# FOUNDATIONS OF PROBABILISTIC LOGIC

H. Guggenheimer
R.S. Freedman

Polytechnic University
Route 110
Farmingdale, NY 11735

## ABSTRACT

We formalize a mathematical approach to probabilistic logic for zero-order logic and derive new inequalities that are necessary and sufficient for consistent probability assignments to propositions. We prove that a complete theory of probabilistic logic requires the a priori assignment of $2^{k-1}$ probabilities for a system with k basic propositions. We also show that a proposal due to Cheeseman, namely, to regard measures of confidence in knowledge systems as expectations that are conditioned on unknown distributions, does not work in general. We demonstrate this by showing that several certainty measures proposed for expert systems are not consistent with the derived inequalities for probabilistic logics.

## I INTRODUCTION; PROBABILISTIC LOGIC

A common practice in machine reasoning systems is the assignment of a measure of plausibility to any statement. Several such assignments exist and have been the subject of experiments (Freedman and Shooman 1985; and Tong et al 1983). Some authors (Cheeseman 1985; Nilsson 1986; and Pearl 1985) have proposed probability theory, based on a classical Bayesian approach, for the basis of a formal theory of plausibility assignment for knowledge systems. (For surveys of other systems of assignment, see (Freedman 1986; Prade 1985; and Reseller 1964). Part of the problem of deriving consistent and complete probability assignments to statements concerns the formalization of the statistical dependencies between statements. Another problem concerns the definition of the correct sample space for statements.

Recently, (Nilsson 1986) has defined a sample space for propositions that results in a fundamental method of assignment, called probabilistic logic. Probabilistic logic is based on classical probability theory, with the standard operators for negation, conjunction, disjunction, and implication. Notation for these operators are given in PROLOG notation as:

**not(A)    A,B    A;B    B :- A**

Let L be a set of propositions At. A probability function on L (a probabilistic logic function) is a map

$$p: L \dashrightarrow [0,1]$$

that is subject to two rules:

R1. If A is a tautology, then $p(A)=1$.

R2. If not(A,B) is a tautology, then $p(A;B) = p(A) + p(B)$.

A classical probability theory is a triple $(S, \Omega, \mu)$ of a sample space S, a Boolean algebra of Borel sets $\Omega$, and a measure $\mu$ defined for $\Omega$. Since we are concerned here with a discrete sample space, the probability theory is given by a "probability" P that is defined on all subsets of S that satisfies

(i) $P(S) = 1$;
(ii) $P(\emptyset) = 0$, where $\emptyset$ denotes the empty set;
(iii) P(A UNION B) + P(A INTERSECTION B) = P(A) + P(B) for all A,B in S.

In order to justify the use of a probability theory in logic, one must show that conditions (1) and (2) imply that p has an interpretation as a classical probability on a relevant sample space. The correct sample space recently defined by (Nilsson 1986) is given by the following:

Definition: The Nilsson Space N(L) of a set L of propositions is the set of all consistent assignments of truth values to the elements of L.

For example, if L = (A, B, B:-A), then N(L) = (x1, x2, x3, x4) where:

x1 := (T,T,T); x2 := (T,F,F); x3 := (F,T,T); x4 := (F,F,T)

Truth value assignment is given by the truth function t, defined by:

t(A) = 1 if A is assigned truth T;
t(A) = 0 if A is assigned false F;

We note that in this example, the truth value of B:-A is uniquely determined by the truth values of A and B. Consequently, N(L) is in one-to-one correspondance with the Nilsson space for L*= (A, B). Here, L* is an example of a basic space associated with L.

Definition: A basic space associated with a set L of propositions is a maximal subset L* of L, such that the elements of L* can be assigned arbitrary truth values and then uniquely determine the truth values of all the other elements of L.

This implies that the cardinality of N(L) can be given as $card(N(L)) = 2^{card(L^*)}$.

Consequently, all basic spaces derived from a given L have the same cardinality.

We note that in the above example, neither (A, B:-A) or (B, B:-A) are basic sets, since in both cases, not all truth assignments for (A, B, B:-A) are possible: (F,F) is excluded in the first case, and (T,F) is excluded in the second case. Moreover, the value of the third proposition cannot be determined: (F,T) does not determine t(B) in the first case, and (T,T) does not determine t(A) in the second case. This implies that not every set of propositions admits a basic space. This leads to the following:

Definition: A set of propositions L is well-formed if has a basic space.

Given a proposition A in L, the indicator function for A, $I_A(x)$, is the function from N(L) to (0,1) defined by:

$I_A(x)$ is the value of $t(A)$ in the assignment x.

In our above example,

$$I_A(x_3) = 1;\ I_{B;-A}(x_2) = 0,\ I_B(x_4) = 0.$$

Definition: A Nilsson probability distribution is a classical probability defined on N(L).

Definition: The induced probability of a proposition A,p(A) is:

$$p(A) = P(t(A) = 1)$$
$$= 1^* P(t(A)=1) + 0^* P(t(A) = 0)$$
$$= \Sigma x^* I_A(x)$$
$$= E\ (I_A)$$

where E is the expectation of IA in the Nilsson probability distribution.

## II INEQUALITIES FOR PROBABILISTIC LOGIC

Properties of the indicator function are:

I1. $I_{A,B}(x) = I_A(x) * I_B(x)$

I2. $I_{A;B}(x) = \max(I_A(x), I_B(x))$

$= I_A(x) + I_B(x) - I_A(x) * I_B(x)$

I3. $I_{not(A)}(x) = 1 - I_A(x)$

These properties are also seen in the induced probabilities:

P1. $p(A,B) = p(A)^*p(B) + cov(A,B)$

P2. $p(A,B) + p(A;B) = p(A) + p(B)$

P3. $p(not(A)) = 1 - p(A)$

Definition: Let L' be a basic set of a well-formed set L of propositions. The full space L^ is the set of all conjunctions of distinct elements of L'.

Any probability assignment on L^ must satisfy a set of inequalities. Given the two tautologies A;not(A) and not(A,not(A)) we have

$$1 = p(A;not(A)) = p(A) + p(not(A))$$

Since A=(A,B);(A,not(B)) and (A,B),(A,not(B)) is the negation of a tautology,

$$p(A) = p(A,B) + p(A,not(B)),\ \text{so that}$$

$$p(A,B) \leq \min(p(A),p(B))$$

By an inductive argument, we have the following inequality constraint: For k = 1..card(L'):

C1: $0 \leq p(A_1,A_2,...,A_k) \leq \min\ (p(A_1),p(A_2),...,p(A_k)) \leq 1$,

This proves the theorem:

Theorem 1: Any probability assignment on a well-formed set L which satisfies the rules R1 and R2, must also satisfy the inequality constraint C1.

## III NECESSARY AND SUFFICIENT CONDITIONS FOR PROBABILISTIC LOGIC

Our main theorem is:

Theorem 2: A probability assignment N(L^) on L^ (that is derived from a probability assignment of a well-formed set L) is consistent with rules R1 and R2, if and only if it is the probability induced by a uniquely defined Nilsson probability distribution.

Our proof is based on classical linear algebraic and geometric methods, see (Guggenheimer 1977).

The condition in Theorem 2 is necessary since C1 is implied by P1-P3. For sufficiency, we have to show that given probabilities for elements in L^ (all conjunctives), and a set of indicator functions, there exists a unique induced Nilsson probability distribution.

Let the elements of L^ be indexed from 1 to N=card(L^). We define the vector r to be the vector of all probabilities assigned to these elements, augmented by a "1". In other words,

$$r = (p(e_1), p(e_2), ..., p(e_N), 1)^T$$

(where T denotes vector transpose). Since $N = 2^{card(L')}-1$, r is a vector in $2^{card(L)}$ dimensional space. Let x be the vector of the (unknown) Nilsson distribution (in the same space), and let H be the matrix whose values are the values of the indicator function

$$h_{ij} = I_i(x_j)$$

where i denotes the index associated with the i-th element in L^, and j denotes the index of the j-th element in N(L^). From the definition of induced Nilsson probabilility (and expectation), the matrix-vector product r = H x holds.

For example, if

L = L' = L^ = (A,B), and N(L) = {x1, x2, x3, x4} = {(T,T), (T,F), (F,T), (F,F)} then we have

$$\begin{bmatrix} p(A) \\ p(B) \\ p(A,B) \\ 1 \end{bmatrix} \begin{bmatrix} I_A(x_1) & I_A(x_2) & I_A(x_3) & I_A(x_4) \\ I_B(x_1) & I_B(x_2) & I_B(x_3) & I_B(x_4) \\ I_{A,B}(x_1) & I_{A,B}(x_2) & I_{A,B}(x_3) & I_{A,B}(x4) \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

The rows of H corresponding to $A_{i1}, A_{i2}, ..., A_{ik}$ contain $2^{card(L')-k}$ ones; all other entries are zero. The last row has all ones. In our above example, this is seen when we evaluate the indicator functions, so

$$\begin{bmatrix} p(A) \\ p(B) \\ p(A,B) \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

The zero indicator function (all propositions assigned F) is the column $(0, 0, ...0,1)^T$. The last row in H implies that $x_1 + x_2 + ... + x_N = 1$.

The matrix H can be reduced to the identity matrix by elementary matrix operations (Gaussian elimination) starting with the row corresponding to

$$A_1, A_2, ..., A_{card(L')}$$

This row is zero except for one "1". H is not singular and consequently, $x = H^{-1}r$, uniquely.

We now show that $0 \le x_k \le 1$, for vector x. All rows of H (except the last row of H that consists of all 1's) can be considered to represent vectors in the faces of the unit cube (in card(L')-1 dimensional space) in the hyperplane $x_{card(L')}=1$. Since the coordinates of these rows consist of only 0 and 1, and the rows (considered as vectors) are linearly independent, we can obtain all vertices of the unit cube by any linear combination of vectors. Consequently, given any Nilsson probability distribution n, the set of all matrix-vector products (H n) is the convex hull of the all vertices of the unit cube. This implies that the vector r is also in the convex hull of the unit cube. Since $x=H^{-1}r$ is unique, x must also be a Nilsson probability distribution. (QED)

## IV APPLICATIONS

We have shown that a complete theory of probabilistic logic based on Nilsson distributions requires the a priori assignment of $2^k-1$ probabilities for a system with k basic propositions. In practice, such an assignment is possible only for small values of k. However, we can always rely on the following:

Corrollary: All inequalities between probabilities defined on a well-formed Boolean logical system are consequences of C1.

Some examples of derived inequalities are Nilssons's inequality (D1).

D1. $p(B:-A) + p(A)-1 \le p(B) \le p(B:-A)$

D2. $1-p(A) \le p(B:-A) \le \min(p(A), p(B)) + 1-p(A)$

D3. $\max(p(not(A)),p(B)) \le p(B:-A) \le \min(1,p(not(A)) + p(B))$

D4. $p(H_o \mid E_1, E_2,...,E_j,... E_m) \le p(H_o) * \min(p(D_j \mid H_o))$.

Cheeseman (1) has proposed to regard measures of confidence attached to statements in expert systems ("uncertainty calculi") as expectations that are conditional upon unknown probability distributions. We show that this approach does not work in general by reviewing some uncertainty calculi that have been proposed and investigated in (Freedman 1986; Freedman and Shooman 1985; Prade 1985; and Tong et al 1983).

The system of Lukasiewicz-Goguen postulates the following measure of confidence c:

LG1. $c(A,B) = \max(0, c(A) + c(B) -1)$
LG2. $c(B:-A) = \min(1, c(B)/c(A))$

Rule LG1 is consistent with C1. In order to check if Rule LG2 is also consistent with probabilistic logic, we consider it as a measure of confidence for modus ponens. According to Cheeseman's approach, we interpret LG2 in the following way:

$$c(B:-A) = E(\ p(B:-A) \mid A,n)$$

where n is some unknown probability distribution. However, from the definition of conditional probability and constraint C1,

$$p(B:-A) \mid A,n) = p(A,B \mid n) / p(A \mid n)$$

$$\le \min(\ 1, p(B \mid n) / p(A \mid n).$$

LG2 then implies $p(B \mid n) = p(A,B \mid n)$ and, by LG1, $p(A \mid n) = 1$, which is not valid for arbitrary probability assignments. Consequently, LG2 with LG1 is not consistent with probabilistic logic.

The system of Dienes is based on

DU1. $c(A,B) = c(A)*c(B)$
DU2. $c(B:-A) = \max(1 - c(A), c(B))$

Rule DU1 implies that for any probabilistic interpretation

$$E(p(A \mid B,n)) = E(\ p(A) \mid n)$$

If this is applied to a Nilsson probability, DU2 shows that

$$p(B:-A) = 1 - p(A) + p(B)*p(B) > 1 - p(A)$$

Nilssons inequality D1 shows that $p(B:-A) \ge p(B)$; therefore, DU2 is valid only if $p(B) > 1 - p(A)$. Consequently, DU2 is not an assignment (compatible with any interpretation as an expectation) that is consistent with probabilistic logic.

The Zadeh-Godel assignment is

Z1. $c(A,B) = \min(c(A), c(B))$
Z2. $c(B:-A) = \min(1, c(B)/c(A))$
Z3. $c(not(A)) = 1 - c(A)$

This assignment is compatible with the interpretation

$$c(A) = \max p(A \mid n)$$

where the maximum ranges over all probability vectors n. This assignment, while consistent, seems overly optimistic. This assignment is also an example of a Dempster upper probability (Dempster 1967).

## V CONCLUSIONS AND FUTURE WORK

Our theory places probabilistic logic on a firm mathematical foundation. On this basis, we already have obtained limit theorems for Dempster theory that greatly reduce computation in Demster-Shafer theory. We are now developing probabilistic limit theorems for Bayesian, fuzzy, and rule based reasoning systems.

## REFERENCES

P. Cheeseman, "In Defense of Probability." In Proc. IJCAI-85. Palo Alto, CA, 1985, pp. 1002-1009.

A. Dempster, "Upper and Lower Probabilities Induced by a Multivalued Mapping." Ann. Math. Stat. 38 (1967) 325-339.

R. Freedman, "Algorithms for Reasoning with Uncertainty," Polytechnic Notes on Artificial Intelligence 2, Polytechnic University, Farmingdale, NY, 1986.

R. Freedman and A. Shooman, "MERLIN, A Tool for Automated Reasoning with Uncertainty." In Proc. Second Conf. on Artificial Intelligence Applications, IEEE Computer Society, December 1985.

H. Guggenheimer, Applicable Geometry. Robert E. Krieger, Huntington, NY, 1977, pp. 1-32.

N. Nilsson, "Probabilistic Logic." Artificial Intelligence, 28 (1986) 71-87.

J. Pearl, "How To Do with Probabilities What People Say You Can't." In Proc. Second Conf. on Artificial Intelligence Applications, IEEE Computer Society, December 1985, pp. 6-11.

H. Prade, "A Computational Approach to Reasoning with Applications to Expert Systems." IEEE Trans. on Pattern Analysis and Machine Intelligence. PAMI-7:3 (1985) 260-283.

N. Rescher, Many-Valued Logic. McGraw-Hill, New York, 1964.

R. Tong, D. Shapiro, J. Dean, and B. McCune, "A Comparison of Uncertainty Calculi in an Expert System for Information Retrieval." in Proc. IJCAI-83. Palo Alto, CA, 1983, pp. 194-197.