June 3, 2012

Pure Inductive Logic

FEWS Tutorial,

Munich, 2012

Jeff Paris,
School of Mathematics,
University of Manchester,
Manchester M13 9PL, UK.

jeff.paris@manchester.ac.uk

[These slides contain some material not covered in the two tutorials]

Context and Notation

We work with a first order language L with relation symbols $R_1, R_2, \ldots, R_q$, say of arities $r_1, r_2, \ldots, r_q$ respectively, and constants a_n for $n \in \mathbb{N}^+ = \{1, 2, 3, \ldots\}$, and no function symbols nor (in general) equality.

The intention here is that the a_i name all the individuals in some population though there is no prior assumption that they necessarily name different individuals.

Let SL denote the set of first order sentences of this language L and $QFSL$ the quantifier free sentences of this language.

Let $\mathcal{T}$ denote the set of structures for L with universe $\{a_1, a_2, a_3, \ldots\}$, with the obvious interpretation of a_i as a_i itself.

The Fundamental Problem for Pure Inductive Logic

Imagine an agent who inhabits some structure $M \in \mathcal{T}$ but knows nothing about what is true in M. Then the problem is,

> $\mathcal{Q}$: *In this situation of zero knowledge, logically, or rationally, what belief should our agent give to a sentence $\theta \in SL$ being true in M?*

Probability Functions

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

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

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

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

All the standard, simple, properties you'd expect of a probability function follow from these (P1-3):

Proposition 1 *Let w be a probability function on SL. Then for $\theta, \phi \in SL$,*

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

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

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

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

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

Theorem 2 *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 on SL satisfying (P1),(P2),(P3) for any $\theta, \phi, \exists x\, \psi(x) \in SL$.*

State Descriptions

As usual let L be our default language with relation symbols $R_1, R_2, \ldots, R_q$ of arities $r_1, r_2, \ldots, r_q$ respectively.

For distinct constants $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ coming from $a_1, a_2, a_3, \ldots$, a *State Description* for $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ is a sentence of L of the form

$$\bigwedge_{k=1}^{q} \bigwedge_{c_1, c_2, \ldots, c_{r_k}} R_k^{\epsilon_{k,\vec{c}}}(c_1, c_2, \ldots, c_{r_k})$$

where the $c_1, c_2, \ldots, c_{r_k}$ range over all (not necessarily distinct) choices from $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$, the $\epsilon_{k,\vec{c}} \in \{0,1\}$ and R^ϵ stands for R if $\epsilon = 1$ and $\neg R$ if $\epsilon = 0$.

In other words, a state description for $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ tells us exactly which of the $R_k(c_1, c_2, \ldots, c_{r_k})$ hold and which do not hold for R_k a relation symbol from our language and any arguments $c_1, c_2, \ldots, c_{r_k}$ from $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$.

Example

Suppose L has just the binary relation symbol R and the unary relation (or predicate) symbol P. Then

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

is a state description for a_1, a_2.

We shall use upper case Θ, Φ, Ψ etc. for state descriptions.

By the DNFT any $\varphi(\vec{a}) \in QFSL$ is logically equivalent to a disjunction of state descriptions

$$\varphi(\vec{a}) \equiv \bigvee_{\Theta(\vec{a}) \models \varphi(\vec{a})} \Theta(\vec{a})$$

so using Theorem 2,

$$w(\varphi(\vec{a})) = w\left(\bigvee_{\Theta(\vec{a}) \models \varphi(\vec{a})} \Theta(\vec{a}) \right) = \sum_{\Theta(\vec{a}) \models \varphi(\vec{a})} w(\Theta(\vec{a}))$$

Hence

Proposition 3 *A probability function is determined by its values on the state descriptions.*

Question $\mathcal{Q}$, it now amounts to:

> $\mathcal{Q}$: *In this situation of zero knowledge, logically, or rationally, what probability function* $w : SL \to [0,1]$ *should our agent adopt when* $w(\theta)$ *is to represent the agent's probability that a sentence* $\theta \in SL$ *is true in the ambient structure* M?

Rational Principles

'Rational Principles' have, to date, largely arisen from considerations of *Symmetry, Relevance* and *Irrelevance.*

The Constant Exchangeability Principle, Ex

For $\phi(a_1, a_2, \dots, a_m) \in SL$ *and (distinct) constants* $a_{i_1}, a_{i_2}, \dots, a_{i_m}$,

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

Henceforth we shall assume that Ex holds for all the probability functions we consider.

Similarly the Principles of *Predicate Exchangeability* (where we require w to give the same probability when we transpose relation symbols of the same arity) and similarly *Strong Negation* where we replace a relation symbol throughout by its negation.

Unary Inductive Logic

In the initial investigations of Johnson and Carnap the language of Inductive Logic was taken to be *unary*.

I.e. the relation symbols $R_1, R_2, \ldots, R_q$ of the L all had arity 1.

Assume for the present that L is unary.

Now a state description for $a_{i_1}, a_{i_2}, \ldots, a_{i_m}$ is of the form

$$\bigwedge_{j=1}^{m} \bigwedge_{k=1}^{q} R_k^{\epsilon_{k,j}}(a_{i_j}),$$

equivalently of the form

$$\bigwedge_{j=1}^{m} \alpha_{h_j}(a_{i_j}), \tag{1}$$

where the $\alpha_{h_j}(x)$ $(1 \leq h_j \leq 2^q)$ are *atoms* of L, that is come from amongst the 2^q formulae of the form

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

where the $\epsilon_1, \epsilon_2, \ldots, \epsilon_q \in \{0, 1\}$.

[There are 2^q of them because there are two choices of ϵ_i for $i = 1, 2, \ldots, q$]

Example If $q = 3$ the conjunction of

$$\begin{array}{ccccccc} R_1(a_1) & R_1(a_2) & \neg R_1(a_3) & R_1(a_4) & \neg R_1(a_5) & \neg R_1(a_6) & \neg R_1(a_7) \\ R_2(a_1) & R_2(a_2) & \neg R_2(a_3) & R_2(a_4) & \neg R_2(a_5) & \neg R_2(a_6) & R_2(a_7) \\ \neg R_3(a_1) & \neg R_3(a_2) & \neg R_3(a_3) & \neg R_3(a_4) & R_3(a_5) & R_3(a_6) & \neg R_3(a_7) \end{array}$$

is a state description for $a_1, a_2, \ldots, a_7$, it tells us everything there is to know about $a_1, a_2, \ldots, a_7$.

The first column, equivalently the atom $R_1(x) \wedge R_2(x) \wedge \neg R_3(x)$ which a_1 satisfies, already tells us everything there is to know about a_1 etc..

The $w_{\vec{x}}$

In this section we construct a family of probability functions. Let $0 \leq x_1, x_2, \ldots, x_{2^q} \leq 1$ with $\sum_{i=1}^{2^q} x_i = 1$.
Define $w_{\vec{x}}$, where $\vec{x} = \langle x_1, x_2, \ldots, x_{2^q} \rangle$, on the state description $\bigwedge_{j=1}^m \alpha_{h_j}(a_{i_j})$ by

$$w_{\vec{x}} \left(\bigwedge_{j=1}^m \alpha_{h_j}(a_{i_j}) \right) = \prod_{j=1}^m x_{h_j},$$

equivalently

$$w_{\vec{x}}(\alpha_{h_1}(a_{i_1}) \wedge \alpha_{h_2}(a_{i_2}) \wedge \ldots \wedge \alpha_{h_m}(a_{i_m})) = x_{h_1} x_{h_2} \ldots x_{h_m}. \qquad (2)$$

So for example if $q = 2$, $\alpha_1(x) = R_1(x) \wedge R_2(x)$, $\alpha_3(x) = \neg R_1(x) \wedge R_2(x)$,
$w_{\vec{x}}(R_1(a_1) \wedge R_2(a_1) \wedge R_1(a_4) \wedge R_2(a_4) \wedge \neg R_1(a_5) \wedge R_2(a_5)) = x_1 x_1 x_3 = x_1^2 x_3$.

$w_{\vec{x}}$, defined as here on state descriptions, extends to a probability function on SL.

Notice that $w_{\vec{x}}$ satisfy Ex on state descriptions, and all sentences in fact.

de Finetti's Representation Theorem

de Finetti's Representation Theorem 4 *A probability function w on a unary language L satisfies Ex just if it is a convex mixture of the $w_{\vec{x}}$.*

More precisely, just if

$$w = \int w_{\vec{x}} \, d\mu(\vec{x}) \qquad (3)$$

where μ is a countably additive measure on the Borel subsets of

$$\{\langle x_1, x_2, \ldots, x_{2^q} \rangle \,|\, 0 \leq x_1, x_2, \ldots, x_{2^q}, \sum_i x_i = 1\},$$

which we shall refer to as the de Finetti prior of w.

Sketch Proof of de Finetti's Theorem – Optional!

Let w satisfy Ex and $\Theta(a_1, \dots, a_m)$ be a state description.

For $m \leq n$,

$$\Theta(a_1, \dots, a_m) = \bigvee_{\Phi(a_1,\dots,a_n) \models \Theta} \Phi(a_1, \dots, a_n).$$

So,

$$w(\Theta) = \sum_{\Phi(a_1,\dots,a_n) \models \Theta} w(\Phi).$$

Let $\bar{\Psi}$ be the set of state descriptions which are the same as the state description $\Psi(a_1, \dots, a_n)$ up to a permutation of the constants $a_1, a_2, \dots, a_n$. By Ex they all get the same probability (via w) and

$$\begin{aligned} w(\Theta) &= \sum_{\bar{\Psi}} \sum_{\substack{\Phi \in \bar{\Psi} \\ \Phi \models \Theta}} w(\Phi) \\ &= \sum_{\bar{\Psi}} v_{\bar{\Psi}}(\Theta) w(\bigvee \bar{\Psi}) \end{aligned}$$

where $v_{\bar{\Psi}}(\Theta)$ is the probability of picking $\Phi \in \bar{\Psi}$ such that $\Phi \models \Theta$.

That is, it is the probability that a random $\Phi(a_1, a_2, \dots, a_n) \in \bar{\Psi}$ will be such that

the atom satisfied by a_1 in Φ will be the same as the atom α_{h_1} satisfied by a_1 in Θ,
the atom satisfied by a_2 in Φ will be the same as the atom α_{h_2} satisfied by a_2 in Θ,
. . . . . .
the atom satisfied by a_m in Φ will be the same as the atom α_{h_m} satisfied by a_m in Θ.

But for n large this is very nearly

$$w_{\vec{x}_{\bar{\Psi}}}(\Theta) = w_{\vec{x}_{\bar{\Psi}}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(a_i)\right)$$

where, if $\vec{x}_{\bar{\Psi}} = \langle x_1, x_2, \dots, x_{2^q} \rangle$, then

x_j is proportion of times that the atom α_i occurs in some/any $\Phi(a_1, \dots, a_n)$ in $\bar{\Psi}$

So

$$\begin{aligned} w(\Theta) &= \sum_{\bar{\Psi}} \sum_{\substack{\Phi \in \bar{\Psi} \\ \Phi \models \Theta}} w(\Phi) \\ &= \sum_{\bar{\Psi}} v_{\bar{\Psi}}(\Theta) w(\bigvee \bar{\Psi}) \\ &\approx \sum_{\bar{\Psi}} w_{\vec{x}_{\bar{\Psi}}}(\Theta) w(\bigvee \bar{\Psi}) \\ &= \sum_{\vec{x}_{\bar{\Psi}}} w_{\vec{x}_{\bar{\Psi}}}(\Theta) w(\bigvee \bar{\Psi}) \end{aligned}$$

since we can reclaim the $\bar{\Psi}$ from $\vec{x}_{\bar{\Psi}}$, and conversely.

So what?

Well, letting n be nonstandard and taking the standard parts of each side, the $\sum_{\vec{x}_{\bar{\Psi}}}$ becomes $\int_{\vec{x}}$, the $w_{\vec{x}_{\bar{\Psi}}}(\Theta)$ becomes $w_{\vec{x}}(\Theta)$, the $w(\bigvee \bar{\Psi})$ becomes the $d\mu(\vec{x})$, and we get

$$w(\Theta) = \int_{\vec{x}} w_{\vec{x}}(\Theta)\, d\mu(\vec{x}).$$

Simple example of how the theorem can be used:

We immediately have that

$$w_{\vec{x}}(\alpha_1(a_1) \wedge \alpha_1(a_2)) + w_{\vec{x}}(\alpha_2(a_1) \wedge \alpha_2(a_2)) = x_1^2 + x_2^2 \geq 2x_1x_2 = 2w_{\vec{x}}(\alpha_1(a_1) \wedge \alpha_2(a_2))$$

Integrating both sides as above gives that for w satisfying Ex,

$$w(\alpha_1(a_1) \wedge \alpha_1(a_2)) + w(\alpha_2(a_1) \wedge \alpha_2(a_2)) \geq 2w(\alpha_1(a_1) \wedge \alpha_2(a_2)).$$

Hence we must have at least one of

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

equivalently (using also Ex to permute constants), we must have one of

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

A more significant application (Gaifman, later simplified by Humburg):

Theorem 5 *Ex implies the:*

Principle of Instantial Relevance, PIR
For $\theta(a_1, a_2, \ldots, a_m) \in SL$,

$$w(\alpha_i(a_{m+2}) \,|\, \alpha_i(a_{m+1}) \wedge \theta(a_1, a_2, \ldots, a_m)) \geq w(\alpha_i(a_{m+2}) \,|\, \theta(a_1, a_2, \ldots, a_m)).$$

Prior Equivalents

We can often usefully characterize properties of probability functions w in terms of properties of their de Finetti priors μ.

Say that w is *super regular* if $w(\theta) > 0$ whenever $\theta \in SL$ is consistent. [The 'super' is added here because in this area just 'regular' is usually taken to mean the weaker condition that $w(\theta) > 0$ for $\theta \in QFSL$.]

Theorem 6 *Let μ be the de Finetti prior of w. Then w is super regular just if for all $\emptyset \subseteq S \subset \{1, 2, \ldots, 2^q\}$,*

$$\mu\{\langle x_1, x_2, \ldots, x_{2^q} \rangle \,|\, x_i = 0 \iff i \in S\} > 0.$$

Inheritance

Idea! w can inherit 'rationality' from it's de Finetti prior μ.

For example it might appear 'rational' in this situation of zero knowledge that μ should be as unassuming, fair, frankly boring, as possible

– in fact we should take μ to be simply the standard Lebesgue measure.

If we do that then w comes out to be the probability function $c^L_{2^q}$ from Carnap's Continuum of Inductive Methods, which for this language L with q unary predicates is characterized by:

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

where m_j is the number of times that the atom α_j occurs amongst

$$\alpha_{h_1}, \alpha_{h_2}, \ldots, \alpha_{h_n}.$$

We shall have more to say about $c_{2^q}^L$ in the next tutorial.

Going back

Say $\vec{x}$ is in the *support* of μ if for all $\epsilon > 0$

$$\mu\{\vec{y} \ : \ |\vec{y} - \vec{x}| < \epsilon\} > 0.$$

Say $\vec{x}$ is an *extreme point* if some $x_i = 1$ (so the rest must be 0).

Pick 'n' Mix Theorem 7 *Let $\vec{b}, \vec{c}$ be non-extreme support points of μ, $0 \leq \lambda \leq 1$ and $\theta(a_1, \ldots, a_m) \in QFSL$.*
Then there are $\phi_n(a_{m+1}, \ldots, a_{m+n}) \in QFSL$ such that

$$\lim_{n \to \infty} w(\theta \,|\, \phi_n(a_{m+1}, \ldots, a_{m+n})) = \lambda w_{\vec{b}}(\theta) + (1 - \lambda) w_{\vec{c}}(\theta).$$

Example

Given atoms α_i, α_j the distance between them,

$$|\alpha_i - \alpha_j|$$

is the number of the predicates $R_1, R_2, \ldots, R_q$ that they give the opposite sign to.

E.g. when $q = 3$

$$|(R_1(x) \wedge R_2(x) \wedge \neg R_3(x)) \ - \ (R_1(x) \wedge \neg R_2(x) \wedge \neg R_3(x))| = 1$$

Say that w satisfies the *Analogy Principle* if for any $\theta(a_1, \ldots, a_n) \in QFSL$, if $|\alpha_i - \alpha_j| < |\alpha_i - \alpha_k|$ then

$$w(\alpha_i(a_{n+2}) \,|\, \alpha_j(a_{n+1}) \wedge \theta(a_1, \ldots, a_n)) > w(\alpha_i(a_{n+2}) \,|\, \alpha_k(a_{n+1}) \wedge \theta(a_1, \ldots, a_n)).$$

Using Pick 'n' Mix we can show that no probability function can satisfy the Analogy Principle for $q > 2$. [Joint work with Alex Hill.]

Polyadic Inductive Logic

[Results stated in this tutorial are joint with various subsets of Alena Vencovská, Jürgen Landes and Chris Nix.]

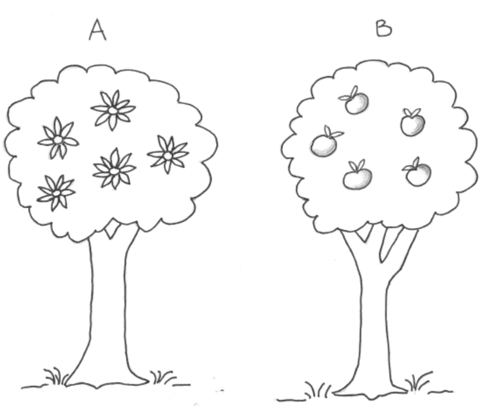
Until the turn of the millennium 'Inductive Logic', with very few exceptions, meant 'Unary Inductive Logic'.

WHY?

Reason 1: Mathematically Polyadic Inductive Logic is much more complicated (and why leave the unary when this seam was still so productive?).

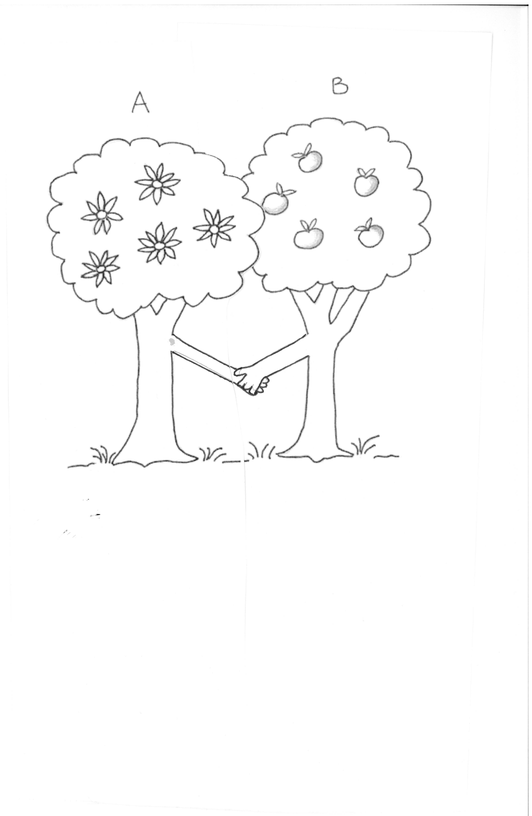
Reason 2: We are not used to doing induction on binary, ternary, . . relations.

Suppose that Adam knows that apples of strain $\mathcal{A}$ are good pollinators and apples of strain $\mathcal{B}$ are easily pollinated.



Then he might conclude that were he to plant them next to each other
. . .

the result would be fruitful



Reason 3: Lack of intuition when it comes to forming beliefs about polyadic relations.

To take an example suppose you are told that

$$R(a_1, a_2) \wedge R(a_2, a_1) \wedge \neg R(a_1, a_3).$$

In this case which of $R(a_3, a_1), \neg R(a_3, a_1)$ would you think the more likely?

This lack of intuition extends too to the problem of proposing rational principles in the polyadic case.

To motivate one such principle we return briefly to Unary Inductive Logic.

Unary Indistinguishability

To motive this recall a state description $\Theta(a_1, a_2, \dots, a_7)$ which we looked at in the unary case:

$$\begin{array}{ccccccc} R_1(a_1) & R_1(a_2) & \neg R_1(a_3) & R_1(a_4) & \neg R_1(a_5) & \neg R_1(a_6) & \neg R_1(a_7) \\ R_2(a_1) & R_2(a_2) & \neg R_2(a_3) & R_2(a_4) & \neg R_2(a_5) & \neg R_2(a_6) & R_2(a_7) \\ \neg R_3(a_1) & \neg R_3(a_2) & \neg R_3(a_3) & \neg R_3(a_4) & R_3(a_5) & R_3(a_6) & \neg R_3(a_7) \end{array}$$

Say a_i, a_j are *indistinguishable w.r.t. this stated description* if it is consistent with Θ that a_i, a_j are actually equal, i.e.

$$\Theta(a_1, a_2, \ldots, a_7) \wedge a_i = a_j$$

is consistent in the Predicate Calculus with equality.

Equivalently a_i, a_j satisfy the same atom.

So here a_1, a_2, a_4 are indistinguishable, as are a_5, a_6, but a_3, a_7 are both distinguishable from all the other a_i.

Indistinguishability with respect to a state description is an equivalence relation.

In this specific case then the equivalence classes are

$$\{a_1, a_2, a_4\}, \quad \{a_5, a_6\}, \quad \{a_3\}, \quad \{a_7\}$$

If we think of the atoms as colours, so $\alpha_i(a_j)$ says that a_j has colour i, then the equivalence classes are just the sets of a_i with the same colour.

Define the *Spectrum of a State Description* to be the multiset of sizes of these equivalence classes.

So for $\Theta(a_1, a_2, \ldots, a_7)$ as above its spectrum is $\{3, 2, 1, 1\}$.

In the unary case both Johnson and Carnap accepted the following principle:

Atom Exchangeability Principle, Ax

For a state description Θ the probability $w(\Theta)$ should only depend on the spectrum of Θ.

This is another symmetry principle because. . .

. . . there is a permutation of atoms/colours and constants which sends $\Theta(\vec{a})$ to $\Phi(\vec{a})$ just if they have the same spectrum.

For example for the state descriptions below the first two have the same spectrum and there is clearly a permutation of colours which sends the first to the second. However the third has a different spectrum ($\{3, 3, 1\}$) and clearly there is no such permutation sending the first to the third.

$$\alpha_1(a_1) \wedge \alpha_1(a_2) \wedge \alpha_8(a_3) \wedge \alpha_1(a_4) \wedge \alpha_7(a_5) \wedge \alpha_7(a_6) \wedge \alpha_6(a_7)$$

$$\alpha_8(a_1) \wedge \alpha_8(a_2) \wedge \alpha_7(a_3) \wedge \alpha_8(a_4) \wedge \alpha_4(a_5) \wedge \alpha_4(a_6) \wedge \alpha_6(a_7)$$

$$\alpha_1(a_1) \wedge \alpha_1(a_2) \wedge \alpha_8(a_3) \wedge \alpha_1(a_4) \wedge \alpha_7(a_5) \wedge \alpha_7(a_6) \wedge \alpha_7(a_7)$$

Polyadic Indistinguishability

Indistinguishabilty extends smoothly to polyadic languages.

Taking L henceforth to be polyadic we define the indistinguishability of a_{i_k}, a_{i_r} w.r.t a state description $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_m})$ as before,

i.e. just if

$$\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_m}) \wedge a_{i_k} = a_{i_r}$$

is consistent (in the Predicate Calculus with Equality).

Example Taking the language L to have just a single binary relation symbol R, with respect to the state description (which is the conjunction of)

$$\begin{array}{cccc} R(a_1,a_1) & \neg R(a_1,a_2) & R(a_1,a_3) & R(a_1,a_4) \\ R(a_2,a_1) & \neg R(a_2,a_2) & R(a_2,a_3) & \neg R(a_2,a_4) \\ R(a_3,a_1) & \neg R(a_3,a_2) & R(a_3,a_3) & R(a_3,a_4) \\ R(a_4,a_1) & R(a_4,a_2) & R(a_4,a_3) & R(a_4,a_4) \end{array}$$

the 'indistinguishability' equivalence classes are

$$\{a_1, a_3\}, \quad \{a_2\}, \quad \{a_4\}$$

and the spectrum is $\{2, 1, 1\}$

In the (possibly) polyadic setting Atom Exchangeability, Ax, now generalizes to

Spectrum Exchangeability Principle, Sx

For a state description Θ the probability $w(\Theta)$ only depends on the spectrum of Θ. If state descriptions Θ, Φ have the same spectrum then $w(\Theta) = w(\Phi)$.

Example When L has just a single binary relation symbol the state descriptions (which are the the conjunctions of)

$$\begin{array}{ccc} R(a_1, a_1) & \neg R(a_1, a_2) & R(a_1, a_3) \\ R(a_2, a_1) & \neg R(a_2, a_2) & R(a_2, a_3) \\ R(a_3, a_1) & \neg R(a_3, a_2) & R(a_3, a_3) \end{array}$$

and of

$$\begin{array}{ccc} \neg R(a_1, a_1) & \neg R(a_1, a_2) & R(a_1, a_3) \\ \neg R(a_2, a_1) & \neg R(a_2, a_2) & R(a_2, a_3) \\ R(a_3, a_1) & R(a_3, a_2) & R(a_3, a_3) \end{array}$$

get the same probability under Sx since both have spectrum $\{2, 1\}$.

Unlike the unary case indistinguishability need not be preserved when we extend state descriptions.

For example the state description

$$\begin{array}{cccc} \neg R(a_1, a_1) & \neg R(a_1, a_2) & R(a_1, a_3) & R(a_1, a_4) \\ \neg R(a_2, a_1) & \neg R(a_2, a_2) & R(a_2, a_3) & \neg R(a_2, a_4) \\ R(a_3, a_1) & R(a_3, a_2) & R(a_3, a_3) & R(a_3, a_4) \\ R(a_4, a_1) & R(a_4, a_2) & R(a_4, a_3) & R(a_4, a_4) \end{array}$$

extends the one immediately above it but has spectrum $\{1, 1, 1, 1\}$, a_4 has distinguished a_1 and a_2.

Before considering Sx further we introduce another rationality requirement which gives considerable extra strength to Sx.

Language Invariance

Language Invariance with Sx, Li+Sx

A probability function w satisfies Language Invariance with Sx if there is a family of probability functions $w^{\mathcal{L}}$, one on each language $\mathcal{L}$, containing w (so $w = w^L$) such that each member of this family satisfies Sx and whenever languages $\mathcal{L}_1, \mathcal{L}_2$ are such that $\mathcal{L}_1 \subseteq \mathcal{L}_2$ then $w^{\mathcal{L}_2} \restriction S\mathcal{L}_1 = w^{\mathcal{L}_1}$.

Aside

Recall the earlier argument for the unary probability function $c^L_{2^q}$ on a unary language L with q predicates.

This was justified in terms of the de Finetti prior μ being standard Lebesgue measure.

If we apply exactly the same reasoning to the language L^+ formed by adding an extra unary predicate to L we obtain $c^{L^+}_{2^{q+1}}$.

But if we restrict this probability function to SL we do not get back our 'favored choice' $c^L_{2^q}$ for that language!

These 'favoured choices' then do not form part of a language invariant family,

if you want language invariance then you can only use the 'Lebesgue measure argument' for one specific language.

- - - - - - - -

It turns out that Li+Sx implies most (maybe even all) of the desirable properties so far proposed for a rational polyadic probability function.

Paradise Gained

From now on assume that w satisfies Li+Sx.

Again we have a de Finetti style representation theorem for such w showing them to be convex mixtures of certain (relatively) simple building block functions $u^{\bar{p},L}$ satisfying Li+Sx: [This next definition of the $u^{\bar{p},L}$ was not given in the tutorial and so is an 'optional extra'.]

The $u^{\bar{p},L}$

Let $\bar{p}$ be a sequence

$$p_0, p_1, p_2, p_3, \ldots$$

of real numbers such that

$$p_1 \geq p_2 \geq p_3 \geq \ldots \geq 0 \quad \text{and} \quad \sum_{i=0}^{\infty} p_i = 1.$$

We think of the subscripts here $0, 1, 2, 3, \ldots$ as *colours*, with 0 being black, and p_i as the probability of picking colour i (with replacement).

Given a state description $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n})$ and a sequence of colours (not necessarily distinct) $c_1, c_2, \ldots, c_n$ (so these are really just natural numbers). We say that $\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n})$ is *consistent* with this sequence if whenever $c_j = c_k \neq 0$ then a_{i_j}, a_{i_k} are indistinguishable with respect to Θ.

Example Suppose that the language has a single binary relation symbol R and $\Theta(a_1, a_2, a_3)$ is the conjunction of

$$\begin{array}{ccc} \neg R(a_1, a_1) & \neg R(a_1, a_2) & R(a_1, a_3) \\ \neg R(a_2, a_1) & \neg R(a_2, a_2) & R(a_2, a_3) \\ R(a_3, a_1) & R(a_3, a_2) & R(a_3, a_3), \end{array}$$

so the indistinguishability equivalence classes are $\{a_1, a_2\}$ and $\{a_3\}$.

Then $\Theta(a_1, a_2, a_3)$ is consistent with the sequence of colours $2, 2, 1$, and with $0, 2, 1$, and with $0, 0, 0$, and with $1, 2, 0$, but not with $1, 2, 1$, nor with $0, 1, 1$.

Define $u^{\bar{p},L}(\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n}))$ as follows:

- Pick a sequence of colours $c_1, c_2, \ldots, c_n$ according to the probabilities $p_0, p_1, p_2, \ldots$, so the probability of picking $c_1, c_2, \ldots, c_n$ is

$$p_{c_1} \times p_{c_2} \times \ldots \times p_{c_n}.$$

- Randomly (i.e. according to the uniform distribution) pick a state description $\Phi(a_{i_1}, a_{i_2}, \ldots, a_{i_n})$ consistent with $c_1, c_2, \ldots, c_n$.
- $u^{\bar{p},L}(\Theta(a_{i_1}, a_{i_2}, \ldots, a_{i_n}))$ is the probability that $\Theta = \Phi$.

The $u^{\bar{p},L}$ satisfy Ex and Li+Sx and they turn out to be *the central building blocks* in the study of Sx. Precisely:

Theorem 8 *The probability function w on SL satisfies Li+Sx if and only if it can be represented in the form*

$$w = \int_{\bar{p}} u^{\bar{p},L} \, d\mu(\bar{p})$$

for some (normalized) countably additive measure μ on the Borel subsets of

$$\{\langle p_0, p_1, p_2, \ldots \rangle \mid p_0 \geq 0, \; p+1 \geq p_2 \geq p_3 \geq \ldots \geq 0, \; \sum_{i=0}^{\infty} p_i = 1\}.$$

Polyadic Relevance

Using the above Li+Sx Representation Theorem we can show the following *Relevance Principle*:

Theorem 9 *Let w be a probability function on L satisfying Li+Sx, let $\Theta(a_1, a_2, \ldots, a_n)$ be a state description and suppose that amongst $a_1, a_2, \ldots, a_n$ there are at least as many a_i which are indistinguishable from a_1 as there are a_i which are indistinguishable from a_2. Then given Θ, the probability that a_{n+1} is indistinguishable from a_1 is greater or equal to the probability that it is indistinguishable from a_2.*

Lack of Intuition Revisited

Recall:

> *Suppose you are told that*
>
> $$R(a_1, a_2) \wedge R(a_2, a_1) \wedge \neg R(a_1, a_3).$$
>
> *Which of $R(a_3, a_1), \neg R(a_3, a_1)$ should you think the more probable?*

– it follows from the above Theorem that if your w satisfies Li+Sx then $\neg R(a_3, a_1)$ will be at least as probable as $R(a_3, a_1)$.

I.e. analogy wins out.

Conformity

For simplicity suppose that the language L has just a single (binary) relation symbol R.

Consider the two 'unary relations' $R(a_1, x)$ and $R(x, x)$ of L.

Which of the two 'state descriptions'

$$R(a_1, a_1) \wedge R(a_1, a_2) \wedge \neg R(a_1, a_3) \wedge R(a_1, a_4)$$

$$R(a_1, a_1) \wedge R(a_2, a_2) \wedge \neg R(a_3, a_3) \wedge R(a_4, a_4)$$

should we think the more probable?

For myself I can see no logical reason why $R(a_1, x)$ and $R(x, x)$ should, in isolation, differ

– so if I was the agent I'd want to give the above 'state descriptions' the same probability, they should *conform*

Sx implies they do get the same probability

Conformity is much more general than this, loosely it specifies a range of pairs of sentences where there is no apparent reason why they should have different probabilities,

– and in all these cases Sx does indeed give them the same probability.

Genetic Variation

I doubt our agent would have expected this:

Theorem 10 *Let w satisfy Sx. Then the probability, according to w, that a_1, a_2 are indistinguishable but distinguishable from all other constants a_i is zero.*

Symmetry Rules?

de Finetti's Theorem tells us that any unary probability function w satisfying the symmetry principle Ex is a mixture of the simple probability functions $w_{\vec{x}}$ (which also satisfy Ex),

$$w = \int_{\vec{x}} w_{\vec{x}} \, d\mu(\vec{x}).$$

From this we can show that w satisfies the relevance principle of Instantial Relevance.

In turn these probability functions $w_{\vec{x}}$ are characterized by satisfying the irrelevance principle:

Constant Irrelevance Principle, IP
If $\theta, \phi \in SL$ have no constants in common then

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

Also it is easy to see that the $w_{\vec{x}}$ are 'extremal' solutions to Ex in the sense that the only mixture of functions satisfying Ex which give you $w_{\vec{x}}$ is the trivial mixture containing that same $w_{\vec{x}}$ alone.

So this Symmetry condition Ex implies a Relevance condition and its extremal solutions are characterized by an Irrelevance condition.

We see a similar phenomenon with Li+Sx.

The 'Symmetry' condition Li+Sx implies a Relevance condition and it's extremal solutions are the $u^{\bar{p},L}$ and they are characterized by the Irrelevance condition:

Weak Irrelevance Principle
If $\theta, \phi \in SL$ have no constants or relations in common then

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

This general phenomenon is well known in Mathematics.

Right now I do not know what we can conclude from it.

Maybe that such rational principles come in families of 3?

Some questions

Are all 'rational' principles ultimately derived from symmetry?

In what sense is Sx a 'symmetry condition'?

What is the limit of symmetry? Is it even consistent?

– in Unary Inductive Logic the limit of symmetry is Carnap's c_0

Can we ever hope to discover rational principles which will completely fix the agent's choice?

Thank to you for your indulgence!

THE END