

ABOUT SEMANTICS OF PROBABILISTIC LOGIC

TRAN DINH QUE

Abstract. The probabilistic logic is a paradigm of handling uncertainty by means of integrating the classical logic and the theory of probability. It makes use of notions such as possible worlds, classes of possible worlds or basic propositions from the classical logic to construct sample spaces on which a probability distribution is performed. When such a sample space is constructed, the probability of a sentence is then defined by means of a distribution on this space.

This paper points out that deductions in the point-valued probabilistic logic via Maximum Entropy Principle as well as in the interval-valued probabilistic logic do not depend on selected sample spaces.

1. INTRODUCTION

In various approaches to handling uncertainty, the paradigm of probabilistic logic has been widely studied in the community of AI researchers (e.g., [1], [4], [5], [6], [8]). The probabilistic logic, an integration of logic and the probability theory, determines a probability of a sentence by means of a probability distribution on some sample space. In order to have a sample space on which a probability distribution is performed, this paradigm has made use of notions of *possible worlds*, *classes of possible worlds* or *basic propositions* from the classical logic. It means that there are three approaches to give semantics of probabilistic logics based on the various sample space: (i) the set of all possible worlds; (ii) classes of possible worlds; (iii) the set of basic propositions.

Based on semantics of probability of a sentence proposed by Nilsson [8], an interval-valued probabilistic logic has been developed by Dieu [4]. Suppose that \mathcal{B} is an interval probability knowledge base (iKB) composed of sentences with their interval values which are closed subinterval of the unit interval $[0, 1]$. From the knowledge base, we can infer the interval value for any sentence. In the special case, in which values of sentences in \mathcal{B} are not interval but point values of $[0, 1]$, i.e., \mathcal{B} is a pointed-valued probabilistic knowledge base (pKB), the value of S deduced from \mathcal{B} , in general, is not a point value [8]. In order to obtain a point value, some constraint has been added to probability distributions. The Maximum Entropy Principle (MEP) is very often used to select such a distribution ([2], [4], [8]).

The purpose of this paper is to examine a relationship of deductions in the point-valued probabilistic logic via MEP as well as in the interval-valued probabilistic logic. We will point out that deductions in these logics do not depend on selected sample spaces. In other words, these approaches are equivalent w.r.t. the deduction of the interval-valued probabilistic logic as well as one of the point-valued probabilistic logic via Maximum Entropy Principle. Section 2 reviews some basic notions: possible worlds, basic propositions and the probability of a sentence according to the selected sample space. Section 3 investigates the equivalence of deductions in the interval-valued probabilistic logic as well as in the point-valued probabilistic logic. Some conclusions and discussions are presented in Section 4.

2. PROBABILITY OF A SENTENCE

2.1. Possible worlds

The construction of logic based on possible worlds has been considered to be a normal paradigm in building semantics of many logics such as probabilistic logic, possibilistic logic, modal logics and so on (e.g., [4], [5], [6], [8]). The notion of possible world arises from the intuition that besides the current world in which a sentence is true there are the other worlds an agent believes that the sentence

may be true. We can consider a set of possible worlds to be a qualitative way for measuring an agent's uncertainty of a sentence. The more possible worlds there are, the more the agent is uncertain about the real state of the world. When such a set of possible worlds is given, the uncertainty of a sentence is quantified by adding a probability distribution on the set.

Suppose that we have a set of sentences $\Sigma = \{\phi_1, \dots, \phi_l\}$ (we restrict to considering propositional sentences in this paper). Let $\mathcal{A} = \{a_1, \dots, a_m\}$ be a set of all atoms or propositional variables in Σ and \mathcal{L}_Σ be a propositional language generated by atoms in \mathcal{A} . Each possible world of Σ or \mathcal{L}_Σ is considered as an interpretation of formulas in the classical propositional logic. That means it is an assignment of truth values *true* (1) or *false* (0) to atoms in \mathcal{A} . Denote Ω to be a set of all such possible worlds and $\omega \models \phi$ to mean that ϕ is true in a possible world ω . Each possible world ω determines a Σ -consistent column vector $\sigma = (\sigma_1, \dots, \sigma_l)^t$, where $\sigma_i = \text{val}_\omega(\phi_i)$ is the truth value of ϕ_i in the possible world ω (we denote here σ^t to be the transpose of vector σ).

Note that two different possible worlds may have the same Σ -consistent vector. We need to consider the set of all possible worlds as well as the set of subsets of possible worlds, which are characterised by Σ -consistent vectors. In the later case, it means that we group all possible worlds with the same Σ -consistent vector into a class. Now we formalise this notion.

Two possible worlds ω_1 and ω_2 of Ω are Σ -equivalence if $\text{val}_{\omega_1}(S_i) = \text{val}_{\omega_2}(S_i)$, for all $i = 1, \dots, l$. This equivalent relation determines a classification on Ω and we call $\bar{\Omega}$ to be the set of all such equivalent classes. Each equivalent class is then characterised by a Σ -consistent column vector σ .

We consider an example.

Example 1. Suppose that $\Sigma = \{S_1 = A, S_2 = A \wedge B, S_3 = A \rightarrow C\}$. Since there are three atoms $\{A, B, C\}$ in Σ , Ω has $2^3 = 8$ possible worlds

$$\begin{aligned} \omega_1 &= (A, B, C), & \omega_5 &= (\neg A, \neg B, C), \\ \omega_2 &= (A, \neg B, C), & \omega_6 &= (\neg A, B, \neg C), \\ \omega_3 &= (A, B, \neg C), & \omega_7 &= (A, \neg B, \neg C), \\ \omega_4 &= (\neg A, B, C), & \omega_8 &= (\neg A, \neg B, \neg C). \end{aligned}$$

The notation $\omega_2 = (A, \neg B, C)$ means that the truth value 1 is assigned to atoms A, C and the value 0 to B and so on.

Truth values of sentences in Σ with respect to possible worlds are given in the following matrix

$$\begin{array}{c} \omega_1 \ \omega_2 \ \omega_3 \ \omega_4 \ \omega_5 \ \omega_6 \ \omega_7 \ \omega_8 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix} \end{array}$$

It is easy to see that there are five classes of possible worlds in $\bar{\Omega}$: $\bar{\omega}_1 = \{\omega_1\}$, $\bar{\omega}_2 = \{\omega_2\}$, $\bar{\omega}_3 = \{\omega_3\}$, $\bar{\omega}_4 = \{\omega_4, \omega_5, \omega_6, \omega_8\}$, $\bar{\omega}_5 = \{\omega_7\}$. Then the truth values of sentences in Σ w.r.t. $\bar{\Omega}$ are represented in the following matrix

$$\begin{array}{c} \bar{\omega}_1 \ \bar{\omega}_2 \ \bar{\omega}_3 \ \bar{\omega}_4 \ \bar{\omega}_5 \\ \begin{matrix} S_1 \\ S_2 \\ S_3 \end{matrix} \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{pmatrix} \end{array}$$

Each column vector in the above matrix characterises truth values of corresponding sentences in a class of possible worlds. For instance, vector $V = (1, 0, 1)^t$ characterises the truth value 1 of S_1 , 0 of S_2 and 1 of S_3 in the class $\bar{\omega}_2 = \{\omega_2\}$ and so on.

The construction of two sets Ω and $\bar{\Omega}$ that we have just discussed plays an important role in giving semantics of probability of a sentence. The set of all possible worlds Ω as well as the set of all

equivalent classes $\bar{\Omega}$ will be sample spaces for a probability distribution. Before going on examining the probability semantics of a sentence, we consider the other sample space that is obtained from basic propositions.

2.2. Basic propositions

In the previous subsection, we have presented notions of possible worlds as well as of their classes which are sample spaces for constructing probability. We now consider the other sample space which is composed of basic propositions.

As presented in subsection 2.1, \mathcal{L}_Σ is denoted to be the propositional language generated by the set of all propositional variables $\mathcal{A} = \{a_1, \dots, a_m\}$ in the set of sentences $\Sigma = \{\phi_1, \dots, \phi_l\}$. A *basic proposition* has the form $\phi = \sigma_1 \wedge \dots \wedge \sigma_m$, in which $\sigma_i = a_i$ or $\sigma_i = \neg a_i$. It is clear that since $|\mathcal{A}| = m$ ($|\cdot|$ is denoted to be the cardinal of a set), there are $n = 2^m$ basic propositions $\mathcal{A}_b = \{b_1, \dots, b_n\}$. The following proposition showed in [9] provides a basis for simplifying operations in the propositional logic.

Proposition 1. *For every sentence ϕ in \mathcal{L}_Σ , there exists a set $\mathcal{A}_\phi \subseteq \mathcal{A}_b$ such that $\phi = \bigvee_{\phi_i \in \mathcal{A}_\phi} \phi_i$.*

Note that it is possible to represent $\phi = \bigvee_{\phi_i \models \phi} \phi_i$. That means ϕ may be represented in the form of disjunction of basic propositions and then ϕ is true if some ϕ_i is true. Consider a simple example.

Example 2. Given $\mathcal{A} = \{A, B\}$, then $\mathcal{A}_b = \{A \wedge B, A \wedge \neg B, \neg A \wedge B, \neg A \wedge \neg B\}$ and for instance, the sentence $A \rightarrow B$ can be represented in the form of disjunctions of some basic propositions in \mathcal{A}_b

$$A \rightarrow B = (A \wedge B) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B).$$

The following proposition points out a closed relationship between the set of all possible worlds and one of basic propositions.

Proposition 2. *There is an one-to-one correspondence between elements of the set of basic propositions $\mathcal{A}_b = \{b_1, \dots, b_n\}$ and ones of the set of possible worlds Ω .*

Proof. In fact, given $b_i \in \mathcal{A}_b$, $b_i = \sigma_1^i \wedge \dots \wedge \sigma_m^i$, where $\sigma_j^i \in \mathcal{A}$ or $\neg \sigma_j^i \in \mathcal{A}$ (\mathcal{A} is the set of atoms defined in Subsection 2.1) consider the column vector $\omega_i = (\omega_1^i, \dots, \omega_m^i)^t$ defined as follows

$$\omega_j^i = \begin{cases} 1 & \text{if } \sigma_j^i = a_j \\ 0 & \text{otherwise } \sigma_j^i = \neg a_j \end{cases}$$

Then ω_i is a possible world corresponding to b_i . In contrast, for every possible world ω_i , it is possible to define a basic proposition b_i corresponding to ω_i . The proposition is proven.

Note that if ϕ is a sentence and ω_j is a possible world w.r.t. $\phi_j \in \mathcal{A}_\phi$, then ϕ is true in ω_j .

2.3. Probability of a Sentence

In this subsection, we consider three ways of determining probability of a sentence according to various sample spaces.

2.3.1. Probability on Possible Worlds

First of all, we remind some notions of the theory of probability [7]. Given a sample space¹ (Ω, \mathcal{E}) , in which Ω is a finite set and $\mathcal{E} = 2^\Omega$, a function $P : \mathcal{E} \rightarrow [0, 1]$ is called to be a probability if it satisfies the following conditions:

¹Since the set \mathcal{E} referred in our considering is always a power set of Ω , for simplicity, we can call Ω to be a sample space.

- (i) $P(A) \geq 0$ for all $A \in \mathcal{E}$;
- (ii) $P(\Omega) = 1$;
- (iii) For every $A, B \in \mathcal{E}$ such that $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

Very often, the probability function is determined by means of a probability distribution on the set Ω . A probability distribution is a function $p : \Omega \rightarrow [0, 1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$. The probability of a set A is then defined to be $P(A) = \sum_{\omega \in A} p(\omega)$.

The semantics of probability of a sentence defined from a probabilistic distribution on classes of possible worlds $\bar{\Omega}$ has been proposed by Nilsson [8] in building his probabilistic logic and utilised later by Dieu [4] in developing the interval-valued probabilistic logic. Suppose that p is a probability distribution on $\bar{\Omega}$, the *probability* of a sentence $\phi \in \Sigma$ is the sum of probabilities on classes in which ϕ is true, i.e.,

$$\bar{P}(\phi) = \sum_{\bar{\omega}_i \models \phi} p(\bar{\omega}_i).$$

Another way of constructing probability is based on a probability distribution on the set of all possible worlds Ω rather than on the classes $\bar{\Omega}$. The probability of a sentence ϕ is then defined

$$P(\phi) = \sum_{\omega_i \models \phi} p(\omega_i).$$

We emphasise here that the probability of a sentence ϕ is not its truth value but its *degree of truth* or *degree of belief* in the truth of the sentence ϕ . Note that P can be defined for any sentence in the language \mathcal{L}_Σ since Ω merely depends on the set of atoms appeared in Σ . Otherwise, \bar{P} , in general, is merely defined for sentences in Σ since $\bar{\Omega}$ may change according to ϕ in the language.

2.3.2. Probability on Basic Propositions

As presented above, \mathcal{A}_b is denoted to be the set of all basic propositions generated from a set Σ of sentences and \mathcal{L}_Σ is its propositional language. Instead of basing on a probability distribution on possible worlds, the probability of a sentence can be given by means of distribution on the set \mathcal{A}_b [6]. Suppose that p is such a probability distribution. Then the probability P_b of a sentence ϕ is defined

$$P_b(\phi) = \sum_{\phi_i \in \mathcal{A}_b, \phi_i \models \phi} p(\phi_i)$$

3. EQUIVALENCE OF DEDUCTIONS IN THE INTERVAL-VALUED AND THE POINT-VALUED PROBABILISTIC LOGICS

In this section, we review the interval-valued and the point-valued probabilistic logics and point out that deductions in these logics do not depend on the selected sample spaces. In other words, deductions in these logics are equivalent.

3.1. Deduction in the Interval-valued Probabilistic Logic

Suppose that $\mathcal{B} = \{(S_i, I_i) \mid i = 1, \dots, l\}$ is an iKB, in which S_i is a sentence and I_i is a closed subinterval of the unit interval $[0, 1]$, and S is a target sentence. We review here a method of deduction developed by Dieu [4] to infer the interval value for the probability of the sentence S . Denote $\Gamma = \{S_1, \dots, S_l, S\}$ and suppose that $\bar{\Omega} = \{\bar{\omega}_1, \dots, \bar{\omega}_k\}$ is the set of all Γ -classes of possible worlds defined by Γ . Each class $\bar{\omega}_i$ is characterised by a consistent vector $(u_{1i}, \dots, u_{li}, u_i)^t$ of truth values of sentences S_1, \dots, S_l, S .

Suppose that $P = (p_1, \dots, p_k)$ is a probability distribution on $\bar{\Omega}$. The truth probability of S_i is defined to be the sum of probabilities of classes of worlds on which S_i is true, i.e.,

$$\pi(S_i) = u_{i1}p_1 + \dots + u_{ik}p_k.$$

The interval value $[\alpha, \beta]$ of S is then defined, by

$$\begin{cases} \alpha = \min_P \pi(S) = \min_P (u_1 p_1 + \dots + u_k p_k) \\ \beta = \max_P \pi(S) = \max_P (u_1 p_1 + \dots + u_k p_k) \end{cases}$$

subject to constraints

$$\begin{cases} \pi_i = u_{i1} p_1 + \dots + u_{ik} p_k \in I_i \\ \sum_{j=1}^k p_j = 1, \quad p_j \geq 0 \quad (j = 1, \dots, k) \end{cases}$$

that can be written in the form of the matrix equation

$$\Pi' = U'P \quad (1)$$

where $\Pi' = (1, \pi_1, \dots, \pi_l)^t$ and U' is the $(l+1) \times k$ -matrix constructed from U by adding a row with values 1. We call the equation (1) to be the *conditional equation*. Denote this interval $[\alpha, \beta]$ to be $F(S, \mathcal{B}, \bar{\Omega})$ - an interval deduced by means of distribution on the sample space $\bar{\Omega}$.

Similarly, let $\Omega = \{\omega_1, \dots, \omega_r\}$ be the set of all possible worlds defined by Γ and

$$\Pi' = W'P \quad (2)$$

be the conditional equation. Let $F(S, \mathcal{B}, \Omega)$ to be an interval value of S deduced from \mathcal{B} by means of distributions on the sample space Ω . The following proposition asserts that these values do not depend on sample spaces.

Proposition 3. *Suppose that \mathcal{B} is iKB and S is a sentence. Ω and $\bar{\Omega}$ are the sets of all possible worlds and classes of possible worlds, respectively, defined by S and sentences in \mathcal{B} . Then $F(S, \mathcal{B}, \Omega) = F(S, \mathcal{B}, \bar{\Omega})$.*

Proof. Suppose that $P = (p_1, \dots, p_k)$ is a probability distribution on $\bar{\Omega}$ and $\pi_i \in I_i$ such that the equation (1) holds. Let $Q = (q_1, \dots, q_r)$ be a distribution on Ω such that

$$p_i = \sum_{\omega_j \in \bar{\Omega}_i} q_j \quad (i = 1, \dots, k) \quad (3)$$

The equation (2) clearly holds. Conversely, $Q = (q_1, \dots, q_r)$ is a distribution on Ω , take P to be a distribution on $\bar{\Omega}$ determined by (3). The equation (1) then holds. From that we can deduce the requirement of proof.

Similarly, we also can define an interval value $F(S, \mathcal{B}, \mathcal{A}_b)$ of S deduced from \mathcal{B} based on probability distributions on \mathcal{A}_b . The following proposition is inferred directly from Proposition 2 and the result of Proposition 3.

Proposition 4. *Let \mathcal{B} be iKB. Then*

$$F(S, \mathcal{B}, \Omega) = F(S, \mathcal{B}, \bar{\Omega}) = F(S, \mathcal{B}, \mathcal{A}_b).$$

3.2. Deduction in Point-valued Probabilistic Logic via MEP

We first review a technique to select a probability distribution via MEP,

Suppose that

$$\mathcal{B} = \{ \langle S_i, \alpha_i \rangle \mid i = 1, \dots, l \}$$

is pKB and S is a sentence ($S \neq S_i, i = 1, \dots, l$). As above, we denote $F(S, \mathcal{B}, \bar{\Omega})$ to be a set of values of $\pi(P) = u_1 p_1 + \dots + u_k p_k$, where P varies in the domain defined by conditional equation

$$\Pi' = U'P, \quad (4)$$

Note that w.r.t. point-valued knowledge base $\Pi' = (1, \alpha_1, \dots, \alpha_l)^t$.

According to MEP, in order to obtain a single value for S , we select a distribution P such that it is the solution of the following optimization problem

$$H(P) = - \sum_{j=1}^k p_j \log p_j \rightarrow \max \quad (5)$$

which subjects to constraints defined by the conditional equation (4).

Suppose that (p_1, \dots, p_k) is a solution of the above problem. Then the probability of S is denoted by

$$F(S, \mathcal{B}, \bar{\Omega}, MEP) = u_1 p_1 + \dots + u_k p_k.$$

The method of solving the problem is given in [8]. We review briefly the way of determining the probability distribution P from the matrix U' . Let a_0, a_1, \dots, a_l be parameters for rows of U' . Each p_i is defined according to a_j by means of i^{th} -column of U'

$$p_i = a_0 \prod_{u_{i,j}=1, 1 \leq j \leq l} a_j \quad (i = 1, \dots, k). \quad (6)$$

For example,

$$U' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

then

$$p_1 = a_0 a_1 a_2, \quad p_2 = a_0 a_1, \quad p_3 = a_0 a_3, \quad p_4 = a_0 a_3.$$

Similarly, suppose that $Q = (q_1, \dots, q_r)$ is a distribution on Ω satisfying MEP, i.e.,

$$H(Q) = - \sum_{j=1}^r p_j \log p_j \rightarrow \max \quad (7)$$

which subjects to constraints defined by the conditional equation

$$\Pi' = W'Q. \quad (8)$$

The probability of S is then defined by

$$F(S, \mathcal{B}, \Omega, MEP).$$

Note that if $Q = (q_1, \dots, q_r)$ is a distribution satisfying MEP on Ω , then q_i 's are determined as similarly as in the expressions (6), and some q_i have the same representation. It is easy to prove the following proposition.

Proposition 5. *Let \mathcal{B} be pKB . Then*

$$F(S, \mathcal{B}, \Omega, MEP) = F(S, \mathcal{B}, \bar{\Omega}, MEP).$$

As stated above, proposition 2 points out that there is an one-to-one corresponding between elements of \mathcal{A}_b and Ω . The following proposition is a direct consequence of Proposition 2 and Proposition 5.

Proposition 6. *Suppose that \mathcal{B} is pKB . The probability value of S deduced from \mathcal{B} via MEP does not depend on the selected sample spaces, i.e.,*

$$F(S, \mathcal{B}, \Omega, MEP) = F(S, \mathcal{B}, \bar{\Omega}, MEP) = F(S, \mathcal{B}, \mathcal{A}_b, MEP).$$

4. CONCLUSIONS

There are various approaches to assigning a probability of a sentence in probabilistic logics. It is able to define a probability of a sentence via probabilistic distributions on the set of all possible worlds, on classes of possible worlds or on the set of basic propositions. We have showed that deductions in the point-valued probabilistic logic via MEP as well as in the interval-valued probabilistic logic do not depend on the selected sample spaces. The obtained results have been presented in Propositions 4 and 6.

Some authors, such as Dieu [4] and Nilsson [8], define the probability of a sentence based on a distribution on classes of possible worlds. Others such as Gaag [6] makes use of basic propositions for constructing the probability of a sentence. On the aspect of semantics, these logics are equivalent. However, the main difference between probabilistic logics proposed by Nilsson as well as Dieu, on one side, and Gaag, on the other side, is a definition of constraints of variables in computing the probability of a sentence. While there is no any constraint in probabilistic logics based on possible worlds given by Nilsson and Dieu, Gaag's approach allows for independency relationships between the propositional variables.

Acknowledgement

The author is grateful to Prof. Phan Dinh Dieu for invaluable criticisms and suggestions. Many thanks to Do Van Thanh for discussions that provided the initial impetus for this work.

REFERENCES

- [1] K. A. Anderson, Characterizing consistency in probabilistic logic for a class of Horn clauses, *Mathematical Programming* **66** (1994) 257-271.
- [2] F. Bacchus, A. J. Grove, J. Y. Halpern, and D. Koller, From statistical knowledge bases to degrees of belief, *Artificial Intelligence* **87** (1-2) (1996) 75-143.
- [3] C. Chang and R. C. Lee, *Symbolic Logic and Mechanical Theorem Proving*, Academic Press, 1973.
- [4] P. D. Dieu, *On a theory of interval-valued probabilistic logic*, Research Report, NCSR Vietnam, Hanoi 1991.
- [5] R. Fagin and J. Y. Halpern, Uncertainty, Belief and Probability, *Computational Intelligence* **7** (1991) 160-173.
- [6] L. Gaag, Computing probability intervals under independency constraints, In P. Bonissone, M. Henrion, L. Kanal and J. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, 1991, 457-466.
- [7] R. Kruse, E. Schwecke, J. Heinsohn, *Uncertainty and Vagueness in Knowledge Based Systems*, Springer-Verlag, Berlin Heidelberg, 1991.
- [8] N. J. Nilsson, Probabilistic logic, *Artificial Intelligence* **28** (1986) 71-78.
- [9] H. S. Stone, *Discrete Mathematical Structures and Their Applications*, Palo Alto, CA: Science Research Associates, 1973.

Received May 4, 1999

Department of Mathematic and Computer Science, Hue University
32, Le Loi, Hue, Vietnam.