J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

PROBABILISTIC LOGIC

A deductive argument is a claim of the form: If P_1, P_2, ..., and P_n are true, then C is true.

The statements P_1, P_2, ..., P_n are the premises of the argument; C is the conclusion. It is the logical structure of the collection of premises and of the conclusion that determines whether the argument is valid (i.e., whether the claim it makes is the case) or invalid.

For example, the argument with premises

P_1: If it is raining, then the ground is wet.

P_2: It is raining.

and conclusion

C: The ground is wet.

is valid. The argument may be expressed symbolically as

P_1: $R \rightarrow G$

P_2: R

C: G

For every consistent assignment of truth values to the propositional variables R and G under which the premises evaluate to true, the conclusion also evaluates to true; the implication with the conjunction of the premises as antecedent and the conclusion as consequent is a tautology, as the following truth table indicates. The argument

P_1: The barometric pressure is 28.09

P_2: The relative humidity is 98%

P_3: The temperature is 81°F

C: It is raining is not valid. It is (logically) possible for the premises to be true and the conclusion to be false simultaneously.

The term "probabilistic logic" is typically used to refer to systems of logic that permit the attachment of probabilities to the premises and the conclusion of an argument. Deduction assumes full belief in the truth of the premises of an argument. However, unless the premises are tautologies (and therefore without empirical content), one cannot be absolutely certain of their truth. Although, for practical purposes, we may accept the truth of the statement "If it is raining, then the ground is wet," we may be wrong. Without our knowing it, someone may have pitched a tent over the piece of ground we are referring to, or put a tarp over it, etc.

R	G	$(R \rightarrow G$	&	$R)$	$\rightarrow$	G
T	T	T	T	T	T	T
T	F	F	F	T	T	F
F	T	T	F	F	T	T
F	F	T	F	F	T	F

Systems of probabilistic logic vary with respect to how the uncertainty associated with the premises of an argument is propagated to the conclusion. We will first discuss the version presented by Nilsson (1).

Suppose that we assign probabilities to the premises of the "ground is wet" argument:

$$p(R \to G) = .9$$
$$p(R) = .6$$

Note that the probabilities of the premises need not sum to 1. (They are not necessarily mutually exclusive and exhaustive.) However, for consistency, it is required that

$$p(R \to G) \geq 1 - p(R)$$

because R → G is logically equivalent to $\tilde{}R$ v G,

$$p(\tilde{}R) = 1 - p(R)$$

and

$$p(\tilde{}R \vee G) \geq p(\tilde{}R)$$

From the probabilities associated with the premises, a range of probabilities can be determined for the conclusion "The ground is wet" (G). The probabilities of the premises are constraints on the probabilities of consistent assignments of truth values to the sentences $R \to G$, R and G. Each consistent assignment of truth values corresponds to a set of possible worlds: the worlds in which R is true, G is true, and $R \to G$ is true; those in which R is true, G is false, and $R \to G$ is false; etc. For these sentences, there are four sets of possible worlds. The worlds in a given set are equivalent with respect to the truth values of the three sentences. The sets of possible worlds are mutually exclusive and exhaustive. (They partition the set of all possible worlds.) Therefore, their probabilities sum to 1. The probabilities of the premises are (linear) constraints on the probabilities of the sets of possible worlds.

Nilsson's method for identifying the sets of possible worlds involves constructing a semantic tree. Each node in the (binary) tree corresponds to an assignment of truth values to some subset (empty, in the case of the root) of the premises and conclusion. The level of the root node is 0. If a node is at level k, then its children are at level $k + 1$. The sentences are ordered arbitrarily. Label them

$$S_1, S_2, \ldots, S_{n+1}$$

(for an argument with n premises). We can associate the set of all possible worlds with the root node. The left child of the root corresponds to the set of possible worlds in which S_1 is true. The right child is the set of worlds in which S_1 is false. In general, for a node X at level k, its left (resp., right) child represents the intersection of the set of worlds in which S_{k+1} is true (resp., false) with the set of worlds associated with X. A terminal node, therefore, represents a simultaneous assignment of truth values to the premises and conclusion of the argument.

Without pruning, the semantic tree will have 2^{n+1} terminal nodes (each corresponding to a truth assignment over all of the sentences of the argument) and $2^{n+2} - 1$ nodes altogether. However, not all the (full or partial) truth assignments are consistent. Each (partial) truth assignment can be checked for consistency. If the assignment is inconsistent (i.e., would require that a propositional variable simultaneously had the values false and true), the children of the node corresponding to it are not generated. For example, the partial assignment

$$R \rightarrow G/F, R/F$$

is not consistent. There is no assignment of truth values to the propositional variables R and G under which both sentences evaluate to false.

In the worst case, when the sentences are logically independent, no pruning will be possible. However, the sentences in real arguments are unlikely to be independent. (Despite this, for arguments of realistic size, the method is intractable. We will later discuss refinements that use more effectively the dependencies among the sentences and that take into account the decision-making context in which the probabilistic inference is being made.)

For the "ground is wet" argument, there are four consistent truth assignments (out of eight combinatorially possible assignments). These four assignments correspond to four sets of possible worlds (W_1, …, W_4):

	W_1	T	T	T
$R \rightarrow G$	W_2	F	T	F
R	W_3	T	F	T
G	W_4	T	F	F

Recall that

$$p(R \rightarrow G) = .9$$

and

$$p(R) = .6$$

We would like to infer the probability of G, "the ground is wet." (We might use the inferred probability to decide whether or not to put on boots before going outside. Modifications to the method that take into account the use that is to be made of the probability of the conclusion will be discussed later.) Although we do not know the probability of G, we know that its probability is the sum of the probabilities of the sets of worlds in which it is true:

$$p(G) = p(W_1) + p(W_3)$$

Similarly,

$$p(R \rightarrow G) = p(W_1) + p(W_3) + p(W_4)$$

and

$$p(R) = p(W_1) + p(W_2)$$

Let p_j abbreviate $p(W_j)$. The set of solutions to the system of linear equations below is the set of probability functions compatible with the probabilities associated with the premises of our argument:

$$\begin{aligned} p_1 + p_2 + p_3 + p_4 &= 1 \\ p_1 + p_3 + p_4 &= .9 \\ p_1 + p_2 &= .6 \end{aligned}$$

Upper and lower bounds on $p(G)$ are determined via two applications of a linear programming algorithm, minimizing and maximizing the objective function $p_1 + p_3$:

$$p(G) \in [.5, .9]$$

In general, as this example illustrates, the probabilities of the premises underdetermine the probability distribution over the partition of possible worlds; thus, there is a range of probability values for the conclusion each of which (due to the linearity of the constraints and the objective function) is compatible with the probabilities of the premises.

The method may be further generalized to allow probability ranges for the premises. Suppose

$$p(R \rightarrow G) \in [.85, .95]$$

and

$$p(R) \in [.55, .65]$$

Then the constraints are in the form of a system of linear inequalities:

$$\begin{aligned} p_1 + p_2 + p_3 + p_4 &= 1 \\ p_1 + p_3 + p_4 &\geq .85 \\ p_1 + p_3 + p_4 &\leq .95 \\ p_1 + p_2 &\geq .55 \\ p_1 + p_2 &\leq .65 \end{aligned}$$

The probability interval is now

$$p(G) \in [.4, .95]$$

The method reduces to ordinary deduction when the conclusion is logically implied by the premises; the probability interval calculated in such cases is [1, 1]. However, when the conclusion is not deductively entailed by the premises, the method does not necessarily return the interval [0, 1]. The probabilities of the premises may arbitrarily strongly constrain the probability of the conclusion. For example, the argument with the single premise $A \rightarrow B$ and conclusion $B \rightarrow A$ is not valid. However, using Nilsson's methods it can be determined that

$$p(B \rightarrow A) \in [1 - p(A \rightarrow B), 1]$$

In addition, the method may be applied to arguments in first-order logic if the sentences are first Skolemized (2). However, for some arguments in first-order logic, it will not be possible to enumerate all of the sets of possible worlds.

Suppose that we are concerned with the probability of the ground being wet because we are trying to decide whether or not to put on boots. (Analogous decision problems could be devised for some autonomous system.) Suppose that the wetness of the ground is the only factor we wish to consider. There are four different relevant outcomes: we put on boots and the ground is wet, we put on boots and the ground is not wet, etc. On the standard approach (3), an agent is assumed to have a preference ranking over the outcomes. Furthermore, the relative desirability of the outcomes can be quantified, on a scale from 0 to 1, where 1 is the utility of the most preferred outcome and 0 is the utility of the least preferred outcome. There are techniques for eliciting from a human decision maker his or her utilities for the remaining outcomes, which involve consideration of hypothetical lotteries (4). Suppose that for our decision maker, the utilities are as follows:

$$\begin{aligned} U(\text{boots, wet}) &= .6 \\ U(\text{boots, not wet}) &= .8 \\ U(\text{no boots, wet}) &= 0 \\ U(\text{no boots, not wet}) &= 1.0 \end{aligned}$$

Our task is to pick the action whose expected utility is maximized. Suppose that a is one of the actions under consideration and that $\{c_1, \ldots, c_m\}$ is a mutually exclusive and exhaustive set of conditions. The possible outcomes of action a are the pairs (a, c_j). On the assumption of independence of the conditions and the actions, the expected utility of action a is

$$\sum_{j=1}^{m} p(c_j) \times U(a, c_j)$$

Simple algebra is sufficient to show that the action “put on boots” has higher expected utility than the action “do not put on boots” whenever the probability of “the ground is wet” exceeds .25. Therefore, either of the two probability intervals calculated here for $p(G)$ is sufficient to determine the appropriate action.

For decision support systems that cannot accommodate interval or other set-valued probability representations, there are alternatives to the calculation of a probability interval for the conclusion of the argument. One possibility is to select the maximum entropy solution to the system of linear constraints. The maximum entropy solution can be calculated using standard optimization techniques (5). The entailed probability is then the sum of the components of the maximum entropy distribution corresponding to the possible worlds in which the conclusion is true. This approach and related alternatives are discussed by Kane (6) and Deutsch-McLeish (7).

Snow (8) explores reductions in the size of the semantic tree that can be achieved by exploiting redundancy in common inference patterns, for example, modus ponens with a conjunctive antecedent:

P_1: A_1
P_2: A_2
P_{n-1}: A_{n-1}
P_n: $(A_1 \;\&\; \cdots \;\&\; A_{n-1}) \rightarrow B$
C: B

The semantic tree method yields 2^n possible worlds for $n-1$ antecedents. Snow uses “don’t care” values to reduce the number of worlds to $n+2$. Similar efficiencies are possible for other patterns of inference. However, in the worst case, there are no exploitable redundancies, and the number of nodes in the semantic tree (number of unknowns in the linear program) is exponential in the number of premises.

Frisch and Haddawy (9) have developed a deductive system for propositional probabilistic logic with an "anytime" property. The terms "anytime" and "flexible" have been used in the computing literature to refer to algorithms that provide useful output even when interrupted and whose output improves, in some sense, as increasing amounts of resources (computing time or space) are allocated to them. In their system, the probabilities associated with the premises of an argument are intervals. The probabilities are conditional probabilities; however, unconditional probabilities can be represented by conditioning on a tautology. The conclusion or "target sentence" initially is given the probability interval [0, 1]. With each inference step (application of an inference rule), the conclusion's probability interval is narrowed.

This system has the advantage of providing a usable interval that may be narrow enough for the purpose at hand relatively quickly. To illustrate, we now discuss a simple example that uses several of Frisch and Haddawy's inference rules.

Consider a decision the outcome of which is contingent on the truth or falsity of a single probabilistically entailed sentence: "It will rain this afternoon." Suppose that the actions under consideration are "Go to the beach" and "Do not go to the beach." Utilities of the four possible outcomes are

$$\begin{aligned} U(\text{Go, Rain}) &= 0 \\ U(\text{Go, No rain}) &= 1 \\ U(\text{Do not, Rain}) &= .8 \\ U(\text{Do not, No rain}) &= .2 \end{aligned}$$

The agent's knowledge is represented in part by the propositions and associated probability intervals:

(1) $p(\text{Temperature} > 85) \in [.95, 1]$
(2) $p(\text{Temperature} > 85 \rightarrow \text{Rain}) \in [.4, .6]$
(3) $p((\text{Barom. pressure} < 30 \text{ \& Humidity} > 80) \rightarrow \text{Rain}) \in [.65, .95]$
(4) $p(\text{Barom. pressure} < 30) \in [.95, 1]$
(5) $p(\text{Humidity} > 80) \in [.95, 1]$
(6) $p(\text{August} \rightarrow \text{Rain}) \in [.2, 1]$
(7) $p(\text{August}) \in [1, 1]$

"Go" maximizes expected utility when $p(\text{Rain}) \leq 0.5$; "Do not go" does so for $p(\text{Rain}) \geq .5$. Neither can be ruled out at this point. However, Frisch and Haddawy's probabilistic inference rules may be applied individually to narrow the interval for $p(\text{Rain})$ until a single admissible action emerges or until it is no longer economical to continue refining (e.g., the last train to the beach is about to leave) and a choice among the admissible actions must be made using some other criterion (e.g., use maximin—pick the action whose minimum utility over the various conditions is greatest, maximize expected utility relative to the midpoint of the probability interval, etc.).

Initially, we can deduce

(1) $p(\text{Rain}) \in [0, 1]$

from the "Trivial derivation" rule: $\vdash p(\alpha|\delta) \in [0, 1]$.

We may next apply "Forward implication propagation,"

$$p(\beta|\delta) \in [x, y], p(\beta \to \alpha|\delta) \in [u, v] \vdash p(\alpha|\delta) \in [\min(0, x + u - 1), v]$$

to statements 1 and 2, yielding

(1) p(Rain) $\in$ [.35, .6]

Although it does not have any effect at this stage, the multiple derivation rule should be applied to maintain the tightest interval for the target sentence:

$$p(\alpha|\delta) \in [x, y], p(\alpha|\delta) \in [u, v] \vdash p(\alpha|\delta) \in [\max(x, u), \min(y, v)]$$

Because .5 $\in$ [.35, .6], both actions remain admissible. Next, "Conjunction introduction",

$$p(\alpha|\delta) \in [x, y], p(\beta|\delta) \in [u, v] \vdash p(\alpha \,\&\, \beta|\delta) \in [\max(0, x + u - 1), \min(y, v)]$$

is applied to statements 4 and 5, yielding

(1) p(Barom. pressure $<$ 30 & Humidity $>$ 80) $\in$ [.9, 1].

Applying forward implication propagation to statements 3 and 10 gives

(1) p(Rain) $\in$ [.55, .95]

Although combining statement 11 with statement 9 via the multiple derivation rule will further narrow the target interval, there is no need to do so; nor is there any need to consider statements 6 and 7. "Do not go" has emerged as uniquely admissible.

Nilsson's methods may themselves be modified to yield an anytime procedure for decision making (10). Rather than construct the linear system corresponding to the full set of sentences, increasingly larger systems may be constructed by adding sentences to the subset currently in use until a single admissible action emerges or it is necessary to choose among the currently admissible actions.

This may be illustrated with the preceding sentences and decision problem. Suppose that sentences 3 and 5 are chosen for the first iteration. Using Nilsson's semantic tree method, five sets of possible worlds are identified. Both actions are admissible. "Go" is admissible because feasible solutions to the following system of linear inequalities exist. Where p_i is the probability of set W_i of possible worlds; "Rain" is true in sets W_3 and

W_5, "Humidity > 80" is true in sets W_1, W_4 and W_5, etc.:

$$\begin{aligned}
&p_1 + p_2 + p_3 + p_4 + p_5 = 1\\
&p_2 + p_3 + p_4 + p_5 \geq .65\\
&p_2 + p_3 + p_4 + p_5 \leq .95\\
&p_1 + p_4 + p_5 \geq .95\\
&(p_3 + p_5) \times 0 + (p_1 + p_2 + p_4) \times 1 \geq (p_3 + p_5) \times 0.8\\
&\qquad + (p_1 + p_2 + p_4) \times 0.2
\end{aligned}$$

"Do not go" is also admissible because the system resulting from reversing the direction of the final inequality also has feasible solutions.

Now add sentence 4. The resulting eight sets of possible worlds may be determined by expanding only the "live" terminal nodes of the semantic tree constructed at the first iteration. (To eliminate the need for a row interchange, the root of the initial tree should represent the target sentence. One may proceed in this way as long as is necessary.) "Do not go" is now identified as uniquely admissible; feasible solutions to the following system exist but do not exist for the corresponding system for "Go":

$$\begin{aligned}
&p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 + p_8 = 1\\
&p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 \geq .65\\
&p_1 + p_2 + p_3 + p_4 + p_5 + p_6 + p_7 \leq .95\\
&p_1 + p_2 + p_6 + p_8 \geq .95\\
&p_1 + p_3 + p_5 + p_8 \geq .95\\
&(p_1 + p_2 + p_3 + p_4) \times 0.8 + (p_5 + p_6 + p_7 + p_8) \times 0.2\\
&\qquad \geq (p_1 + p_2 + p_3 + p_4)\\
&\qquad\quad \times 0 + (p_5 + p_6 + p_7 + p_8) \times 1
\end{aligned}$$

Frisch and Haddawy's system is applicable to decision problems with an arbitrary number m of mutually exclusive conditions. The $[\frac{1}{2}\, m(m-1) + 1]$ statements

$$\begin{aligned}
&p(c_1 v \cdots v c_m) \in [1, 1]\\
&p(c_1 \& c_2) \in [0, 0]\\
&\vdots\\
&p(c_{m-1} \& c_m) \in [0, 0]
\end{aligned}$$

must be included to express the mutual exclusivity of the conditions c_j. Intervals must be maintained for each of the conditions. The soundness of Frisch and Haddawy's inference rules guarantees that, at any time, the interval $[l_j, u_j]$ associated with any c_j is a superset of the tightest interval entailed (algebraically) by the full collection of sentences. Thus, the sharpest intervals available at any time yield a linear system from which it can be determined whether an action would not be admissible relative to the sharper probability bounds computable at any later time; action a_i is (ultimately) admissible only if there exist feasible solutions to

$$\begin{aligned}
&p(c_1) + \cdots + p(c_m) = 1\\
&p(c_1) \geq l_1\\
&p(c_1) \leq u_1\\
&\vdots\\
&p(c_m) \geq l_m\\
&p(c_m) \leq u_m\\
&p(c_1) \times U(a_i, c_1) + \cdots + p(c_m) \times U(a_i, c_m)\\
&\qquad \geq p(c_1) \times U(a_1, c_1) + \cdots + p(c_m) \times U(a_1, c_m)\\
&\vdots\\
&p(c_1) \times U(a_i, c_1) + \cdots + p(c_m) \times U(a_i, c_m)\\
&\qquad \geq p(c_1) \times U(a_k, c_1) + \cdots + p(c_m) \times U(a_k, c_m)
\end{aligned}$$

where l_j and u_j are the current bounds on $p(c_j)$ and there are k actions.

Nilsson's semantic tree method can be adapted to take into account the mutual exclusivity and exhaustiveness of multiple (i.e., more than two) conditions in a decision problem. The first m levels of the tree will correspond to the m conditions. (This facilitates the anytime adaptation of Nilsson's methods discussed previously.) At level m, there will be m live nodes, one for each of the assignments in which exactly one of the conditions is true. The remaining levels of the tree are constructed as usual.

For example, with conditions c_1, c_2, and c_3, an arbitrary number $k \geq 2$ of actions a_i, and data $p(B \to c_1) \in [0.9, 1]$ and $p(B) \in [0.8, 1]$, there are six sets of possible worlds, corresponding to the matrix

$$\begin{array}{ll}
110000 & (c_1)\\
001100 & (c_2)\\
000011 & (c_3)\\
111010 & (B \to c_1)\\
100101 & (B)
\end{array}$$

Action a_i is E-admissible iff there exist feasible solutions to the system of linear inequalities:

$$\begin{aligned}
&p_1 + \cdots + p_6 = 1\\
&p_1 + p_2 + p_3 + p_5 \geq .9\\
&p_1 + p_4 + p_6 \geq .8\\
&(p_1 + p_2) \times U(a_i, c_1) + (p_3 + p_4) \times U(a_i, c_2) + (p_5 + p_6)\\
&\qquad \times U(a_i, c_3) \geq (p_1 + p_2) \times U(a_1, c_1) + (p_3 + p_4) \times U(a_1, c_2)\\
&\qquad + (p_5 + p_6) \times U(a_1, c_3)\\
&\vdots\\
&(p_1 + p_2) \times U(a_i, c_1) + (p_3 + p_4) \times U(a_i, c_2) + (p_5 + p_6)\\
&\qquad \times U(a_i, c_3) \geq (p_1 + p_2) \times U(a_k, c_1) + (p_3 + p_4) \times U(a_k, c_2)\\
&\qquad + (p_5 + p_6) \times U(a_k, c_3)
\end{aligned}$$

Although they are not the concern of this article, it should be mentioned that there also exist logics for reasoning *about* probability (11,12). In these systems, it is possible to infer a probability statement; for example,

$$p(G) \in [.4, .95]$$

from a given collection of probability statements. However, the inferred probability statement must also be given; it is not generated by the rules of inference.

Other systems for reasoning with and about probability include Bundy's Incidence Calculus (13) and Quinlan's Inferno (14). The Incidence Calculus, although developed independently, is quite similar to Nilsson's probabilistic logic. This similarity, unfortunately, includes intractability, in the form of a "legal assignment finder" step. The Inferno system calculates probability intervals by constraint propagation over a network whose nodes are statements and whose links are relations among (not necessarily pairs of) statements. In this respect, the Inferno system resembles systems for constructing and reasoning with Bayesian and Markov networks (15).

BIBLIOGRAPHY

1. N. Nilsson, Probabilistic logic, *Artif. Intell.*, **28** (1): 71–87, 1986.
2. J. A. Robinson, *Logic: Form and Function*, New York: North-Holland, 1979.
3. P. Gardenfors, N.-E. Sahlins (eds.), *Decision, Probability, and Utility*, Cambridge, UK: Cambridge Univ. Press, 1988.
4. R. Clemen, *Making Hard Decisions*, Belmont, CA: Wadsworth, 1991.
5. P. Cheeseman, A method of computing generalized Bayesian probability values for expert systems, *Proc. Int. Joint Conf. Artif. Intell.*, 1983, pp. 198–202.
6. T. Kane, Enhancing the inference mechanism of Nilsson's probabilistic logic, *Int. J. Intell. Syst.*, **5**: 487–504, 1990.
7. M. Deutsch-McLeish, An investigation of the general solution to entailment in probabilistic logic, *Int. J. Intell. Syst.*, **5**: 477–486, 1990.
8. P. Snow, Compressed constraints in probabilistic logic and their revision, *Proc. 7th Conf. Uncertainty Artif. Intell.*, 1991, pp. 386–391.
9. A. Frisch, P. Haddawy, Anytime deduction for probabilistic logic, *Artif. Intell.*, **69** (1–2): 93–122, 1994.
10. M. Pittarelli, Anytime decision making with imprecise probabilities, *Proc. 10th Conf. Uncertainty Artif. Intell.*, 1994, pp. 470–477.
11. F. Bacchus, *Representing and Reasoning with Probabilistic Knowledge*, Cambridge, MA: MIT Press, 1990.
12. J. Halpern, An analysis of first-order logics of probability, *Proc. Int. Joint Conf. Artif. Intell.*, 1989, pp. 1375–1381.
13. A. Bundy, Incidence calculus: A mechanism for probabilistic reasoning, *J. Autom. Reasoning*, **1**: 263–283, 1985.
14. J. R. Quinlan, Inferno: A cautious approach to uncertain inference, *Comput. J.*, **26** (3): 255–267, 1983.
15. J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, San Francisco: Morgan Kaufmann, 1988.

MICHAEL PITTARELLI
SUNY Institute of Technology