Axiomatizing Noisy-OR

Fabio G. Cozman[1]

Abstract. The Noisy-OR function is extensively used in probabilistic reasoning, and usually justified with heuristic arguments. This paper investigates sets of conditions that imply the Noisy-OR function.

1 INTRODUCTION

This paper examines the foundations of a rather popular pattern of probabilistic reasoning, the *Noisy-OR* combination function.

When building a probabilistic model, one must often deal with a variable X that depends directly on several variables $Y_1, \ldots, Y_n$. The Bayesian network fragment in Figure 1 (left) shows a possible situation, where variables Y_i are parents of X. In this case call X a *collider* [11]. To specify a Bayesian network, each collider must be associated with a probability distribution $p(X|Y_1, \ldots, Y_n)$. If all variables are binary, the complete specification of $p(X|Y_1, \ldots, Y_n)$ requires 2^n probability values. An attractive strategy is then to find methods that specify $p(X|Y_1, \ldots, Y_n)$ using relatively few parameters. The Noisy-OR function is a compact representation for the distribution of colliders when all involved variables are binary: Assume all variables have values T (for "true") and F (for "false"), and start with n probability values p_i, where p_i is the probability that $\{X = T\}$ conditional on the event that *only* Y_i is equal to T. That is, $p_i = p\big(X = T|Y_i = T, \{Y_j = F\}_{j=1,j\neq i}^{n}\big)$. The probabilities p_i are called *link probabilities*. The distribution of X conditional on $Y_1, \ldots, Y_n$ is

$$p(X = T|Y_1, \ldots, Y_n) = 1 - \prod_{i:Y_i=T} (1 - p_i). \qquad (1)$$

Given its history of good service, the Noisy-OR has been the object of intense investigation in the literature. However it does not seem that the following question has been asked so far: Is there a simple set of conditions on colliders that forces the Noisy-OR function to be adopted? An axiomatic characterization of the Noisy-OR is the purpose of this paper.

2 ARGUMENTS FOR NOISY-OR

The Noisy-OR function was proposed by Pearl (apparently at the same time as similar proposals appeared in other fields) [10]. The argument for Noisy-OR was detailed by Henrion [6], who added a *leak* probability p_L that X is T even when all Y_i are F. Other extensions of the Noisy-OR model ensued, for example Noisy-AND, Noisy-MAX [3, 12].

A particularly relevant characteristic of the Noisy-OR function is its *explaining away* property. Assuming variables $Y_1, \ldots, Y_n$ are (unconditionally) independent of each other, explaining away occurs when, for any distinct Y_i and Y_j,

$$p(Y_i = T|X = T, Y_j = T, \mathbf{Y}_{-ij}) < p(Y_i = T|X = T, \mathbf{Y}_{-ij}), \qquad (2)$$

where $\mathbf{Y}_{-ij}$ indicates arbitrary instantiations of variables $\{Y_k\}_{k=1,k\neq i,k\neq j}^{n}$ (note that Wellman and Henrion define "explaining away" in a slightly different form). Explaining away is one of several qualitative patterns of probabilistic reasoning that can be referred to as *synergy* patterns among parents of a variable [13, 14]. We note that Lucas has also put forward a thorough analysis of the interaction between qualitative patterns and several "noisy" combination functions [8].

The basic independence relations of the Noisy-OR function and generalizations were captured by Heckerman and Breese; their basic structure is depicted in Figure 1 (right) [5]. Zhang and Poole have also investigated this structure, additionally assuming that g is itself a combination of two-place functions [15].

A slightly different stream of research has focused on the use of Noisy-OR functions for combination of logical/probabilistic rules [4, 7, 9].Hence the Noisy-OR function has become a central element in the interface between logical and probabilistic reasoning.

3 AXIOMS FOR NOISY-OR

Assume that $Y_1, \ldots, Y_n$ are binary variables that are (unconditionally) independent of each other, and X is a binary variable that directly depends on $Y_1, \ldots, Y_n$. All binary variables have values T and F. Pearl argues that the following properties are desirable for a combination function [10]:

Accountability: The probability of $\{X = T\}$ must be zero if all Y_i are set to F.

Exception independence: The value of X is affected by each Y_i only through $(Y_i \wedge I_i)$, where the $I_1, \ldots, I_n$ are inhibitory variables that are (unconditionally) independent of each other and of the variables $Y_1, \ldots, Y_n$.

Exception independence implies the structure in Figure 1 (right), where g is a function of $\{Y_i \wedge I_i\}_{i=1}^{n}$. Now, g may be any function, including a probabilistic one. An additional, and rather substantive, condition that is present in the literature is to take g as a *deterministic* function:

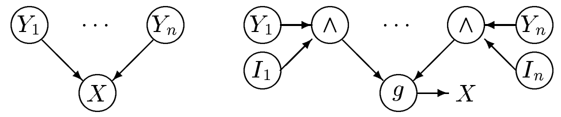

Figure 1. Left: A collider X with parents $Y_1, \ldots, Y_n$ (or the purposes of this paper, nodes and variables are equivalent). Right: Inhibitory variables and the combination function g.

[1] Escola Politécnica, Univ. de São Paulo, São Paulo, Brazil. Email: fgcozman@usp.br

Determinism: The value of X is produced by a deterministic function $g(Y_1 \wedge I_1, \ldots, Y_n \wedge I_n)$.

A quite reasonable requirement is that g be "associative" and "commutative" [15]:

Associativity: For any $\{Y_1, \ldots, Y_n\}$, $g(Y_1, \ldots, Y_n) = g(\ldots g(g(Y_i, Y_j), Y_k) \ldots, Y_l)$ (for distinct i,j,k,l), and the value of g does not change for any permutation of its inputs.

We obtain, using direct results by Lucas [8]:[2]

Theorem 1 *Accountability, exception independence, determinism, and associativity, are satisfied only by four two-place functions* $g(A, B)$*:* $\{F, A \wedge B, A \oplus B, A \vee B\}$.

Exception independence, associativity and determinism are "structural" properties that define the nature of combination functions. To proceed, further conditions must be adopted — conditions that capture the intended meaning for combination functions. A possible condition to impose on any combination function is that the function satisfies the explaining away property (2).

Explaining Away: Property (2) is satisfied as long as $Y_1, \ldots, Y_n$ are (unconditionally) independence of one another, when link probabilities and prior probabilities for the Y_i are in the open interval (0,1).

We obtain:

Theorem 2 *Accountability, exception independence, associativity, determinism and explaining away are satisfied only by two two-place functions* $g(A, B)$*:* $\{A \oplus B, A \vee B\}$.
Proof can be found in [2] (with additional discussion on the role of zero probabilities).

One possible alternative to the explaining away property is the following property, first explored by Agosta [1]:

Reverse independence If X is produced by a combination of (unconditionally) independence parents $Y_1, \ldots, Y_n$, then the parents are independent conditional on $\{X = F\}$.

Reverse independence "almost" implies the Noisy-OR function:

Theorem 3 *Accountability, exception independence, associativity, determinism and reverse independence are satisfied only by two two-place functions* $g(A, B)$*:* $\{F, A \vee B\}$.
Proof can be found in [2].

Explaining away and reverse independence do not uniquely imply the Noisy-OR function when adopted separately. However, we uniquely obtain the Noisy-OR function by adopting both properties:

Theorem 4 *Accountability, exception independence, associativity, determinism, explaining away and reverse independence are only satisfied by the Noisy-OR function.*

Thus we have identified one possible path for axiomatizing the Noisy-OR function, in the form of the six conditions in Theorem 4 (a similar conclusion can be derived from results by Lucas [8]). One may wonder whether an alternative axiomatization is possible with less conditions. Consider then the following intuitive property that is satisfied by the Noisy-OR function:

Cumulativity: If two configurations $\mathbf{Y}_1$ and $\mathbf{Y}_2$ of parents $Y_1, \ldots, Y_n$ are identical, except that some variables are set to T in $\mathbf{Y}_1$ and to F in $\mathbf{Y}_2$, then $p(X = T|\mathbf{Y}_1) > p(X = T|\mathbf{Y}_2)$ for link probabilities in the open interval $(0, 1)$.

[2] $A \oplus B$ denotes the XOR operation for A and B.

While the explaining away and reverse independence properties are found in the literature and reflect standard facts about the Noisy-OR, cumulativity is a new (albeit straightforward) assumption on combination functions. Note that, once cumulativity is assumed, accountability loses most of its appeal. In fact, accountability is not even necessary in the presence of cumulativity:

Theorem 5 *Exception independence, associativity, determinism and cumulativity are only satisfied by Noisy-OR functions.*
Proof can be found in [2].

4 CONCLUSION

To summarize, there are two sets of properties that imply the Noisy-OR function. First, accountability, exception independence, associativity, determinism, explaining away and reverse independence. Second, exception independence, associativity, determinism and cumulativity. The first set of properties contains conditions that have been long associated with the Noisy-OR function, usually in connection to causal models. The second set is more compact, and is perhaps appropriate as a common foundation for both "causal" and "rule-based" applications of the Noisy-OR function.

Acknowledgements

This work has received generous support from HP Labs and HP Brazil. The work has also been partially supported by CNPq (through grant 3000183/98-4). Thanks to the reviewers, in particular to a reviewer that pointed out the important work by Lucas [8].

REFERENCES

[1] J. M. Agosta. "Conditional inter-causally independent" node distributions, a property of "noisy-or" models. In *UAI*, pp. 9–16, 1991. Morgan Kaufmann.

[2] F. G. Cozman. Axiomatizing Noisy-OR. Technical Report USP-BT/PMR, São Paulo, Brazil, 2004.

[3] F. J. Diez. Parameter adjustment in Bayes networks: The generalized noisy OR-gate. In *UAI*, pp. 99–105, 1993. Morgan Kaufmann.

[4] S. Glesner and D. Koller. Constructing flexible dynamic belief networks from first-order probalistic knowledge bases. In *ECSQARU*, pp. 217–226, 1995.

[5] D. Heckerman and J. S. Breese. Causal independence for probability assessment and inference using Bayesian networks. Technical Report MSR-TR-94-08, Microsoft Research, March 1994.

[6] M. Henrion. Some practical issues in constructing belief networks. In *Uncertainty in Artificial Intelligence 3*. North-Holland, 1989.

[7] M. Jaeger. Relational Bayesian networks. In *UAI*, pp. 266–273, 1997. Morgan Kaufmann.

[8] P. Lucas. Bayesian network modelling by qualitative patterns. In *ECAI*, 2002.

[9] L. Ngo and P. Haddawy. Answering queries from context-sensitive probabilistic knowledge bases. *Theoretical Computer Science*, 171(1–2):147–177, 1997.

[10] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, 1988.

[11] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search (second edition)*. MIT Press, 2000.

[12] S. Srinivas. A generalization of the noisy-OR model. In *UAI*, pp. 208–215, 1993. Morgan Kaufmann.

[13] M. P. Wellman. Qualitative probabilistic networks for planning under uncertainty. *Uncertainty in Artificial Intelligence 2*, pp. 197–208, 1988.

[14] M. P. Wellman and M. Henrion. Explaining 'explaining away'. *Trans. Pattern Analysis and Machine Intelligence*, 15(3):287–307, 1993.

[15] N. L. Zhang and D. Poole. Exploiting causal independence in Bayesian network inference. *Journal of Artificial Intelligence Research*, pp. 301–328, 1996.