Pure Inductive Logic

Nesin Maths Village
Summer School Course Notes

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 $M \in T\mathcal{L}$ 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 $[\vDash \theta]$ then $w(\theta) = 1$.

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

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

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 instantations of the predicates or their negations to be stochastically independent. That is,

$$c_\infty\left(\bigwedge_{j=1}^{k} \pm R_{i_j}(b_1^j, \ldots, b_{r_{i_j}}^j)\right) = \frac{1}{2^k}$$

(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".

Properties of probability functions

Let w be a probability function on SL. 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 $\theta \vee \phi \equiv \theta \vee (\neg\theta \wedge \phi)$ and $\models \neg(\theta \wedge (\neg\theta \wedge \phi))$, (Pd) and (P2) give

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

Also $\phi \equiv (\theta \wedge \phi) \vee (\neg\theta \wedge \phi)$ and $\models \neg((\neg\theta \wedge \phi) \wedge (\theta \wedge \phi))$ so by (Pd) and (P2) again,

$$w(\phi) = w((\theta \wedge \phi) \vee (\neg\theta \wedge \phi)) = w(\theta \wedge \phi) + w(\neg\theta \wedge \phi). \tag{3}$$

Eliminating $w(\neg\theta \wedge \phi)$ from (2), (3) now gives

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

Note that all the above properties follow by (P1) and (P2); the condition (P3) was not used.

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

$$w(\theta\,|\,\phi) = \frac{w(\theta \wedge \phi)}{w(\phi)}. \tag{4}$$

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 $w(\theta \mid \phi) = 1$. $\square$

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: *In the situation of zero knowledge, logically, or rationally, what probability function $w : SL \to [0,1]$ should a rational agent adopt when $w(\theta)$ is to represent the agent's probability that a sentence $\theta \in SL$ is true in his 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 $b_1, b_2, \ldots, b_m$,

$$w(\theta(a_1, a_2, \ldots, a_m)) = w(\theta(b_1, b_2, \ldots, b_m)). \tag{5}$$

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 predicates 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.

We will consider these principles and some equivalent formulations of them again in Chapter 3. Before that, however, we will discuss some justification of assuming that belief functions are *probability* functions.

Problems 1

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.

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

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

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

4. Show that Ex is equivalent to:

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

$$w(\theta(a_1, a_2, \dots, a_m)) = w(\theta(a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(m)})). \qquad (6)$$

5. 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 to Problems 1

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

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

$$= \int_D f(d) d\mu = \int_D w_d(\exists x \psi(x)) d\mu = w(\exists x \psi(x))$$

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.

5. 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 $i \in \{1, \ldots, 2^q\}$ 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.

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

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

Justifications of probability

Why use probability? Some other approaches would be much simpler to work with: for example *truth-functional belief like fuzzy logic*, where belief values of any sentence can be worked out from belief values of atomic sentences (instantiations of predicates or their negations).

The most frequently quoted justification for bellief as probability is the Dutch Book argument (Ramsey, de Finetti). It is based on identiying '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 it

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

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 Bet1_p is acceptable to the agent and $0 \leq q < p$ then Bet1_q is acceptable to him (with Bet1_q: larger gain if θ is true in M and smaller loss if θ is false).

• Similarly if Bet2_p is acceptable to the agent and $p < q \leq 1$ then Bet2_q is acceptable.

Consequently, there is some $P \in [0,1]$ such that for all $p < P$, Bet1_p is acceptable to the agent and for all $p > P$, Bet2_p is acceptable. Define $Bel(\theta)$ to be that P:

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

$Bel(\theta)$ 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 $M \in \mathcal{T}$ turns out to be.

Example Assume that L contains two unary predicates P and Q, and let $\theta = P(a_1) \wedge Q(a_1)$, $\quad \phi = \neg P(a_1) \vee \neg Q(a_1)$. Assume that $Bel(\theta) = 0.75$ and $Bel(\phi) = 0.45$. Then for θ the agent would accept $\mathrm{Bet1}_{0.7}$ and for ϕ he would accept $\mathrm{Bet1}_{0.4}$. Take the stake $s > 0$ to be the same for both bets. Depending on the ambient structure, the outcome for the agent is shown below:

$$\begin{array}{cccccc}
P(a_1) & Q(a_1) & \overbrace{P(a_1) \wedge Q(a_1),}^{\theta} & \overbrace{\neg P(a_1) \vee \neg Q(a_1)}^{\phi} & \overbrace{\text{Payoff}}^{Bet\,on\,\theta} & \overbrace{\text{Payoff}}^{Bet\,on\,\phi} \\
T & T & T & F & 0.3s & -0.4s \\
T & F & F & T & -0.7s & 0.6s \\
F & T & F & T & -0.7s & 0.6s \\
F & F & F & T & -0.7s & 0.6s
\end{array}$$

Overall, the agent certainly loses so the two bets acceptable to the agent are an example of a Dutch Book.

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 the above example, the combined effect of the two bets in the case that the ambient structure turns out to be M is for the agent to gain

$$s(V_M(\theta) - 0.7) + s(V_M(\phi) - 0.4) = s(V_M(\theta) + V_M(\phi) - 1.1).$$

Since we have $\models \neg(\theta \wedge \phi)$ it follows that $V_M(\theta) + V_M(\phi) = V_M(\theta \vee \phi) \leq 1$ which makes it plain that the agent loses no matter what.

Theorem 2 *Suppose that for* $Bel : SL \to [0,1]$ *there are no sets*[1] A, B, *sentences* $\theta_i \in SL$, $p_i \in [0, Bel(\theta_i))$, *stakes* s_i *for* $i \in A$, *and sentences* ϕ_j,

[1]Here we mean finite or countably infinite sets. If A and/or B are infinite, a convergence

stakes $t_j > 0$, $q_j \in (Bel(\phi_j), 1]$ *for* $j \in B$, *such that*

$$\sum_{i \in A} s_i(V_M(\theta_i) - p_i) + \sum_{j \in B}(-t_j)(V_M(\phi_j) - q_j) < 0 \tag{8}$$

for all $M \in \mathcal{T}$. *Then* Bel *satisfies (P1-3).*

Note that the conditions of the theorem express that there is no Dutch book against Bel (the convergence condition from the footnote giving a limit on the loss that the agent or the bookmaker can be exposed to in the worst possible case).

Proof of Theorem 2: For (P1) suppose that $\theta \in SL$ and $\vDash \theta$ 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 (8), contradiction.

Suppose that (P2) fails, say $\theta, \phi \in SL$ are such that $\vDash \neg(\theta \wedge \phi)$ 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 (8) and contradicting our assumption. A similar argument when

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

condition is assumed to be satisfied, namely that there is $K > 0$ such that

$$\left|\sum_{i \in A} s_i(V_M(\theta_i) - p_i)\right|, \quad \left|\sum_{j \in B} t_j(V_M(\phi_j) - q_j)\right| < K \tag{7}$$

for each M.

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 (8) 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 (4). 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 $M \not\models \phi$.

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 (cf Problem II.2) that the requirement of no Dutch book against the agent still forces Bel to satisfy (P1),(P2),(P3) and moreover that for al 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 (4). We remark that, conversely, if $Bel : SL \to [0,1]$ is a probability function then Bel cannot be (conditionally) Dutch Booked. See for example [1] for a proof.

Problems 2

1. Let L contains 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.

2. (i) Write down what the gain/loss of the agent is after accepting $\mathrm{CBet1}_p$ or $\mathrm{CBet2}_p$ 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 $\theta, \phi \in SL$, $Bel(\theta \,|\, \phi) \cdot Bel(\phi) = Bel(\theta \wedge \phi)$.

Solutions to Problems 2

1. (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 $s > 0$: $\mathrm{Bet1}_{0.75}$ on θ and $\mathrm{Bet2}_{0.35}$ 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 the remark concluding Section II 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).

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

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.

2. (i) Accepting CBet1_p means gaining

$$sV_M(\phi)(V_M(\theta) - p)$$

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 $M \in \mathcal{T}$ we have

$$\begin{gathered} \textstyle\sum_{i\in A} s_i(V_M(\theta_i) - p_i) + \sum_{i\in B}(-t_i)(V_M(\phi_i) - q_i) + \\ \sum_{i\in C} u_i V_M(\psi_i)(V_M(\eta_i) - r_i) + \sum_{i\in D}(-v_i)V_M(\xi_i)(V_M(\zeta_i) - m_i) \;<\; 0 \end{gathered} \tag{9}$$

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

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

$$Bel(\theta \,|\, \phi) \cdot Bel(\phi) < Bel(\theta \wedge \phi). \tag{10}$$

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

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

for any M, since $V_M(\phi)V_M(\theta) = V_M(\theta \wedge \phi)$, contradicting the given no Dutch Book condition. Hence with (10), $Bel(\phi) < 1$. 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 (10) cannot hold. So if the required equality fails it must be because

$$Bel(\theta \,|\, \phi) \cdot Bel(\phi) > Bel(\theta \wedge \phi). \tag{11}$$

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

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

Specifying Probability Functions

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 $\theta(b_1, b_2, \ldots, b_m) \in QFSL$ is logically equivalent to a disjunction of state descriptions for $b_1, b_2, \ldots, b_m$

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

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

$$w(\theta(\vec{b})) = \sum_{\Theta \in S} w(\Theta(\vec{b})). \tag{12}$$

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 $a_1, a_2, \ldots, a_n$ $(n \in \mathbb{N})$.

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

***

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

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

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

(iii) $$w(\Theta(a_1, a_2, \ldots, a_m)) = \sum_{\Phi(a_1,\ldots,a_{m+1}) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_1, a_2, \ldots, a_{m+1})). \tag{13}$$

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

$$w(\theta(b_1, b_2, \ldots, b_m)) = \sum_{\Theta(a_1,\ldots,a_k) \models \theta(b_1,\ldots,b_m)} w(\Theta(a_1, a_2, \ldots, a_k)) \tag{14}$$

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

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

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

$$w(\Phi(a_1, \ldots, a_n)) = w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)}))\,. \tag{15}$$

Then w satisfies Ex on quantifier-free formulas.

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

$$\begin{aligned}
w(\Theta(a_1, \ldots, a_m)) &= \sum_{\Phi(a_1,\ldots,a_n) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_1, \ldots, a_n)) \\
&= \sum_{\Phi(a_1,\ldots,a_n) \models \Theta(a_1,\ldots,a_m)} w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)})), \quad \text{by (15)}, \\
&= \sum_{\Phi(a_{\tau(1)},\ldots,a_{\tau(n)}) \models \Theta(a_{\tau(1)},\ldots,a_{\tau(m)})} w(\Phi(a_{\tau(1)}, \ldots, a_{\tau(n)})) \\
&= \sum_{\Psi(a_1,\ldots,a_n) \models \Theta(a_{\tau(1)},\ldots,a_{\tau(m)})} w(\Psi(a_1, \ldots, a_n)) \\
&= w(\Theta(a_{\tau(1)}, \ldots, a_{\tau(m)})) \\
&= w(\Theta(a_{i_1}, \ldots, a_{i_m})) \\
&= w(\Theta(b_1, \ldots, b_m)) \,.
\end{aligned}$$

It follows that w satisfies (5) on state descriptions and hence by virtue of (12) on $QFSL$. ■

Extending Probability Functions from QFSL to all sentences

Theorem 3 *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

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

of $\mathcal{T}$ 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

$$\bigcup_{i \in \mathbb{N}} [\phi_i] = [\theta]. \tag{16}$$

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 $\mathcal{T}$ its substructure with universe the $\{a_1, a_2, a_3, \ldots\}$ will be, and will satisfy the same quantifier free sentences, thus contradicting (16). 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 $\psi(x)$)

$$\begin{aligned}
[\exists x\, \psi(x)] &= \{ M \in \mathcal{T} \mid M \vDash \exists x\, \psi(x) \} \\
&= \{ M \in \mathcal{T} \mid M \vDash \psi(a_i), \text{ some } i \in \mathbb{N}^+ \} \\
&= \bigcup_{i \in \mathbb{N}^+} \{ M \in \mathcal{T} \mid M \vDash \psi(a_i) \} \\
&= \bigcup_{i \in \mathbb{N}^+} [\psi(a_i)] \qquad (17)
\end{aligned}$$

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 (17) 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.

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.

$\blacksquare$

Problems 3

1. Let $L = \{R\}$, where R is binary. A state description for m constants, $\Theta(b_1, b_2, \dots, b_m)$, 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.

,

2. (*Trurh functional belief versus probability*) We say that a function

$$Bel : SL \to [0,1]$$

is truth-functional if there are functions

$$\begin{aligned} &F_\neg : [0,1] \to [0,1], \\ &F_\wedge : [0,1] \times [0,1] \to [0,1], \\ &F_\vee : [0,1] \times [0,1] \to [0,1] \end{aligned}$$

such that for $\phi, \psi \in SL$,

$$\begin{aligned} &Bel(\neg\phi) = F_\neg(Bel(\phi), \\ &Bel(\phi \wedge \psi)) = F_\wedge(Bel(\phi), Bel(\psi)), \\ &Bel(\phi \vee \psi)) = F_\vee(Bel(\phi), Bel(\psi)) \end{aligned}$$

(a) Give an example of a probability function which is truth functional.

(b) Show that no probability function w which gives value $\frac{1}{2}$ to some sentence can be truth-functional.

(c) Let L be as in Problem 1 and let Bel be a truth-functional belief function with $F_\neg(x) = 1 - x$ and $F_\wedge(x, y) = \min\{x, y\}$ and such that $Bel(R(a_i, a_j)) = 0.4$ for each $i, j \in \{1, 2\}$. Write down

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

(note that it does not matter in which order the conjunctions are applied) and suggest how to find two probability functions w_1, w_2 which agree with Bel on $R(a_i, a_j)$ $(i, j \in \{1, 2\})$ but disagree with Bel and each other on

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

(w_1, w_2 are not required to satisfy Ex - the exercise is harder with Ex).

3. 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 3

1. 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 $\{1, 2, \dots, m\}$ 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 $w(D) = w(\overline{D})$.

2. (a) If M is any structure from $\mathcal{T}L$ then V_M is truth functional with

$$F_\neg(x) = 1 - x, \quad F_\wedge(x, y) = \min\{x, y\}, \quad F_\vee(x, y) = \max\{x, y\}.$$

(b) Assume that w is a probability function and $w(\phi) = \frac{1}{2}$. Then

$$w(\neg\phi) = 1 - \frac{1}{2} = \frac{1}{2}$$

and if w was truth-functional, we would have

$$0 = w(\phi \wedge \neg(\phi)) = F_\wedge\left(\frac{1}{2}, \frac{1}{2}\right) = w(\phi \wedge \phi) = \frac{1}{2},$$

contradiction.

(c)We have $Bel(\neg R(a_i, a_j)) = 1 - 0.4 = 0.6$ so

$$Bel(\neg R(a_1, a_1) \wedge \neg R(a_1, a_2) \wedge R(a_2, a_1) \wedge R(a_2, a_2)) = 0.4.$$

We can specify various probability functions w satisfying $w(R(a_i, a_j)) = 0.4$ for $i, j \in \{1, 2\}$ using the scheme (13) (and then Gaifman's theorem). We shall use the matrix notation to represent state descriptions $\Theta(a_1, a_2)$. Let $x_1, x_2, \dots, x_{16}$ stand, in this order, for w of the state descriptions represented by

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \dots, \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

(so $w \begin{pmatrix} d_{1,1} & d_{1,2} \\ d_{2,1} & d_{2,2} \end{pmatrix} = x_{(1+d_{2,2}+2d_{2,1}+4d_{1,2}+8d_{1,1})}$).

The conditions $w(\top) = 1$ and $w(R(a_i, a_j) = 0.4$ for $i, j \in \{1, 2\}$ amount to

$$\sum_{i=1}^{16} x_i = 1$$

and

$$\sum_{i=9}^{16} x_i = \sum_{i=5}^{8} x_i + \sum_{i=13}^{16} x_i = x_3 + x_4 + x_7 + x_8 + x_{11} + x_{12} + x_{15} + x_{16} = \sum_{i=1}^{8} x_{2i} = 0.4$$

Any solution satisfying $x_i \geq 0$ for all i extends to probability functions on SL (for example, we can define w for larger state descriptions to satisfy (13) by giving each extension of a state description for a_1, a_2 equal value). Note that

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

is x_4.

So for example, two possible solutions are obtained as above from

$$x_1 = x_2 = x_4 = x_{13} = x_{17} = 0.2; \quad x_i = 0 \text{ otherwise}$$

and

$$x_1 = 0.2, \quad x_8 = x_9 = 0.4; \quad x_i = 0 \text{ otherwise.}$$

3. w_1 satisfies conditions (13): (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,

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

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

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

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

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; in this section, all relation symbols $R_1, R_2, \ldots, R_q$ in L are assumed to be unary. As such they are referred to more often as predicate rather than relation symbols.

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

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

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

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

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 $\Theta(b_1, b_2, \ldots, b_m)$ 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 $\langle m_1, m_2, \ldots, m_{2^q} \rangle$, 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 $b_1, \ldots, b_m$ 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 1 holds since for a permutation τ of $\{1, 2, \ldots, m\}$, 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 $a_{k_1}, \ldots, a_{k_m}$, $a_{j_1}, \ldots, a_{j_m}$ 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 $\bigwedge_{i=1}^m \alpha_{g_i}(a_{j_i}) \equiv \bigwedge_{i=1}^m \alpha_{h_i}(a_{\sigma(k_i)})$. By Problem I.4, 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 (13) 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:

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

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

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

Clearly, $w_{\vec{c}}$ satisfies Ex. However, Px and SN hold only for special choices of $\vec{c}$, see Problem IV.2. 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

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

Proposition 4 *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

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

Conversely if w satisfies Ex and IP then by repeated application

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

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. ■

de Finetti's Representation Theorem

Let $L = \{R_1, \ldots, R_q\}$ 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

$$
\begin{aligned}
w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) &= \int_{\mathbb{D}_{2^q}} \prod_{j=1}^{2^q} x_j^{m_j} \, d\mu(\vec{x}) \\
&= \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) d\mu(\vec{x}),
\end{aligned} \tag{19}
$$

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

Conversely, given a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ the function w defined by (19) 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 (as far as values of any w satisfying Ex are concerned) state descriptions $\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$ are fully characterized by two numbers

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

If w is a probability function satisfying Ex and m_1, m_2 as above, we define

$$w(m_1, m_2) = w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)\right) \tag{20}$$

For fixed m_1, m_2 there are $\binom{m}{m_1}$ distinct possibilities for the ordering of the $h_1, h_2, \ldots, h_m$. Let $r > m$. Since state descriptions for $b_1, \ldots, b_r$ are exclusive and exhaustive,

$$1 = w(\top) = \sum_{r_1+r_2=r} \binom{r}{r_1} w(r_1, r_2). \tag{21}$$

Also, considering which of them extend a given $\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$ as above,

$$w(m_1, m_2) = \sum_{\substack{r_1+r_2=r \\ m_1 \leq r_1,\, m_2 \leq r_2}} \binom{r-m}{r_1-m_1} w(r_1, r_2). \tag{22}$$

From (21) 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 $\langle r_1/r, r_2/r \rangle \in \mathbb{D}_2$. Note that from (22) we obtain that $w(m_1, m_2)$ equals

$$\sum_{\substack{r_1+r_2=r \\ m_1 \leq r_1,\, m_2 \leq r_2}} \binom{r-m}{r_1-m_1} \binom{r}{r_1}^{-1} \binom{r}{r_1} w(r_1, r_2). \tag{23}$$

We shall show that

$$\left| \binom{r-m}{r_1-m_1} \binom{r}{r_1}^{-1} - \left(\frac{r_1}{r}\right)^{m_1} \left(\frac{r_2}{r}\right)^{m_2} \right| \tag{24}$$

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

Notice that the left hand term in (24) can be written as

$$\left(\frac{r_1}{r}\right)^{m_1} \left(\frac{r_2}{r}\right)^{m_2} \frac{(1-r_1^{-1})\cdots(1-(m_1-1)r_1^{-1})(1-r_2^{-1})\cdots(1-(m_2-1)r_2^{-1})}{(1-r^{-1})\cdots(1-(m-1)r^{-1})}. \tag{25}$$

We now consider cases.

• If $m_1 = m_2 = 0$ then (24) is zero.

• If $m_2 > 0$ and $r_2 \leq \sqrt{r}$ then both terms in (24) are less than $r^{-m_2/2} \leq r^{-1/2}$ (this is justified for large r e.g because for any $r_1 \leq r$, $s \leq m_1$ we have $\frac{1-sr_1^{-1}}{1-sr^{-1}} \leq 1$ and for $\sqrt{r} > m$, $s \leq m_2$, we have $\frac{1-sr_2^{-1}}{1-(m_1+s)r^{-1}} \leq 1$)

and similarly if $m_1 > 0$ and $r_1 \leq \sqrt{r}$. If $m_2 > 0$ and $r_2 > \sqrt{r}$ and either $m_1 = 0$ or $r_1 > \sqrt{r}$ then using (25) and the fact that $r_1/r, r_2/r \leq 1$ we see that (24) is at most

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

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

$$\sum_{\substack{r_1+r_2=r \\ m_1 \leq r_1,\, m_2 \leq r_2}} \left(\frac{r_1}{r}\right)^{m_1} \left(\frac{r_2}{r}\right)^{m_2} \mu_r(\{\langle r_1/r, r_2/r\rangle\}). \tag{26}$$

In turn this equals the limit of the same expressions but summed simply over $0 \leq r_1, r_2$, $r_1 + r_2 = r$ since from (21) (or trivially if $m_1 = 0$),

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

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

In other words,

$$w(m_1, m_2) = \lim_{r\to\infty} \int_{\mathbb{D}_2} x_1^{m_1} x_2^{m_2}\, d\mu_r(\langle x_1, x_2\rangle). \tag{27}$$

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 (27).

Finally the converse result, that functions w defined by (19) extend to probability functions on SL satisfying Ex follows by Theorem 3. $\blacksquare$

From (19) 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 (19). 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 5 *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 $\theta \in SL$,*

$$w(\theta) = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}(\theta)\, d\mu(\vec{x}). \tag{28}$$

Conversely given a measure μ on $\mathbb{D}_{2^q}$, w defined by (28) is a probability function on SL satisfying Ex.

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

$$w = \int_{\mathbb{D}_{2^q}} w_{\vec{x}}\, d\mu(\vec{x}), \tag{29}$$

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. $\blacksquare$

Problems 4

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

$$w\left(\alpha_j(a_{m+1}) \,|\, \bigwedge_{i=1}^m \alpha_{h_i}(a_i)\right) = \frac{m_j + \lambda 2^{-q}}{m + \lambda} \tag{30}$$

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

2. 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.

3. (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

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

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 $w = c_\infty$.

Solutions to Problems 4

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

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

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 (13), 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 condition (i) and (ii) from (13) 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 (cf Section 3). Furthermore, the extension satisfies Ex by Lemma 1 since from the above it can be see 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 $\bigwedge_{i=1}^n \alpha_{h_i}(a_i)$.

2. 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 $\left\langle \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right\rangle$.

(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.

3. 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

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

we have

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

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

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

which simplifies to give

$$(x_1 - y_1)(x_1 y_2 - y_1 x_2) < 0.$$

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 Finettti representation and addiing suitable pairs of equalities, we obtain

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

whih means that $x_1 = x_2$, $x_3 = x_4$ and $x_1 = x_3$ on $\mathbb{D}_4$ so $x_1 = x_2 = x_3 = x_4$ except possibly on a set of μ measure 0, which means that w is c_∞ as required.

Further Unary Principles

As in the previous section, all relation symbols $R_1, R_2, \ldots, R_q$ in L are assumed to be unary in this one.

The following principle is an attempt at formalizing 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).

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

$$w(\alpha(a_{n+2}) \mid \alpha(a_{n+1}) \wedge \theta(a_1, a_2, \ldots, a_n)) \geq w(\alpha(a_{n+2}) \mid \theta(a_1, a_2, \ldots, a_n)). \quad (31)$$

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

Theorem 6 *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 (31), 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 4 the $w_{\vec{x}}$ satisfy IP)

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

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

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

and (31) amounts to

$$\left(\int_{\mathbb{D}_{2q}} w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right) \cdot \left(\int_{\mathbb{D}_{2q}} x_1^2 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right) \;\geq\; \left(\int_{\mathbb{D}_{2q}} x_1 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x})\right)^2 . \quad (32)$$

[1] In what follows we use a 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.

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 (32) is equivalent to

$$\int_{\mathbb{D}_{2^q}} \left(x_1 A - \int_{\mathbb{D}_{2^q}} x_1 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x}) \right)^2 w_{\vec{x}}(\theta(\vec{a}))\, d\mu(\vec{x}) \;\geq\; 0 \tag{33}$$

as can be seen by multiplying out the square and dividing by A. But obviously, being an integral of a non-negative function, (33) holds, as required. ■

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 τ of $\{1, 2, \ldots, 2^q\}$ and constants $b_1, b_2, \ldots, b_m$,

$$w\left(\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i) \right) = w\left(\bigwedge_{i=1}^{m} \alpha_{\tau(h_i)}(b_i) \right). \tag{34}$$

Equivalently, in the presence of Ex, Ax asserts that the left hand side of (34) depends only on the *spectrum* of the state description $\bigwedge_{i=1}^{m} \alpha_{h_i}(b_i)$, that is on the *multiset* $\{m_1, m_2, \ldots, m_{2^q}\}$, where, again, $m_j = |\{i \mid h_i = j\}|$.

Quite diffferent 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,

$$\lim_{n \to \infty} \left(w\left(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i) \right) - \frac{u_j(n)}{n} \right) = 0 \tag{35}$$

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

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.

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

$$w\left(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) \tag{36}$$

depends only on n *and* $r = |\{\, i \mid 1 \leq i \leq n \text{ and } h_i = j \,\}|$ *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 (36) does not depend on j, all atoms are treated in the same way.

The functions defined in Problem 4.1 clearly satisfy JSP. We refer to them as Carnap continuum functions and denote them c_λ^L, so for $\lambda > 0$,

$$c_\lambda^L\left(\alpha_j(a_{n+1}) \mid \bigwedge_{i=1}^{n} \alpha_{h_i}(a_i)\right) = \frac{m_j + \lambda 2^{-q}}{n + \lambda}$$

where $m_j = |\{i \mid 1 \leq i \leq n \text{ and } h_i = j\}|$. The same expressions with 0 or ∞ in place of λ lead us to define c_∞^L by

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

and c_0^L by

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

Theorem 7 *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.

For the other direction assume that w satisfies JSP. Then w satisfies Ax (Problem 5.3) 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

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

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

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

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

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

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

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

for all n, r. From

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

we get

$$g(1,1) + (2^q - 1)g(0,1) = 1. \tag{37}$$

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 5.4 and (37)).

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

$$g(r,n) = \frac{r + \lambda 2^{-q}}{n+\lambda} \qquad (r = 0, 1, \ldots, n). \tag{38}$$

We have already shown it for $n = 0, 1$. Assume that $n \geq 1$ and (38) holds for n. For $u + v = n + 1$, and distinct m, k,

$$1 = w\left(\bigvee_{h=1}^{2^q} \alpha_h(a_{n+1}) \mid \bigwedge_{i=1}^{u} \alpha_m(a_i) \wedge \bigwedge_{i=u+1}^{n+1} \alpha_k(a_i)\right)$$

so

$$1 \;=\; g(u, n+1) + g(v, n+1) + (2^q - 2)g(0, n+1). \tag{39}$$

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

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

For $r+s+t=n$ and distinct m, j, k,

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

so[1]

$$g(r,n+1)g(s,n) = g(s,n+1)g(r,n). \tag{40}$$

Using $s=0$ and the inductive hypothesis gives

$$g(r,n+1) = (r\lambda^{-1}2^q+1)g(0,n+1). \tag{41}$$

Taking $u=1,\ v=n$ in (39) and using $r=1,n$ in (41) gives

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

and hence

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

Substituting in (41) now gives (38) too for $n+1$ and $r=1,2,\dots,n$, and finally also for $r=n+1$ using (39) with $u=0,\ v=n+1$. ■

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

Problems 5

1. Let L be a unary language.

(a) Show that any probability function w which satisfies SN also satisfies (34) for $m = 1$, that is, for any constant b and any two atoms α_k, α_j,

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

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

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

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

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

(b) For functions satisfying Ex and Ax, conjecture and prove a representation theorem, using the functions $v_{\vec{c}}$ from (a) .

(c) Conjecture and prove representation theorem for functions satisfying Ex and Px.

3. Show that JSP implies Ax.

4. 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

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

Solutions to Problems 5

1.(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 $\alpha_2(a_1) \wedge \alpha_3(a_2)$.

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

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

(b) **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*

$$
w = \int_{\mathbb{D}_{2^q}} v_{\vec{x}}\, d\mu(\vec{x}). \tag{42}
$$

Conversely, given a measure μ on the Borel subsets of $\mathbb{D}_{2^q}$ the probability function w on SL defined by (42) satisfies Ax (and Ex).

Proof 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 $\sigma \in \mathsf{S}_{2^q}$,

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

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 (42) 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.

3. 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.

4. 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.

Polyadic Pure Inductive Logic

To start with, we shall restrict our considerations to the case of L containing a single, binary, relation symbol R. It reduces notational difficulties while still allowing insight into the polyadic context.

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

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

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

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

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 $n \times n$ $\{0, 1\}$-matrix $T = (t_{i,j})$.[1] (Recall that we have already used such a representation in Problems III.)

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 $D = (d_{i,j})$ 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

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

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 $i, j \leq n$,

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

This does uniquely determines a probability function on SL satisfying Ex, see Problem 6.1(b).

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

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

of such w^D with non-standard D. Remaining within standard mathematics, 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:

Lemma 2 *For a probability function w on SL satisfying Ex, $n \in \mathbb{N}$ and $\epsilon > 0$ there is an $N \in \mathbb{N}$ and $\lambda_D \geq 0$ for each $N \times N$ $\{0,1\}$-matrix D such that $\sum_D \lambda_D = 1$ and for any $\theta(a_1, \ldots, a_n) \in QFSL$,*[1]

$$|w(\theta) - \sum_D \lambda_D w^D(\theta)| < \epsilon.$$

Proof Let $\theta(a_1, \ldots, a_n) \in QFSL$ and let $N > n$. We have

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

For a state description $\Phi(a_1, \ldots, a_N)$ 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

$$\Psi(a_1, \ldots, a_N) = \Phi(a_{\sigma(1)}, \ldots, a_{\sigma(N)}) \tag{44}$$

(where σ is a permutation of $\{1, \ldots, N\}$). Collecting state description from the same equivalence classes together, we can write

$$w(\theta(a_1, \ldots, a_n)) = \sum_{\overline{\Phi}} w(\Phi(a_1, \ldots, a_N)) \cdot K(\Phi, \theta), \tag{45}$$

where $K(\Phi, \theta)$ is $|\Psi \in \bar{\Phi};\ \Psi(a_1, \ldots, a_N) \models \theta(a_1, \ldots, a_n)|$.

Let $K(\Phi)$ be the number of all state descriptions that can be obtained from Φ by permuting constants. (Note that this is not necessarily $N!$ because some permutation may yield the same state descriptions, but every state description in $K(\Phi)$ is obtained from Φ by the same number of permutations.) Then

$$\frac{K(\Phi, \theta)}{K(\Phi)} \tag{46}$$

[1]This notation means that all the constants appearing in $\theta(a_1, \ldots, a_n)$ are amongst the $a_1, \ldots, a_n$.

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. (46) is also the probability that when (uniformly) randomly picking, *without* replacement, $h(1), h(2), \ldots, h(n)$ from $\{1, 2, \ldots, N\}$,

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

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

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

if there are no repeats in the $h(1), h(2), \ldots, h(n)$ (and hence the difference is of order $N^{-(n+1)}$) or N^{-n} if there are repeats. There are N^n n-tuples $h(1), h(2), \ldots, h(n)$ altogether and less than $\binom{n}{2} N^{n-1}$ of them are with repeats, so - appealing to Problem 6.1(b) - the difference between $\frac{K(\Phi,\theta)}{K(\Phi)}$ and $w^{D_\Phi}(\theta)$ is of order N^{-1}.

For $\Psi \in \overline{\Phi}$, let D_Ψ be the $N \times N$ matrix representing Ψ (note that D_Ψ obtains from D_Φ by simultaneously permuting rows and columns and that $w^{D_\Phi} = w^{D_\Psi}$). Let

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

From (45) we have

$$\begin{aligned} w(\theta(a_1, \ldots, a_n)) &= \sum_{\overline{\Phi}} w(\Phi(a_1, \ldots, a_N)) \cdot K(\Phi, \theta) \\ &= \sum_{\overline{\Phi}} K(\Phi) \cdot w(\Phi(a_1, \ldots, a_N)) \cdot \frac{K(\Phi,\theta)}{K(\Phi)}, \end{aligned}$$

so since there are $K(\Phi)$ state descriptions Ψ in $\overline{\Phi}$, we obtain

$$w(\theta(a_1, \ldots, a_n)) = \sum_{\overline{\Phi}} \sum_{\Psi \in \overline{\Phi}} \lambda_{D_\Psi} \cdot \frac{K(\Phi,\theta)}{K(\Phi)}.$$

Furthermore:

- with Ψ ranging over all state descriptions for $a_1, \dots, a_N$, we have

$$\sum_{\Psi} w(\Psi(a_1, \dots, a_N)) = \sum_{\overline{\Phi}} \sum_{\Psi \in \overline{\Phi}} \lambda_{D_\Psi} = 1,$$

- $\left| \frac{K(\Phi,\theta)}{K(\Phi)} - w^{D_\Phi}(\theta) \right|$ is of order N^{-1},
- For any $\Psi \in \bar{\Phi}$, $w^{D_\Psi}(\theta)$ equals $w^{D_\Phi}(\theta)$.

It follows that for N large enough

$$|w(\theta) - \sum_{\Psi} \lambda_{D_\Psi} w^{D_\Psi}(\theta)| = |w(\theta) - \sum_{D} \lambda_D w^D(\theta)| < \epsilon,$$

as required.

■

To illustrate how the above theorem can be useful, we will prove the following lemma, which is important for the study of analogy in inductive logic.

Lemma 3 *For a probability function w on SL 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 the previous lemma 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

$$e_i = |\{j \,|\, d_{i,j} = 1\}| \,.$$

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.

■

Next we consider some principles that arguably capture some of our intuition about analogy. L now stands again for a general language.

The Counterpart Principle, CP

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

$$w(\theta \,|\, \theta') \geq w(\theta). \tag{47}$$

A stronger version, SCP: *If 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*

$$w(\theta \,|\, \theta') \geq w(\theta \,|\, \theta'') \geq w(\theta). \tag{48}$$

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 consistent family of probability functions satisfying Ex and Px, in which there is a probability function for any language. Formally, this requirement is expressed as

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 $S\mathcal{L}$.

It is easier to show that Li is enough for the basic version of the Counterpart Principle:

Theorem 8 *If w satisfies Li then w satisfies the CP.*

Proof Assume that w satisfies Li. Taking the 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 6. In particular then

$$v(R(a_1) \mid R(a_2)) \geq v(R(a_1)). \qquad (49)$$

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 can proceed similarly. It is convenient to introduce the following notation: for a natural number c, define

$$\overline{c} = 2c - 1, \qquad \underline{c} = 2c\,.$$

Let

$$\begin{aligned}
\theta = \theta(&a_1, \dots, a_m, a_{\overline{m+1}}, \dots, a_{\overline{m+t}}, a_{\underline{m+k+1}}, \dots, a_{\underline{m+2k}},\\
&R_1, \dots, R_p, R_{\overline{p+1}}, \dots, R_{\overline{p+s}}, R_{\underline{p+j+1}}, \dots, R_{\underline{p+2j}}),\\
\theta' = \theta(&a_1, \dots, a_m, a_{\overline{m+1}}, \dots, a_{\overline{m+t}}, a_{\underline{m+3k+1}}, \dots, a_{\underline{m+4k}},\\
&R_1, \dots, R_p, R_{\overline{p+1}}, \dots, R_{\overline{p+s}}, R_{\underline{p+3j+1}}, \dots, R_{\underline{p+4j}}),\\
\theta'' = \theta(&a_1, \dots, a_m, a_{\overline{m+2t+1}}, \dots, a_{\overline{m+3t}}, a_{\underline{m+3k+1}}, \dots, a_{\underline{m+4k}},\\
&R_1, \dots, R_p, R_{\overline{p+2s+1}}, \dots, R_{\overline{p+3s}}, R_{\underline{p+3j+1}}, \dots, R_{\underline{p+4j}}).
\end{aligned}$$

(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

$$\begin{aligned}
\theta(&a_1, \dots, a_m, a_{\overline{m+it+1}}, \dots, a_{\overline{m+(i+1)t}}, a_{\underline{m+lk+1}}, \dots, a_{\underline{m+(l+1)k}},\\
&R_1, \dots, R_p, R_{\overline{p+is+1}}, \dots, R_{\overline{p+(i+1)s}}, R_{\underline{p+lj+1}}, \dots, R_{\underline{p+(l+1)j}}),
\end{aligned}$$

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 3.

$$v(R(a_1,a_2)|R(a_1,a_4)) \geq v(R(a_1,a_2)|R(a_3,a_4))$$

so

$$w(\theta_{1,2} \mid \theta_{1,4}) \geq w(\theta_{1,2} \mid \theta_{3,4})$$

and the result follows.

Problems 6

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

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

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

$$d_{h(i)} = t_i.$$

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

(b) 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, $h(1), h(2), \dots, h(n)$ from $\{1, 2, \dots, N\}$ 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 $\theta(a_1, \dots, a_n) \in QFSL$, $w^D(\theta)$ is the probability that when (uniformly) randomly picking, with replacement, $h(1), h(2), \dots, h(n)$ from $\{1, 2, \dots, N\}$,

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

Solutions to Problems 6

1. (a) 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

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

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

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

(b) First note the conditions (13) clearly hold since the picking is with replacement. Explicitly, for

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

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

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

and

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

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

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

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

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

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

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

as required.

Ex follows by Lemma 1 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 $b_1, \dots, b_n$ is also the probability of (uniformly) randomly picking, with replacement, $h(1), h(2), \dots, h(n)$ from $\{1, 2, \dots, N\}$ such that for each $i, j \leq n$,
$$d_{h(i),h(j)} = t_{i,j}.$$

Since any $\theta(b_1, \dots, b_n) \in QFSL$ is logically equivalent to a disjunction of state descriptions, $w_D(\theta(b_1, \dots, b_n))$ is the sum of $w^D(\Theta(b_1, \dots, b_n))$ over those $\Theta(b_1, \dots b_n)$ that logically imply it and hence the probability that when (uniformly) randomly picking, with replacement, $h(1), h(2), \dots, h(n)$ from $\{1, 2, \dots, N\}$,
$$\bigwedge_{i,j\leq n} R^{d_{h(i),h(j)}}(a_i, a_j) \models \theta(a_1, \dots, a_n),$$

as required.

References

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

[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.