Logic and Probability*

(manuscript in progress)

Nebojša Ikodinović
University of Belgrade, Faculty of Mathematics
nebojsa.ikodinovic@matf.bg.ac.rs

Contents

We begin in philosophy by discussing the question 'Why specify probability functions on sentences of predicate languages?', and at the same time, reviewing some basic concepts of mathematical logic and probability theory. Then, we touch on important mathematical topics such as model theory of finite structures, and model theory of probability spaces. Finally, we take a proof-theoretic approach and present some techniques for developing formal systems that may be of interest to theoretical computer science and, in particular, artificial intelligence. These seemingly separate themes are presented in a way that highlights their common grounds and opens up new possibilities for their mutual connections.

*These lecture notes cover part of the course:
Nebojša Ikodinović, Dragan Doder, *Logics with Probability Operators and Quantifiers*
ESSLLI, 34th EUROPEAN SUMMER SCHOOL IN LOGIC, LANGUAGE AND INFORMATION
University of Ljubljana, Faculty of Computer and Information Science, July 31 – August 11, 2023

1. *Why specify probability functions on sentences of predicate languages?*

MOTIVATING EXAMPLE. Inductive logic is an emerging discipline within Mathematical logic.[1] In contrast, Inductive logic has long been an important area of philosophy which emphasize the importance of inductive reasoning in attempts to understand human cognition. In the recent years, Artificial Intelligence is very interested in such attempts, as it is interested in the formalization of all forms of reasoning. Mathematics, especially mathematical logic, provides a natural link between philosophical discussions and machine implementations.

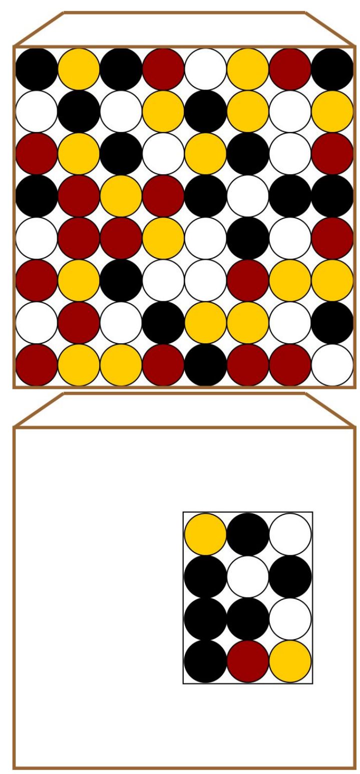
[1] J. Paris, A. Vencovská, *Pure Inductive Logic*, Cambridge University Press, 2015

Carnap was probably the first to attempt to give a mathematically rigorous foundation for Inductive Logic. He held that important aspects of the scientific method are irreducibly inductive. His leading idea was to develop systems of inductive logic based on the concept of probability:

> 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)*)

It is hard to say what logical probability is.[2] Instead, we give a hopefully illustrative example, as simple as possible, that highlights two faces of the concept of probability.

[2] Carnap himself changed his intuitive point of view several times.



Let us consider the following question in two cases: What is the probability of drawing a black ball from the box?

In the first case, We know the contents of the box. In this case, answering the question is a simple high school task.

In the second case, we had the opportunity to see only 12 balls and find that among them there are 6 black balls. In addition to what we have seen, we know or believe that: in general, the colors of the balls are uniformly distributed in the box; but exceptions are possible. It is clear that the question in the second case could trigger an extensive debate, and the answer could be a combination of our experimental observation and our knowledge, i.e., a mixture of an empirical factor and a logical factor. The question of how to combine these two factors is at the heart of Carnap’s attempt to explicate the notion of inductive probability.

Later we will return to this example. Now, the reader is invited to write down his/her estimation in the second case. At the end of the section it will be interesting to compare the filling-based answers and the solution within Carnap’s Basic System of Inductive Logic.

1.1. ... Sentences of Predicate languages

Although we assume that the reader is familiar with the basics of predicate languages, a brief overview is given to fix notations and terminology.

Predicate languages form a very important class of formal languages. We restrict ourselves to so-called *pure predicate language* determined by 1) a universe of discourse and 2) predicates with which we want to describe the universe.

The term **universe (of discourse)** generally refers to the collection of objects discussed in a particular discourse. The elements of a universe will be called **individuals**.

For example, let the universe consist of six individuals (persons): Ann, Ben, Cam, Deb, Eva, Fox

In general, a **predicate** is a word or phrase which can be combined with one or more names of individuals to yield meaningful sentences. A unary (1-place) predicate refers to a property of single individuals; A binary (2-place) predicate refers to relationship between pairs of individuals; etc.

E.g.

```
...is an artist.
...is a barber.
...is following ...(on a social networking service)
```

We greatly simplify the writing of sentences by introducing a **vocabulary** consisting of symbols used to denote:

1. 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, $a_1, \ldots$ we always assume that the names exhausted the universe;
2. predicates; upper case letters, with or without numeric subscripts will be used to denote predicates: A, B, C, $\ldots$, $A_1, \ldots$

Instead of the full names of persons from our universe (Ann, Ben, Cam, Deb, Eva, Fox) we use the abbreviations: a, b, c, d, e, f.
We write
$A(\cdot)$ for '$\cdots$ `is an artist`',
$B(\cdot)$ for '$\cdots$ `is a barber`', and
$F(\cdot,\cdot)$ for '$\cdots$ `is following` $\cdots$'.

The choice of vocabulary depends on the situation we want to describe. When we specify a vocabulary, each predicate symbol will be supplied with superscript that determines the arity of the predicate (e.g. $A^{(1)}$, $B^{(1)}$, $F^{(2)}$). The basic building blocs of a predicate language are atomic sentences.

► An **atomic sentence** is made of a k-place predicate symbol followed by k individual names. The individuals to which a predicate is applied are to the right of the predicate, separated by commas, and the entire list is in parentheses.

$$\texttt{Predicate}(\underbrace{\_\_\_\_,\ldots,\_\_\_\_}_{\text{places for individual names}})$$

Atomic sentences are the simplest sentences of the above vocabulary:

$A(a)$	`Ann is a artist.`
$A(e)$	`Eva is a artist.`
$B(b)$	`Ben is a barber.`
$F(f,c)$	`Fox is following Cam.`

If the universe contains n individuals, then

- a unary predicate determines n atomic sentences;
- a binary predicate determines n^2 atomic sentences;

  etc.

If $\mathtt{L}$ is a set of predicate symbols, over a universe $\mathbb{U}$, $\mathtt{L}(\mathbb{U})$ will be the set of all atomic sentences. If $\mathtt{L}$ and $\mathbb{U}$ are finite sets, then: $\#\mathtt{L}(\mathbb{U}) = \sum_{P\in\mathtt{L}} \#\mathbb{U}^{\mathrm{arity}(P)}$. ($\#X$ denotes the number of elements in X.) If $\mathtt{L}$ and $\mathbb{U}$ are at most countably infinite, then $\mathtt{L}(\mathbb{U})$ is at most countably infinite too.

Using the 1-placed predicate A over the universe with 6 individuals, we make 6 atomic sentences:

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

Using the 2-placed predicate F over the same universe, we make 36 atomic sentences:

$$\begin{array}{llllll} F(a,a) & F(a,b) & F(a,c) & F(a,d) & F(a,e) & F(a,f) \\ F(b,a) & F(b,b) & F(b,c) & F(b,d) & F(b,e) & F(b,f) \\ F(c,a) & F(c,b) & F(c,c) & F(c,d) & F(c,e) & F(c,f) \\ \ldots \end{array}$$

▶ An **atomic truth assignment** (a valuation) is any function $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$. An $\mathtt{L}$-model 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. We will sometimes call the function $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$ itself an $\mathtt{L}$-model. If $\mathtt{L}$ and $\mathbb{U}$ are finite, there are also finitely many truth assignments: $\prod_{P\in\mathtt{L}} 2^{\#\mathbb{U}^{\mathrm{arity}(R)}}$.
A truth assignment is often called:

- an $\mathtt{L}$-interpretation over $\mathbb{U}$;
- a relational $\mathtt{L}$-structure over $\mathbb{U}$.

Relational structures arise in many branches of mathematics. Graphs (undirected, directed), orderings, databases are well-known examples of relational structures.
If $\mathbb{S} \subseteq \mathbb{U}$ (and hence $\mathtt{L}(\mathbb{S}) \subseteq \mathtt{L}(\mathbb{U})$), and $\mathbf{M}|_{\mathtt{L}(\mathbb{S})}$ is the restriction of $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$ to $\mathtt{L}(\mathbb{S})$, then $(\mathbb{S},\mathbf{M}|_{\mathtt{L}(\mathbb{S})})$ is the **substructure** of $(\mathbb{U},\mathbf{M})$.

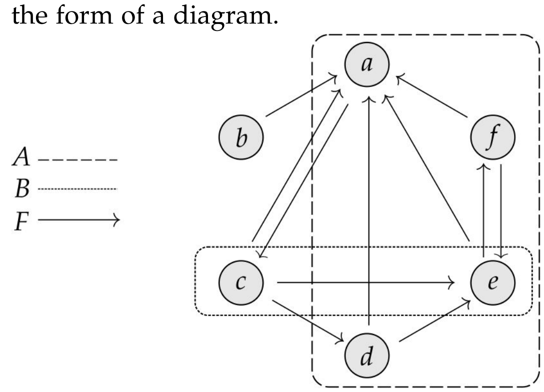
For $\mathbb{U} = \{a,b,c,d,e,f\}$ and $\mathtt{L} = \{A^{(1)},B^{(1)},F^{(2)}\}$, there are

$$2^6 \cdot 2^6 \cdot 2^{36} = 2^{48} = 281\,474\,976\,710\,656$$

truth assignments of atomic sentences. One of them is:

$$\begin{array}{llllll} A(a) & A(b) & A(c) & A(d) & A(e) & A(f) \\ B(a) & B(b) & B(c) & B(d) & B(e) & B(f) \\ F(a,a) & F(a,b) & F(a,c) & F(a,d) & F(a,e) & F(a,f) \\ F(b,a) & F(b,b) & F(b,c) & F(b,d) & F(b,e) & F(b,f) \\ F(c,a) & F(c,b) & F(c,c) & F(c,d) & F(c,e) & F(c,f) \\ F(d,a) & F(d,b) & F(d,c) & F(d,d) & F(d,e) & F(d,f) \\ F(e,a) & F(e,b) & F(e,c) & F(e,d) & F(e,e) & F(e,f) \\ F(f,a) & F(f,b) & F(f,c) & F(f,d) & F(f,e) & F(f,f) \end{array}$$

It is convenient to present an assignment in the form of a diagram.



Pure predicate languages get their full expressive power by extending vocabularies with so-called **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 $\forall$.

Variables are used as placeholders for arbitrary individuals. A symbol which is either a variable or an individual name is called an individual symbol.

▶ An **atomic formula** is made of a k-place predicate symbol followed by k individual symbols.

$$\texttt{Predicate(}\ \underbrace{\_\_\_\_,\ldots,\_\_\_\_}_{\text{places for individual symbols}}\ \texttt{)}$$

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

▶ Starting from atomic formulas we inductively build up all **formulas** of the chosen vocabulary:

- 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.

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

▶ When a quantifier and a variable are placed in front of a formula, all occurrences of the variable in the formula become bound. If the appearance of a variable is not bound by a quantifier, we say that it is free. Thus, in a formula, a variable may occur **free** or **bound** (or both).

We used upper case letters to denote formulas. A formula F will be denoted by $F(x_1,\ldots,x_k)$ when we want to emphasize that all free variables of F are among $x_1,\ldots,x_k$. If F is a formula, x is a variable, and c is a symbol for an individual, then $F[x/c]$ denotes the formula obtained by replacing all free occurrences of the variable x with the symbol c. Of course, if x is not free in F, then the formula $F[x/c]$ is identical to the formula F.

The formula F:

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

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

▶ A **sentence** is a formula with no free occurrence of a variable. Let $\mathtt{L}_{\omega\omega}(\mathbb{U})$ be the set of all sentences. The set of all quantifier free sentences $\mathtt{L}_{\omega 0}(\mathbb{U})$ can be regarded as the set of propositional formulas whose 'propositional letters' are the atomic formulas of $\mathtt{L}$.

$\neg \exists x\, (A(x) \wedge B(x))$
$\forall x \exists y\, (A(x) \Rightarrow B(y) \wedge F(x,y))$

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

In the notation $\mathtt{L}_{\omega\omega}$, the letter ω denotes the first infinite ordinal, that is the standard well-ordered set of natural numbers. The first occurrence of ω indicates that the language permits conjunctions and disjunctions with less then ω (i.e., finitely many) constituents. The second occurrence of ω indicates that the language permits simultaneous quantification over fewer than then ω (i.e., finitely many) variables. Later we will deal with the languages $\mathtt{L}_{\omega_1\omega}$, where ω_1 is the first uncountable ordinal number.

▶ Every atomic truth assignment $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$ has a unique extension to the **truth assignment** over all sentences $\mathbf{M} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to \{0,1\}$.

Given an atomic truth assignment $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0,1\}$, the truth value of any sentences could be calculated using the truth tables of logical connectives,

	$\neg$
0	1
1	0

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

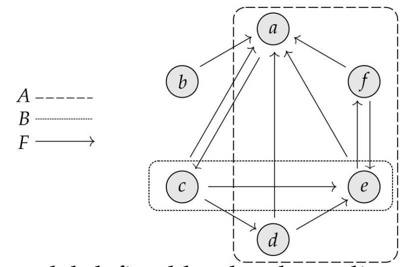
$\vee$	0	1
0	0	1
1	1	1

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

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

and the following rules for quantifiers:

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



Note that after replacing all free occurrence of variables in a formula by individuals, we obtain the sentence that can be true or false.

▶ We say that an assignment **M models** a sentence S, and write $\mathbf{M} \models S$, if $\mathbf{M}(S) = 1$.

If $\mathbf{M}$ is the $\mathtt{L}$-model defined by the above diagram, then:
$\mathbf{M} \models F(a,b) \wedge \neg F(b,a)$
$\mathbf{M} \models \forall x\, \neg F(x,b)$
$\mathbf{M} \models \exists y \forall x\, \neg F(x,y)$ etc.

▶ Finally, we introduce the fundamental logical notions.
A sentence A is **valid**, denoted by $\models A$, if A is true in all models.
If every assignment that models P also models Q, then Q is a (semantical or logical) **consequence** of P, denoted by $P \models Q$.
If both, Q is a consequence of P and P is a consequence of Q, then P and Q are **equivalent**, denoted by by $A \equiv B$.

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

▶ A **literal** is an atomic formula (also known as a positive literal) or its negation (a negative literal):

$$\pm\mathtt{Predicate}(\ \underbrace{\_\_\_, \dots, \_\_\_}_{\text{places for individual and variables}}\ )$$

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

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.

THEOREM 1. **[DNF]** Every quantifier free sentence F is equivalent to a disjunction of atoms:

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

1.2. ... *Specify Probability functions* ...

By the standard Kolmogorov's foundation of the theory of probability, a probabilistic model is determined by 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* $\mu : \mathcal{B} \to [0,1]$.[6] One thinks of Ω as being the set of all possible outcomes (so called elementary events) of a given random phenomenon.

[6] One of the most familiar examples is the roll of two dice. The sample space Ω is the set of 36 pairs of numbers,

$$\Omega = \{(1,1), (1,2), \ldots, (1,6), (2,1), \ldots, (6,6)\}$$

Each possible outcome is equally likely, $\mu(i,j) = 1/36$, for any (i,j). If E is the event *at least one six is rolled*, then $\mu(E) = 11/36$. The probability of not rolling a six is $\mu(E^{\complement}) = 25/36$.

DEFINITION 1. A collection $\mathcal{B}$ of subsets of Ω is called an **algebra** if it satisfies the following properties:

- $\Omega \in \mathcal{B}$;
- if $E \in \mathcal{B}$ then $E^{\complement} = \Omega \setminus E \in \mathcal{B}$;
- if $E_1, E_2 \in \mathcal{B}$ then $E_1 \cup E_2 \in \mathcal{B}$.

An algebra $\mathcal{B}$ is a σ-algebra if:

- for any sequence $\langle E_n : n \geqslant 1 \rangle$ of sets in $\mathcal{B}$, $\bigcup\limits_{n=1}^{\infty} E_n \in \mathcal{B}$.[7]

[7] This property is called σ-additivity. There are two trivial examples of σ-algebras. The first is the power set 2^{Ω} which is also called the discrete σ-algebra. The second is the so-called trivial σ-algebra consisting only of the two sets Ω and $\emptyset$.

A pair $(\Omega, \mathcal{B})$ is called a **measurable space**.

DEFINITION 2. A function $\mu : \mathcal{B} \to [0,1]$ is a **finitely-additive probability measure** on $(\Omega, \mathcal{B})$ if it satisfies the following properties:

- $\mu(\Omega) = 1$;
- if E_1, E_2 are disjoint sets ($E_1 \cap E_2 = \emptyset$), then $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$

A finitely-additive probability measure is a **probability measure** if:

- for any sequence $\langle E_n : n \geqslant 1 \rangle$ of pairwise disjoint sets, $\mu(\bigcup\limits_{i=1}^{\infty} E_i) = \sum\limits_{i=1}^{\infty} \mu(E_i)$.

We will call the triple $(\Omega, \mathcal{B}, \mu)$ a **probability space**.

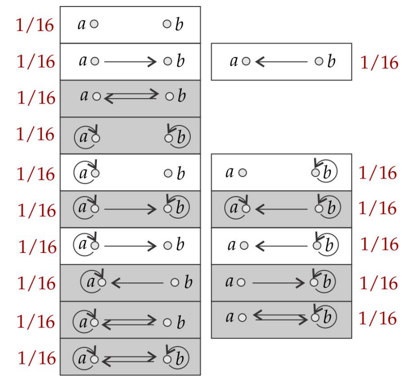
PROPOSITION 1. (a) If $\langle E_n : n \geqslant 1 \rangle$ is an increasing sequence of sets in $\mathcal{B}$, i.e. if $E_1 \subseteq E_2 \subseteq \cdots$, then $\mu(\bigcup\limits_{n=1}^{\infty} E_n) = \lim\limits_{n\to\infty} \mu(E_n)$.

(b) If $\langle E_n : n \geqslant 1 \rangle$ is a decreasing sequence of sets in $\mathcal{B}$, i.e. $E_1 \supseteq E_2 \supseteq \cdots$, then $\mu(\bigcap\limits_{n=1}^{\infty} E_n) = \lim\limits_{n\to\infty} \mu(E_n)$.

EXAMPLE 1. Let us consider a very simple probability space. We are given an urn containing all L-structures over $\mathbb{U}$. The random experiment consists in choosing one structure from the urn. For each sentence from $\mathrm{L}_{\omega\omega}(\mathbb{U})$ we could naturally define the probability of being true in a randomly chosen structure.

To simplify the presentation, let $\mathtt{L} = \{F\}$ consist of just one 2-placed predicate F. There are 16 L-structures over $\{a, b\}$:

$$(\{a,b\}, \mathbf{M}_1),\ (\{a,b\}, \mathbf{M}_2),\ \cdots (\{a,b\}, \mathbf{M}_{16}).$$



These structure form the sample space $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$. Assume that all structures are equally likely to be chosen.

What is the probability that we choose a structure in which $\forall x \exists y\, F(x, y)$ is true? The sentence $\forall x \exists y\, F(x, y)$ is true in 9 of 16 structures:

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

In the same way, we find the probability of any sentence from $\mathtt{L}_{\omega\omega}(\mathbb{U})$. A sentence $S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$ determines the event

$$[S] = \{(\{a,b\}, \mathbf{M}_i) \mid \mathbf{M}_i(S) = 1\},$$

i.e. the set of all structures in which the sentence S is true, and

$$\mathbf{P}(S) = \frac{\#[S]}{16},$$

where $\#[S]$ denotes the cardinality (the number of elements) of $[S]$.

Exercise 1. Considering the urn from the preceding example, and find the probability of:
$F(a, b)$, $\forall x\, F(x, a)$, $\exists y \forall x\, F(x, y)$, $\forall x \forall y\, (F(x, y) \Rightarrow F(y, x))$

DEFINITION 3. A **probability** is a function $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0, 1]$ 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 $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;[9]

P3 $\mathbf{P}(\exists x\, A) = \lim_{n\to\infty} \mathbf{P}(A(c_1) \vee \cdots \vee A(c_n))$, in case that the universe is countable $\mathbb{U} = \{c_1, c_2, c_3, \ldots\}$. [10]

PROPOSITION 2. Let $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0, 1]$ be a probability. Then for $A, B \in \mathtt{L}_{\omega\omega}(\mathbb{U})$.

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

Exercise 2. Prove Proposition 2.

PROPOSITION 3. For $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0, 1]$ satisfying P1 and P2, condition P3 is equivalent to:

$$(P3')\quad \mathbf{P}(\exists x\, A) = \sum_{n=1}^{\infty} \mathbf{P}\left(A(c_n) \wedge \neg \bigvee_{i=1}^{n-1} A(c_i)\right)$$

[9] Equivalently,
(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
$\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B)$;

[10] Condition P3 is often referred to as **Gaifman's Condition**, and it will be discussed later.

$(P3'')$ $\mathbf{P}(\forall x\, A) = \lim_{n\to\infty} \mathbf{P}\left(\bigwedge_{i=1}^{n} A(c_i)\right)$

Exercise 3. Prove Proposition 3. (See Preposition 1.)

A particularly simple example of a probability is any truth assignment $\mathbf{M} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to \{0,1\}$ ($\{0,1\} \subset [0,1]$).

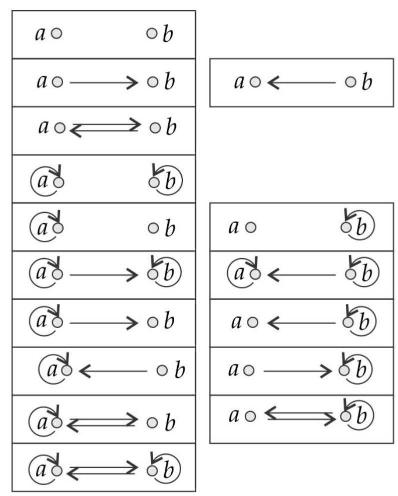
PROPOSITION 4. Convex sums of probability functions are also probability function. In other words, for a family of probabilities $\mathbf{P}_i : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$, $i \in I$, and numbers $a_i \geqslant 0$, $i \in I$ such that $\sum_i a_i = 1$, the function $S \mapsto \sum_i a_i \mathbf{M}(S)$ is a probability.

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

$$S \mapsto \sum_i a_i \mathbf{M}(S)$$

is a probability. It turns out that every probability is a linear combination of some classical structures.

EXAMPLE 2. As in the previous example, let $\mathtt{L}$ contain only one 2-place predicate F, and $\mathbb{U} = \{a,b\}$. Let $\mathbf{P} : \{F\}_{\omega\omega}(\mathbb{U}) \to [0,1]$ be a probability.



There is a bijection between all structures and so-called *complete atoms* ($\mathtt{L}$-atoms over $\mathbb{U}$).

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} $F(a,a) \wedge F(a,b) \wedge F(b,a) \wedge F(b,b)$

Then, for any sentence S:

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

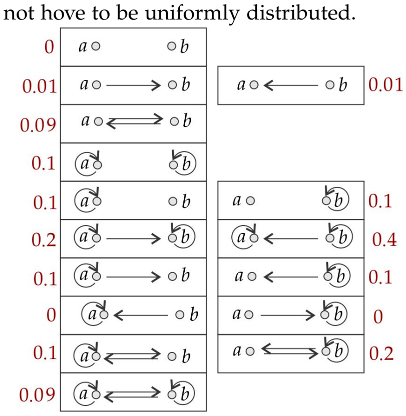
Moreover, each sentence is a disjunction of atoms:

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

which is just a reformulation of the earlier observation: each $\mathtt{L}$-sentence could be identify with the set all structures in which it is true. If $a_i = \mathbf{P}(A_i)$, $i = 1, \ldots, 16$, it is obvious that $\sum_{i=1}^{16} a_i = 1$, and for any S:

$$\mathbf{P}(S) = \mathbf{P}\left(\bigvee_{\mathbf{M}_i(S)=1} A_i\right) = \sum_{\mathbf{M}_i(S)=1} \mathbf{P}(A_i) = \sum_{\mathbf{M}_i(S)=1} a_i = \sum_{i=1}^{16} a_i \mathbf{M}_i(S)$$

Probabilities assigned to outcomes do not hove to be uniformly distributed.



Exercise 4. Find a probability $\mathbf{P} : \{F\}_{\omega\omega}(\{a,b\}) \to [0,1]$, if it exists, such that:

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

If the universe is infinite, there are infinitely many possible models, but an analogous result holds. The proof relies on the ideas of the previous example. The next theorem precisely states a general result. The theorem shows how a probability on the set of sentences is related to the classical definition of a probability measures on events. In the theorem, we use the following notations:

- $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$ is the set of all $\mathtt{L}(\mathbb{U})$-truth assignments;
- $[S] = \{\mathbf{M} \in \mathcal{M}_{\mathtt{L}(\mathbb{U})} \mid \mathbf{M}(S) = 1\}$, $S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$;

  $\mathbf{1}_{[S]}$ is the characteristic (indicator) function of $[S]$, that is the function $\mathbf{1}_{[S]} : \mathcal{M}_{\mathtt{L}(\mathbb{U})} \to \{0,1\}$, such that $\mathbf{1}_{[S]}(\mathbf{M}) = \mathbf{M}(S)$.
- $\mathcal{A}_{\mathtt{L}(\mathbb{U})} = \{[S] \mid S \in \mathtt{L}_{\omega 0}(\mathbb{U})\}$ is an algebra of subsets of $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$;
- $\mathcal{B}_{\mathtt{L}(\mathbb{U})}$ is a σ-algebra extending $\mathcal{A}_{\mathtt{L}(\mathbb{U})}$.

THEOREM 2. [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:[13]

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

[13] In finite cases:

$$\mathbf{P}(S) = \sum_i \mathbf{M}_i(S) a_i,$$

where $\mu(\mathbf{M}_i) = a_i$, as in Example 2.

The representation theorem shows that a probability on a set of formulas can be viewed as a kind of model consisting of a family of classical models (called worlds) equipped with their 'weights'. The semantics of many probabilistic logics, which will be discussed later, is based on this view.

The proof of the preceding theorem will be omit. It would be proved using a logical analog of Carathéodory's extension theorem.

THEOREM 3. [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:

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

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.

THEOREM 4. [Gaifman's Extension Theorem][14] Assume that a function $\mathbf{P} : \mathtt{L}_{\omega 0}(\mathbb{U}) \to [0,1]$ satisfies (P1) and (P2). Then $\mathbf{P}$ has a unique extension to a probability function satisfying (P1), (P2), (P3) for all sentences $\mathtt{L}_{\omega\omega}(\mathbb{U})$.

[14] H. Gaifman, *Concerning measures in first order calculi*, Israel journal of mathematics 2 (1), 1-18, 1964

PROOF. It is easy to check that

$$\mathcal{A}_{\mathtt{L}(\mathbb{U})} = \{[S] \mid S \in \mathtt{L}_{\omega 0}(\mathbb{U})\}$$

is an algebra of subsets of $\mathcal{M}_{\mathtt{L}(\mathbb{U})}$, and $\mu_{\mathbf{P}}$ defined by

$$\mu_{\mathbf{P}}([A]) = \mathbf{P}(A)$$

is a finitely additive measure on this algebra.

The condition $(*)$ of Carathéodory's Extension Theorem is trivially satisfied:

(∗) Suppose $S, A_i \in \mathrm{L}_{\omega 0}$, $i = 1, 2, \ldots$ and $(\star)$ $\bigcup_{i=1}^{\infty}[A_i] = [S]$. Then, it must be the case that for some finite n: $\bigcup_{i \leqslant n}[A_i] = [S]$, otherwise $\{\neg A_i \mid i = 1, 2, \ldots\} \cup \{S\}$ would be finitely satisfiable and hence by Compactness[15] would be satisfiable in some $\mathbf{M}$, contradicting $(\star)$.

[15] Compactness theorem is a fundamental property of first-order logic: *A set of* L*-formulas has a model iff every finite subset of it does.*

By Carathéodory's Extension Theorem there is a unique extension $\mu^+_{\mathbf{P}}$ of $\mu_{\mathbf{P}}$ defined on the σ-algebra $\mathcal{B}$ generating by $\mathcal{A}$. Notice that since

$$[\exists x\, S] = \{\mathbf{M} \mid \mathbf{M} \models \exists x\, S\} = \bigcup_{i \geqslant 1}\{\mathbf{M} \mid \mathbf{M} \models S[x/c_i]\} = \bigcup_{i \geqslant 1}[S[x/c_i]]$$

and $\mathcal{B}$ is closed under complements and countable unions, the algebra $\mathcal{B}$ contains all sets $[S]$, $S \in \mathrm{L}_{\omega\omega}(\mathbb{U})$. Now define $\mathbf{P}^+$ on $\mathrm{L}_{\omega\omega}(\mathbb{U})$:

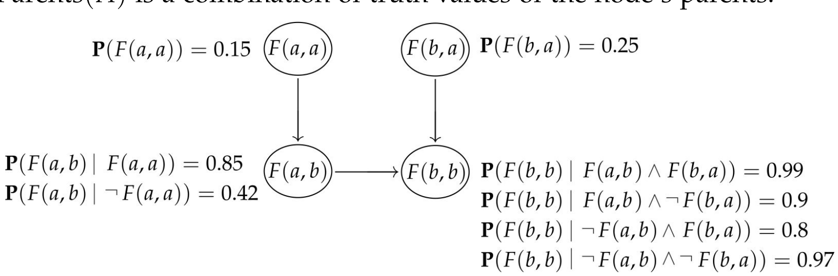
$$\mathbf{P}^+(S) \stackrel{\text{def}}{=} \mu^+_{\mathbf{P}}([S]).$$

Since $\mu^+_{\mathbf{P}}$ is a measure, $\mathbf{P}^+$ satisfies P1, P2 and also P3 from the fact that $\mu^+_{\mathbf{P}}$ is countably additive.

Finally, $\mathbf{P}^+$ must be the unique extension of $\mathbf{P}$ satisfying P1-3. For if there was another such probability function, say $\mathbf{Q}$, then it is easy to see that $\mu^+_{\mathbf{Q}}$ would have to agree with $\mu^+_{\mathbf{P}}$ on $\mathcal{B}$, and then $\mathbf{P}$ would have to agree with $\mathbf{Q}$ □

By Gaifman's theorem, to specify a probability function on $\mathrm{L}_{\omega\omega}(\mathbb{U})$ it is enough to say how it acts on the quantifier free sentences. Moreover, it is enough to define probability function on conjunctions of literals (atomic formulas and their negations), i.e., on atoms in $\mathrm{L}_{\omega 0}(\mathbb{U})$. The next example comes from a very vital area of research with large applications.

EXAMPLE 3. Bayesian networks are closed acyclic graphs (DAG's) whose nodes represent assertions and edges represents some kind of conditional dependencies. Nodes that are not connected are assertions that are conditionally independent of each other. In many cases, the assertions associated with the nodes are represented by atomic formulas; say in $\{F\}_{\omega 0}(\{a, b\})$. In addition, each node is associated with a probability function that takes, as input, a particular set of truth-values for the node's parents, and gives (as output) the probability of the sentence represented by the node. More precisely, each note A is associated with the values $\mathbf{P}(A \mid \pm\text{Parents}(A)) \in [0, 1]$, where $\pm\text{Parents}(A)$ is a combination of truth values of the node's parents.



The join probability function determines probabilities of atoms, by so-called the chain rule:

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

By Gaifman's Extension theorem, the given Bayesian network uniquely determines the probability on $\{F\}_{\omega\omega}(\{a,b\})$.

Similar ideas may be applied to undirected and possibly cyclic graphs such as Markov networks. This will be discussed later.

Any probability on the set of all sentences can be regarded as a natural generalization of the notion of classical structure. We recall that a truth assignment $\mathbf{M} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to \{0,1\}$ is a special case of probability. This analogy motivates Gaifman to introduce the notion of a probability structure.

DEFINITION 4. A probability structure is a pair $(\mathbb{U}, \mathbf{P})$, where $\mathbf{P}$ is a probability on $\mathtt{L}_{\omega\omega}(\mathbb{U})$.

The work of Gaifman was extended by D. Scott and P. Krauss[16]. Among other things, they have developed a model theory of probability structures. It turns out that various probability-model-theoretic concepts can be defined by analogy with the standard concepts of ordinary model theory. The following examples present two important constructions of probability structures. Details are left to the reader.

[16] D. Scott and P. Krauss, *Assigning Probabilities to Logical Formulas*, In Jaakko Hintikka, Patrick Suppes (eds.), Aspects of Inductive Logic. Elsevier: Amsterdam. pp. 219 – 264, 1966

EXAMPLE 4. **[Independent union]** Let I be a finite index set. For each $i \in I$, let $\mathtt{L}_i$ be a set of predicate symbols, and $\mathbf{P}_i : \mathtt{L}_{i\omega\omega}(\mathbb{U}) \to [0,1]$ a probability. Let $\mathtt{L} = \bigcup_{i \in I} \mathtt{L}_i$. Any atom A from $\mathtt{L}_{\omega 0}(\mathbb{U})$ can be written as $\bigwedge_{i \in I} A_i$, where A_i is an atom from $\mathtt{L}_{i\omega 0}(\mathbb{U})$. A probability $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ can be defined by

$$\mathbf{P}(A) = \prod_{i \in I} \mathbf{P}_i(A_i)$$

The probability structure $(\mathbb{U}, \mathbf{P})$ is the **independent union** of probabilistic structures $(\mathbb{U}, \mathbf{P}_i)$, $i \in I$.

EXAMPLE 5. **[Ultraproduct]** Let I be a finite index set, and $\mathtt{L}$ a set of predicate symbols. For each $i \in I$ let $\mathbf{P}_i : \mathtt{L}_{\omega\omega}(\mathbb{U}_i) \to [0,1]$ be a probability. Let $\mathbb{U} = \prod_{i \in I} \mathbb{U}_i$ be the Cartesian product of the family of universes $\mathbb{U}_i$, $i \in I$. For $S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$ and $i \in I$, let $S \mid i$ be the projection of S onto the i^{th} coordinate: replace in S every $c = (c_i)_{i \in I} \in \mathbb{U}$ by $c_i \in \mathbb{U}_i$. Finally, let μ be a probability on the power set of I. Define for all $S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$ a function $\mathbf{P}$ by the equation

$$\mathbf{P}(S) = \sum_{i \in I} \mathbf{P}_i(S \mid i) \cdot \mu(\{i\}).$$

$(\mathbb{U}, \mathbf{P})$ is the **ultraproduct** with respect to μ of the family of probability structure $(\mathbb{U}_i, \mathbf{P}_i)$, $i \in I$.

EXAMPLE 6. There is an interesting 'game' that also leads us to the concept of probability.

Imagine that a bookmaker invites you to participate in the following game. Several sentences S_1, S_2, ..., S_k are written on the ticket. You have to guess which sentences will be true and which will be false in a mysterious structure $\mathbf{M}^?$, which will be randomly chosen later. The ticket also contains the list of betting quotients $0 \leqslant p_1, \dots, p_k \leqslant 1$. For each sentence S_i you should choose one out of the two options:

(Opt_0) You give the bookmaker p_i €, and if it becomes $\mathbf{M}^?(S_i) = 1$, you receive 1 € from the bookmaker;

(Opt_1) You receive p_i € from the bookmaker, and if it becomes $\mathbf{M}^?(S_i) = 0$, you have to give the bookmaker 1 €.

Sentence	Quotient	Your choice	Cost	Payout
S_1	p_1	☐ (Opt_0)	$-p_i$	1 if $\mathbf{M}^?(S_1) = 1$
		☐ (Opt_1)	p_i	-1 if $\mathbf{M}^?(S_1) = 0$
$\vdots$				
S_k	p_k	☐ (Opt_0)	$-p_i$	1 if $\mathbf{M}^?(S_k) = 1$
		☐ (Opt_1)	p_i	-1 if $\mathbf{M}^?(S_k) = 0$

Let us define a function:

$$\mathbf{B}(S) = \begin{cases} 0, & \text{you choose } \mathrm{Opt}_0 \text{ for } S \\ 1, & \text{you choose } \mathrm{Opt}_1 \text{ for } S \end{cases}$$

When the mysterious structure is revealed, your total gain will be:

$$\sum_{i=1}^{k} (-1)^{\mathbf{B}(S_i)} (\mathbf{M}^?(S_i) - p_i).$$

How to make a rational strategy? The strategy could be based on the function $\mathbf{Bel} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ which would determine a threshold of your belief in sentences, in the following sense:

- if $p_i > \mathbf{Bel}(S_i)$, you choose (Opt_0) for S_i;
- if $p_i < \mathbf{Bel}(S_i)$, you choose (Opt_1) for S_i;
- ako je $p_i = \mathbf{Bel}(S_i)$, you choose either (Opt_0) or (Opt_1) for S_i.

Of course, you will accept the game only when there is a structure for which your gain is positive. It turns out that only in these cases you can find a rational strategy, and moreover, such strategies are determined by probability functions.

THEOREM. Suppose that for $\mathbf{Bel} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ there are no a most countable set I, sentences S_i, $i \in I$, and real numbers $p_i \in [0,1]$, $i \in I$, such that

$$\sum_{i \in I} (-1)^{b_i} (\mathbf{M}(S_i) - p_i) < 0, \text{ for all } \mathbf{M},$$

where

$$b_i = \begin{cases} 0, & \mathbf{Bel}(S_i) > p_i, \\ 1, & \mathbf{Bel}(S_i) \leqslant p_i. \end{cases}$$

Then **Bel** is probability.

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)

In this subsection, we focus on the second point of the quote – Carnap's inductive logic.[17] Before we introduce the Carnap's Basic System of Inductive logic, let us recall the notion of conditional probability.

[17] Gaifman was a Carnap's student and his work originally started with a study of Carnap's inductive logic.

PROPOSITION 5. Given a probability function $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$, and $C \in \mathtt{L}_{\omega\omega}(\mathbb{U})$ with $\mathbf{P}(C) > 0$, the function $\mathbf{P}(\cdot \mid C) : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ defined by

$$\mathbf{P}(S \mid C) = \frac{\mathbf{P}(S \wedge C)}{\mathbf{P}(C)},\ S \in \mathtt{L}_{\omega\omega}(\mathbb{U})$$

is a probability on $\mathtt{L}_{\omega\omega}(\mathbb{U})$.

DEFINITION 5. [18] Given a probability function $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$, and $C \in \mathtt{L}_{\omega\omega}(\mathbb{U})$ with $\mathbf{P}(S) > 0$, the **conditional probability** is a function $\mathbf{P}(\cdot \mid C) : \mathtt{L}_{\omega\omega}(\mathbb{U}) \to [0,1]$ (said **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 $\mathbf{P}(S_1 \wedge S_2) = \mathbf{P}(S_1) \cdot \mathbf{P}(S_2)$.

[18] There are also attempts to define the conditional probability directly. E.g. following De Finetti approach, we could define $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}) \times \mathtt{L}_{\omega\omega}(\mathbb{U}) \setminus \{\bot\} \to [0,1]$ by the following axioms:
$\mathbf{P}(A \mid A) = 1$;
$\mathbf{P}(\cdot \mid A)$ is a probability;
$\mathbf{P}(A \wedge B \mid C) = \mathbf{P}(A \mid C) \cdot \mathbf{P}(B \mid A \wedge C)$.

Carnap's last and probably the best exposition of inductive probability was published posthumously in two parts, in 1971 and 1980, **A Basic System of Inductive Logic I/II**. Here, we present just a fragment, as simple as possible, of the Basic System. We work with:

- a vocabulary contains only a finite number of unary predicate symbols $\mathtt{L} = \{F_1, F_2, \ldots, F_k\}$, $(k > 2)$,
- countably many constant symbols for individuals, $\mathbb{U} = \{c_1, c_2, \ldots\}$.

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 $\mathtt{L}(\mathbb{S})$. For each individual from the sample, $D_{\mathbb{S}}$ decides which properties it has, and which it doesn't have.[19]

Carnap looked for a probability $\mathbf{P} : \mathtt{L}_{\omega 0}(\mathbb{U}) \to [0,1]$ satisfying the following additional axioms:[20]

[19] We allow the empty set to be a sample. We choose a valid sentence to be a sample description for the empty set $D_{\emptyset}$.

[20] Patrick Maher, *Explication of Inductive Probability*, Journal of Philosophical Logic volume 39, pages 593–616, 2010

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

C2 **[Symmetry]** $\mathbf{P}(D_{\mathbb{S}})$ is not changed by permuting individual symbols from $\mathbb{S}$;

C3 **[Instantial relevance]** $\mathbf{P}(F_i(c_k) \mid F_i(c_\ell)) > \mathbf{P}(F_i(c_k))$ roughly says that *we learn from experience*;

C4 λ-condition: if $\mathbb{S}$ does not 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.

THEOREM 5. [**$\lambda - \gamma$ theorem**] If $\mathbf{P}$ is a probability which satisfies C1-4 and $k > 2$, then there exist $\lambda > 0$ and $\gamma_1, \dots, \gamma_k \in (0, 1)$ 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:

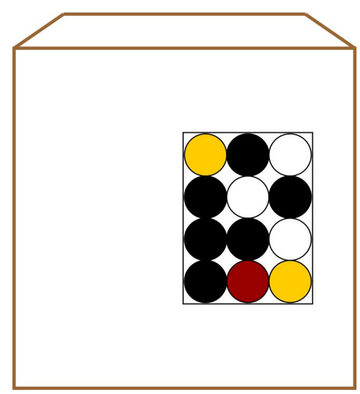
$$\mathbf{P}(F_i(c_k) \mid D_{\mathbb{S}}) = \frac{n_i + \lambda\gamma_i}{n + \lambda},$$

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.

The meaning of γ_i could be regarded as the a priori probability that something has the property F_i. The meaning of λ is interesting. From the main equation of the previous theorem we see: the probability that c_k has the property F_i, given the sample description of $\mathbb{S}$ is a mixture of the **empirical factor**, $\frac{n_i}{n}$, and the **logical factor**, γ_i.

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



Let us find the meaning of λ in our box example from the beginning of this section. The empirical factor is determined by our experimental observation: we see 12 balls, and 6 black balls. The logical factor is determined by our knowledge: Generally, four colors are uniformly distributed in the box; but, exceptions are possible.

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

It is obvious, the larger λ is, the more weight is put on the logical factor, and the slower one learns from experience.

Exercise 5. Solve the equation

$$\left(\frac{12}{12+\lambda}\right)\frac{6}{12} + \left(\frac{\lambda}{12+\lambda}\right)\frac{1}{4} = a,$$

where a is your estimation from the beginning of the section.

The above considerations illustrate two general issues related to inductive logic:[22]

[22] We refer to J. Paris, A. Vencovská, *Six Problems in Pure Inductive Logic*, Journal of Philosophical Logic volume 48, pages 731–747, 2019

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

2. *Very large finite phenomena*

MOTIVATING EXAMPLE. Interesting examples of probability models (in the sence of Definition 4, page 12) are known as random graphs.[23] We consider simple random graphs determined by two parameters: a positive integer n, and a real number $p \in [0,1]$.

[23] The subject of Random Graph began with a paper Erdős, Rényi, *On the Evolution of the random graph*, 1958.

Consider a vocabulary that contains only one 2-placed predicate symbol R, over the universe of size n, $\mathbb{U}_n = \{1, \dots, n\}$. Let us define a probability on $\{R\}_{\omega\omega}(\mathbb{U}_n)$ by the following rules:

- the probability of each atomic sentence is p,

$$\mathbf{P}_p(R(i,j)) = p,\ i,j \in \mathbb{U}_n,$$

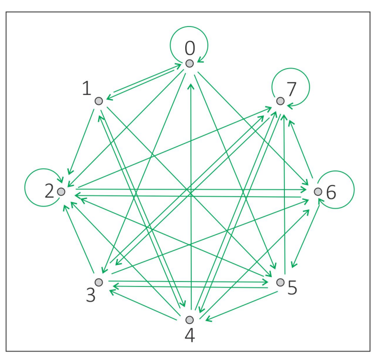
- all atomic sentence are mutually independent, i.e. if I and J make a partition of the set $\{1,\dots,n\} \times \{1,\dots,n\}$, and $|I| = k$, then

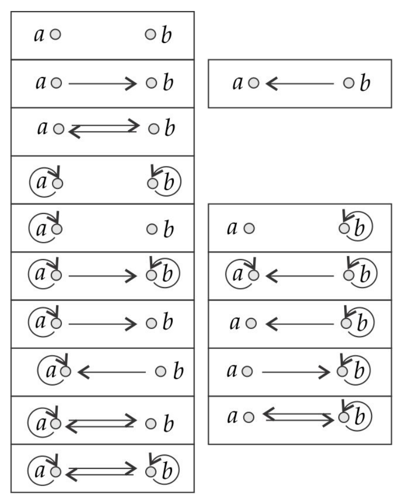
$$\mathbf{P}_p\left(\bigwedge_{(i,j)\in I} R(i,j) \wedge \bigwedge_{(m,\ell)\in J} \neg R(k,\ell)\right) = p^k(1-p)^{n^2-k}.$$

This setting completely determines a probability structure $(\mathbb{U}_n, \mathbf{P}_p)$. In a sense, this probabilistic model represents a randomly generated classical structure.

For example, let us generate the structure $(\mathbb{U}_8, \mathbf{P}_{0.5})$ performing the following experiment. We start with 8 nodes. For each pair of nodes, we flip a fair coin to decide whether these nodes should be adjacent or not. If head shows up, we draw an arrow between the nodes (or a loop if a single node appears in the pair); if tail shows up, we do draw no arrow between the nodes.

	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





Now, let us compare this probability structure with the family of all $\{R\}$-structures over two-element universe. Our random structure contains all, up to isomorphism, two-element structures. Considering pairs of individuals of the randomly generated $(\mathbb{U}_8, \mathbf{P}_{0.5})$, all possible 'types' of two-member structures can be observed; moreover, all of them are almost equally distributed inside $(\mathbb{U}_8, \mathbf{P}_{0.5})$.

The probability that the sentence $\forall x \exists y\, R(x,y)$ is true in an arbitrarily chosen two-member structure is approximately equal to 0.56 (see Example 1). Determining this probability is very close to counting pairs $(i,j) \in \mathbb{U}_8 \times \mathbb{U}_8$ for which the following open formula is true:

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

By counting, we find: the probability that a randomly chosen pair from $(\mathbb{U}_8, \mathbf{P}_{0.5})$ satisfies $F(x,y)$ is approximately equal to 0.57. We obtain almost the same result when counting two-element structures satisfying the sentence and when counting pairs in our random structure satisfying the corresponding open formula. The following observation will be of a central importance in this section: If H is much greater than n, then randomly generated $(\mathbb{U}_H, \mathbf{P}_{0.5})$ 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.

2.1. *Random structures*

Just as in the above example, Geifman defined an infinite probability model. We present this model in the simplest possible form. Let us consider a vocabulary (without equality) with a binary predicate symbol R over the countable universe: $\mathbb{U} = \{c_1, c_2, \ldots\}$. Let $p \in (0,1)$ be a fixed real number. Let us define a measure on the set $\{R\}_{\omega 0}(\mathbb{U})$ as follows: for each atomic sentence $R(c_i, c_j)$ let $\mathbf{P}(R(c_i, c_j)) = p$. Next, for every conjunction $\bigwedge_{i=1}^{k} A_i \wedge \bigwedge_{i=k+1}^{n} \neg A_i$, where A_i's are different atomic sentences,

$$\mathbf{P}\left(\bigwedge_{i=1}^{k} A_i \wedge \bigwedge_{i=k+1}^{n} \neg A_i\right) = p^k (1-p)^{n-k}.$$

$(\mathbb{U}, \mathbf{P})$ is a probability structure symmetric in $\mathbb{U}$.[25]

Gaifman has noted an interesting property of the probability defined in this way: *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 $\mathbf{P}(A_1 \wedge A_2) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2)$. Gaifman proved an important consequence of this property: if $S_1, S_2 \in \{R\}_{\omega\omega}(\mathbb{U})$ and no individual constant occurs both in S_1 and S_2, then $\mathbf{P}(S_1 \wedge S_2) = \mathbf{P}(S_1) \cdot \mathbf{P}(S_2)$. Hence,

> 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.

Therefore, $(\mathbb{U}, \mathbf{P})$ is a probability structure that determine a **complete theory** in the usual, classical sense.[26]

We would reach the same result if the starting vocabulary contains more predicate symbols; if $\mathtt{L} = \{R_1, \ldots, R_k\}$, we should assign a real number $p_i \in (0,1)$ to each predicate symbol R_i, $1 \leqslant i \leqslant k$. The complete theory determined by such probability structure defines a classical $\mathtt{L}$-structure $\mathbf{R}$. For any sentence S: $\mathbf{R}(S) = 1$ iff $\mathbf{P}(S) = 1$. We call $\mathbf{R}$ the countable **random structure** over the vocabulary $\mathtt{L}$.[27]

[25] A probability model $(\mathbb{U}, \mathbf{P})$ is **symmetric** in $\mathbb{S} \subseteq \mathbb{U}$, if for every sentence F in $\mathtt{L}_{\omega\omega}(\mathbb{S})$,

$$\mathbf{P}(F(c_{i_1}, \ldots, c_{i_k})) = \mathbf{P}(F(\pi(c_{i_1}), \ldots, \pi(c_{i_k}))),$$

whenever π is a permutation of $\mathbb{S}$.

[26] A theory is any set of sentences. A theory is **complete** if it is consistent and for every sentence, either that sentence or its negation is provable.

[27] This structure $\mathbf{R}$ is unique up to isomorphism.

Based on the intuition of the introductory example, a random L-structure contains in some sense all finite L-structures as its substructures. In other words, when we consider n-element subsets of the universe of random structures, we will observe all possible types of n-element structures. In such a situation, it is natural to describe the types of n-element L-structures with open formulas using n variables. We use the following notations:

- $\mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var})$ is the set of all formulas with predicates from L, constant symbols for individuals from $\mathbb{U}$, and variables from Var;
- $\mathtt{L}^k_{\omega 0}$ and $\mathtt{L}^k_{\omega\omega}$ are the abbreviations for the sets $\mathtt{L}_{\omega 0}(\emptyset, \{x_1, \ldots, x_k\})$ and $\mathtt{L}_{\omega\omega}(\emptyset, \{x_1, \ldots, x_k\})$, respectively; in particular, $\mathtt{L}^0_{\omega 0}$ is a propositional language (which is non-trivial only if L contains 0-ary predicates, i.e. propositional letters); we allow $\mathtt{L}^0_{\omega 0}$ to include the two quantifier-free sentences $\top$ and $\bot$, denoting the always true and always false sentence, respectively.
- $\mathtt{L}_{\omega\omega} = \bigcup\limits_{k \geqslant 0} \mathtt{L}^k_{\omega\omega}$ is actually the set $\mathtt{L}_{\omega\omega}(\emptyset, \mathtt{Var})$, where $\mathtt{Var} = \{x_1, x_2, \ldots\}$

EXAMPLE 7. [28] Assume that the vocabulary consists of a binary predicate symbol $<$, and consider only the structures in which the interpretation of $<$ is a total order. Let A_n be a first-order sentence asserting *there are at least n elements*. Using n distinct variables, A_n can be written as

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

However, on total orders, A_n is equivalent to a sentence in $\{<\}^2_{\omega\omega}$. For example, A_4 can be written as

$$\exists x \exists y \, (x < y \wedge \exists x \, (y < x \wedge \exists y \, (x < y))).$$

It follows that on total orders the sentence B_n asserting that *there are exactly n elements* is also in $\{<\}^2_{\omega\omega}$, since it is equivalent to $A_n \wedge \neg A_{n+1}$.

[28] P. Kolaitis, M. Vardi, *Infinitary Logics and 0-1 Laws*. Information and Computation 98, pp. 258-294, 1992

Exercise 6. Let the vocabulary consist of a single binary predicate F. Using $n+1$ distinct variables it is easy to write the sentence asserting that *there is a path of length n from x to y*:

$$\exists x_1 \cdots \exists x_{n-1} \, (F(x, x_1) \wedge F(x_1, x_2) \wedge \cdots \wedge F(x_{n-1}, y)).$$

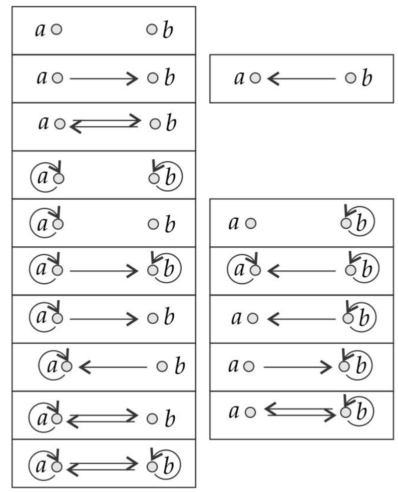
Prove that this property is expressible in $\{F\}^3_{\omega\omega}$.

DEFINITION 6. 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 L is the conjunction of all the formulas in a maximally consistent set of atomic formulas and negated atomic formulas in variables $\overline{x}$.[29] The logical constant $\top$ is the type in the empty tuple ov variables.

[29] In the presence of the equality sing, a type contains also equalities $x_i = x_j$ and inequalities $x_i \neq x_j$.

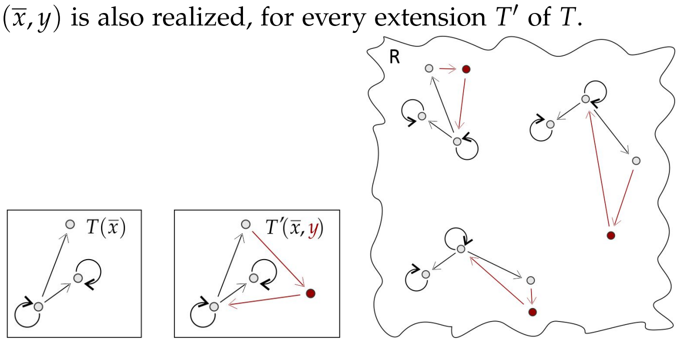
Every type $T(x_1,\dots,x_k)$ is a quantifier-free formula in $L^k_{\omega 0}$. Note that there are only finitely many distinct types in the variables $x_1,\dots,x_k$, if the vocabulary $\mathtt{L}$ is finite.

EXAMPLE 8. *Complete atoms* and *types* differ only in that the former use individual constants, and the latter use variables. Types in the variables $x_1,\dots,x_k$ over $\mathtt{L}$ correspond to the complete atoms of $\mathtt{L}(\{x_1,\dots,x_k\})$. As in the example 2 (page 9), let $\mathtt{L}$ contains only one 2-place predicate R. There are 16 types in the variables x_1, x_2 that corresponds to all $\mathtt{L}$-structures over a two-element universe.



$T_1(x_1,x_2)$ $\quad \neg R(x_1,x_1)\wedge\neg R(x_1,x_2)\wedge\neg R(x_2,x_1)\wedge\neg R(x_2,x_2)$
$T_2(x_1,x_2)$ $\quad \neg R(x_1,x_1)\wedge R(x_1,x_2)\wedge\neg R(x_2,x_1)\wedge\neg R(x_2,x_2)$
$T_3(x_1,x_2)$ $\quad \neg R(x_1,x_1)\wedge\neg R(x_1,x_2)\wedge R(x_2,x_1)\wedge\neg R(x_2,x_2)$
$\vdots$
$T_{16}(x_1,x_2)$ $\quad R(x_1,x_1)\wedge R(x_1,x_2)\wedge R(x_2,x_1)\wedge R(x_2,x_2)$

The random structure (i.e., the complete theory determined by the probability from Gaifman's simple example) is characterized by an infinite set of *extension axioms* which intuitively state that any type can be extended to any other possible type. More precisely, for every $k\geqslant 1$, the extension axiom E_k is a first-order sentence with k variables asserting that any $\mathtt{L}$-structure with less than k elements can be extended to any possible $\mathtt{L}$-structure with k elements. Wherever within the random structure $\mathbf{R}$ we recognize a type $T(\bar{x})$, we can find an element so that $T'(\bar{x},y)$ is also realized, for every extension T' of T.



DEFINITION 7. Let x_k be a new variable, different from all the variables in $\bar{x}=(x_1,\dots,x_{k-1})$. A type $T'(\bar{x},x_k)$ extends a type $T(\bar{x})$ if every conjunct of $T(\bar{x})$ is also a conjunct of $T'(\bar{x},x_k)$.[31] With each pair of types that T' extends T we associate an **extension axiom** $E_{T,T'}$ stating that

$$\forall\bar{x}\,\big(T(\bar{x})\Rightarrow\exists x_k\,T'(\bar{x},x_k)\big)\,.$$

If k is a positive integer, we write E_k for the conjunction of all extension axioms $E_{T,T'}$ with at most k variables. Note that E_k is a sentence in $\mathtt{L}^k_{\omega\omega}$.

Exercise 7. Write down the sentence E_2, if the vocabulary contains single 2-placed predicate symbol R.

[31] For example, $T'(x_1,x_2,x_3)$:
$\neg R(x_1,x_1)\wedge R(x_1,x_2)\wedge R(x_1,x_3)\wedge$
$\wedge\neg R(x_2,x_1)\wedge R(x_2,x_2)\wedge\neg R(x_2,x_3)\wedge$
$\wedge R(x_3,x_1)\wedge\neg R(x_3,x_2)\wedge\neg R(x_3,x_3)$
extends $T(x_1,x_2)$:
$\neg R(x_1,x_1)\wedge R(x_1,x_2)\wedge$
$\wedge\neg R(x_2,x_1)\wedge R(x_2,x_2)$

In each model of the formula E_k all possible types determined by formulas with $\leqslant k$ variables are realized. A classical structure satisfying E_k 'includes' all possible finite $\mathtt{L}$-structures with $\leqslant k$ elements, i.e, realize every type $T(x_1, \ldots, x_m)$, $m \leqslant k$. Of course, a type in variables $\overline{x}$ decides 'the truth' of every open formula $F(\overline{x})$.

THEOREM 6. Let k and m be two positive integers such that $m \leqslant k$. If $T(x_1, \ldots, x_m)$ is type over the vocabulary $\mathtt{L}$ and $F(x_1, \ldots, x_m)$ is a formula in $\mathtt{L}^k_{\omega\omega}$, 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 1. If S is a sentence of $L^k_{\omega\omega}$, then either $E_k \models S$ or $E_k \models \neg S$.[32]

[32] As a result, if $\mathbf{A}$ and $\mathbf{B}$ are two models of E_k, then $\mathbf{A} \equiv^k_{\omega\omega} \mathbf{B}$, which means $\mathbf{A}$ and $\mathbf{B}$ satisfy the same $L^k_{\omega\omega}$-sentences.

It is interesting to emphasize that the previous theorem also holds for formulas of a richer, infinitary language $\mathtt{L}^k_{\infty\omega}$. We use the notation $\mathtt{L}_{\infty\omega}$ to indicate infinitary languages which allow arbitrary, possibly infinite conjunctions (and hence disjunctions); $\mathtt{L}_{\omega_1\omega}$ refers to languages that allow only countable conjunctions and disjunctions. The set of formulas $\mathtt{L}_{\infty\omega}(\mathbb{U}, \mathtt{Var})$ is an extension of $\mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var})$, defined by the usual formation rules, and the following additional rule:

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

Given an $\mathtt{L}$-structure $(\mathbb{U}, \mathbf{M})$, the truth assignment $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$ extends to infinitary formulas by the rules:

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

We use $\mathtt{L}^k_{\infty\omega}$ as an abbreviation for $\mathtt{L}_{\infty\omega}(\emptyset, \{x_1, \ldots, x_k\})$; $\mathtt{L}^\omega_{\infty\omega} = \bigcup_{k \geqslant 1} \mathtt{L}^k_{\infty\omega}$.[33]

[33] The infinitary languages $\mathtt{L}^k_{\infty\omega}$, $k \geqslant 1$, and $\mathtt{L}^\omega_{\infty\omega}$ were introduced by Barwise (1977) as a tool for studying positive fixed-point logic on infinite structures. Since that time, however, these languages have had numerous uses and applications in theoretical computer science.

EXAMPLE 9. We continue Example 8. On total orders, properties such as *there are an even number of elements*, *the universe is finite*, etc., are expressible in $\{<\}^2_{\infty\omega}$. In general, if X is any set of positive integers, then the property *the cardinality of the total order is a member of X* is expressible in $\{<\}^2_{\infty\omega}$ by $\bigvee_{n \in X} B_n$.

Exercise 8. If F is a binary predicate, prove that the properties:
• *there is no cycle;*
• *x and y are connected by a path whose length is an even number* (see Exercise 6);
• *there is a path between every pair of elements* (connectivity);
are expressible in $\{F\}^3_{\infty\omega}$.

Fagin[34] realized that the extension axioms are relevant to the study of probabilities on finite structures. More precisely, given a finite set $\mathtt{L}$ of predicate symbols, he considered the sequence of *finite probability spaces with counting probability measure*[35] over the sample space $\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}$ consisting of all $\mathtt{L}$-structures over $\mathbb{U}_n = \{c_1, \ldots, c_n\}$, $n \geqslant 1$. The probability $\mathbf{P}_n : \mathtt{L}_{\omega\omega}(\emptyset) \to [0,1]$ is defined such that $\mathbf{P}_n(S)$ is the fraction of 'outcomes' from $\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}$ for which S is true:

[34] R. Fagin, *Probabilities on Finite Models*, The Journal of Symbolic Logic Vol. 41, No. 1, pp. 50-58, 1976

[35] A finite probability spaces with counting probability measure is any space of the form $(\Omega, 2^{\Omega}, \mu)$, where Ω is a finite set and $\mu(S) = \frac{\#S}{\#\Omega}$, $S \subseteq \Omega$.

$$\mathbf{P}_n(S) = \frac{\#[S]}{\#\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}} = \frac{\#\{\mathbf{M} \in \mathcal{M}_{\mathtt{L}(\mathbb{U}_n)} \mid \mathbf{M}(S) = 1\}}{\#\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}},\ S \in \mathtt{L}_{\omega\omega}(\emptyset).$$

The **asymptotic probability** $\mathbf{P}(S)$ is defined to be equal to the limit

$$\mathbf{P}(S) = \lim_{n\to\infty} \mathbf{P}_n(S),$$

provided this limit exists. It turns out that the asymptotic probability coincides with the Gaifmain's probability determining the random structure.

All the considerations in this subsection are related to the uniform probability on $\mathcal{M}_{\mathtt{L}(\mathbb{U}_n)}$. However, there is a well developed study of random structures under different probability measures. . . .

PROPOSITION 6. The asymptotic probability of all extension axioms is equal to 1, that is $\mathbf{P}(E_k) = 1$, for every $k \geqslant 1$. (E_k is the conjunction of all extension axioms $E_{T,T'}$ with at most k variables.)

EXERCISE 9. Estimate $\mathbf{P}_{10}(E_3)$; $\mathbf{P}_k(E_3)$, for $k > 10$, etc.

THEOREM 7. [the 0-1 low] If S is a sentence in $\mathtt{L}_{\omega\omega}(\emptyset)$, then the asymptotic probability $\mathbf{P}(S)$ exists and it is equal either 0 or 1.

PROOF. If S is a sentence, then by Corollary 1 (page 21),

$$E_k \models S \text{ or } E_k \models \neg S.$$

In the first case we have that $\mathbf{P}(S) = 1$ and in the second $\mathbf{P}(\neg S) = 1$, since $\mathbf{P}(E_k) = 1$. □

In fact, Glebskii *et al.*[37] proved the 0-1 low, several years earlier than Fagin, using a different approach – a certain quantifier elimination method. This approach is based on a quantifier-elimination theorem for $\mathtt{L}^k_{\omega\omega}$ on models of E_k.

[37] Y. Glebskii, D. Kogan, M. Liogonkh, V. Talanov, *Range and degree of realizability of formulas in the restricted predicate calculus*, Cybernetics 5, pp. 142-154, 1969

THEOREM 8. Let k be a positive integer and let $F(x_1, \ldots, x_m)$ be a formula of $\mathtt{L}^k_{\omega\omega}$ ($m \leqslant k$). Then there is a quantifier-free formula $B(x_1, \ldots, x_m)$ of $\mathtt{L}^k_{\omega 0}$ such that:

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

PROOF. Let $\mathcal{B}_F$ be the set of all types $T(\overline{x})$ for which there is a structure $\mathbf{D}$ such that

$$\mathbf{D} \models E_k \wedge \exists \overline{x}\, (F(\overline{x}) \wedge T(\overline{x})).$$

We claim that the required formula $B(x_1, \ldots, x_m)$ iz

$$\bigvee_{T \in \mathcal{B}_F} T(x_1, \ldots, x_m).$$

Note that $B(x_1, \ldots, x_m)$ is a quantifier-free formula of $L^k_{\omega\omega}$, because the vocabulary L is finite, and consequently there are finitely many distinct types in the variables $x_1, \ldots, x_m$. Moreover, it follows from the definitions that

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

For the other direction, let $\mathbf{D}$ be a model of E_k, and let $a_1, \ldots, a_m$ be elements from the universe of $\mathbf{D}$, such that $\mathbf{D} \models B(a_1, \ldots, a_m)$. Then, there is a type T in the set $\mathcal{B}_F$ such that $\mathbf{D} \models T(a_1, \ldots, a_m)$. By Theorem 6 (page 21), exactly one of the following two statements holds:

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

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

Since, T is a type in the set $\mathcal{B}_F$, the second statement (2) is impossible and, hence $\mathbf{D} \models F(a_1, \ldots, a_m)$. □

Taking $\overline{x}$ to be empty, the previous theorem says that each first-order sentence with k variables collapses to $\top$ or $\bot$ almost everywhere, leading to the 0-1 law. The 0-1 low leads to many *almost everywhere* variants of important properties of logics.

When we restrict our considerations only on finite structures, Theorem 8 implies that the language $\mathrm{L}^k_{\omega\omega}$ in a sense reduces to the language $\mathrm{L}^k_{\omega 0}$ (weakly) almost everywhere[38], i.e. $\mathrm{L}^k_{\omega\omega}$ admits almost everywhere quantifier elimination: every formula F in $\mathrm{L}^k_{\omega\omega}$ is equivalent almost everywhere to a formula F' in $\mathrm{L}^k_{\omega 0}$.[39] With this observation, we touch on an interesting direction of research in final model theory – the comparative study of the expressive power of different logics.

It is worth mention one very important consequence of the 0-1 low, although we do not go into details:[40] By well-known Trakhtenbrot's Theorem we *cannot* effectively decide whether a first-order sentence is valid in all finite models. But we *can* deside whether a first-order sentence is valid in almost all models.

All previous results indicate that for random structures it makes more sense to consider the so-called *counting quantifiers* instead of the classical quantifiers $\forall$ and $\exists$.[41] In the next subsection we introduce languages with a special kind of such quantifiers – *probabilistic quantifiers*.

[38] in symbols $\mathrm{L}^k_{\omega\omega} \leq_{w.a.e.} \mathrm{L}^k_{\omega 0}$

[39] Two formulas $F_1(\overline{x})$ and $F_2(\overline{x})$ are **equivalent almost everywhere**, written $F_1(\overline{x}) \equiv_{a.e.} F_2(\overline{x})$ if the sentence $\forall \overline{x}\,(F_1(\overline{x}) \Leftrightarrow F_2(\overline{x}))$ has asymptotic probability 1.

[40] J. Väänänen, A Short Course on Finite Model Theory, http://www.math.helsinki.fi/logic/people/jouko.vaananen/shortcourse.pdf

[41] For a deeper study of finite model theory and the corresponding logics (Logics with Counting etc.), we recommend the book: L. Libkin, *Elements of Finite Model Theory*, Springer, 2012, https://homepages.inf.ed.ac.uk/libkin/fmt/

2.2. *Probabilistic quantifiers*

The random L-structure $\mathbf{R}$ has the following nice properties: a sentence S from $\mathrm{L}_{\omega\omega}$ is true in $\mathbf{R}$ iff it is true in almost all finite models. A weaker version of this property is used to define a special class of structures, called pseudo-finite structures, that gives an 'infinitary' approach to finite model theory

DEFINITION 8. [42] A model is **pseudo-finite** if every first-order sentence true in the model is also true in a finite model.

[42] A. Pillay, *Pseudofinite Model Theory*, 2015

Any random structure is an example of a pseudo-finite structure. Pseudo-finiteness can be defined in several equivalent ways. The proposition below exhibits three ways. If $\mathtt{L}$ is a vocabulary, we use $\Gamma_{\mathtt{L}}$ to denote the first-order theory of all finite $\mathtt{L}$-structures, i.e. the set of $\mathtt{L}$-sentence that are true in all finite $\mathtt{L}$-structure.

PROPOSITION 7. For every $\mathtt{L}$-structure $\mathbf{M}$, the following conditions are equivalent:

- $\mathbf{M} \models \Gamma_{\mathtt{L}}$;
- $\mathbf{M}$ is a pseudo-finite structure;
- There is a set $\{\mathbf{M}_i \mid i \in I\}$ of finite $\mathtt{L}$-structures and an ultrafilter[43] $\mathfrak{U}$ on I such that $\mathbf{M} \equiv \prod_{\mathfrak{U}} \mathbf{M}_i$.

[43] A ultrafilter on I can be identified with a finitely additive $\{0,1\}$-valued probablity measure on the power set 2^I.

The third item of the previous theorem shows that certain parts of nonstandard analysis will become useful. We will not go into details, but only highlight some intuitive basics. One can think of pseudo-finite structures as structures with a huge (hyperfinite) number of elements, much larger relative to the 'sizes' that can be described by quantifier-free formulas.

What is particularly important for our story is that pseudofinite structures admit a useful class of probability measures that reflect counting measures on finite structures. Roughly speaking, this class of probability measures is determined by a probability on the set of all formulas $\mathtt{L}_{\omega\omega}$, including formulas with free variables. The introductory example to this section shows that:

counting models satisfying a sentence

can be reduced to

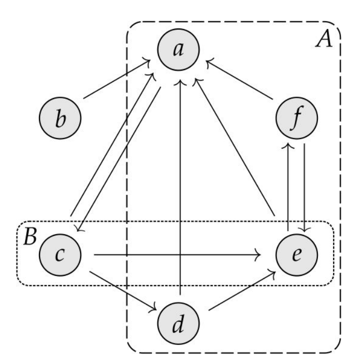
counting tuples in a 'huge' random structure satisfying an open formula.

Hence, there are good reasons to extend (without any essential change) Definition 3 (page 8) to the set of all formulas:

A **probability** is a function $\mathbf{P} : \mathtt{L}_{\omega\omega}(\mathbb{U}, \mathtt{Var}) \to [0,1]$ that satisfies the properties P1, P2, P3.

The following examples illustrate a typical way of defining probability on a set of all formulas.

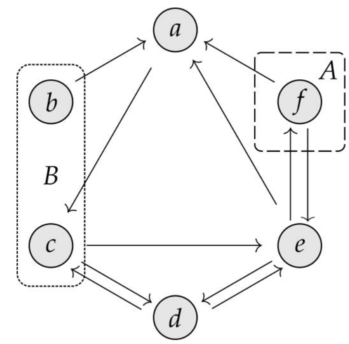
EXAMPLE 10. Let us back to the structure from the very beginning: $\mathtt{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$, $\mathbb{U} = \{a,b,c,d,e,f\}$ and $\mathbf{M}$ is defined by the diagram on the right. Using the counting probability measure $\mathbf{P}_n$ on $\mathbb{U}^n$, we can find the probability of all open formula with n free variables, for all $n \geqslant 1$. For instance:
$\mathbf{P}_1(A(x)) = \frac{4}{6}$, $\mathbf{P}_2(F(x,y)) = \frac{11}{36}$, $\mathbf{P}_1(\exists x\, F(x,y)) = \frac{5}{6}$,
$\mathbf{P}_2(F(x,y) \Rightarrow F(y,x)) = \frac{4}{36}$, $\mathbf{P}_3(B(x) \wedge B(y) \wedge B(z)) = \frac{8}{216}$ etc.



L-structure $\mathbf{M}$

Let us consider one more L-structure $(\mathbb{U}, \mathbf{M}')$, and set a non-uniform distribution of the individual 'weights':

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



L-structure $\mathbf{M}'$

For $n > 1$, we assume the product distribution on n-tuples:

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

Now, we can calculate the probabilities of open formulas:
$\mathbf{P}'_1(A(x)) = 0$,
$\mathbf{P}'_1(\exists x\, F(x, y)) = \mathbf{P}'_1(\{a, c, d, e, f\}) = 0.1 + 0.3 + 0.4 + 0.1 + 0 = 0.9$,
$\mathbf{P}'_2(B(x) \wedge B(y)) = \mathbf{P}'_2(\{(b, b), (b, c), (c, b), (c, c)\}) = 0.16$ etc.

Finally, suppose that our 'confidence' in these two structures is determined by the following distribution:

$$\mu : \begin{pmatrix} \mathbf{M} & \mathbf{M}' \\ 75\% & 25\% \end{pmatrix}$$

We obtain a third class of probability measures which is a mixture of the previous two classes:

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

In the case when we have several classical models, each of which is assigned the weight, the probability of open formula could be regarded as the 'expected' (average) probability of that formula.

A representation theorem, similar to Theorem 2 (page 10), shows that any probability on the set of formulas may be understood as a result of two consecutive drawings[46]:

- 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 $\mathbf{P}^i_n$, $n \geqslant 1$.

[46] J. Łoś, *Semantical interpretation of the probability of formulas*, Studia Logica T. 14, pp. 183-196, 1963

Having in mind Definition 4 (page 12), any probability on the set of formulas can also be considered as a generalization of the classical structure. An important class of such generalized structures consists of a special kind of multidimensional probability space with a probability for each dimension, $(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$, where:

- $(\mathbb{U}, \mathbf{M})$ 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 most) 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[47] which we omit.

Such structures are named **graded probability structures**, and we omit details for the reason that will be discussed later. Languages suitable for such structures are introduced by Keisler.[48]

[47] Fubini's properties, etc. We refer to H. J. Keisler, *Probability quantifiers*, In J. Barwise and S. Feferman, editors, Model Theoretical Logics, Chapiter XIV, Springer, 1985
[48] J. Keisler, *Hyperpinite Model Theory*, Studies in Logic and the Foundations of Mathematics, Volume 87, pp. 5-110, 1977

Logics $\mathsf{L}_{\omega \mathsf{P}}$ *and* $\mathsf{L}_{\omega_1 \mathsf{P}}$

First we introduce a probability language $\mathsf{L}_{\omega \mathsf{P}}$, and its infinitary extension $\mathsf{L}_{\omega_1 \mathsf{P}}$, in which the classical quantifiers are replaced by probability quantifiers of the form $(\mathsf{P}\overline{x} \geqslant r)$ and $(\mathsf{P}\overline{x} > r)$, 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 $(\mathsf{P}\overline{x} \geqslant r)F$ and $(\mathsf{P}\overline{x} > r)F$ are formulas too.

Graded probability structures provide semantics for such set of formulas. The probability formula $(\mathsf{P}\overline{x} > r)F(\overline{x})$ is true in a graded probability structure, denoted by $(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1} \models (\mathsf{P}\overline{x} > r)F(\overline{x})$, iff $\mathbf{P}_n(F(\overline{x})) > r$; $(\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1} \models (\mathsf{P}\overline{x} \geqslant r)F(\overline{x})$ iff $\mathbf{P}_n(F(\overline{x})) \geqslant r$.

EXAMPLE 11. Here is a list of several formulas which are true in any graded structure:
• $(\mathsf{P}\overline{x} \geqslant r)\, F(\overline{x}) \Rightarrow (\mathsf{P}\overline{x} \geqslant s)\, F(\overline{x}),\ s < r$
• $(\mathsf{P}\overline{x} > r)\, F(\overline{x}) \Rightarrow (\mathsf{P}\overline{x} \geqslant r)\, F(\overline{x})$
• $(\mathsf{P}\overline{x} \geqslant r)\, F(\overline{x}) \Rightarrow (\mathsf{P}\overline{y} \geqslant r)\, F(\overline{y})$
• $(\mathsf{P}\overline{x} \geqslant 0)\, F(\overline{x})$
• $(\mathsf{P}\overline{x} \geqslant 1)\, \neg(A(\overline{x}) \wedge B(\overline{x})) \wedge (\mathsf{P}\overline{x} \geqslant r)\, A(\overline{x}) \wedge (\mathsf{P}\overline{x} \geqslant s)\, B(\overline{x}) \Rightarrow$
$\Rightarrow (\mathsf{P}\overline{x} \geqslant r + s)\, (A(\overline{x}) \vee B(\overline{x}))$
• $(\mathsf{P}\overline{x} \leqslant r)\, A(\overline{x}) \wedge (\mathsf{P}\overline{x} \leqslant s)\, B(\overline{x}) \Rightarrow (\mathsf{P}\overline{x} \leqslant r + s)\, (A(\overline{x}) \vee B(\overline{x}))$
Note also:
• If $S \Rightarrow F(\overline{x})$ is true in a graded structure, then $S \Rightarrow (\mathsf{P}\overline{x} \geqslant 1)\, F(\overline{x})$ is also true in that structure.

The reason we do not detail the definition and variety of graded probability structures is that these structures can be 'approximated', *up to* $\mathsf{L}_{\omega_1 \mathsf{P}}$, by almost any sequence of finite structures, i.e. by a hyperfinite model. This result is a consequence of a model-theoretic form of the well-known law of large numbers.

EXAMPLE 12. Assume we are given an urn containing 1000 different balls, $\mathbb{U} = \{b_1, \ldots, b_{1000}\}$. 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). In this way, we obtain an ordered sample

$$\overline{b} = (b_1, \ldots, b_{100}),$$

and for each color, we find the percentage of balls in this color:

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

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

The lows of large numbers support the intuition that it should be $\mathbf{p}_{\overline{b}}(R(x)) \approx 0.5$, $\mathbf{p}_{\overline{b}}(B(x)) \approx 0.3$, $\mathbf{p}_{\overline{b}}(G(x)) \approx 0.2$. In other words the grad. structure $(\{b_1,\dots,b_{1000}\},\cdots,\mathbf{P})$ is very 'similar' to $(\{b_1,\dots,b_{100}\},\cdots,\mathbf{p}_{\overline{b}})$.

To state the lows of large numbers for our logic, we need the notion of a finite sample of a graded probability structure.

DEFINITION 9. Let $\mathcal{G} = (\mathbb{U},\mathbf{M},\mathbf{P}_n)_{n\geqslant 1}$ be a graded probability structure for L, and let $\overline{a}^k = (a_1,\dots,a_k) \in \mathbb{U}^k$ 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 $\{a_1,\dots,a_k\}$, classical part is the substructure over $\{a_1,\dots,a_k\} \subseteq \mathbf{M}$, and probabilities $\mathbf{p}_n$ are given by:

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

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

$$\mathbf{p}_3(F(x,y,z)) = \frac{\sum_{1\leqslant i,j,\ell\leqslant k} F(a_i,a_j,a_\ell)}{k^3} \text{ etc.}$$ [49]

[49] Thus, the finite set $\{a_1,\dots,a_k\}$ has measure one in $\mathbf{M}(\overline{a}^k)$, and the measure of a singleton $\{a\}$ is $\frac{1}{k}$ multiplied by the number of occurrences of a in the sequence $\overline{a}^k$.

Let us first illustrate the logical variant of the weak law of large numbers in the simplest case.

PROPOSITION 8. Let $\mathcal{G} = (\mathbb{U},\mathbf{M},\mathbf{P}_n)_{n\geqslant 1}$ be a graded probability structure for L that satisfies $(\mathsf{P}x > r)\, F(x)$, where $F(x)$ is a quantifier-free formula. Then

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

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

SKETCH OF PROOF. Let $\varepsilon > 0$ be arbitrary. We have to show that there exists k_0 such that for all $k \geqslant k_0$, it holds

$$\mathbf{P}_k \left\{ \overline{a}^k \in \mathbb{U}^k \mid \mathcal{G}(\overline{a}^k) \models (\mathsf{P}x > r)\, F(x) \right\} > 1 - \varepsilon$$

If $\mathbf{P}_1(F(x)) = q$, then $q > r$ (since $\mathcal{G}$ satisfies $(\mathsf{P}x > r)\, F(x)$). Let δ be any number such that $0 < \delta < q - r$. By the weak low of large numbers, there is k_0 such that for any $k \geqslant k_0$, and any k-tuple $\overline{a}^k$,

$$\mathbf{P}_k\{\overline{a}^k \mid \left| \frac{F(a_1)+\cdots+F(a_k)}{k} - q \right| < \delta\} > 1 - \varepsilon.$$

Thus,

$$\mathbf{P}_k\{\overline{a}^k \mid r < \frac{F(a_1)+\cdots+F(a_k)}{k}\} > 1 - \varepsilon,$$

that is

$$\mathbf{P}_k\{\overline{a}^k \mid \mathcal{G}(\overline{a}^k) \models (\mathsf{P}x > r)\, F(x)\} > 1 - \varepsilon$$

□

Applying the induction on the length of formulas, we obtain a general result.

THEOREM 9. [Weak Law of Large numbers for $\mathsf{L}_{\omega_1\mathsf{P}}$] Let $\mathcal{G} = (\mathbb{U}, \mathbf{M}, \mathbf{P}_n)_{n \geqslant 1}$ be a graded probability structure for L satisfying

$$(\mathsf{P}\overline{x}_1 > r_1) \cdots (\mathsf{P}\overline{x}_n > r_n)B,$$

where B is a finite quantifier-free formula of L. Then

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

The property $\mathcal{G}(\overline{a}^k) \models (\mathsf{P}x \geqslant r)\, F(x)$ is expressible by a finite quantifier-free formula $B(\overline{x}^k)$ of L. If $\ell = [kr] + 1$, let $B(x_1, \ldots, x_k)$ be the formula

$$\bigvee_{1 \leqslant i_1 < \cdots < i_\ell \leqslant k} (F(x_{i_1}) \wedge \cdots \wedge F(x_{i_\ell})).$$

Of course: $\mathbf{M} \models B(a_1, \ldots, a_k)$ iff $\mathcal{G}(\overline{a}^k) \models (\mathsf{P}x \geqslant r)\, F(x)$. This observation, together with the weak law of large numbers for $\mathsf{L}_{\omega_1\mathsf{P}}$, forms the basis for the proof of an important theorem which has numerous consequences.[50]

[50] D. Hoover, *A Normal Form Theorem for $\mathsf{L}_{\omega_1\mathsf{P}}$, with Applications*, The Journal of Symbolic Logic, Vol. 47, No. 3, pp. 605-624, 1982

THEOREM 10. [Normal Form Theorem] Every formula $F(\overline{x})$ of $L_{\omega_1\mathsf{P}}$ is equivalent to a countable boolean combination of formula of the form $(\mathsf{P}\overline{x} \geqslant r)\, B(\overline{x}, \overline{y})$, where $B(\overline{x}, \overline{y})$ is a finite conjunction of atomic formulas of L.

Logic $\mathsf{L}_{\mathcal{C}\mathsf{E}}$ ($\mathsf{L}_{\mathcal{C}\int}$)

Besides the language $\mathsf{L}_{\omega\mathsf{P}}$ Keisler has developed another language which is much more suitable for the description of probabilistic (physical) phenomena. The great expressive power of this new language is achieved by using **real-valued expressions** of the form $\mathsf{E}_x\, F(x)$, which represent the probability of the formula $F(x)$, i.e., 'the expected truth value of the formula $F(x)$'. The logic of 'expected truth values' can be regarded as a multivalued logic.

The language $\mathsf{L}_{\mathcal{C}\,\mathsf{E}}$, where $\mathcal{C}$ is a set of continuous real functions, has the following symbols (in addition to predicate and individual symbols):

- individual variables: $x, y, z, x_1, \ldots$;
- equality sign: $=$ (optional);
- connectives: an n-ary connective C for each continuous real function $C : \mathbb{R}^n \to \mathbb{R}$ from $\mathcal{C}$;[51]
- thee quantifiers: E_x, $\sup_x$, $\inf_x$.[52]

These symbols are used to build the set of terms (or *real-valued formulas*).

[51] There are uncountable many connectives, and this may be problematic. To avoid this anomaly, one often considers a countable set of connectives which allows to construct arbitrarily good approximations of every continuous connective.

[52] Keisler used $\int \cdots \mathrm{d}x$ instead of E_x

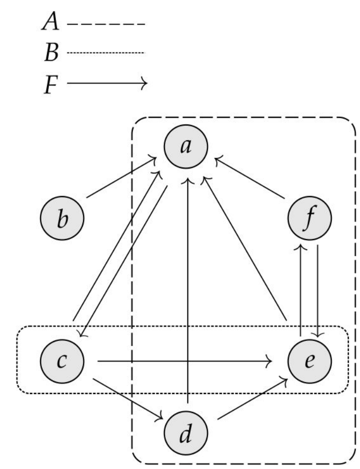
DEFINITION 10. The set of **terms** is the smallest set which satisfies the following rules:

- 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 $\overline{x}$. A term with no free variables is called a constant term.

EXAMPLE 13. Let us back to the graded structure $\mathcal{G}$ for $\mathtt{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$ defined in Example 10:

- $\mathbb{U} = \{a, b, c, d, e, f\}$;
- $\mathbf{M}$ is defined by the diagram on the right;
- $\mathbf{P}_n$ is the counting probability measures on $\mathbb{U}^n$, for all $n \geqslant 1$.



Let us now find the values of several terms:
$\mathcal{G}(\mathsf{E}_x A(x)) = \frac{A(a)+A(b)+A(c)+A(d)+A(e)+A(f)}{6} = \frac{4}{6}$
$\mathcal{G}(\sup_x (A(x) + B(x))) = 2$
$\mathcal{G}(\mathsf{E}_x F(x,y)) = \mathcal{G}\left(\frac{F(a,y)+F(b,y)+F(c,y)+F(d,y)+F(e,y)+F(f,y)}{6}\right)$;
$\mathcal{G}(\mathsf{E}_x F(x,a)) = \frac{5}{6} \quad \mathcal{G}(\mathsf{E}_x F(x,b)) = 0 \quad \mathcal{G}(\mathsf{E}_x F(x,c)) = \frac{1}{6} \ldots$
$\mathcal{G}(\max_y \mathsf{E}_x F(x,y)) = \frac{5}{6}$ etc.

As illustrated in the previous example, any graded probability structure defines the values of all $\mathtt{L}_{\mathcal{C}\,\mathsf{E}}$-terms. However, the language $\mathtt{L}_{\mathcal{C}\,\mathsf{E}}$ can also be used to describe so-called real-valued structures. Let us remind that a (classical) two-valued L-structure is of the form $(\mathbb{U}, \mathbf{M})$, where $\mathbf{M}$ is a truth assignment to atomic formulas, $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \{0, 1\}$. A **real-valued L-structure** is a pair $(\mathbb{U}, \mathbf{M})$, where $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \mathbb{R}$, i.e. $\mathbf{M}$ assigns a real number to every atomic L-sentence.

This kind of generalization also leads to a corresponding generalization of the notion of graded real-valued structure. The rules for calculating the values of terms remain the same: E.g.,

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

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

EXAMPLE 14. The symbols from $\mathtt{L} = \{A^{(1)}, B^{(1)}, F^{(2)}\}$ could be interpreted as *real-valued properties* of individuals from a universe; say:

- $A(i)$ represents a quantitative characteristic (e.g., 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;

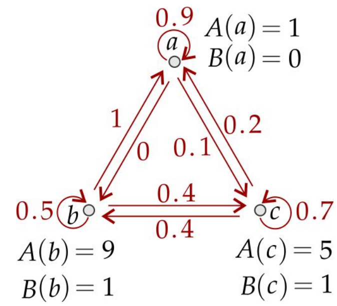
- $F(i,j)$ represents a degree of 'confidence' of i in j.

Consider the following simple graded real-valued structure:

- $\mathbb{U} = \{a, b, c\}$;
- $\mathbf{M} : \mathtt{L}(\mathbb{U}) \to \mathbb{R}$ is defined by:

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

  which is represented by the diagram in the margin;



- $\mathbf{P}_n$ is the counting probability measures on $\mathbb{U}^n$, for all $n \geqslant 1$.

Let us determine the values of several terms:

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

$\sup_x (A(x) + B(x)) = \sup\{A(a) + B(a), A(b) + B(b), A(c) + B(c)\} = 10$

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

$\mathsf{E}_x F(x,b) = \frac{F(a,b)+F(b,b)+F(c,b)}{3} = \frac{0+0.5+0.4}{3} = 0.3$

$\mathsf{E}_x F(x,c) = \frac{F(a,c)+F(b,c)+F(c,c)}{3} = \frac{0.1+0.4+0.7}{3} \approx 0.4$

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

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

Now, we briefly summarize this section and announce the issues that we will deal with in the next section.

✔ Defining probability on the set of all formulas for a vocabulary $\mathtt{L}$ led us to graded probabilistic $\mathtt{L}$-structures – classical $\mathtt{L}$-structures $(\mathbb{U}, \mathbf{M})$ extended by a sequence of probabilities $\mathbf{P}_n$, $n \geqslant 1$, for each dimension $\mathbb{U}^n$.

✔ There are several languages suitable for describing graded prob. structures: $\mathtt{L}_{\omega\mathsf{P}}$, $\mathtt{L}_{\omega_1\mathsf{P}}$, $\mathtt{L}_{\mathcal{C}\mathsf{E}}$ etc.

✔ Any graded structure can be well approximated by a hyperfinite structure, so that $\mathtt{L}_{\omega_1\mathsf{P}}$- properties are preserved. If $(\mathbb{U}, \cdots, \mathbf{P})$ is a grad. structure, and $\bar{b} = (b_1, \ldots, b_H)$ an ordered sample, where H is a 'huge' integer, with

$$\mathbf{p}_{\bar{b}}(F(x)) = \frac{F(b_1) + \cdots + F(b_H)}{H},$$

then $(\mathbb{U}, \cdots, \mathbf{P})$ and $(\{b_1, \ldots, b_H\}, \cdots, \mathbf{p}_{\bar{b}})$ satisfy 'almost' the same sentence from $\mathtt{L}_{\omega_1\mathsf{P}}$.

? 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.[55] Hoover (1978) found a simple and natural set of axiom and rule schemes for $\mathtt{L}_{\omega_1\mathsf{P}}$ ($\mathtt{L}_{\mathcal{C}\mathsf{E}}$), and proved the completeness theorem with respect to the class of graded probability structures.

[55] Several valid formulas from $\mathtt{L}_{\omega\mathsf{P}}$ are given in the example 11 (page 26).

> It is conventional to begin the study of new logic by proving a completeness theorem. By giving a complete set of axioms and rules one shows that *whether a given sentence is a theorem* is to some degree independent of the model of set theory one works in. Once this is done, the model construction used to prove the completeness theorem is generally of more practical value than the theorem itself. (D. N. Hoover, *Probability logic*, Ann. Math Logic 14, 287-313, 1978)

In the next section, we will consider completeness problems in a somewhat simpler context.

3. *Axiomatization issues*

MOTIVATING EXAMPLE. First-order language is the most common language for talking about relational structures. However, there are many alternative languages suitable for describing certain, specific aspects of relational structures. Modal languages provide an internal, local perspective on relational structure.[56]. Let us briefly present the *basic modal language* whose symbols are: 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.[57]

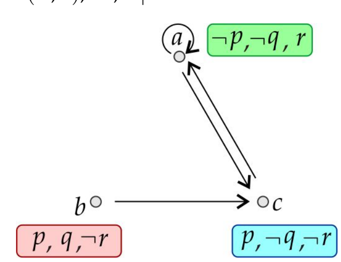
A model for this language is a triple $(\mathbb{W}, R, [\cdot])$, where
• $\mathbb{W}$ is a non-empty set (of words);
• R is a binary relation on $\mathbb{W}$;
• $[\cdot]$ is a function (a valuation) assigning to each propositional letter p a subset $[p] \subseteq \mathbb{W}$.

Such an model can be viewed as a relational structure consisting of a domain, a single binary relation, and the unary relation given by $[\cdot]$:

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

The modal language is interpreted by means of the following inductive definition of the notion *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 $R(w, v)$, $\mathbf{W}, v \models F$.[58]

For example, consider the model $\mathbf{W}$ defined by the figure on the right.



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

In this section we introduce a probability propositional logic and a class of appropriate models that are quite similar to the models of modal logic. The difference is that a binary relation is replaced by its probabilistic generalization, a special kind of real-valued binary relation.

[56] P. Blackburn, M. de Rijke, Y. Venema, *Modal logic*, Cambridge University Press, 2010

[57] We also have a dual operator $\Box$ ('box'; *necessary*) for the diamond which is defined by $\Box A :\equiv \neg\Diamond\neg A$.

[58] $\mathbf{W}, w \models \Box F$ iff for **all** $v \in \mathbb{W}$ such that $R(w, v)$, $\mathbf{W}, v \models F$.

3.1. *Markov process*

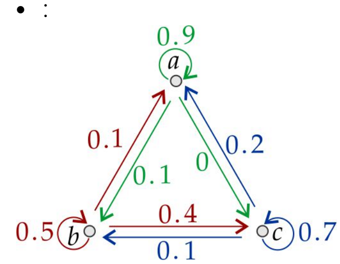
There are a wide range of probabilistic structures that come from different areas such as theoretical computer science, artificial intelligence, economics, game theory and so on.[60] Although there are differences among these structures, roughly speaking, in each case we deal with a structure that is a Markov process. We begin with a very simple example of the finite process that illustrates what kind of structures we deal with.

[60] Here are some influential references.

- 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
- $\vdots$

EXAMPLE 15. Consider the discrete space $(\mathbb{W}, 2^{\mathbb{W}})$ with three elements, $\mathbb{W} = \{a, b, c\}$. The elements of $\mathbb{W}$ will be called **worlds**. Assume that each world has its own probability measure over $\mathbb{W}$; i.e., a probability distribution is attached to each world:

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



Each world separately measures subsets (events) of $\mathbb{W}$:

$$\mu_a(\{a, c\}) = 0.9,\ \mu_b(\{a, c\}) = 0.5,\ \mu_c(\{a, c\}) = 0.9.$$

It is useful to imagine that this structure represents the successive execution of an experiment with possible elementary outcomes a, b, c. The value $\mu_i(j)$ can be viewed as the probability that outcome j occurs after i. For $X \subseteq \mathbb{W}$, $\mu_i(X)$ represents the probability that the event X occurs after the outcome i.

We can also consider so-called iterated probabilistic statements. For example, one can ask what is the probability that after a an outcome occurs after which c occurs with at least 40% of probability:

$$\mu_a\{i \mid \mu_i\{c\} \geqslant 0.4\} = \mu_a\{b, c\} = 0{,}1$$

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 function (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, the probabilities $\mu_w : 2^{\mathbb{W}} \to [0, 1]$, $w \in \mathbb{W}$, are defined by $\mu_w(X) \stackrel{\text{def}}{=} \sum_{w' \in X} \mu(w, w')$.

In the general case, in addition to the space $(\mathbb{W}, \mathcal{F})$ we consider one more measurable space – the space over the set $\Pi_\sigma(\mathbb{W}, \mathcal{F})$ of all probability measures defined on $(\mathbb{W}, \mathcal{F})$ which is endowed with the σ-field generated by the sets $M_{X,r} = \{m \in \Pi_\sigma(\mathbb{W}, \mathcal{F}) \mid m(X) \geq r\}$, for $X \in \mathcal{F}$ and $r \in [0,1]$.

DEFINITION 11. A Markov process is a triple $(\mathbb{W}, \mathcal{F}, \mu)$, where

- $\mathbb{W}$ is a nonempty set (of individuals; agents; worlds; states; atomic events; outcomes; etc.)
- $\mathcal{F}$ is a σ-algebra over $\mathbb{W}$.
- $\mu : \mathbb{W} \to \Pi_\sigma(\mathbb{W}, \mathcal{F})$ is a measurable function[62] from $(\mathbb{W}, \mathcal{F})$ to the space $\Pi_\sigma(\mathbb{W}, \mathcal{F})$.

[62] Given two measurable spaces $(W, \mathcal{F})$ and $(W', \mathcal{F}')$, a function $f : W \to W'$ is measurable if $f^{-1}(Y) \in \mathcal{F}$, for every $Y \in \mathcal{F}'$.

PROPOSITION 9. $\mu : (\mathbb{W}, \mathcal{F}) \to \Pi_\sigma(\mathbb{W}, \mathcal{F})$ is measurable iff for all $X \in \mathcal{F}$, the function $w \mapsto \mu_w(X)$ (from $\mathbb{W}$ to $[0,1]$) is measurable.

For simplicity, in what follows we ignore the prefix/index 'σ' and consider only processes whose measures are finite-additive only.

Markov process can be regarded in a number of ways – as a *probabilistic machine*, a *transition system*, a *real-valued relation structure* etc. These different views approve practical importance of such systems, and also various terminology that is in use. Our intuition could be based on the following:

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.

In point (2), $\mathcal{F}$ could be a set of sentences of any predicate language $\mathtt{L}$, but we will stick to the simplest case when $\mathtt{L}$ contains only 0-placed predicates, i.e. propositional letters. In this case, the set of $\mathtt{L}$-sentences is actually a set of propositional formulas. Point (3) suggests that in addition to the $\mathtt{L}$-formulas, we should also consider estimates of the probabilities of formulas: *the probability of X is at least r*. Such an intuition leads us naturally to a language $\mathtt{L}_\mathsf{P}$ that is suitable for formalizing the agents' beliefs.[63]

[63] $\mathtt{L}_{\mathcal{C}\mathsf{E}}$ could be a suitable, but more complex language for Markov processes.

The symbols for $\mathtt{L}_\mathsf{P}$ are:

- propositional letters from a countable vocabulary $\mathtt{L} = \{p^{(0)}, q^{(0)}, r^{(0)}, p_1^{(0)}, \ldots\}$;
- the logical constant 'true' $\top$:
- the classical connectives: $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$
- modal-like probabilistic operators $\mathsf{P}_{\geqslant r}$, for every $r \in [0,1] \cap \mathbb{Q}$,[64] with the intended meaning *the probability is at least r*.

[64] $\mathbb{Q}$ is the set of rational numbers.

DEFINITION 12. The set of formulas $\mathtt{L}_\mathsf{P}$ 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 $\mathsf{P}_{\geqslant r}F$ is a formula.

DEFINITION 13. An LPP-structure is a quadruple $\mathbf{W} = (\mathbb{W}, \mathcal{F}, \mu, [\cdot])$, where $(\mathbb{W}, \mathcal{F}, \mu)$ 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 $[\cdot] : \mathtt{L} \to \mathcal{F}$ 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.
- $[\mathsf{P}_{\geqslant r}F] = \{w \in \mathbb{W} \mid \mu(w, [F]) \geqslant r\}$

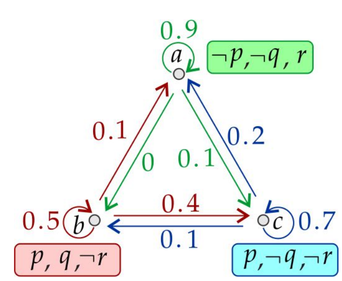
Thus, every $\mathtt{L}_\mathtt{P}$-formula F defines the measurable set $[F] \in \mathcal{F}$.

DEFINITION 14. 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 $w \in [F]$ ($w \notin [F]$).

The following properties are obvious:

- $\mathbf{W}, w \models \top$
- $\mathbf{W}, w \models \neg F$ 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 $\mathbf{W}, w \models F_2$
- $\mathbf{W}, w \models \mathsf{P}_{\geqslant r}F$ iff $\mu(w, [F]) \geqslant r$

EXAMPLE 16. An $\mathtt{L}_\mathtt{P}$-structure $\mathbf{W}$ is shown in the figure on the right margin. Note, that $\mathbf{W}$ consists of the process from Example 15 which is extended with the following valuation of prop. letters:



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

Let us consider the truth of formulas containing the letters p, q, r:

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

$[p \Rightarrow q] = \{a, b\}$

$\mu(a, \{a, b\}) = 0.9 \quad \mu(b, \{a, b\}) = 0.6 \quad \mu(c, \{a, b\}) = 0.3$

$\mathbf{W}, a \models \mathsf{P}_{\geqslant 0.8}(p \Rightarrow q) \quad \mathbf{W}, b \not\models \mathsf{P}_{\geqslant 0.8}(p \Rightarrow q) \quad \mathbf{W}, c \not\models \mathsf{P}_{\geqslant 0.8}(p \Rightarrow q)$

$[\mathsf{P}_{\geqslant 0.8}(p \Rightarrow q)] = \{a\}$

$\mu(a, \{a\}) = 0.9 \quad \mu(b, \{a\}) = 0.1 \quad \mu(c, \{a\}) = 0.2$

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

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

EXERCISE 10. Compare structures from the previous example and from Example 14 (page 29).

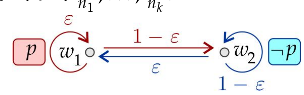
DEFINITION 15. Let Γ be a set of $\mathsf{L_P}$-formulas. $\mathbf{W}, w \models \Gamma$ iff $\mathbf{W}, w \models F$, for all $F \in \Gamma$. An $\mathsf{L_P}$-structure $\mathbf{W}$ is a model of Γ if $\mathbf{W}, w \models \Gamma$, for all $w \in \mathbb{W}$.

EXERCISE 11. Show that any finite subset of

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

has a model, but Γ have no model.

$\{\neg\mathsf{P}_{\geqslant 1}\neg p, \neg\mathsf{P}_{\geqslant \frac{1}{n_1}} p, \ldots, \neg\mathsf{P}_{\geqslant \frac{1}{n_k}} p\}$ has the following model, where $0 < \varepsilon < \frac{1}{n_1}, \ldots, \frac{1}{n_k}$:



From the model-theoretic point of view, the most important sets of formulas are those which completely describe worlds of an $\mathsf{L_P}$-structure $\mathbf{W}$:

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

We will call such sets *complete descriptions*.[67] One of the main logical problems is the reverse one: to investigate which sets of formulas could be (extended to) complete descriptions, i.e. could determine the worlds of a model. To deal with such problems, we take the proof-theoretic approach, and try to discover:

[67] Note the closeness of the concepts: Complete atoms (from section 1), types (from section 2) and complete descriptions (from section 3).

- *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.

DEFINITION 16. A formula F is **valid**, denote by $\models F$, if it satisfied at every world of every model.

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 $\mathbf{W}$.[68]

[68] $\models F$ coincides with $\varnothing \models F$.

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

3.2. *The Completeness problem for* $\mathsf{L_P}$

EXAMPLE 17. We distinguish three main groups of valid formulas. All tautologies (all $\mathsf{L_P}$-instances of tautologies) are valid:

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

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

the consequences of the ordering properties of the rational numbers are valid:

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

the consequences of the basic properties of probability are valid:

$\mathsf{P}_{\geqslant 1}(p \vee \neg p);\ \mathsf{P}_{\geqslant 0}(p \vee q),\ \mathsf{P}_{\geqslant 0.5}p \Rightarrow \mathsf{P}_{\geqslant 0.5}(p \vee q),\ \mathsf{P}_{\geqslant 0.6}p \Rightarrow \neg\mathsf{P}_{\geqslant 0.6}\neg p$, etc.;

In the next proposition we emphasize some important valid formulas.

PROPOSITION 10. For all formulas A, B:

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

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

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

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

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

PROPOSITION 11. For all formulas A, B,

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

The two preceding propositions represent the probabilistic part of a complete axiomatization that determines the notion of deductive consequence, i.e., a syntactic characterization of the concept of semantic consequence. Of course, infinitary inference rules are unavoidable because of the lack of compactness of the semantic consequence relation.

The axiomatic system $\mathbf{Ax}(\mathsf{L_P})$ contains the following axiom schemata:

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

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

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

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

and inference rules:

(MP) From A and $A \Rightarrow B$ infer B

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

(EQ$_r$) From $A \Rightarrow (B \Leftrightarrow C)$ infer $A \Rightarrow (\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C)$, for all $r \in (0,1] \cap \mathbb{Q}$

(A$_r$) From $A \Rightarrow \mathsf{P}_{\geqslant t}B$, for all $t < r$ infer $A \Rightarrow \mathsf{P}_{\geqslant r}B$, for all $r \in (0,1] \cap \mathbb{Q}$

(EQ$_r$) $\dfrac{A \Rightarrow (B \Leftrightarrow C)}{A \Rightarrow (\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C)}$

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

Note that the rules (A$_r$), $r \in [0,1] \cap \mathbb{Q}$ are infinitary in the sense that they have an infinite number of premises. The intuition behind these rules is the following: the infinitely long 'formula' $\bigwedge\limits_{t<r} \mathsf{P}_{\geqslant t}B$ must be equivalent to $\mathsf{P}_{\geqslant r}B$. The language $\mathsf{L_P}$ does not allow infinitary formulas, so the above equivalence may be replaced by an infinitary rule of inference: given the set of premises $\{\mathsf{P}_{\geqslant t}B \mid t < r\}$, one may infer $\mathsf{P}_{\geqslant r}B$. In order to be able to prove Deduction theorem (see the proof of Theorem 11), we modify this rule by adding a prefix '$A \Rightarrow$' to the premises and to the conclusion: given the set of premises $\{A \Rightarrow \mathsf{P}_{\geqslant t}B \mid t < r\}$ one may infer $A \Rightarrow \mathsf{P}_{\geqslant r}B$. Of course, it is usual to omit the prefix '$\top \Rightarrow$' and just write:

$$(\mathrm{EQ}_r) \quad \frac{B \Leftrightarrow C}{\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C} \qquad (\mathrm{A}_r) \quad \frac{\mathsf{P}_{\geqslant t}B \quad t < r}{\mathsf{P}_{\geqslant r}B}$$

DEFINITION 17. A formula F is a **syntactical consequence** of Γ,[72] denoted by $\Gamma \vdash F$ iff there are a countable ordinal κ and a sequence of formulas $\langle F_i : i \leqslant \kappa \rangle$ such that $F_\kappa = F$ and 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.

[72] F can be deduced (derived, inferred etc.) from Γ if there is a **derivation** of the form:

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

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

EXAMPLE 18. Let us prove $\{\neg \mathsf{P}_{\geqslant \frac{1}{n}} p \mid n \geqslant 1\} \vdash \mathsf{P}_{\geqslant 1} \neg p$. (See Exercise 11.)

First, we make derivations for all $\mathsf{P}_{\geqslant t}\neg p$, $t \in [0,1)$. Choose $t \in [0,1)$ and a natural number $n_t \geqslant 1$ such that $1 - t \geqslant \frac{1}{n_t}$; then $t' = 1 - t - \frac{1}{n_t} > 0$. We give sketches of the derivations, omitting many details of classical propositional reasoning.

1. $p \Leftrightarrow p \wedge \top$ [A0]
2. $\neg\top \Leftrightarrow p \wedge \neg\top$ [A0]
3. $\mathsf{P}_{\geqslant \frac{1}{n_t}} p \Leftrightarrow \mathsf{P}_{\geqslant \frac{1}{n_t}} (p \wedge \top)$ 1, (EQ$_{1/n_t}$),
4. $\mathsf{P}_{\geqslant t'} \neg\top \Leftrightarrow \mathsf{P}_{\geqslant t'}(p \wedge \neg\top)$ 2, (EQ$_{t'}$)
5. $\neg\mathsf{P}_{\geqslant \frac{1}{n_t}} p \Leftrightarrow \neg\mathsf{P}_{\geqslant \frac{1}{n_t}} (p \wedge \top)$ 3, (A0) $(A \Leftrightarrow B) \Rightarrow (\neg A \Leftrightarrow \neg B)$
6. $\neg\mathsf{P}_{\geqslant t'} \neg\top \Leftrightarrow \neg\mathsf{P}_{\geqslant t'}(p \wedge \neg\top)$ 4, (A0) $(A \Leftrightarrow B) \Rightarrow (\neg A \Leftrightarrow \neg B)$
7. $\mathsf{P}_{\geqslant 1}\top$ (A2)
8. $\mathsf{P}_{\geqslant 1}\top \Rightarrow \neg\mathsf{P}_{\geqslant t'}\neg\top$ $1 + t' > 1$, (A5)
9. $\neg\mathsf{P}_{\geqslant t'}(p \wedge \neg\top)$ 7, 8, 6, (MP)
10. $\neg\mathsf{P}_{\geqslant \frac{1}{n_t}} p$ [the premise]
11. $\neg\mathsf{P}_{\geqslant \frac{1}{n_t}} (p \wedge \top)$ 5, 10, (MP)

12. $\neg\mathsf{P}_{\geqslant\frac{1}{n_t}}(p\wedge\top)\wedge\neg\mathsf{P}_{\geqslant t'}(p\wedge\neg\top)\Rightarrow\neg\mathsf{P}_{\geqslant 1-t}p$ $\quad \frac{1}{n_t}+t'=\frac{1}{n_t}+(1-t-\frac{1}{n_t})=1-t$, (A4)

13. $\neg\mathsf{P}_{\geqslant 1-t}p$ $\quad$ 11, 9, 12, (MP)

14. $\neg\mathsf{P}_{\geqslant 1-t}(\top\wedge p)\wedge\neg\mathsf{P}_{\geqslant t}(\top\wedge\neg p)\Rightarrow\neg\mathsf{P}_{\geqslant 1}\top$ $\quad$ (A4)

15. $\mathsf{P}_{\geqslant 1}\top\Rightarrow\mathsf{P}_{\geqslant 1-t}p\vee\mathsf{P}_{\geqslant t}\neg p$ $\quad$ [the low of contraposition and De Morgan's lows]

16. $\mathsf{P}_{\geqslant t}\neg p$ $\quad$ 7, 13, 15

All these derivations of $\mathsf{P}_{\geqslant t}\neg p$, $t<1$, extended by an application of the infinitary rule (A_1), make the derivation of $\mathsf{P}_{\geqslant 1}\neg p$.

DEFINITION 18. A set of formulas Γ is **consistent** iff $T\nvdash\bot$, 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).

EXAMPLE 19. $\{\neg\mathsf{P}_{\geqslant 1}\neg p\}\cup\{\neg P_{\geqslant\frac{1}{n}}p\mid n\geqslant 1\}$ is not consistent set of formulas, by the previous example shows.

THEOREM 11. [Deduction theorem] If $\Gamma, F\vdash G$ then $\Gamma\vdash F\Rightarrow G$.

PROOF. We use the induction on the length of the derivation of G from Γ, F. The cases when (1°) G is an axiom, or (2°) F coincides with G, or (3°) $G\in\Gamma$, or (4°) G is derived from Γ, F by (MP) are standard.

Suppose that $G = A\Rightarrow(\mathsf{P}_{\geqslant r}B\Leftrightarrow\mathsf{P}_{\geqslant r}C)$ is derived from Γ, F by (EQ_r) with the premise $A\Rightarrow(B\Leftrightarrow C)$:

$$\underline{\Gamma, F,\dots, A\Rightarrow(B\Leftrightarrow C)},\dots A\Rightarrow(\mathsf{P}_{\geqslant r}B\Leftrightarrow\mathsf{P}_{\geqslant r}C)$$

Applying Induction hypothesis on the underlined subderivation, we obtain:

$\Gamma\vdash F\Rightarrow(A\Rightarrow(B\Leftrightarrow C))$

Now it is easy to proceed to the desired result:

$\Gamma\vdash(F\wedge A)\Rightarrow(B\Leftrightarrow C)$

$\Gamma\vdash(F\wedge A)\Rightarrow(\mathsf{P}_{\geqslant r}B\Leftrightarrow\mathsf{P}_{\geqslant r}C)$ $\quad (\mathrm{EQ}_r)$

$\Gamma\vdash F\Rightarrow(A\Rightarrow(\mathsf{P}_{\geqslant r}B\Leftrightarrow\mathsf{P}_{\geqslant r}C))$

Suppose that $G = A\Rightarrow\mathsf{P}_{\geqslant r}B$ is derived from Γ, F by (A_r) with the premises $A\Rightarrow\mathsf{P}_{\geqslant t}B$, $t<r$:

$$\underline{\Gamma, F,\dots, A\Rightarrow\mathsf{P}_{\geqslant t}B}, t<r,\dots A\Rightarrow\mathsf{P}_{\geqslant r}B$$

Then:

$\Gamma\vdash F\Rightarrow(A\Rightarrow\mathsf{P}_{\geqslant t}B)$, $t<r$

$\Gamma\vdash(F\wedge A)\Rightarrow\mathsf{P}_{\geqslant t}B$, $t<r$

$\Gamma\vdash(F\wedge A)\Rightarrow\mathsf{P}_{\geqslant r}B$ $\quad (\mathrm{A}_r)$

$\Gamma\vdash F\Rightarrow(A\Rightarrow\mathsf{P}_{\geqslant r}B)$ $\quad\square$

THEOREM 12. [Extension theorem] Every consistent set of formulas Γ can be extended to a maximal consistent set Γ^+.

PROOF Let $\langle F_k : k \geqslant 1\rangle$ be an arbitrary enumeration of all $\mathsf{L_P}$-formulas.

We define a sequence Γ_n, $n \geqslant 0$, inductively as follows:

- $\Gamma_0 = \Gamma$
- $n \geqslant 1$
  - 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:
    * $F_n = A \Rightarrow \mathsf{P}_{\geqslant r}B$. 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

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

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

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

Lemma 1. Γ_n is consistent for each n.

Proof of Lemma 1. By induction on n.

$\Gamma_0 = \Gamma$ is consistent by the assumption of the theorem.

Assume that Γ_n is consistent.

If $\Gamma_n \cup \{F_n\}$ is consistent, then surely Γ_{n+1} is consistent.

When $\Gamma_n \cup \{F_n\}$ is not consistent, then $\Gamma_n \cup \{\neg F_n\}$ must be consistent. If $\Gamma_n \cup \{\neg F_n\}$ were inconsistent, then it would be $\Gamma_n \cup \{\neg F_n\} \vdash \bot$ and $\Gamma_n \cup \{F_n\} \vdash \bot$, i.e., by Deduction theorem, $\Gamma_n \vdash \neg F_n \Rightarrow \bot$ and $\Gamma_n \vdash F_n \Rightarrow \bot$, and hence $\Gamma_n \vdash \bot$, by using (A0) $(F_n \Rightarrow \bot) \Rightarrow ((\neg F_n \Rightarrow \bot) \Rightarrow \bot)$, which contradicts the inductive hypothesis.

If $\Gamma_n \cup \{A \Rightarrow \mathsf{P}_{\geqslant r}B\}$ is not consistent, then there must be $t < r$ such that $\Gamma_n \cup \{\neg(A \Rightarrow \mathsf{P}_{\geqslant r}B), \neg(A \Rightarrow \mathsf{P}_{\geqslant t}B)\}$ is consistent. Assume the opposite:

$\Gamma_n, \neg(A \Rightarrow \mathsf{P}_{\geqslant r}B), \neg(A \Rightarrow \mathsf{P}_{\geqslant t}B) \vdash \bot,$ for all $t < r$. Then,

$\Gamma_n, \neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \vdash \neg(A \Rightarrow \mathsf{P}_{\geqslant t}B) \Rightarrow \bot$, for all $t < r$, [Deduction theorem]

$\Gamma_n, \neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \vdash A \Rightarrow \mathsf{P}_{\geqslant t}B$, for all $t < r$, [(A0) $(\neg(A \Rightarrow \mathsf{P}_{\geqslant t}B) \Rightarrow \bot) \Rightarrow (A \Rightarrow \mathsf{P}_{\geqslant t}B)$]

$\Gamma_n, \neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \vdash A \Rightarrow \mathsf{P}_{\geqslant r}B$, [$\mathrm{A}_r$]

$\Gamma_n \vdash \neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \Rightarrow (A \Rightarrow \mathsf{P}_{\geqslant t}B)$, [Deduction theorem]

$\Gamma_n \vdash \bot$ [(A0) $(\neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \Rightarrow (A \Rightarrow \mathsf{P}_{\geqslant r}B)) \Rightarrow \bot$]

which contradicts the induction hypothesis.

Lemma 2. For each formula F, $\neg F \in \Gamma^+$ iff $F \notin \Gamma^+$.

Proof of Lemma 2. By the construction, for each formula F, either $F \in \Gamma^+$ or $\neg F \in \Gamma^+$, but not both, and thus either $F \notin \Gamma^+$ or $\neg F \notin \Gamma^+$.

Lemma 3. If $\Gamma^+ \vdash F$ then $F \in \Gamma^+$.

Proof of Lemma 3. It is obvious that Γ^+ contains all instances of the axiom schemata. We prove that Γ^+ is closed for the inference rules: is the premises of a rule belong to Γ, then the conclusion is in Γ^+ too.

[MP] If $A, A \Rightarrow B \in \Gamma^+$, but $B \notin \Gamma^+$, then there would be n such that

$$A, A \Rightarrow B, \neg B \in \Gamma_n,$$

and Γ_n would be inconsistent, contrary to Lemma 1.
[EQ$_r$] If $A \Rightarrow (B \Leftrightarrow C) \in \Gamma^+$, but $A \Rightarrow (\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C) \notin \Gamma^+$, then there would be n such that

$$A \Rightarrow (B \Leftrightarrow C), \neg(A \Rightarrow (\mathsf{P}_{\geqslant r}B \Leftrightarrow \mathsf{P}_{\geqslant r}C)) \in \Gamma_n,$$

and Γ_n would be inconsistent, contrary to Lemma 1.
[A$_r$] If $A \Rightarrow \mathsf{P}_{\geqslant t}B \in \Gamma^+$, for all $t < r$, but $A \Rightarrow \mathsf{P}_{\geqslant r}B \notin \Gamma^+$, then $\neg(A \Rightarrow \mathsf{P}_{\geqslant r}B) \in \Gamma^+$, and by the construction, there is $t' < r$ such that $\neg(A \Rightarrow \mathsf{P}_{\geqslant t'}B) \in \Gamma^+$, and hence, for some n

$$A \Rightarrow \mathsf{P}_{\geqslant t'}B, \neg(A \Rightarrow \mathsf{P}_{\geqslant t'}B) \in \Gamma_n$$

which contradicts the consistency of Γ_n.
Lemma 4. Γ^+ is a maximal consistent set.
Proof of Lemma 4. Γ^+ is consistent set. Really, if $\Gamma^+ \vdash \bot$, then $\bot \in \Gamma^+$, and hence $\bot \in \Gamma_n$, for some n, which contradicts Lemma 1.

If $\Gamma^+ \subsetneqq \Gamma^\#$, then there is a formula F such that $F \in \Gamma^\#$ and $F \notin \Gamma^+$. By Lemma 2, $\neg F \in \Gamma^+$, and also $\neg F \in \Gamma^\#$, which gives that $\Gamma^\#$ is inconsistent. Thus, Γ^+ is a maximal consistent set. □

THEOREM 13. [Completeness theorem] Every consistent set of formulas has an $\mathsf{L_P}$-model.

PROOF Let Γ be a consistent set of formulas. The required model will be constructed over the set of all maximal consistent extensions of Γ:

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

It should be shown:

1) $\mathcal{F}$ is an algebra of subsets of $\mathbb{W}$;

2) μ_Δ is finitely additive;

3) $(\mathbb{W}, \mathcal{F}, \mu_\Delta) \models \Gamma$.

1) $\top \in \Delta$, for every Δ; hence $[\top] = \mathbb{W} \in \mathcal{F}$.

Let A be an $\mathsf{L_P}$-formula. For every Δ:

$$\neg A \in \Delta \text{ iff } A \notin \Delta, \text{ i.e. } \Delta \in [\neg A] \text{ iff } \Delta \notin [A],$$

which gives $\mathbb{W} \setminus [A] = [\neg A] \in \mathcal{F}$.

Let A, B be $\mathsf{L_P}$-formulas. For every Δ:

$$A \vee B \in \Delta \text{ iff } A \in \Delta \text{ or } B \in \Delta, \text{ i.e. } \Delta \in [A \vee B] \text{ iff } \Delta \in [A] \cup [B],$$

which gives $[A] \cup [B] = [A \vee B] \in \mathcal{F}$.
2) First, we prove the equivalence: $\mu_\Delta[F] \geqslant r$ iff $\mathsf{P}_{\geqslant t}F \in \Delta$.

Assume $\mu_\Delta[F] \geqslant r$. For each $t < r$, by the definition of the supremum, there exists $s_t > t$ such that $\mathsf{P}_{\geqslant s_t}F \in \Delta$. The following derivation (where we omit many details) shows that it must be $\mathsf{P}_{\geqslant t}F \in \Delta$.

1. $\mathsf{P}_{\geqslant s_t} F$ [in Δ]
2. $\mathsf{P}_{\geqslant t+(s_t-t)} F \Rightarrow \mathsf{P}_{\geqslant t}(F \wedge \top) \vee \mathsf{P}_{\geqslant s_t-t}(F \wedge \neg\top)$ the contraposition law, $(A4)$
3. $\mathsf{P}_{\geqslant t}(F \wedge \top) \vee \mathsf{P}_{\geqslant s_t-t}(F \wedge \neg\top)$ 1, 2, (MP)
4. $\mathsf{P}_{\geqslant t} F \Leftrightarrow \mathsf{P}_{\geqslant t}(F \wedge \top)$ $F \Leftrightarrow F \wedge \top$, (EQ_t)
5. $\mathsf{P}_{\geqslant s_t-t}\neg\top \Leftrightarrow \mathsf{P}_{\geqslant s_t-t}(F \wedge \neg\top)$ $\neg\top \Leftrightarrow F \wedge \neg\top$, (EQ_{s_t-t})
6. $\mathsf{P}_{\geqslant 1}\top \Rightarrow \neg\mathsf{P}_{\geqslant s_t-t}\neg\top$ $1+s_t-t>1$
7. $\mathsf{P}_{\geqslant 1}\top$ (A2)
8. $\neg\mathsf{P}_{\geqslant s_t-t}\neg\top$ 7, 6, (MP)
9. $\neg\mathsf{P}_{\geqslant s_t-t}(F \wedge \neg\top)$ 8, 5
10. $\mathsf{P}_{\geqslant t}(F \wedge \top)$ 9, 3, (A0) $(A \vee B) \wedge \neg B \Rightarrow A$
11. $\mathsf{P}_{\geqslant t} F$ 4, 10

Thus, $\mathsf{P}_{\geqslant t} F \in \Delta$ for every $t < r$, and by (A_r), $\mathsf{P}_{\geqslant r} F \in \Delta$. The converse is obviously true.

Finally, we prove that μ_Δ is finitely additive.

$\mu_\Delta(\mathbb{W}) = \mu_\Delta[\top] = 1$, since $\mathsf{P}_{\geqslant r}\top \in \Delta$, for all $r \in [0,1] \cap \mathbb{Q}$.

Assume $[A] \cap [B] = \emptyset$, $\mu_\Delta[A] = a$ and $\mu_\Delta[B] = b$, for $a, b \in [0,1]$. We have to prove that $\mu_\Delta([A] \cup [B]) = a + b$.

If $[A] \cap [B] = \emptyset$, then for all Δ, $A \wedge B \notin \Delta$, and $\neg(A \wedge B) \in \Delta$. The tautology $\neg(A \wedge B) \Rightarrow (A \Leftrightarrow (A \wedge \neg B))$ gives $A \Leftrightarrow (A \wedge \neg B) \in \Delta$. From $\mathsf{P}_{\geqslant a} A \in \Delta$ we have $\mathsf{P}_{\geqslant a}(A \wedge \neg B) \in \Delta$.

It cannot be $a + b > 1$. Otherwise, using Axiom $(A5)$, we would have

$$\mathsf{P}_{\geqslant a}(A \wedge \neg B) \Rightarrow \neg\mathsf{P}_{\geqslant b}\neg(A \wedge \neg B)$$

i.e. $\neg\mathsf{P}_{\geqslant b}(\neg A \vee B) \in \Delta$. From

(A1) $\mathsf{P}_{\geqslant 0}(\neg A \wedge \neg B)$,

$\neg\mathsf{P}_{\geqslant b}(\neg A \vee B) \in \Delta$ and

(A3) $\mathsf{P}_{\geqslant b}((\neg A \vee B) \wedge B) \wedge \mathsf{P}_{\geqslant 0}((\neg A \vee B) \wedge \neg B) \to \mathsf{P}_{\geqslant b}(\neg A \vee B)$,[73]

we have

$\neg\mathsf{P}_{\geqslant b}((\neg A \vee B) \wedge B) \in \Delta$, i.e. $\neg\mathsf{P}_{\geqslant b} B \in \Delta$, which is a contradiction. Hence, $a + b \leqslant 1$

[73] $(\neg A \vee B) \wedge B \Leftrightarrow B$
$(\neg A \vee B) \wedge \neg B \Leftrightarrow \neg A \wedge \neg B$

Since $\mathsf{P}_{\geqslant b} B \in \Delta$, by Axiom $(A3)$[74]

[74] $(A \vee B) \wedge \neg B \Leftrightarrow A \wedge \neg B$
$(A \vee B) \wedge B \Leftrightarrow B$

$$\mathsf{P}_{\geqslant a}((A \vee B) \wedge \neg B) \wedge \mathsf{P}_{\geqslant b}((A \vee B) \wedge B) \Rightarrow \mathsf{P}_{\geqslant a+b}(A \vee B)$$

Thus, $\mathsf{P}_{\geqslant a+b}(A \vee B) \in \Delta$, i.e. $\mu_\Delta([A] \cup [B]) \geqslant a + b$.

If there were $\varepsilon > 0$ such that $a + b + \varepsilon \leqslant 1$ and $\mathsf{P}_{\geqslant a+b+\varepsilon}(A \vee B) \in \Delta$, then by Axiom $(A4)$ in the 'contrapositive form':[75]

[75] $(A \vee B) \wedge B \Leftrightarrow B$
$(A \vee B) \wedge \neg B \Leftrightarrow A \wedge \neg B$

$$\mathsf{P}_{\geqslant a+b+\varepsilon}(A \vee B) \Rightarrow \mathsf{P}_{\geqslant b+\varepsilon/2} B \vee \mathsf{P}_{\geqslant a+\varepsilon/2}(A \wedge \neg B),$$

we would have $\mathsf{P}_{\geqslant a+\frac{\varepsilon}{2}} A \in \Delta$ or $\mathsf{P}_{\geqslant b+\frac{\varepsilon}{2}} B \in \Delta$, which is impossible. $\square$

THEOREM 14. $\Gamma \vdash F$ iff $\Gamma \models F$

Appendix

Proof of Preposition 2

Each of (1)-(4) holds if we only assume P1 and P2.

1. We have that $\models A \vee \neg A$ and $A \models \neg\neg A$, so by P1 and P2,

$$1 = \mathbf{P}(A \vee \neg A) = \mathbf{P}(A) + \mathbf{P}(\neg A).$$

2. If $A \models B$ then $\neg B \models \neg A$, so from P2 and 1), and the fact that $\mathbf{P}$ takes values in $[0,1]$,

$$1 \geq \mathbf{P}(\neg B \vee A) = \mathbf{P}(\neg B) + \mathbf{P}(A) = 1 - \mathbf{P}(B) + \mathbf{P}(A),$$

from which the required inequality follows.

3. If $A \equiv B$ then $A \models B$ and $B \models A$. By 2), $\mathbf{P}(A) \leq \mathbf{P}(B)$ and $\mathbf{P}(B) \leq \mathbf{P}(A)$, so $\mathbf{P}(A) = \mathbf{P}(B)$.

4. Since $A \vee B \equiv A \vee (\neg A \wedge B)$ and $A \models \neg(\neg A \wedge B)$, P2 and 3) give

$$\mathbf{P}(A \vee B) = \mathbf{P}(A \vee (\neg A \wedge B)) = \mathbf{P}(A) + \mathbf{P}(\neg A \wedge B) \tag{1}$$

Also, $B \equiv (A \wedge B) \vee (\neg A \wedge B)$, and $A \wedge B \models \neg(\neg A \wedge B)$, so by P2 and 3)

$$\mathbf{P}(B) = \mathbf{P}((A \wedge B) \vee (\neg A \wedge B)) = \mathbf{P}(A \wedge B) + \mathbf{P}(\neg A \wedge B)) \tag{2}$$

Eliminating $\mathbf{P}(\neg A \wedge B)$ from (1), (2) gives $\mathbf{P}(A \vee B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \wedge B)$. $\square$

Proof of Theorem in Example 6

For P1 suppose that $\models S$, but $\mathbf{Bel}(S) < 1$. Then $\mathbf{M}(S) = 1$, for all $\mathbf{M}$. If $\mathbf{Bel}(S) < p < 1$, we have

$$(-1)(\mathbf{M}(S) - p) = p - 1 < 0, \text{ for all } \mathbf{M}.$$

Contradiction.

Now suppose that P2 fails, and there are S_1, S_2 such that $S_1 \models \neg S_2$, but

$$\mathbf{Bel}(S_1) + \mathbf{Bel}(S_2) < \mathbf{Bel}(S_1 \vee S_2).$$

Then $S_1 \models \neg S_2$ forces that at most one of S_1, S_2 can be true in any $\mathbf{M}$ su:

$$\mathbf{M}(S_1 \vee S_2) = \mathbf{M}(S_1) + \mathbf{M}(S_2).$$

Pick $p_1 > \mathbf{Bel}(S_1)$, $p_2 > \mathbf{Bel}(S_2)$, $p < \mathbf{Bel}(S_1 \vee S_2)$ such that $p_1 + p_2 < p$. Then

$$(-1)(\mathbf{M}(S_1) - p_1) + (-1)(\mathbf{M}(S_2) - p_2) + (\mathbf{M}(S_1 \vee S_2) - p) = (p_1 + p_2) - p < 0, \text{ for all } \mathbf{M}.$$

A similar argument when $\mathbf{Bel}(S_1) + \mathbf{Bel}(S_2) > \mathbf{Bel}(S_1 \vee S_2)$ shows that this cannot hold either se we must have equality here.

Finally suppose that $\mathbf{Bel}(\exists x\, A) > \lim_{n\to\infty} \mathbf{Bel}(A(c_1) \vee \cdots \vee A(c_n))$. Then we could pick p and p_n, $n = 1, 2, \ldots$ such that

$$\mathbf{Bel}(A(c_1) \vee \cdots \vee A(c_n)) < p_n < p < \mathbf{Bel}(\exists x\, A) \text{ and } \sum_{n\geqslant 1} p_n < p.$$

Since for $\mathbf{M}$

$$\mathbf{M}(\exists x\, A) = \max_i \mathbf{M}(A(c_i)) = \max_n \mathbf{M}(A(c_1) \vee \cdots \vee A(c_n)),$$

- if $\mathbf{M}(\exists x\, A) = 0$, then $\mathbf{M}(A(c_1) \vee \cdots \vee A(c_n)) = 0$, for all n;
- if $\mathbf{M}(\exists x\, A) = 1$, then there is k such that $\mathbf{M}(A(c_1) \vee \cdots \vee A(c_n)) = 1$, for $n \geqslant k$.

Then:

$$(\mathbf{M}(\exists x\, A) - p) + \sum_{n\geqslant 1} (-1)(\mathbf{M}(A(c_1) \vee \cdots \vee A(c_n)) - p_n) \leqslant p - \sum_{n\geqslant 1} p_n < 0, \text{ for all } \mathbf{M}. \quad \square$$

Proof of Theorem 6.

This theorem will be proved by induction on the construction of formula is $\mathrm{L}^k_{\omega\omega}$ whose variables are among $x_1, \ldots, x_k$ and whose free variables are among $x_1, \ldots, x_m$, simultaneously for all $m \leqslant k$ and for all types $s(x_1, \ldots, x_m)$.

The base case of the induction (atomic formulas and equalities) and the induction step for the negation $\neg$ are obvious.

Assume that $F(x_1, \ldots, x_m)$ is a conjunction $F_1(x_1, \ldots, x_m) \wedge F_2(x_1, \ldots, x_m)$. We distinguish two cases.

1. $E_k \models \forall \overline{x}(s(\overline{x}) \Rightarrow F_i(\overline{x}))$, for both $i = 1, 2$, then $E_k \models \forall \overline{x}(s(\overline{x}) \Rightarrow F_1(\overline{x}) \wedge F_2(\overline{x}))$
2. $E_k \models \forall \overline{x}(s(\overline{x}) \Rightarrow \neg F_i(\overline{x}))$, for at least one $i = 1, 2$, then $E_k \models \forall \overline{x}(s(\overline{x}) \Rightarrow \neg(F_1(\overline{x}) \wedge F_2(\overline{x})))$

A crucial use of the extension axioms will be made in the case where the formula $F(x_1, \ldots, x_m)$ starts with an existential quantifier. Assume that F is a formula $\exists y\, G(x_1, \ldots, x_m, y)$, and that induction hypotheses holds for $G(x_1, \ldots, x_m, y)$.

If $E_k \models \forall \overline{x}(s(\overline{x}) \Rightarrow \neg\exists y\, G(\overline{x}, y))$, then 2. holds for F. Otherwise,

$$E_k \not\models \forall \overline{x}(s(\overline{x}) \Rightarrow \neg\exists y\, G(\overline{x}, y)).$$

We show that in the latter case $E_k \not\models \forall \overline{x}(s(\overline{x}) \Rightarrow \exists y\, G(\overline{x}, y))$. By our assumption about variables of F, we must have that the variable y is the variable x_j, for some j such that $1 \leqslant j \leqslant k$. We now distinguish two cases> the case where $j > m$ and the case where $j \leqslant m$.

CASE 1. $j > m$, which means that the variable y is different from all the variables $x_1, \ldots, x_m$. Note that in this case m must be less than k. Since, $E_k \not\models \forall \overline{x}(s(\overline{x}) \Rightarrow \neg\exists y\, G(\overline{x}, y))$, there is a structure $\mathbf{D}$ such that $(D) \models E_k$, and $(D) \models \exists \overline{x}(s(\overline{x}) \wedge \exists y\, G(\overline{x}, y))$, i.e.

$$\mathbf{D} \models \exists \overline{x} \exists y\, (s(\overline{x}) \wedge\, G(\overline{x}, y))$$

Let $a_1, \ldots, a_m, b$ be elements of the universe D of $\mathbf{D}$ such that:

$$\mathbf{D} \models (s(a_1, \ldots, a_m) \wedge\, G(a_1, \ldots, a_m, b))$$

Let $t(x_1, \dots, x_m, y)$ be the unique type determined by $(a_1, \dots, a_m, b)$ in $\mathbf{D}$. Thus,

$$\mathbf{D} \models \exists \overline{x} \exists y \, (t(\overline{x}, y) \wedge \, G(\overline{x}, y))$$

By applying induction hypotheses to the formula G and the type $t(\overline{x}, y)$, we infer that

$$E_k \models \forall \overline{x} \forall y (t(\overline{x}, y) \Rightarrow G(\overline{x}, y)).$$

Since the type t is an extension of s and E_k is the conjunction of all extension axioms with at most k variables, it follows that

$$E_k \models \forall \overline{x} (s(\overline{x}) \Rightarrow \exists y t(\overline{x}, y)).$$

We can now conclude that

$$E_k \models \forall \overline{x} (s(\overline{x}) \Rightarrow \exists y G(\overline{x}, y)).$$

Note that $t(x_1, \dots, x_m, y)$ extends the type $s(x_1, \dots, x_m)$.

CASE 2. $j \leqslant m$, which means that the variable z is the variable x_j for some $j \leqslant m$. WLOG, assume $j = 1$. ... $\square$

Proof of Proposition 6.

Assume for convenience that $\mathtt{L} = \{R\}$, where R is a binary predicate symbol. We show that $\mathbf{P}(\neg E_{T,T'}) = 0$, where $T'(\overline{x}, y)$ is a type which extends $T(\overline{x})$. Let $\overline{x} = (x_1, \dots, x_k)$, and let $A(\overline{x}, y)$ be

$$\bigwedge_{\pm R(x_i, y) \text{ are in } T'} \pm R(x_i, y).$$

Assume $\mathbb{U} = \{c_1, \dots, c_n\}$, $[\mathbb{U}]^k$ is the set of k-tuples of mutually different individuals, and $n > k$. Then

$$\begin{aligned}
\mathbf{P}_n(\neg E_{T,T'}) &= \mathbf{P}_n(\exists \overline{x} (T(\overline{x}) \wedge \forall y \, \neg T'(\overline{x}, y))) \\
&\leqslant \sum_{\overline{c} \in [\mathbb{U}]^k} \mathbf{P}_n(T(\overline{c}) \wedge \forall y \, \neg T'(\overline{c}, y)) \\
&= n(n-1) \cdots (n-k+1) \mathbf{P}_n(T(c_1, \dots, c_k) \wedge \forall y \, \neg T'(c_1, \dots, c_k, y)), \text{ by symmetry} \\
&\leqslant n^k \mathbf{P}_n(T(c_1, \dots, c_k) \wedge \forall y \, \neg T'(c_1, \dots, c_k, y)) \\
&\leqslant n^k \mathbf{P}_n(\forall y \, \neg T'(c_1, \dots, c_k, y)) \\
&\leqslant n^k \mathbf{P}_n(\bigwedge_{i=k+1}^{n} \neg T'(c_1, \dots, c_k, c_i)) \\
&\leqslant n^k \mathbf{P}_n(\bigwedge_{i=k+1}^{n} \neg A(c_1, \dots, c_k, c_i)) \\
&= n^k \prod_{i=k+1}^{n} \mathbf{P}_n(\neg A(c_1, \dots, c_k, c_i)), \text{ by independence} \\
&= n^k \prod_{i=k+1}^{n} \left(1 - \frac{1}{2^{2k+1}}\right) \\
&= n^k \left(1 - \frac{1}{2^{2k+1}}\right)^{n-k} \\
&\to 0 \text{ as } n \to \infty
\end{aligned}$$

Since $\mathbf{P}_n(E_{T,T'}) + \mathbf{P}_n(\neg E_{T,T'}) = 1$, we have $\mathbf{P}_n(E_{T,T'}) \to 1$, $n \to \infty$. $\square$

References

1. F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller, *From Statistical Knowledge Bases to Degrees of Belief*, Artif. Intell. 87(1-2), 75-143, 1996

2. H-D. Ebbinghaus, J. Flum, *Finite Model Theory*, Springer, 1995

3. N. Ikodinović, Z. Ognjanović, A. Perović, M. Rašković, *Completeness theorems for σ–additive probabilistic semantics*, Annals of Pure and Applied Logic, Volume 171, Issue 4, 2020.

4. N. Ikodinović, *Some Notes on Finite and Hyperfinite model theory*, Logic and Applications – LAP2016, Dubrovnik, Croatia, 2016

5. N. Ikodinović, M. Rašković, Z. Marković, Z. Ognjanović, *Logics with Generalized Measure Operators*, Journal of Multiple-Valued Logic and Soft Computing, Volume 20, Number 5-6, 527-555, 2013

6. Z. Ognjanović *Probabilistic Extensions of Various Logical Systems*, Springer, 2020.

7. Z. Ognjanović, M. Rašković, Z. Marković, *Probability Logics: Probability-Based Formalization of Uncertain Reasoning*, Springer, 2016.

8. Z. Ognjanović, M. Rašković, and Z. Marković, *Probability logics*, in Zbornik radova, subseries Logic in computer science 12(20), 35-111, Mathematical institute, 2009

9. M. Rašković, *Classical logic with some probability operators*, Publication de l'Institut Math. 53, 1-3, 1993