The Imprecise Noisy-OR Gate

Alessandro Antonucci
IDSIA
Lugano (Switzerland)
Email: alessandro@idsia.ch

***Abstract*—The noisy-OR gate is an important tool for a compact elicitation of the conditional probabilities of a Bayesian network. An imprecise-probabilistic version of this model, where sets instead of single distributions are used to model uncertainty about the inhibition of the causal factors, is proposed. This transforms the original Bayesian network into a so-called credal network. Despite the higher computational complexity generally characterizing inference on credal networks, it is possible to prove that, exactly as for Bayesian networks, the local complexity to update probabilities on an imprecise noisy-OR gate takes only linear, instead of exponential, time in the number of causes. This result is also extended to fault tree analysis and allows for a fast fusion of the causal effects on models with an imprecise-probabilistic quantification of the initiating events.**

Keywords: Noisy-OR gates, Bayesian networks, credal networks, imprecise probability, message propagation, fault trees.

I. Introduction

Bayesian networks are probabilistic graphical models particularly suited for the implementation knowledge-based expert systems. The directed graph associated to a Bayesian network depicts conditional independence relations, inducing a factorization into the joint probability mass function over the network variables.[1] The joint probabilities can be therefore expressed as a product of conditional probabilities, one for each variable given the corresponding values of the parent variables (i.e., the immediate predecessors according to the graph). Accordingly, to assess the model probabilities, instead of the (exponential number of) joint states over all the variables, we can only consider the *conditional probability tables* (CPTs) for each variable given its parents.

Yet, the number of elements in a CPT growths exponentially with the number of parents of the variable under consideration, this making potentially demanding the knowledge acquisition for Bayesian networks where there are nodes with many parents. Similarly, the number of local computations to *update* a Bayesian network by message propagation is exponential in the number of parents of the variable, this making slow inferences on a network with nodes with many parents.

A remarkable exception to this problem is when a causal relation between a variable and its parents holds.[2] If the parents of a variable can be regarded as, independent and sufficient, causes of it, an OR gate might provide a parameter-free quantification of the CPT, with the conditional probability functions assigning all the mass to the state corresponding to the Boolean disjunction of the parents. A less naive and much more realistic description can be achieved by assuming that each cause (parent) has a non-negligible probability of being *inhibited*, i.e., even if a sufficient cause is active (true), it might not be able trigger the effect (its child). This is the idea behind the so-called *noisy-OR gate* originally proposed by Pearl [11]. Under these assumptions, the CPT can be quantified only on the basis of the inhibition probabilities for each parent, these being a linear instead of exponential number of parameters. This is valuable even when computing inferences: Pearl's message-propagation algorithm for Bayesian networks updating has local complexity exponential in the number of parents, while if a noisy-OR gate is specified instead of a generic CPT, only linear time is needed [11].

The Noisy-OR gate has been indeed extended by Henrion [8], who introduced the notion of *leak probability*, i.e., a non-zero probability for the effect to be triggered even if all the causes are inactive (false). Afterward, Díez [6] and Srinivas [12] extended the model to non-Boolean variables, while a recursive version of the gate able to model synergies among the causes was proposed by Lemmer and Gossnik [9]. As a matter of fact, none of these important developments considers the elicitation of the inhibition probabilities as a critical issue.[3] Yet, these probabilities refer to scenarios where only a single cause is active, this being quite a rare situation when many causes are present. Thus, in real cases, this quantification is either based on very few data, or should be demanded to an expert domain.

Walley's seminal work in the field of *imprecise probability* [14] has pointed out how, in these situations, the standard Bayesian theory of probability could be ineffective in modeling uncertainty. Sets of, instead of single, distributions might provide a more realistic (although less informative in general) picture of the uncertainty associated to a small data set or to expert knowledge, which is mostly qualitative. Following these ideas, we propose an imprecise-probabilistic formulation

[1] We focus on models over Boolean variables. A discussion on the extension to multi-valued, but still categorical, variables is in the conclusions.

[2] A causal interpretation for the directed edges of the graph associated to a Bayesian network might be appealing but is generally wrong. The graph depicts conditional independence relations, and the edges model correlations, while causal relations are just a special case of them.

[3] A remarkable exception is the work in [10], where both Zadeh's possibility theory and Dubois' possibilistic logic approaches are considered to achieve a more reliable quantification of the inhibition process. Notably, this work was motivated by practical issues: when trying to quantify the inhibition probabilities in a network for satellite faults diagnosis, possibilistic approaches appeared as a more realistic model of experts' knowledge. Yet, unlike the work presented in this paper, that approach was not focused on algorithmic issues, this making the applicability of their possibilistic noisy-OR gates very limited.

of the noisy-OR gate where the inhibition probabilities are free to vary in an interval, this corresponding to a convex set of conditional distributions for the inhibition process.[4] Moving to the imprecise-probabilistic framework transforms the original Bayesian network into an imprecise-probabilistic graphical model called *credal network* [4]. We prove that, even in the credal networks setting, inference based on message propagation has local complexity which grows linearly, instead of exponentially, with the number of parents. This makes therefore possible to take advantage of the higher expressiveness of the imprecise noisy-OR for practical problems. The result easily extends to propagation on any kind of (deterministic, noisy, and imprecise noisy) Boolean gate. This makes it possible an application to *fault tree analysis* [13], by allowing for an efficient computation of the fused causal effects of the initiator variables, even when their quantification is based on an imprecise-probabilistic model.

As an informal introduction to these ideas, let us open the discussion with the following simple example.

Example 1: *Sadly looking at your flat tire, you identify two possible (sufficient) causes: air escape and cap deflating. Yet, assuming that the tire is flat if and only if there is an air escape* or *the cap is deflating seems to be a too naive picture. It could be much more realistic to assume that sometimes a deflating cap cannot make the tire flat, and similarly for air escape. Accordingly, you distinguish between these causes and their "effective" counterparts, the latter variables corresponding to the fact that the causal factor is not only active, but also effective in making the tire flat (see Fig. 1). Concerning data about the effectiveness of a deflating cap, you have only four records, and only in one case the deflation of the cap was ineffective. No data are available about air escapes, but on the basis of your expertise, you estimate as "very unlikely" the fact that an air escape cannot make the tire flat. On the basis of these pieces of knowledge, a sharp (frequentist) probabilistic assessment like* $\frac{1}{4}$ *for the deflating cap and* $.2$ *for the air escape seems to be somehow arbitrary and potentially inaccurate. Interval-valued estimates for these probabilities, like for instance* $[\frac{1}{5}, \frac{2}{5}]$ *and* $[.05, .25]$*, may still contain some arbitrariness, but are clearly more robust, this giving higher realism to the model.*[5]

These ideas are going to be formalized and developed in the rest of the paper. In particular, in Section II, we introduce the necessary background material. The imprecise noisy-OR is presented in Section III, while message propagation is specialized to this model in Section IV. An application to fault tree analysis in Section V. Conclusions and a number of possible directions for the development of the proposed model are discussed in Section VI.

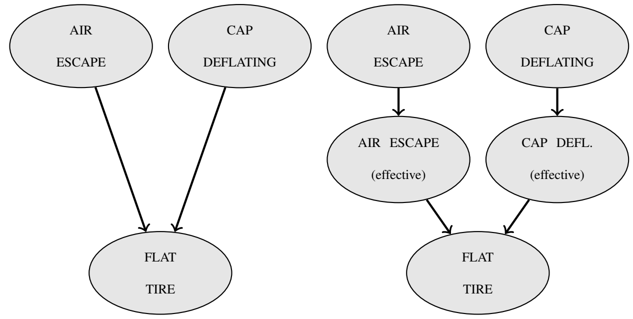

Fig. 1. The causal relation discussed in Example 1 depicted as a direct influence of the causes on the effect (left), and with the intermediation of two auxiliary variables modeling the effectiveness of the causes (right).

II. BACKGROUND

First let us introduce the necessary formalism. Given a Boolean variable X, we denote by x its generic element, while $+x$ and $-x$ are respectively the true and the false state. A probability mass function over X is denoted by $P(X)$, while $K(X)$ is *credal set*, i.e., a convex set of distributions over X. In particular we denote by $K_{l,u}(X)$ the credal set induced by the linear constraint $l \leq P(+x) \leq u$.[6] We denote by $\mathrm{ext}[K(X)]$ the *extreme points* of the credal set (i.e., the probability mass functions which cannot be expressed as a convex combination of other ones). As an example note that $\mathrm{ext}[K_{l,u}(X)] = \{[l, 1-l], [u, 1-u]\}$, where the array notation $[P(+x), P(-x)]$ is used for the elements of a mass function.

A. Bayesian and credal networks

Let us consider a directed acyclic graph $\mathcal{G}$ whose nodes are in one-to-one correspondence with a set of Boolean variables $\mathbf{X} := (X_1, \ldots, X_n)$. The Markov condition gives the semantics of $\mathcal{G}$: *every node is conditionally independent of its non-parents non-descendants given its parents*.[7] Under these assumptions, the joint probabilities factorize as follows:

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|\pi_i), \quad (1)$$

where the (joint) variable Π_i includes all the parents of X_i according to $\mathcal{G}$ and π_i is the state of this variables consistent with $(x_1, \ldots, x_n)$. A Bayesian network specification is therefore intended as the assessment of the conditional probability mass functions $P(X_i|\pi_i)$, for each $i = 1, \ldots, n$ and π_i. If, in the above specification, a credal set $K(X_i|\pi_i)$ is provided instead of $P(X_i|\pi_i)$, we say that a *credal network* has been specified instead. While a Bayesian network defines a joint probability mass function as in (1), a credal network defines a joint credal

[4]For Boolean variables, an interval constraint for one of the two states is a fully general specification of the corresponding imprecise-probabilistic model.

[5]These two interval-based imprecise-probabilistic assessments are obtained respectively with the *imprecise Dirichlet model* [15] (cautiousness parameter $s = 1$), and following the verbal-numerical scale proposed by Walley in [16].

[6]Note that any credal set over a Boolean X can be specified in this way.

[7]We say that X is (stochastically) independent of Y, if the joint probability mass function $P(X, Y)$ is such that $P(x, y) = P(x)P(y)$ for each x, y. This concept can be extended to credal sets by the notion of *strong independence*: X and Y are strongly independent if, for each $P(X, Y) \in \mathrm{ext}[K(X, Y)]$, stochastic independence holds between X and Y.

set, i.e., the following convex set of joint distributions:

$$K(\mathbf{X}) = \mathrm{CH}\left\{ P(\mathbf{X}) \;\middle|\; \begin{array}{l} P(\mathbf{x}) = \prod_{i=1}^{n} P(x_i|\pi_i), \\ P(X_i|\pi_i) \in K(X_i|\pi_i) \\ \forall i = 1,\ldots,n \quad \forall \pi_i \end{array} \right\}, \quad (2)$$

where CH denotes the convex hull operation. This credal set is called the *strong extension* of the credal network [4].

B. Pearl's message propagation and 2U algorithm

A typical inference problem we may want to consider in a Bayesian or credal network is *updating*. This corresponds to compute the posterior probabilities for a variable of interest X_q given evidential information x_E about some other variables $X_E \subset \mathbf{X}$. In the case of Bayesian networks, this means to compute the posterior probability $P(x_q|x_E)$, which can be obtained from the joint in (1) by marginalization and conditioning. Of course, in the case of credal networks only the lower and upper bounds of this posterior probability with respect to the strong extension (2) can be evaluated. We denote as $\underline{P}(x_q|x_E)$ and $\overline{P}(x_q|x_E)$ these bounds. Updating is a NP-hard task in both Bayesian and credal networks [5].

For polytree-shaped Bayesian networks, Pearl's message-propagation algorithm can efficiently compute updated probabilities. We point the reader to [11] for details about this algorithm. This propagation scheme has been extended to polytree-shaped Boolean credal networks in [7]. The resulting algorithm, called *2-Updating* (2U), is the only known exact procedure for efficient updating on credal networks.

Here we briefly review some features of the 2U algorithm by assuming the reader familiar with the main ideas of belief propagation. Let us therefore consider a polytree-shaped credal network over Boolean variables. To each $X_i \in \mathbf{X}$ associate the *lower messages* $\underline{\mu}(X_i)$ and $\underline{\Lambda_i}$, which are received respectively from the parents Π_i and the children Γ_i of X_i. 2U performs a distributed calculation of these messages, from which any updated probability can be computed as follows:[8]

$$\underline{P}(x_q|x_E) = \left(1 + \left(\frac{1}{\underline{\mu}(x_q)} - 1\right)\frac{1}{\underline{\Lambda_q}}\right)^{-1}. \quad (3)$$

In order to implement the distributed computation, also the arcs of $\mathcal{G}$ should be equipped with lower and upper messages. Thus, for each pair of nodes (X_j, X_i) such that $\mathcal{G}$ has an arc from X_j to X_i, the lower messages $\underline{\mu}_i(X_j)$ and $\underline{\Lambda}_j^i$ are also provided. The messages associated to the incoming and outgoing arcs of X_i are used to compute those associated to X_i, according to the following equations:

$$\underline{\mu}(x_i) = \min_{\substack{j:X_j\in\Pi_i \\ \mu_i(x_j)\in \\ \{\underline{\mu}_i(x_j),\overline{\mu}_i(x_j)\}}} \sum_{\pi_i} \underline{P}(x_i|\pi_i) \prod_{j:X_j\in\Pi_i} \mu_i(x_j), \quad (4)$$

$$\underline{\Lambda}_i = \prod_{k:X_k\in\Gamma_i} \underline{\Lambda}_i^k, \quad (5)$$

where x_j is the state of X_j consistent with π_i. The messages associated to the arcs are indeed computed as follows:

$$\underline{\mu}_i(x_j) = \left(1 + \left(\frac{1}{\underline{\mu}(x_j)} - 1\right)\frac{1}{\prod_{l:X_l\in\Gamma_j, l\neq i}\underline{\Lambda}_j^l}\right)^{-1}, \quad (6)$$

$$\underline{\Lambda}_i^k = \min_{\substack{l:X_l\in\Pi_i, l\neq k \\ \mu_i(x_l)\in\{\underline{\mu}_i(x_l),\overline{\mu}_i(x_l)\}}} \left(\min_{\Lambda_i\in\{\underline{\Lambda}_i,\overline{\Lambda}_i\}} \underline{c}_i^k(\Lambda_i)\right), \quad (7)$$

where the function to be minimized is

$$\underline{c}_i^k(\Lambda_i) = \begin{cases} \frac{1+(\Lambda_i-1)\overline{\rho}(x_i|x_k)}{1+(\Lambda_i-1)\underline{\rho}(x_i|x_k)} & \text{if} \quad \Lambda_i \geq 1 \\ \frac{1+(\Lambda_i-1)\underline{\rho}(x_i|x_k)}{1+(\Lambda_i-1)\overline{\rho}(x_i|x_k)} & \text{otherwise} \end{cases}, \quad (8)$$

with

$$\underline{\rho}(x_i|x_k) = \sum_{\substack{l:X_l\in\Pi_i, l\neq k \\ x_l\in\{+x_l,-x_l\}}} \underline{P}(x_i|\pi_i) \prod_{l:X_l\in\Pi_i, l\neq k} \mu_i(x_l). \quad (9)$$

Note that $\underline{\rho}(x_i|x_k)$, and hence $\underline{c}_i^k(\Lambda_k)$, are also functions of $\{\mu_i(x_l)\}_{l:X_l\in\Pi_i}$. This dependence is left implicit for the sake of notation.

Overall, (3)–(9) define a distributed algorithm that obeys the same principles of Pearl's belief updating algorithm. The global computation is carried out in discrete steps. At each step, some nodes are sending messages from a certain subset of active nodes, which modifies the set of active nodes for the next step and the procedure is repeated until no node is active. This condition is satisfied when some node has been updated about the global state of the network. In this state, the messages associated to the nodes are the final result of the computation.

Regarding computational complexity, the computation of the messages in X_i is dominated by (4) and (7). Let $p_i := |\Pi_i|$ denote the *indegree* for the node X_i. These equations require an optimization over 2^{p_i} different configurations of the messages $\mu_i(x_j)$ (for each parent X_j, we can choose the upper or lower $\mu_i(x_j)$). Moreover, the functions to be minimized are sums with 2^{p_i} terms, and for each term p_i multiplications must be performed (we need to multiply the messages of all the parents). An additional factor p_i arises in (7), because it is computed for each k. Overall, this means a time complexity $O(p_i^2 2^{2p_i})$ locally to X_i ($p_i 2^{p_i}$ is the time to compute the function for a fixed configuration of the parent messages, 2^{p_i} is the number of combinations of parent messages and p_i is the number of times (7) needs to be evaluated). As noted in [1], instead of computing (7) separately for each k, we can reuse the computations that are performed in (9) for distinct k's, that is, evaluations of (9) altogether (for all k) can be computed in time $O(p_i 2^{p_i})$ given that a configuration of the messages is fixed (just note that from a x_k to another, we can reuse the computations and spend just constant time per term of the summation instead of $O(p_i)$ time). Hence, the final complexity of 2U implementation is $O(p_i 2^{p_i})$ times the number of configurations of the messages, that is, $O(p_i 2^{2p_i})$. This is a fast implementation as the belief propagation in standard Bayesian networks already takes time $O(p_i 2^{p_i})$ to evaluate the functions.

[8] An analogous relation is provided for the upper probability $\overline{P}(x_q|x_E)$ by simply replacing the lower with the corresponding *upper messages* $\overline{\mu}(X_q)$ and $\overline{\Lambda}_q$. Similarly, any equation involving lower messages and probabilities has a corresponding upper formulation [7].

III. From precise to imprecise noisy-OR gates

Let X be a node of a Bayesian network with parents $(Y_1, \ldots, Y_m)$, where all the variables are Boolean. Assume that a causal interpretation of the effect on X of its parents can be assumed, each parent being a sufficient cause to make X true. The relative CPT can be therefore quantified by the following Boolean disjunction:

$$P(+x|y_1, \ldots, y_m) = \begin{cases} 1 & \text{if} \quad \exists\, i \,/\, y_i = +y_i \\ 0 & \text{otherwise} \end{cases} . \tag{10}$$

In practical situations, such a purely deterministic relation is often unrealistic: each cause has non-negligible probability of being *inhibited*, i.e., even if we assume the cause to be in its true state, the cause might be unable to trigger the effect to be true as well. This can be modeled by augmenting the network with an auxiliary Boolean variable, say Z_i, for each cause X_i. Variable Z_i is interpreted as the fact that variable Y_i is effective in triggering X. Accordingly, the state of X is (deterministically) affected by that of $(Z_1, \ldots, Z_m)$, while the state of each Z_i is (probabilistically) affected only by Y_i. This independence structure can be depicted in the network graph by putting Z_i in between Y_i and X (see Fig. 2). Regarding the quantification of the conditional probabilities, $P(x|z_1, \ldots, z_m)$ should be clearly defined as in (10), while for the relation between Z_i and X_i we have that an inactive cause cannot be effective, i.e., $P(+z_i|-y_i) = 0$, and

$$P(+z_i|+y_i) = p_i, \tag{11}$$

where $1 - p_i$ denotes the *inhibition probability* for the cause Y_i. In this situation, if all the causes are inactive, none of them can trigger X, which is therefore inactive too. As noted in [8], a situation where, even if all the causes are inactive, there is a non-negligible "leak" probability p_0 for X to be active, can be easily described by adding to the model another, parentless, parent of X, say Z_0 such that $P(+z_0) = p_0$. Overall, we have augmented the original Bayesian network with $m+1$ auxiliary variables. Yet, these can be marginalized out from the joint in (1) and return a Bayesian network quantification for the original set of variables and graph, for which the CPT $P(-x|y_1, \ldots, y_m)$ is:

$$\sum_{z_0, z_1, \ldots, z_m} P(-x|z_0, \ldots, z_m) P(z_0) \prod_{i=1}^{m} P(z_i|y_i) = (1 - p_0) \prod_{\substack{i=1,\ldots,m: \\ z_i = +z_i}} (1 - p_i). \tag{12}$$

The CPT quantification in (12), called (leaky) *noisy-OR gate*, allows for the specification of 2^m conditional probabilities by means of only $m+1$ parameters. Although unnecessary for its derivation, the auxiliary variables $\{Z_i\}_{i=0}^m$ facilitate a correct interpretation of the model, and are going to be crucial for its extension to imprecise probabilities.

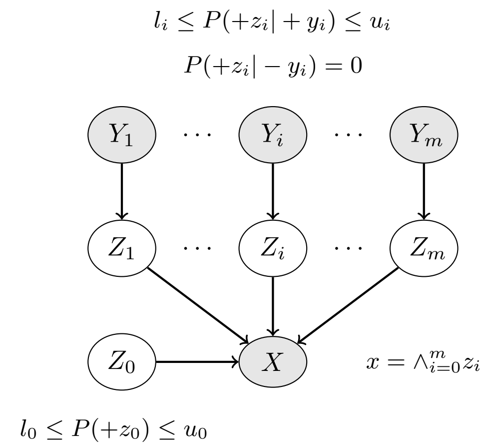

Fig. 2. The imprecise noisy-OR gate. White nodes denote auxiliary variables.

Imprecise noisy-OR gate

Assume that, when assessing the conditional probabilities $P(+z_i|+y_i)$ the single, "sharp", probabilistic assessment p_i seems to be potentially unreliable. In order to gain robustness, the linear constraint $l_i \leq P(+z_i|+y_i) \leq u_i$ could represent a more cautious choice. Similarly we can proceed for the leak probability $l_0 \leq P(+z_0) \leq u_0$. This implicitly defines the credal set $K_{l_i,u_i}(Z_i|+y_i)$ for the auxiliary variables when the cause is active, while the quantification when the cause is inactive is "precise" (being in particular deterministic). Accordingly, if a noisy-OR quantification appears as a suitable model for the CPT of a Bayesian network, but we are unable to provide reliable sharp estimates for the inhibition probabilities, we should augment the model with the auxiliary variables $\{Z_i\}_{i=0}^m$ according to the topology in Fig. 2 and adopt the above described quantification. This transforms the original Bayesian network into a credal network (defined over a wider set of variables). We call this model *imprecise noisy-OR gate*. Note that to compute inferences the auxiliary variables should be marginalized out from the credal network strong extension (2). As an example, for the bounds of the CPT we have:

$$\underline{P}(-x|y_1, \ldots, y_m) = \min_{\substack{l_0 \leq P(+z_0) \leq u_0 \\ l_i \leq P(+z_i|+y_i) \leq u_i \\ i=1,\ldots,m}} \sum_{\substack{z_i = \pm z_i \\ i=0,\ldots,m}} P(-x|z_0, \ldots, z_m) P(z_0) \prod_{i=0}^{m} P(z_i|y_i) = (1 - u_0) \prod_{\substack{i=1,\ldots,m: \\ y_i = +y_i}} (1 - u_i), \tag{13}$$

and, with a similar derivation,

$$\overline{P}(-x|y_1, \ldots, y_m) = (1 - l_0) \prod_{\substack{i=1,\ldots,m: \\ y_i = +y_i}} (1 - l_i). \tag{14}$$

An example of the computation of these bounds and a comparison with the estimated provided by the precise models are in Table I. It is worth noting that the credal sets $K(X|y_1, \ldots, y_m)$ defined by the bounds in (13) and (14), for each $(y_1, \ldots, y_m)$, cannot be used to express the credal network strong extension over the original variables of the Bayesian network without the auxiliary variables. In fact, in (2) each conditional probability mass function is free to take its values from its credal set independently of the others. Yet, it is easy to see that if $P(-x|+y_1,\ldots,+y_m)$ takes its lower bound $\prod_{i=0}^{m}(1-u_i)$, any conditional probability $P(+z_i|+y_i)$ should take its upper bound u_i, and hence the value of any other conditional probability for X is forced to take a specific value as in (12) with $p_i = u_i$ for each $i = 0,\ldots,m$. In other words, this corresponds to say that, after the marginalization of the auxiliary variables, the credal network modeling the imprecise noisy-OR has *non-separately specified* credal sets [2]. In order to avoid the presence of the auxiliary variables in the expression of the strong extension of the credal network as in (2), the only option is therefore require the whole CPT $P(X_i|Y_1,\ldots,Y_n)$ to take values in a collection of 2^{m+1} possible specification obtained by considering all the possible choice of the parameters $p_i \in \{l_i,u_i\}$, with $i = 0,\ldots,m$. Thus, among the elements of the strong extension of this, so-called *extensive* [2], specification of the credal network, we have joint probability mass functions corresponding to convex combinations of joint mass functions associated to different specification of the CPT. Although, each CPT is a (precise) noisy-OR gate as in (12), it is a simple exercise to verify that a convex combination of two noisy-OR is not ensured to be a noisy-OR as well. This basically means that the imprecise noisy-OR is not only providing a *sensitivity analysis* of the precise noisy-OR, as the strong extension does not only includes noisy-OR gates. This point suggests the possibility of approximating generic CPTs by means of imprecise noisy-OR gates. This direction will be discussed in the conclusions.

AIR ESCAPE	+	+	−	−
CAP DEFLATE	+	−	+	−
OR	1.00	1.00	1.00	.00
Noisy-OR	.95	.80	.75	.00
Imprecise Noisy-OR	[.90, .99]	[.75, .95]	[.60, .80]	.00

TABLE I
THE CONDITIONAL PROBABILITIES FOR THE TIRE TO BE FLAT ACCORDING ON THE BASIS OF THE ASSESSMENTS IN EXAMPLE 1.

Here we have considered the modeling properties of the imprecise noisy-OR gate. An important feature characterizing the model when doing inference is detailed in the next section.

IV. MESSAGE PROPAGATION ON THE IMPRECISE NOISY-OR

After having proposed the noisy-OR gate for a compact quantification of CPTs modeling causal effects, Pearl has proved that the local computations required by his message propagation updating algorithm for polytrees take only linear time in the number of parents if the CPT is a noisy-OR gate, while exponential time is required by general CPTs. Here, we prove the analogous result for credal networks when inferences are computed with 2U on an imprecise noisy-OR gate.

First, as we noted in Section II-B, the two equations making the local complexity of 2U exponential in the indegree are (4) and (7). Let us therefore consider (4) when X_i is quantified by an imprecise noisy-OR. Thus, according to (13), for $X_i = -x_i$, the expression inside the minima in (4) rewrites as follows:[9]

$$\sum_{\pi_i} \prod_{\substack{k:X_k\in\Pi_i\\ x_k=+x_k}} (1-u_k) \prod_{j:X_j\in\Pi_i} \mu_i(x_j), \tag{15}$$

where this can be done despite the non-separate specification of the conditional credal set, as the lower bound of the conditional probabilities for $-x_i$ is simultaneously attained when all the inhibition probabilities are $1-u_k$, $k=1,\ldots,m$. We can easily see that (15) rewrites as:

$$\sum_{x_j:X_j\in\Pi_i} \left[\prod_{\substack{j:X_j\in\Pi_i\\ x_j=+x_j}} (1-u_j)\mu_i(+x_j) \right] \left[\prod_{\substack{j:X_j\in\Pi_i\\ x_j=-x_j}} \mu_i(-x_j) \right]. \tag{16}$$

From which, by starting to sum out a single parent X_k of X_i:

$$[(1-u_k)\mu_i(+x_k)+\mu_i(-x_k)] \cdot \sum_{\substack{x_j:\\ X_j\in\Pi_i,\\ j\neq k}} \left[\prod_{\substack{j:X_j\in\Pi_i,\\ j\neq k\\ x_j=+x_j}} (1-u_j)\mu_i(+x_j) \right] \left[\prod_{\substack{j:X_j\in\Pi_i,\\ j\neq k\\ x_j=-x_j}} \mu_i(-x_j) \right] \tag{17}$$

The same can be clearly done for all the other parents of X_i. Thus, we can consider again the minima as in (4) and exploit the fact that $\underline{\mu}_i(+x_l) = 1-\overline{\mu}_i(-x_l)$, and hence it is possible to write $\mu_i(-x_l) := 1-\mu_u(+x_l)$:

$$\underline{\mu}(x_i) = \min_{\substack{k:X_k\in\Pi_i\\ \mu_i(x_k)\in\\ \{\underline{\mu}_i(x_k),\overline{\mu}_i(x_k)\}}} \prod_{k:X_k\in\Pi_i} [1-u_k\mu_i(+x_k)], \tag{18}$$

where the minimization can be performed separately for each term of the product. Similar benefits occur when computing $\overline{\mu}(x_i)$ and the Λ messages by means of (7). The computational advantages of message propagation on imprecise noisy-OR gate can be therefore summarized by the following result.

Proposition 1: *2U-based inference on imprecise noisy-OR gates takes local complexity linear in the number of parents.*

As a simple corollary to this result, let us note that if $l_i = u_i = 0$, for each $i = 0,\ldots,m$, the imprecise noisy-OR collapses into a deterministic OR gate, while de Morgan's laws makes it possible to convert a noisy-OR into a noisy-AND by simple negations. Overall, we can easily prove an analogous of Proposition 1 for any kind of (deterministic, noisy, imprecise noisy) Boolean gate. Although simple to derive, this result is not trivial: even if a particular gate might have been precisely quantified, the message propagation should be generally computed by 2U instead of Pearl's algorithm as some other node could be quantified in an imprecise way.

[9] In this derivation, we also assume the model non-leaky, i.e., $l_0 = u_0 = 0$, which makes it possible to remove Z_0 from the model. This makes the notation more readable, without affecting the validity of our conclusions.

These results suggests the possibility of an imprecise-probabilistic approach to fault tree analysis, which is discussed in the next section.

V. Imprecise fault trees

Fault tree analysis has become very popular in dependability analysis and safety studies of large critical systems [13]. Although Bayesian networks have been proposed after, it is possible to regard fault trees as a special class of Bayesian networks [3], with all the variables Boolean, polytree-shaped graph, and CPTs deterministically quantified by Boolean operators. In most of the cases, the fault tree has a single leaf (i.e., without children) node (e.g., node *brake fails* in Fig. 3), which is called the *top event* and represents the failure of the overall system. The root (i.e., parentless) nodes of the network, which are also called *initiating events* (e.g., nodes *pads, sensor, controller, actuator fails* in Fig. 3), represent events whose potential failure can induce the global failure of the system. What makes the model probabilistic is the uncertain information about the initiating events, whose *failure rates*, generally expressed in number of failures per unit of time, are assessed. Given a failure rate λ_X for initiating event X, a simple dynamics of the failure probability could be:

$$P_t(+x) = 1 - e^{\lambda_X t} \simeq \lambda_X t, \tag{19}$$

where the true state of X denotes the failure of the initiating event. For each time t, message propagation algorithms for Bayesian networks can be adopted to compute the corresponding marginal probability of failure for the top event. By exploiting the results in the previous section, we can easily do the same kind of computations even if for the failure rate λ_X only a lower and an upper estimate, say $\underline{\lambda}_X$ and $\overline{\lambda}_X$, are available, and hence a credal set $K_{\underline{\lambda}_X t, \overline{\lambda}_X t}(X)$ models our uncertainty about the probability that the initiator X fails.

As an example of this kind of computations, Fig. 4 reports the dynamic behavior of the probability of failure for the top event of the fault tree in Fig. 3, with an imprecise quantification of the failure rates, compared with the results obtained with a precise value. The time shift between the moment when the upper probability (pessimistic estimate) of failure becomes higher than .5 and the same happens for the precise estimate is considerable (> 500hrs).

Finally, let us note that one of the major weaknesses of fault tree analysis is the fact that all the initiating events are considered as statistically independent, so that the methodology has a scarce modeling power counterbalanced by the ability to deal with large scale. As a possible attempt to overcome this limitation, we can try to use imprecise-probabilistic modeling to describe our lack of information about the independence between two or more events. Let us present this idea by means of the following simple example.

Example 2: *Consider the fault tree in Fig. 3. Assume the failure rates for the four initiating events have been precisely assessed, and they are all the same. Let $p = .1$ the corresponding probability of failure. Assume also that the independence between the actuator (A) and the controller (C) could be doubtful. A very conservative approach consists in determining the credal set $K(C, A)$ including all the joint mass probability functions:*

$$P(A, C) = [P(+a, +b), P(+a, -b), P(-a, +b), P(-a, -b)],$$

such that $P(+a) = P(+b) = .1$. Standard polyhedra algorithms return $\mathrm{ext}[K(A, C)] = \{[.0, .1, .1, .8], [.1, .0, .0, .9]\}$. *By applying the* binarization *procedure of credal sets described in [1], it is possible to describe $K(C, A)$ as the strong extension of a credal network with a single link from C to A and conditional credal sets $K_{.1,.1}(C)$, $K_{0,1}(A|+c)$, $K_{0,1/9}(A|-c)$.*[10] *Overall, this induces the specification of a credal network over the topology in Fig.3, with an additional link between nodes* controller *and* actuator. *This makes the network multiply-connected, and hence 2U unusable. Yet, a loopy version of 2U (called L2U, see [1]), which inherits the same computational advantages of 2U when Boolean gates are present, but in generally only returns an approximate inferences, can be considered.*

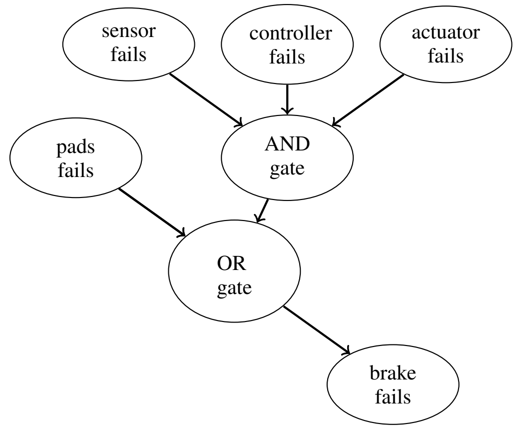

Fig. 3. A fault tree modeling the failure of a brake according to the Boolean formula brake = [pads $\vee$ (sensor $\wedge$ controller $\wedge$ actuator)].

VI. Conclusions and outlooks

An imprecise-probabilistic version of the noisy-OR gate has been proposed. The goal is to gain in realism and robustness for the modeling of causal relations of a noisy-OR gate when a sharp assessment of the inhibition probabilities seems unreliable. Notably, the higher expressive power provided by the imprecise modeling is not compromising the computational efficiency: exactly as in the case of Bayesian networks with the standard noisy-OR gate, message propagation on the imprecise gate only takes a linear complexity in the number of causes.

As a future work, we intend to explore the possibility of approximating general CPTs in polytree-shaped Bayesian networks over Boolean variables by means of imprecise noisy-OR gates. The corresponding credal network should be finally

[10] Note that, in general, this credal network has non-separately specified credal sets. So, assuming that the network has separately specified credal sets induces an outer approximation.

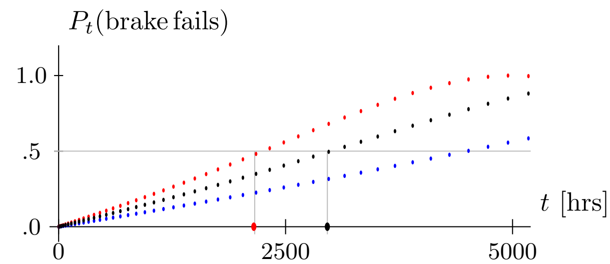

Fig. 4. Probability of failure for the top node of the fault tree in Fig.3. Red and blue points are the inferences obtained by assuming for the failure rates of all the four initiating events a lower bound corresponding to a failure every $10'000$ hours and $5'000$ hours for the upper bound. Inferences are computed by exploiting the techniques described in Sect. IV. Black points report the estimates based on the precise estimates based on the average between the lower and the upper failure rates.

updated by the 2U algorithm, this returning an interval containing the true value of the inference on the original Bayesian network.

An extension to multiple-valued, non-Boolean, variables should be also regarded as a future development. This would basically correspond to a generalization to imprecise probabilities of the noisy-MAX gate [6]. This task seems to be almost straightforward for what concerns the specification of the model, while an analogous of Proposition 1 for the imprecise noisy-MAX seems to be much more problematic, because of the lack of efficient time algorithm for non-Boolean polytree-shaped credal networks. Finally, we also intend to evaluate the practical advantages in terms of robustness of the proposed extension of fault trees to imprecise models.

References

[1] A. Antonucci, Y. Sun, C. P. de Campos, and M. Zaffalon. Generalized loopy 2U: a new algorithm for approximate inference in credal networks. *International Journal of Approximate Reasoning*, 51(5):474–484, 2010.

[2] A. Antonucci and M. Zaffalon. Decision-theoretic specification of credal networks: a unified language for uncertain modeling with sets of Bayesian networks. *International Journal of Approximate Reasoning*, 49(2):345–361, 2008.

[3] L. Bobbio, A.and Portinale, M. Minichino, and E. Ciancamerla. Comparing fault trees and bayesian networks for dependability analysis. In *Computer Safety, Reliability and Security, 18th International Conference, SAFECOMP'99*, pages 310–322, 1999.

[4] F. G. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[5] C. P. de Campos and F. G. Cozman. The inferential complexity of Bayesian and credal networks. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 1313–1318, Edinburgh, 2005.

[6] F.J. Díez. Parameter adjustment in Bayes networks. the generalized noisy or-gate. In *UAI '93: Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, pages 99–105, 1993.

[7] E. Fagiuoli and M. Zaffalon. 2U: an exact interval propagation algorithm for polytrees with binary variables. *Artificial Intelligence*, 106(1):77–107, 1998.

[8] M. Henrion. Some practical issues in constructing belief networks. In *Proceedings of the Annual Conference on Uncertainty in Artificial Intelligence (UAI-87)*. Elsevier Science, 1987.

[9] J.F. Lemmer and D.E. Gossink. Recursive noisy or - a rule for estimating complex probabilistic interactions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 34(6):2252–2261, 2004.

[10] S. Parsons and Bigham J. Possibility theory and the generalised noisy or model. Technical report, Queen Mary and Westfield College.

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

[12] S. Srinivas. A generalization of the noisy-or model. In *UAI '93: Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence*, pages 208–218, 1993.

[13] W. E. Vesely, F. F. Goldberg, N. H. Roberts, and D. F. Haasl. *Fault Tree Handbook*. U.S. Nuclear Regulatory Commission, 1981.

[14] P. Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, New York, 1991.

[15] P. Walley. Inferences from multinomial data: learning about a bag of marbles. *J. R. Statist. Soc. B*, 58(1):3–57, 1996.

[16] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.