From: AAAI-88 Proceedings. Copyright ©1988, AAAI (www.aaai.org). All rights reserved.

# A Note on Probabilistic Logic

by
Dr. Mary McLeish
Departments of Computing Science and Mathematics
University of Guelph
Guelph, Ontario, Canada
N1G 2W1

**Abstract**

This paper answers a question posed by Nils Nilsson in his paper [9] on Probabilistic Logic: When is the maximum entropy solution to the entailment problem equal to the solution obtained by the projection method? Conditions are given for the relevant matrices and vectors which can be tested without actually computing the two solutions and comparing them. Examples are discussed and some comments made concerning the use and computational problems of probabilistic logic.

## 1 Introduction

Reasoning with uncertain information has received much attention lately. The central problem becomes that of combining several pieces of inexact information. A number of different schemes have been proposed ranging from systems using Bayes' rule [8], quasi-probabilistic schemes [1], the Fuzzy approach [12] and the use of Belief functions developed first by A. Dempster [3] and later by G. Shafer [11].

A recent model proposed by N. Nilsson [9] is an extension of first-order logic in which the truth values of sentences can range between 0 and 1. This author has done some earlier work investigating nonmonotonicity in this setting. [c.f. 5-7]. Nilsson develops a combination or entailment scheme for his probabilistic logic. Usually the equations that need to be solved to obtain an answer to a particular entailment problem are underconstrained. Nilsson proposes two methods of obtaining an exact solution: one involving a maximum entropy approach discussed in [2] and the other an approximation using the projection of the final entailment vector on the row space of the others. Nilsson gives an example where the two values obtained by applying these methods are equal and one where they differ. He suggests one reason which will make them differ and puts forward the question of general conditions for equality.

The next section discusses the answer to this question and Section 3 provides a detailed explanation of the examples used originally by Nilsson. It also examines another example, related to Nilsson's examples both for its properties concerning the two solutions and for its relevance in handling a common entailment problem. The solution is compared with earlier results in [9]. In the course of these examples, an alternate method of finding the maximum entropy solution is also proposed. In the remainder of this introduction, some necessary terms from [9] are explained.

This research has been supported by NSERC grant #A4515

### Definitions:

The definitions follow those given in [9] and are reviewed quickly here.

$S$ represents a finite sequence $L$ of sentences arranged in arbitrary order, e.g. $S = \{S_1, S_2 \cdots S_L\}$.

$V' = \{v_1, v_2, \ldots v_L\}$ is a valuation vector for $S$, where $'$ denotes transpose and $v_1 = 1$ if $S_i$ has value true, $= 0$ otherwise.

$V$ is consistent if it corresponds to a consistent valuation of the sentences of $S$. $v$ is the set of all consistent valuation vectors for $S$ and let $K = |v|$ (cardinality). (Note $K \leq 2^L$). Each consistent $V$ corresponds to an equivalence class of "possible worlds" in which the sentences in $S$ are true or false according to the components of $V$. Let $M$ (sentence matrix) be the $L \times M$ matrix whose columns are the vectors in $v$. Its rows will be denoted by $S$. If $P$ is the $i$'th unit column vector, $MP_i = V_i$, where $V_i$ is the $i$th vector of $v$.

Example 1.1: Let $S = (A, A \supset B, B)$ $v = \left\{\begin{pmatrix}1\\1\\1\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}\begin{pmatrix}0\\1\\1\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix}\right\}$

and $M_1 = \begin{Bmatrix}1 & 1 & 0 & 0\\1 & 0 & 1 & 1\\1 & 0 & 1 & 0\end{Bmatrix}$

However, if each of the sentences' truth values are uncertain in some sense, a probability distribution over classes of possible worlds is introduced. $P' = \{P_1, P_2, \ldots P_k\}$ with $0 \leq P_i \leq 1$ and $\sum_i P_i = 1$.

Here the $i$'th component of $P$ represents the probability that 'our' world is a member of the $i$'th class of worlds. Now a consistent probabilistic valuation vector $V$ over the sentences in $S$ is computed from the equation $V = MP$. The components of $V$ are the probabilities of the $S_i$ being true (or the probability that 'our' world is a member of one of those classes of possible worlds in which sentence $S_i$ is true).

Returning to example 1, we find that even if consistent valuations are known for the sentences $A$ and $A \supset B$, the probability of $B$ is not necessarily uniquely defined. One can determine the bounds $p(A \supset B) + p(A) - 1 \leq p(B) \leq p(A \supset B)$, which provide some restrictions. However, often a more precise value for $p(B)$ needs to be predicted. One method for doing this is explained below.

**Probabilistic Entailment**: A method using a maximum entropy approach (borrowed from P. Cheeseman [5]) is used to obtain an exact solution for $p(B)$.

The entropy function becomes $H = -P.\log P + l_1 (v_1 - S_1.P) + l_2(v_2 - S_2.P) + \cdots l_L(v_L - S_L.P)$, where the $l_i$ are Lagrange multipliers. Following Cheeseman, the solution for maximum entropy becomes

$$P_i = e^{-1}.e^{-(l_1 S_{1i})} \ldots\ldots e^{-(l_L S_{Li})}$$

If one employs this method, at least for example 1, the solution for $P' = \{p+q-1, 1-q, (1-p)/2, (1-p)/2\}$ , when $V' = \{1, p, q\}$ and thus $p(B) = p/2 + q - 1/2$.

**Projection Method**: Another method uses an approximation of $B$ by $S^*$, the projection of $B$ onto the row space of $M' = M_1$ with the last row deleted and a row of 1's inserted at the top. Then $S^* = \sum c_i S_i$ and $S^* . P = \sum c_i V_i$. Applied to the example given, $S^* = (1,0,½ , ½)$ and $P(S^*) = \frac{p}{2} + q - \frac{1}{2} = p(B)$ using the max entropy solution.

However, the entailment example of $A, B, A \cap B$ is given in [9] and here the two methods give different solutions. The following section addresses the problem of why this happens.

## 2 Conditions for Equality of the Two Solutions

From the example used in [9] of the entailment of $A, B$ and $A \cap B$, it is seen that $p(A \cap B)$ from max entropy $= pq$ (where $p(A) = p$ , $p(B) = q$) and from the projection method the result is $\frac{p}{2} + \frac{q}{2} - \frac{1}{4}$. Nilsson rightly states that these two quantities cannot be equal when the max entropy solution contains a product because the solution obtained from the projection method will always be a linear combination of the $V_i$ (here, $1, p, q$). Thus, knowledge of the maximum entropy solution is sufficient to answer the equality question. Let us state this formally as a first condition. Let $P$ stand for the solution vector for $M'X = V$ obtained using maximum entropy.

**Theorem 2.1:** Only if the maximum entropy solution $P$ can be written as a linear combination of the vectors $V_i$ can $S \cdot P = \sum c_i V_i$, where $S^* = \sum c_i S_i$ is the projection of $S$ on the row space of $M'$.

To explain the reasons for this more fully and find conditions not requiring the computation of $P$, let us look at some matrix algebra theory.

If the row vectors of $M'$ are not linearly independent, they can be reduced to an independent set by dropping any unessential ones. Consider then $M(M'M)^{-1}$. This matrix has the properties of a generalized inverse of $M'$ (referred to as $(M')^-$. [c.f. 10]. Now if $P = M(M'M)^{-1}V$, it will solve $M'X = V$. (This means $P$ is a linear combination of the $V_i$). As $S^* \cdot Q = \sum c_i V_i$ for any solution $Q$ to $M'X = V$, then $S^* \cdot M(M'M)^{-1}V$ must be the solution from the projection method to the entailment problem. Now $S \cdot P = SM(M'M)^{-1}V$ $= S \cdot M(M'M)^{-1}M'P = S^*P$ ($S$ and $S^*$ are now row vectors). This shows that if $P$ is a particular linear combination of the $v_i$, the solutions are the same. Some other obvious results hold and will be stated before more general results are given.

**Lemma 2.2**: If $P = (M')^- V$, the 2 solutions are identical.

**Lemma 2.3**: If $S^* \cdot P = S \cdot P$ the 2 solutions are identical.
This follows from a remark above with $P = Q$.

**Lemma 2.4**: If $S = S^*$ the 2 solutions are identical.

**Lemma 2.5**: If $P = P^*$ (the projection of the max entropy vector on the row space of $M'$), the two solutions are equal.
This f .5ollows from $S \cdot P = S \cdot M(M'M)^{-1}M'P$
$= S^*P$.

What is the general solution to $M'X = V$? From [4,10], it can be seen to be a particular solution plus any linear combination of solutions to the homogeneous equation. One way to express this is as $(M')^- + (H-I)Z$, where $Z$ is arbitrary and $H = (M')^- M'$. Solutions may also be obtained to the homogeneous equation by first row reducing $M'$ to a form in which the first $r$ column vectors are independent.

Then $W_1 =$ (the $r + 1$'st column, $-1, 0, \cdots, 0$)
$W_2 =$ (the $r + 2$nd column, $0, -1, \cdots, 0$)
etc.
$W_{n-r} =$ (the $n$th column, $0, 0, \cdots, -1$).

Here $dim(M') = m \times n$ and $M'$ is assumed to have rank

$r$.

Thus $P$ and $P_1 = M(M'M)^{-1}V$ must be related according to $P = P_1 + (w_1, w_2, \cdots, w_{n-r}).Z$. Thus $S\,P = S{\cdot}P_1 + S{\cdot}(w_1, w_2, \cdots, w_{n-r}).Z$. But $S{\cdot}P_1 = S^*{\cdot}P_1$ (as the projection of $P_1$ on the row space of $M_1$ equals itself). Therefore the 2 solutions are equal if and only if $S{\cdot}(w_1, w_2, \cdots, w_{n-r}) = 0$ or $S$ is orthogonal to the space of homogeneous solutions - unless of course $P = P_1$, when the 2 solutions are clearly equal ($Z$ is then identically 0). From matrix theory this implies $S$ lies in the row space of $M'$. The above actually proves the following result:

**Theorem 2.6**: The max entropy and projection solutions are equal if and only if either $P = (M')^{-}V$ or $S^* = S$.

Clearly the second condition is easy to check, however the first requires the full computation of $P$. However, if we consider how the maximum entropy solution is formed we can find an easier condition to check. Suppose $M'$ can be transformed by simple row operations (not the full Gram-Schmidt process which involves vector products of the rows) to the form $(I\ C)$, where $I$ is an $r{\times}r$ unit matrix and $C$ contains at most one 1 in any column. The row vectors are the orthogonal. $M'$ may always be row reduced to $M_s' = (I\ C)$ (if its rank is $r$), but $C$ may not have this special form.

In this situation then, the max entropy solution may be written as $M_s[a_1 \cdots a_n]'$, where the $a_i$ are the special exponential variables used in the max entropy solution [2]. Now $M_s[a_1 \cdots a_n]'$ may be made equal to $M_s(M_s'M_s)^{-1}V_s$ by letting $(a_1 \cdots a_n) = V_s$ ($V_s$ is $V$ transformed appropriately when $M'$ was converted to $(I\ C)$.) The solution for the $a_i$ is unique and this solution gives $M'P = V$. Thus $P = (M_s')^{-}V_s = (M')^{-}V$.

If $M'$ cannot be row reduced to the form just described, one obtains $p_i = a_1, i = i, \cdots r$ and then $p_j =$ products of the $a_i$ and hence the $p_i$, for $j > r$ and $i \leq r$. If $P = M(M'M)^{-1}V$, this is not possible as all the members of $P$ are linear combinations of the $V_i$. Thus we have result 2.7:

**Theorem 2.7**: $P = (M'^{-})V$ if and only if $M'$ is row reducible to the form $(I\ C)$, where each column in $C$ contains at most one 1. (So $C' = C$ transpose is in echelon form).

The next section discusses some examples in the light of these results. The final results 2.6 and 2.7 make the discovery of the equality of the 2 solutions easy to verify. The property required of $M'$ could be stated in terms of the existence of a transformation matrix which will turn $M'$ into this form, but the conditions as given are just as easy to verify. There are standard methods of row-reducing $M'$ to $(I\ C)$ and then it is just a question of checking whether or not $C$ has the desired properties.

## 3 Examples

**3.1** Consider the example of $M_1 = \begin{bmatrix}1&1&0&0\\1&0&1&1\\1&0&1&0\end{bmatrix}$ and $M' = \begin{bmatrix}1&1&1&1\\1&1&0&0\\1&0&1&1\end{bmatrix}$ used in [9]. It can be shown that $M(M'M)^{-1}\begin{bmatrix}1\\p\\q\end{bmatrix}$ does give the max entropy solution $[p+q-1, 1-q, (1-p)/2, (1-p)/2]'$.

Indeed, this could also be discovered by the fact that $M'$ can be row reduced to the echelon form $\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\end{bmatrix}$ where the column vectors never contain more than one non-zero entry (row vectors are orthonormal). Actually $B \neq S^*$ and $B{\cdot}(0\ 0\ 1\ {-1})$, where $(0\ 0\ 1\ {-1}) = w_1$ is not zero. However, the row reduction immediately gives the max entropy variables $(a_1, a_2, a_3)$

$$\text{as } (M'M)^{-1}\begin{pmatrix}q+p-1\\1-q\\1-p\end{pmatrix} = \begin{pmatrix}1&0&0\\0&1&0\\0&0&½\end{pmatrix}\begin{pmatrix}q+p-1\\1-q\\1-p\end{pmatrix} = \begin{pmatrix}q+p-1\\ \frac{1-q}{2}\\ \frac{1-p}{2}\end{pmatrix}$$

Then $M(a_1a_2a_3)' = (p_1p_2p_3p_4)'$ produces the max entropy solution shown above.

**3.2.** Consider the example of $S_1 = A$, $S_2 = B$, $S_3 = A \cap B$ given in [9]. Let $S = S_3$. $M' = \begin{bmatrix}1&1&1&1\\1&1&0&0\\1&0&1&0\end{bmatrix}$, which has a row-echelon form $\begin{bmatrix}1&0&0&-1\\0&1&0&1\\0&0&1&1\end{bmatrix}$ where $(v_1, v_2, v_3)$ have become $((v_3 - v_1 + v_2), (v_1 - v_3), (v_1 - v_2))$. A solution to the homogeneous system may be obtained from the last column of $M' : \beta = (-1, 1, 1, -1)$ (See page 6, [10]). The dimension of the solution space should be $n-r$ where $M'$ is an $m{\times}n$ matrix of rank $r$. So all homogeneous solutions are of the form $k\beta$ for some constant $\beta$.

Thus the max entropy solution and the solution using the generalized inverse $M'^{-}$ introduced earlier differ by $k(-1, 1, 1, -1)$. The vector $S$ is $(1\ 0\ 0\ 0)$, which is not orthogonal to $k\beta$. Thus the solutions for $p(S)$ will not be the same. One could also check that $S$

is not in the subspace generated by the row vectors of $M'$ as $N$, the matrix obtained by adding the row vector $S$ to the matrix $M'$ has non-zero determinant.

It is interesting to see how the different solutions are related.

The particular solution found by using the generalized inverse $(M')^{-}$ is:

$$M(M'M)^{-1}\begin{pmatrix} q-1+p \\ 1-q \\ 1-p \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \\ -1 & 1 & 1 \end{pmatrix}\begin{pmatrix} q-1+p \\ 1-q \\ 1-p \end{pmatrix}$$

$$= \frac{1}{4}\begin{pmatrix} 2q+2p-1 \\ -2q+2p+1 \\ 2q-2p+1 \\ -2q-2p+3 \end{pmatrix} = P_1 \ .$$

Now $P = k[-1\ 1\ 1\ -1]' + P_1$ for some $k$. Indeed if $k = \dfrac{4pq - 2p - 2p + 1}{4}$, a solution is found.

This computation suggests that a max entropy solution can always be found for $k$ from the particular and homogeneous solutions: (i.e.) using the reduced form of $M'$, the max entropy solution for the $p_i$ in terms of $a_i$ (notation as in [9]), becomes

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dfrac{a_2 a_3}{a_1} \end{bmatrix} = k\begin{pmatrix} -1 \\ 1 \\ 1 \\ -1 \end{pmatrix} + P_1$$

Let the rows of $P'$ be $b_1$, $b_2$, $b_3$, $b_4$. Then $a_1 = -k + b_1$, $a_2 = k + b_2$, $a_3 = k + b_3$ and then $(k + b_1)(k + b_3) = (-k + b_4)(-k + b_1)$, from which $k$ can be easily found and thus $a_1$, $a_2$ and $a_3$.

For this $M'$ matrix then, the only way in which $S^*{\cdot}P_1 = S{\cdot}P$ is for $S{\cdot}(-1\ 1\ 1\ -1)'$ to be zero. This also means that $S$ should lie in the row space of $M'$ and equal its own projection on this row space. If $N$ is a square matrix, then finding $\det(N)$ quickly produces an answer. Otherwise, it is probably easier to compute the reduced form of $M'$ and check if $S{\cdot}W_i = 0$ for each of the $W_i$ computed from the last $n-r$ column vectors of the reduced form of $M'$ by adding $\underbrace{(-1\ 0\ 0\ \cdots\ 0)}_{n-r}$, $(0, -1, 0\ \cdots\ 0) \ldots (0, 0, \cdots -1)$ to them.

**3.3** As another example, consider the following scheme where $Ax : x$ is a bird, $Bx : x$ flies, and consider the entailment:

$$\begin{array}{l} A\ (\text{Tweety}) \\ \forall\ x\ [Ax \rightarrow Bx] \\ B\ (\text{Tweety}) \end{array}$$

Note that this is not the same as the sequence: $\exists y A(y)$, $\forall x[A(x) \rightarrow B(x)]$, $(\exists z)(Bz)$ represented by the $M$ matrix in [9]. Nor is it the same as

$$\begin{array}{l} S_1 : A\ (\text{Tweety}) \\ S_2 : A\ (\text{Tweety}) \rightarrow B\ (\text{Tweety}) \\ S_3 : B\ (\text{Tweety})\ , \end{array}$$

which was also investigated in [9] this. Let $p$, $q$ and $r$ be the probabilities of $S_1$, $S_2$ and $S_3$ respectively. So $p$ represents the probability Tweety is a bird (we are assuming a universe of animals for example), $q$ represents the probability that all birds fly and $r$ represents the probability that Tweety flies.

Now $M_1$ the matrix of consistent vectors becomes,

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

and $M' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$, which can be reduced to the echelon form: $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$. Therefore $P_1$ will not equal $P$.

Consider $B{\cdot}w_i$, where $w_1 = (0\ 0\ 1\ -1\ 0\ 0\ 0)$, $w_2 = (0\ 1\ 0\ 0\ -1\ 0\ 0)$, $w_3 = (-1, 1, 1, 0\ 0\ -1\ 0)$, $w_4 = (-1\ 1\ 1\ 0\ 0\ 0\ -1)$ are the solutions to the homogeneous equations. ($B = (1\ 0\ 1\ 0\ 1\ 1\ 0)$, following Nilsson's terminology here). $B{\cdot}w_1 = 0$, $B{\cdot}w_2 = -1$, $B{\cdot}w_3 = -1$ and $B{\cdot}w_4 = 0$. Therefore $B{\cdot}P'$ will not equal $B{\cdot}P$. Indeed, the projection of $B$ on the row vectors of $M'$ is (.8 .6 .6 .6 .6 .4 .4), (not a very good approximation).

The solution using the generalized inverse discussed earlier is $r' = \dfrac{2 + p + q}{5}$. The max entropy solution is very complicated resulting in

$$r = \frac{(p+q)+\sqrt{(p+q)^2+4(1-p)(1-q)}}{4} \ .$$

When $p = q = 1$, $r = 1$, unlike the solution $r'$, which equals .8, not a very sensible result. When $p = q = 0$, $r' = \dfrac{2}{5}$ and $r = \dfrac{1}{2}$. These values represent the probability tweety flies if it is not a bird and not all birds fly. When $q = 1, r' = \dfrac{3 + p}{5}$ and $r = \dfrac{1}{2}(1 + p)$. The fact that $r$ is greater than $p$ is reasonable as this model allows for a non-zero chance that non-birds fly. It is bounded below by ½ because if $p = 0$ (Tweety is not a bird), $r = ½$.

Note that the maximum entropy solution here does have some properties which are more desirable than using the system

$$S_1 : A \text{ (Tweety)}$$
$$S_2 : A \text{ (Tweety)} \rightarrow B \text{ (Tweety)}$$
$$S_3 : B \text{ (Tweety)}$$

and the scheme in [9] giving $r = \frac{p}{2} + q - \frac{1}{2}$. When $p = q = 0$, $r$ is undefined and when $p = 0$, $r = q - \frac{1}{2}$, which could be negative. Of course, which solution is the more acceptable depends on the chosen interpretation in probabilistic logic of the problem being considered. If you have an animal and know the probability of its being a bird, and know the probability that all birds fly and wish to discover if that animal flies, then the scheme given here in example 3.3 is reasonable.

## 4 Conclusions

This note has given the conditions under which the maximum entropy and projection methods discussed in [9] produce the some entailment results. These conditions can be applied without actually computing the two solutions and comparing them.

It might be interesting to obtain an idea from the general solution to the type of underconstrained equations encounted here to study $p(S)$. That is, the possible solutions for $p(S)$ are $S \cdot (M(M'M)^{-1}V + (H - I)Z$ for any vector $Z$ and this bracketed value may be computed with relative ease. The maximum entropy solution will be included amongst these values.

## References.


[1] Buchanan, B.G., Shortliffe, E.H., *Rule-based expert systems,* Addison-Wesley, (May, 1985).

[2] Cheeseman, P., "A method of computing generalized Bayesian probability values for expert systems". Proc. Eighth International joint conference on Artificial Intelligence, Karlsruhe, (1983) pp. 198-292.

[3] Dempster, A.P., "A generalization of Bayesian inference", Journal of the Royal Statistical Society, series B, vol. 30, (1968), pp. 205-247.

[4] Hoffmann, K., Kunze, R., "Linear Algebra", Prentice-Hall, 1970.

[5] McLeish, M., "Dealing with uncertainty in nonmonotonic reasoning". Proceedings of the Conference on Intelligent Systems with Machines, April (1986), Oakland U., Michigan pp. 1-5.

[6] McLeish, M., "Nilsson's probabilistic entailment extended to Dempster-Shafer Theory", Proc. of the Uncertainty Management Workshop, Seattle, 1987, (submitted to Uncertainty in A.I. Book, North Holland)

[7] McLeish, M. "Probabilistic logic: some comments and possible use for nonmonotonic reasoning", to appear in *Uncertainty in Artificial Intelligence* edited by J. Lemmer, North Holland, April, 1988.

[8] Duda, R.O., Hart, P.E., Nilsson, N.J., "Subjective Bayesian methods for rule-based inference systems", Proc 1976 National Computer Conference, AFIPS, vol. 45, pp. 1075-1082.

[9] Nilsson, N., "Probabilistic Logic", SRI technical report #321 (1984) and Artificial Intelligence, vol. 28, Feb. (1986), pp. 71-87.

[10] Rao, C.R., "Linear Statistical Inference and Its Applications", John Wiley, 1973.

[11] Shafer, G.A., *Mathematical Theory of Evidence,* Princeton University Press, Princeton, N.J. (1979).

[12] Zadeh, L.A., "Fuzzy logic and approximate reasoning", Synthese 30: (1975), pp. 407-428.