Maximum Entropy in Nilsson's Probabilistic Logic

Thomas B. Kane
Department of Computer Science,
Heriot-Watt University,
79, Grassmarket,
Edinburgh EH1 2HJ,
Scotland.

Abstract

Nilsson's Probabilistic Logic is a set theoretic mechanism for reasoning with uncertainty. We propose a new way of looking at the probability constraints enforced by the framework, which allows the expert to include conditional probabilities in the semantic tree, thus making Probabilistic Logic more expressive. An algorithm is presented which will find the maximum entropy point probability for a rule of entailment without resorting to solution by iterative approximation. The algorithm works for both the propositional and the predicate logic. Also presented are a number of methods for employing the conditional probabilities.

1. Introduction

A recent trend in reasoning with uncertainty has been to move away from representing the uncertainty of a sentence with a point probability, towards more complex mechanisms. Most notably with Probabilistic Logic [Nilsson 1986a, Grosofl986a, Guggcnheimer 1987a], Incidence Calculus [Bundy 1986a, Bundyl986b], and stochastic simulation [Pearl 1987a]. All of these systems involve explicit knowledge of possible world scenarios. In Incidence Calculus, the probability of a sentence is based on a sample space of points. Each of these points can be regarded as a possible world. In Pearl's stochastic simulation probabilities of events are computed by recording the fraction of time that events occur in a random series of scenarios generated from some causal model. Probabilistic Logic is a generalisation of the ordinary true-false semantics for logical sentences to a semantics that allows sentences to be uncertain, and consequently to have more than one possible state.

The consequences of their set-theoretic nature leaves these systems prey to complexity problems in space and time. Bundy's Legal assignment finder, which finds all the legal specialisations of an initial probability assignment is at least exponential. The number of runs it takes to approximate correct probability values in the stochastic simulator is of the same order, as is entailment inside probabilistic logic. Nilsson reports that implementation of the full procedure for probabilistic entailment would usually be impracticable.

The maximum entropy [Levine 1979a, Bard 1982a] principle also needs knowledge of all the possible states of uncertain information, and in this respect it is related to the possible world listed above, and shares the same complexity problems.

These methods form part of what appears to be a formidable family of conceptually compelling theories of reasoning with uncertainty which suffer from the same problem: intractibility. This paper addresses this problem for Nilsson's probabilistic logic, and discusses it's use of the maximum entropy method. It is the coupling of this method to the semantic framework of probabilistic logic which is at the core of this paper. The system produced is very fast, and allows the expert to use conditional probabilities in designing the statistical distribution.

2. Entropy

Entropy [Harris 1982a], is a statistical term which has evolved from a study of thermodynamics [1977a]. It is related to the probability of a thermodynamic system being in a given state as related to the number of different molecular configurations that the system can assume in that state. Since in general a system changes spontaneously toward a more probable stale, the entropy increases accordingly. Equilibrium, or maximum entropy, is the state in which the molecules can occupy the greatest number of configurations.

More formally, the entropy of the probability mass function $p_x(x)$ may be regarded as a descriptive quantity, just as the median, mode, variance and coefficient of skewness may be regarded as descriptive parameters. The entropy of a distribution is a measure of the extent to which the probability is concentrated on a few points or dispersed over many. It is an expression of the degree of disorder of a system.

Definition

If $p_i = \Pr(X = x_i)$, where X is a discrete random variable, and $p_i \geq 0$, i=1,2,...,n, and $\sum_{i=1}^{n} p_i = 1$, then the entropy of X is H(X) where

$$H(X) = H(p_1, p_2, \ldots, p_n)$$

$$= -\sum_{i=1}^{n} p_i \log p_i . \qquad \textbf{(Equation 1)}$$

In the examples here we will use 2 as the logbase; although any base can actually be used [Harrisl982a].

Example

We are provided with four coins and told that one of the coins is counterfeit Below there are four probability distributions given, where the coins are labelled 1 to 4, and p_n represents the probability that coin n is the counterfeit coin. The distributions are labelled Dl to D4, and the entropy is labelled H.

Distribution	p_1	p_2	p_3	p_4	H
D1	1/4	1/4	1/4	1/4	2
D2	1/2	1/6	1/6	1/6	1.8
D3	3/4	1/12	1/12	1/12	1.2
D4	1	0	0	0	0

Figure 1: Coin Example.

In this case, the distribution with maximum entropy is DI. The reduction in entropy from DI to D4 demonstrates the effect of having more information about the change in probabilistic likelihood of one of the coins over the others. D4 represents the case where there is no uncertainty as to which coin is counterfeit.

Information is embodied in each of the distributions, and we can see that the distribution which says least about the identity of the counterfeit coin is DI. This equation of information with entropy leads to the maximum entropy principle: *Of all probability distributions which satisfy the constraints imposed by the known aggregate probabilities, choose the one which has the maximum entropy or, equivalently, contains the least information.*

3. Probabilistic Entailment and Context

Nilsson defines probabilistic entailment as an analogue of logical entailment in classical logic. The rule of modus ponens allows us to use the set $(A_1\ A_1$,=>B) to deduce (B). When we have uncertainty about whether or not A_1 or A_1=>B is true, the real world, which has the real value of B, becomes a random variable, and can be one of a number of possible states. These states (possible worlds) can be produced mechanically by an exhaustive theorem prover [Changl973a], and the collected group represented called a semantic tree. In conventional set theoretic terms, this set of all possibilities is the *universal set.* In statistical terms, this set is called the *sample space* or *possibility space.*

As an example, Nilsson uses the set $(A_1, A_1$=>B) to estimate the probability of logically entailed sentence B. A complete interpretation table for the worlds which form the base set for the inference is:

Sentence	a	b	c	Probability
τ	1	1	1	1
A_1	1	1	0	π_1
A_1=>B	1	0	1	π_2

Figure 2: Interpretation Table for $(A_1, A_1$=>B).

The possible worlds are labelled with small letters a, b and c. Each possible world must eventually be assigned a non-zero probability such that if the probability of a sentence S is S, and S is true in worlds a and b, then p(a) + p(b) = S. The tautology T is true in all possible worlds and is included in the set to ensure that all the probabilities sum to 1.

Structurally, world c in this example is the world which causes concern. Nilsson presented the states for the semantic tree as:

Sentence	a	b	c	d	Probability
τ	1	1	1	1	1
A_1	1	1	0	0	π_1
A_1=>B	1	0	1	1	π_2

Figure 3: Interpretation Table Reduced from $(A_1, A_1$=>B,B).

The reasoning behind this being that in worlds a and b, B can only assume one logical value, 1 and 0 respectively. But in world c, B can logically assume either of the values 1 or 0. Hence, in figure 3, c represents the world where B is false, and d represents the world where B is true.

However, in the semantic tree for figure 3 there is no way of distinguishing between the worlds c and d, because they are the same world: i.e. where A_1 is false and the rule A_1=>B is true. Figure 3 also imposes an unnecessary condition on the relationship between the possible worlds $(-A_1, A_1$=>B, B) and (-iA_1A_1=>B, -.B), namely that they have the same probability. In this sense, figure 3 incorporates information into our reasoning process which is not necessarily true.

In appendix B we show that 2"+l possible worlds are created from the tree in figure 2, where n is the number of propositions in the antecedent list of the rule. Effectively, we are left with n equations and 2^n possible worlds to solve for. One way to remove the additional degrees of freedom is to maximise the entropy of the system. Bard [Bardl982a ,Bardl980a] presents examples which employs the notion of a semantic tree, and which illustrates the following solution methods with clear examples.

Each possible world is rewritten in terms of a multiplication of factors [Bardl982a ,Cheeseman 1983a ,Nilsson 1986a], where an unknown factor is associated with each of the sentences in the database. We shall use the following notation, with a_1 representing the factor for the tautology; and the factors a_j being associated with proposition j; and factor a_R being associated with the rule. We include the factor in the multiplication list for a possible world only if the world has a one in the corresponding row of the semantic tree. So, in figure 2, we have:

$$a=a_\tau a_1 a_R$$
$$b=a_\tau a_1$$
$$c=a_\tau a_R$$

In the case of figure 2 the equations are:

$$\begin{aligned} 1.\quad & a_\tau a_1 a_R + a_\tau a_1 + a_\tau a_R = 1 \\ 2.\quad & a_\tau a_1 a_R + a_\tau a_1 = \pi_1 \\ 3.\quad & a_\tau a_1 a_R + a_\tau a_R = \pi_2 \end{aligned}$$

Equations 2

Wheras the equations which need to be solved from figure 3 are:

$$\begin{aligned} 1.\quad & a_\tau a_1 a_R + a_\tau a_1 + 2a_\tau a_R = 1 \\ 2.\quad & a_\tau a_1 a_R + a_\tau a_1 = \pi_1 \\ 3.\quad & a_\tau a_1 a_R + 2a_\tau a_R = \pi_2 \end{aligned}$$

Equations 3: M.E. Probabilistic Equations for Fig. 3.

An attractive aspect of the distribution shown in figure 3 is that because of the assumption of equal probabilities between worlds c and d, once the relevant equations are solved, the probability of Q is simply $a_T a_1 a_R + a_T a_R$. However, in the case of figure 2, the probability of Q is $a_T a_1 a_R + x a_T a_R$, where x is some real number between 0 and 1, which indicates how likely it is that Q will be true in the context of P being false, and P=>Q being true.

It may be that the expert feels that when P is false, there is little chance of Q being true, and may therefore wish x to be small. Another way of estimating x would be to use the ambient prior probability of Q. Another way may be to base the calculation of x on how many conditions are being met in the context in which x is being applied. Another way would be just to assume that x is 0.5, and force equal probabilities between worlds c and d as in figure 3. Some of these assignment techniques are discussed in this paper, but a more complete discussion can be found in (2).

The tree reproduced in figure 2 suggests the following new method of entailment to infer probabilistically from the set $\{A_1, A_2, \cdots, A_n A_1 \wedge A_2 \wedge \cdots \wedge A_n => B\}$ to the set $\{ B \}$, where $\pi_1, \pi_2, \cdots \pi_n$ are the probabilities of the premises $A_1, A_2, \cdots, A_n$, and π_R is the probability of the rule:

1. Make the semantic tree for the base set to find the number of possible worlds.

2. Assign the probabilities $\pi_1, \pi_2, \cdots, \pi_n, \pi_R$ amongst the worlds consistently.

3. Each of these worlds provides a context in which to test the consistency of B. Find in which of these worlds B can be true, false, or either. Allow the expert to provide context splits for the worlds where B can consistently be either true or false.

The entailed probability of B is then the sum of the probabilities of all the worlds in which it can only be true, plus the respective context split proportions of the worlds where it can be either true or false.

4. The Update Method of Solution

The update method of solution, proposed in Cheeseman [Cheeseman1983a], is a particular case of the general one-point method of solution which covers all iterative methods of the form:

$$x_{n+1} = F(X_n)$$

and the method can be expressed as follows.

1. Number the equations 1...n.
2. For each a_i solve equation i in terms of the other variables.
3. Assume initial values for each of the a_i.
4. Choose the next a_i, and recalculate it in terms of the others.
5. If the change in any of the a_i's is more than ε then back to 4.
6. Otherwise stop with success.

Figure 4: Algorithm for Update Solution Method.

Where ε is a small number representing the tolerance in the approximation. This method will always converge on a solution, but the time taken is dependent on two things: (1) the number of a_i's to be solved for, and (2) the initial starting values for the a_i's.

Paris et al [Paris1988a] shows that the problem of computing these factors to a reasonable accuracy is NP-hard, and consequently such methods are probably unfeasible. The method proposed in this paper is possible because of the internal structure of entailment problems, which is described in the next section.

5. A Semantic Tree Template for the Entailment Rule

In the spirit of maximum entropy we include only the premises involved with a rule, and the rule itself- since it is only these sentences we have probabilities for. The semantic tree for any rule of the form $A_1 \wedge A_2 \wedge \cdots \wedge A_n => B$ which involves n + 1 sentences, will produce 2^n+1 worlds, rather than 2^{n+1} worlds. For completeness, the effects of the worlds on the conclusion B is shown in brackets.

1. All $A_1, A_2, \cdots A_n$ true, Rule true. (B true.)
2. All $A_1, A_2, \cdots A_n$ true, Rule false. (B false.)
3. At least one of $A_1, A_2, \cdots A_n$ false, Rule true. (B true or false.)

Figure 5: The Possible Worlds for an Entailment Base Set.

In a predicate calculus rule of the form: $\forall (x_1, x_2, \cdots, x_n). A_1(x_1) \wedge A_2(x_2) \wedge \cdots \wedge A_n(x_n) => B(x_1, x_2, \cdots x_n)$ the same number of worlds is produced in direct analogy with the above cases, except that in case 2 the conclusion can be either true or false, and so the expert is given the opportunity of providing another context split. These cases are worked through in appendix B.

6. Consistency in the Semantic Tree

The inclusion of the implication rule imposes a consistency constraint on the probabilities of the premises $A_1, A_2, \cdots A_n$. Namely, that in the world where the rule is false, (case 2), all of the premises are true. This logical necessity imposes the following probabilistic constraint:

For any probabilistic rule of the form $p(A_1 \wedge A_2 \wedge \cdots \wedge A_n => B) = \pi_R$, the probabilities of the premises are consistent iff $\forall i.(i = 1 \cdots n) p(A_i) > 1 - \pi_R$.

Simply because the probability of the world where the rule is false is $(1-\pi_R)$, and this is a possible world shared by all the premises. Consequently, the probability assigned to each premise must be at least this large, with some probability left over to be dispersed amongst it's other possible worlds. An exactly analagous constraint holds for the probabilities of the antecedents in a predicate calculus rule.

7. A New Algorithm for Solving the Maximum Entropy Equations

The algorithm can be expressed as follows. For any probabilistic rule of entailment of the form: $p(A_1 \wedge A_2 \wedge ... \wedge A_n => B) < 1$, the corresponding aggregate factors are labelled a_T, a_1,... a_n, a_R as in section 3. The probabilites are labelled $1, \pi_1, \pi_2, \ldots, \pi_n, \pi_R$ such that 1 is the probability of the tautology; $\pi_1, \ldots, \pi_n$ are the probabilities or propositions $A_1, \ldots, A_n$; and π_R is the probability of the rule. The solution is as follows:

$$a_{j(j=1,\ldots,n)}=\frac{\pi_j+\pi_R-1}{1-\pi_j} \qquad a_\tau=\frac{1-\pi_R}{\prod_{j=1}^{n}a_j} \qquad a_R=\frac{\pi_R}{a_\tau\prod_{j=1}^{n}(1+a_j)}$$

Figure 6: Algorithm for Rapid Calculations of Aggregate Factors.

Once these aggregate factors are found for any consistent probability problem, the possible worlds can be rebuilt from the appropriate multiplication of factors. Not only will we have the probability of a conclusion but also a detailed breakdown of the probabilities of the contributing possible worlds.

The proof that this algorithm is correct is inductive, and can be found in appendix A. For the case where the probability of the rule is one, we use the same reasoning on a simplified version of the semantic tree.

8. Using Context Splits

We are now in a position to use conditional probabilities in the Probabilistic Logic in the fonn of the context splits. Obviously, if the expert provides all of the context splits, Probabilistic Logic is now able to produce point probabilities from the probabilistic entailment. If the expert wishes to specify all of these for a rule of the form $A_1 \wedge A_2 \wedge \cdots \wedge A_n => B$, the system requires:

1. the probabilistic strength of the rule $p(A_1, A_2, \cdots A_n => B)$
2. 2^n-l context splits.

Figure 10: Information the Expert can Provide.

As the number n increases, providing reliable context splits will become impracticable. This section suggests some mechanisms for dealing with this problem.

1/. If we assume that the probability of B and ~B is the same in each split context, then the probability of the conclusion is simply:

$$p(B)=p(A_1 \wedge A_2 \wedge \cdots \wedge A_n)+\frac{1}{2}(\pi_R-p(A_1 \wedge A_2 \wedge \cdots \wedge A_n))$$

$$=\frac{1}{2}(\pi_R+p(A_1 \wedge A_2 \wedge \cdots \wedge A_n))$$

$$=\frac{1}{2}(\pi_R+\prod_{i=1}^{n+2}a_i) \qquad \text{Equation 4}$$

This method of assigning probability gives a very quick result.

2/. We can assign a contextual weight of 1/n to each of the sentences in the rule. e.g. for the case of n = 3, the contextual weight to each sentence is 1/3, and for the tree:

Sentence	Worlds								
	a	b	c	d	e	f	g	h	i
A_1	1	1	0	0	0	0	1	1	1
A_2	1	1	0	0	1	1	0	0	1
A_3	1	1	0	1	0	1	0	1	0
$A_1 \wedge A_2 \wedge A_3 => B$	1	0	1	1	1	1	1	1	1
B	1	1	1	1	1	1	1	1	1
split:	1	0	0	1/3	1/3	2/3	1/3	2/3	2/3

Figure 7: Rapid calculation of Probability.

$$p(B)=a+\frac{1}{3}(d+e+g)+\frac{2}{3}(f+h+i)$$

3/. Another method is to get the expert to assign contextual weights to each of the propositions A_j, to A_m as an indicator of how much the entailment of the conclusion depends on each of the individual A_i,.

So, lor example, in the rule given, the expert may assign weights: $(A_1=0.5, A_2=0.3, A_3=0.2)$, and from the tree,

$$p(B)=a+0.2d+0.3e+0.5f+0.5g+0.7h+0.8i$$

This system of Probabilistic Logic gives the expert all the necessary tools to fully design a subjective probability distribution to describe their level of expertise. All three of the above methods have been implemented in Prolog and may conveniently be used in meta-interprctcd expert systems [Sterling 1986a] for deducing maximum entropy point probabilities from uncertain information and uncertain rules of inference. A fully operational expert system using these mechanisms is described in [Kane 1989a].

9. Combination Problem

We have presented a completely sound method of providing the maximum entropy result from a probabilistic rule of inference, within the constraints of consistency. All of these rules may exist in a database independently of each other and be called on only when needed. A problem which arises is how best to combine the results of two reasoning processes, both of which it is consistent to fire, and both of which entail the same conclusion?

One solution would be to join the two rules together using the logical or-operator (since the conclusion can be entailed from either of the rules), and join the two probabilities together using the maximum entropy principle.

Example

The rules are:

Rule	Simplification	Probability
$A_1 \wedge A_2 => Z$	$\neg A_1 \vee \neg A_2 \vee B$	π_1
$A_3 \wedge A_4 => Z$	$\neg A_3 \vee \neg A_4 \vee B$	π_2

Figure 8: A Method of Combining Rules.

Logical connection of the two rules using the or-operator gives $\neg A_1 \vee \neg A_2 \vee \neg A_3 \vee \neg A_4 \vee B = A_1 \wedge A_2 \wedge A_3 \wedge A_4 => B$. And the probability of the new rule is $\pi_1+\pi_2-\pi_1 * \pi_2$, using the independence assumption which is built into the maximum entropy principle.

10. Conclusion

What we have presented is a manner of getting a point estimate from a probabilistic rule of entailment using the maximum entropy distribution. The equations are not solved iteratively, and consequently, the result is achieved very quickly. This result has more mathematical significance than a point produced by the MYCIN inference mechanism, and can be computed just as quickly.

The system as presented might be thought of as a probabilistic Prolog, and is fully operational.

We have implemented these ideas in an expert system for vision recognition of 2-dimensional objects, and the system works succesfully.

Acknowledgement

Dedicated to my dad, who showed me perseverance. With thanks to Heriot-Watt university's computer science department for help and encouragement. Particular thanks to Steven Salvini, Patrick McAndrew, Andrew Wallace, Hunter Davis; and the reviewers for pertinent suggestions.

Appendix A: Proof of Algorithm by Induction

The proof proceeds in four stages. First, to derive the expression for the world where the rule is false. Second, to show that for each of the terms $a_{(j=1,2,\cdots,n)}$ there is a direct match of terms on the numerator and denominator of the expression: $(\pi_j+\pi_R-1)\ /\ (1-\pi_j)$, i.e. all the unknown worlds where sentence A_j, is true divided by all the worlds where A_j, is false. The third stage is related to the first and allows us to solve for a_t. The fourth stage is for the final factor a_R and is based on the worlds in which the rule is true, and a recursive expression for describing the contribution of each of the possible worlds to this probability: $a_\tau a_R \prod_{j=2}^{n+1}(1+a_j) = \pi_R$

Base Case

Sentence	Possible Worlds			Equations	Probability
τ	1	1	1	$a_\tau a_1 a_R + a_\tau a_1 + a_\tau a_R =$	1
A_1	1	0	1	$a_\tau a_1 a_R + a_\tau a_1 =$	π_1
$A_1 => B$	0	1	1	$a_\tau a_1 a_R + a_\tau a_R =$	π_R

From the equations we get:

1/. $a_\tau a_1 = 1-\pi_R$

2/. $a_1 = \dfrac{a_\tau a_1 a_R}{a_\tau a_R} = \dfrac{\pi_1-(1-\pi_R)}{(1-\pi_1)} = \dfrac{\pi_1+\pi_R-1}{1-\pi_1}$

3/. $a_\tau = \dfrac{a_\tau a_1}{a_1} = \dfrac{1-\pi_R}{a_1}$

4/. $a_R = \dfrac{a_\tau a_1 a_R + a_\tau a_R}{a_\tau a_1 + a_\tau} = \dfrac{a_\tau a_R(1+a_1)}{a_\tau(1+a_1)} \quad \dfrac{\pi_R}{a_\tau(1+a_1)}$

And the above equations satisfy the algorithm with n = 1.

Step

Algorithm is true for n, prove true for n+1.

The new premise $A_{(n+1)}$ is added to the antecedent arm of the rule, and placed after premise A_n in the premise list. We now have aggregate factors $a_{(n+1)} a_R$.

1/. $a_\tau a_1 \cdots a_{(n+1)} = 1-\pi_R$

2/. In each row there are now 2^{n+1} possible worlds, where there used to be 2^n The difference between the tree for n+1 propositions and n propositions, being that in row n+1 there are now 2" l's and 2" 0\s, and the rule is pushed down to position n+2.

For the half of the tree with 0's in row n+1 we proved that there is a direct match to give each of the previous a_i 's. For the other half, we use the same enumeration, and find that the factor for proposition n+1 cancels out on top and bottom. Furthermore the numerator still only holds the worlds where sentence A_j is true, and the denominator the worlds where A_j is false. Therefore the equation still holds.

Is the formula true for new row n+1 ?

The new tree was made up of two identical copies of the old tree, one of which has a 1 in row n+1, the other of which has a zero in row n+1. Consequently, again it is possible to cancel the terms of the true worlds divided by the false worlds so that there is only a factor of $a_{(n+1)}$ left.

3/. a_τ is trivially $\dfrac{1-\pi_R}{a_1 a_2 \ldots a_n}$

4/. The expression for all the worlds in terms of the factors for n worlds is $a_\tau \prod_{j=2}^{n+1}(1+a_j)$

When we include the new row, we have a new multiplicative factor: We have two copies: one with an a_{n+1} in row n+1, and one with a 1. So the new expression for all the worlds is:

$$a_{(n+1)} a_\tau \prod_{j=2}^{n+1}(1+a_j) + a_\tau \prod_{j=2}^{n+1}(1+a_j) = (1+a_{(n+1)}) a_\tau \prod_{j=2}^{n+1}(1+a_j)$$

$$= a_\tau a_R \prod_{j=2}^{(n+1)+1}(1+a_j)$$

In the event that the probability of the rule is 1, we lose the world where the rule is false, and consequently the rule is subsumed into the tautology. In this case we have only n+1 factors, where $a_{(j=1,2,\cdots,n)} = \dfrac{\pi_j}{1-\pi_j}$, and a_τ is $\dfrac{1}{\prod_{j=1}^{n}(1+a_j)}$. The proof is a generalisation from steps 2 and 4 above, and is omitted in the interests of brevity.

Appendix B: Semantic Tree Case Analysis

Firstly consider the propositional case: { $A_1, A_2, \cdots, A_n, A_1 \wedge A_2 \wedge \cdots \wedge A_n => B$ }, {B}.

Case 1: All $A_1, A_2, \cdots, A_n$ true, and rule true. With the rule clause having all the $A_1, A_2, \cdots, A_n$ negated, the rule is continuously resolved away by each literal, eventually releasing the literal B. Only the inclusion of ~B in the set, would produce a contradiction.

Case 2: All $A_1, A_2, \cdots, A_n$ true, and rule false. The negation of the rule means that n+1 clauses replace the rule, where all of the $A_1, A_2, \cdots, A_n$ will be true, and the consequent is false. The inclusion of B in the set, would produce a contradiction.

Case 3: At least one of AA_1A_2, A_n false, and rule true. The literals in the rule cannot all be resolved away from the premises, and so no statement can be made about B from the rule. Consequently, either B on ~B will be consistent with the set. The number of worlds produced is 2"~1. (i.e. only removing the all true world from the list of posiblitics.)

There is no analogous case for case 3 where the rule is false. This is because, the rule will split into n+1 clauses, with A_1A_2, A_n all true; and B false. Consequently, if any of the A_1A_2, A_n premises are false, a contradiction is immediately produced.

Now consider the predicate calculus case:

$$\{ \exists(x_1).A_1(x_1), \exists(x_2).A_2(x_2), \cdots, \exists(x_n).A_n(x_n), \forall(x_1,x_2,\cdots,x_n).$$
$$A_1(x_1) \wedge A_2(x_2) \wedge \cdots \wedge A_n(x_n) => B(x_1,x_2,\cdots,x_n) \},$$
$$\{\exists(a_1,a_2,\cdots a_n).B(a_1,a_2,\cdots a_n)\}.$$

When this conclusion is negated, the clause produced is: $\neg B(v_1,v_2,\cdots,v_n)$, where $v_1,v_2,\cdots,v_n$ are variables. When it is just simplified the clause produced is: $B(g_1,g_2,\cdots g_n)$ where $g_1,g_2,\cdots g_n$ are constants.

Case 1: All $\exists(x_1).A_1(x_1), \exists(x_2)A_2(x_2), \cdots, \exists(x_n).A_n(x_n)$ true, and rule true.

All the antecedent propositions being true gives the list $A_1(c_1), A_2(c_2), \cdots, A_n(c_n)$ where $c_1,c_2,\cdots,c_n$ are constants. The rule being true gives $\neg A_1(x_1) \vee \cdots \vee \neg A_n(x_n) \vee B(x_1,x_2,\cdots x_n)$ where $x_1,x_2,\cdots x_n$ are variables. From these we can resolve away to produce from the rule clause: $B(c_1,c_2,\cdots c_n)$ and the inclusion of $\neg B(v_1,v_2,\cdots,v_n)$ produces nil.

Case 2: All $\exists(x_1).A_1(x_1), \exists(x_2).A_2(x_2), \cdots, \exists(x_n).A_n(x_n)$ true, and rule false. The rule converts to n+1 clauses: the $A_n(c_n)$, for i=1 to n, where c_n is a unique constant for each predicate functor A_n; plus a final clause which is $\neg B(c_1,c_2,\cdots,c_n)$. The inclusion of $B(g_1,g_2,\cdots g_n)$ would not produce nil, and neither would the inclusion of $B(v_1,v_2,\cdots,v_n)$.

Case 3: At least one of $\exists(x_1).A_1(x_1), \exists(x_2).A_2(x_2), \cdots, \exists(x_n).A_n(x_n)$ false, and rule true. A directly analogous case to case 3 above. The rule does not free any information about the consequent, and so the second set can be true or false. Again we get 2^n-1 possible worlds.

References

Nilssonl986a. N. Nilsson, "Probabilistic Logic," *Artificial Intelligence,* Vol. 28 Nr. 1, (Feb 1986).

Grosofl986a. B. Grosof, "An Inequality Paradigm for Probabilistic Logic," *Uncertainty in Artificial Intelligence,* pp. 259-275, Elsevier Science Publishers, (1986).

Guggcnheimerl987a. H. Guggenhcimcr and R.S. Freedman, "Foundations of Probabilistic Logic," *Proceedings of National Conference on Artificial Intelligence,* pp. 939-941, (1987).

Bundy 1986a. A. Bundy, *Incidence Calculus: A Mechanism for Probabilistic Reasoning,* 2, pp. 109-126, (Apr 1986).

Bundy 1986b. A. Bundy, "Correctness Criterea of Some Algorithms for Uncertain Reasoning Using Incidence Calculus," *Journal of Automated Reasoning,* Vol. 2, pp. 109-126, D. Reidel, Dordrecht, (1986).

Pearl 1987a. J. Pearl, "Evidential Reasoning Using Stochasuc Simulation of Causal Models," *Artificial Intelligence,* Vol. 32, pp. 245-257, Elsevier Science Publishers, North Holland, (1987).

Levinel979a. R.D. Levine and M. Tribus, *The Maximum Entropy Formalism,* pp. 15-118, MIT Press, Cambridge, Mass, (1979).

Bard 1982a. Y. Bard, "Maximum Entropy Principle, Classical Approach," *Encyclopedia of Statistical Sciences,* Vol.5, pp. 336-338, John Wiley & Sons, U.S.A., (1982). (Kotz and Johnson eds.) .

Harris 1982a. B. Harris, "Entropy," *Encyclopedia of statistical Sciences,* Vol. 2, pp. 512-516, John Wiley & Sons, U.S.A., (1982). (Kotz & Johnsons eds.) .

1977a. K. Mendelssohn , *The Quest for Absolute Zero (the meaning of low temperature physics),* Taylor & Francis, London, (1977).

Changl973a. C.L. Chang and R.C.T. Lee, *Symbolic Logic and Mechanical Theorem Proving,* Academic Press, New York, (1973).

Bard 1980a. Y. Bard, "A Model of Shared DASD and Multipathing," *CACM,* Vol. 23, pp. 564-572, (1980).

Cheesemanl983a. P. Cheeseman, "A Method for Computing Generalised Bayesian Probability Values for Expert Systems," *Proceedings of the International Joint Conference on Artificial Intelligence,* William Kaufman, Los Angeles, (Aug 1983).

Parisl988a. J.B. Paris and A. Vencovska, "On the Applicability of Maximum Entropy to Inexact Reasoning," *International Journal of Approximate Reasoning,* Vol. 3, pp. 1-34, (1988).

Sterling 1986a. L. Sterling and E. Shapiro, *The Art of Prolog,* pp. 311-319, MIT Press, Cambridge, Mass, (1986).

Kanel989a. T.B. Kane, P. McAndrew, and A.M. Wallace, "Object Recognition using Probabilistic Logic and Maximum Entropy," *Internal Report: 8904,* Heriot-Watt University, Computer Science, Edinburgh, (Jan 1989).