,*

$$d_{h_i} = t_i.$$

Show that this uniquely determines a probability function on SL *satisfying Ex, and that this function is one of the* $w_{\vec{x}}$*.*

Solutions

9 For the unconditional result, suppose on the contrary that there were countable sets A, B etc. such that (7), (6) held. Let μ_{Bel} be the measure such that for $\theta \in SL$

$$Bel(\theta) = \int_{\mathcal{T}L} V_M(\theta)\, d\mu_{Bel}$$

We have

$$\sum_{i \in A} s_i(Bel(\theta_i) - p_i) + \sum_{i \in B}(-t_i)(Bel(\phi_i) - q_i)$$

$$= \sum_{i \in A} s_i \int_{\mathcal{T}L} (V_M(\theta_i) - p_i)\, d\mu_{Bel}(M) \;+ \sum_{i \in B}(-t_i) \int_{\mathcal{T}L} (V_M(\phi_i) - q_i)\, d\mu_{Bel}(M)\,.$$

By (7) and Lebesgue's Dominated Convergence Theorem, this further equals

$$\int_{\mathcal{T}L} \left(\sum_{i \in A} s_i(V_M(\theta_i) - p_i) + \sum_{i \in B}(-t_i)(V_M(\phi_i) - q_i) \right) d\mu_{Bel}$$

By (6) this is strictly negative. Hence also the expression we have started with is strictly negative, but that is impossible since

$$Bel(\theta_i) - p_i > 0, \quad Bel(\phi_i) - q_i < 0$$

The proof for the conditional version is similar.

10 Consider the values such a function must give to state descriptions. Note that any probability function v satisfies

$$v\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i) \right) = \prod_{j=1}^{n} v\left(\alpha_{h_j}(a_j) \mid \bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i) \right)$$

(where $\bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i)$ for $j = 1$ stands for $\top$). Accordingly, for a state description

$$\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)$$

let r_j be the number of times that h_j occurs amongst $h_1, h_2, \ldots, h_{j-1}$ and with a view to use (9), define $w(\top) = 1$ and

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \prod_{j=1}^{n}\left(\frac{r_j + \lambda 2^{-q}}{j - 1 + \lambda}\right)$$

Then the conditions (i) and (ii) from (9) are clearly satisfied. For (iii) note that if

$$\Theta(a_1, a_2, \ldots, a_n) = \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)$$

then

$$\begin{aligned}\sum_{\Phi(a_1,\ldots,a_{n+1}) \models \Theta(a_1,\ldots,a_n)} w(\Phi(a_1, a_2, \ldots, a_{n+1})) &= \sum_{k=1}^{2^q} w\left(\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) \wedge \alpha_k(a_{n+1})\right) \\ &= \sum_{k=1}^{2^q} w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) \cdot \left(\frac{m_k + \lambda 2^{-q}}{n + \lambda}\right)\end{aligned}$$

so since the m_k sum to n, (iii) holds, too. The existence and uniqueness of an extension to a probability function on SL follows. Furthermore, the extension satisfies Ex by Lemma 4 since from the above it can be seen that

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \frac{\prod_{k=1}^{2^q} \prod_{j=0}^{m_k - 1}(j + \lambda 2^{-q})}{\prod_{j=0}^{n-1}(j + \lambda)}$$

and this expression depends only on the signature $\langle m_1, \ldots, m_{2^q}\rangle$ of $\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)$.

11 Let $L = \{P, Q\}$ and, as usual,

$\alpha_1(x) = P(x) \wedge Q(x)$, $\alpha_2(x) = P(x) \wedge \neg Q(x)$, $\alpha_3(x) = \neg P(x) \wedge Q(x)$, $\alpha_4(x) = \neg P(x) \wedge \neg Q(x)$

Let $\vec{c} = \langle c_1, c_2, c_3, c_4\rangle$. If $w_{\vec{c}}$ satisfies Px we must have

$$w_{\vec{c}}(\alpha_2(a_1)) = w_{\vec{c}}(\alpha_3(a_1)).$$

Hence $c_2 = c_3$. This condition is sufficient since permuting P and Q in a state description amounts to swapping α_2 and α_3, so any $w_{\vec{c}}$ with $c_2 = c_3$ satisfies Px for state descriptions, and hence on SL.

If $w_{\vec{c}}$ satisfies SN then it gives equal value to all atoms (since any atom can be transformed to any other by adding or removing negations). Since, moreover,

$$1 = w(\top) = w\left(\bigwedge_{i=1}^{4} \alpha_i(a_1)\right)$$

$\vec{c}$ must be $\langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4} \rangle$.

(b) For any $\vec{c}$ with $c_2 \neq c_3$, $w_{\vec{c}}$ does not satisfy Px but if $\vec{d} = \langle c_1, c_3, c_2, c_4 \rangle$ then $\frac{1}{2}(w_{\vec{c}} + w_{\vec{d}})$ does: It suffices to check it for state descriptions, so if $\Theta = \bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)$ has signature $\langle m_1, m_2, m_3, m_4 \rangle$ then swapping P and Q produces a state description Θ' obtained from Θ by swapping α_2 with α_3 everywhere and hence a state description with signature $\langle m_1, m_3, m_2, m_4 \rangle$. We have

$$\frac{1}{2}(w_{\vec{c}} + w_{\vec{d}})(\Theta) = \frac{1}{2}(c_1^{m_1} c_2^{m_2} c_3^{m_3} c_4^{m_4} + c_1^{m_1} c_3^{m_2} c_2^{m_3} c_4^{m_4}) = \frac{1}{2}(w_{\vec{c}} + w_{\vec{d}})(\Theta')$$

as required.

12 Let

$$c = \frac{|h \in \{1, \dots, N\} : \quad d_h = 1.\}|}{N}.$$

For

$$\Theta(a_1, \dots, a_m) = \bigwedge_{i=1}^{m} Q^{t_i}(a_i),$$

$w^{\vec{d}}(\Theta(a_1, \dots, a_m))$ is the ratio

$$\frac{|\langle h_1, \dots, h_m \rangle \in \{1, \dots, N\}^m : \text{ for all } i \leq m, \ d_{h_i} = t_i\}|}{N^m}$$

$$= \prod_{i=1}^{m} \frac{|h \in \{1, \dots, N\} : \quad d_h = t_i.\}|}{N} = c^{m_1}(1-c)^{m_2}$$

where $m_1 = |\{i \in \{1, \dots, m\} : \ t_i = 1\}|$ and $m_2 = |\{i \in \{1, \dots, m\} : \ t_i = 0\}|$. Hence $w^{\vec{d}} = w_{\langle c, 1-c \rangle}$.

4 de Finetti's Representation Theorem

Let $L = \{R_1, \dots, R_q\}$ be a unary language and let w be a probability function on SL satisfying Ex. Then there is a (normalized, countably additive) measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ such that

$$\begin{aligned} w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) &= \int_{\mathbb{D}_{2^q}} \prod_{j=1}^{2^q} x_j^{m_j}\, d\mu(\vec{x}) \\ &= \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) d\mu(\vec{x}), \end{aligned} \tag{21}$$

where $m_j = |\{i \,|\, h_i = j\}|$ for $j = 1, 2, \dots, 2^q$.

Conversely, given a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ the function w defined by (21) extends uniquely to a probability function on SL satisfying Ex.

Proof We will prove the result for $q = 1$, the full case being similar. For $q = 1$ there are just two atoms, $\alpha_1(x) = R_1(x)$ and $\alpha_2(x) = \neg R_1(x)$. Hence the signature of a state descriptions $\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$ is $\langle m_1, m_2 \rangle$ where

$$m_1 = |\{i \,|\, h_i = 1\}|, \quad m_2 = |\{i \,|\, h_i = 2\}| \quad (\text{with } m_1 + m_2 = m).$$

In such case we write unambiguously

$$w(m_1, m_2) = w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right). \tag{22}$$

Let $r > m$. For fixed r_1, r_2 with $r_1 + r_2 = r$ there are $\binom{r}{r_1}$ state descriptions for $b_1, \dots, b_r$ with signature $\langle r_1, r_2 \rangle$. Since state descriptions for $b_1, \dots, b_r$ are exclusive and exhaustive, we have

$$1 = w(\top) = \sum_{r_1 + r_2 = r} \binom{r}{r_1} w(r_1, r_2). \tag{23}$$

Also, noting that when $r_1 + r_2 = r$ and $r_1 \geq m_1$, $r_2 \geq m_2$ there are $\binom{r-m}{r_1 - m_1}$ state descriptions extending any given state description for $b_1, \dots, b_m$ with signature $\langle m_1, m_2 \rangle$,

$$w(m_1,m_2)=\sum_{\substack{r_1+r_2=r\\ m_1\le r_1,\, m_2\le r_2}} \binom{r-m}{r_1-m_1} w(r_1,r_2). \tag{24}$$

From (23) let μ_r be the discrete measure on $\mathbb{D}_2$ which puts measure

$$\binom{r}{r_1} w(r_1,r_2)$$

on the point $\langle r_1/r, r_2/r\rangle \in \mathbb{D}_2$. From (24) we obtain

$$w(m_1,m_2)=\sum_{\substack{r_1+r_2=r\\ m_1\le r_1,\, m_2\le r_2}} \binom{r-m}{r_1-m_1}\binom{r}{r_1}^{-1}\binom{r}{r_1} w(r_1,r_2). \tag{25}$$

We shall show that

$$\left|\binom{r-m}{r_1-m_1}\binom{r}{r_1}^{-1} - \left(\frac{r_1}{r}\right)^{m_1}\left(\frac{r_2}{r}\right)^{m_2}\right| \tag{26}$$

tends to 0 as $r\to\infty$ uniformly in r_1, r_2. Notice that

$$\left(\frac{r_1}{r}\right)^{m_1}\left(\frac{r_2}{r}\right)^{m_2} = w_{\langle r_1/r, r_2/r\rangle}(m_1,m_2).$$

The left hand term in (26) can be written as

$$\left(\frac{r_1}{r}\right)^{m_1}\left(\frac{r_2}{r}\right)^{m_2} \frac{(1-r_1^{-1})\cdots(1-(m_1-1)r_1^{-1})(1-r_2^{-1})\cdots(1-(m_2-1)r_2^{-1})}{(1-r^{-1})\cdots(1-(m-1)r^{-1})}. \tag{27}$$

We now consider cases.

- If $m_1=m_2=0$ then (26) is zero.
- If $m_2>0$ and $r_2\le\sqrt{r}$ then both terms in (26) are less than $r^{-m_2/2}\le r^{-1/2}$

(this is justified for large r e.g because for any $r_1\le r$, $s\le m_1$ we have $\frac{1-sr_1^{-1}}{1-sr^{-1}}\le 1$ and for $\sqrt{r}>m$, $s\le m_2$, we have $\frac{1-sr_2^{-1}}{1-(m_1+s)r^{-1}}\le 1$)

and similarly if $m_1>0$ and $r_1\le\sqrt{r}$. If $m_2>0$ and $r_2>\sqrt{r}$ and either $m_1=0$ or $r_1>\sqrt{r}$ then using (27) and the fact that $r_1/r, r_2/r\le 1$ we see that (26) is at most

$$1-\frac{(1-\sqrt{r}^{\,-1})\dots(1-(n-1)\sqrt{r}^{\,-1})(1-\sqrt{r}^{\,-1})\dots(1-(k-1)\sqrt{r}^{\,-1})}{(1-r^{-1})\dots(1-(n+k-1)r^{-1})}.$$

Similarly if $m_1 > 0$ and $r_1 > \sqrt{r}$ and either $m_2 = 0$ or $r_2 > \sqrt{r}$, and together we have covered all cases.

Hence from (23) and (25) $w(m_1, m_2)$ equals the limit as $r \to \infty$ of

$$\sum_{\substack{r_1+r_2=r \\ m_1 \le r_1,\, m_2 \le r_2}} \left(\frac{r_1}{r}\right)^{m_1} \left(\frac{r_2}{r}\right)^{m_2} \mu_r(\{\langle r_1/r, r_2/r\rangle\}). \tag{28}$$

In turn this equals the limit of the same expressions but summed simply over $0 \le r_1, r_2$, $r_1 + r_2 = r$ since from (23) (or trivially if $m_1 = 0$),

$$\sum_{\substack{r_1+r_2=r \\ r_1 < m_1,\, m_2 \le r_2}} \left(\frac{r_1}{r}\right)^{m_1} \left(\frac{r_2}{r}\right)^{m_2} \mu_r(\{\langle r_1/r, r_2/r\rangle\}), \quad \text{etc.}$$

tends to zero as $r \to \infty$.

In other words,

$$w(m_1, m_2) = \lim_{r \to \infty} \int_{\mathbb{D}_2} x_1^{m_1} x_2^{m_2}\, d\mu_r(\langle x_1, x_2\rangle). \tag{29}$$

By Prohorov's Theorem, see for example [4, Theorem 5.1], since $\mathbb{D}_2$ is compact the μ_r have a subsequence μ_{i_r} weakly convergent to a countably additive measure μ, meaning that for any continuous function $f(x_1, x_2)$

$$\lim_{r \to \infty} \int_{\mathbb{D}_2} f(x_1, x_2)\, d\mu_{i_r}(\langle x_1, x_2\rangle) = \int_{\mathbb{D}_2} f(x_1, x_2)\, d\mu(\langle x_1, x_2\rangle).$$

Using this the required result follows from (29).

Finally the converse result, that functions w defined by (21) extend to probability functions on SL satisfying Ex follows by checking conditions (9) and by Theorem 5. □

From (21) it follows that the integrals

$$\int_{\mathbb{D}_2} f(x_1, x_2)\, d\mu(\langle x_1, x_2\rangle)$$

are uniquely determined by w for *any* polynomial $f(x_1, x_2)$, and hence (see for example [3]) that μ must be the unique measure satisfying (21). We shall call this measure the *de Finetti prior of* w.

de Finetti's Theorem generalizes directly to SL and indeed in what follows we shall use that name in this extended sense. Precisely:

Corollary 8 *Let w be a probability function on SL satisfying Ex. Then there is a measure μ on $\mathbb{D}_{2^q}$ (the de Finetti prior of w in fact) such that for $\theta \in SL$,*

$$w(\theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta)\, d\mu(\vec{x}). \tag{30}$$

Conversely given a measure μ on $\mathbb{D}_{2^q}$, w defined by (30) is a probability function on SL satisfying Ex.

In other words every probability function w on SL is a convex mixture

$$w = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\, d\mu(\vec{x}), \tag{31}$$

of the $w_{\vec{c}}$ for $\vec{c} \in \mathbb{D}_{2^q}$.

Proof de Finetti Theorem gives this for θ a state description, hence for $\theta \in QFSL$, and then in turn for any $\theta \in SL$ by induction on quantifier complexity and Lebesgue's Dominated Convergence Theorem. The converse follows by checking (P1-3) noting that the functions $\vec{x} \mapsto w_{\vec{x}}(\theta)$ are measurable. $\Box$

Further Unary[1] Principles

One well known principle, employed already by Carnap is

The Principle of Regularity, Reg

If $\theta \in QFSL$ is satisfiable then $w(\theta) > 0$.

Note that since any satisfiable $\theta \in QFSL$ is logically equivalent to a non-empty disjunction of state descriptions, Reg is equivalent to requiring that the probability of any state description is non-zero.

Requiring that *any* satisfiable sentence whatsoever has a non-zero probability is stronger than Reg. It is referred to as

The Principle of Super Regularity (Universal Certainty), SReg

If $\theta \in SL$ is satisfiable then $w(\theta) > 0$.

[1] We continue to work in the unary context, but the first two principles below, Reg ad SReg have the same formulation in general.

The principles have interesting characterisation in terms of the de Finetti priors of w: let μ be the de Finetti prior of w. Then w satisfies Reg just if

$$\mu(\{\,\vec{c} \in \mathbb{D}_{2^q} \mid c_1, c_2, \ldots, c_{2^q} > 0\,\}) > 0 \tag{32}$$

and w satisfies SReg just if for every $\emptyset \neq T \subseteq \{1, 2, \ldots, 2^q\}$,

$$\mu(\{\,\vec{c} \in \mathbb{D}_{2^q} \mid c_i > 0 \iff i \in T\,\}) > 0.$$

We omit the proof, see [1].

The next principle is an attempt at formalising the requirement that upon witnessing an instance of something occurring, one's belief in encountering it again should increase (or at least stay the same). We continue to assume that the language is unary.

The Principle of Instantial Relevance, PIR[1] *For* $\theta(a_1, a_2, \ldots, a_n) \in SL$ *and atom* $\alpha(x)$ *of* L*,*

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \theta(a_1, a_2, \ldots, a_n)). \tag{33}$$

Using de Finetti's theorem we can show that PIR is in fact is a consequence of Ex.

Theorem 9 *Ex implies PIR*

Proof We will write $\vec{a}$ for $a_1, a_2, \ldots, a_n$. Let the probability function w on SL satisfy Ex. Employing the notation of (33), let $\alpha(x) = \alpha_1(x)$ and denote $A = w(\theta(\vec{a}))$. Then for μ the de Finetti prior for w (using the fact that by Proposition 7 the $w_{\vec{x}}$ satisfy IP)

$$A = w(\theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\,,$$

$$w(\alpha_1(a_{n+1}) \wedge \theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x}),$$

$$w(\alpha_1(a_{n+2}) \wedge \alpha_1(a_{n+1}) \wedge \theta(\vec{a})) = \int_{\mathbb{D}_{2^q}} x_1^2 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})$$

[1]Recall the convention that expressions like $\frac{w(\phi)}{w(\psi)} = \frac{w(\theta)}{w(\eta)}$ stand for $w(\phi)w(\eta) = w(\theta)w(\psi)$ so denominators can be 0.

and (33) amounts to

$$\left(\int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right) \cdot \left(\int_{\mathbb{D}_{2^q}} x_1^2 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right) \;\geq\; \left(\int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right)^2 . \tag{34}$$

If $A = 0$ then this clearly holds (because the other two integrals are less or equal to A and greater equal zero) so assume that $A \neq 0$. In that case (34) is equivalent to

$$\int_{\mathbb{D}_{2^q}} \left(x_1 A - \int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right)^2 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x}) \;\geq\; 0 \tag{35}$$

as can be seen by multiplying out the square and dividing by A. But obviously, being an integral of a non-negative function, (35) holds, as required.

□

We remark that the above proof can be modified to show that Ex implies also

The Extended Principle of Instantial Relevance, EPIR

For $\theta(a_1, a_2, \ldots, a_n), \psi(a_1) \in SL$,

$$w(\psi(a_{n+2}) \,|\, \psi(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\psi(a_{n+2}) \,|\, \theta(a_1, a_2, \ldots, a_n)). \tag{36}$$

The next principle is justified on the grounds of symmetry, similarly as Ex. Rather than symmetry between constants though in this case the claim is that in the situation of zero knowledge the atoms are interchangeable. Precisely:

The Atom Exchangeability Principle, Ax

For any permutation[1] τ *of* $\{1, 2, \ldots, 2^q\}$ *and constants* $b_1, b_2, \ldots, b_m$,

$$w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) = w\left(\bigwedge_{i=1}^{m} \alpha_{\tau(h_i)}(b_i)\right). \tag{37}$$

[1]We refer to a permutation τ of $\{1, 2, \ldots, 2^q\}$ also as permutation *of atoms*, meaning the permutation which sends α_i to $\alpha_{\tau(i)}$.

Equivalently, in the presence of Ex, Ax asserts that the left hand side of (37) depends only on the *spectrum*[1] of the state description $\bigwedge_{i=1}^m \alpha_{h_i}(b_i)$, that is on the *multiset* $\{m_1, m_2, \dots, m_{2^q}\}$, where, again, $m_j = |\{i \mid h_i = j\}|$.

For example, if $\vec{c} \in \mathbb{D}_{2^q}$ then

$$v_{\vec{c}} = |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma \in \mathsf{S}_{2^q}} w_{\langle c_{\sigma(1)}, c_{\sigma(2)}, \dots, c_{\sigma(2^q)} \rangle}, \tag{38}$$

where S_{2^q} is the set of all permutations of $\{1, 2, \dots, 2^q\}$, satisfies Ax (cf. Problem 16(a)). de Finetti's representation theorem can be modified to yield a representation theorem for probability functions satisfying Ex and Ax in terms of these $v_{\vec{c}}$ - see Problem 16(b) for a proof.

Theorem 10 *(Representation Theorem for Ax) Let L be a unary language with q relation symbols and let w be a probability function on SL satisfying Ax (and Ex). Then there is a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ such that*

$$w = \int_{\mathbb{D}_{2^q}} v_{\vec{x}} \, d\mu(\vec{x}). \tag{39}$$

Conversely, given a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ the probability function w on SL defined by (39) satisfies Ax (and Ex).

Quite different in motivation is the following principle (intended to be considered in the presence of Ex):

Reichenbach's Axiom, RA

Let $\alpha_{h_i}(x)$ for $i = 1, 2, 3, \ldots$ be an infinite sequence of atoms of L. Then for $\alpha_j(x)$ an atom of L,

$$\lim_{n \to \infty} \left(w \left(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^n \alpha_{h_i}(a_i) \right) - \frac{u_j(n)}{n} \right) = 0 \tag{40}$$

[1]Note the difference between spectrum and signature of a state description: unlike the signature, spectrum does not code which atoms are which. Spectra are usually listed in decreasing order (not-necessarily strictly). If convenient we omit the zeros from the spectrum so spectra of state descriptions for m constants are multisets $\{n_1, n_2, \dots, n_k\}$ of strictly positive natural numbers with $\sum_{i=1}^k n_i = m$, denoted $\tilde{m}, \tilde{n}$ etc.

where $u_j(n) = |\{i \,|\, 1 \leq i \leq n \text{ and } h_i = j\}|$.

Informally, this asserts that as the number of constants, of which everything is known, grows, w should see this information as a *statistical sample* so that the probability that the next constant will satisfy the atom α_j and the frequency of past instances of $\alpha_j(a_i)$ get closer and closer.

We remark that although this may seem very common sense in situations where the the sequences $u_j(n)/n$ converge, the principle does not assume it.

RA again has an interesting characterisation in terms of the de Finetti prior: for w satisfing Reg, w satisfies RA if and only if every point in $\mathbb{D}_{2^q}$ is a support point of the de Finetti prior μ of w. We omit the proof, see [1].

The next principle draws on the idea that irrelevant information can/should be ignored. It has played a crucial role in Inductive Logic since its inception.

Johnson's Sufficientness Postulate, JSP

$$w\left(\alpha_j(a_{n+1}) \,|\, \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) \tag{41}$$

depends only on n *and* $r = |\{\, i \,|\, 1 \leq i \leq n \text{ and } h_i = j \,\}|$ *i.e. the number of times that* α_j *occurs amongst the* α_{h_i} *for* $i = 1, 2, \ldots, n$.

Note in particular that (41) treats all atoms α_j in the same way.

The functions defined in Problem 10 clearly satisfy JSP. We refer to them as Carnap continuum functions and denote them c^L_λ, so for $\lambda > 0$,

$$c^L_\lambda\left(\alpha_j(a_{n+1}) \,|\, \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \frac{m_j + \lambda 2^{-q}}{n + \lambda}$$

where $m_j = |\{i \,|\, 1 \leq i \leq n \text{ and } h_i = j\}|$.

The same expressions with 0 or ∞ in place of λ lead us to define[1] c_∞^L by

$$c_\infty^L\left(\bigwedge_{i=1}^n \alpha_{h_i}(a_i)\right) = 2^{-qn}$$

and c_0^L by

$$c_0^L\left(\bigwedge_{i=1}^n \alpha_{h_i}(a_i)\right) = \begin{cases} 2^{-q} & \text{if } h_1 = h_2 = \ldots = h_n, \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

[1]By taking $\lambda = \infty$ here and in similar situations we mean taking the limit as λ tends to ∞. Note that the result is in agreement with the general definition of c_∞^L we gave in the first example of the course. Both in case of c_∞^L and c_0^L, the definition is given for state descriptions but observing that conditions (9) hold and applying Gaifman's Theorem ensures that they are defined uniquely as probability functions on SL.

Problems

Problem 13 *(a) Let L contain just two unary predicates and assume that w satisfy Ex and that $w(\alpha_1(a_1)), w(\alpha_2(a_1)) > 0$. By considering*

$$\int_{\mathbb{D}_4} (bx_1 - cx_2)^2 \, d\mu(x)$$

for a suitable choice of constants c, b show that we cannot have both

$$w(\alpha_1(a_2) \,|\, \alpha_1(a_1)) < w(\alpha_1(a_2) \,|\, \alpha_2(a_1)) \ \ \textit{and} \ \ w(\alpha_2(a_2) \,|\, \alpha_2(a_1)) < w(\alpha_2(a_2) \,|\, \alpha_1(a_1)).$$

Give an example of a probability function w satisfying Ex for which the first of these does hold.

(b) Assume now that w satisfies Ex+SN. Show that

$$w(\alpha_1(a_1) \wedge \alpha_1(a_2)) = w(\alpha_2(a_1) \wedge \alpha_2(a_2)) \geq w(\alpha_1(a_1) \wedge \alpha_2(a_2))$$

and hence that in this case

$$w(\alpha_1(a_2) \,|\, \alpha_1(a_1)) \geq w(\alpha_1(a_2)) \,|\, \alpha_2(a_1)).$$

Show that if w satisfies Ex+SN+Px then we can only have equality here if $w = c_\infty$.

Problem 14 *Show that Ax implies Px and SN.*

Problem 15 *Let L be a unary language.*

(a) Show that any probability function w which satisfies SN also satisfies (37) *for $m = 1$, that is, for any constant b and any two atoms α_k, α_j,*

$$w\left(\alpha_k(b)\right) = w\left(\alpha_j(b)\right).$$

(b) Find a probability function which satisfies SN but not Ax.

Problem 16 *(a) Let L be a unary language and $\vec{c} \in \mathbb{D}_{2^q}$. Show that the function*

$$v_{\vec{c}} = |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma \in \mathsf{S}_{2^q}} w_{\langle c_{\sigma(1)}, c_{\sigma(2)}, \ldots, c_{\sigma(2^q)} \rangle},$$

where S_{2^q} is the set of all permutations of $\{1, 2, \ldots, 2^q\}$, satisfies Ax.

(b) Prove Theorem 10.

Solutions

13 Assume that

$w(\alpha_1(a_2)\,|\,\alpha_1(a_1)) < w(\alpha_1(a_2)\,|\,\alpha_2(a_1))$ and $w(\alpha_2(a_2)\,|\,\alpha_2(a_1)) < w(\alpha_2(a_2)\,|\,\alpha_1(a_1))$

do hold, and let μ be the de Finetti prior of w. The above inequalities yield

$$\frac{\int_{\mathbb{D}_4} x_1^2\,d\mu}{\int_{\mathbb{D}_4} x_1\,d\mu} < \frac{\int_{\mathbb{D}_4} x_1x_2\,d\mu}{\int_{\mathbb{D}_4} x_2\,d\mu}, \qquad \frac{\int_{\mathbb{D}_4} x_2^2\,d\mu}{\int_{\mathbb{D}_4} x_2\,d\mu} < \frac{\int_{\mathbb{D}_4} x_1x_2\,d\mu}{\int_{\mathbb{D}_4} x_1,d\mu}$$

so setting

$$b = \int_{\mathbb{D}_4} x_2\,d\mu, \qquad c = \int_{\mathbb{D}_4} x_1\,d\mu$$

we have

$$b\int_{\mathbb{D}_4} x_1^2\,d\mu \,<\, c\int_{\mathbb{D}_4} x_1x_2\,d\mu, \qquad c\int_{\mathbb{D}_4} x_2^2\,d\mu \,<\, b\int_{\mathbb{D}_4} x_1x_2\,d\mu,$$

Multiplying the first inequality by b, the second one by c and adding them yields

$$\int_{\mathbb{D}_4} (bx_1 - cx_2)^2\,d\mu(x) \,<\, 0,$$

contradiction.

To find a required example, after checking that the strict inequality fails for the $w_{\vec{x}}$, try $\frac{1}{2}(w_{\vec{x}} + w_{\vec{y}})$; in this case the first inequality anounts to

$$\frac{x_1^2 + y_1^2}{x_1 + y_1} < \frac{x_1x_2 + y_1y_2}{x_2 + y_2}$$

which simplifies to give

$$(x_1 - y_1)(x_1y_2 - y_1x_2) < 0.$$

This holds for example when $\vec{x} = \langle 0.1, 0.2, 0.3, 0.4\rangle$ and $\vec{y} = \langle 0.2, 0.6, 0.1, 0.1\rangle$.

(b) By SN, we can see that

$$w(\alpha_1(a_1)) = w(\alpha_2(a_1)), \quad w(\alpha_1(a_1) \wedge \alpha_1(a_2)) = w(\alpha_2(a_1) \wedge \alpha_2(a_2))$$

and by Ex,

$$w(\alpha_1(a_1) \wedge \alpha_2(a_2)) = w(\alpha_2(a_1) \wedge \alpha_1(a_2))$$

so the first claim follows from (a).

Assume

$$w(\alpha_1(a_2) \,|\, \alpha_1(a_1)) = w(\alpha_1(a_2)) \,|\, \alpha_2(a_1)),$$

so since $w(\alpha_1(a_1)) = w(\alpha_2(a_1))$,

$$w(\alpha_1(a_2) \wedge \alpha_1(a_1)) = w(\alpha_1(a_2)) \wedge \alpha_2(a_1)).$$

By SN, also

$$w(\alpha_2(a_2) \wedge \alpha_2(a_1)) = w(\alpha_2(a_2)) \wedge \alpha_1(a_1)),$$
$$w(\alpha_3(a_2) \wedge \alpha_3(a_1)) = w(\alpha_3(a_2)) \wedge \alpha_4(a_1)),$$
$$w(\alpha_4(a_2) \wedge \alpha_4(a_1)) = w(\alpha_4(a_2)) \wedge \alpha_3(a_1))$$

and by Px moreover

$$w(\alpha_1(a_2) \wedge \alpha_1(a_1)) = w(\alpha_1(a_2)) \wedge \alpha_3(a_1)),$$

$$w(\alpha_3(a_2) \wedge \alpha_3(a_1)) = w(\alpha_3(a_2)) \wedge \alpha_1(a_1)).$$

Writing out what these mean in terms of the se Finetti representation and adding suitable pairs of equalities, we obtain

$$\int_{\mathbb{D}_4} (x_1 - x_2)^2 \, d\mu(x) \; = \int_{\mathbb{D}_4} (x_3 - x_4)^2 \, d\mu(x) \; = \int_{\mathbb{D}_4} (x_1 - x_3)^2 \, d\mu(x) \; = \; 0,$$

which means that $x_1 = x_2$, $x_3 = x_4$ and $x_1 = x_3$ on $\mathbb{D}_4$ except possibly on a set of μ measure 0, so $x_1 = x_2 = x_3 = x_4$ except possibly on a null set, so w is $w_{\langle 1/4,1/4,1/4,1/4 \rangle} = c_\infty^L$ as required.

14 Assume that w satisfies Ax. Swapping R and $\neg R$, or R_i and R_j respectively each generates a permutation σ of atoms. If

$$\Theta(b_1, \ldots, b_n) = \bigwedge \alpha_{h_i}(b_i)$$

is a state description then $\Theta'(b_1, \ldots, b_n)$ obtained from $\Theta(b_1, \ldots, b_n)$ by swapping R and $\neg R$ or R_i and R_j respectively is

$$\Theta'(b_1, \ldots, b_n) = \bigwedge \alpha_{\sigma(h_i)}(b_i),$$

Hence by Ax, $w(\Theta'(b_1,\dots,b_n)) = w(\Theta'(b_1,\dots,b_n))$. Each of Px and SN follow e.g. by induction on quantifier complexity of sentences in disjunctive normal form.

15 (a) If w satisfies SN then it gives equal value to all atoms (since any atom can be transformed to any other by adding or removing negations).

(b) Let $L = \{R, Q\}$ where R, Q are unary. Let

$$w = \frac{1}{2}(w_{\langle 1/2,1/2,0,0\rangle} + w_{\langle 0,0,1/2,1/2\rangle}).$$

Recall that to check that w satisfies SN it suffices to check that is satisfies it for state descriptions. Note that

$$w_{\langle 1/2,1/2,0,0\rangle}\left(\bigwedge_{i=1}^{m} \alpha_{k_i}(b_i)\right) = \begin{cases} \frac{1}{2}\left(\frac{1}{2}\right)^m & \text{if } k_i \in \{1,2\} \text{ for all } m \\ 0 & \text{otherwise} \end{cases}$$

and similarly

$$w_{\langle 0,0,1/2,1/2\rangle}\left(\bigwedge_{i=1}^{m} \alpha_{k_i}(b_i)\right) = \begin{cases} \frac{1}{2}\left(\frac{1}{2}\right)^m & \text{if } k_i \in \{3,4\} \text{ for all } m \\ 0 & \text{otherwise} \end{cases}$$

Replacing R by $\neg R$ throughout a state description

$$\bigwedge_{i=1}^{m} \alpha_{k_i}(b_i)$$

means swapping α_1 with α_3 and α_2 with α_4 everywhere, so clearly w gives the resulting state description the same value. Similarly replacing Q by $\neg Q$ everywhere. Hence w satisfies SN. However, w does not satisfy Ax as apparent for example by considering the value it gives to $\alpha_1(a_1) \wedge \alpha_2(a_2)$ and $\alpha_2(a_1) \wedge \alpha_3(a_2)$.

16 (a) Let $\tau \in \mathsf{S}_{2^q}$.

$$\begin{aligned} v_{\vec{c}}\left(\bigwedge_{i=1}^{m} \alpha_{\tau(h_i)}(b_i)\right) &= |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma \in \mathsf{S}_{2^q}} w_{\langle c_{\sigma(1)}, c_{\sigma(2)}, \dots, c_{\sigma(2^q)}\rangle} \left(\bigwedge_{i=1}^{m} \alpha_{\tau(h_i)}(b_i)\right) \\ &= |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma \in \mathsf{S}_{2^q}} \prod_{i=1}^{m} c_{\sigma(\tau(h_i))} \\ &= |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma\tau \in \mathsf{S}_{2^q}} \prod_{i=1}^{m} c_{\sigma(\tau(h_i))}, \\ &\qquad \text{since } \sigma \mapsto \sigma\tau \text{ just permutes } \mathsf{S}_{2^q}, \\ &= |\mathsf{S}_{2^q}|^{-1} \sum_{\sigma \in \mathsf{S}_{2^q}} \prod_{i=1}^{m} c_{\sigma(h_i)}, \\ &= v_{\vec{c}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right). \end{aligned}$$

(b) Suppose that w satisfies Ax. By de Finetti's Representation Theorem there is a measure μ such that for a state description $\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$ and $\sigma \in \mathsf{S}_{2^q}$,

$$\begin{aligned} w\left(\bigwedge_{i=1}^{m} \alpha_{\sigma(h_i)}(b_i)\right) &= \int_{\mathbb{D}_{2^q}} w_{\langle x_1, x_2, \dots, x_{2^q}\rangle} \left(\bigwedge_{i=1}^{m} \alpha_{\sigma(h_i)}(b_i)\right) d\mu(\vec{x}) \\ &= \int_{\mathbb{D}_{2^q}} w_{\langle x_{\sigma(1)}, x_{\sigma(2)}, \dots, x_{\sigma(2^q)}\rangle} \left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) d\mu(\vec{x}) \qquad (43) \end{aligned}$$

Since w satisfies Ax,

$$w\left(\bigwedge_{i=1}^{m} \alpha_{\sigma(h_i)}(b_i)\right)$$

is the same for any $\sigma \in \mathsf{S}_{2^q}$ so averaging both sides of (43) over all $\sigma \in \mathsf{S}_{2^q}$ gives (39) when we restrict w and $v_{\vec{x}}$ to state descriptions. The general version follows as de Finetti's Theorem. The converse result is straightforward.

5 Carnap's Continuum

We continue to work in unary languages.

Theorem 11 *Suppose that the unary language L has at least two relation symbols, i.e. $q \geq 2$. Then the probability function w on SL satisfies Ex and JSP if and only if $w = c_\lambda^L$ for some $0 \leq \lambda \leq \infty$.*

Proof It is clear from their defining equations that the c_λ^L satisfy JSP. By Problem 10 they satisfy Ex.

For the other direction assume that w satisfies Ex and JSP. Then w satisfies Ax (Problem 17) so since

$$1 = w\left(\bigvee_{i=1}^{2^q} \alpha_i(a_1)\right) = \sum_{i=1}^{2^q} w(\alpha_i(a_1)),$$

we have $w(\alpha_i(a_1)) = 2^{-q}$ for all i. Now suppose that

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = 0$$

for some state description. We may assume that n is minimal; clearly $n > 1$. If $h_1 = h_2$ then by PIR

$$0 = w\left(\alpha_{h_1}(a_1) \mid \bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\right) \geq w\left(\alpha_{h_1}(a_1) \mid \bigwedge_{i=3}^{n} \alpha_{h_i}(a_i)\right)$$

so

$$w\left(\bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\right) = 0$$

etc., contradicting the minimality of n. Hence all the h_i must be different. So by JSP

$$0 = w\left(\alpha_{h_1}(a_1) \mid \bigwedge_{i=2}^{n} \alpha_{h_i}(a_i)\right) = w\left(\alpha_1(a_1) \mid \bigwedge_{i=2}^{n} \alpha_2(a_i)\right)$$

and we must have

$$w\left(\alpha_1(a_1) \wedge \bigwedge_{i=2}^{n} \alpha_2(a_i)\right) = 0.$$

Hence $n = 2$. This means that for any n, whenever the h_i are not all equal, we have

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = 0$$

and consequently $w = c_0^L$.

So now assume that w is non-zero on all state descriptions. Let

$$g(r,n) = w\left(\alpha_j(a_{n+1}) \,|\, \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right).$$

where $r = |\{\, i \,|\, h_i = j \,\}|$. Note that $g(0,0) = 2^{-q}$ and

$$1 > g(r,n) > 0$$

for all n, r. From

$$1 = w\left(\bigvee_{i=1}^{2^q} \alpha_i(a_2) \,|\, \alpha_j(a_1)\right) = \sum_{i=1}^{2^q} w(\alpha_i(a_2) \,|\, \alpha_j(a_1))$$

we get

$$g(1,1) + (2^q - 1)g(0,1) = 1. \tag{44}$$

By PIR, $g(1,1) \geq g(0,0)$ so

$$1 > g(1,1) \geq 2^{-q}.$$

Hence for some $0 < \lambda \leq \infty$,

$$g(1,1) = \frac{1 + 2^{-q}\lambda}{1+\lambda}, \quad g(0,1) = \frac{2^{-q}\lambda}{1+\lambda},$$

(by Problem 18 and (44)).

We now show by induction on $n \in \mathbb{N}$ that for this same λ

$$g(r,n) = \frac{r + \lambda 2^{-q}}{n+\lambda} \qquad (r = 0, 1, \ldots, n). \tag{45}$$

We have already shown it for $n = 0, 1$. Assume that $n \geq 1$ and (45) holds for n.

We shall write $\alpha_{h_1}\alpha_{h_2}\ldots\alpha_{h_n}$ etc. for

$$\bigwedge_{i=1}^{n} \alpha_{h_i}(b_i),$$

collecting the same atoms if repeated, writing for example $\alpha_2^3\alpha_4^2$ for $\alpha_2\alpha_2\alpha_4\alpha_2\alpha_4$. It is understood that the instantiating constants are distinct; by Ex it does not matter which constants they are.

For $r+s+t = n$ and distinct m, j, k we obtain,[1] by expressing $w(\alpha_m\alpha_j \mid \alpha_m^r\alpha_j^s\alpha_k^t)$ in two possible ways, that

$$\begin{aligned} w(\alpha_m \mid \alpha_j\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_j \mid \alpha_m^r\alpha_j^s\alpha_k^t) &= w(\alpha_m\alpha_j \mid \alpha_m^r\alpha_j^s\alpha_k^t) \\ &= w(\alpha_j \mid \alpha_m\alpha_m^r\alpha_j^s\alpha_k^t) \cdot w(\alpha_m \mid \alpha_m^r\alpha_j^s\alpha_k^t). \end{aligned}$$

Hence

$$g(r, n+1)g(s, n) = g(s, n+1)g(r, n). \tag{46}$$

Using $s = 0$ and the inductive hypothesis gives

$$g(r, n+1) = (r\lambda^{-1}2^q + 1)g(0, n+1). \tag{47}$$

This expresses all the $g(r, n+1)$ for $r = 1, \ldots, n$ in terms of $g(0, n+1)$. To find $g(0, n+1)$, note that distinct m, k,

$$1 = w\left(\bigvee_{i=1}^{2^q} \alpha_i \mid \alpha_m^n\,\alpha_k\right)$$

so

$$1 \;=\; g(n, n+1) + g(1, n+1) + (2^q - 2)g(0, n+1). \tag{48}$$

This yields

$$g(0, n+1) = \frac{\lambda 2^{-q}}{n+1+\lambda}.$$

Substituting in (47) gives

$$g(r, n+1) = \frac{r + \lambda 2^{-q}}{n+1+\lambda}$$

[1] Note this is where we need $q \geq 2$.

for $r = 1, 2, \ldots, n$. Finally, from

$$1 = w\left(\bigvee_{i=1}^{2^q} \alpha_i \,|\, \alpha_m^{n+1}\right)$$

we have

$$1 \;=\; g(n+1, n+1) + (2^q - 1)g(0, n+1).$$

so

$$g(n+1, n+1) = \frac{n+1+\lambda 2^{-q}}{n+1+\lambda}$$

This concludes the induction step. □

Carnap's Continuum functions have a number of attractive properties. They satisfy Ex, JSP, Ax (and hence Px and SN). Moreover, as we shall now show, they satisfy

The Unary Language Invariance Principle, ULi

A probability function w for a unary language L satisfies Unary Language Invariance if there is a family of probability functions $w^{\mathcal{L}}$, one for each (finite) unary language $\mathcal{L}$, satisfying Px and Ex such that $w^L = w$ and whenever $\mathcal{L} \subseteq \mathcal{L}'$, $w^{\mathcal{L}} = w^{\mathcal{L}'} \restriction S\mathcal{L}$ (i.e. $w^{\mathcal{L}'}$ restricted to $S\mathcal{L}$).

(Furthermore, we say that w satisfies Unary Language Invariance with $\mathcal{P}$, where $\mathcal{P}$ is some property, if the members $w^{\mathcal{L}}$ of this family also all satisfy the property $\mathcal{P}$.)

Note that for such a family we must have $w^{\mathcal{L}} = w \restriction S\mathcal{L}$ for $\mathcal{L} \subseteq L$. To show that c_λ^L satisfy Language Invariance, we need the following result.

Proposition 12 *Let $\theta_1(x), \theta_2(x), \ldots, \theta_k(x)$ be disjoint quantifier free formulae of L. Then for $0 < \lambda \leq \infty$*

$$c_\lambda^L\left(\theta_j(a_{n+1}) \,|\, \bigwedge_{i=1}^n \theta_{h_i}(a_i)\right) = \frac{m_j + t_j \lambda 2^{-q}}{n+\lambda}$$

and for $n > 0$,

$$c_0^L(\bigwedge_{i=1}^n \theta_{h_i}(a_j)) = \begin{cases} t_j 2^{-q} & \text{if } h_1 = h_2 = \ldots = h_n, \\ 0 & \text{otherwise,} \end{cases}$$

where $m_j = |\{i \,|\, h_i = j\}|$ *and* $t_j = |\{r \,|\, \alpha_r(x) \models \theta_j(x)\}|$.

(For a proof see solution to problem 19.)

Corollary 13 *For a fixed* λ *the* c_λ^L *form a ULi family.*

Proof Let $L = \{R_1, R_2, \dots, R_q\}$ and $L' = \{R_1, R_2, \dots, R_{q+1}\}$. By iterating the argument it is enough to prove that $c_\lambda^{L'} \restriction SL = c_\lambda^L$. Using the previous proposition and applying it to $c_\lambda^{L'}$, taking $\theta_1(x), \dots, \theta_{2^q}(x) \in QFFL'$ to be the atoms of L, $\theta_s(x) = \alpha_s(x)$, we can see that $c_\lambda^{L'} \restriction SL$ and c_λ^L agree on state descriptions and hence they are equal.

□

Furthermore, c_λ^L satisfy Reichenbach's Axiom: with $u_j(n)$ as in (40),

$$\begin{aligned} c_\lambda^L\left(\alpha_j(a_{n+1}) \,|\, \bigwedge_{i=1}^n \alpha_{h_i}(a_i)\right) - \frac{u_j(n)}{n} &= \frac{u_j(n) + 2^{-q}\lambda}{n+\lambda} - \frac{u_j(n)}{n} \\ &= \frac{2^{-q}\lambda}{n+\lambda} - \frac{\lambda u_j(n)}{n(n+\lambda)} \;\to 0 \quad \text{as } n \to \infty. \end{aligned}$$

Possibly on the negative side, c_λ^L for $\lambda > 0$ give value 0 to all non-tautologous universal sentences $\forall x\, \theta(x)$, with $\theta(x)$ quantifier free (see e.g. [1, Chapter 18]). Hence

$$c_\lambda^L\left(\forall x \theta(x) \,|\, \bigwedge_{i=1}^n \theta(a_i)\right) = 0$$

no matter how big n is. It has been argued that since (by Proposition 12) for $0 < \lambda < \infty$ and consistent quantifier free $\theta(x)$,

$$\lim_{n\to\infty} c_\lambda^L\left(\theta(a_{n+1}) \,|\, \bigwedge_{i=1}^n \theta(a_i)\right) = 1, \tag{49}$$

such universal confirmation is not important.

Another Continuum of Inductive Methods

The Generalised Principle of Instantial Relevance, GPIR

For $\psi(a_1, a_2, \ldots, a_n), \phi(a_1), \theta(a_1) \in QFSL$, *if* $\theta(x) \models \phi(x)$

$$w(\theta(a_{n+2}) \mid \phi(a_{n+1}) \wedge \psi(a_1, a_2, \ldots, a_n)) \geq w(\theta(a_{n+2}) \mid \psi(a_1, a_2, \ldots, a_n)). \quad (50)$$

The principle's claim to rationality is based on observing that a rational agent's probability of something occurring 'next time' should arguably increase (not decrease) upon witnessing a consequence of it 'this time'.

It can be shown that when $\psi(a_1, \ldots, a_n)$ is missing, (50) follows from Ex+Ax.

Let $-(2^q - 1)^{-1} \leq \delta \leq 1$ and $\gamma = 2^{-q}(1 - \delta)$ and let w_δ^L be the average sum of the 2^q probability functions $w_{\vec{e}}$ where $\vec{e}$ are those vectors from $\mathbb{D}_{2^q}$ which have one coordinate $\gamma + \delta$ and the rest γ, that is $\vec{e} = \langle \gamma, \ldots, \gamma, \gamma+\delta, \gamma, \ldots, \gamma \rangle$.

As the following results[1] show, just as the c_λ^L are determined by the requirements of Ex and JSP, the w_δ^L are determined by the requirements of Ex, Ax and GPIR.

Theorem 14 *Let* $-(2^q - 1)^{-1} < \delta < 1$. *Then* w_δ^L *satisfies GPIR, Reg, Ax, Ex.*

Theorem 15 *Let the probability function* w *on* SL *satisfy GPIR, Ex, Ax and Reg and* $q \geq 2$. *Then* $w = w_\delta^L$ *for some* $-(2^q - 1)^{-1} < \delta < 1$.

Furthermore, $w_0^L = c_\infty^L$ and $w_1^L = c_0^L$; the latter does not satisfy Reg but it does satisfy Ex, Ax and GPIR. No other w_δ^L belongs to the Carnap Continuum.

Theorem 16 *For a fixed* $0 \leq \delta \leq 1$ *and unary languages* $\mathcal{L}$, *the functions* $w_\delta^{\mathcal{L}}$ *form a language invariant family.*

These functions are referred to as the *NP-Continuum of Inductive Methods* (NP standing for Nix-Paris). They have a number of arguably desirable properties, which are described below. We omit the proofs, see again [1].

[1] Quoted without proofs, see [1].

Recovery *A probability function w on SL is* recoverable *(satisfies* Recovery*) if whenever $\Psi(a_1, a_2, \ldots, a_n)$ is a state description then there is another state description $\Phi(a_{n+1}, a_{n+2}, \ldots, a_h)$ such that $w(\Phi \wedge \Psi) \neq 0$ and for any $\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g}) \in QFSL$,*

$$w(\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g}) \mid \Phi \wedge \Psi) = w(\theta(a_{h+1}, a_{h+2}, \ldots, a_{h+g})). \qquad (51)$$

Hence w is *recoverable* if given a 'past history' in the form of a state description Ψ there is a possible 'future' state description Φ which makes matters look just as they were to start with as regards the quantifier free properties of the hitherto unobserved constants $a_{h+1}, a_{h+2}, \ldots$.[1]

Proposition 17 *For $0 \leq \delta < 1$ the w_δ^L satisfy Recovery.*

So each w_δ^L for $0 \leq \delta < 1$ satisfies Reg, Recovery and Uli with Ax. Somewhat remarkably, these properties even with Recovery required for just one (non-trivial) state description determine the w_δ^L:

Theorem 18 *A probability function w on SL satisfying Reg and ULi with Ax has the property that for some state description $\Phi(a_1, a_2, \ldots, a_n)$ with $n > 0$, and for all $\theta(a_{n+1}, a_{n+2}, \ldots, a_{n+g}) \in QFSL$,*

$$w(\theta(a_{n+1}, a_{n+2}, \ldots, a_{n+g}) \mid \Phi(a_1, a_2, \ldots, a_n)) = w(\theta(a_{n+1}, a_{n+2}, \ldots, a_{n+g}))$$

just if $w = w_L^\delta$ for some $0 \leq \delta < 1$.

The last property of the w_δ^L that we mention, again without proof[2]

The Weak Irrelevance Principle, WIP *Suppose that $\theta, \phi \in QFSL$ are such that they have no constant nor relation symbols in common. Then*

$$w(\theta \wedge \phi) = w(\theta) \cdot w(\phi).$$

[1] Replacing QFSL by SL yields an equivalent Principle.

[2] We will return to this in the polyadic context, since the w_δ^L are the unary members of certain fundamentally important language invariant families, which are in in a sense characterised by WIP.

Unlike the the Carnap Continuum functions, the NP-continuum functions do not satisfy Reichenbach axiom, see Problem 21. They behave in the same way as the Carnap Continuum functions as regards universal sentences, giving 0 to all non-tautologous universal sentences $\forall x\, \theta(x)$, with $\theta(x)$ quantifier free but unlike the Carnap Continuum functions, they do not redeem themselves, as the Carnap's functions do via (49), either.

It has been argued that the principles underlying the Carnap Continuum functions come from assuming that the agent imagines his ambient structure M as the result of some regular statistical process, for example the picking of balls from an urn as is often quoted to motivate JSP. Being rational then means expecting regularity, predictability, and pattern. On the other hand the version of rationality underlying the NP-continuum functions is that of simplicity or economy. For example the consideration that information might be ignored where possible can motivate GPIR (without the background evidence $\psi(a_1, \ldots, a_n)$ the inequality is a consequence of Ex+Ax), Recoverability and WIP.

Problems

Problem 17 *Show that JSP implies Ax.*

Problem 18 *Let* $1 > x > a$. *Show that there is* $\lambda > 0$ *such that*

$$x = \frac{1 + a\lambda}{1 + \lambda}$$

and that, consequently, if $x + (a^{-1} - 1)y = 1$ *then*

$$y = \frac{a\lambda}{1 + \lambda}\,.$$

Problem 19 *Prove Proposition 12.*

Problem 20 *Show that for* $\phi(a_1), \theta(a_1), \psi(\vec{a}) \in SL$ *where* $\vec{a}$ *stands for* $a_1, \dots, a_n$,

$$\begin{aligned} & w(\theta(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \psi(\vec{a})) \geq w(\theta(a_{n+2}) \,|\, \psi(\vec{a})) \\ & \qquad \iff \quad w(\neg\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1}) \wedge \psi(\vec{a})) \geq w(\neg\theta(a_{n+2}) \,|\, \psi(\vec{a})), \end{aligned}$$

and use it to show that changing $\theta(x) \models \phi(x)$ *to* $\phi(x) \models \theta(x)$ *in GPIR yields an equivalent principle, that is, GPIR is equivalent to :*

For $\psi(a_1, a_2, \dots, a_n), \phi(a_1), \theta(a_1) \in SL$, *if* $\phi(x) \models \theta(x)$ *then*

$$w(\theta(a_{n+2}) \,|\, \phi(a_{n+1}) \wedge \psi(a_1, a_2, \dots, a_n)) \geq w(\theta(a_{n+2}) \,|\, \psi(a_1, a_2, \dots, a_n)).$$

Problem 21 *By considering the sequence of atoms* $\alpha_{h_i}(x)$ *where* $h_i = 1$ *for all* i, *or another sequence, show that* w_δ^L *for* $\delta \neq 0, 1$, *fail Reichenbach's Axiom.*

Solutions

17 Since

$$w\left(\bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \prod_{j=1}^{n} w(\alpha_{h_j}(a_j) \mid \bigwedge_{i=1}^{j-1} \alpha_{h_i}(a_i))$$

(with both sides zero if not all the conditional probabilities are defined) JSP gives that this right hand side is invariant under permutations of atoms. Hence so is the left hand side and this yields the result.

18 Differentiating shows that the continuous function

$$f(\lambda) = \frac{1 + a\lambda}{1 + \lambda}$$

is decreasing from 1 to a for $\lambda \in (0, \infty)$, so for any $x \in (a, 1)$ there must be some $\lambda \in (0, \infty)$ such that $f(\lambda) = x$. The rest is obvious.

19 The result for $\lambda = 0$ is straightforward so assume that $\lambda > 0$. For $s = 1, 2, \ldots, k$ let

$$\Gamma_s = \{\alpha_r(x) \mid \alpha_r(x) \models \theta_s(x)\},$$

so $t_s = |\Gamma_s|$. Let P be the set of all state descriptions $\psi(a_1, a_2, \ldots, a_n)$ of the form

$$\alpha_{g_1}(a_1) \wedge \alpha_{g_2}(a_2) \wedge \ldots \wedge \alpha_{g_n}(a_n)$$

where $\alpha_{g_i}(x) \in \Gamma_{h_i}$ for $i = 1, 2, \ldots, n$.

Then

$$\begin{aligned} c_\lambda^L\left(\theta_j(a_{n+1}) \mid \bigwedge_{i=1}^{n} \theta_{h_i}(a_i)\right) &= \sum_{\alpha_r(x) \in \Gamma_j} c_\lambda^L\left(\alpha_r(a_{n+1}) \mid \bigwedge_{i=1}^{n} \bigvee_{\alpha_{g_i}(x) \in \Gamma_{h_i}} \alpha_{g_i}(a_i)\right) \\ &= \sum_{\alpha_r(x) \in \Gamma_j} c_\lambda^L\left(\alpha_r(a_{n+1}) \mid \bigvee_{\psi \in P} \psi\right) \\ &= \sum_{\alpha_r(x) \in \Gamma_j} \sum_{\psi \in P} \frac{c_\lambda^L(\alpha_r(a_{n+1}) \mid \psi) \cdot c_\lambda^L(\psi)}{c_\lambda^L(\bigvee_{\psi \in P} \psi)}\,. \end{aligned} \tag{52}$$

But for any particular such ψ

$$
\begin{aligned}
\sum_{\alpha_r(x)\in\Gamma_j} c_\lambda^L(\alpha_r(a_{n+1}) \,|\, \psi)\cdot\, c_\lambda^L(\psi) &= \sum_{\alpha_r(x)\in\Gamma_j} \frac{s_r+\lambda 2^{-q}}{n+\lambda}\cdot\, c_\lambda^L(\psi)\\
&= \frac{m_j+\lambda 2^{-q}|\Gamma_j|}{n+\lambda}\cdot\, c_\lambda^L(\psi),
\end{aligned}
$$

where s_r is the number of times that $\alpha_r(x)$ is instantiated in ψ; the last equality follows since the s_r sum to m_j for any ψ under consideration. Substituting this into (52) gives

$$
\begin{aligned}
c_\lambda^L\left(\theta_j(a_{n+1}) \,|\, \bigwedge_{i=1}^{n}\theta_{h_i}(a_i)\right) &= \sum_{\psi\in P}\frac{m_j+\lambda 2^{-q}|\Gamma_j|}{n+\lambda}\cdot\frac{c_\lambda^L(\psi)}{c_\lambda^L(\bigvee_{\psi\in P}\psi)}\\
&= \frac{m_j+\lambda 2^{-q}|\Gamma_j|}{n+\lambda},
\end{aligned}
$$

as required.

20 We have

$$
\begin{aligned}
w(\theta(a_{n+2}) \,|\, \psi(\vec{a})) &= w(\theta(a_{n+2}) \,|\, \phi(a_{n+1})\wedge\psi(\vec{a}))\cdot w(\phi(a_{n+1}) \,|\, \psi(\vec{a}))\\
&\quad + w(\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1})\wedge\psi(\vec{a}))\cdot w(\neg\phi(a_{n+1}) \,|\, \psi(\vec{a}))
\end{aligned}
$$

and

$$
w(\phi(a_{n+1}) \,|\, \psi(\vec{a})) + w(\neg\phi(a_{n+1})|\psi(\vec{a})) = 1,
$$

so

$$
\begin{aligned}
&w(\theta(a_{n+2}) \,|\, \phi(a_{n+1})\wedge\psi(\vec{a})) \geq w(\theta(a_{n+2}) \,|\, \psi(\vec{a}))\\
&\qquad\iff w(\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1})\wedge\psi(\vec{a})) \leq w(\theta(a_{n+2}) \,|\, \psi(\vec{a}))\\
&\qquad\iff w(\neg\theta(a_{n+2}) \,|\, \neg\phi(a_{n+1})\wedge\psi(\vec{a})) \geq w(\neg\theta(a_{n+2}) \,|\, \psi(\vec{a})),
\end{aligned}
$$

The equivalence follows since $\phi(x_1)\models\theta(x_1)$ just if $\neg\theta(x_1)\models\neg\phi(x_1)$.

21 Recall that

$$
w_L^\delta\left(\bigwedge_{i=1}^{m}\alpha_{h_i}(a_i)\right) = 2^{-q}\sum_{j=1}^{2^q}\gamma^{m-m_j}(\gamma+\delta)^{m_j}
$$

where as usual $m_j = |\{i \mid h_i = j\}|$. Hence - in the notation of the statement of RA, for the given sequence and with $j = 1$, $u_j(n)/n \to 1$ as $n \to \infty$. However

$$w_L^\delta \left(\alpha_1(a_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_1(a_i) \right) = \frac{(\gamma+\delta)^{n+1} + (2^q - 1)\gamma^{n+1}}{(\gamma+\delta)^n + (2^q - 1)\gamma^n}$$

which tends to $\gamma + \delta$ when $\delta > 0$ and to γ when $\delta < 0$ as $n \to \infty$, so RA fails.

6 Polyadic Pure Inductive Logic

To start with, throughout this section, we shall restrict our considerations to the case of L containing a single, binary, relation symbol R. It reduces notational difficulties and it will help us to gain some intuition about the non-unary context.

Note that in this case state descriptions for $a_1, \ldots, a_m$ have the form

$$\Theta(a_1, \ldots, a_m) = \bigwedge_{i,j=1}^{m} \pm R(a_i, a_j)$$

where as before $\pm R$ stands for R or $\neg R$. To make this easier to work with, we can also write

$$\Theta(a_1, \ldots, a_m) = \bigwedge_{i,j=1}^{m} R^{t_{i,j}}(a_i, a_j)$$

where $t_{i,j} \in \{0, 1\}$ and R^0 stands for $\neg R$, R^1 stands for R. This allows us to represent Θ by the $m \times m$ $\{0, 1\}$-matrix $T = (t_{i,j})$.[1] (Recall that we have already used such a representation in Problem 7.)

We now introduce probability functions w^D which play a role similar to that played in the unary case by the $w_{\vec{x}}$. Let $N \in \mathbb{N}$ and let $D = (d_{i,j})$ be an $N \times N$ $\{0, 1\}$-matrix (it is best to think of N as large although it can be any nonzero natural number).

Define a probability function w^D on SL by setting

$$w^D \left(\bigwedge_{i,j \leq m} R^{t_{i,j}}(a_i, a_j) \right)$$

to be the probability of (uniformly) randomly picking, with replacement, $h(1), h(2), \ldots, h(m)$ from $\{1, 2, \ldots, N\}$ such that for each $i, j \leq m$,

$$d_{h(i),h(j)} = t_{i,j}.$$

This does uniquely determine a probability function on SL satisfying Ex, see Problem 22.

[1] Information about the exact constants involved is lost, but since we will only consider probability function satisfying Ex, it will not matter.

Clearly convex mixtures of these w^D also satisfy Ex. Conversely, any probability function satisfying Ex can be expressed as an integral of standard parts of such w^D with non-standard D. Remaining within standard mathematics for the present, we shall just sketch how to show that any probability function w satisfying Ex can be approximated arbitrarily closely on $QFSL$ by convex mixtures of the w^D. More precisely:

Theorem 19 *Let w be a probability function w on SL satisfying Ex. For each $m \in \mathbb{N}$ and $\epsilon > 0$ there are*

- $N \in \mathbb{N}$,
- *a set $\mathcal{D}$ of $N \times N$ $\{0,1\}$-matrices,*
- $\lambda_D \geq 0$ *for each* $D \in \mathcal{D}$

such that $\sum_{D \in \mathcal{D}} \lambda_D = 1$ and for any $\theta(a_1, \ldots, a_m) \in QFSL$,[1]

$$|w(\theta) - \sum_{D \in \mathcal{D}} \lambda_D w^D(\theta)| < \epsilon.$$

Proof It suffices to prove the lemma for state desctiptions $\Theta(a_1, \ldots, a_m)$ since any quantifier free sentence $\theta(a_1, \ldots, a_m)$ is a disjunction of some of them (and there are just 2^{m^2} of them).

For a state description $\Theta(a_1, \ldots, a_m)$ and $N > m$ we have

$$w(\Theta(a_1, \ldots, a_m)) = \sum_{\Psi(a_1, \ldots, a_N) \models \Theta(a_1, \ldots, a_m)} w(\Psi(a_1, \ldots, a_N)).$$

We wish to collect together those contributions to the above sum which must by Ex be equal. For a state description $\Phi(a_1, \ldots, a_N)$ we define $\overline{\Phi}$ to be the set (equivalence class) of all state descriptions that can be obtained from $\Phi(a_1, \ldots, a_N)$ by permuting constants, that is, state descriptions of the form

$$\Psi(a_1, \ldots, a_N) = \Phi(a_{\sigma(1)}, \ldots, a_{\sigma(N)}) \tag{53}$$

(where σ is a permutation of $\{1, \ldots, N\}$). Note that for such Ψ we have $w(\Psi) = w(\Phi)$ by Ex, and that $\overline{\Psi} = \overline{\Phi}$. Let

$$K(\Phi, \Theta) = |\{\Psi \in \overline{\Phi};\ \Psi(a_1, \ldots, a_N) \models \Theta(a_1, \ldots, a_m)\}|$$

[1] This g means that all the constants actually appearing in $\theta(a_1, \ldots, a_m)$ are amongst the $a_1, \ldots, a_m$; some need not appear.

and let $K(\Phi) = |\overline{\Phi}|$.[1]

We choose one representative of each class and we denote the set of these representatives as $\mathcal{R}$. Now we can write

$$w(\Theta(a_1, \ldots, a_m)) = \sum_{\Phi \in \mathcal{R}} w(\Phi(a_1, \ldots, a_N)) \cdot K(\Phi, \Theta),$$

$$= \sum_{\Phi \in \mathcal{R}} w(\Phi(a_1, \ldots, a_N)) K(\Phi) \cdot \frac{K(\Phi, \Theta)}{K(\Phi)}, \qquad (54)$$

The ratio

$$\frac{K(\Phi, \Theta)}{K(\Phi)} \qquad (55)$$

is the probability that a random permutation σ yields Ψ that belongs to $K(\Phi, \Theta)$. Let $D_\Phi = (d_{i,j})$ be the $N \times N$ matrix representing Φ. (55) is also the probability that when (uniformly) randomly picking, *without* replacement, $h(1), h(2), \ldots, h(m)$ from $\{1, 2, \ldots, N\}$,

$$\bigwedge_{i,j \leq m} R^{d_{h(i),h(j)}}(a_i, a_j) \models \Theta(a_1, \ldots, a_m).$$

That is, such that for each $i, j \leq m$,

$$d_{h(i),h(j)} = t_{i,j}.$$

Recall that the same definition except that the picking is *with* replacement, gives $w^{D_\Phi}(\Theta(a_1, \ldots, a_m))$.

The difference in the probability of picking particular $h(1), h(2), \ldots, h(m)$ from $\{1, 2, \ldots, N\}$ with and without replacement is

$$\prod_{i=0}^{m-1} (N - i)^{-1} - N^{-m},$$

if there are no repeats in the $h(1), h(2), \ldots, h(m)$ (and hence the difference is of order $N^{-(m+1)}$) or N^{-m} if there are repeats. There are N^m m-tuples

[1]Note that this is not necessarily $N!$ because several permutations may yield the same state descriptions, but every state description obtainable from Φ is obtained from Φ by the same number of permutations.

$h(1), h(2), \ldots, h(m)$ altogether and less than $\binom{m}{2} N^{m-1}$ of them are with repeats, so the difference between $\frac{K(\Phi,\Theta)}{K(\Phi)}$ and $w^{D_\Phi}(\Theta)$ is of order N^{-1}.

Let

$$\lambda_{D_\Phi} = K(\Phi) w(\Phi(a_1, \ldots, a_N)).$$

From (54) we have

$$w(\Theta(a_1, \ldots, a_m)) \;=\; \sum_{\Phi \in \mathcal{R}} \lambda_{D_\Phi} \cdot \frac{K(\Phi, \Theta)}{K(\Phi)}.$$

Furthermore:

$$\sum_{\Phi \in \mathcal{R}} \lambda_{D_{\overline{\Phi}}} = 1$$

and as argued above,

$$\left| \frac{K(\Phi, \Theta)}{K(\Phi)} - w^{D_\Phi}(\Theta) \right|$$

is of order N^{-1}. It follows that for N large enough

$$\left| \sum_{\Phi \in \mathcal{R}} \lambda_{D_\Phi} \cdot \frac{K(\Phi, \Theta)}{K(\Phi)} - \sum_{\Phi \in \mathcal{R}} \lambda_{D_\Phi} w^{D_\Phi}(\Theta) \right| \;=\; \left| w(\Theta) - \sum_{\Phi \in \mathcal{R}} \lambda_{D_\Phi} w^{D_\Phi}(\Theta) \right| < \epsilon,$$

so it suffices to take $\mathcal{D} = \{D_\Phi : \Phi \in \mathcal{R}\}$. $\square$

Problems

Problem 22 *Let L be a language with a single, binary, predicate R. Let $D = (d_{i,j})$ be an $N \times N$ $\{0,1\}$-matrix. Define w^D on SL by setting $w^D(\top) = 1$ and*

$$w^D\left(\bigwedge_{i,j\leq n} R^{t_{i,j}}(a_i, a_j)\right)$$

to be the probability of (uniformly) randomly picking, with replacement, $h(1), h(2), \ldots, h(n)$ from $\{1, 2, \ldots, N\}$ such that for each $i, j \leq n$,

$$d_{h(i),h(j)} = t_{i,j}.$$

Show that this uniquely determines a probability function on SL satisfying Ex. Moreover, show that for $\theta(a_1, \ldots, a_n) \in QFSL$, $w^D(\theta)$ is the probability that when (uniformly) randomly picking, with replacement, $h(1), h(2), \ldots, h(n)$ from $\{1, 2, \ldots, N\}$,

$$\bigwedge_{i,j\leq n} R^{d_{h(i),h(j)}}(a_i, a_j) \models \theta(a_1, \ldots, a_n).$$

Solutions to Problems

22 First note that from the conditions (9), (i) and (ii) clearly hold. (iii) also holds, since the picking is with replacement. In detail, for

$$\Theta(a_1, \ldots, a_m) = \bigwedge_{i,j \leq m} R^{t_{i,j}}(a_i, a_j)$$

$w^D(\Theta(a_1, \ldots, a_m))$ is the ratio

$$\frac{|\langle h_1, \ldots, h_m \rangle \in \{1, \ldots, N\}^m : \text{ for all } i, j \leq m,\ d_{h_i, h_j} = t_{i,j}.\}|}{N^m} \geq 0$$

and

$$\sum_{\Phi(a_1, \ldots, a_{m+1}) \models \Theta(a_1, \ldots, a_m)} w^D(\Phi(a_1, a_2, \ldots, a_{m+1}))$$

$$= \sum_{\vec{s} \in \{0,1\}^{2m+1}} w^D \left(\bigwedge_{i,j \leq m} R^{t_{i,j}}(a_i, a_j) \wedge \bigwedge_{i=1}^{m+1} R^{s_{i,m+1}}(a_i, a_{m+1}) \wedge \bigwedge_{j=1}^{m} R^{s_{m+1,j}}(a_{m+1}, a_j) \right)$$

where

$$\vec{s} = \langle s_{1,m+1}, s_{2,m+1}, \ldots, s_{m+1,m+1}, s_{m+1,1} \ldots, s_{m+1,m} \rangle.$$

For a given $\vec{s}$ the summand above is

$$\frac{\left| \left\{ \langle h_1, \ldots, h_m, h_{m+1} \rangle \in \{1, \ldots, N\}^{m+1} : \forall i, j \leq m,\ d_{h_i, h_j} = t_{i,j} \ \& \ \begin{array}{l} d_{h_i, h_{m+1}} = s_{i,m+1} \\ d_{h_{m+1}, h_j} = s_{m+1,j} \\ d_{h_{m+1}, h_{m+1}} = s_{m+1,m+1} \end{array} \right\} \right|}{N^{m+1}}$$

Since for any given $h_1, \ldots, h_m$, each $h_{m+1} \in \{1, \ldots, N\}$ adds to precisely one such summand, the summands add to

$$w^D(\Theta(a_1, \ldots, a_m))$$

as required. Hence the definition does uniquely determine a probability function on SL

Ex follows by Lemma 4 and Gaifman's Theorem so $w^D\left(\bigwedge_{i,j\leq n} R^{t_{i,j}}(b_i,b_j)\right)$ with any other distinct $b_1,\ldots,b_n$ is also the probability of (uniformly) randomly picking, with replacement, $h(1),h(2),\ldots,h(n)$ from $\{1,2,\ldots,N\}$ such that for each $i,j\leq n$,

$$d_{h(i),h(j)} = t_{i,j}.$$

Since any $\theta(b_1,\ldots,b_n)\in QFSL$ is logically equivalent to a disjunction of state descriptions, $w_D(\theta(b_1,\ldots,b_n))$ is the sum of $w^D(\Theta(b_1,\ldots,b_n))$ over those $\Theta(b_1,\ldots b_n)$ that logically imply it and hence the probability that when (uniformly) randomly picking, with replacement, $h(1),h(2),\ \ldots,h(n)$ from $\{1,2,\ldots,N\}$,

$$\bigwedge_{i,j\leq n} R^{d_{h(i),h(j)}}(a_i,a_j)\models\theta(a_1,\ldots,a_n),$$

as required.

General Representation Theorem for Ex

(Reading from [1, Chapter 24], for those interested in nonstandard methods.)

Let w be a probability function on SL satisfying Ex. Let U^* be a nonstandard ω_1-saturated elementary extension of a sufficiently large portion U of the set theoretic universe containing w.

For $c \in U$ let c^* denote its image in U^* where this differs from c. Let $n \in \mathbb{N}$, let $\nu \in \mathbb{N}^*$ be nonstandard and let $\Theta(a_1, a_2, \dots, a_n)$ be a state description.

Working in U^*, we have:

$$w^*(\Theta(a_1, \dots, a_n)) = \sum_{\Phi(a_1,\dots,a_\nu) \models \Theta(a_1,\dots,a_n)} w^*(\Phi(a_1, \dots, a_\nu)). \qquad (56)$$

Since w^* satisfies Ex in U^*, for $\sigma \in U^*$ a permutation of $\{1, 2, \dots, \nu\}$ and $\Phi(a_1, a_2, \dots, a_\nu)$ a state description in U^*,

$$w^*(\Phi(a_1, a_2, \dots, a_\nu)) = w^*(\Phi(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(\nu)})).$$

Pick a representative $\Psi(a_1, \dots, a_\nu)$ from each of the classes

$$\{\Phi(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(\nu)}) \,|\, \sigma \text{ a permutation of } \{1, 2, \dots, \nu\}\}$$

and denote the corresponding class H_Ψ. w^* is constant on H_Ψ so from (56),

$$w^*(\Theta(a_1, \dots, a_n)) = \sum_{\Psi(a_1,\dots,a_\nu)} \omega^\Psi(\Theta(a_1, \dots, a_n))\, w^*(\bigvee H_\Psi) \qquad (57)$$

where $\omega^\Psi(\Theta(a_1, a_2, \dots, a_n))$ is the proportion of $\Phi(a_1, \dots, a_\nu) \in H_\Psi$ which logically imply $\Theta(a_1, \dots, a_n)$ (i.e. *are* $\Theta(a_1, \dots, a_n)$ when restricted just to $a_1, a_2, \dots, a_n$). Note that for any $\Theta(a_1, a_2, \dots, a_n)$, $\omega^\Psi(\Theta(a_1, a_2, \dots, a_n))$, as a function of Ψ, is internal (i.e. is in U^*), and that $\mu_0\{\Psi\} = w^*\left(\bigvee \mathcal{H}_\Psi\right)$ determines a finitely additive measure on the algebra of (internal) subsets of the set A of the chosen representatives Ψ. So we can rewrite (57) as

$$w^*(\Theta(a_1, \dots, a_n)) = \int_A \omega^\Psi(\Theta(a_1, \dots, a_n))\, d\mu_0(\Psi). \qquad (58)$$

There is another way to picture $\omega^\Psi(\Theta(a_1, a_2, \dots, a_n))$. Namely imagine randomly picking, in U^*, *without replacement* and according to the uniform distribution, $a_{h_1}, a_{h_2}, \dots, a_{h_n}$ from $\{a_1, \dots, a_\nu\}$. Then $\omega^\Psi(\Theta(a_1, a_2, \dots, a_n))$ is the probability that $\Psi(a_1, \dots, a_\nu)$ restricted to $a_{h_1}, \dots, a_{h_n}$ is $\Theta(a_{h_1}, a_{h_2}, \dots, a_{h_n})$, equivalently that

$$\Psi(a_1, \dots, a_\nu) \models \Theta(a_{h_1}, a_{h_2}, \dots, a_{h_n}).$$

This is because $\Psi_\sigma(a_1, \dots, a_\nu) = \Psi(a_{\sigma(1)}, \dots, a_{\sigma(\nu)})$ restricted to $a_1, \dots, a_m$ is Θ just when for $h_1, \dots, h_m$ such that

$$h_1 = \sigma^{-1}(1), \;\; \dots, \;\; h_m = \sigma^{-1}(m),$$

that is,

$$\sigma(h_1) = 1, \;\; \dots, \;\; \sigma(h_m) = m,$$

Ψ restricted to $a_{h_1}, \dots, a_{h_m}$ is $\Theta(a_{h_1}, \dots, a_{h_m})$. Hence the proportion of ordered m tuples of distinct numbers $\langle h_1, \dots, h_m \rangle$ from $\{1, \dots, \nu\}$ such that the restriction of Ψ to $a_{h_1}, \dots, a_{h_m}$ is $\Theta(a_{h_1}, \dots, a_{h_m})$ is the same as the proportion of $\sigma \in S_\nu$ for which Ψ_σ restricted to $a_1, \dots, a_m$ is Θ . It may be that different σ yield the same Ψ_σ but classes of those σ which do yield the same Ψ_σ have the same size, so the proportion of Ψ_σ such that their restriction to $a_1, \dots, a_m$ is Θ is also the same.

Viewed in this way we see that as a function of standard state descriptions $\Theta(a_1, \dots, a_n)$ (so $n \in \mathbb{N}$), ω^Ψ satisfies (9) and extends to satisfy (P1), (P2) and Ex on $QFSL$, *except* that it takes values in $[0,1]^*$ rather than just $[0,1]$.

Using Loeb Integration Theory (see e.g. [**?**, p17-20]) we can take standard parts of (58), denoted by $\circ$, to obtain

$$w(\Theta(a_1, \dots, a_n)) = \int_A {}^\circ\omega^\Psi(\Theta(a_1, \dots, a_n)) \, d\mu(\Psi) \tag{59}$$

where μ is a countably additive measure on a σ-algebra of subsets of the set A of $\Psi(a_1, \dots, a_\nu)$. The functions ${}^\circ\omega^\Psi$ now satisfy (P1), (P2) and Ex with standard values so extend uniquely by Theorem 5 to probability functions satisfying Ex on SL. Observe that ${}^\circ\omega^\Psi(\Theta(a_1, \dots, a_n))$ would be the same if the picking of the $a_{h_1}, a_{h_2}, \dots, a_{h_n}$ in the definition of ω^Ψ had been *with replacement* since the difference in the probability of picking $a_{h_1}, a_{h_2}, \dots, a_{h_n}$ with and without replacement is

$$\prod_{i=0}^{n-1} (\nu - i)^{-1} - \nu^{-n},$$

if there are no repeats in the $a_{h_1}, a_{h_2}, \ldots, a_{h_n}$ (and hence the difference is of order $\nu^{-(n+1)}$) or ν^{-n} if there are repeats. There are ν^n n-tuples $a_{h_1}, a_{h_2}, \ldots, a_{h_n}$ altogether and less than $\binom{n}{2}\nu^{n-1}$ of them are with repeats so this could only produce an infinitesimal change in $\omega^\Psi(\Theta(a_1, a_2, \ldots, a_n))$, and hence no change at all once we take standard parts.

Apart from Ex the $^\circ\omega^\Psi$ satisfy the polyadic version of the *Constant Irrelevance Principle, IP* given on page 33, namely that if $\theta, \phi \in QFSL$ have no constants in common then

$$^\circ\omega^\Psi(\theta \wedge \phi) = {}^\circ\omega^\Psi(\theta) \cdot {}^\circ\omega^\Psi(\phi).$$

To see this it is enough to show, as in the proof of Proposition 7, that for $n, m \in \mathbb{N}$ and state descriptions $\Theta(a_1, a_2, \ldots, a_n)$, $\Phi(a_{n+1}, a_{n+2}, \ldots, a_m)$

$$\begin{aligned} &^\circ\omega^\Psi(\Theta(a_1, a_2, \ldots, a_n) \wedge \Phi(a_{n+1}, a_{n+2}, \ldots, a_m)) = \\ &\qquad\qquad {}^\circ\omega^\Psi(\Theta(a_1, a_2, \ldots, a_n)) \cdot {}^\circ\omega^\Psi(\Phi(a_{n+1}, a_{n+2}, \ldots, a_m)). \end{aligned}$$

But this clearly holds by the above observation since if the picking of the $a_{h_1}, a_{h_2}, \ldots, a_{h_n}$ in the definition of ω^Ψ had been *with replacement* all these choices would have been independent of each other and hence the sum of $\omega^\Psi(\Delta(a_1, \ldots, a_m))$ for the (finitely many) $\Delta(a_1, \ldots, a_m)$ extending $\Theta(a_1, \ldots, a_n) \wedge \Psi(a_{n+1} \ldots, a_m)$ would be equal to the product

$$\omega^\Psi(\Theta(a_1, a_2, \ldots, a_n)) \cdot \omega^\Psi(\Phi(a_{n+1}, a_{n+2}, \ldots, a_m)).$$

In summary then we have arrived at a result due to Peter Krauss (by an entirely different proof):

Theorem 20 *If the probability function w on SL satisfies Ex then w can be represented in the form*

$$w = \int_A {}^\circ\omega^\Psi \, d\mu(\Psi) \tag{60}$$

for some countably additive measure μ on an algebra of subsets of A and probability functions $^\circ\omega^\Psi$ on SL satisfying IP (and Ex).

Conversely if w is of this form then it satisfies Ex.

To further illuminate this theorem we note (in line with Krauss' results) that these $^\circ\omega^\Psi$ are precisely the probability functions satisfying IP (and Ex).

Proposition 21 *If the probability function w on SL satisfies IP (and Ex) then $w = {}^\circ\omega^\Psi$ for some Ψ (within U^*), and conversely.*

Proof Suppose that w is a probability function on the sentences of the language $L = \{R_1, R_2, \ldots, R_q\}$ which satisfies IP (and Ex). By Theorem 20 w can be represented in the form (60). Let $\theta = \theta(a_1, \ldots, a_n) \in QFSL$ and let $\theta' = \theta(a_{n+1}, \ldots, a_{2n})$. Then by IP,

$$\begin{aligned}
0 &= 2(w'(\theta \wedge \theta') - w(\theta) \cdot w(\theta')) \\
&= \int_A {}^\circ\omega^\Psi(\theta \wedge \theta')\, d\mu(\Psi) + \int_A {}^\circ\omega^\Lambda(\theta \wedge \theta')\, d\mu(\Lambda) \\
&\quad -2\left(\int_A {}^\circ\omega^\Psi(\theta)\, d\mu(\Psi)\right) \cdot \left(\int_A {}^\circ\omega^\Lambda(\theta')\, d\mu(\Lambda)\right) \\
&= \int_A {}^\circ\omega^\Psi(\theta) \cdot {}^\circ \omega^\Psi(\theta')\, d\mu(\Psi) + \int_A {}^\circ\omega^\Lambda(\theta) \cdot {}^\circ \omega^\Lambda(\theta')\, d\mu(\Lambda) \\
&\quad -2\left(\int_A {}^\circ\omega^\Psi(\theta)\, d\mu(\Psi)\right) \cdot \left(\int_A {}^\circ\omega^\Lambda(\theta')\, d\mu(\Lambda)\right) \\
&= \int_A \int_A ({}^\circ\omega^\Psi(\theta) - {}^\circ \omega^\Lambda(\theta))^2\, d\mu(\Psi) d\mu(\Lambda)
\end{aligned}$$

since by Ex, ${}^\circ\omega^\Psi(\theta) = {}^\circ \omega^\Psi(\theta')$ etc..

Using the countable additivity of μ we can see that there must be a subset C of A with μ measure 1 such that ${}^\circ\omega^\Psi(\theta)$ is constant on C for *each* $\theta \in QFSL$. Picking a particular $\Lambda \in C$ then we have that

$$w(\theta) = \int_A {}^\circ\omega^\Psi(\theta)\, d\mu(\Psi) = {}^\circ\omega^\Lambda(\theta)$$

for $\theta \in QFSL$, so $w = {}^\circ\omega^\Lambda$. $\square$

Note that using Proposition 7 in the case of unary L gives de Finetti's Representation Theorem, at least after taking a suitable coarsening of the measure μ under the map

$$\Psi = \bigwedge_{i=1}^{\nu} \alpha_{h_i}(a_i) \mapsto \langle {}^\circ(\nu_1/\nu), {}^\circ(\nu_2/\nu), \ldots, {}^\circ(\nu_{2^q}/\nu) \rangle$$

where (in U^*) $\nu_j = |\{i \,|\, h_i = j\}|$.

7 Analogy

In this section we shall consider how reasoning *by analogy* could be captured within inductive logic and employ the theory developed last time to shed some light on this question.

Reasoning by analogy is widely used in our common everyday reasoning, probably more so than the rules of the classical logic. To give a random example, knowing that one's son greatly enjoyed a birthday party at the local swimming pool when he was 10 year old might lead to a parent recommending this possibility to a friend who is wishes to arrange a party for his 11 year old daughter and her friends to mark the end of their Junior School years.

One common approach to analogy *in the unary context* is to argue that, for example, atoms

$$\alpha(x) = R_1(x) \wedge \neg R_2(x) \wedge R_3(x) \wedge R_4(x),$$
$$\alpha'(x) = R_1(x) \wedge \neg R_2(x) \wedge \neg R_3(x) \wedge R_4(x),$$

are somewhat 'analogous' since they agree on the signs (i.e. negated or unnegated) of R_1, R_2, R_4 and only disagree in sign on the single relation symbol R_3 and hence learning $\alpha(a_{n+1})$ should add some support to one's believing $\alpha'(a_{n+2})$.

More generally, we define the *Hamming distance* $\| \alpha_k - \alpha_j \|$ of atoms

$$\alpha_k(x) = \bigwedge_{i=1}^{q} R_i^{t_i}(x), \qquad \alpha_j(x) = \bigwedge_{i=1}^{q} R_i^{s_i}(x)$$

to be $\sum_{i=1}^{q} |t_i - s_i|$, that is, the number of R_i on which α_k and α_j differ. (So for the Hamming distance between α and α' above is 1.) The principle suggested above becomes

Strong Analogy Principle, SAP (Analogy by Proximity)

For $\alpha_m(x), \alpha_j(x), \alpha_k(x)$ atoms of L, $j \neq k$ and consistent $\psi(a_1, a_2, \ldots, a_n) \in QFSL$, if

$$\| \alpha_m(x) - \alpha_j(x) \| < \| \alpha_m(x) - \alpha_k(x) \|$$

then

$$w(\alpha_m(a_{n+2}) \mid \alpha_j(a_{n+1}) \wedge \psi(\vec{a})) > w(\alpha_m(a_{n+2}) \mid \alpha_k(a_{n+1}) \wedge \psi(\vec{a})). \qquad (61)$$

Unfortunately, even when $q = 2$, hardly any probability functions satisfy this principle together with the basic symmetry principles Ex, Px and SN. For languages with three or more predicates it is even worse, there are no solutions. Modifying the principle in various ways (for example, taking state descriptions in place of arbitrary quantifier free sentence $\psi(\vec{a})$) yields more solutions, even when $q \geq 3$, but no clear picture emerges as was the case for example with JSP.

An alternative approach could be based on arguing that objects are more likely to be found similar in certain respects if they are already known to be similar in other respects. More generally, *probabilistic support for future similarities should be an increasing function of known past similarities.*

There are various ways in which this could be formulated within inductive logic. An obvious basic possibility, again within the unary context, is

The General Analogy Principle, GAP

For $\vec{a} = \langle a_3, a_4, \ldots, a_k \rangle$ *and* $\psi(a_1, \vec{a}), \phi(a_1, \vec{a}) \in QFSL$,

$$w(\phi(a_2, \vec{a}) \,|\, \psi(a_1, \vec{a}) \wedge \psi(a_2, \vec{a}) \wedge \phi(a_1, \vec{a})) \geq w(\phi(a_2, \vec{a}) \,|\, \phi(a_1, \vec{a}))$$

Hwever, it can be shown that the only probability function satisfying this and Px, Ex and SN, is c_0^L, see [5].

These two ways of formalising analogical reasoning do not seem to work. However, there is another way, applicable both for unary and polyadic languages, which appears to capture analogy better. In the rest of this section, L stands for a general language, polyadic or unary.

The Counterpart Principle, CP (Analogy by Structural Similarity)

Let $\theta, \theta' \in SL$ *be such that* θ' *is the result of replacing some constant/relation symbols in* θ *by new constant/relation symbols of the same arity not occurring in* θ. *Then*

$$w(\theta \,|\, \theta') \geq w(\theta). \tag{62}$$

A stronger version, SCP: *If moreover* θ'' *is the result of replacing the same and possibly also other constant/relation symbols in* θ *by new constant/relation* *symbols of the same arity not occurring in θ then*

$$w(\theta \,|\, \theta') \geq w(\theta \,|\, \theta'') \geq w(\theta). \tag{63}$$

Remarkably, it turns out that the Counterpart principle as well as its stronger version hold for any probability function which is a member of a language invariant family of probability functions with a probability function for any language (rather than just unary languages, as in ULI). Formally,

Language Invariance Principle, Li

A probability function w for a language L satisfies Language Invariance if there is a family of probability functions $w^{\mathcal{L}}$, one on each language $\mathcal{L}$, all satisfying Px and Ex, and such that $w^L = w$ and if $\mathcal{L} \subseteq \mathcal{L}'$ then $w^{\mathcal{L}}$ is $w^{\mathcal{L}'}$ restricted to $S\mathcal{L}$.

Showing that Li is enough for the basic version of the Counterpart Principle is easier, so we begin with that:

Theorem 22 *If w satisfies Li then w satisfies the CP.*

Proof Assume that w satisfies Li. Taking functions of the family together we can obtain a probability function w^+ for the infinite language L^+ which contains infinitely many relation symbols of each arity, extends w and satisfies Px and Ex. Let θ, θ' be as in the statement of the principle. Assume without loss of generality that the constant symbols appearing in θ are amongst $a_1, a_2, \ldots, a_t, a_{t+1}, \ldots, a_{t+k}$, all the relation symbols appearing in θ are amongst $R_1, R_2, \ldots, R_s, R_{s+1}, \ldots, R_{s+j}$, and that to form θ', $a_{t+1}, \ldots, a_{t+k}$ were replaced by $a_{t+k+1}, a_{t+k+2}, \ldots, a_{t+2k}$, and $R_{s+1}, \ldots, R_{s+j}$ were replaced by $R_{s+j+1}, \ldots, R_{s+2j}$ respectively. So with the obvious notation we can write

$$\theta = \theta(a_1, \ldots, a_t, a_{t+1}, \ldots, a_{t+k}, R_1, \ldots, R_s, R_{s+1}, \ldots, R_{s+j}),$$
$$\theta' = \theta(a_1, \ldots, a_t, a_{t+k+1}, \ldots, a_{t+2k}, R_1, \ldots, R_s, R_{s+j+1}, \ldots, R_{s+2j}).$$

With this notation let θ_{i+1} be

$$\theta(a_1, \ldots, a_t, a_{t+ik+1}, \ldots, a_{t+(i+1)k}, R_1, \ldots, R_s, R_{s+ij+1}, \ldots, R_{s+(i+1)j}),$$

so $\theta_1 = \theta$, $\theta_2 = \theta'$. (It is understood that relation symbols in the blocks $R_{s+ij+1}, \ldots, R_{s+(i+1)j}$ are of appropriate arities.)

Let $\mathcal{L}$ be the unary language with a single unary relation symbol R and define $\tau : QFS\mathcal{L} \to QFSL^+$ by

$$\tau(R(a_i)) = \theta_i, \quad \tau(\neg\phi) = \neg\tau(\phi), \quad \tau(\phi \wedge \psi) = \tau(\phi) \wedge \tau(\psi), \qquad \text{etc.}$$

for $\phi, \psi \in QFS\mathcal{L}$.

Now define $v : QFS\mathcal{L} \to [0,1]$ by

$$v(\phi) = w^+(\tau(\phi)).$$

Then since w^+ satisfies (P1-2) (on SL^+) so does v (on $QFS\mathcal{L}$). Also since w^+ satisfies Ex and Px, for $\phi \in QFS\mathcal{L}$, permuting the θ_i in $\tau(\phi)$ will leave $w^+(\tau(\phi))$ unchanged so permuting the a_i in ϕ will leave $v(\phi)$ unchanged. Hence v satisfies Ex.

By Gaifman's Theorem, v has an extension to a probability function on $S\mathcal{L}$ satisfying Ex and hence satisfying PIR by Theorem 9. In particular then

$$v(R(a_1) \,|\, R(a_2)) \geq v(R(a_1)). \tag{64}$$

But since $\tau(R(a_1)) = \theta$, $\tau(R(a_2)) = \theta'$ this amounts to just the Counterpart Principle

$$w(\theta \,|\, \theta') \geq w(\theta).$$

□

To show that Li is enough also for the stronger version of CP, we need a lemma. I will be proved via the approximate representation theorem for probability functions satisfying Ex on the language $\mathcal{L}$ containing just one binary relations symbol, shown in the previous section.

Lemma 23 *For a probability function w on $S\mathcal{L}$ satisfying Ex*

$$w(R(a_1, a_2) \,|\, R(a_1, a_4)) \geq w(R(a_1, a_2) \,|\, R(a_3, a_4)).$$

Proof Since w satisfies Ex, we have $w(R(a_1, a_4)) = w(R(a_3, a_4))$ so it suffices to show that

$$w(R(a_1, a_2) \wedge R(a_1, a_4)) \geq w(R(a_1, a_2) \wedge R(a_3, a_4)).$$

In view of Theorem 19 it suffices to prove it for the functions w^D.

Let $D = (d_{i,j})$ be an $N \times N$ $\{0,1\}$-matrix and let

$$e_i = |\{j \mid d_{i,j} = 1\}|\,.$$

Then

$$w^D(R(a_1,a_2) \wedge R(a_1,a_4)) = \left(\sum_{i=1}^{N} e_i^2\right) N^{-3} = \left(\sum_{i=1}^{N} (e_i/N)^2\right) \left(\sum_{i=1}^{N} (1/N)^2\right),$$

$$w^D(R(a_1,a_2) \wedge R(a_3,a_4)) = \left(\sum_{i=1}^{N}\sum_{j=1}^{N} e_i e_j\right) N^{-4} = \left(\sum_{i=1}^{N} (e_i/N)(1/N)\right)^2.$$

By the Cauchy-Schwarz Inequality

$$\left(\sum_{i=1}^{N} (e_i/N)^2\right) \left(\sum_{i=1}^{N} (1/N)^2\right) \geq \left(\sum_{i=1}^{N} (e_i/N)(1/N)\right)^2$$

so the result follows.

□

Theorem 24 *If w satisfies Li then w satisfies the SCP.*

Proof We can proceed similarly as in the case of SP. It is convenient to introduce the following notation: for a natural number c, define

$$\overline{c} = 2c - 1, \qquad \underline{c} = 2c\,.$$

Without loss of generality, assume that $\theta, \theta', \theta''$ have the following form which will be convenient for the proof:

$$\begin{aligned}
\theta = \theta(&a_1, \ldots, a_m, a_{\overline{m+1}}, \ldots, a_{\overline{m+t}}, a_{\underline{m+k+1}}, \ldots, a_{\underline{m+2k}}, \\
&R_1, \ldots, R_p, R_{\overline{p+1}}, \ldots, R_{\overline{p+s}}, R_{\underline{p+j+1}}, \ldots, R_{\underline{p+2j}}), \\
\theta' = \theta(&a_1, \ldots, a_m, a_{\overline{m+1}}, \ldots, a_{\overline{m+t}}, a_{\underline{m+3k+1}}, \ldots, a_{\underline{m+4k}}, \\
&R_1, \ldots, R_p, R_{\overline{p+1}}, \ldots, R_{\overline{p+s}}, R_{\underline{p+3j+1}}, \ldots, R_{\underline{p+4j}}), \\
\theta'' = \theta(&a_1, \ldots, a_m, a_{\overline{m+2t+1}}, \ldots, a_{\overline{m+3t}}, a_{\underline{m+3k+1}}, \ldots, a_{\underline{m+4k}}, \\
&R_1, \ldots, R_p, R_{\overline{p+2s+1}}, \ldots, R_{\overline{p+3s}}, R_{\underline{p+3j+1}}, \ldots, R_{\underline{p+4j}}).
\end{aligned}$$

(where the relation symbols in the same positions have the same arities). Assume w satisfies Li , θ, θ' and θ'' are in SL and L^+, w^+ are as above.

Let $\theta_{i+1,l+1}$ stand for

$$\theta(a_1, \ldots, a_m, a_{\overline{m+it+1}}, \ldots, a_{\overline{m+(i+1)t}}, a_{\underline{m+lk+1}}, \ldots, a_{\underline{m+(l+1)k}},$$

$$R_1, \ldots, R_p, R_{\overline{p+is+1}}, \ldots, R_{\overline{p+(i+1)s}}, R_{\underline{p+lj+1}}, \ldots, R_{\underline{p+(l+1)j}}),$$

so $\theta = \theta_{1,2}$, $\theta' = \theta_{1,4}$ and $\theta'' = \theta_{3,4}$.

Let $\mathcal{L}$ be the binary language with a single binary relation symbol R. Define $\tau : QFS\mathcal{L} \to QFSL^+$ by

$$\tau(R(a_i, a_l)) = \theta_{i,l}, \quad \tau(\neg\phi) = \neg\tau(\phi), \quad \tau(\phi \wedge \psi) = \tau(\phi) \wedge \tau(\psi), \qquad \text{etc.}$$

and define $v : QFS\mathcal{L} \to [0, 1]$ by

$$v(\phi) = w^+(\tau(\phi)).$$

The v extends to a a probability function on $S\mathcal{L}$ which satisfies Ex. By Lemma 23.

$$v(R(a_1, a_2)|R(a_1, a_4)) \geq v(R(a_1, a_2)|R(a_3, a_4))$$

so

$$w(\theta_{1,2} \,|\, \theta_{1,4}) \geq w(\theta_{1,2} \,|\, \theta_{3,4})$$

and the result follows.

$\square$

8 Spectrum Exchangeability

Let L be a general polyadic language again. For a state description $\Theta(b_1, b_2, \ldots, b_m)$ of L and $i, j \in \{1, \ldots, n\}$ define $b_i \sim_\Theta b_j$ if b_i and b_j behave in exactly the same way within Θ, meaning that replacing some arbitrary occurrence(s) of b_i by b_j and/or vice versa does not render θ inconsistent. In other words, for any r-ary relation symbol from L and not necessarily distinct $b_{k_1}, \ldots, b_{k_u}, b_{k_{u+2}}, \ldots, b_{k_r}$ from $\{b_1, b_2, \ldots, b_m\}$,

$$\Theta(b_1, b_2, \ldots, b_m) \models R(b_{k_1}, \ldots, b_{k_u}, b_i, b_{k_{u+2}}, \ldots, b_{k_r}) \\ \leftrightarrow \; R(b_{k_1}, \ldots, b_{k_u}, b_j, b_{k_{u+2}}, \ldots, b_{k_r}).$$

Another way of expressing this is to say that upon adding equality with its axioms to our framework, $\Theta(b_1, \ldots, b_r)$ and b_i, b_j are such that $b_i = b_j$ is consistent with Θ. Clearly, $\sim_\Theta$ is an equivalence relation on $\{b_1, \ldots, b_m\}$.

Define the *Spectrum of* Θ, $\mathcal{S}(\Theta)$ to be the multiset of sizes of the equivalence classes of $\sim_\Theta$. The set of possible spectra for n constants, that is, the set of of multisets of positive natural numbers summing up to n, is denoted $Spec(n)$.

Note that for a unary language L and state description $\Theta(b_1, \ldots, b_m) = \bigwedge_{i=1}^m \alpha_{h_i}(b_i)$, $b_i \sim_\Theta b_j$ just when $h_i = h_j$, that is, just when b_i and b_j satisfy the same atom. Hence this new definition of spectrum generalizes the unary definition given earlier.

The Spectrum Exchangeability Principle, Sx

For state descriptions $\Theta(b_1, b_2, \ldots, b_m), \Phi(b_1, b_2, \ldots, b_m) \in SL$, *if* $\mathcal{S}(\Theta) = \mathcal{S}(\Phi)$ *then* $w(\Theta) = w(\Phi)$.

Clearly, Sx implies Ex. For a unary language L it is - in the presence of Ex - equivalent to Ax.

Example Let L contain just a single binary relation symbol R. The conjunction $\Theta(a_1, a_2, a_3, a_4)$ of

$$\begin{array}{cccc} R(a_1,a_1) & \neg R(a_1,a_2) & R(a_1,a_3) & R(a_1,a_4) \\ R(a_2,a_1) & \neg R(a_2,a_2) & R(a_2,a_3) & \neg R(a_2,a_4) \\ R(a_3,a_1) & \neg R(a_3,a_2) & R(a_3,a_3) & R(a_3,a_4) \\ R(a_4,a_1) & R(a_4,a_2) & R(a_4,a_3) & R(a_4,a_4) \end{array}$$

is a state description, $\sim_\Theta$ has equivalence classes $\{a_1, a_3\}$, $\{a_2\}$, $\{a_4\}$ and the spectrum $\mathcal{S}(\Theta)$ is $\{2, 1, 1\}$.

The conjunction $\Phi(a_1, a_2, a_3, a_4)$ of

$$\begin{array}{cccc}
\neg R(a_1,a_1) & \neg R(a_1,a_2) & R(a_1,a_3) & R(a_1,a_4) \\
\neg R(a_2,a_1) & \neg R(a_2,a_2) & \neg R(a_2,a_3) & \neg R(a_2,a_4) \\
R(a_3,a_1) & R(a_3,a_2) & R(a_3,a_3) & R(a_3,a_4) \\
R(a_4,a_1) & R(a_4,a_2) & R(a_4,a_3) & R(a_4,a_4)
\end{array}$$

is also a state description, $\sim_\Phi$ has equivalence classes $\{a_1\}$, $\{a_2\}$, $\{a_3, a_4\}$ and the spectrum is again $\{2, 1, 1\}$. According to Sx, $w(\Theta) = w(\Phi)$.

(As we did before, if we represent the above state descriptions by matrices, the spectra are easier to see.)

Sx is quite strong. We shall see that it sheds light on some obvious questions one might ask, most notably as regards heeding relevant information. Before studying it in some detail, consider the following natural principle, which does not arise in the unary context:

The Variable Exchangeability Principle, Vx

Let R be an r-ary relation symbol of L, $\sigma \in \mathsf{S}_r$ and for $\theta \in SL$ let θ' be the result of replacing each $R(t_1, t_2, \ldots, t_r)$ appearing in θ by $R(t_{\sigma(1)}, t_{\sigma(2)}, \ldots, t_{\sigma(r)})$. Then $w(\theta) = w(\theta')$.

The t_i above are terms - constant symbols or variables. For example, for binary R by Vx we should have

$$w(R(a_1, a_1) \wedge \neg R(a_2, a_2) \wedge R(a_1, a_2)) = w(R(a_1, a_1) \wedge \neg R(a_2, a_2) \wedge R(a_2, a_1)).$$

The argument for Vx as a rational principle is that there is no reason why $R(x_1, x_2, \ldots, x_r)$ should be treated differently from $R(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(r)})$. Note that Vx *does not follow* from Ex, see Problem 26.

It is easy to see that Vx for just state descriptions implies Vx as above. Furthermore, since for a state description $\Theta(b_1, \ldots, b_m)$, Θ' has the same spectrum as Θ, we have:

Proposition 25 *Sx implies Vx.*

The following lemma will be useful. We omit the proof as it is somewhat technical, see [1, Chapter 27].

Lemma 26 *Suppose that the state descriptions* $\Theta(a_1, \ldots, a_m), \Phi(a_1, \ldots, a_m)$ *of* L *have the same spectrum and let* $\tilde{n} = \{n_1, n_2, \ldots, n_s\}$ *be a multiset with* $n_1, \ldots, n_s \in \mathbb{N}^+$ *and* $n = \sum_{i=1}^{s} n_i > m$. *Then the number of state descriptions for* $a_1, a_2, \ldots, a_n$ *with spectrum* $\tilde{n}$ *extending* Θ *is the same as the number with this spectrum extending* Φ.

In fact, the result holds even when we compare the numbers of state descriptions for $a_1, a_2, \ldots, a_n$ with spectrum $\tilde{n}$ extending Θ, where these state descriptions are *for another, larger language* $\mathcal{L} \supset L$.

Next we shall introduce the quintessential functions satisfying Sx, namely the $u^{\bar{p},L}$ which can be used to provide an integral representations of all all probability functions satisfying Sx.

Let $\mathbb{B}$ be the set of infinite sequences

$$\bar{p} = \langle p_0, p_1, p_2, p_3, \ldots \rangle$$

of non-negative reals such that $p_1 \geq p_2 \geq p_3 \geq \ldots$ and $\sum_{i=0}^{\infty} p_i = 1$. We think of the subscripts $0, 1, 2, \ldots$ as *colours* and p_i as the probability of picking colour i. 0 stands for 'black' and it has a special status.

Let $c_1, c_2, \ldots, c_m$ be a sequence of colours, so $\vec{c} = \langle c_1, c_2, \ldots, c_m \rangle \in \mathbb{N}^m$. We say that a state description $\Theta(b_1, b_2, \ldots, b_m)$ is *consistent with* $\vec{c}$ if whenever $c_s = c_t \neq 0$ then $b_s \sim_\Theta b_t$. Note that $c_s = c_t = 0$ imposes no requirement on b_s being equivalent to b_t or not.

Let $\mathcal{C}(\vec{c}, \vec{b})$ be the set of all state descriptions for $\vec{b} = \langle b_1, b_2, \ldots, b_m \rangle$ consistent with $\vec{c}$. Note that $\mathcal{C}(\emptyset, \emptyset)$ contains just $\top$. When considering $\Theta(\vec{b}) \in \mathcal{C}(\vec{c}, \vec{b})$ we say that b_s has colour c_s.

The number of state descriptions in $\mathcal{C}(\vec{c}, \vec{b})$ can be expressed as follows. Let $b_{g_1}, b_{g_2}, \ldots, b_{g_t}$ include exactly one representative for each of the non-black colours, along with all the black b_s. Any state description Θ in $\mathcal{C}(\vec{c}, \vec{b})$ is determined by choosing all the $\pm R_i(b_{j_1}, \ldots, b_{j_{r_i}})$ with $i \in \{1, \ldots, q\}$ and $\langle b_{j_1}, \ldots, b_{j_{r_i}} \rangle \in \{b_{g_1}, b_{g_2}, \ldots, b_{g_t}\}^{r_i}$ since after that everything in Θ is fixed by the requirement of indistinguishability for b_s, b_t with the same non-black colour. This observation yields the following lemma.

Lemma 27 *Let $b_1, \ldots, b_m, b_{m+1}$ be distinct and let*

$$\begin{aligned}\vec{b} &= \langle b_1, \ldots, b_m \rangle\\ \vec{b^+} &= \langle b_1, \ldots, b_{m+1} \rangle\\ \vec{c} &= \langle c_1, \ldots, c_m \rangle\\ \vec{c}^+ &= \langle c_1, \ldots, c_{m+1} \rangle\end{aligned}$$

Let $\Theta(\vec{b}) \in \mathcal{C}(\vec{c}, \vec{b})$. Then the number of $\Phi(\vec{b}^+) \in \mathcal{C}(\vec{c}^+, \vec{b}^+)$ extending Θ equals

$$\frac{|\,\mathcal{C}(\vec{c}^+, \vec{b}^+)|}{|\,\mathcal{C}(\vec{c}, \vec{b})|}$$

(and as such it is independent of the choice of Θ).

Proof If $c_{m+1} = c_s \neq 0$ for some $s \leq m$ then the result is clear since $|\,\mathcal{C}(\vec{c}^+, \vec{b}^+)| = |\,\mathcal{C}(\vec{c}, \vec{b})|$ and there is exactly one extension of $\Theta(\vec{b})$ in $\mathcal{C}(\vec{c}^+, \vec{b}^+)$, namely the one in which b_{m+1} joins the same equivalence class as b_s. If c_{m+1} is black or a new colour then - using the observation and notation preceding the lemma - extensions of Θ that lie in $\mathcal{C}(\vec{c}^+, \vec{b}^+)$ are formed by choosing $\pm R_i(b_{j_1}, \ldots, b_{j_{r_i}})$ where b_{m+1} *does appear* in $b_{j_1}, \ldots, b_{j_{r_i}}$ and otherwise only the $b_{g_1}, \ldots, b_{g_t}$ are allowed to appear. Since any state description in $\mathcal{C}(\vec{c}^+, \vec{b}^+)$ is determined by making these choices when the $b_{j_1}, \ldots, b_{j_{r_i}}$ are simply from $\{b_{g_1}, \ldots, b_{g_t}, b_{m+1}\}$, the result follows. $\square$

Now define for a state description $\Theta(a_1, a_2, \ldots, a_m)$,

$$u^{\bar{p},L}(\Theta) = \sum_{\substack{\vec{c} \in \mathbb{N}^m \\ \Theta \in \mathcal{C}(\vec{c}, \vec{a})}} |\,\mathcal{C}(\vec{c}, \vec{a})|^{-1} \prod_{s=1}^{m} p_{c_s}. \tag{65}$$

Lemma 28 *$u^{\bar{p},L}$ as defined by (65) extends uniquely to a probability function on SL.*

Proof It suffices to show that conditions (i),(ii),(iii) from (9) hold. The first two are obvious. For (iii) notice that for $\vec{c} = \langle c_1, c_2, \ldots, c_m \rangle$, $\vec{c}^+ = \langle c_1, c_2, \ldots, c_{m+1} \rangle$, $\vec{a} = \langle a_1, \ldots, a_m \rangle$, $\vec{a}^+ = \langle a_1, \ldots, a_{m+1} \rangle$,

$$\sum_{\substack{\Phi(\vec{a}^+) \models \Theta(\vec{a}) \\ \Phi \in \mathcal{C}(\vec{c}^+, \vec{a}^+)}} |\,\mathcal{C}(\vec{c}^+, \vec{a}^+)|^{-1} p_{c_{m+1}} \prod_{s=1}^{m} p_{c_s}$$

is equal to 0 if $\Theta(\vec{a}) \notin \mathcal{C}(\vec{c},\vec{a})$ and

$$|\,\mathcal{C}(\vec{c},\vec{a})|^{-1} p_{c_{m+1}} \prod_{s=1}^{m} p_{c_s}$$

otherwise, by Lemma 27. Hence

$$\begin{aligned}\textstyle\sum_{\vec{c}^+\in\mathbb{N}^{m+1}} \sum_{\substack{\Phi(\vec{a}^+)\models\Theta(\vec{a})\\ \Phi\in\mathcal{C}(\vec{c}^+,\vec{a}^+)}} |\,\mathcal{C}(\vec{c}^+,\vec{a}^+)|^{-1} p_{c_{m+1}} \prod_{s=1}^{m} p_{c_s} \\ = \textstyle\sum_{\substack{\vec{c}\in\mathbb{N}^m\\ \Theta\in\mathcal{C}(\vec{c},\vec{a})}} \sum_{c_{m+1}\in\mathbb{N}} |\,\mathcal{C}(\vec{c},\vec{a})|^{-1} p_{c_{m+1}} \prod_{s=1}^{m} p_{c_s}\end{aligned}$$

which by summing out the c_{m+1} gives the required result. □

The functions c_0^L, defined in the general polyadic case analogously to the unary case (see Problem 23), and c_∞^L, are special cases of the $u^{\bar{p},L}$: c_0^L when $\bar{p} = \langle 0,1,0,0,\ldots\rangle$ and c_∞^L when $\bar{p} = \langle 1,0,0,\ldots\rangle$, as apparent from comparing values given to state descriptions.

Lemma 29 *$u^{\bar{p},L}$ satisfies Ex and Sx.*

Proof Ex follows[1] by Lemma 4 and Theorem 5.

It remains to show that $u^{\bar{p},L}$ satisfies Sx. Note that if state descriptions $\Theta(a_1,\ldots,a_m), \Phi(a_1,\ldots,a_m)$ have the same spectrum then there is a permutation of $\{1,\ldots,m\}$ such that for

$$\Psi(a_1,\ldots,a_m) = \Phi(a_{\sigma(1)},\ldots,a_{\sigma(m)}),$$

Θ and Ψ have the same equivalences of indiscernibility, that is, $\sim_\Theta = \sim_\Psi$. Since we already know that $u^{\bar{p},L}$ satisfies Ex, it suffices to show that for Θ, Ψ with the same equivalence of indiscernibility,

$$u^{\bar{p},L}(\Theta(a_1,\ldots,a_m)) = u^{\bar{p},L}(\Psi(a_1,\ldots,a_m)).$$

But this clearly holds since for any $\vec{c} \in \mathbb{N}^m$ we have $\Theta \in \mathcal{C}(\vec{c},\vec{a})$ just if $\Psi \in \mathcal{C}(\vec{c},\vec{a})$, so Sx follows. □

Moreover, as discussed below, the $u^{\bar{p},L}$ functions satisfy WIP and - with $\bar{p}$ fixed - also Li.

[1] The condition in Lemma 4 is apparent upon noticing that if $\Psi(a_1,\ldots,a_m) = \Phi(a_{\tau(1)},\ldots,a_{\tau(m)})$ then $\Psi \in \mathcal{C}(\tau^{-1}\vec{c},\vec{a})$ just when $\Phi \in \mathcal{C}(\vec{c},\vec{a})$ where $\tau^{-1}\vec{c} = \langle c_{\tau^{-1}(1)},\ldots,c_{\tau^{-1}(m)}\rangle$, and that $|\mathcal{C}(\tau^{-1}\vec{c},\vec{a})| = |\mathcal{C}(\vec{c},\vec{a})|$.

Theorem 30 *The $u^{\bar{p},L}$ satisfy Li with Sx and WIP.*

Proof To show that the $u^{\bar{p},L}$ satisfy Li let $\mathcal{L}$ be a language properly extending L. It is enough to show that $u^{\bar{p},\mathcal{L}}$ agrees with $u^{\bar{p},L}$ on state descriptions $\Theta(a_1,\ldots,a_m)$ of L.

We have

$$u^{\bar{p},L}(\Theta) = \sum_{\substack{\vec{c}\in\mathbb{N}^m \\ \Theta\in\mathcal{C}^L(\vec{c},\vec{a})}} |\,\mathcal{C}^L(\vec{c},\vec{a})|^{-1} \prod_{s=1}^m p_{c_s},$$

and

$$u^{\bar{p},\mathcal{L}}(\Theta) = \sum_{\Phi} \sum_{\substack{\vec{c}\in\mathbb{N}^m \\ \Phi\in\mathcal{C}^{\mathcal{L}}(\vec{c},\vec{a})}} |\,\mathcal{C}^{\mathcal{L}}(\vec{c},\vec{a})|^{-1} \prod_{s=1}^m p_{c_s}, \tag{66}$$

where the $\Phi(a_1,\ldots,a_m)$ are the state descriptions of $\mathcal{L}$ which logically imply $\Theta(a_1,\ldots,a_m)$ and the appropriate superscripts have been added to the $\mathcal{C}(\vec{c},\vec{a})$.

Any such Φ can be expressed as

$$\Theta(a_1,\ldots,a_m) \wedge \Phi'(a_1,\ldots,a_m)$$

where $\Phi'(a_1,\ldots,a_m)$ is a state description of the language $\mathcal{L}-L$. Furthermore

$$\Phi \in \mathcal{C}^{\mathcal{L}}(\vec{c},\vec{a}) \iff [\Theta \in \mathcal{C}^L(\vec{c},\vec{a}) \ \text{ and } \ \Phi' \in \mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})].$$

Hence

$$|\mathcal{C}^{\mathcal{L}}(\vec{c},\vec{a})| = |\mathcal{C}^L(\vec{c},\vec{a})| \cdot |\mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})|$$

and from (66) we obtain

$$\begin{aligned}
u^{\bar{p},\mathcal{L}}(\Theta) &= \sum_{\Phi'} \sum_{\substack{\vec{c}\in\mathbb{N}^m,\, \Theta\in\mathcal{C}^L(\vec{c},\vec{a}) \\ \Phi'\in\mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})}} |\,\mathcal{C}^L(\vec{c},\vec{a})|^{-1} |\,\mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})|^{-1} \prod_{s=1}^m p_{c_s}, \\
&= \sum_{\substack{\vec{c}\in\mathbb{N}^m \\ \Theta\in\mathcal{C}^L(\vec{c},\vec{a})}} |\,\mathcal{C}^L(\vec{c},\vec{a})|^{-1} \left(\sum_{\Phi'\in\mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})} |\mathcal{C}^{\mathcal{L}-L}(\vec{c},\vec{a})|^{-1} \right) \prod_{s=1}^m p_{c_s}, \\
&= \sum_{\substack{\vec{c}\in\mathbb{N}^m \\ \Theta\in\mathcal{C}^L(\vec{c},\vec{a})}} |\,\mathcal{C}^L(\vec{c},\vec{a})|^{-1} \prod_{s=1}^m p_{c_s} \ = \ u^{\bar{p},L}(\Theta),
\end{aligned}$$

as required.

We have already shown that the $u^{\bar{p},L}$ satisfy Sx in Lemma 29. For WIP see Problem 27.

□

Remarkably, the converse to Theorem 30 also holds: the $u^{\bar{p},L}$ are the only probability functions satisfying Li with WIP and Sx. (This can be proved using the representation theorem below and the same method as that of Proposition 21 itn he extra reading above.)

We shall now state the representation theorem for functions satisfying Li with Sx and indicate how it can be proved. The full proof is long and technical, see [1, Chapter 31].

Theorem 31 *Let w^L be a probability function on SL satisfying Li with Sx. Then there is a measure μ on $\mathbb{B}$ such that*

$$w^L = \int_{\mathbb{B}} u^{\bar{p},L}\, d\mu(\bar{p}). \tag{67}$$

Conversely given such a measure μ, w^L defined by (67) is a probability function on SL satisfying Li with Sx.

Proof For the converse, note that by Theorem 30 for any fixed $\bar{p}$ the $u^{\bar{p},\mathcal{L}}$ form a language invariant family, so $\int_{\mathbb{B}} u^{\bar{p},\mathcal{L}}\, d\mu(\bar{p})$ is a language invariant family containing w.

Now let w^L be a probability function on SL satisfying Li with Sx, and denote other members of a fixed language invariant family satisfying Sx and containing w^L accordingly by $w^{\mathcal{L}}$ (so $w^{\mathcal{L}}$ is on $S\mathcal{L}$).

By the remark following Lemma 26, we see that we can unambiguously define $\mathcal{N}^{L,\mathcal{L}}(\tilde{m},\tilde{n})$ to be the number of state descriptions *in the language* $\mathcal{L}$ for $a_1, a_2, \ldots, a_n$ with spectrum $\tilde{n}$ that extend some/any fixed state description $\Theta(a_1, \ldots, a_m)$ in the language L with spectrum $\tilde{m}$.

For a state description $\Theta(a_1, \ldots, a_m)$ in the language L, $n > m$ and $L \subseteq \mathcal{L}$ we have

$$w^L(\Theta(a_1, \ldots, a_m)) = \sum_{\Psi(a_1,\ldots,a_n) \models \Theta(a_1,\ldots a_m)} w^{\mathcal{L}}(\Psi(a_1, \ldots, a_n))$$

where the Ψ are state descriptions in the language $\mathcal{L}$. We can group them according to their spectrum and make the above

$$w^L(\Theta(a_1,\ldots,a_m)) = \sum_{\tilde{n}\in Spec(n)} \sum_{\substack{\Psi\models\Theta \\ \mathcal{S}(\Psi)=\tilde{n}}} w^{\mathcal{L}}(\Psi(a_1,\ldots,a_n)).$$

Since the $w^{\mathcal{L}}$ all satisfy Sx, we can define $w^{\mathcal{L}}(\tilde{n})$ to be $w^{\mathcal{L}}(\Psi(a_1,\ldots,a_n))$ where $\Psi(a_1,\ldots,a_n)$ is some/any state description in the language $\mathcal{L}$ with $\mathcal{S}(\Psi)=\tilde{n}$. If the spectrum of $\Theta(a_1,\ldots,a_m)$ is $\tilde{m}$, the above gives

$$\begin{aligned} w^L(\Theta(a_1,\ldots,a_m)) \;=\; w^L(\tilde{m}) \;&=\; \sum_{\tilde{n}\in Spec(n)} \mathcal{N}^{L,\mathcal{L}}(\tilde{m},\tilde{n})w^{\mathcal{L}}(\tilde{n}) \\ &= \sum_{\tilde{n}\in Spec(n)} \frac{\mathcal{N}^{L,\mathcal{L}}(\tilde{m},\tilde{n})}{\mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{n})}\mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{n})w^{\mathcal{L}}(\tilde{n})\,. \end{aligned}$$

(where $\mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{n})$ is the number of all state descriptions in the language $\mathcal{L}$ with spectrum $\tilde{n}$).

Now note that for a fixed $\mathcal{L}$ and n,

$$\sum_{\tilde{n}\in Spec(n)} \mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{n})w^{\mathcal{L}}(\tilde{n}) = 1.$$

For $\tilde{n}=\{n_1,\ldots,n_s\}\in Spec(n)$, assuming the $n_1,\ldots,n_s$ are listed in (not-necessarily-strictly) decreasing order,

$$\bar{p}(\tilde{n}) = \left\langle 0, \frac{n_1}{n},\ldots,\frac{n_s}{n},0,0,\ldots\right\rangle \in \mathbb{B}.$$

Let $\mu_n^{\mathcal{L}}$ be the discrete measure on $\mathbb{B}$ which puts measure

$$\mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{n})w^{\mathcal{L}}(\tilde{n})\,.$$

on this point and let

$$U_{\tilde{n}}^{\mathcal{L}}(\Theta(a_1,\ldots,a_m)) \;=\; \frac{\mathcal{N}^{L,\mathcal{L}}(\tilde{m},\tilde{\nu})}{\mathcal{N}^{\mathcal{L}}(\emptyset,\tilde{\nu})}\,.$$

Then

$$w^L(\Theta(a_1,\ldots,a_m)) \;=\; w^L(\tilde{m}) \;=\; \sum_{\tilde{n}\in Spec(n)} U^{\mathcal{L}}_{\tilde{n}}(\Theta(a_1,\ldots,a_m))\,\mu^{\mathcal{L}}_n(\bar{p}(\tilde{n}))\,. \quad (68)$$

It turns out that for large n and for $\mathcal{L}$ with a large number of unary predicaes, and for $\Theta(a_1,\ldots,a_m)$ with spectrum $\tilde{m}$, $U^{\mathcal{L}}_{\tilde{n}}(\Theta(a_1,\ldots,a_m))$ is close to $u^{\bar{p}(\tilde{n}),L}(\tilde{m})$. Summing over $\tilde{n} \in Spec(n)$ is the same as summing over the $\bar{p}(\tilde{n}) \in \mathbb{B}$. Taking the limit, or working in a nonstandard universe and using the Loeb theory, *eventually* yields the result. Note that in the process, also the $u^{\bar{p},L}$ with $p_0 > 0$ become involved.

$\square$

Problems

Problem 23 *(i) Show that c_∞^L satisfies Sx and decide if any V_M can satisfy it.*

(ii) Recall we defined c_0^L via (42) for unary languages. Define it similarly in general and show that it satisfies Sx.

Problem 24 *Noting that if L is the disjoint union of two languages L_1, L_2 and $\Theta(a_1, a_2, \ldots, a_m)$ a state description in L then*

- $\Theta(a_1, a_2, \ldots, a_m) = \Theta_1(a_1, a_2, \ldots, a_m) \wedge \Theta_2(a_1, a_2, \ldots, a_m)$ *for some state descriptions Θ_1, Θ_2 of L_1, L_2 respectively,*
- *for $i, j \in \{1, \ldots, m\}$ we have $a_i \sim_\Theta a_j \iff a_i \sim_{\Theta_1} a_j$ and $a_i \sim_{\Theta_2} a_j$,*

show that if w is a probability function on SL satisfying Sx and L_1 is a sublanguage of L, then w restricted to L_1, $w \restriction L_1$, also satisfies Sx.

Problem 25 *Let $m \in \mathbb{N}$, $m > 1$. Give an example of a language L and state descriptions for $\Theta(a_1, \ldots, a_m)$, $\Phi(a_1, \ldots, a_m, a_{m+1})$ satisfying*

$$\Phi(a_1, \ldots, a_m, a_{m+1}) \models \Theta(a_1, \ldots, a_m)$$

such that $\mathcal{S}(\Theta) = \{m\}$, $\mathcal{S}(\Phi) = \{1, 1, \ldots, 1\}$.

Problem 26 *Let L be the language with a single binary relation symbol R and let L_1 be the language with a single unary relation symbol P. For $\phi \in SL$ let ϕ^* be the result of replacing each occurrence of $R(t_1, t_2)$ in ϕ, where t_1, t_2 are terms of L, by $P(t_1)$. Define $w : SL \to [0, 1]$ by $w(\phi) = c_\infty^{L_1}(\phi^*)$.*

Show that w is a probability functionn SL satisfying Ex but not Vx.

Problem 27 *Show that $u^{\bar{p},L}$ satisfies WIP.*

Solutions

23 (i) c_∞^L gives *all* state descriptions for m constants the same value, so it satisfies Sx. No V_M can satisfy Sx because Sx clearly implies SN (recall Problem 4).

(ii) For L as usual with relations symbols R_i of arity r_i ($i = 1, 2, \dots, q$) and any distinct constants $b_1, \dots, b_m$ define

$$c_0^L \left(\bigwedge_{i=1}^{q} \bigwedge_{\langle c_1,\dots,c_{r_i}\rangle \in \{b_1,\dots,b_m\}^{r_i}} R_i^{\epsilon_i}(c_1, \dots, c_{r_i}) \right) = 2^{-q},$$

for any of the 2^q choices of the $\epsilon_i \in \{0,1\}$ (where R^1, R^0 stand for R and $\neg R$ respectively), with all other state descriptions getting value 0. This extends uniquely to a probability function on SL. The state descriptions for m constants which have non-zero probability (2^{-q}) are just those with spectrum $\{m\}$, so c_0^L satisfies Sx.

24 Let L be the disjoint union of L_1 and L_2. Let $\Theta(a_1, \dots, a_m), \Phi(a_1, \dots, a_m)$ be state descriptions in L_1 with the same spectrum. Assume for the present that $\sim_\Theta$ and $\sim_\Phi$ have exactly the same equivalence classes.

Any state description for $a_1, \dots, a_m$ in L which extends Θ can be expressed as

$$\Theta(a_1, \dots, a_m) \wedge \Psi(a_1, \dots, a_m) \tag{69}$$

where Ψ is a state description in L_2 and can be paired with the state description

$$\Phi(a_1, \dots, a_m) \wedge \Psi(a_1, \dots, a_m). \tag{70}$$

of L which extends Φ. Furthermore, since $\sim_\Theta$ and $\sim_\Phi$ have exactly the same equivalence classes, (69) and (70) will also have the same equivalence classes and spectrum. Hence, since w satisfies Sx,

$$w(\Theta(a_1, \dots, a_m) \wedge \Psi(a_1, \dots, a_m)) = w(\Phi(a_1, \dots, a_m) \wedge \Psi(a_1, \dots, a_m)).$$

Summing this over all choices of $\Psi(a_1, \dots, a_m)$ gives as required that $w(\Theta) = w(\Phi)$.

Finally, if Θ and Φ have the same spectrum but not the same equivalence classes, then we can find a permutation $\sigma \in \mathsf{S}_m$ such that $\Theta(a_1, a_2, \dots, a_m)$

and $\Phi(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)})$ do have the same equivalence classes and so, as above,

$$w(\Theta(a_1, a_2, \ldots, a_m)) = w(\Phi(a_{\sigma(1)}, a_{\sigma(2)}, \ldots, a_{\sigma(m)})) = w(\Phi(a_1, a_2, \ldots, a_m))$$

by Ex.

25 Let L contain m binary relation symbols $R_1, \ldots, R_m$ and let

$$\Theta(a_1, \ldots, a_m) = \bigwedge_{i=1}^{m} \bigwedge_{j=1}^{m} \bigwedge_{k=1}^{m} R_i(a_j, a_k),$$

with $\Phi(a_1, \ldots, a_m, a_{m+1})$ being

$$\Theta(a_1, \ldots, a_m) \wedge \bigwedge_{i=1}^{m} \left(\bigwedge_{j=1}^{m+1} R_i(a_j, a_{m+1}) \wedge \bigwedge_{k=1}^{i} \neg R_i(a_{m+1}, a_k) \wedge \bigwedge_{k=i+1}^{m} R_i(a_{m+1}, a_k) \right).$$

Then clearly

$$\Phi(a_1, \ldots, a_m, a_{m+1}) \models \Theta(a_1, \ldots, a_m),$$

$\mathcal{S}(\Theta) = \{m\}$ and also $\mathcal{S}(\Phi) = \{1, 1, \ldots, 1\}$ since a_{m+1} is not $\sim_\Phi$ -equivalent with any a_j for $j \leq m$, for example on the grounds of

$$\Phi \models R_1(a_j, a_1) \wedge \neg R_1(a_{m+1}, a_1)$$

and for $j, k \leq m$, $j < k$, a_j, a_k are not $\sim_\Phi$ -equivalent, for example on the grounds of

$$\Phi \models \neg R_j(a_{m+1}, a_j) \wedge R_j(a_{m+1}, a_k).$$

26 To show that w is a probability function satisfying Ex, check that P1-P3 and Ex are inherited from $c_\infty^{L_1}$. However, w does not satisfy Vx since, for example,

$$w(R(a_1, a_2) \wedge \neg R(a_1, a_3)) = c_\infty^{L_1}(P(a_1) \wedge \neg P(a_1)) = 0,$$

$$w(R(a_2, a_1) \wedge \neg R(a_3, a_1)) = c_\infty^{L_1}(P(a_2) \wedge \neg P(a_3)) = 1/4.$$

27 Let θ, ϕ and L be as in the statement of WIP. Suppose that L_1, L_2 are disjoint sublanguages of L, $L = L_1 \cup L_2$ and θ, ϕ each involve only relation symbols from L_1, L_2 respectively and have no constants in common. Arguing as in the proof of Proposition 7, see equation (19), it suffices to consider state descriptions, and appealing to Ex we can take the state descriptions of L_1, L_2 respectively to be $\Theta(a_1, \ldots, a_m)$ and $\Phi(a_{m+1}, \ldots, a_{m+n})$. Setting $\vec{a} = a_1, \ldots, a_m$ and $\vec{b} = a_{m+1}, \ldots, a_{m+n}$, by Li,

$$\begin{aligned}
u^{\bar{p},L}(\Theta) \cdot u^{\bar{p},L}(\Phi) &= u^{\bar{p},L_1}(\Theta) \cdot u^{\bar{p},L_2}(\Phi) && (71)\\
&= \sum_{\substack{\vec{c} \in \mathbb{N}^m \\ \Theta \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})}} |\mathcal{C}^{L_1}(\vec{c}, \vec{a})|^{-1} \prod_{s=1}^{m} p_{c_s} \times \sum_{\substack{\vec{d} \in \mathbb{N}^n \\ \Phi \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})}} |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|^{-1} \prod_{s=1}^{n} p_{d_s} \\
&= \sum_{\substack{\vec{c} \in \mathbb{N}^m \\ \Theta \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})}} \sum_{\substack{\vec{d} \in \mathbb{N}^n \\ \Phi \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})}} |\mathcal{C}^{L_1}(\vec{c}, \vec{a})|^{-1} |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|^{-1} \prod_{s=1}^{m+n} p_{e_s} && (72)
\end{aligned}$$

where $\vec{e} = \vec{c}\,\hat{}\,\vec{d} = c_1, \ldots, c_m, d_1, \ldots, d_n$, etc..

Now suppose that $\Theta(\vec{a}) \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})$ and $\Phi(\vec{b}) \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})$ and consider forming a state description $\Psi(a_1, \ldots, a_{m+n}) \in \mathcal{C}^{L}(\vec{e}, \vec{a}\,\hat{}\,\vec{b})$ extending $\Theta(a_1, \ldots, a_m) \wedge \Phi(a_{m+1}, \ldots, a_{m+n})$. The only constraints on Ψ are that it must be consistent with the colouring $\vec{e}$ and with the 'choices' already determined by Θ and Φ, for example if a_2, a_{m+1} get the same non-black colour, R is a binary relation symbol of L_1 and $\Theta \models R(a_1, a_2)$ then we must have that $\Psi \models R(a_1, a_{m+1})$. But clearly the free choices in forming Ψ do not depend on these particular $\Theta(\vec{a}) \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})$ and $\Phi(\vec{b}) \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})$ only on $\vec{e}$. In other words the number, N say, of $\Psi(a_1, \ldots, a_{m+n}) \in \mathcal{C}^{L}(\vec{e}, \vec{a}\,\hat{}\,\vec{b})$ extending $\Theta(a_1, \ldots, a_m) \wedge \Phi(a_{m+1}, \ldots, a_{m+n})$ will be the same no matter which $\Theta(\vec{a}) \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})$ and $\Phi(\vec{b}) \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})$ we started from. Hence also

$$|\mathcal{C}^{L}(\vec{e}, \vec{a}\,\hat{}\,\vec{b})| = N \cdot |\mathcal{C}^{L_1}(\vec{c}, \vec{a})| \cdot |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|. \tag{73}$$

Consequently, with the $\Psi(a_1, \ldots, a_{m+n})$ ranging over state descriptions of L extending $\Theta \wedge \Phi$, if $\Theta \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})$ and $\Phi \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})$,

$$\begin{aligned}
&\sum_{\Psi \in \mathcal{C}^{L}(\vec{e}, \vec{a}\,\hat{}\,\vec{b})} |\mathcal{C}^{L}(\vec{e}, \vec{a}\,\hat{}\,\vec{b})|^{-1} \\
&\qquad = N \cdot (N \cdot |\mathcal{C}^{L_1}(\vec{c}, \vec{a})| \cdot |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|)^{-1} \quad \text{by (73)} \\
&\qquad = |\mathcal{C}^{L_1}(\vec{c}, \vec{a})|^{-1} |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|^{-1}, && (74)
\end{aligned}$$

whilst if not both $\Theta \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})$ and $\Phi \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})$ then the right hand side of (74) is zero. Hence, with the $\Psi(a_1, \ldots, a_{m+n})$ ranging over state descriptions of L extending $\Theta \wedge \Phi$,

$$\begin{aligned} u^{\bar{p},L}(\Theta \wedge \Phi) &= \sum_{\Psi} \sum_{\substack{\vec{e} \in \mathbb{N}^{m+n} \\ \Psi \in \mathcal{C}^L(\vec{e}, \vec{a}\,\hat{}\,\vec{b})}} |\mathcal{C}^L(\vec{e}, \vec{a}\,\hat{}\,\vec{b})|^{-1} \prod_{s=1}^{m+n} p_{e_s} \\ &= \sum_{\substack{\vec{c} \in \mathbb{N}^m \\ \Theta \in \mathcal{C}^{L_1}(\vec{c}, \vec{a})}} \sum_{\substack{\vec{d} \in \mathbb{N}^n \\ \Phi \in \mathcal{C}^{L_2}(\vec{d}, \vec{b})}} |\mathcal{C}^{L_1}(\vec{c}, \vec{a})|^{-1} |\mathcal{C}^{L_2}(\vec{d}, \vec{b})|^{-1} \prod_{s=1}^{m+n} p_{e_s} \quad \text{by (74)} \\ &= u^{\bar{p},L}(\Theta) \cdot u^{\bar{p},L}(\Phi) \quad \text{by (72),} \end{aligned}$$

as required.

9 Principle of Induction

When L is unary, the proof of Theorem 31 works even if we only have ULi (rather than Li) with Sx, equivalently ULi with Ax:

Theorem 32 *Let w be a probability function on a unary language L satisfying ULi with Ax. Then there is a measure μ on $\mathbb{B}$ such that*

$$w = \int_{\mathbb{B}} u^{\bar{p},L} \, d\mu(\bar{p}) \tag{75}$$

and hence w satisfies Li with Sx.

It turns out that a single w on a *unary* language that satisfies ULi may be a member of different *unary* language invariant families. An example is the Carnap function c_2^L, where L contains just one unary predicate.

On the other hand, if w is a probability function on SL for L that contains a non-unary relation symbol and w satisfies Li with Sx then the language invariant family containing w *is* unique. Also,

Lemma 33 *Let $\{U_L\}$ and $\{V_L\}$ be two language invariant families satisfying Sx which agree on unary languages. Then they agree on all languages.*

See [1] for proofs and discussion of these claims.

Now recall that we have met two parameterised families of purely unary probability functions satisfying ULi with Ax, namely the c_λ^L of Carnap's continuum and the w_L^δ of NP-continuum. Each of these (with a fixed λ or δ respectively) must be part of a unique language invariant family with Sx extending over all (finite) languages, which leads to the question of what are the polyadic members of these families.

It is an easy exercise to answer this question for the w_L^δ (with $0 \leq \delta \leq 1$). Recall that it has been defined so that

$$w_L^\delta\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right) = 2^{-q} \sum_{j=1}^{2^q} (\delta + \gamma)^{m_j} \gamma^{m-m_j}$$

(where $2^q\gamma + \delta = 1$ and $m_j = |\{i\,|h_i = j\}|$). w_L^δ coincides with $u^{\bar{p},L}$ for

$$\bar{p} = \langle 1-\delta, \delta, 0, 0, 0, \ldots \rangle \in \mathbb{B}.$$

Hence the family is that of $u^{\bar{p},L}$, that is, with the coresponding measure on $\mathbb{B}$ giving all to this single point.

Carnap's c_λ^L agrees with w_L^1 when $\lambda = 0$ and with w_L^0 when $\lambda = \infty$. For $0 < \lambda < \infty$ the prior on $\mathbb{B}$ giving the family of c_λ^L has been shown by Kingman to be the Poisson-Dirichlet distribution for parameter λ. As this is not very easy to work with, the properties of the corresponding polyadic family in terms of which principles they might satisfy remain to be clarified. However, they both do satisfy a natural principle of relevance which arguably has just as good a claim to be a formalisation of inductive thinking as PIR, and which moreover generalises well to the polyadic context. We conclude this course with a discussion of it.

Principle of Induction

Let L be a general language which can also be purely unary. Considering how we should formalise, as a principle to impose on a probability function, the desideratum that more observed occurrences of something in the past should increase the agent's belief that it will happen in the future, we are first led to consider generalising the Principle of Instantial Relevance, PIR.

However, a straightforward generalisation of PIR which might recommend that when considering probabilities of state descriptions and their extensions, the probability of a new constant joining an equivalence class should increase after another constant has joined this equivalence class (as compared to the probability of a new constant just joining the class) does not make much sense because after adding one constant to a state description, the number of possibilities to add yet another one increases considerably in view of how the newest one might relate not only to the original ones but also to the previously added one.

In the unary context, with L containing q unary predicates, there are always 2^q possibilities to add a new constants to a state description, just as many as there are atoms. So on average the (conditional) probabilities of extensions are 2^{-q}. When there is a non-unary relation symbol in the language, the number of possibilities how to add a new constant increases with the size of the state description. Hence the conditional probabilities of extensions of larger state descriptions on average grow lower and lower and it makes little sense comparing them

However, requiring that, when adding a new constant to a state description, the conditional probability of those extensions where the new constant joins a larger equivalence class as opposed to the conditional probability of extensions where the new constant joins a smaller class arguably captures the same intuition. The corresponding principle is called the Principle of Induction and we shall first consider it in the unary context.

The Unary Principle of Induction, UPI

Let L contain only unary predicates. If $\Theta(b_1, \dots, b_m) = \bigwedge_{i=1}^m \alpha_{h_i}(b_i)$ and g and j are such that

$$|\{i : h_i = g\}| \geq |\{i : h_i = j\}|$$

(there are more occurrences of α_g amongst the α_{h_i} than there are of α_j) then

$$w(\alpha_g(b_{m+1}) \,|\, \Theta(\vec{b})) \geq w(\alpha_j(b_{m+1}) \,|\, \Theta(\vec{b})).$$

Somewhat remarkably, the principle holds for any probability function satisfying Sx (equivalently - since the language is unary - Ax+Ex). It follows from a more general property spelled out in the following theorem.

Theorem 34 *Let L be unary and let $\Theta(b_1, \dots, b_m)$ and $\Psi(b_1, \dots, b_m)$ be state descriptions with spectra $\{n_1, \dots, n_r\}$ and $\{m_1, \dots, m_t\}$ respectively.*[1] *If r or t is less that 2^q, define n_j and m_j for $r < j \leq 2^q$ or $t < j \leq 2^q$ respectively to be 0. A necessary and sufficient condition for*

$$w(\Theta(\vec{b})) \leq w(\Psi(\vec{b})) \tag{76}$$

to hold for all probability functions w on SL satisfying Ex+Ax is that

$$\sum_{j \leq i} n_j \,\leq\, \sum_{j \leq i} m_j \quad \textit{for } i = 1, 2, \dots, 2^q. \tag{77}$$

[1] Recall the convention of listing spectra in order, so $n_1 \geq n_2 \geq \ldots \geq n_r$ and $m_1 \geq m_2 \geq \ldots \geq m_t$.

Proof Let $\tilde{n}$, $\tilde{m}$ pertain to Θ, Ψ as in the theorem. Note that since we are only concerned with probability functions satisfying Ex and Ax, we may, employing the notation from page 56, assume that Θ, Ψ are $\alpha_1^{n_1}\alpha_2^{n_2}\dots\alpha_{2^q}^{n_{2^q}}$ and $\alpha_1^{m_1}\alpha_2^{m_2}\dots\alpha_{2^q}^{m_{2^q}}$ respectively.

Assume that (77) fails for $\tilde{m}, \tilde{n}$. Let i be the least such that it fails so

$$\sum_{j\leq i} n_j > \sum_{j\leq i} m_j,$$

that is,

$$\sum_{j>i} n_j < \sum_{j>i} m_j.$$

It can be checked that (76) fails for $v_{\vec{c}}$,[1] when $\vec{c} \in \mathbb{D}_{2^q}$ is defined as follows:

$$c_j = \begin{cases} (1-(2^q-i)\epsilon)/i & \text{for } j \leq i \\ \epsilon & \text{for } j > i \end{cases}$$

where $\epsilon > 0$ is sufficiently small. This is because $v_{\vec{c}}(\Theta)$, $v_{\vec{c}}(\Psi)$ are polynomials in ϵ with lowest power in that for Θ being $\sum_{j>i} n_j$ and in that of Ψ being $\sum_{j>i} m_j$. We omit the details.

Now suppose that (77) holds. By Muirhead's Inequality, see [6, p44], for such $\vec{n}, \vec{m}$ and real numbers $x_1, x_2, \dots, x_{2^q} \geq 0$,

$$\sum_{\sigma\in \mathsf{S}_{2^q}} x_{\sigma(1)}^{m_1}x_{\sigma(2)}^{m_2}\dots x_{\sigma(2^q)}^{m_{2^q}} \geq \sum_{\sigma\in \mathsf{S}_{2^q}} x_{\sigma(1)}^{n_1}x_{\sigma(2)}^{n_2}\dots x_{\sigma(2^q)}^{n_{2^q}}, \tag{78}$$

If $\vec{x} \in \mathbb{D}_{2^q}$ then (78) expresses that $|\mathsf{S}_{2^q}|v_{\vec{x}}(\Psi) \geq |\mathsf{S}_{2^q}|v_{\vec{x}}(\Theta)$ so the result follows by Theorem 10.

□

Corollary 35 *If L is unary and w is a probability function on SL satisfying Ex and Ax then w satisfies UPI.*

Proof Let $\Theta(\vec{b}), g$ and j be as in UPI. Let $\Theta(\vec{b})$ have signature $\langle k_1, \dots, k_{2^q}\rangle$. Since w satisfies Ax, we can assume without loss of generality that $k_1 \geq k_2 \geq \dots \geq k_{2^q}$, so the specrum, with the zeros left in, is $\{k_1, \dots, k_{2^q}\}$.

[1]See page 45 for the definition of these functions.

The condition from UPI amounts to $k_g \geq k_j$. Comparing conditional probabilities $w(\alpha_g(b_{m+1}) \,|\, \Theta(\vec{b}))$ and $w(\alpha_j(b_{m+1}) \,|\, \Theta(\vec{b}))$ is the same as comparing $w(\alpha_g(b_{m+1}) \wedge \Theta(\vec{b}))$ and $w(\alpha_j(b_{m+1}) \wedge \Theta(\vec{b}))$. The spectra $\tilde{m}$ and $\tilde{n}$ of $\alpha_g(b_{m+1}) \wedge \Theta(\vec{b})$ and $\alpha_j(b_{m+1}) \wedge \Theta(\vec{b})$ respectively are as $\{k_1, \dots, k_{2^q}\}$ except that in the former, 1 is added to k_g and in the latter to k_j. Since $g \geq j$, (77) is satisfied and the result follows. $\square$

For a general language that contains some non-unary relation symbol we need to be more careful when formulating a principle of induction because the intuition behind it works only when the new constant joins an existing class. Otherwise, a new constant could either be different from all the previous constants but respect the original indiscernibilities, or it could differentiate between previously indistinguishable constants and thus refine the original equivalence classes. Note that in the unary context, a new constant cannot do this. Clearly, the probability of a constant not joining an existing class should further depend on how much havoc it makes of equivalences defined by the original state description, and we should not require these probabilities to be the same. Hence the general statement of the Principle of Induction is as follows:

The Principle of Induction, PI

Let $\Theta(b_1, \dots, b_m)$ *be a state description and let*

$$\Theta_1(b_1, \dots, b_m, b_{m+1}), \quad \Theta_2(b_1, \dots, b_m, b_{m+1})$$

be extensions of Θ. *Then*

$$w(\Theta_1 \,|\, \Theta) \;\geq\; w(\Theta_2 \,|\, \Theta) \tag{79}$$

whenever

$$0 \neq |\{i \,|\, 1 \leq i \leq m \text{ and } b_{m+1} \sim_{\Theta_1} b_i\}| \;\geq\; |\{i \,|\, 1 \leq i \leq m \text{ and } b_{m+1} \sim_{\Theta_2} b_i\}| \,.$$

The Principle of Induction holds for any probability function satisfying Li with Sx. It can be proved similarly as in the unary case, only the mathematics is more complicated, utilizing the functions $u^{\bar{p},L}$ in place of the $v_{\vec{x}}$, the representation theorem for Li+Sx in place of Theorem 10 and a generalized Muirhead Theorem.

References

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

[2] Ash, R.B., *Probability and Measure Theory,* 2nd edition, Academic Press, 1999.

[3] Billingsley, P.P., *Probability and Measure,* John Wiley & Sons, New York, 1995.

[4] Billingsley, P.P., *Convergence of Probability Measures,* Second Edition, John Wiley & Sons, New York, 1999.

[5] Howarth, E., Paris, J.B. & Vencovská, A., An examination of the SEP Candidate Analogical Inference Rule within Pure Inductive Logic, *Journal of Applied Logic*, 2016, 14:22-45.

[6] Hardy, G.H., Littlewood, J.E. & Pólya, G., *Inequalities*, Cambridge University Press, 1967.