Logics with Probability Operators and Quantifiers

N. Ikodinović[1] D. Doder[2]

[1]University of Belgrade
Faculty of Mathematics

[2]Utrecht Univeristy
Department of Information and Computing Science

34th EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION, University of Ljubljana, Faculty of Computer and Information Science, July 31 – August 11, 2023

Course Overview and Introduction

Logics with Probability Operators and Quantifiers
Area: Logic and Computation (LoCo)
Level: Advanced
Lecturers: Nebojša Ikodinović and Dragan Doder
Goals:

- Present the major areas of research that connect probability theory and mathematical logic
- Provide guidelines for developing probabilistic (formal) logics

`http://www.matf.bg.ac.rs/p/nebojsa-ikodinovic/cas/4948/esslli-2023/`

`http://enastava.matf.bg.ac.rs/~ikodinovic/Logic_and_Probability_Ikodinovic.pdf`

Table of Contents

Table of Contents

1. Why specify probability functions on sentences of predicate languages?
    - classical logic – classical structures
    - inductive logic – probability structures
    - probability structures → a class of classical structures with weights
2. Very large finite phenomena
    - a class of classical structures with weights →
      → classical (random) structure
    - new logics
3. Axiomatization issue
    - Markov process
      (probabilistic machines, transition systems, real-valued structures etc.)
    - L_P-completeness

Table of Contents

Motivating example

- Deductive reasoning (Mathematics, Mathematical logic)
- Inductive reasoning (Philosophy, Logic)

The connections between logic and computer science are growing rapidly and are becoming deeper.

Carnap (1891–1970) was probably the first who tried to give a mathematically rigorous foundation of inductive reasoning.

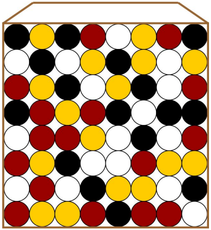
Motivating example

By 'inductive logic' I understand a theory of logical probability providing rules for inductive thinking. (R. Carnap, R. C. Jeffrey, Studies in Inductive Logic and Probability (I))

Exercise

What is the probability of picking a black ball from the box?

1)



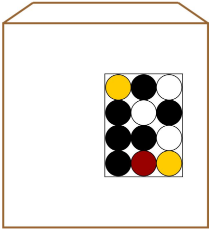
Motivating example

By 'inductive logic' I understand a theory of logical probability providing rules for inductive thinking. (R. Carnap, R. C. Jeffrey, Studies in Inductive Logic and Probability (I))

Exercise

What is the probability of picking a black ball from the box?

2)



experience:

12 balls, 6 of which are black

knowledge:

The colors of the balls are uniformly distributed in the box, but exceptions are possible.

Write down your 'feeling-based' answer.

Universe (of discourse):

Ann	Ben	Cam	Deb	Eva	Fox
a	b	c	d	e	f

Predicates:

'$\cdots$ is an artist'	'$\cdots$ is a barber'	'$\cdots$ is following $\cdots$'
$A(\cdot)$	$B(\cdot)$	$F(\cdot,\cdot)$

Atomic sentences:

Ann is a artist.	Ann is a barber.	Fox is following Cam.
$A(a)$	$B(a)$	$F(f,c)$

$\mathtt{L}(\mathbb{U})$ is the set of all atomic sentences, where $\mathtt{L}$ is the set of predicate symbols, and $\mathbb{U}$ is the universe.

How many atomic sentences are there in $\{A,B,F\}(\{a,b,c,d,e,f\})$?

$\#\{A,B,F\}(\{a,b,c,d,e,f\}) = 6^1 + 6^1 + 6^2 = 48$

If $\mathtt{L}$ and $\mathbb{U}$ are finite sets, then $\#\mathtt{L}(\mathbb{U}) = \sum_{P\in\mathtt{L}} \#\mathbb{U}^{\mathrm{arity}(P)}$.

Truth assignments

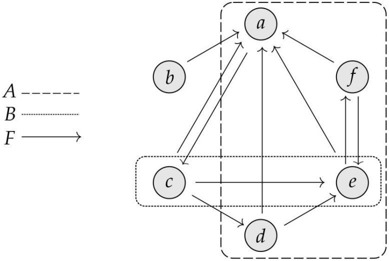
An **atomic truth assignment** (a valuation) is any function $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$.

$A(a)$	$A(b)$	$A(c)$	$A(d)$	$A(e)$	$A(f)$
$B(a)$	$B(b)$	$B(c)$	$B(d)$	$B(e)$	$B(f)$

$F(a,a)$	$F(a,b)$	$F(a,c)$	$F(a,d)$	$F(a,e)$	$F(a,f)$
$F(b,a)$	$F(b,b)$	$F(b,c)$	$F(b,d)$	$F(b,e)$	$F(b,f)$
$F(c,a)$	$F(c,b)$	$F(c,c)$	$F(c,d)$	$F(c,e)$	$F(c,f)$
$F(d,a)$	$F(d,b)$	$F(d,c)$	$F(d,d)$	$F(d,e)$	$F(d,f)$
$F(e,a)$	$F(e,b)$	$F(e,c)$	$F(e,d)$	$F(e,e)$	$F(e,f)$
$F(f,a)$	$F(f,b)$	$F(f,c)$	$F(f,d)$	$F(f,e)$	$F(f,f)$

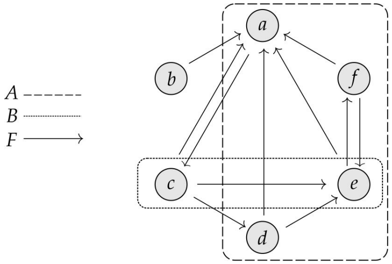
An **atomic truth assignment** (a valuation) is any function $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$.

$A(a)$	$A(b)$	$A(c)$	$A(d)$	$A(e)$	$A(f)$
$B(a)$	$B(b)$	$B(c)$	$B(d)$	$B(e)$	$B(f)$
$F(a,a)$	$F(a,b)$	$F(a,c)$	$F(a,d)$	$F(a,e)$	$F(a,f)$
$F(b,a)$	$F(b,b)$	$F(b,c)$	$F(b,d)$	$F(b,e)$	$F(b,f)$
$F(c,a)$	$F(c,b)$	$F(c,c)$	$F(c,d)$	$F(c,e)$	$F(c,f)$
$F(d,a)$	$F(d,b)$	$F(d,c)$	$F(d,d)$	$F(d,e)$	$F(d,f)$
$F(e,a)$	$F(e,b)$	$F(e,c)$	$F(e,d)$	$F(e,e)$	$F(e,f)$
$F(f,a)$	$F(f,b)$	$F(f,c)$	$F(f,d)$	$F(f,e)$	$F(f,f)$



Truth assignments

An **atomic truth assignment** (a valuation) is any function $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$.



How many truth assignments are there for $\{A, B, F\}(\{a, b, c, d, e, f\})$?

$2^{48} = 281\,474\,976\,710\,656$

$$(\#\{A, B, F\}(\{a, b, c, d, e, f\}) = 6^1 + 6^1 + 6^2 = 48)$$

Relational structures and Pure predicate language

An $\mathtt{L}$-**model** (structure, interpretation) with the universe $\mathbb{U}$ is a pair $(\mathbb{U}, \mathbf{M})$, where $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$ is a truth assignment.

Pure predicate language

non logical symbols (vocabulary):

- names for individuals; lower case letters, with or without numeric subscripts, from the beginning of the alphabet will be used to denote individual names: a, b, c, d, a_1, ... we always assume that the names exhausted the universe;
- names for predicates; upper case letters, with or without numeric subscripts will be used to denote predicates: A, B, C, ..., A_1, ...

logical symbols:

- variables; lower case letters, with or without numeric indices, from the end of the alphabet will be used for variables – x, y, z, x_1, ...
- connectives – not $\neg$, and $\wedge$, or $\vee$, if ... then $\Rightarrow$, iff $\Leftrightarrow$;
- quantifiers – there exists $\exists$, for all $\forall$

Formulas

Atomic formulas: `Predicate(` $\underbrace{\_\_\_\_, \ldots, \_\_\_\_}_{\text{places for individuals and variables}}$ `)`

E.g. $A(a)$, $A(x)$, $B(x)$, $F(x, y)$, $F(x_2, y_1)$, etc. $F(z_1, z_1)$

Formulas:

- Each atomic formula is a formula;
- If P and Q are formulas, then $\neg P$, $P \wedge Q$, $P \vee Q$, $P \Rightarrow Q$, $P \Leftrightarrow Q$ are formulas too.
- If x is a variable and P is a formula, then $\forall x P$ and $\exists x P$ are formulas too.

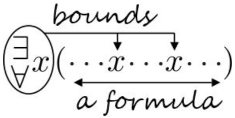
E.g.

$\neg F(a, b)$	$A(x) \wedge \neg F(a, x)$	$\forall y\, B(y)$
$\exists x\, F(a, x) \wedge \neg \exists x F(x, y)$	$B(x) \vee \forall x\, B(x)$	$\forall x \exists y\, F(x, y)$
$\forall x(\forall y F(x, y) \vee F(y, x))$	$A(z) \vee \exists y\, B(y) \Rightarrow \forall z(F(x, z) \wedge F(z, y))$	

etc.

Formulas

A variable may occur **free** or **bound** in a formula.



E.g.

$\neg F(a, b)$ $\quad$ $A(x) \wedge \neg F(a, x)$ $\quad$ $\forall y\, B(y)$

$\exists x\, F(a, x) \wedge \neg \exists x F(x, y)$ $\quad$ $B(x) \vee \forall x\, B(x)$ $\quad$ $\forall x \exists y\, F(x, y)$

$\forall x(\forall y F(x, y) \vee F(y, x))$ $\quad$ $A(z) \vee \exists y\, B(y) \Rightarrow \forall z(F(x, z) \wedge F(z, y))$

etc.

A formula F is denoted by $F(x_1, \ldots, x_k)$ when we want to emphasize that all free variables of F are among $x_1, \ldots, x_k$.

E.g. The formula F:

$$A(x) \vee B(y) \Rightarrow \forall z(F(y, z) \wedge F(z, y))$$

could be denoted $F(x, y)$, but also $F(x, y, z)$, $F(x, y, x_1, y_2)$ etc.

Sentences

A **sentence** is a formula with no free occurrence of a variable.

E.g.

$\neg F(a,b)$	~~$A(x) \wedge \neg F(a,x)$~~	$\forall y\, B(y)$
~~$\exists x\, F(a,x) \wedge \neg\exists x F(x,y)$~~	~~$B(x) \vee \forall x\, B(x)$~~	$\forall x \exists y\, F(x,y)$
~~$\forall x(\forall y F(x,y) \vee F(y,x))$~~	~~$A(z) \vee \exists y\, B(y) \Rightarrow \forall z(F(x,z) \wedge F(z,y))$~~	
$\forall x \forall y\, (F(x,y) \vee F(y,x))$	$A(b) \vee \exists y\, B(y) \Rightarrow \forall z(\forall x\, F(x,z) \wedge F(z,e))$	

etc.

$\mathrm{L}(\mathbb{U})$	$\subset$	$\mathrm{L}_{\omega 0}(\mathbb{U})$	$\subset$	$\mathrm{L}_{\omega\omega}(\mathbb{U})$
$P(c_{i_1}, \ldots, c_{i_k})$		$\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$		$\forall$, $\exists$, No free variables!

Truth assignments

Every atomic truth assignment $\mathbf{M} : \mathrm{L}(\mathbb{U}) \to \{0,1\}$ has a unique extension to a **truth assignment** that determines the truth-values of all the sentences $\mathbf{M} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to \{0,1\}$.

	$\neg$
0	1
1	0

$\wedge$	0	1
0	0	0
1	0	1

$\vee$	0	1
0	0	1
1	1	1

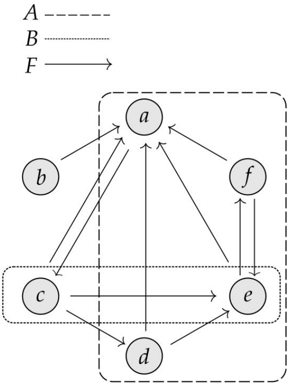
$\Rightarrow$	0	1
0	1	1
1	0	1

$\Leftrightarrow$	0	1
0	1	0
1	0	1

- $\mathbf{M}(\exists x\, F) = \max_{c \in \mathbb{U}} \mathbf{M}(F[x/c])$,
- $\mathbf{M}(\forall x\, F) = \min_{c \in \mathbb{U}} \mathbf{M}(F[x/c])$,

An assignment **M models** a sentence S ($\mathbf{M} \models S$) if $\mathbf{M}(S) = 1$.

Truth assignments



$\mathbf{M}(F(b,a) \wedge \neg F(a,b)) = 1$

$\quad \mathbf{M} \models F(b,a) \wedge \neg F(a,b)$

$\mathbf{M}(\forall x\, F(x,a)) = 0$

$\quad \mathbf{M} \not\models \forall x\, F(x,a)$

$\mathbf{M}(\exists x\, F(x,b)) = 0$

$\quad \mathbf{M} \not\models \exists x\, F(x,b)$

$\mathbf{M}(\forall x \forall y\, (F(x,y) \Rightarrow F(y,x))) = 0$

$\quad \mathbf{M} \not\models \forall x \forall y\, (F(x,y) \Rightarrow F(y,x))$

$\mathbf{M}(\exists x \exists y\, (F(x,y) \wedge F(y,x))) = 1$

$\quad \mathbf{M} \models \exists x \exists y\, (F(x,y) \wedge F(y,x))$

$\mathbf{M}(\exists y \forall x\, \neg F(x,y)) = 1$

$\quad \mathbf{M} \models \exists y \forall x\, \neg F(x,y)$

$\mathbf{M}(\forall x\, (A(x) \vee \exists y\, (F(x,y) \wedge A(y)))) = 1$

$\quad \mathbf{M} \models \forall x\, (A(x) \vee \exists y\, (F(x,y) \wedge A(y)))$

Fundamental logical notions

- A is **valid** ($\models A$) if A is true in all models.
- Q is a (semantical or logical) **consequence** of P ($P \models Q$) if every assignment that models P also models Q.
- P and Q are **equivalent** ($A \equiv B$) if both, Q is a consequence of P and P is a consequence of Q.

E.g:

- $\models A \Rightarrow (B \Rightarrow A \wedge B)$ $\quad \models \exists x \forall y\, F \Rightarrow \forall y \exists x\, F$ etc.
- $\exists x(A \wedge B) \models \exists x\, A \wedge \exists x\, B$ $\quad \forall x\, A \vee \forall x\, B \models \forall x(A \vee B)$ etc.
- $\forall x(A \wedge B) \equiv \forall x\, A \wedge \forall x\, B$ $\quad \exists x(A \vee B) \equiv \exists x\, A \vee \exists x\, B$
  $\neg(A \wedge B) \equiv \neg A \vee \neg B$ $\quad \neg(A \vee B) \equiv \neg A \wedge \neg B$ etc.

DNF

A **literal** is an atomic formula (also known as a positive literal) or its negation (a negative literal): $\pm\texttt{Predicate}(\underbrace{\_\_\_\_,\ldots,\_\_\_\_}_{\text{places for individuals and variables}})$

Positive literals: $A(a)$, $A(x)$, $B(x)$, $F(x, y)$, $F(x_2, y_1)$, etc.

Negative literals: $\neg A(a)$, $\neg A(x)$, $\neg B(x)$, $\neg F(x, y)$, $\neg F(x_2, y_1)$, etc.

An **atoms** is a conjunction of literals.

DNF theorem

Every quantifier free sentence F is equivalent to a disjunction of atoms:

$$\bigvee_{i=1}^{m}\left(\bigwedge_{j=1}^{n}\pm\texttt{Atomic}_{ij}\right)$$

A typical probabilistic model consists of:

- a sample space Ω,
- a family of events $\mathcal{B} \subseteq 2^{\Omega}$ (2^{Ω} is the power set of Ω, i.e. the set of all subsets of Ω) and
- a probability $\mu : \mathcal{B} \to [0, 1]$.

Rolling two dice:

- $\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\}$
- $\mathcal{B} = 2^{\Omega}$
  $E = \{(5,5)\}$ $F = \{(1,3), (3,1), (2,2)\}$
  $G = \{\text{at least one six is rolled}\} = \{(6,1), (6,1), (6,2), \ldots, (6,6)\}$ etc.
- $\mu(X) = \frac{\#X}{36}$, where $\#X$ is the number of elements in X
  $\mu(E) = \frac{1}{36}$ $\mu(F) = \frac{3}{36}$ $\mu(G) = \frac{11}{36}$

Probability spaces

$(\Omega, \mathcal{B})$ is a **measurable space**:

- $\mathcal{B}$ is a σ-algebra of subsets of Ω
  - $\Omega \in \mathcal{B}$
  - if $E \in \mathcal{B}$ then $E^{\complement} = \Omega \setminus E \in \mathcal{B}$
  - if $E_1, E_2 \in \mathcal{B}$ then $E_1 \cup E_2 \in \mathcal{B}$
  - from any sequence (E_n) of sets in $\mathcal{B}$, $\bigcup_{n=1}^{\infty} E_n \in \mathcal{B}$.

$(\Omega, \mathcal{B}, \mu)$ is a **probability space**:

- $\mu : \mathcal{B} \to [0, 1]$ is a σ-additive probability measure
  - $\mu(\Omega) = 1$
  - if $E_1 \cap E_2 = \emptyset$, then $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$
  - for any sequence $\langle E_n : n \geqslant 1 \rangle$ of pairwise disjoint sets, $\mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$.

Discrete probability spaces

$(\Omega, 2^\Omega, \mu)$ is a **discrete probability space** if

- Ω is countable and
- μ is defined by a function $m : \Omega \to [0,1]$ that assigns a real number from $[0,1]$ to each member of Ω in such a way that $\sum_{w\in\Omega} m(w) = 1$:

$$\mu(E) = \sum_{w\in E} m(w), E \subseteq \Omega.$$

The number $m(w)$ is called the **weight** of the point w.

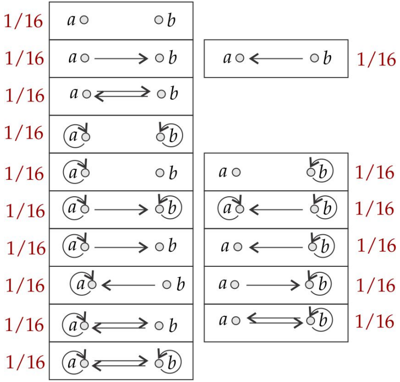
If Ω is finite and $m(w) = \frac{1}{\#\Omega}$, for all $w \in \omega$, then μ is called the **counting measure**.

Probability of sentences

$\mathtt{L} = \{F\}$, where F is a 2-placed predicate symbol; $\mathbb{U} = \{a, b\}$

What is the probability that a $\{F\}(\{a, b\})$-sentence is true?

E.g. $\mathbf{P}(\forall x \exists y\, F(x, y)) =?$



$$\mathbf{P}(\forall x \exists y\, F(x, y)) = \frac{9}{16} = 0.56$$

$$\mathbf{P}(\exists y \forall x\, F(x, y)) = \frac{7}{16} = 0.4375$$

$$\mathbf{P}(\exists z\, F(a, z)) = \frac{12}{16} = 0.75$$

$$\mathbf{P}(\exists x\, F(x, x) \vee \forall x\, \neg F(x, x)) = \frac{16}{16} = 1$$

$$\mathbf{P}(\exists x\, (F(x, x) \wedge \neg F(x, x))) = \frac{0}{16} = 0$$

Probability of sentences

Definition

A **probability** is a function $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ that satisfies:

- P1 if $\models A$, $\mathbf{P}(A) = 1$; any valid sentence is a certain (sure) sentence;
- P2 if $A \models \neg B$, then $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;
- P3 $\mathbf{P}(\exists x\, A) = \lim_{n\to\infty} \mathbf{P}(A(c_1) \vee \cdots \vee A(c_n))$, in that case that the universe is countable $\mathbb{U} = \{c_1, c_2, c_3, \ldots\}$.

(P2′) if $A \models B$, then $\mathbf{P}(B \Rightarrow A) = 1 + \mathbf{P}(A) - \mathbf{P}(B)$
(P2″) if $\models \neg(A \wedge B)$, then $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$
(P3) Gaifman's condition that could be omit, since ...

Probability of sentences

Definition

A **probability** is a function $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ that satisfies:

- P1 if $\models A$, $\mathbf{P}(A) = 1$; any valid sentence is a certain (sure) sentence;
- P2 if $A \models \neg B$, then $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;
- P3 $\mathbf{P}(\exists x\, A) = \lim_{n\to\infty} \mathbf{P}(A(c_1) \vee \cdots \vee A(c_n))$, in that case that the universe is countable $\mathbb{U} = \{c_1, c_2, c_3, \ldots\}$.

Proposition

Let $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ be a probability. Then for $A, B \in \mathrm{L}_{\omega\omega}(\mathbb{U})$.

1. $\mathbf{P}(\neg A) = 1 - \mathbf{P}(A)$
2. if $A \models B$, $\mathbf{P}(A) \leqslant \mathbf{P}(B)$
3. if $A \equiv B$, $\mathbf{P}(A) = \mathbf{P}(B)$
4. $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \wedge B)$

Convex sums of probabilities

Definition

A **probability** is a function $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ that satisfies:

- P1 if $\models A$, $\mathbf{P}(A) = 1$; any valid sentence is a certain (sure) sentence;
- P2 if $A \models \neg B$, then $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;
- P3 $\mathbf{P}(\exists x\, A) = \lim_{n\to\infty} \mathbf{P}(A(c_1) \vee \cdots \vee A(c_n))$, in that case that the universe is countable $\mathbb{U} = \{c_1, c_2, c_3, \ldots\}$.

Any truth assignment $\mathbf{M} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to \{0,1\}$ is a probability.

For a family of structures $\mathbf{M}_i$, $i \in I$, and numbers $a_i \geqslant 0$, $i \in I$ such that $\sum_i a_i = 1$,

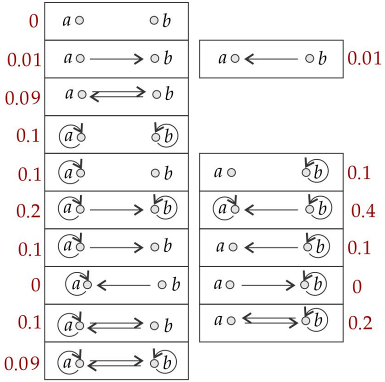
$$S \mapsto \sum_i a_i \mathbf{M}(S) \ \text{ is a probability.}$$

Every probability is a convex combination of some classical structures.

Probability of sentences

$\mathtt{L} = \{F\}$, where F is a 2-placed predicate symbol; $\mathbb{U} = \{a, b\}$

What is the probability that a $\{F\}(\{a, b\})$-sentence is true?



$$\mathbf{P}(\forall x \forall y\,(F(x,y) \Rightarrow F(y,x)))$$
$$= 0+0.09+0.1+0.1+0.1+0.1+0.2+0.09$$
$$= 0.78$$

$$\mathbf{P}(\neg\forall x \forall y\, F(x,y)) = 1 - 0.09 = 0.91$$

$$\mathbf{P}(\forall x \forall y\, \neg F(x,y)) = 0$$

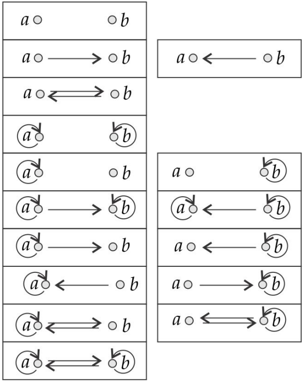
$$\mathbf{P}(\exists x\, F(x,x) \vee \forall x\, \neg F(x,x)) = \frac{16}{16} = 1$$

$$\mathbf{P}(\exists x\,(F(x,x) \wedge \neg F(x,x))) = \frac{0}{16} = 0$$

Probability of sentences

$\mathtt{L} = \{F\}$, $\mathbb{U} = \{a, b\}$; $\mathbf{P} : \{F\}_{\omega\omega}(\mathbb{U}) \to [0, 1]$

There is a bijection between all structures and so-called *complete atoms*.



$$
\begin{array}{ll}
A_1 & \neg F(a,a) \wedge \neg F(a,b) \wedge \neg F(b,a) \wedge \neg F(b,b) \\
A_2 & \neg F(a,a) \wedge F(a,b) \wedge \neg F(b,a) \wedge \neg F(b,b) \\
A_3 & \neg F(a,a) \wedge \neg F(a,b) \wedge F(b,a) \wedge \neg F(b,b) \\
\vdots & \\
A_{16} & F(a,a) \wedge F(a,b) \wedge F(b,a) \wedge F(b,b)
\end{array}
$$

$\mathbf{M}_i(S) = 1$ iff $A_i \models S$

$S \equiv \bigvee_{\mathbf{M}_i(S)=1} A_i$

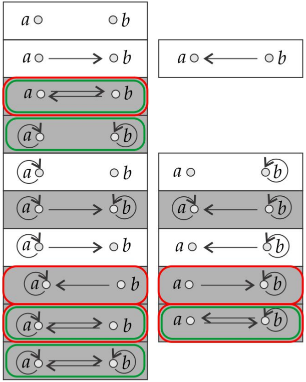
$$\mathbf{P}(S) = \sum_{\mathbf{M}_i(S)=1} \mathbf{P}(A_i) = \sum_{i=1}^{16} \mathbf{P}(A_i) \cdot \mathbf{M}_i(S)$$

Probability of sentences

Exercise

Find a probability $\mathbf{P} : \{F\}_{\omega\omega}(\{a, b\}) \to [0, 1]$, if it exists, such that:

$$\mathbf{P}(\forall x \exists y\, F(x, y)) = 1$$
$$\mathbf{P}(\forall x \forall y\, (F(x, y) \Rightarrow F(y, x))) = 0.9$$
$$\mathbf{P}(\exists x\, \neg F(x, x)) = 0.01$$



Representation theorem

If $\mathtt{L}$ or $\mathbb{U}$ are infinite, then $\mathtt{L}_{\omega\omega}(\mathbb{U})$ are infinite too.

- $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$ – all $\mathtt{L}$-structures over $\mathbb{U}$;
- $[S] = \{\mathbf{M} \in \mathcal{M}_{\mathtt{L}(\mathbb{U})} \mid \mathbf{M}(S) = 1\},\ S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$;
  $\mathbf{1}_{[S]} : \mathcal{M}_{\mathtt{L}(\mathbb{U})} \to \{0,1\}$, $\mathbf{1}_{[S]}(\mathbf{M}) \stackrel{\mathrm{def}}{=} \mathbf{M}(S)$
- $\mathcal{A}_{\mathtt{L}(\mathbb{U})} = \{[S] \mid S \in \mathtt{L}_{\omega 0}(\mathbb{U})\}$ – an algebra of subsets of $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$;
- $\mathcal{B}_{\mathtt{L}(\mathbb{U})}$ – an σ-algebra extending $\mathcal{A}_{\mathtt{L}(\mathbb{U})}$

Representation theorem

For any probability function $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ there is a probability measure $\mu_{\mathbf{P}}$ on $\mathcal{B}_{\mathtt{L}(\mathbb{U})}$ such that for any S:

$$\mathbf{P}(S) = \int_{\mathcal{M}_{\mathtt{L}(\mathbb{U})}} \mathbf{1}_{[S]}(\mathbf{M})\,\mathrm{d}\mu_{\mathbf{P}}(\mathbf{M})$$

In finite cases: $\mathbf{P}(S) = \sum_i \mathbf{1}_{[S]}(\mathbf{M}_i)\mu(\mathbf{M}_i) = \sum_i \mathbf{M}_i(S)\mu(\mathbf{M}_i)$

Extension theorems

Carathéodory's Extension Theorem

Let $\mathcal{A}$ be an algebra of subsets of Ω and $\mu_0 : \mathcal{A} \to [0,1]$ be a finitely-additive probability measure such that:

$$(*) \quad \mu(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i), \text{ for any sequence } \langle E_n : n \geqslant 1 \rangle \text{ of disjoint sets from } \mathcal{A} \text{ such that } \bigcup_{n=1}^{\infty} E_i \in \mathcal{A}.$$

Then there exists a unique measure $\mu : \mathcal{B} \to [0,1]$, on the σ-algebra $\mathcal{B}$ generated by $\mathcal{A}$ such that its restriction to $\mathcal{A}$ coincides with μ_0.

Gaifman's Extension Theorem

Assume that a function $\mathbf{P} : \mathrm{L}_{\omega 0}(\mathbb{U}) \to [0,1]$ satisfies (P1) and (P2). Then $\mathbf{P}$ has a unique extension to a probability function satisfying (P1), (P2), (P3) for all sentences $\mathrm{L}_{\omega\omega}(\mathbb{U})$.

Extension theorems

Gaifman's Extension Theorem

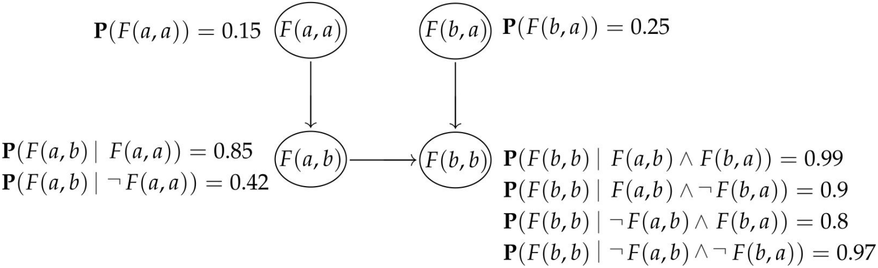
Assume that a function $\mathbf{P} : \mathrm{L}_{\omega 0}(\mathbb{U}) \to [0, 1]$ satisfies (P1) and (P2). Then $\mathbf{P}$ has a unique extension to a probability function satisfying (P1), (P2), (P3) for all sentences $\mathrm{L}_{\omega\omega}(\mathbb{U})$.

Corollary

Gaifman's and DNF theorems: to specify a probability function on $\mathrm{L}_{\omega\omega}(\mathbb{U})$ it is enough to define probability function on atoms (conjunctions of literals) in $\mathrm{L}_{\omega 0}(\mathbb{U})$.

Bayesian networks

Bayesian networks are closed acyclic graphs (DAG's) whose nodes represent assertions and edges represents some kind of conditional dependencies.



The Chain Rule:

$$\mathbf{P}(F(a,a) \wedge \neg F(a,b) \wedge \neg F(b,a) \wedge F(b,b))$$
$$= \mathbf{P}(F(a,a)) \times \mathbf{P}(\neg F(a,b) \mid F(a,a)) \times \mathbf{P}(F(b,a)) \times \mathbf{P}(F(b,b) \mid \neg F(a,b) \wedge \neg F(b,a))$$
$$= 0.15 \times (1 - 0.85) \times (1 - 0.25) \times 0.97 \approx 0.02$$

Probability structures

Definition

A classical structure is a pair $(\mathbb{U}, \mathbf{M})$, where $\mathbf{M}$ is a truth-assignment on $\mathrm{L}_{\omega\omega}(\mathbb{U})$.

Definition

A probability structure is a pair $(\mathbb{U}, \mathbf{P})$, where $\mathbf{P}$ is a probability on $\mathrm{L}_{\omega\omega}(\mathbb{U})$.

The representation theorem shows that a probability on a set of formulas, i.e., a probability structure can be viewed as a kind of model consisting of a family of classical models (called worlds) equipped with their 'weights'.

$$\mathbf{P}(S) = \sum_i \mathbf{world}_i(S) \times \mathtt{weight}(\mathbf{world}_i)$$

Probability structures

Definition

A probability structure is a pair $(\mathbb{U}, \mathbf{P})$, where $\mathbf{P}$ is a probability on $\mathtt{L}_{\omega\omega}(\mathbb{U})$.

Many constructions for probability models can be developed by analogy with ordinary model theory.

Indipendent union

$\mathbf{P}_i : \mathtt{L}_{i\omega\omega}(\mathbb{U}) \to [0,1]$, $i \in I$

$\mathtt{L} = \bigcup_{i \in I} \mathtt{L}_i$; $\mathtt{L}_{\omega 0}(\mathbb{U}) \ni A \equiv \bigwedge_{i \in I} A_i$, $A_i \in \mathtt{L}_{i\omega 0}(\mathbb{U})$

$$\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1],\ \mathbf{P}(A) \stackrel{\text{def}}{=} \prod_{i \in I} \mathbf{P}_i(A_i)$$

Ultraproduct

$\mathbf{P}_i : \mathtt{L}_{\omega\omega}(\mathbb{U}_i) \to [0,1]$, $i \in I$

$\mathbb{U} = \prod_{i \in I} \mathbb{U}_i$, $\mu : 2^I \to [0,1]$ is a probability

$$\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1],\ \mathbf{P}(S) \stackrel{\text{def}}{=} \sum_{i \in I} \mathbf{P}_i(S \mid i) \cdot \mu(\{i\})$$

1.3 Why ... ?

We have reason to hope that the results of probability logic may have useful applications to deductive logic, inductive logic and to probability theory. (D. Scott and P. Krauss, 1966)

Definition

Given a probability function $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$, and $C \in \mathrm{L}_{\omega\omega}(\mathbb{U})$ with $\mathbf{P}(S) > 0$, the **conditional probability** is a function $\mathbf{P}(\cdot \mid C) : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ (said $\mathbf{P}$ conditioned on C) defined by:

$$\mathbf{P}(S \mid C) = \frac{\mathbf{P}(S \wedge C)}{\mathbf{P}(C)}.$$

Sentences S_1 and S_2 are **independent** if $\mathbf{P}(S_1 \wedge S_2) = \mathbf{P}(S_1) \cdot \mathbf{P}(S_2)$.

A Basic System of Inductive Logic I/II

- The vocabulary contains only a finite number of unary predicate symbols $\mathtt{L} = \{F_1, F_2, \ldots, F_k\}$, $(k > 2)$;
- The universe is countable, $\mathbb{U} = \{c_1, c_2, \ldots\}$;
- A **sample** $\mathbb{S}$ is a finite set of individuals, $\mathbb{S} \subseteq \mathbb{U}$. A **sample description** $D_{\mathbb{S}}$ is a complete atom (a maximal consistent conjunction of literals) determined by atomic sentences from $\mathtt{L}(\mathbb{S})$.

Carnap looked for a probability $\mathbf{P} : \mathtt{L}_{\omega 0}(\mathbb{U}) \to [0, 1]$ ...

C1 [**Regularity**] $\mathbf{P}(A) > 0$ if A is not a contradiction

C2 [**Symmetry**] $\mathbf{P}(D_{\mathbb{S}})$ isn't changed by permuting individuals from $\mathbb{S}$.

C3 [**Instantial relevance**] $\mathbf{P}(F_i(c_k) \mid F_i(c_\ell)) > \mathbf{P}(F_i(c_k))$

C4 [λ-**condition**] if $\mathbb{S}$ doesn't involve c_k then $\mathbf{P}(F_i(c_k) \mid D_{\mathbb{S}})$ depends only on the number of individuals from $\mathbb{S}$ and the number of individual from $\mathbb{S}$ having the property F_i.

$\lambda - \gamma$ theorem

$\lambda - \gamma$ theorem

If **P** is a probability which satisfies C1-4 and $k > 2$, then there exist $\lambda > 0$ and $\gamma_1, \ldots, \gamma_k \in (0, 1)$ such that the probability that c_k has the property F_i, given the simple description $D_{\mathbb{S}}$ is given by the following equation:

$$\mathbf{P}(F_i(c_k) \mid D_{\mathbb{S}}) = \frac{n_i + \lambda\gamma_i}{n + \lambda},$$

where

- c_k is any individual constant not in the sample $\mathbb{S}$
- n is the number of individuals in $\mathbb{S}$
- n_i is the number of individuals from $\mathbb{S}$ having F_i.

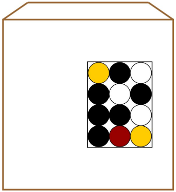
empirical factor vs. logical factor

$$\mathbf{P}(F_i(c_k) \mid D_{\mathbb{S}}) = \frac{n_i + \lambda\gamma_i}{n + \lambda} = \left(\frac{n}{n+\lambda}\right)\frac{n_i}{n} + \left(\frac{\lambda}{n+\lambda}\right)\gamma_i$$

γ_i – logical factor, i.e. a priori probability that something has F_i

$\frac{n_i}{n}$ – empirical factor

The larger λ is, the more weight is put on the logical factor, and the slower one learns from experience!



The colors of the balls are uniformly distributed in the box, but exceptions are possible.

$$\mathbf{P}(\texttt{Black}(c_k) \mid D_{\mathbb{S}}) = \left(\frac{12}{12+\lambda}\right)\frac{6}{12} + \left(\frac{\lambda}{12+\lambda}\right)\frac{1}{4}$$

Exercise

Solve the equation $\left(\frac{12}{12+\lambda}\right)\frac{6}{12} + \left(\frac{\lambda}{12+\lambda}\right)\frac{1}{4} = \textit{your esitamtion}.$

Pure Inductive Logic

Problem 1

Find a probability $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ (if it exists) such that $\mathbf{P}(S_i) = r_i$, $i \in I$.

Problem 2

Find a probability $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ (if it exists) that satisfies some general conditions (such as regularity, symmetry etc.)

- Studying additional properties of probabilities that are related to various principles and laws of indicative logic, and
- Searching for suitable representations of such probabilities.

J. Paris, A. Vencovská, *Six Problems in Pure Inductive Logic*, Journal of Philosophical Logic volume 48, pages 731–747, 2019

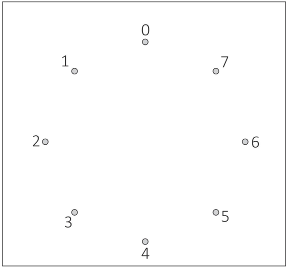
Table of Contents

Motivating example

$\mathbf{P} : \{R\}_{\omega\omega}(\mathbb{U}_n) \to [0,1]$, $\mathbb{U}_n = \{1, 2 \ldots, n\}$ $(n \geqslant 1,\ p \in [0,1])$

- $\mathbf{P}_p(R(i,j)) = p,\ i,j \in \mathbb{U}_n$
- $\mathbf{P}_p\left(\bigwedge_{(i,j)\in I} R(i,j) \wedge \bigwedge_{(m,\ell)\in J} \neg R(k,\ell)\right) = p^k(1-p)^{n^2-k}$, where I and J make a partition of the set $\{1,\ldots,n\} \times \{1,\ldots,n\}$, and $|I| = k$.

$(\mathbb{U}_8, \mathbf{P}_{0.5})$



	0	1	2	3	4	5	6	7
0								
1								
2								
3								
4								
5								
6								
7								

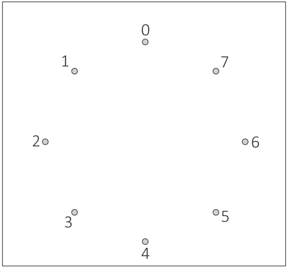
For every pair of nodes we flip a fair coin to decide whether the nodes should be adjacent or not.

Motivating example

$\mathbf{P} : \{R\}_{\omega\omega}(\mathbb{U}_n) \to [0,1]$, $\mathbb{U}_n = \{1, 2 \ldots, n\}$ $(n \geqslant 1,\ p \in [0,1])$

- $\mathbf{P}_p(R(i,j)) = p,\ i,j \in \mathbb{U}_n$
- $\mathbf{P}_p\left(\bigwedge_{(i,j)\in I} R(i,j) \wedge \bigwedge_{(m,\ell)\in J} \neg R(k,\ell)\right) = p^k(1-p)^{n^2-k}$, where I and J make a partition of the set $\{1,\ldots,n\} \times \{1,\ldots,n\}$, and $|I| = k$.

$(\mathbb{U}_8, \mathbf{P}_{0.5})$



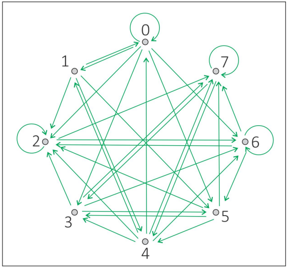
	0	1	2	3	4	5	6	7
0	H	H	H	H	T	H	H	T
1	H	T	H	T	H	H	T	T
2	T	T	H	T	T	T	H	T
3	T	T	H	T	T	H	H	H
4	H	H	H	H	T	T	H	T
5	T	T	T	H	H	T	T	H
6	T	T	H	T	T	H	H	H
7	T	T	T	H	H	T	T	H

Motivating example

$\mathbf{P} : \{R\}_{\omega\omega}(\mathbb{U}_n) \to [0,1]$, $\mathbb{U}_n = \{1, 2 \ldots, n\}$ $(n \geqslant 1,\ p \in [0,1])$

- $\mathbf{P}_p(R(i,j)) = p,\ i, j \in \mathbb{U}_n$
- $\mathbf{P}_p\left(\bigwedge_{(i,j)\in I} R(i,j) \wedge \bigwedge_{(m,\ell)\in J} \neg R(k,\ell)\right) = p^k(1-p)^{n^2-k}$, where I and J make a partition of the set $\{1,\ldots,n\} \times \{1,\ldots,n\}$, and $|I| = k$.

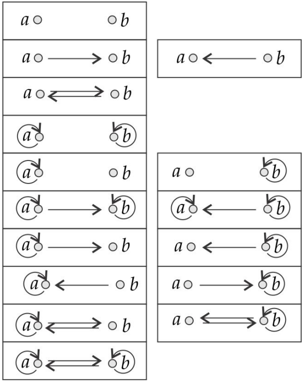
$(\mathbb{U}_8, \mathbf{P}_{0.5})$



	0	1	2	3	4	5	6	7
0	H	H	H	H	T	H	H	T
1	H	T	H	T	H	H	T	T
2	T	T	H	T	T	T	H	T
3	T	T	H	T	T	H	H	H
4	H	H	H	H	T	T	H	T
5	T	T	T	H	H	T	T	H
6	T	T	H	T	T	H	H	H
7	T	T	T	H	H	T	T	H

Motivating example

What is the probability that the sentence $\forall x \exists y\, R(x,y)$ is true in a randomly chosen two-member structure?



$\mathbf{P}(\forall x \exists y\, R(x,y)) \approx 0.56$

The assertion 'a sentence S with no individual names has a model with k elements' is expressible by a quantifier-free formula $F_k(x_1, \ldots, x_k)$, where variables are placeholders for individuals of $\mathbb{U}_k = \{c_1, \ldots, c_k\}$.

$F_2(x,y)$ for $\forall x \exists y\, R(x,y)$:

$$(R(x,x) \vee R(x,y)) \wedge (R(y,x) \vee R(y,y))$$

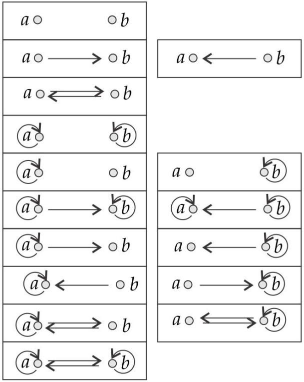
$F_3(x,y,z)$ for $\forall x \exists y\, R(x,y)$:

$$\begin{aligned} &(R(x,x) \vee R(x,y) \vee R(x,z)) \wedge \\ \wedge &(R(y,x) \vee R(y,y) \vee R(y,z)) \wedge \\ \wedge &(R(z,x) \vee R(z,y) \vee R(z,z)) \end{aligned}$$

etc.

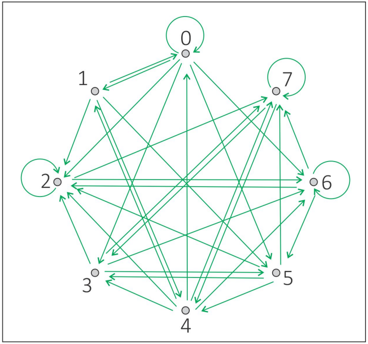
Motivating example

What is the probability that the sentence $\forall x \exists y\, R(x, y)$ is true in a randomly chosen two-member structure?



$\mathbf{P}(\forall x \exists y\, R(x, y)) \approx 0.56$

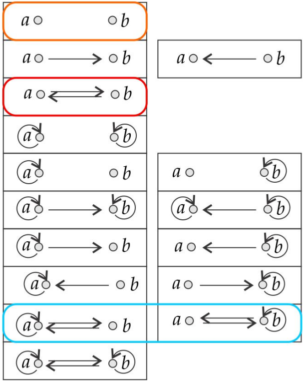
What is the probability that a randomly chosen pair from $(\mathbb{U}_8, \mathbf{P}_{0.5})$ satisfies $F_2(x, y)$: $(R(x, x) \vee R(x, y)) \wedge (R(y, x) \vee R(y, y))$?



$\mathbf{P}(F(x, y)) = ?$

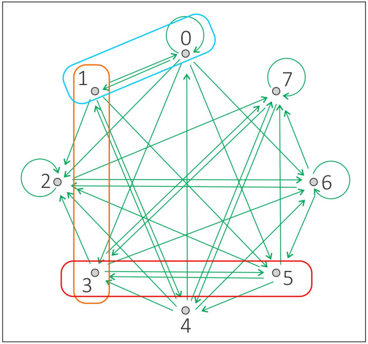
Motivating example

What is the probability that the sentence $\forall x \exists y\, R(x,y)$ is true in an arbitrarily chosen two-member structure?



$\mathbf{P}(\forall x \exists y\, R(x,y)) \approx 0.56$

What is the probability that a randomly chosen pair from $(\mathbb{U}_8, \mathbf{P}_{0.5})$ satisfies $F(x,y)$: $(R(x,x) \vee R(x,y)) \wedge (R(y,x) \vee R(y,y))$?



$\mathbf{P}(F(x,y)) = 0.56$

Motivating example

If H is much greater than n, then $(\mathbb{U}_H, \mathbf{P}_{0.5})$ contains (almost) all n-structures as its substructures. Moreover, the distribution of n-structures inside $(\mathbb{U}_H, \mathbf{P}_{0.5})$ is (almost) uniform.

Gaifman's example

$\mathbf{P} : \{R\}_{\omega\omega}(\mathbb{U}) \to [0,1]$, $\mathbb{U} = \{c_1, c_2, \ldots\}$, $p \in [0,1]$

- $\mathbf{P}_p(R(c_i, c_j)) = p$, $c_i, c_j \in \mathbb{U}$
- $\mathbf{P}\left(\bigwedge_{i=1}^{k} A_i \wedge \bigwedge_{i=k+1}^{n} \neg A_i\right) = p^k(1-p)^{n-k}$

$(\mathbb{U}, \mathbf{P})$ is a probability structure (symmetric in $\mathbb{U}$).

Logical independence implies statistical independence!

- If the atoms $A_1, A_2 \in \{R\}_{\omega 0}(\mathbb{U})$ are logically independent, in the sense that no atomic sentence is a part of both A_1 and A_2, then $\mathbf{P}(A_1 \wedge A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2)$.
- If $S_1, S_2 \in \{R\}_{\omega\omega}(\mathbb{U})$ and no individual constant occurs both in S_1 and S_2, then $\mathbf{P}(S_1 \wedge S_2) = \mathbf{P}(S_1) \cdot \mathbf{P}(S_2)$.
- If S does not contain individual constants, then $\mathbf{P}(S) = \mathbf{P}(S \wedge S) = \mathbf{P}(S)^2$, hence $\mathbf{P}(S)$ is either 0 or 1.

$*$ The same result holds if the starting vocabulary contains more predicate symbols, $\mathtt{L} = \{R_1, \ldots, R_k\}$, and a real number $p_i \in (0,1)$ is assigned to each predicate symbol R_i.

Random structures

$\mathbf{P} : \{R\}_{\omega\omega}(\mathbb{U}) \to [0,1]$, $\mathbb{U} = \{c_1, c_2, \ldots\}$, $p \in [0,1]$

- $\mathbf{P}_p(R(c_i, c_j)) = p$, $c_i, c_j \in \mathbb{U}$
- $\mathbf{P}\left(\bigwedge_{i=1}^{k} A_i \wedge \bigwedge_{i=k+1}^{n} \neg A_i\right) = p^k(1-p)^{n-k}$

$(\mathbb{U}, \mathbf{P})$ is a probability structure (symmetric in $\mathbb{U}$).
Logical independence implies statistical independence!

- If S does not contain individual constants, then $\mathbf{P}(S) = \mathbf{P}(S \wedge S) = \mathbf{P}(S)^2$, hence $\mathbf{P}(S)$ is either 0 or 1.

There is a classical $\mathtt{L}$-structure $\mathbf{R}$ such that for any sentence S:

$$\mathbf{R}(S) = 1 \text{ iff } \mathbf{P}(S) = 1$$

We call $\mathbf{R}$ the countable **random structure** over the vocabulary $\mathtt{L}$.

$*$ A random $\mathtt{L}$-structure contains in some sense all finite $\mathtt{L}$-structures as its substructures.

Languages $\mathtt{L}^k_{\omega 0}$, $\mathtt{L}^k_{\omega\omega}$, $k \geqslant 1$

- $\mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var})$ = the set of all formulas with predicates from $\mathtt{L}$, constant symbols for individuals from $\mathbb{U}$, and variables from $\mathtt{Var} = \{x_1, x_2, \ldots\}$;
- $\mathtt{L}^k_{\omega 0} = \mathtt{L}_{\omega 0}(\emptyset, \{x_1, \ldots, x_k\})$
  $\mathtt{L}^k_{\omega\omega} = \mathtt{L}_{\omega\omega}(\emptyset, \{x_1, \ldots, x_k\})$
  $\mathtt{L}^0_{\omega 0}$ = the propositional language which always include $\top$ and $\bot$
- $\mathtt{L}_{\omega\omega} = \bigcup_{k \geqslant 0} \mathtt{L}^k_{\omega\omega} = \mathtt{L}_{\omega\omega}(\emptyset, \mathtt{Var})$

Exercise

The vocabulary consists of a binary predicate symbol $<$, and consider only the structures in which the interpretation of $<$ is a total order.
A_n = 'there are at least n elements':

$$\exists x_1 \cdots \exists x_n \, (x_1 < x_2 \land x_2 < x_3 \land \cdots \land x_{n-1} < x_n).$$

Prove that, on total orders, A_n is equivalent to a sentence in $\{<\}^2_{\omega\omega}$.

Types (complete atoms over a tuple of variables)

Definition

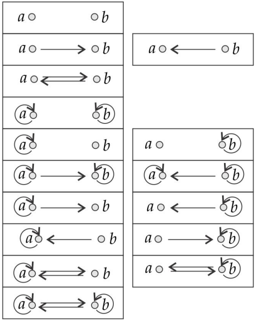
If $\overline{x} = (x_1, \ldots, x_k)$ is a sequence of distinct variables, then **a type** $T(\overline{x})$ in the variables $\overline{x}$ over $\mathtt{L}$ is the conjunction of all the formulas in a maximally consistent set of atomic formulas and negated atomic formulas in variables $\overline{x}$.

The logical constant $\top$ is the type in the empty tuple of variables.

- $T(x_1, \ldots, x_k)$ is a quantifier-free formula from $L^k_{\omega 0}$.
- There are only finitely many distinct types in the variables $x_1, \ldots, x_k$, if the vocabulary $\mathtt{L}$ is finite.

Types (complete atoms over a tuple of variables)

There are 16 types in variables $\{x_1, x_2\}$ that corresponds to all L-structures over two-element universe.



$$
\begin{array}{ll}
T_1(x_1,x_2) & \neg R(x_1,x_1) \wedge \neg R(x_1,x_2) \wedge \neg R(x_2,x_1) \wedge \neg R(x_2,x_2) \\
T_2(x_1,x_2) & \neg R(x_1,x_1) \wedge R(x_1,x_2) \wedge \neg R(x_2,x_1) \wedge \neg R(x_2,x_2) \\
T_3(x_1,x_2) & \neg R(x_1,x_1) \wedge \neg R(x_1,x_2) \wedge R(x_2,x_1) \wedge \neg R(x_2,x_2) \\
\vdots & \\
T_{16}(x_1,x_2) & R(x_1,x_1) \wedge R(x_1,x_2) \wedge R(x_2,x_1) \wedge R(x_2,x_2)
\end{array}
$$

Types

Definition

A type $T'(x_1, \ldots, x_{k-1}, x_k)$ extends a type $T(x_1, \ldots, x_{k-1})$ if every conjunct of $T(x_1, \ldots, x_{k-1})$ is also a conjunct of $T'(x_1, \ldots, x_{k-1}, x_k)$.

E.g. $T'(x_1, x_2, x_3)$:

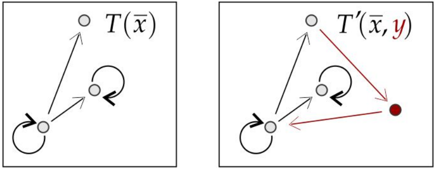
$$\neg R(x_1, x_1) \wedge R(x_1, x_2) \wedge R(x_1, x_3) \wedge$$
$$\wedge \neg R(x_2, x_1) \wedge R(x_2, x_2) \wedge \neg R(x_2, x_3) \wedge$$
$$\wedge R(x_3, x_1) \wedge \neg R(x_3, x_2) \wedge \neg R(x_3, x_3)$$

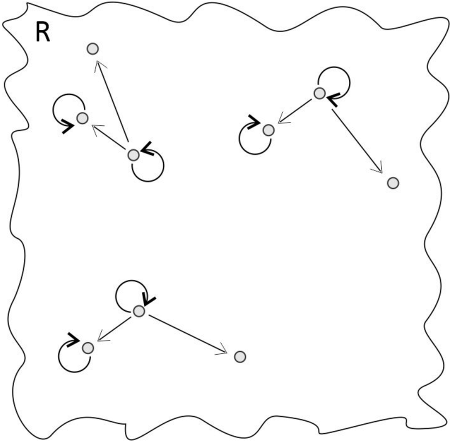
extends $T(x_1, x_2)$:

$$\neg R(x_1, x_1) \wedge R(x_1, x_2) \wedge$$
$$\wedge \neg R(x_2, x_1) \wedge R(x_2, x_2)$$

Extension axioms

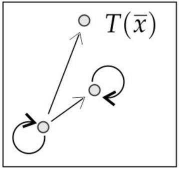
A 'huge' random structure contains all 'small' structures;
and even more it is true
any small substructure be extended to any other possible small substrukture.

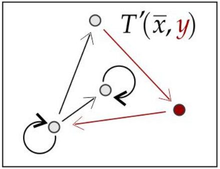


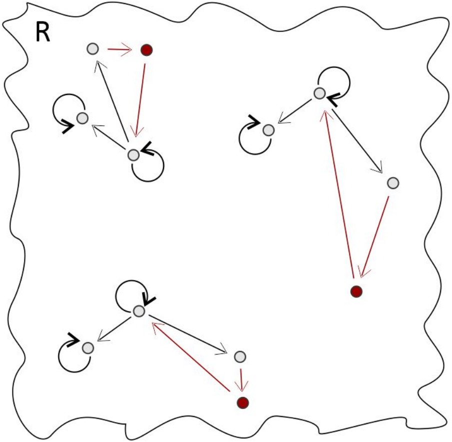


Extension axioms

A 'huge' random structure contains all 'small' structures;
and even more it is true
any small substructure be extended to any other possible small substrukture.







Extension axioms

Let $T'(\overline{x}, x_k)$ and $T(\overline{x})$ be a pair of types such that T' extends T.
$E_{T,T'}$ – the **extension axiom** $\forall \overline{x} \, (T(\overline{x}) \Rightarrow \exists x_k \; T'(\overline{x}, x_k))$

E_k – the conjunction of all extension axioms $E_{T,T'}$ with at most k variables; E_k is a *sentence* in $\mathrm{L}^k_{\omega\omega}$.

Exercise

Write down the sentence E_2, if the vocabulary contains single 2-placed predicate symbol R.

Exercise

Find the smallest model of E_2, if the vocabulary contains single 2-placed predicate symbol R.

Extension axioms

Theorem

Let k and m be two positive integers such that $m \leqslant k$. If $T(x_1, \dots, x_m)$ is type over the vocabulary $\mathtt{L}$ and and $F(x_1, \dots, x_m)$ is a formula of $\mathtt{L}^k_{\omega\omega}$ with free variables among $\overline{x} = (x_1, \dots, x_m)$, then exactly one of the following two statements holds:

1. $E_k \models \forall \overline{x}(T(\overline{x}) \Rightarrow F(\overline{x}))$
2. $E_k \models \forall \overline{x}(T(\overline{x}) \Rightarrow \neg F(\overline{x}))$

Corollary

If S is a sentence of $L^k_{\omega\omega}$, then either $E_k \models S$ or $E_k \models \neg S$.

Theorem also holds for infinitary languages $\mathtt{L}^k_{\infty\omega}$

Infinitary languages

$\mathtt{L}_{\infty\omega}(\mathbb{U}, \mathtt{Var})$ is an extension of $\mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var})$, defined by the usual formation rules for pure predicate languages, and the following additional rule:

- If Γ is a set of formulas, then $\bigwedge \Gamma$ and $\bigvee \Gamma$ are also formulas.

The semantics of infinitary formulas is a direct extension of the semantics of pure predicate languages. Given an $\mathtt{L}$-structure $(\mathbb{U}, \mathbf{M})$:

- $\mathbf{M}(\bigvee \Gamma) = \max_{\gamma \in \Gamma} \mathbf{M}(\gamma)$
- $\mathbf{M}(\bigwedge \Gamma) = \min_{\gamma \in \Gamma} \mathbf{M}(\gamma)$

Exercise

On total orders, express in $\{<\}^2_{\infty\omega}$ the properties:

- there are an even number of elements
- the cardinality of the total order is a prime number
- the universe is finite

Asymptotic probability

The extension axioms are relevant to the study of probabilities on finite structures. (R. Fagin, *Probabilities on Finite Models*, 1976)

Given a finite set $\mathtt{L}$ of predicate symbols, Fagin considered the sequence of *finite probability spaces with counting probability measure*:

- $\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}$ = all $\mathtt{L}$-structures over $\mathbb{U}_n = \{c_1, \ldots, c_n\}$, $n \geqslant 1$;
- $\mathbf{P}_n : \mathrm{L}_{\omega\omega}(\emptyset) \to [0, 1]$
  $\mathbf{P}_n(S) = \frac{\#[S]}{\#\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}} = \frac{\#\{\mathbf{M} \in \mathcal{M}_{\mathtt{L}(\mathbb{U}_n)} | \mathbf{M}(S)=1\}}{\#\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}}$, $S \in \mathrm{L}_{\omega\omega}(\emptyset)$

$\mathbf{P}(S) = \lim\limits_{n\to\infty} \mathbf{P}_n(S)$ – the asymptotic probability

Exercise

Find (estimate):

1) $\mathbf{P}_{10}(E_2)$ 2) $\mathbf{P}_k(E_2)$, $k > 10$ 3) $\lim\limits_{k\to\infty} \mathbf{P}_k(E_2)$

4) $\mathbf{P}_{100}(E_3)$ 5) $\mathbf{P}_k(E_3)$, $k > 100$ 6) $\lim\limits_{k\to\infty} \mathbf{P}_k(E_3)$

0-1 low

Proposition

The asymptotic probability of all extension axioms is equal to 1, that is $\mathbf{P}(E_k) = 1$, for every $k \geqslant 1$.

Corollary

If S is a sentence of $L^k_{\omega\omega}$, then either $E_k \models S$ or $E_k \models \neg S$.

0-1 low

If S is a sentence in $\mathrm{L}_{\omega\omega}(\emptyset)$, then the asymptotic probability $\mathbf{P}(S)$ exists and is equal either 0 or 1.

0-1 low

Glebskii *et al.* (*Range and degree of realizability of formulas in the restricted predicate calculus*, 1969) proved the 0-1 low, using a different approach – a certain quantifier elimination method.

Proposition

Let k be a positive integer and let $F(x_1, \ldots, x_m)$ be a formula of $\mathrm{L}^k_{\omega\omega}$ $(m \leqslant k)$. Then there is a quantifier-free formula $B(x_1, \ldots, x_m)$ of $\mathrm{L}^k_{\omega 0}$ such that:

$$E_k \models \forall \overline{x}\, (F(\overline{x}) \Leftrightarrow B(\overline{x})).$$

Taking $\overline{x}$ to be empty, the previous theorem says that each sentence with k variables collapses to $\top$ or $\bot$ almost everywhere, leading to the 0-1 law.

almost everywhere

Proposition

Let k be a positive integer and let $F(x_1, \dots, x_m)$ be a formula of $\mathrm{L}^k_{\omega\omega}$ $(m \leqslant k)$. Then there is a quantifier-free formula $B(x_1, \dots, x_m)$ of $\mathrm{L}^k_{\omega 0}$ such that:

$$E_k \models \forall \overline{x}\,(F(\overline{x}) \Leftrightarrow B(\overline{x})).$$

... almost everywhere variants of important theorems:

- $\mathrm{L}^k_{\omega\omega} \leq_{w.a.e.} \mathrm{L}^k_{\omega 0}$.
  $\mathrm{L}^k_{\omega\omega}$ admits almost everywhere quantifier elimination.
- We can effectively deside whether a first-order sentence is valid in almost all models.
  (By Trakhtenbrot's Theorem we cannot effectively decide whether a first-order sentence is valid in all finite models!)

⋮

Probabilities of formulas

counting small models satisfying a sentence
vs.
counting tuples in a 'huge' models satisfying an open formula

Definition

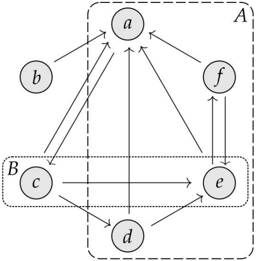
A **probability** is a function $\mathbf{P} : \mathrm{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var}) \to [0, 1]$ that satisfies the properties

P1 if $\models A$, $\mathbf{P}(A) = 1$; any valid formula is a certain (sure) formula;

P2 if $A \models \neg B$, then $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;

P3 $\mathbf{P}(\exists x\, A) = \lim_{n\to\infty} \mathbf{P}(A(c_1) \vee \cdots \vee A(c_n))$, in that case that the universe is countable $\mathbb{U} = \{c_1, c_2, c_3, \ldots\}$.

Probabilities of formulas



$\mathbf{P}_n$ are counting measures

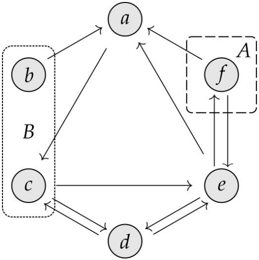
$\mathbf{P}_1(A(x)) = \frac{4}{6}$

$\mathbf{P}_2(F(x,y)) = \frac{11}{36}$

$\mathbf{P}_1(\exists x\, F(x,y)) = \frac{5}{6}$

$\mathbf{P}_2(F(x,y) \wedge F(y,x)) = \frac{4}{36}$

$\mathbf{P}_3(B(x) \wedge B(y) \wedge B(z)) = \frac{8}{216}$



$\mathbf{P}'_1 = \begin{pmatrix} a & b & c & d & e & f \\ 0.1 & 0.1 & 0.3 & 0.4 & 0.1 & 0 \end{pmatrix}$

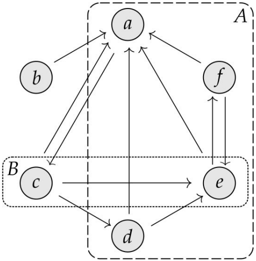
$\mathbf{P}'_n(i_1, \cdots, i_n) = \mathbf{P}'_1(i_1) \cdots \mathbf{P}'_1(i_n)$

$\mathbf{P}'_1(A(x)) = 0$

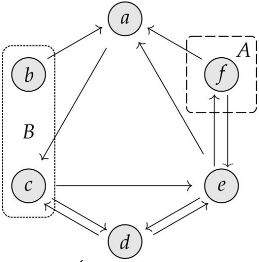
$\mathbf{P}'_1(B(x)) = 0.4$

$\mathbf{P}'_1(\exists x\, F(x,y)) = 0.9$

$\mathbf{P}'_2(F(x,y) \wedge \neg F(y,x)) = 0.08$



$\mathbf{P}_n$ are counting measures



$$\mathbf{P}_1' = \begin{pmatrix} a & b & c & d & e & f \\ 0.1 & 0.1 & 0.3 & 0.4 & 0.1 & 0 \end{pmatrix}$$

$$\mathbf{P}_n'(i_1, \cdots, i_n) = \mathbf{P}_1'(i_1) \cdots \mathbf{P}_1'(i_n)$$

$$\mu : \begin{pmatrix} \mathbf{M}_{\mathrm{left}} & \mathbf{M}_{\mathrm{right}} \\ 75\% & 25\% \end{pmatrix}$$

$$\mathbf{P}_n^{\mathrm{mix}}(F) = 0.75 \cdot \mathbf{P}_n(F) + 0.25 \cdot \mathbf{P}_n'(F)$$

Representation theorem

A probability on the set of formulas may be understood as a result of two consecutive drawings:

- First a structure $\mathbf{M}_i$ will be chosen at random from among all structures of a given class $\{\mathbf{M}_i \mid i \in I\}$ following a given probability measure μ on I; and then
- having the so obtained structure $\mathbf{M}_i$, a sequence of elements is selected from it, following probabilities $\mathbf{P}^i_n$, $n \geqslant 1$.

Graded probability structures

- $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$ is a classical structure.
- $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ is a probability structure.
- $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var}) \to [0,1]$ could be an even more general kind of structure.

A **graded probability structures** $(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$ is a kind of multidimensional probability space with a probability for each dimension, where:

- $\mathbf{M}$ is a classical structure, $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$
- each $\mathbf{P}_n$ is a probability on the set of formulas with at n-free variables, $\mathbf{P}_n : \mathtt{L}_{\omega\omega}(\mathbb{U}, \{x_1, \ldots, x_n\}) \to [0,1]$ and $\mathbf{P}_n$, $n \geqslant 1$ are related by some additional requirements (Fubini's properties, etc.) which we omit.

Values of terms

R. Fagin (1976)
Probabilities on Finite Models

$\downarrow$

Finite model theory

$L^k_{\infty\omega}$; $L^\omega_{\infty\omega}$; $L^k_{\omega_1\omega}$; $L^{II}_{\omega\omega}$; etc.

J. Keisler (1977)
Hyperpinite Model Theory

$\downarrow$

Model theory of stochastic processes

$L_{\omega P}$; $L_{\omega_1 P}$; $L_{\mathcal{C}\int}$, $L^-_{\mathcal{C}\int}$ etc.

$L_{\omega P}, L_{\omega_1 P}$

$L_{\omega P}, L_{\omega_1 P}$ – the classical quantifiers are replaced by probability quantifiers of the form $(P\overline{x} \geqslant r)$ and $(P\overline{x} > r)$, where $r \in [0, 1]$. The set of formulas is defined as usual, with the following rule for probability quantifiers:

- if F is a formula, then $(P\overline{x} \geqslant r)F$ and $(P\overline{x} > r)F$ are formulas too.

$(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1} \models (P\overline{x} > r)F(\overline{x})$ iff $\mathbf{P}_n(F(\overline{x})) > r$
$(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1} \models (P\overline{x} \geqslant r)F(\overline{x})$ iff $\mathbf{P}_n(F(\overline{x})) \geqslant r$

Formulas which are true in any graded structure:
- $(P\overline{x} \geqslant r)\, F(\overline{x}) \Rightarrow (P\overline{x} \geqslant s)\, F(\overline{x})$, $s < r$
- $(P\overline{x} > r)\, F(\overline{x}) \Rightarrow (P\overline{x} \geqslant r)\, F(\overline{x})$
- $(P\overline{x} \geqslant r)\, F(\overline{x}) \Rightarrow (P\overline{y} \geqslant r)\, F(\overline{y})$
- $(P\overline{x} \leqslant r)\, A(\overline{x}) \wedge (P\overline{x} \leqslant s)\, B(\overline{x}) \Rightarrow (P\overline{x} \leqslant r + s)\, (A(\overline{x}) \vee B(\overline{x}))$

Note also:
- If $S \Rightarrow F(\overline{x})$ is true in a graded structure, then $S \Rightarrow (P\overline{x} \geqslant 1)\, F(\overline{x})$ is also true in that structure.

Graded probability structures

Why do not we detail the definition and variety of graded probability structures.

Example.

An urn contains 1000 different balls, $\mathbb{U} = \{b_1, \dots, b_{1000}\}$.
There are 50% of red (R) balls; 30% of blue (B) balls, and 20% of green (G) balls: $\mathbf{P}(R(x)) = 0.5$, $\mathbf{P}(B(x)) = 0.3$, $\mathbf{P}(G(x)) = 0.2$.

We randomly choose 100 of them, one at a time, returning each ball to the urn before choosing the next one (the same ball could be drawn several times): $\overline{b} = (b_1, \dots, b_{100})$

$$\mathbf{p}_{\overline{b}}(R(x)) = \frac{R(b_1)+R(b_2)+\cdots+R(b_{100})}{100} \approx ?$$

$$\mathbf{p}_{\overline{b}}(B(x)) = \frac{B(b_1)+B(b_2)+\cdots+B(b_{100})}{100} \approx ?$$

$$\mathbf{p}_{\overline{b}}(G(x)) = \frac{G(b_1)+G(b_2)+\cdots+G(b_{100})}{100} \approx ?$$

$$\boxed{(\{b_1, \dots, b_{1000}\}, \cdots, \mathbf{P}, \dots) \approx (\{b_1, \dots, b_{100}\}, \cdots, \mathbf{p}_{\overline{b}}, \dots)}$$

Finite samples

Definition

Let $\mathcal{G} = (\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$ be a graded probability structure for $\mathtt{L}$, and let $\overline{a}^k = (a_1, \ldots, a_k) \in \mathbb{U}^k$ be a k-tuple of elements of $\mathbb{U}$. The **finite sample** $\mathcal{G}(\overline{a}^k)$ is the graded probability structure whose universe is $\{a_1, \ldots, a_k\}$, classical part is the substructure over $\{a_1, \ldots, a_k\} \subseteq \mathbf{M}$, and probabilities $\mathbf{p}_n$ are given by:

$$\mathbf{p}_1(F(x)) = \frac{\#\{m \leqslant k \mid \mathbf{M}(F(a_m)) = 1\}}{k} = \frac{F(a_1) + \cdots + F(a_k)}{k}$$

$$\mathbf{p}_2(F(x,y)) = \frac{\sum_{1 \leqslant i,j \leqslant k} F(a_i, a_j)}{k^2}$$

$$\mathbf{p}_3(F(x,y,z)) = \frac{\sum_{1 \leqslant i,j,\ell \leqslant k} F(a_i, a_j, a_\ell)}{k^3} \text{ etc.}$$

Weak Law of Large numbers for $\mathrm{L}_{\omega_1 \mathrm{P}}$

Proposition

Let $\mathcal{G} = (\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$ be a graded probability structure for L that satisfies $(\mathrm{P}x > r)\, F(x)$, where $F(x)$ is a quantifier-free formula. Then

$$\lim_{k \to \infty} \mathbf{P}_k \left\{ \overline{a}^k \in \mathbb{U}^k \mid \mathcal{G}(\overline{a}^k) \models (\mathrm{P}x > r)\, F(x) \right\} = 1.$$

In other words, for large enough k, in almost all samples $\mathcal{G}(\overline{a}^k)$, the formula $(\mathrm{P}x > r)\, F(x)$ is true.

Weak Law of Large numbers

Let $\mathcal{G} = (\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$ be a graded probability structure for L, satisfying $(\mathrm{P}\overline{x}_1 > r_1) \cdots (\mathrm{P}\overline{x}_n > r_n) B$, where B is a finite quantifier-free formula of L. Then

$$\lim_{k \to \infty} \mathbf{P}_k \left\{ \overline{a}^k \in \mathbb{U}^k \mid \mathcal{G}(\overline{a}^k) \models (\mathrm{P}\overline{x}_1 > r_1) \cdots (\mathrm{P}\overline{x}_n > r_n) B \right\} = 1.$$

$\mathrm{L}_{\mathcal{C}\mathsf{E}}$

$\mathrm{L}_{\mathcal{C}\mathsf{E}}$ uses the following symbols:

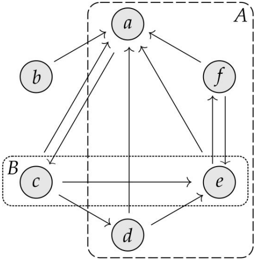
- individual variables: $x, y, z, x_1, \ldots$
- connectives: an *n*-ary connective C for each continuous real function $C : \mathbb{R}^n \to \mathbb{R}$ from $\mathcal{C}$
- thee quantifiers: E_x, $\sup_x$, $\inf_x$

The set of terms is the smallest set which satisfies the following rules:

- Every L-atomic formula is a term.
- If C is an *n*-ary connective and $T_1, \ldots, T_n$ are terms, then $\mathrm{C}(T_1, \ldots, T_n)$ is a term.
- If T is a term and x is a variable, $\mathsf{E}_x T$, $\sup_x T$, $\inf_x T$ are terms.

Free and bound variables are defined as usual, with quantifiers binding the variables. $T(x_1, \ldots, x_n)$ denotes a term with at most the free variables $\overline{x}$.

Values of terms



$\mathbf{P}_n$ – counting measures

The values of terms:

$\mathrm{E}_x A(x) = \frac{A(a)+A(b)+A(c)+A(d)+A(e)+A(f)}{6} = \frac{4}{6}$

$\sup_x (A(x) + B(x)) = 2$

$\mathrm{E}_x F(x, y)$

$= \frac{F(a,y)+F(b,y)+F(c,y)+F(d,y)+F(e,y)+F(f,y)}{6}$

$\mathrm{E}_x F(x, a) = \frac{5}{6}$

$\mathrm{E}_x F(x, b) = 0$

$\mathrm{E}_x F(x, c) = \frac{1}{6}$

$\max_y \mathrm{E}_x F(x, y) = \frac{5}{6}$ etc.

Real-valued structures

$\mathtt{L}_{\mathcal{C}\,\mathsf{E}}$ can be used to describe so-called real-valued structures.

- A (classical) two-valued $\mathtt{L}$-structure is of the form $(\mathbb{U}, \mathbf{M})$, where **M** is a truth assignment to atomic formulas, $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$.
- A **real-valued** $\mathtt{L}$**-structure** is a pair $(\mathbb{U}, \mathbf{M})$, where $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \mathbb{R}$, i.e. **M** assigns a real number to every atomic $\mathtt{L}$-sentence.

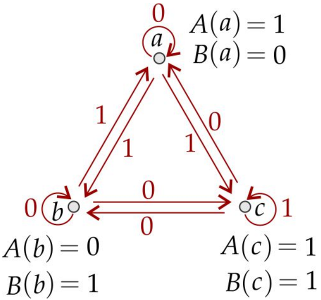
The rules for calculating the values of terms remain the same: $\mathsf{E}_x T$:

$$\mathbf{M}(\mathsf{E}_x\, T) = \int_{\mathbb{U}} T[x/c]\mathrm{d}\mu(c)$$

In discrete case: $\mathbf{M}(\mathsf{E}_x\, T) = \sum_{c \in \mathbb{U}} T[x/c] \cdot \mu(\{c\})$

$\{0,1\}$-valued structures

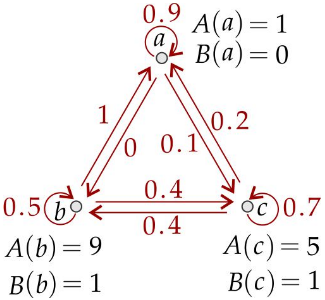
$\mathrm{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$, $\mathbb{U} = \{a, b, c\}$, $\mathbf{M} : \mathrm{L}(\mathbb{U}) \to \{0,1\}$



$$\begin{array}{lll}
\mathbf{M}(A(a)) = 1 & \mathbf{M}(A(b)) = 0 & \mathbf{M}(A(c)) = 1 \\
\mathbf{M}(B(a)) = 0 & \mathbf{M}(B(b)) = 1 & \mathbf{M}(B(c)) = 1 \\
\mathbf{M}(F(a,a)) = 0 & \mathbf{M}(F(a,b)) = 1 & \mathbf{M}(F(a,c)) = 1 \\
\mathbf{M}(F(b,a)) = 1 & \mathbf{M}(F(b,b)) = 0 & \mathbf{M}(F(b,c)) = 0 \\
\mathbf{M}(F(c,a)) = 0 & \mathbf{M}(F(c,b)) = 0 & \mathbf{M}(F(c,c)) = 1
\end{array}$$

$\mathbb{R}$-valued structures

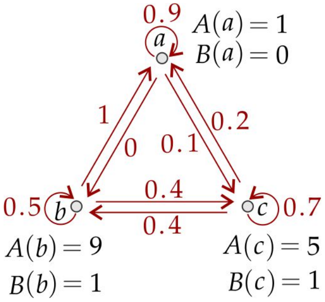
$\mathrm{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$, $\mathbb{U} = \{a, b, c\}$, $\mathbf{M} : \mathrm{L}(\mathbb{U}) \to [0,1]$



$\mathbf{M}(A(a)) = 1$	$\mathbf{M}(A(b)) = 9$	$\mathbf{M}(A(c)) = 5$
$\mathbf{M}(B(a)) = 0$	$\mathbf{M}(B(b)) = 1$	$\mathbf{M}(B(c)) = 1$
$\mathbf{M}(F(a,a)) = 0.9$	$\mathbf{M}(F(a,b)) = 0$	$\mathbf{M}(F(a,c)) = 0.1$
$\mathbf{M}(F(b,a)) = 1$	$\mathbf{M}(F(b,b)) = 0.5$	$\mathbf{M}(F(b,c)) = 0.4$
$\mathbf{M}(F(c,a)) = 0.2$	$\mathbf{M}(F(c,b)) = 0.4$	$\mathbf{M}(F(c,c)) = 0.4$

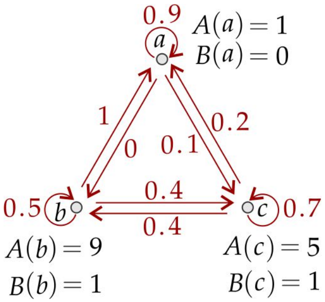
Real-valued structures

$\mathtt{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$, $\mathbb{U} = \{a, b, c\}$, $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to [0, 1]$



- $A(i)$ represents a quantitative characteristic (weight, height, temperature, IQ etc.) of an individual i;
- $B(i)$ represents a qualitative characteristic (a classical yes/no, i.e 1/0 property) of an individual i;
- $F(i, j)$ represents a degree of 'confidence' of i in j.
- $\mathbf{P}_n$ is the counting probability measures on $\mathbb{U}^n$, for all $n \geqslant 1$.

Real-valued structures



$$\mathsf{E}_x A(x) = \frac{A(a)+A(b)+A(c)}{3} = \frac{15}{3} = 5$$

$$\max_x (A(x) + B(x)) = 10$$

$$\mathsf{E}_x F(x,a) = \frac{F(a,a)+F(b,a)+F(c,a)}{3} = \frac{0.9+1+0.2}{3} = 0.7$$

$$\mathsf{E}_x F(x,b) = \frac{F(a,b)+F(b,b)+F(c,b)}{3} = 0.3$$

$$\mathsf{E}_x F(x,c) = \frac{F(a,c)+F(b,c)+F(c,c)}{3} \approx 0.4$$

$$\max_y \mathsf{E}_x F(x,y) = 0.7$$

$$\mathsf{E}_y \mathsf{E}_x F(x,y) = \frac{0.7+0.3+0.4}{3} = \frac{1.4}{3} \approx 0.47 \text{ etc.}$$

Summary

- ✔ Probabilities on the set of all L-formulas led to graded probabilistic L-structures – classical L-structures $(\mathbb{U}, \mathbf{M})$ extended by a sequence of probabilities $\mathbf{P}_n$, $n \geqslant 1$, for each dimension $\mathbb{U}^n$.
- ✔ We have roughly sketched several languages suitable for describing such structures: $\mathrm{L}_{\omega\mathrm{P}}$, $\mathrm{L}_{\omega_1\mathrm{P}}$, etc.
- ✔ Any graded structure can be well approximated by a hyperfinite structure, so that they satisfy 'almost' the same sentence from $\mathrm{L}_{\omega_1\mathrm{P}}$.
- ? One of the most important tasks of Mathematical logic is to characterize the set of all formulas that are true for all structures of a given type. It is conventional to begin the study of new logic by proving a completeness theorem.

Table of Contents

Motivating example

First-order language is the most common language for talking about relational structures. However, there are many alternative languages . . .

Modal languages provide an internal, local perspective on relational structure. (P. Blackburn, M. de Rijke, Y. Venema, *Modal logic*, Cambridge University Press, 2010)

Motivating example

Modal languages provide an internal, local perspective on relational structure.

The basic modal language uses symbols:

- a countably many propositional letters p, q, r etc.,
- the classical connectives $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$,
- and a unary modal operator $\Diamond$ ('diamond'; *possibly*).

The well-defined modal formulas are given by the following rules:

- a propositional letter is a formula;
- if A is a formula, then $\neg A$ is a formula, too
- if A_1 and A_2 are formulas, then $A_1 \wedge A_2$, $A_1 \vee A_2$, $A_1 \Rightarrow A_2$, $A_1 \Leftrightarrow A_2$ are formulas
- if A is a formula, then $\Diamond A$ is a formula.

Motivating example

A model for this language is a triple $(\mathbb{W}, R, [\cdot])$, where
• $\mathbb{W}$ is a non-empty set (of words);
• R is a binary relation on $\mathbb{W}$;
• $[\cdot]$ is a function (a valuation) assigning to each propositional letter p a subset $[p] \subseteq \mathbb{W}$.
Such an model can be viewed as a relational structure consisting of a domain, a single relation binary relation, and the unary relation given by $[\cdot]$:

$$\mathbf{W} = (\mathbb{W}, R, [p], [q], \cdots)$$

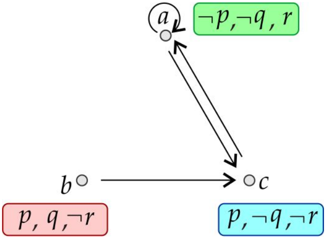
A formula F is satisfied (or true) in $\mathbf{W}$ at $w \in \mathbb{W}$:

- $\mathbf{W}, w \models p$ iff $w \in [p]$
- $\mathbf{W}, w \models \neg F$ iff $\mathbf{W}, w \not\models \neg F$
- $\mathbf{W}, w \models F_1 \wedge F_2$ iff $\mathbf{W}, w \models F_1$ and $\mathbf{W}, w \models F_2$ etc.
- $\mathbf{W}, w \models \Diamond F$ iff for some $v \in \mathbb{W}$, with $R(w, v)$, $\mathbf{W}, v \models F$.

Motivating example

...

- $\mathbf{W}, w \models \Diamond F$ iff for some $v \in \mathbb{W}$, with $R(w, v)$, $\mathbf{W}, v \models F$
  iff from w there is an 1-arrow-long path to a world where F is true.



$$\mathbf{W}, w \models \underbrace{\Diamond \cdots \Diamond}_{k \text{ times}} F$$

iff

from w there is a k-arrows-long path to a world where F is true.

$\mathbf{W}, a \models p \Rightarrow q$	$\mathbf{W}, b \models p \Rightarrow q$	$\mathbf{W}, c \not\models p \Rightarrow q$
$\mathbf{W}, a \models \Diamond(p \Rightarrow q)$	$\mathbf{W}, b \not\models \Diamond(p \Rightarrow q)$	$\mathbf{W}, c \models \Diamond(p \Rightarrow q)$
$\mathbf{W}, a \models \Diamond\Diamond(p \Rightarrow q)$	$\mathbf{W}, b \models \Diamond\Diamond(p \Rightarrow q)$	$\mathbf{W}, c \models \Diamond\Diamond(p \Rightarrow q)$

etc.

Markov process

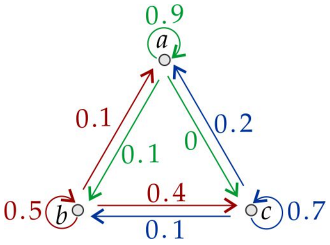
There are a wide range of probabilistic structures that come from different, mostly practical areas such as theoretical computer science, artificial intelligence, economics, game theory and so on.

- CAUSAL REASONING J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, 1988.
- LEARNING FROM DATA R. E. Neapolitan, *Probabilistic Reasoning in Expert Systems*, Wiley, New York, 1990.
- MULTI-AGENT SYSTEMS R. Fagin, J. Y. Halpern, Y. Moses, M. Y. Vardi, *Reasoning About Knowledge*, MIT Press, Cambridge, 2003.
- ROBOTICS S. Thrun, W. Burgard, D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, 2005.
- LOGIC PROGRAMMING K. Kersting, L. D. Raedt, *Bayesian logic programming: Theory and tool*, in Getoor, L. and Taskar, B., editors, Introduction to Statistical Relational Learning, MIT Press, Cambridge, 2007
- $\vdots$

Markov process

$(\mathbb{W}, 2^{\mathbb{W}})$, where $\mathbb{W} = \{a, b, c\}$ is a three-world set

$$\mu_a = \begin{pmatrix} a & b & c \\ 0.9 & 0.1 & 0 \end{pmatrix} \mu_b = \begin{pmatrix} a & b & c \\ 0.1 & 0.5 & 0.4 \end{pmatrix} \mu_c = \begin{pmatrix} a & b & c \\ 0.2 & 0.1 & 0.7 \end{pmatrix}$$



$$\mu_a(\{c\}) = 0 \qquad \mu_a(\{a, c\}) = 0.9$$
$$\mu_b(\{c\}) = 0.4 \qquad \mu_b(\{a, c\}) = 0.5$$
$$\mu_c(\{c\}) = 0.7 \qquad \mu_c(\{a, c\}) = 0.9$$

What is the probability that after a an outcome occurs after which c occurs with at least 40% of probability?

$$\{i \in \mathbb{W} \mid \mu_i(\{c\}) \geqslant 0.4\} = \{b, c\}$$
$$\mu_a\{i \mid \mu_i(\{c\}) \geqslant 0.4\} = \mu_a(\{b, c\}) = 0,1$$

Markov process

Definition

A **Markov process** consists of a family of probability space $(\mathbb{W}, \mathcal{F}, \mu_w)_{w \in \mathbb{W}}$ over the same measurable space $(\mathbb{W}, \mathcal{F})$, with an additional requirement related to 'the measurability of probabilistic assertions':

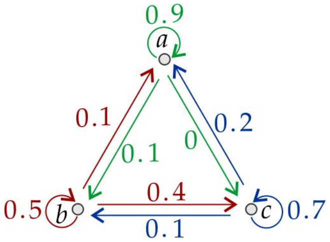
$(*)$ for all $X \in \mathcal{F}$ and $r \in [0,1]$, $\{w \in \mathbb{W} \mid \mu_w(X) \geqslant r\} \in \mathcal{F}$.

A discrete Markov process is a triple $(\mathbb{W}, 2^{\mathbb{W}}, \mu)$, where $\mathbb{W}$ is at most countable, and $\mu : \mathbb{W} \times \mathbb{W} \to [0,1]$ is a real-valued binary relation which satisfies the property:

$$\sum_{w' \in \mathbb{W}} \mu(w, w') = 1, \text{ for all } w \in \mathbb{W}.$$

In this case, $\mu_w(X) \stackrel{\text{def}}{=} \sum_{w' \in X} \mu(w, w')$.

$(\mathbb{W}, \mathcal{F}, \mu_w)_{w \in \mathbb{W}}$



1. individuals from $\mathbb{W}$ can be viewed as agents (with different beliefs and different degrees of trust in each other)
2. events from $\mathcal{F}$ can be considered as statements; $w \in X$ means: *the agent w believes that the statement X is true.*
3. $\mu_w(X)$ could be regarded as a degree of the w's opinion about the general belief in X.

$\mathrm{L_P}$

The symbols for $\mathrm{L_P}$ are:

- propositional letters from a countable vocabulary $\mathrm{L} = \{p, q, r, p_1, \ldots\}$;
- the logical constant 'true' $\top$:
- the classical connectives: $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$
- modal-like probabilistic operators $\mathrm{P}_{\geqslant r}$, for every $r \in [0,1] \cap \mathbb{Q}$, with the intended meaning *the probability is at least* r.

The set of formulas $\mathrm{L_P}$ is the smallest set such that:

- all prop. letters are formulas; $\top$ is a formula;
- if F is a formula, then $\neg F$ is a formula;
- if F_1 and F_2 are formulas, and $\star \in \{\wedge, \vee, \Rightarrow, \Leftrightarrow\}$, then $F_1 \star F_2$ is a formula;
- if F is a formula and $r \in [0,1] \cap \mathbb{Q}$, then $\mathrm{P}_{\geqslant r} F$ is a formula.

$\mathrm{L_P}$-models

Definition

An $\mathrm{L_P}$-structure is a quadruple $\mathbf{W} = (\mathbb{W}, \mathcal{F}, \mu_w, [\cdot])$, where $(\mathbb{W}, \mathcal{F}, \mu_w)$ is a Markov process, and $[\cdot] : \mathtt{L} \to \mathcal{F}$ is a function (valuation) which valued propositional letters by members form $\mathcal{F}$.

Each valuation $[\cdot] : \mathtt{V} \to \mathcal{F}$ can be extended inductively to all formulas:

- $[\top] = \mathbb{W}$, $[\neg F] = \mathbb{W} \setminus [F]$, $[F_1 \wedge F_2] = [F_1] \cap [F_2]$, etc.
- $[\mathrm{P}_{\geqslant r} F] = \{w \in \mathbb{W} \mid \mu(w, [F]) \geqslant r\}$

Thus, every $\mathrm{L_P}$-formula F defines the measurable set $[F] \in \mathcal{F}$.

Definition

A formula F is **true** (**false**) in $\mathbf{W}$ at w, denoted by $\mathbf{W}, w \models F$ ($\mathbf{W}, w \not\models F$), if $w \in [F]$ ($w \notin [F]$).

$\mathrm{L_P}$-models

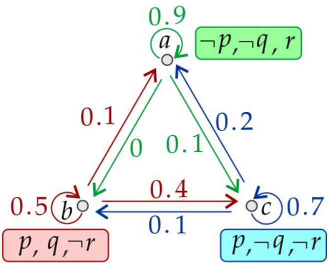
Definition

A formula F is **true** (**false**) in $\mathbf{W}$ at w, denoted by $\mathbf{W}, w \models F$ ($\mathbf{W}, w \not\models F$), if $w \in [F]$ ($w \notin [F]$).

- $\mathbf{W}, w \models \top$
- $\mathbf{W}, w \models \neg F$ iff $\mathbf{W}, w \not\models F$
- $\mathbf{W}, w \models F_1 \wedge F_2$ iff $\mathbf{W}, w \models F_1$ and $\mathbf{W}, w \models F_2$
- $\mathbf{W}, w \models F_1 \vee F_2$ iff $\mathbf{W}, w \models F_1$ or $\mathbf{W}, w \models F_2$
- $\mathbf{W}, w \models F_1 \Rightarrow F_2$ iff $\mathbf{W}, w \not\models F_1$ or $\mathbf{W}, w \models F_2$
- $\mathbf{W}, w \models \mathrm{P}_{\geqslant r} F$ iff $\mu(w, [F]) \geqslant r$

L_P-models

$(\mathbb{W}, \mathcal{F}, \mu_w, [\cdot])_{w \in \mathbb{W}}$



$[p] = \{b\}$
$[q] = \{b, c\}$
$[r] = \{a, c\} \dots$

$\mathbf{W}, a \models p \Rightarrow q$
$\mathbf{W}, b \models p \Rightarrow q$
$\mathbf{W}, c \not\models p \Rightarrow q$

$[p \Rightarrow q] = \{a, b\}$
$\mu_a(\{a, b\}) = 0.9$
$\mu_b(\{a, b\}) = 0.6$
$\mu_c(\{a, b\}) = 0.3$

$\mathbf{W}, a \models P_{\geqslant 0.8}(p \Rightarrow q)$
$\mathbf{W}, b \not\models P_{\geqslant 0.8}(p \Rightarrow q)$
$\mathbf{W}, c \not\models P_{\geqslant 0.8}(p \Rightarrow q)$

$[P_{\geqslant 0.8}(p \Rightarrow q)] = \{a\}$
$\mu_a(\{a\}) = 0.9 \ \mu_b(\{a\}) = 0.1 \ \mu_c(\{a\}) = 0.2$

$\mathbf{W}, a \models P_{\geqslant 0.15}P_{\geqslant 0.8}(p \Rightarrow q)$
$\mathbf{W}, b \not\models P_{\geqslant 0.15}P_{\geqslant 0.8}(p \Rightarrow q)$
$\mathbf{W}, c \models P_{\geqslant 0.15}P_{\geqslant 0.8}(p \Rightarrow q)$

$[P_{\geqslant 0.15}P_{\geqslant 0.8}(p \Rightarrow q)] = \{a, c\}$ etc.

Having a/no model

Definition

Let Γ be a set of L_P-formulas. $\mathbf{W}, w \models \Gamma$ iff for all $F \in \Gamma$, $\mathbf{W}, w \models F$. An L_P-structure $\mathbf{W}$ is a model of Γ if $\mathbf{W}, w \models \Gamma$, for all $w \in \mathbb{W}$.

Exercise

Show that any finite subset of

$$\Gamma = \{\neg \mathrm{P}_{\geqslant 1} \neg p\} \cup \{\neg P_{\geqslant \frac{1}{n}} p \ : \ n \geqslant 1\}$$

has a model, but Γ have no model.

Complete descriptions

From the model-theoretic point of view, the most important sets of formulas are those which are *complete descriptions* of worlds of an $\mathrm{L_P}$-structure $\mathbf{W}$:

$$\Gamma_{\mathbf{W},w} \stackrel{\text{def}}{=} \{F \in \mathrm{L_P} \mid \mathbf{W}, w \models F\}$$

Which sets of formulas could be (extended to) complete descriptions, i.e. could determine the worlds of a model?

We take the proof-theoretic approach, and try to discover:

- *valid formulas* belonging to each complete description;
- *closure properties* of complete descriptions: if a set of formulas Γ is a part of a complete description, which formulas must also belong to that complete description.

semantic consequence relation

Definition

A formula F is **valid**, denote by $\models F$, if it satisfied at every world of every model.

Definition

A formula F is a **semantic consequence** of Γ, denote by $\Gamma \models F$ iff

$$\mathbf{W}, w \models \Gamma \text{ implies } \mathbf{W}, w \models F,$$

for all worlds of every model $\mathbf{W}$.

$\emptyset \models F$ coincides with $\models F$.

Our main main objective is to axiomatize the relation $\models$ by constructing a **deducibility relation** $\vdash$ and showing: $\Gamma \models F$ iff $\Gamma \vdash F$.

Valid formulas

We distinguish three main groups of valid formulas.

- All $\mathrm{L_P}$-instances of the tautologies:

  $p \vee \neg p$; $p \wedge q \Rightarrow q$, $p \wedge (p \Rightarrow q) \Rightarrow q$, etc.

  $P_{\geqslant 0.3} p \vee \neg P_{\geqslant 0.3} p$; $P_{\geqslant 0.3} p \wedge P_{\geqslant 0.1} \neg q \Rightarrow P_{\geqslant 0.1} \neg q$, etc.

- the consequences of the ordering properties of the rational numbers:

  $P_{\geqslant 0.3} p \Rightarrow P_{\geqslant 0.2} p$; $P_{\geqslant 0.3}(p \wedge q) \Rightarrow P_{\geqslant 0.12}(p \wedge q)$ etc.

- the consequences of the basic properties of probability:

  $P_{\geqslant 1}(p \vee \neg p)$; $P_{\geqslant 0}(p \vee q)$, $P_{\geqslant 0.5} p \Rightarrow P_{\geqslant 0.5}(p \vee q)$,

  $P_{\geqslant 0.6} p \Rightarrow \neg P_{\geqslant 0.6} \neg p$, etc.

Valid formulas

Proposition

For all formulas A, B:

(A1) $\models \mathrm{P}_{\geqslant 0} A$

(A2) $\models \mathrm{P}_{\geqslant r} \top$, for all $r \in [0,1]_{\mathbb{Q}}$

(A3) $\models \mathrm{P}_{\geqslant r}(A \wedge B) \wedge \mathrm{P}_{\geqslant s}(A \wedge \neg B) \Rightarrow \mathrm{P}_{\geqslant r+s} A$, $r + s \leq 1$

(A4) $\models \neg \mathrm{P}_{\geqslant r}(A \wedge B) \wedge \neg \mathrm{P}_{\geqslant s}(A \wedge \neg B) \Rightarrow \neg \mathrm{P}_{\geqslant r+s} A$, $r + s \leq 1$

(A5) $\models \mathrm{P}_{\geqslant r} A \Rightarrow \neg \mathrm{P}_{\geqslant s} \neg A$, $r + s > 1$

Closure properties

Proposition

For all formulas A, B,

- $A \Leftrightarrow B \models P_{\geqslant r}A \Leftrightarrow P_{\geqslant r}B$, for all $r \in [0,1] \cap \mathbb{Q}$
- $\{P_{\geqslant t}A \mid t < r\} \models P_{\geqslant r}A$, for all $r \in (0,1] \cap \mathbb{Q}$
- $P_{\geqslant t_1}A, \ldots, P_{\geqslant t_k}A \not\models P_{\geqslant r}A$, for every choice of finitely many rationals $t_1, \ldots, t_k < r$.

$\mathbf{Ax}(\mathrm{L_P})$

(A0) every $\mathrm{L_P}$-instance of a tautology

(A1) $\mathsf{P}_{\geqslant 0}A$

(A2) $\mathsf{P}_{\geqslant r}\top$, za svako $r \in [0,1]_{\mathbb{Q}}$

(A3) $\mathsf{P}_{\geqslant r}(A \wedge B) \wedge \mathsf{P}_{\geqslant s}(A \wedge \neg B) \Rightarrow \mathsf{P}_{\geqslant r+s}A$, $r + s \leq 1$

(A4) $\neg\mathsf{P}_{\geqslant r}(A \wedge B) \wedge \neg\mathsf{P}_{\geqslant s}(A \wedge \neg B) \Rightarrow \neg\mathsf{P}_{\geqslant r+s}A$, $r + s \leq 1$

(A5) $\mathsf{P}_{\geqslant r}A \Rightarrow \neg\mathsf{P}_{\geqslant s}\neg A$, $r + s > 1$

(MP) $$\frac{A \quad A \Rightarrow B}{B}$$

(EQ_r) $$\frac{A \Rightarrow (B \Leftrightarrow C)}{A \Rightarrow (\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C)}$$

(A_r) $$\frac{A \Rightarrow \mathsf{P}_{\geqslant t}B \quad t < r}{A \Rightarrow \mathsf{P}_{\geqslant r}B}$$

Syntactical consequence relation

Definition

A formula F can be deduced (derived, inferred etc.) from Γ ($\Gamma \vdash F$) if there is a **derivation** of the form:

$$F_1, F_2, \underbrace{\ldots \qquad\qquad \ldots}_{\text{possibly infinite sequence}} F_\kappa$$

such that for all $i \leqslant \kappa$, F_i is either an instance of some axiom, or $F_i \in \Gamma$, or it can be inferred from some of its predecessors by application of some inference rule.

Exercise

$\{\neg \mathrm{P}_{\geqslant \frac{1}{n}} p \mid n \geqslant 1\} \vdash \mathrm{P}_{\geqslant 1} \neg p$

Consistency

Definition

A set of formulas Γ is **consistent** iff $\Gamma \nvdash \bot$, where $\bot$ is the abbreviation for $\neg\top$; Γ is **maximal consistent** iff it is consistent and it is not contained in any other consistent theory (i.e. it is maximal in the sense of inclusion).

$\{\neg P_{\geqslant 1}\neg p\} \cup \{\neg P_{\geqslant \frac{1}{n}} p \mid n \geqslant 1\}$ is not consistent.

Deduction theorem

If $\Gamma, F \vdash G$ then $\Gamma \vdash F \Rightarrow G$.

PROOF. The induction on the length of the derivation $\Gamma, F \vdash G$ ...

Extension theorem

Every consistent set of formulas Γ can be extended to a maximal consistent set Γ^+.

PROOF. $\langle F_k : k \geqslant 1\rangle$ – an arbitrary enumeration of all L_P-formulas.

- $\Gamma_0 = \Gamma$
- $n \geqslant 1$
  - If $\Gamma_n \cup \{F_n\}$ is consistent, then $\Gamma_{n+1} = \Gamma_n \cup \{F_n\}$;
  - Let $\Gamma_n \cup \{F_n\}$ is not consistent. Then we have the following cases:
    - $F_n = A \Rightarrow P_{\geqslant r} B$. Then, there is $t_n < r$ such that $\Gamma_n \cup \{\neg F_n, \neg(A \Rightarrow P_{\geqslant t_n} B)\}$ is consistent. In this case we define Γ_{n+1} by

      $$\Gamma_{n+1} = \Gamma_n \cup \{\neg F_n, \neg(A \Rightarrow P_{\geqslant t_n} B)\}.$$

      The existence of such t_n is provided by (A_r);
    - Otherwise, $\Gamma_{n+1} = \Gamma_n \cup \{\neg F_n\}$.

Let $\Gamma^+ = \bigcup_{n \geqslant 0} \Gamma_n$.

Extension theorem

Every consistent set of formulas Γ can be extended to a maximal consistent set Γ^+.

PROOF. $\langle F_k : k \geqslant 1 \rangle$ – an arbitrary enumeration of $\mathrm{L_P}$-formulas.

$\vdots$

Let $\Gamma^+ = \bigcup_{n \geqslant 0} \Gamma_n$.

Calim 1 Γ_n is consistent for each n;
Calim 2 Γ^+ is deductively closed, i.e. if $\Gamma^+ \vdash F$ then $F \in \Gamma^+$.
Calim 3 Γ^+ is a maximal consistent set.

Completeness theorem

Every consistent set of formulas has an $\mathrm{L_P}$-model.

PROOF.

- $\mathbb{W}$ is the set of all maximal consistent extensions of Γ; according to the previous theorem, $\mathbb{W}$ is not empty.
- $\mathcal{F} = \{[F] \mid F \in \mathrm{L_P}\}$, where $[F] = \{\Delta \in \mathbb{W} \mid F \in \Delta\}$;
- for all $\Delta \in \mathbb{W}$, $\mu_\Delta[F] \stackrel{\text{def}}{=} \sup\{t \in [0,1] \cap \mathbb{Q} \mid \mathrm{P}_{\geqslant t} F \in \Delta\}$.

1) $\mathcal{F}$ is an algebra of subsets of $\mathbb{W}$?
2) μ_Δ is finitely additive?
3) $(\mathbb{W}, \mathcal{F}, \mu_\Delta) \models \Gamma$?

Completeness theorem

Corollary

$\Gamma \vdash F$ iff $\Gamma \models F$