Evaluating interval-valued influence diagrams ☆

Rafael Cabañas [a,*], Alessandro Antonucci [b], Andrés Cano [a],
Manuel Gómez-Olmedo [a]

[a] *Department of Computer Science and Artificial Intelligence CITIC, University of Granada, Spain*
[b] *Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA), Galleria 2, 6928 Manno (Lugano), Switzerland*

ARTICLE INFO

ABSTRACT

Influence diagrams are probabilistic graphical models used to represent and solve sequential decision problems under uncertainty. Sharp numerical values are required to quantify probabilities and utilities. This might be an issue with real models, whose parameters are typically obtained from expert judgments or partially reliable data. We consider an interval-valued quantification of the parameters to gain realism in the modeling and evaluate the sensitivity of the inferences with respect to perturbations in the sharp values of the parameters. An extension of the classical influence diagrams formalism to support such interval-valued potentials is presented. The variable elimination and arc reversal inference algorithms are generalized to cope with these models. At the price of an outer approximation, the extension keeps the same complexity as with sharp values. Numerical experiments show improved performances with respect to previous methods. As a natural application, we propose these models for practical sensitivity analysis in traditional influence diagrams. The maximum perturbation level on single or multiple parameters preserving the optimal strategy can be computed. This allows the identification of the parameters deserving a more careful elicitation.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Influence diagrams (IDs) are popular probabilistic graphical models intended to represent and solve decision problems with uncertainty. IDs represent a sophistication of Bayesian networks to cope with sequential decision tasks. The parameters of an ID are not only, as in Bayesian networks, probabilities of conditional states of single variables given other variables, but also utilities of joint states of sets of variables. As Bayesian networks, IDs demand a sharp estimation of their parameters. Both probabilities and utilities might be quantified by expert knowledge or statistical data processing. Yet, sharp values can be unable to express a qualitative expert judgment or a statistical analysis based on scarce or missing data. E.g., which is the number modeling the probability for an option more probable than its negation? And the negative utility of a disaster scenario that never occurred in the past?

For reasons of this sort, in the last two decades, various extensions of Bayesian networks intended to support generalized probabilistic statements have been proposed. These models have been developed in the field of possibility theory [4], evi-

---

dence theory [28], and imprecise probability [9]. The latter models, called *credal networks*, offer a direct sensitivity-analysis interpretation: a credal network is a collection of Bayesian networks, all over the same variables and with the same graph, whose parameters are consistent with constraints (e.g., interval specifications) modeling a limited ability in the assessment of sharp estimates. Something similar has been also done with decision trees [16,17,19], while the situation is different for IDs. The early attempts of Fertig and Breese [13] first, and Zaffalon [11] after, to extend these models to non-sharp quantification are among the few works in this direction.[1,2] This sounds unfortunate as the above considerations about the difficulty of assessing sharp estimates for probabilities are even more compelling for utilities, which are supposed to model intrinsically qualitative objects such as preferences.[3]

We extend to the interval-valued case the formalism of IDs by keeping the same sensitivity-analysis interpretation of credal networks. Such a generalized ID is therefore equivalent to a collection of classical (i.e., "precise") IDs whose parameters are consistent with the interval constraints. In this framework also the expected utility of a policy becomes interval-valued. A decision criterion to detect the optimal decision when comparing intervals is therefore needed. We adopt a conservative approach, called *interval dominance* in the imprecise-probability jargon [27], which rejects all the decisions leading to certainly sub-optimal strategies.

The standard approaches to IDs evaluation, namely *variable elimination* [18,31] and *arc reversal* [25], are generalized in order to cope with the interval-valued case. The extension to intervals does not increase the computational complexity which remains the same as with sharp parameters for both the algorithms. This is achieved at the price of an outer approximation in the inferences, which is required to preserve the interval-valued modeling. An experimental comparison against the arc reversal technique proposed by Fertig and Breese in [13,12,6] (i.e., the only practical approach proposed so far these models) shows a clear improvement in terms of both evaluation time and accuracy.

The proposed algorithms can be also used for practical sensitivity analysis in (standard) IDs. By replacing the sharp values of some parameters with intervals, we can decide whether or not the original optimal strategy is robust with respect to a perturbation consistent with the intervals. The maximal level of perturbation leaving the strategy unchanged can be therefore regarded as a robustness descriptor. This allows to identify the more critical parameters of the model and, for instance, deciding which ones deserve a more careful elicitation.

The results originally presented in a conference version of this work [8] are presented here with more accurate descriptions and examples. This extended version also includes novel material such as the arc reversal algorithm, an improved version of the variable elimination algorithm, an empirical validation against the algorithm in [13], and the application to sensitivity analysis of standard IDs.

The paper is organized as follows. Section 2 introduces the notation and some basic concepts. Section 3 defines interval-valued potentials and the corresponding algebraic structure. The algorithms to evaluate interval-valued IDs are in Section 4, while the procedure for sensitivity analysis is in Section 5. The empirical analysis is presented in Section 6. Finally, the conclusions of the paper are given in Section 7.

2. Basics

Let us first define the basic notation. We use upper-case letters for variables and lower-case for their possible values. Given a variable X, x is an element of the domain of X, which we denote as Ω_X. Given a set of n variables $\boldsymbol{X} := \{X_1, \ldots, X_n\}$, and a multi-valued index $J \subseteq [1, n]$, X_J is the joint variable including any X_i such that $i \in J$. Thus, $\Omega_{X_J} = \times_{i \in J} \Omega_{X_i}$, where $\times$ is the Cartesian product. Given a second index I, the notation $x_I \sim x_J$ is used to express *consistency*, i.e., to denote the fact that the two states have the same values on $X_{I \cap J}$. *Chance* variables are those whose actual value might be unknown, *decision* variables are those whose actual value can be set by a decision maker.

2.1. Influence diagrams

Influence diagrams (IDs) [15] are a class of graphical models designed to formalize sequential decision problems with uncertainty. Compared with decision trees [23], IDs offer a compact encoding of the independence relations between variables, which prevents an exponential growth in the problem representation.

An ID over a set of chance variables $\boldsymbol{X}$ and a set of decisions $\boldsymbol{D}$ is made of a qualitative and a quantitative part. The qualitative part is an acyclic directed graph $\mathcal{G}$ with three types of nodes. *Chance* nodes are depicted as circles and are in one-to-one correspondence with the chance variables, i.e., those in $\boldsymbol{X}$. *Decision* nodes are depicted as squares and associated to decision variables, i.e., those in $\boldsymbol{D}$. Utility nodes are depicted as diamonds and should be barren, i.e., they have no children. The terms node and variable are used interchangeably for both chance and decision variables. Utility nodes are not

---

[1] The work of Zhou et al. [32] combining sharp probabilities with interval-valued utilities is just a trivial special case of the general framework we present here.

[2] Sensitivity analysis does not require the specification of more general classes of models, being only focused on the results of the inferences. Thus, it should be regarded as a different topic, which, as a matter of fact, received more attention (e.g., [22]).

[3] The possibilistic influence diagrams proposed in [14] can be regarded as a successful attempt in the same direction we explore. Yet, the possibilistic framework does not allow for interval-valued specifications, thus lacking a direct sensitivity-analysis interpretation.

associated to variables. Yet, these nodes are jointly denoted as $\boldsymbol{U}$. The parents of an arbitrary node Y according to $\mathcal{G}$ are also called *direct predecessors* and denoted Π_Y. Similarly, the children of Y are called *direct successors* and denoted Γ_X.

The quantitative part is made of a set of *probability potentials* (PPs) that represents the uncertainty, and a set of *utility potentials* (UPs) that represents the user preferences. A PP over two disjoint sets of variables X_I and X_J, denoted as $\phi(X_I|X_J)$, is a map $\phi : \Omega_{X_{I \cup J}} \to [0, 1]$ such that $\sum_{x_I \in \Omega_{X_I}} \phi(x_I|x_J) = 1$ for each $x_J \in \Omega_{X_J}$. Similarly, a UP over X_K, denoted as $\psi(X_K)$, is a map $\psi : \Omega_{X_K} \to \mathbb{R}$. Notice that UPs are not necessarily normalized. For each chance node, a PP over the corresponding variable and its direct predecessors is defined, while, for each utility node, an UP over the parents should be assessed. The complete definition of ID is as follows.

Definition 1 (*Influence diagram*). An influence diagram is a tuple $\langle \mathcal{G}, \boldsymbol{X}, \boldsymbol{D}, \boldsymbol{U}, \Phi, \Psi \rangle$, where $\mathcal{G}$ is an acyclic directed graph over $\boldsymbol{X} \cup \boldsymbol{D} \cup \boldsymbol{U}$, while $\Phi = \{\phi(X|\Pi_X)\}_{X \in \boldsymbol{X}}$ and $\Psi = \{\psi(\Pi_U)\}_{U \in \boldsymbol{U}}$ are collections of, respectively, PPs and UPs.

To model sequential decision problems with IDs two additional assumptions are required. An ID is *regular* if the graph $\mathcal{G}$ contains at least a directed path connecting all the decision nodes. Because of the acyclicity of $\mathcal{G}$, this defines a complete topological ordering over the decision variables. Without lack of generality, the indexes of the decision nodes, say $\boldsymbol{D} := \{D_1, \ldots, D_n\}$, can be assumed to reflect such order, i.e., $D_1 \prec \ldots \prec D_n$, with the symbol $\prec$ denoting topological precedence. We partition the chance variables $\boldsymbol{X}$ in $n + 1$ sets denoted as $\{\mathcal{I}_i\}_{i=0}^n$. For each $i = 0, \ldots, n - 1$, $\mathcal{I}_i$ includes the chance nodes directly preceding D_{i+1} but not D_i. If a chance variable is a direct predecessor of more than a decision node, it belongs to the set associated to the decision variable with the smallest index. The remaining chance variables, i.e., those not having decision nodes among their direct successors, belong to $\mathcal{I}_n$. This forms a partition of $\boldsymbol{X}$, i.e., $\cup_{i=0}^n \mathcal{I}_i = \boldsymbol{X}$ and $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$ for each $i, j = 0, 1, \ldots, n$, $i \neq j$. Overall, regular IDs are characterized by the following partial order of the variables: $\mathcal{I}_0 \prec D_1 \prec \mathcal{I}_1 \prec \cdots \prec D_n \prec \mathcal{I}_n$. It is easy to see that, apart from the variables in $\mathcal{I}_n$, any topological order extracted from $\mathcal{G}$ is consistent with the above partial order. Such order reflects a temporal interpretation: the chance variables in $\mathcal{I}_i$ are observed by the decision maker before decision D_{i+1} is taken, and the ordering over $\boldsymbol{D}$ reflects the order in which the different decisions are taken. According to the *non-forgetting* assumption, previous decisions and observations are known at each decision. Arcs included to satisfy this assumption are called *non-forgetting arcs*. Here we only consider regular IDs and assume the non-forgetting arcs to be present. A classical ID is considered here below.

Example 1 (*The oil wildcatter [23,26]*). An oil wildcatter must decide whether or not to drill. He/she is uncertain whether the amount of oil (O) in the place is empty (e), wet (w) or soaking (s). The wildcatter can make seismic tests (S) that will give a closed reflection pattern (c) indicating much oil, an open pattern (o) indicating for some oil, or a diffuse pattern (d) denoting almost no hope for oil. These two are chance variables, while the decision variables are T [to test (t) or not (nt)] and D [to drill (d) or not (nd)]. The utility nodes P and C describe respectively the profit possibly obtained from the presence of oil and the cost of the tests. Fig. 1 depicts the graph of an ID modeling this problem. The graph is regular as decision T precedes decision D. In this case the partial order is complete being $T \prec S \prec D \prec O$. The numerical values of the potentials, reported below in a matrix form with the corresponding states depicted in gray, are

$$\phi(O) = \begin{bmatrix} .5 \\ .3 \\ .2 \end{bmatrix} \begin{matrix} e \\ w \\ s \end{matrix} \quad , \quad \psi(T) = \begin{bmatrix} -10 \\ 0 \end{bmatrix} \begin{matrix} t \\ nt \end{matrix} \quad , \quad \psi(O, D) = \begin{matrix} d & nd \\ \begin{bmatrix} -70 & 0 \\ 50 & 0 \\ 200 & 0 \end{bmatrix} & \begin{matrix} e \\ w \\ s \end{matrix} \end{matrix} \quad ,$$

$$\phi(S|OT) = \begin{matrix} & \begin{matrix} & t & & & nt & \\ e & w & s & e & w & s \end{matrix} \\ \begin{bmatrix} .1 & .3 & .5 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ .3 & .4 & .4 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ .6 & .3 & .1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} & \begin{matrix} c \\ o \\ d \end{matrix} \end{matrix} \quad .$$

2.2. IDs evaluation (definitions)

A *policy* for a decision variable D_i is a mapping $\delta_{D_i} : \Omega_{\Pi_{D_i}} \to \Omega_{D_i}$ associating a state of D_i (i.e., a decision) to its past observations and decisions. A *strategy* Δ is a collection of policies, one for each decision variable, i.e., $\Delta := \{\delta_{D_1}, \delta_{D_2} \ldots, \delta_{D_n}\}$. Evaluating IDs consists in the identification of an optimal strategy Δ^*, which maximizes the expected value of the sum of the UPs. The (optimal) policies of an optimal strategy and the maximum expected utility are defined as follows.

Definition 2 (*Optimal policy and maximum expected utility [18]*). Consider an ID as in Definition 1 which is also regular and satisfies the non-forgetting assumption. Let the temporal order be described as $\mathcal{I}_0 \prec D_1 \prec \mathcal{I}_1 \prec \cdots \prec D_n \prec \mathcal{I}_n$. For each $i = 1, \ldots, n$, the optimal policy for decision D_i is
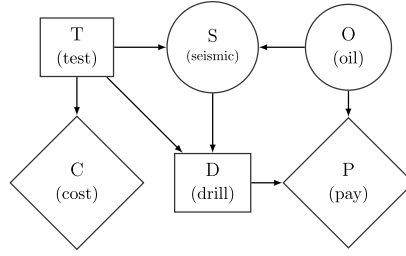
Fig. 1. Graph of an ID modeling the oil wildcatter's decision.

$$\delta^*_{D_i}(\mathcal{I}_0, D_1, \ldots, \mathcal{I}_{i-1}) := \arg\max_{D_i} \sum_{\mathcal{I}_i} \max_{D_{i+1}} \cdots \max_{D_n} \sum_{\mathcal{I}_n} \prod_{X \in \boldsymbol{X}} \phi(X|\Pi_X) \sum_{U \in \boldsymbol{U}} \psi(\Pi_U), \tag{1}$$

and the maximum expected utility is

$$\mathrm{MEU} := \sum_{\mathcal{I}_0} \max_{D_1} \cdots \max_{D_n} \sum_{\mathcal{I}_n} \prod_{X \in \boldsymbol{X}} \phi(X|\Pi_X) \sum_{U \in \boldsymbol{U}} \psi(\Pi_U), \tag{2}$$

with $(X, \Pi_X, \Pi_U) \sim (\{D_i\}_{i=1}^n, \{\mathcal{I}_j\}_{j=0}^n)$ for each $X \in \boldsymbol{X}$ and $U \in \boldsymbol{U}$.

Eq. (1) returns the value of D_i maximizing the (unnormalized) expected value of the sum of the UPs. At the moment of that decision, all the previous decisions have been already taken and all the chance variables in the past observed. The maximization is indeed achieved with respect to D_i and the subsequent decisions, while the expectation is computed with respect to the uncertainty about the chance variables in the future of D_i. The MEU in Eq. (2) can be regarded as the expected value of the sum of the utilities when the decision maker takes his/her decisions on the basis of the optimal policies in Eq. (1). These basic concepts are demonstrated in the following example.

Example 2 *(Wildcatter's policy).* In the oil wildcatter's ID (Example 1), the optimal policies as in Eq. (1) associated to decisions T and D are

$$\delta^*_D(S, T) = \begin{matrix} & t & nt & \\ \begin{bmatrix} d & d \\ d & d \\ nd & d \end{bmatrix} & \begin{matrix} c \\ o \\ d \end{matrix} \end{matrix}, \quad \delta^*_T = [\, t \,],$$

while the maximum expected utility as in Eq. (2) is MEU $= 22.5$. Note that the optimal policy for T has no arguments, since there are no variables in its past (T is the first decision and $\mathcal{I}_0 = \emptyset$). Accordingly, doing the tests is always the best decision for T, while, regarding D, it is always better to drill apart from the case in which the test was done and a diffusive pattern found.

2.3. IDs evaluation (algorithms)

Here we review two of the main algorithms for computing the optimal strategy Δ^* and the maximum expected utility MEU in IDs, namely variable elimination (Section 2.3.1) and arc reversal (Section 2.3.2).

2.3.1. Variable elimination

Variable elimination (VE) is a typical approach to inference in graphical models. VE algorithms for IDs [18,31] are commonly used to solve Eq. (2). Unlike VE for Bayesian networks, in regular IDs the elimination order is not arbitrary: it should be the inverse of an order consistent with the partial order associated to the ID [20]. When the last variable is eliminated, the algorithm returns a potential with no arguments (i.e., a single value) corresponding to the MEU as in Eq. (2). Every time a decision variable is eliminated, the corresponding optimal policy is also obtained. Algorithm 1 depicts the general scheme of VE for IDs.

While chance variables are removed by sum, as in Bayesian networks, decision variables are instead eliminated by maximization. Algorithm 2 shows how to remove a single variable, no matter whether chance or decision, from an ID. Let us clarify some of the notation used to describe this procedure. The operator dom returns the variables in the argument of a potential. Sums in line 4 and maxima in line 6 are two different forms of *marginalization* (respectively for decision and chance variables), i.e., removing the variable from the argument of the potential. The division (line 4) is intended element-wise. In the PP in line 6, Y can be eliminated by instantiating an arbitrary value $y \in \Omega_Y$: when removing a decision, usually there are not PPs containing and if any, the decision is not affecting the values of such PP (any decision is d-separated from

---

Algorithm 1 VarElim – Variable elimination scheme.

---

input: regular ID with temporal order $\{\mathcal{I}_0, D_1, \mathcal{I}_1, \ldots, D_n, \mathcal{I}_n\}$

1: **for** $k \leftarrow n$ **to** 0 **do**
2:     **while** $\mathcal{I}_k \neq \emptyset$ **do**
3:         Select $X \in \mathcal{I}_k$ ▷ Pick a chance variable to eliminate
4:         $(\Phi, \Psi) \leftarrow \text{ElimVar}(X, \Phi, \Psi)$ ▷ Chance variable elimination (Algorithm 2)
5:         $\mathcal{I}_k \leftarrow \mathcal{I}_k \backslash \{X\}$
6:     **end while**
7:     **if** $k > 0$ **then**
8:         $(\Phi, \Psi) \leftarrow \text{ElimVar}(D_k, \Phi, \Psi)$ ▷ Decision variable elimination (Algorithm 2)
9:     **end if**
10: **end for**

Algorithm 2 ElimVar – Elimination of a single variable.

---

input: Y (variable to remove), Φ, Ψ (sets of current potentials)

1: $(\Phi_Y, \Psi_Y) \leftarrow (\{\phi \in \Phi | Y \in \text{dom}(\phi)\}, \{\psi \in \Psi | Y \in \text{dom}(\psi)\})$ ▷ Select
2: $(\phi_Y, \psi_Y) \leftarrow (\otimes_{\phi \in \Phi_Y} \phi, \otimes_{\psi \in \Psi_Y} \psi)$ ▷ Combine
3: **if** $Y \in \boldsymbol{X}$ **then**
4:     $(\phi'_Y, \psi'_Y) \leftarrow (\sum_Y \phi_Y, \frac{\sum_Y \phi_Y \otimes \psi_Y}{\sum_Y \phi_Y})$ ▷ Remove by sum (chance vars)
5: **else**
6:     $(\phi'_Y, \psi'_Y) \leftarrow (\phi_{Y=y}, \max_Y \psi_Y)$ ▷ Remove by max (decision vars)
7:     $\delta_Y^* \leftarrow \arg\max_Y \psi_Y$ ▷ Optimal policy (as a byproduct)
8: **end if**
9: $(\Phi, \Psi) \leftarrow (\Phi \backslash \Phi_Y \cup \{\phi'_Y\}, \Psi \backslash \Psi_Y \cup \{\psi'_Y\})$ ▷ Update
10: **return** (Φ, Ψ)

---

its predecessors [18] and any successor has already been removed). When eliminating a decision variable, the maximization of the UP also gives the corresponding optimal policy (line 7).

Finally, the implementation of the $\otimes$ operator to combine potentials is detailed in the following statement. It is easy to check that these definitions are well-posed and the operator is associative and commutative.

Definition 3 *(Combining potentials)*. The combination $\psi \otimes \psi'$ of two UPs, say $\psi(X_I)$ and $\psi'(X_J)$, is a UP over $X_{I \cup J}$ obtained by element-wise sums, i.e.,

$$(\psi \otimes \psi')(x_{I \cup J}) := \psi(x_I) + \psi'(x_J), \tag{3}$$

for each $x_{I \cup J} \in \Omega_{X_{I \cup J}}$, with $x_I, x_J \sim x_{I \cup J}$. The combination $\phi \otimes \psi$ of a PP $\phi(X_I | X_J)$ with a UP $\psi(X_K)$ is a UP over $X_L := X_{I \cup J \cup K}$ defined by element-wise products, i.e.,

$$(\phi \otimes \psi)(x_{I \cup J \cup K}) := \phi(x_I | x_J) \cdot \psi(x_K), \tag{4}$$

for each $x_{I \cup J \cup K} \in \Omega_{X_{I \cup J \cup K}}$, with $x_I, x_J, x_K \sim x_{I \cup J \cup K}$. Finally, the combination $\phi \otimes \phi'$ of two PPs, say $\phi(X_I | X_J)$ and $\phi'(X_K | X_L)$, is a PP over $X_{I \cup K}$ given $X_{(J \cup L) \backslash (I \cup K)}$ defined by element-wise products, i.e.,

$$(\phi \otimes \phi')(x_{I \cup K} | x_{(J \cup L) \backslash (I \cup K)}) := \phi(x_I | x_J) \cdot \phi'(x_K | x_L), \tag{5}$$

for each $x_{I \cup K} \in \Omega_{X_{I \cup K}}$ and $x_{(J \cup L) \backslash (I \cup K)} \in \Omega_{X_{(J \cup L) \backslash (I \cup K)}}$, with $x_I, x_J, x_K, x_L \sim x_{I \cup K}, x_{(J \cup L) \backslash (I \cup K)}$.

VE complexity is linear in the size of the largest potential generated during the evaluation [21]. This corresponds to the combinations in line 4 of Algorithm 2. Suppose we have an ID with n variables (chance or decision) with no more of d states, then the complexity of VE for evaluating it is bounded by $\mathcal{O}(n \cdot d^w)$ where w is the arity of the largest potential ever created during the evaluation. Note that, if the optimal elimination order is followed, w corresponds to the *treewidth* of the graph [5,24].

2.3.2. Arc reversal

Arc reversal (AR) [25] is an evaluation algorithm for IDs alternative to VE. In AR the orientation of an arc among two chance nodes can be reversed by Bayes rule. AR is based on a simple observation: the elimination of a variable, no matter whether chance or decision, having a utility node as unique direct successor involves only two potentials, thus does not affect the overall inferential complexity. Such patterns can be always created by properly changing some arc orientations. Algorithm 3 details the AR scheme. The basic transformations required by AR are described here below. The overall procedure is indeed demonstrated in Example 3. It is a trivial exercise to check that these transformations map the original ID into an *equivalent* one (two IDs are equivalent if they have the same expected utility and the same optimal policies for the remaining decisions). AR copes with IDs with a single utility node. If this is not the case, it is sufficient to apply the following transformation.
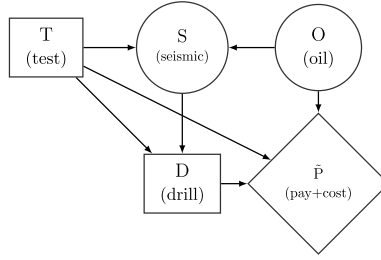
Fig. 2. Merging the utility nodes in the oil wildcatter's decision problem.

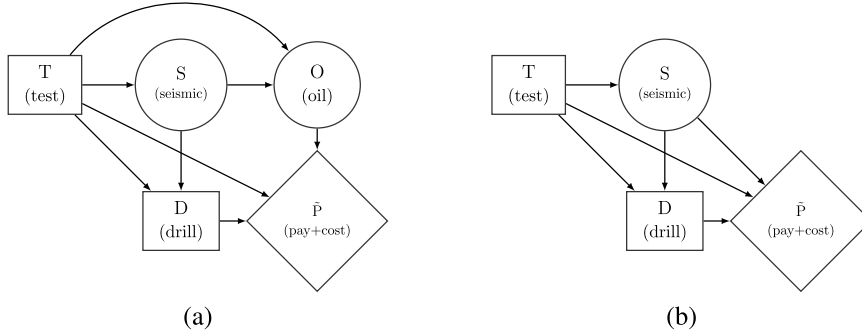(a)   (b)

Fig. 3. Reversal the $O \rightarrow S$ arc (left) and removal of node O (right).

Transformation 1 *(Merging utilities). Given an ID with at least two utility nodes, add a new utility node $\tilde{U}$, which is a barren child of all the parents of the utility nodes, i.e., $\Pi_{\tilde{U}} := \bigcup_{U \in \boldsymbol{U}} \Pi_U$. Define an UP associated to $\tilde{U}$ as $\psi(\pi_{\tilde{U}}) := \sum_{U \in \boldsymbol{U}} \psi(\pi_U)$. Finally, remove all the utility nodes different from $\tilde{U}$ and the corresponding UPs.*

Transformation 2 *(Chance nodes removal). Assume that a chance node Y has the utility node U as unique direct successor. Let X_I denote the direct predecessors of Y, and X_J those of U others than Y. To eliminate Y replace the PP $\phi(Y|X_I)$ and the UP $\psi(Y, X_J)$, with the UP $\psi(X_{I \cup J}) := \sum_{y \in \Omega_Y} \psi(y, X_J) \cdot \phi(y|X_I)$. Finally, remove Y from $\mathcal{G}$ and add new arcs connecting the nodes in X_I with U.*

Transformation 3 *(Decision nodes removal). Assume that a decision node Y has the utility node U as unique direct successor. Let X_I denote the direct predecessors of Y. Assume that the direct predecessors of U others than Y, and denoted as X_J, are also direct predecessors of Y. To eliminate Y, replace the UP $\psi(Y, X_J)$ with the UP $\psi(X_J) := \max_{y \in \Omega_Y} \psi(y, X_J)$. Finally, remove Y from $\mathcal{G}$.*

Transformation 4 *(Arc reversal). Assume that the chance nodes Y and X are directly connected by an arc, but not by other directed paths. Let $\phi(Y|X_I)$ and $\phi(X|Y, X_J)$ be the relative PPs, which means that X_I are the direct predecessors of Y and X_J those of X others than Y. Change the orientation of the arc and add arcs from X_I towards X and from X_J towards Y. The new PP for X is $\phi(X|X_I, X_J) := \sum_y \phi(y|X_I) \cdot \phi(X|y, X_J)$. The PP for Y is such that $\phi(y|x, X_I, X_J) \propto \phi(y|X_I) \cdot \phi(x|Y, X_J)$ with the proportionality constants obtained by normalization.*

Example 3 *(Reversing the oil wildcatter's arcs).* Consider the ID in Example 1 with the graph in Fig. 1. We first apply Transformation 1 to merge the two utility nodes C and P. The resulting equivalent ID with a single utility node is in Fig. 2. Then we reverse the arc from O to S by Transformation 4. As shown in Fig. 3.a, this makes the utility node $\tilde{P}$ the unique direct successor of O. The PPs $\phi(O)$ and $\phi(S|T, O)$ have been replaced by the new PPs $\phi'(O|S, T)$ and $\phi'(S|T)$. E.g.,

$$\phi'(O = e|S = c, T = t) := \frac{\phi(O = e)\phi(S = c|O = e, T = t)}{\sum_{O \in e, w, s} \phi(O = o)\phi(S = c|O = o, T = t)} \ . \tag{6}$$

Transformation 2 can be therefore used to eliminate the chance node O. This corresponds to replace the potentials $\phi(O|S, T)$ and $\psi(O, D)$ with the UP $\psi(S, T, D) := \sum_o \phi(o|S, T)\psi(o, D)$. In the resulting model (see Fig. 3.b), Transformation 3 can be used to remove D since the other direct predecessors of $\tilde{P}$ are also predecessors of D. By similarly continuing we eliminate all the decision and chance nodes and end up with the utility node only, whose constant UP corresponds to the MEU.

The above procedure can be easily extended to the general case. Algorithm 3 outlines the whole scheme. Compared to VE, the complexity of AR is not reduced because of the additional arcs added when reversing the arcs. Unlike VE, each step

---

Algorithm 3 ArcRev – Arc reversal scheme.

1: merge utility nodes in U ▷ Transformation 1
2: **while** $\Pi_U \neq \emptyset$ **do**
3:    **if** $\exists X \in \mathbf{X} \cap \Pi_U : \Gamma_X = \{ U \}$ **then**
4:      remove X ▷ Transformation 2
5:    **else if** $\exists D \in \mathbf{D} \cap \Pi_U : \Pi_U \subset \Pi_D \cup \{D\}$ **then**
6:      remove D ▷ Transformation 3
7:      remove barren nodes
8:    **else**
9:      find $X \in \mathbf{X} \cap \Pi_U : \mathbf{D} \cap \Gamma_X = \emptyset$
10:      **while** $\mathbf{X} \cap \Gamma_X \neq \emptyset$ **do**
11:        find $Y \in \mathbf{X} \cap \Pi_X : \nexists$ other directed path from X to Y
12:        replace arc $X \rightarrow Y$ with $X \leftarrow Y$ ▷ Transformation 4
13:      **end while**
14:      remove X
15:    **end if**
16: **end while**

---

of AR can be regarded as a transformation of an ID in an equivalent one with fewer variables. It has been shown empirically that the complexity of VE is never higher than the complexity of AR [3,7].

3. Interval-valued potentials

In real applications of IDs, the values of the potentials should be elicited from expert knowledge or learned from data. More realism in this quantification of the parameters can be obtained by replacing the sharp estimates with intervals. Such generalization is formalized below.

3.1. Interval-valued utility and probability potentials

The notions of UP and PP in Section 2.1 can be extended to intervals as follows.

Definition 4 (*Interval utilities*). An *interval-valued utility potential* (IUP) over X_I is a pair of UPs over X_I. We use the compact notation $\overline{\underline{\psi}}(X_I)$ for an IUP over X_I, $\underline{\psi}$ and $\overline{\psi}$ are the two UPs involved in the specification and are called, respectively, the lower and upper *bounds* of the IUP. The *extension* $\overline{\underline{\psi}}^*(X_I)$ of this IUP is the set of UPs consistent with the bounds, i.e.,

$$\overline{\underline{\psi}}^*(X_I) := \left\{ \psi : \Omega_{X_I} \rightarrow \mathbb{R} \, \middle| \, \underline{\psi}(x_I) \leq \psi(x_I) \leq \overline{\psi}(x_I), \forall x_I \in \Omega_{X_I} \right\}. \tag{7}$$

The extension of an IUP $\overline{\underline{\psi}}$ is non-empty if and only if $\underline{\psi}(x_I) \leq \overline{\psi}(x_I) \; \forall x_I \in \Omega_{X_I}$.

Definition 5 (*Interval probabilities*). An *interval-valued probability potential* (IPP) over X_I given X_J is a pair of (in general not normalized) PPs over X_I given X_J. We denote such an IPP as $\overline{\underline{\phi}}(X_I|X_J)$, where $\underline{\phi}(X_I|X_J)$ and $\overline{\phi}(X_I|X_J)$ are the two (unnormalized) bounds. The *extension* $\overline{\underline{\phi}}^*(X_I|X_J)$ of this IPP is the set of PPs consistent with the bounds, i.e.,

$$\overline{\underline{\phi}}^*(X_I|X_J) := \left\{ \phi : \Omega_{X_I} \times \Omega_{X_J} \rightarrow \mathbb{R}_0^+ \, \middle| \, \begin{array}{l} \sum_{x_I} \phi(x_I|x_J) = 1 \\ \underline{\phi}(x_I|x_J) \leq \phi(x_I|x_J) \leq \overline{\phi}(x_I|x_J) \\ \forall(x_I, x_J) \in \Omega_{X_I} \times \Omega_{X_J} \end{array} \right\}. \tag{8}$$

Condition $\underline{\phi}(x_I|x_J) \leq \overline{\phi}(x_I|x_J)$ for each x_I, x_J, together with $\sum_{x_I} \underline{\phi}(x_I|x_J) \leq 1 \leq \sum_{x_I} \overline{\phi}(x_I|x_J)$, for each $x_J \in \Omega_{X_J}$ is necessary and sufficient for the extension of the IPP to be non-empty. The additional condition $\overline{\phi}(x_i') + \sum_{x_i \neq x_i'} \underline{\phi}(x_i) \leq 1$ and the analogous expression for the lower instead of the upper bounds is called *reachability* [10]. The meaning is that for each $p \in [\underline{\phi}(x_I|x_J), \overline{\phi}(x_I|x_J)]$, there is at least a PP $\phi \in \overline{\underline{\phi}}^*$ such that $\phi(x_I|x_J) = p$. Note also that an IPP with non-empty extension can be always reduced to a reachable one by shrinking its bounds and this has no effect on its extension. Given an IPP, we always check whether or not it is reachable and, if not, we apply the shrinking. This operations will be iterated after any modification of the IPPs. We call *one-sided* an IPP whose upper bound is constantly set equal to one. The reachability constraint makes the upper bound of one-sided IPPs equal to $\overline{\phi}(x_I|x_J) = 1 - \sum_{x_I' \neq x_I} \underline{\phi}(x_I'|x_J)$. Being defined by linear constraints, both the extensions of an IUP and an IPP are convex sets of, respectively, UPs and PPs. Convex sets of UPs and PPs which are not extensions of IUPs and IPPs can be also considered, but this topic would be beyond the scope of this paper.

Example 4. Consider the following interval-valued potentials:

$$\underline{\overline{\phi}}(O) = \begin{bmatrix} [.475, .525] \\ [.285, .335] \\ [.190, .240] \end{bmatrix}, \ \underline{\overline{\psi}}(T) = \begin{bmatrix} [-10, -5] \\ [-5, 5] \end{bmatrix}, \ \underline{\overline{\psi}}(O, D) = \begin{bmatrix} [-75, -65] & [-5, 5] \\ [45, 55] & [-5, 5] \\ [195, 205] & [-5, 5] \end{bmatrix},$$

$$\underline{\overline{\phi}}(S|O, T) = \begin{bmatrix} [.095, .145] & [.285, .335] & [.475, .525] & [.317, .367] & [.317, .367] & [.317, .367] \\ [.288, .335] & [.380, .430] & [.380, .430] & [.317, .367] & [.317, .367] & [.317, .367] \\ [.570, .620] & [.285, .335] & [.095, .145] & [.317, .367] & [.317, .367] & [.317, .367] \end{bmatrix},$$

where the same variables and matrix notation as in Example 1 is used. It is a trivial exercise to check that these potentials have non-empty extensions, the UPs and PPs in Example 1 are included in these extensions, and the IPPs are reachable.

3.2. Operations with interval-valued potentials

The combination operator $\otimes$ considered in Definition 3 for standard potentials can be extended to intervals as follows.

Definition 6 (*Combining interval-valued potentials*). The combination $\overline{\underline{\psi \otimes \psi'}}$ of two IUPs, say $\underline{\overline{\psi}}(X_I)$ and $\underline{\overline{\psi}}'(X_J)$, is an IUP over $X_{I \cup J}$ such that

$$\underline{\psi \otimes \psi'}(x_{I \cup J}) := \underline{\psi}(x_I) + \underline{\psi}'(x_J), \tag{9}$$

for each $x_{I \cup J} \in \Omega_{X_{I \cup J}}$, with $x_I, x_J \sim x_{I \cup J}$; and similarly for the upper bounds. The combination $\overline{\underline{\phi \otimes \psi}}$ of an IPP $\underline{\overline{\phi}}(X_I|X_J)$ with an IUP $\underline{\overline{\psi}}(X_K)$ is an IUP over $X_{I \cup J \cup K}$ such that

$$\underline{\phi \otimes \psi}(x_{I \cup J \cup K}) := \underline{\phi}(x_I|x_J) \cdot \underline{\psi}(x_K), \tag{10}$$

for each $x_{I \cup J \cup K} \in \Omega_{X_{I \cup J \cup K}}$, with $x_I, x_J, x_K \sim x_{I \cup J \cup K}$; if $\underline{\psi}(x_K) < 0$ the lower bound of the combination is obtained by multiplying the lower bound of the IUP for the upper bound of the IPP (and vice versa for the upper bound). Finally, the combination $\overline{\underline{\phi \otimes \phi'}}$ of two IPPs, say $\underline{\overline{\phi}}(X_I|X_J)$ and $\underline{\overline{\phi}}'(X_K|X_L)$ is an IPP over $X_{I \cup K}$ given $X_{(J \cup L) \setminus (I \cup K)}$ such that

$$\underline{\phi \otimes \phi'}(x_{I \cup K}|x_{(J \cup L) \setminus (I \cup K)}) := \underline{\phi}(x_I|x_J) \cdot \underline{\phi}'(x_K|x_L), \tag{11}$$

for each $x_{I \cup K} \in \Omega_{X_{I \cup K}}$ and $x_{(J \cup L) \setminus (I \cup K)} \in \Omega_{X_{(J \cup L) \setminus (I \cup K)}}$, with $x_I, x_J, x_K, x_L \sim x_{I \cup K}, x_{(J \cup L) \setminus (I \cup K)}$.

Example 5 (*Interval-valued potential combination*). Consider the IPPs and IUPs in Example 4 associated to the oil wildcatter's ID. It is a straightforward exercise to check that the following combined potentials

$$\underline{\overline{\psi}}(T, O, D) := \underline{\overline{\psi}}(T) \otimes \underline{\overline{\psi}}(O, D),$$

$$\underline{\overline{\phi}}(S, O|T) := \underline{\overline{\phi}}(O) \otimes \underline{\overline{\phi}}(S|O, T),$$

$$\underline{\overline{\psi}}(S, O, T, D) := \underline{\overline{\phi}}(S, O|T) \otimes \underline{\overline{\psi}}(O, D),$$

where $\otimes$ is the operator introduced in Definition 6, are

$$\underline{\overline{\psi}}(T, O, D) = \begin{matrix} & & t & & nt & \\ & d & nd & d & nd & \\ \begin{bmatrix} [-90, -70] & [-20, 0] & [-80, -60] & [-10, 10] \\ [30, 50] & [-20, 0] & [40, 60] & [-10, 10] \\ [180, 200] & [-20, 0] & [190, 210] & [-10, 10] \end{bmatrix} & \begin{matrix} e \\ w \\ s \end{matrix} \end{matrix},$$

$$\underline{\overline{\phi}}(S, O|T) = \begin{matrix} & & t & & & nt & \\ & e & w & s & e & w & s & \\ \begin{bmatrix} [.045, .076] & [.081, .112] & [.090, .126] & [.150, .192] & [.090, .123] & [.060, 0.088] \\ [.135, .176] & [.108, .144] & [.072, .103] & [.150, .192] & [.090, .123] & [.060, 0.088] \\ [.271, .326] & [.081, .112] & [.018, .035] & [.150, .192] & [.090, .123] & [.060, 0.088] \end{bmatrix} & \begin{matrix} c \\ o \\ d \end{matrix} \end{matrix},$$

$$\overline{\underline{\psi}}(S, O, T, D) = \begin{bmatrix} [-5.709, -2.933] & [-.381, .381] & [-14.437, -9.777] & [-.962, .962] \\ [3.655, 6.172] & [-.561, .561] & [4.061, 6.756] & [-.614, .614] \\ [17.599, 25.83] & [-.630, .630] & [11.732, 18.040] & [-.440, .440] \\ [-13.191, -8.799] & [-.879, .879] & [-14.437, -9.777] & [-.962, .962] \\ [4.873, 7.923] & [-.720, .720] & [4.061, 6.756] & [-.614, .614] \\ [14.079, 21.156] & [-.516, .516] & [11.732, 18.040] & [-.440, .440] \\ [-24.413, -17.599] & [-1.627, 1.627] & [-14.437, -9.777] & [-.962, .962] \\ [3.655, 6.172] & [-.561, .561] & [4.061, 6.756] & [-.614, .614] \\ [3.520, 7.134] & [-.174, .174] & [11.732, 18.040] & [-.440, .440] \end{bmatrix} \begin{matrix} e \\ w \ \ c \\ s \\ e \\ w \ \ o \\ s \\ e \\ wd \\ s \end{matrix} .$$

The above combination operator generalizes the combination of precise potentials in Definition 3 by keeping the same commutative and associative properties. A deeper characterization is provided by the following result, which provides a sensitivity-analysis justification for the proposed generalization of the combination operator.

Proposition 1. *Given two potentials (no matter whether IUPs or IPPs) $\overline{\underline{\psi}}$ and $\overline{\underline{\phi}}$, the combination of the elements of their extensions is included in the extension of their combination, i.e.,*

$$\left\{ \psi \otimes \phi \,\middle|\, \psi \in \overline{\underline{\psi}}^*, \phi \in \overline{\underline{\phi}}^* \right\} \subseteq \overline{\underline{\psi \otimes \phi}}^* . \tag{12}$$

The proof easily follows from the fact that potentials consistent with the bounds on the right-hand side of Eqs. (9)–(11) cannot produce bounds not consistent with those on the left-hand side. The other operations over sharp-valued potentials required by the algorithms VE and AR (division, sum-marginalization and max-marginalization) can also be generalized to intervals and similar characterizations provided.

Definition 7 *(Dividing interval-valued potentials).* The ratio between an IUP $\overline{\underline{\psi}}(X_I)$ and an IPP $\overline{\underline{\phi}}(X_J)$ is an IUP $\overline{\underline{\psi/\phi}}$ over $X_{I \cup J}$ such that, for each $x_{I \cup J} \in \Omega_{X_{I \cup J}}$,

$$\overline{\psi/\phi}(x_{I \cup J}) := \overline{\psi}(x_I)/\underline{\phi}(x_J) , \tag{13}$$

$$\underline{\psi/\phi}(x_{I \cup J}) := \underline{\psi}(x_I)/\overline{\phi}(x_J) , \tag{14}$$

with $x_I, x_J \sim x_{I \cup J}$.

The ratio of two IPPs is analogously defined. With zero denominators, the result is set to $+\infty$ for positive numerators and $-\infty$ for negative ones. When both numerator and denominator are zero, we set $\frac{0}{0} = 0$.

Definition 8 *(Sum-marginalization).* The sum-marginalization $\overline{\underline{\sum_X \psi}}$ of an IUP $\overline{\underline{\psi}}(X, X_I)$ is an IUP over X_I such that

$$\overline{\sum_X \psi}(x_I) := \sum_{x \in \Omega_X} \overline{\psi}(x, x_I) , \tag{15}$$

$$\underline{\sum_X \psi}(x_I) := \sum_{x \in \Omega_X} \underline{\psi}(x, x_I) , \tag{16}$$

for each $x_I \in \Omega_{X_I}$.

The sum-marginalization of an IPP is analogously defined.

Definition 9 *(Max-marginalization).* The max-marginalization $\overline{\underline{\max_D \psi}}$ of an IUP $\overline{\underline{\psi}}(D, X_I)$ is an IUP over X_I such that

$$\overline{\max_D \psi}(x_I) := \max_{d \in \Omega_D} \overline{\psi}(d, x_I) , \tag{17}$$

$$\underline{\max_D \psi}(x_I) := \max_{d \in \Omega_D} \underline{\psi}(d, x_I) , \tag{18}$$

for each $x_I \in \Omega_{X_I}$.

The max-marginalization of an IPP is analogously defined. Envelope theorems analogous to that in Proposition 1 can be proved for all the operators defined in this section.

Example 6 *(Marginalization and division).* Consider the interval-valued potentials $\overline{\underline{\psi}}(S, O, T, D)$ and $\overline{\underline{\phi}}(S, O|T)$ obtained in Example 5. By sum-marginalizing out the variable O, we obtain

$$\overline{\underline{\psi}}(S, T, D) = \sum_O \overline{\underline{\psi}}(S, O, T, D) =$$

$$
\begin{array}{cccccc}
 & \multicolumn{2}{c}{t} & \multicolumn{2}{c}{nt} & \\
 & d & nd & d & nd & \\
\begin{bmatrix}
[15.544, 29.069] & [-1.572, 1.572] & [1.356, 15.019] & [-2.017, 2.017] \\
[5.762, 20.279] & [-2.116, 2.116] & [1.356, 15.019] & [-2.017, 2.017] \\
[-17.238, -4.292] & [-2.363, 2.363] & [1.356, 15.019] & [-2.017, 2.017]
\end{bmatrix} & & & &
\begin{array}{c} c \\ o \\ d \end{array}
\end{array} \quad ,
$$

$$\overline{\underline{\phi}}(S|T) = \sum_O \overline{\underline{\phi}}(S, O|T) =
\begin{array}{ccc}
 & t & nt \\
\begin{bmatrix}
[0.217, 0.314] & [0.301, 0.403] \\
[0.316, 0.423] & [0.301, 0.403] \\
[0.37, 0.473] & [0.301, 0.403]
\end{bmatrix} &
\begin{array}{c} c \\ o \\ d \end{array}
\end{array} \quad .$$

The division of the IUP and IPP previously obtained gives the IUP

$$\overline{\underline{\psi}}_2(S, T, D) = \frac{\overline{\underline{\psi}}(S, T, D)}{\overline{\underline{\phi}}(S|T)} =$$

$$
\begin{array}{cccccc}
 & \multicolumn{2}{c}{t} & \multicolumn{2}{c}{nt} & \\
 & d & nd & d & nd & \\
\begin{bmatrix}
[49.45, 134.207] & [-7.256, 7.256] & [3.363, 49.924] & [-6.704, 6.704] \\
[13.617, 64.201] & [-6.698, 6.698] & [3.363, 49.924] & [-6.704, 6.704] \\
[-46.585, -9.084] & [-6.385, 6.385] & [3.363, 49.924] & [-6.704, 6.704]
\end{bmatrix} & & & &
\begin{array}{c} c \\ o \\ d \end{array}
\end{array} \quad .
$$

Finally, the max-marginalization of D from $\overline{\underline{\psi}}_2(S, T, D)$ is

$$\overline{\underline{\psi}}(S, T) = \max_D \overline{\underline{\psi}}_2(S, T, D) =
\begin{array}{ccc}
 & t & nt \\
\begin{bmatrix}
[49.45, 134.207] & [3.363, 49.924] \\
[13.617, 64.201] & [3.363, 49.924] \\
[-6.385, 6.385] & [3.363, 49.924]
\end{bmatrix} &
\begin{array}{c} c \\ o \\ d \end{array}
\end{array} \quad .$$

4. Interval-valued influence diagrams

IDs can be extended to intervals by replacing the PPs and UPs in Definition 1 with an equal number of IPPs and IUPs defined over the same domains. A model of this kind is called an *interval-valued influence diagram* (IID). As an example, the interval-valued potentials in Example 4 can be used to transform into an IID the ID in Example 1.

IIDs offer a direct sensitivity analysis interpretation. An IID can be regarded as a collection of so-called *consistent* IDs, all with the same graph and set of variables, with PPs and UPs taking their values from the extensions of the IPPs and IUPs of the IID. IID evaluation is therefore intended as the calculation of the interval spanned by the MEU values of the consistent IDs. We similarly define the optimal policies of an IID as the union of those optimal, in the sense of Eq. (1), for at least a consistent ID.

Both the VE and AR schemes can be adopted for IIDs evaluation by replacing the operations over sharp potentials with the analogous operations for interval-valued potentials defined in Section 3.2. We show that this approach might produce unnecessarily large outer approximations. To avoid that, we propose a sophistication of these algorithms based on linear programming (Section 4.1 for VE and Section 4.3 for AR) as well as an alternative VE which gives faster but less accurate inferences (Section 4.2). The latter approach gives an outer approximation analogous to the generalization of the AR algorithm proposed by Fertig and Breese [13,12,6].

4.1. Variable elimination in IIDs by linear programming

Consider the VE scheme outlined by Algorithm 1. The procedure to eliminate a variable, detailed by Algorithm 2, is based on two sequential steps: first the potentials including the variable to eliminate in their arguments are combined (line 2), then the elimination is performed on the combined potential (lines 4 or 6). When coping with IIDs, we perform the last combination together with the elimination. This corresponds to a linear program, that avoids unnecessary additional approximations.

4.1.1. Chance variables elimination from IPPs

Because of Definition 1, only one of the PPs to be combined in line 2 of Algorithm 2 has Y on the left-hand side of its argument. The same holds with the IPPs of an IID. When eliminating a chance variable from the IPPs of an IID, we proceed as follows. First we combine with the operator in Definition 6 all the IPPs apart from the one having Y on the left. The corresponding IPP is combined indeed with the only IPP having Y on the left-hand side and, simultaneously, the variable is sum-marginalized (first term in line 4) as described by Definition 8. The procedure is detailed here below.

Definition 10 (*Eliminating chance variables from IPPs*). Consider the elimination of the chance variable Y during VE. Let $\overline{\underline{\phi}}(X_I|X_J, Y)$ denote the IPP obtained by combining all the IPPs with Y on the right-hand side, and $\overline{\underline{\phi}}(Y, X_K|X_L)$ the only IPP with Y on the left. The elimination of Y from the combination of these two IPPs generates an IPP $\overline{\underline{\phi}}(X_K, X_I|X_L, X_J)$. For each $x_{I \cup K} \in \Omega_{X_{I \cup K}}$ and $x_{L \cup J} \in \Omega_{X_{L \cup J}}$, an outer approximation of the lower bound $\underline{\phi}(x_{K \cup I}|x_{L \cup J})$ is the solution of the following task:

$$\text{minimize} \quad \sum_{y \in \Omega_Y} \phi(x_I|x_J, y) \cdot \phi(y, x_K|x_L),$$

$$\text{subject to} \quad \underline{\phi}(x_I|x_J, y) \le \phi(x_I|x_J, y) \le \overline{\phi}(x_I|x_J, y),$$

$$\underline{\phi}(y, x_K|x_L) \le \phi(y, x_K|x_L) \le \overline{\phi}(y, x_K|x_L), \forall y \in \Omega_Y.$$

The optimization variables $\{\phi(x_I|x_J, y)\}_{y \in \Omega_Y}$ are free to vary one independently of the other. Each one of these variables is in a different term of the objective function. Thus, we can easily minimize with respect to these variables and replace $\phi(x_I|x_J, y)$ with the lower bound $\underline{\phi}(x_I|x_J, y)$. This reduces the task to a linear program over the optimization variables $\{\phi(y, x_K|x_L)\}_{y \in \Omega_Y}$. Yet, it should be noticed that the optimization variables $\{\phi(y, x_K|x_L)\}_{y \in \Omega_Y}$ are not only required to satisfy the separate constraints reported in the above task, but also the normalization constraint of the PPs consistent with the IPP $\overline{\underline{\phi}}(y, x_K|x_L)$. These are constraints among the different tasks corresponding to the different values of x_K. By considering the reachability constraints for $\overline{\underline{\phi}}(X_K|X_L) := \sum_Y \overline{\underline{\phi}}(X_K, Y|L)$ (see Definition 8), we have

$$1 - \sum_{x'_K \ne x_K, y} \overline{\phi}(y, x'_K|x_L) \le \sum_y \phi(y, x_K|x_L) \le 1 - \sum_{x'_K \ne x_K, y} \underline{\phi}(y, x'_K|x_L). \tag{19}$$

Note that if $X_K = \emptyset$ the constraint in Eq. (19) degenerates in the trivial normalization of the potential and becomes useless.

Example 7. Consider the elimination of the chance variable O from the IID associated to the graph in Fig. 1 with the IPPs and IUPs as in Example 4. This consists in the combination $\overline{\underline{\phi}}(S, O|T) := \overline{\underline{\phi}}(O) \otimes \overline{\underline{\phi}}(S|O, T)$, and then the sum-marginalization $\overline{\underline{\phi}}(S|T) := \sum_o \overline{\underline{\phi}}(S, o|T)$. To show how this works in practice, let us compute the upper bound $\overline{\phi}(S = c|T = t)$. The corresponding linear program is:

$$\text{maximize} \quad .145 \cdot \phi(e) + .335 \cdot \phi(w) + .525 \cdot \phi(s),$$

$$\text{subject to} \quad .475 \le \phi(e) \le .525,$$

$$.285 \le \phi(w) \le .335,$$

$$.190 \le \phi(s) \le .240,$$

$$\phi(e) + \phi(w) + \phi(s) = 1.$$

The objective function is maximized when $\phi(e) = .475$, $\phi(w) = .285$ and $\phi(s) = .240$, which gives $\overline{\phi}(c|t) \simeq .290$. By iterating this procedure for all the values of S and T and for the lower bounds too, the following IPP is obtained:

$$\overline{\underline{\phi}}(S|T) = \begin{matrix} & t & nt & \\ \begin{bmatrix} [.221, .290] & [.317, .367] \\ [.330, .385] & [.317, .367] \\ [.375, .444] & [.317, .367] \end{bmatrix} & \begin{matrix} c \\ o \\ d \end{matrix} \end{matrix}$$

It is worth noticing that not including the additional constraints in Eq. (19) would have produced larger intervals.

4.1.2. Chance variables elimination from IUPs

Let us consider here how the second term of line 4 of Algorithm 2, i.e., the elimination of a chance variable Y from the utility potentials, can be achieved with IIDs. We first combine all the IUPs including Y in their arguments as in Definition 6. For the IPPs, we proceed as in the previous section by first combining all the IPPs with Y on the right-hand side. The remaining combinations and the division are performed simultaneously as described in the following definition.

Definition 11 *(Eliminating chance variables from IUPs).* Let $\overline{\phi}(X_I | X_J, Y)$ be the IPP obtained by combining all the IPP with Y on the right-hand side, $\overline{\phi}(Y, X_K | X_L)$ be the only IPP with Y on the left-hand side and $\overline{\psi}(Y, X_M)$ the combination of all the IUPs with Y in the argument. The elimination of a chance variable Y from the combination of these potentials produces a new IPP $\overline{\psi}(X_I, X_J, X_K, X_L, X_M)$. For each $x_{I \cup J \cup K \cup L \cup M} \in \Omega_{X_{I \cup J \cup K \cup L \cup M}}$, an outer approximation of the lower bound $\underline{\psi}(x_{I \cup J \cup K \cup L \cup M})$ is the solution of the task

$$
\text{minimize} \quad \frac{\sum_{y \in \Omega_Y} \phi(x_I | x_J, y) \cdot \phi(y, x_K | x_L) \cdot \psi(y, x_M)}{\sum_{y \in \Omega_Y} \phi(x_I | x_J, y) \cdot \phi(y, x_K | x_L)} \,,
$$

$$
\text{subject to} \quad \underline{\phi}(x_I | x_J, y) \le \phi(x_I | x_J, y) \le \overline{\phi}(x_I | x_J, y) \,,
$$

$$
\underline{\phi}(y, x_K | x_L) \le \phi(y, x_K | x_L) \le \overline{\phi}(y, x_K | x_L) \,,
$$

$$
\underline{\psi}(y, x_M) \le \psi(y, x_M) \le \overline{\psi}(y, x_M) \,.
$$

The task has a linearly constrained cubic fractional objective function. The minimization with respect to the optimization variables associated to an IUP can be trivially achieved by setting $\psi(y, x_M) = \underline{\psi}(y, x_M)$. Unlike the task in Definition 10, the optimization with respect to the optimization variables $\{\phi(x_I | x_J, y)\}$ is not trivial as the variables appear both in the numerator and in the denominator of the objective function. Nevertheless we can regard the product $\phi(y, x_K | x_L) \cdot \phi(x_I | x_J, y)$ as a single optimization variable subject to

$$
\underline{\phi}(y, x_K | x_L) \cdot \underline{\phi}(x_I | x_J, y) \le \phi(y, x_K | x_L) \cdot \phi(x_I | x_J, y) \le \overline{\phi}(y, x_K | x_L) \overline{\phi}(x_I | x_J, y) \,. \tag{20}
$$

In this way the task becomes a linear-fractional program which can be reduced to a linear program using the classical Charnes-Cooper transformation. This introduces an outer approximation, which can be partially mitigated by additional reachability constraints as in the previous section. In this case the constraints are

$$
1 - \sum_{\{x'_K, x'_I\} \ne \{x_K, x_I\}, y} \overline{\phi}(y, x'_K | x_L) \cdot \overline{\phi}(x_I |, x_J, y) \le \sum_y \phi(y, x_K | x_L) \cdot \phi(x_I | x_J, y)
$$

$$
\le 1 - \sum_{\{x'_K, x'_I\} \ne \{x_K, x_I\}, y} \underline{\phi}(y, x'_K | x_L) \cdot \underline{\phi}(x'_I | x_J, y) \,. \tag{21}
$$

As in the previous section, if $X_K = X_I = \emptyset$ the constraint becomes ineffective and the problem becomes linear (instead of linear-fractional).

Example 8. Consider the VE scheme applied to the oil wildcutter's IID. To remove the chance variable O from the IUPs, we should consider the IPPs $\overline{\phi}(O)$ and $\overline{\phi}(S | O, T)$ and the IUP $\overline{\psi}(O, D)$. A new IUP $\underline{\psi}(S, T, D)$ is obtained. The upper bound $\overline{\psi}(e, t, d)$ requires the solution of the fractional task

$$
\text{maximize} \quad \frac{-65 \cdot \phi(e) \cdot \phi(c | e, t) + 55 \cdot \phi(w) \cdot \phi(c | w, t) + 205 \cdot \phi(s) \cdot \phi(c | s, t)}{\phi(e) \cdot \phi(c | e, t) + \phi(w) \cdot \phi(c | w, t) + \phi(s) \cdot \phi(c | s, t)} \,,
$$

$$
\text{subject to} \quad .475 \cdot .095 \le \phi(e) \cdot \phi(c | e, t) \le .525 \cdot .145 \,,
$$

$$
.285 \cdot .285 \le \phi(w) \cdot \phi(c | w, t) \le .335 \cdot .335 \,,
$$

$$
.190 \cdot .475 \le \phi(s) \cdot \phi(c | s, t) \le .240 \cdot .525 \,,
$$

$$
.217 \le \phi(e) \cdot \phi(c | e, t) + \phi(w) \cdot \phi(c | w, t) + \phi(s) \cdot \phi(c | s, t) \le .314 \,.
$$

The maximum is 108.44 which is achieved when the first two variables takes their minimum value and the third its maximum. By solving similar tasks for each joint state in $\Omega_S \times \Omega_T \times \Omega_D$, we obtain the IUP

$$
\overline{\psi}(S, T, D) = \begin{bmatrix} (t, d) & (t, nd) & (nt, d) & (nt, nd) \\ [60.8, 108.44] & [-5.0, 5.0] & [3.96, 41.57] & [-5.0, 5.0] \\ [16.17, 53.0] & [-5.0, 5.0] & [3.96, 41.57] & [-5.0, 5.0] \\ [-40.58, -10.27] & [-5.0, 5.0] & [3.96, 41.57] & [-5.0, 5.0] \end{bmatrix} \begin{matrix} c \\ o \\ d \end{matrix} \,.
$$

In this particular case, as $.475 \cdot .095 + .285 \cdot .285 + .19 \cdot .475 \simeq .217$, not including the additional constraints in Eq. (21) does not make the result less accurate.

4.1.3. Decision variables elimination

Here we discuss how to extend the operations in lines 6 and 7 of Algorithm 2 to IIDs. The arg max operation is intrinsically related to the fact that a UP has sharp values. To decide the optimal options when comparing intervals, we adopt a conservative approach, called *interval dominance* in the imprecise-probability jargon [27], which rejects all the decisions leading to certainly sub-optimal strategies. The procedure is described here below.

Definition 12 (Interval optimality). Let $\overline{\underline{\psi}}$ be an IUP over $Y \cup X_I$. An element $y \in \Omega_Y$ is *interval-optimal* given $x_I \in \Omega_{X_I}$ if there is no $y' \in \Omega_Y \setminus \{y\}$ such that $\underline{\psi}(y', x_I) > \overline{\psi}(y, x_I)$.

Let D be a decision variable to be eliminated from $\overline{\underline{\psi}}(D, X_I)$ during the VE (if there are multiple potentials they are combined using Definition 6). To detect the optimal policy $\delta_D^*(X_I)$ we compute the interval-optimal states of D given each $x_I \in \Omega_{X_I}$. This corresponds to a so-called *credal* policy allowing for indecision between two or more possible options. Finally the maximization of the IUP giving as result a new IUP $\overline{\underline{\psi}}(X_I)$ is done as explained in Definition 9, i.e., by acting separately on the two bounds. Note that the elimination of a decision does not involve any computation over IPPs.

Example 9. In the oil wildcatter's IID, the removal of the decision D involves the IUP $\overline{\underline{\psi}}(S, T, D)$ obtained in Example 8. The resulting IUP and the corresponding optimal policy are

$$\overline{\underline{\psi}}(S, T) = \begin{bmatrix} [60.8, 108.44] & [3.96, 41.57] \\ [16.17, 53.0] & [3.96, 41.57] \\ [-5.0, 5.0] & [3.96, 41.58] \end{bmatrix} \begin{matrix} c \\ o \\ d \end{matrix} \, , \quad \delta_D^*(S, T) = \begin{bmatrix} \{d\} & \{d, nd\} \\ \{d\} & \{d, nd\} \\ \{nd\} & \{d, nd\} \end{bmatrix} \begin{matrix} c \\ o \\ d \end{matrix} \, .$$

Note that for some configurations, the interval resulting from the maximization corresponds with only one of the alternatives given as optimal.

4.2. A faster outer approximation

In this section we propose an alternative approach to the generalization of VE to IIDs, which does not require any linear program to be solved. This corresponds to a heuristic approach to the solution of those optimizations, that introduces an additional outer approximation. The only difference with respect to the VE described in the previous section is the chance variable elimination from IUPs for which we propose a procedure different from that in Definition 11. Suppose that we aim to remove a variable Y from an IPP $\overline{\underline{\phi}}(Y, X_I | X_J)$ and an $\overline{\underline{\psi}}(Y, x_K)$. The elimination of Y generates a new IUPs $\overline{\underline{\psi}}(X_{I \cup J \cup K})$ such that, for each $x_{I \cup J \cup K} \in \Omega_{X_{I \cup J \cup K}}$,

$$\overline{\psi}(x_{I \cup J \cup K}) := \sum_{y \in \Omega_Y} \frac{\overline{\phi}(y, x_I | x_J)}{\overline{\phi}(y, x_I | x_J) + \sum_{y' \neq y} \underline{\phi}(y', x_I | x_J)} \cdot \overline{\psi}(y, x_K) \,, \tag{22}$$

$$\underline{\psi}(x_{I \cup J \cup K}) = \sum_{y \in \Omega_Y} \frac{\underline{\phi}(y, x_I | x_J)}{\underline{\phi}(y, x_I | x_J) + \sum_{y' \neq y} \overline{\phi}(y', x_I | x_J)} \cdot \underline{\psi}(y, x_K) \,. \tag{23}$$

If there is more than an IPP with Y (either on the left or on right), all the IPPs are combined using Definition 6. Compared to what is done by Definition 11, the above considered potential provides an outer approximation based on the *one-sided* potentials defined in Section 3. An example is shown here below.

Example 10. The removal of the chance variable O (from the IUPs) involves the IPPs $\overline{\underline{\phi}}(O)$ and $\overline{\underline{\phi}}(S | O, T)$ and the IUP $\overline{\underline{\psi}}(O, D)$. A new IUP $\overline{\underline{\psi}}(S, T, D)$ is obtained. E.g., according to Eq. (23), $\overline{\psi}(c, t, d) = 107.302$, which corresponds to

$$\frac{.525 \cdot .145 \cdot (-65)}{.525 \cdot .145 + .285 \cdot .285 + .19 \cdot .475} + \frac{.335 \cdot .335 \cdot 55}{.475 \cdot .095 + .335 \cdot .335 + .19 \cdot .475} + \frac{.24 \cdot .525 \cdot 205}{.475 \cdot .095 + .285 \cdot .285 + .24 \cdot .525} \,.$$

By similarly proceeding for all the joint states in $\Omega_S \times \Omega_T \times \Omega_D$, we obtain the IUP shown here below. Note that the intervals in Example 8 are included in those of the current IUP.

$$\overline{\underline{\psi}}(S, T, D) = \begin{bmatrix} [64.124, 107.302] & [-3.849, 6.3] & [10.971, 38.662] & [-4.1, 5.988] \\ [21.95, 51.444] & [-4.088, 6.003] & [10.971, 38.662] & [-4.1, 5.988] \\ [-32.605, -15.972] & [-4.358, 5.681] & [10.971, 38.662] & [-4.1, 5.988] \end{bmatrix} \begin{matrix} c \\ o \\ d \end{matrix}$$

with columns (t, d), (t, nd), (nt, d), (nt, nd).

4.3. Arc reversal in IIDs by linear programming

The AR scheme outlined in Algorithm 3 evaluates IDs by performing three basic operations: elimination of chance (Transformation 2) and decision (Transformation 3) variables and arc reversal (Transformation 4). The first operation is just a particular case of the chance variable elimination from IUPs explained in Section 4.1.2. Yet, in AR, the removal of a chance variable Y always involves an IPP $\overline{\phi}(Y|X_I)$ such that Y is the only variable on the left-hand side and an IUP $\overline{\psi}(Y, X_J)$. Because of the normalization constraint, $\overline{\sum_Y \phi(Y|X_J)}$ is a constant potential always equal to one and the removal corresponds to a simple linear program without fractional terms. Similarly, the second operation consists in removing a decision variable D from an IUP, say $\overline{\psi}(D, X_I)$, and deciding the optimal policy $\delta_{D_i}^*$. This is done exactly in the same way as for the VE algorithm (see Section 4.1.3). Finally, to reverse arcs, we generalize Transformation 4 to intervals as follows.

Transformation 5 *(Interval arc reversal). Assume that the chance nodes Y and X of an IID are directly connected by an arc, but not by other directed paths. Let $\overline{\phi}(Y|X_I)$ and $\overline{\phi}(X|Y, X_J)$ be the relative PPs, which means that X_I are the direct predecessors of Y and X_J those of X others than Y. Change the orientation of the arc and add arcs from X_I towards X and from X_J towards Y. Then replace the original IPPs with $\overline{\phi}(X|X_I, X_J)$ and $\overline{\phi}(Y|X, X_I, X_J)$, where the first IPP is obtained by sum-marginalization and the second is such that $\underline{\phi}(y|x, x_I, x_J)$ is the minimum of*

$$\frac{\phi(y|x_I) \cdot \phi(x|y, x_J)}{\sum_{y' \in \Omega_Y} \phi(y'|x_I) \cdot \phi(x|y', x_J)}, \tag{24}$$

with respect to the interval constraints induced by the two original IPPs, for each $x \in \Omega_X$, $y \in \Omega_Y$, and $x_{I \cup J} \in \Omega_{X_{I \cup J}}$.

To solve the above optimization, we can use the same optimization strategy considered by Zaffalon in his naive credal classifier [29]. Accordingly, we divide the denominator by the numerator and rewrite Eq. (24) as

$$\left[1 + \sum_{y' \neq y} \frac{\phi(y'|x_I) \cdot \phi(x|y', x_J)}{\phi(y|x_I) \cdot \phi(x|y, x_J)} \right]^{-1}. \tag{25}$$

Then, as $f(t) = [1 + t]^{-1}$ is a monotone decreasing function of $t \in \mathbb{R}$, we reduce the minimization of the objective function in Eq. (24) or Eq. (25), to the maximization of

$$\sum_{y' \neq y} \frac{\phi(y'|x_I) \cdot \phi(x|y', x_J)}{\phi(y|x_I) \cdot \phi(x|y, x_J)}. \tag{26}$$

As there are no constraints over the optimization variables $\{\phi(x|y, x_J)\}_{y \in \Omega_Y}$ we perform the optimization with respect to these variable and obtain the objective function

$$\sum_{y' \neq y} \frac{\phi(y'|x_I) \cdot \overline{\phi(x|y', x_J)}}{\phi(y|x_I) \cdot \underline{\phi(x|y, x_J)}}, \tag{27}$$

which is a linearly constrained linear-fractional objective function. The task can be therefore reduced to a linear program.

Example 11. The reversal of the arc from O to S involves the IPPs $\overline{\phi}(S|O, T)$ and $\overline{\phi}(O)$. The resulting IPPs are $\overline{\phi}(S|T)$ and $\overline{\phi}(O|S, T)$. Computing the upper bound $\overline{\phi}(e|c, t)$ of the IPP attached to O corresponds to the following linear program

$$\begin{aligned} \text{minimize} \quad & \frac{\phi(w) \cdot .285}{\phi(e) \cdot .145} + \frac{\phi(s) \cdot .475}{\phi(e) \cdot .145}, \\ \text{subject to} \quad & .475 \leq \phi(e) \leq .525, \\ & .285 \leq \phi(w) \leq .335, \\ & .19 \leq \phi(s) \leq .24, \\ & \phi(e) + \phi(w) + \phi(s) = 1. \end{aligned}$$

The minimum value of the previous function is 2.2525, and hence

$$\overline{\phi}(e|c, t) = \frac{1}{1 + 2.2525} \simeq .307.$$

By solving similar linear programs for each bound and state of $\Omega_O \times \Omega_S \times \Omega_T$, we obtain the IPP

$$\overline{\underline{\psi}}(S,T,D) = \begin{bmatrix} [0.169, 0.307] & [0.439, 0.561] & [0.375, 0.494] & [0.439, 0.561] & [0.66, 0.766] & [0.439, 0.561] \\ [0.294, 0.453] & [0.256, 0.368] & [0.292, 0.41] & [0.256, 0.368] & [0.187, 0.28] & [0.256, 0.368] \\ [0.333, 0.499] & [0.168, 0.268] & [0.192, 0.298] & [0.168, 0.268] & [0.041, 0.09] & [0.168, 0.268] \end{bmatrix} \begin{matrix} c \\ o \\ d \end{matrix}$$

with column labels $(e,t)\ (e,nt)\ (w,t)\ (w,nt)\ (s,t)\ (s,nt)$.

The upper bound $\overline{\phi}(c|t)$ of the IPP attached to the final parent is the solution of a linear program with the same constraints as in the previous optimization and objective function to be maximized: $\phi(e) \cdot .145 + \phi(w) \cdot .335 + \phi(s) \cdot .525$. This function is maximized for $\phi(e) = .475$, $\phi(w) = .285$ and $\phi(s) = .24$. Thus, $\overline{\phi}(c|t) = .29$. If similar programs are solved for each bound and state of $\Omega_S \times \Omega_T$, the following IPP is obtained

$$\overline{\underline{\phi}}(S|T) = \begin{bmatrix} [.221, .290] & [.330, .385] & [.375, .449] \\ [.317, .367] & [.317, .367] & [.317, .367] \end{bmatrix} \begin{matrix} t \\ nt \end{matrix} \quad .$$

with column labels $c \quad o \quad d$.

4.4. Complexity analysis

The asymptotic complexity of the algorithms proposed in this section for IIDs evaluation is just the same as that of their IDs counterparts. Roughly speaking, with intervals, we perform the double of the number of arithmetic operations required with sharp values. The time required to run the linear programs is polynomial in the number of variables and constraints, which in turn depends on the arity of the local potentials involved during the elimination.

5. Sensitivity analysis

The algorithms for IIDs evaluation developed in Section 4 can be used for practical sensitivity analysis in standard IDs. The sharp-valued potentials of an ID can be replaced by interval-valued potentials whose extensions include the original potentials (e.g., as the extensions of the interval-valued potentials in Example 4 contain the potentials in Example 1). These sets of potentials are intended to describe the possible effects of a perturbation of the sharp values of the parameters. In particular we want to parametrize the level of perturbation as described in the following definition.

Definition 13 (*Nested perturbations*). Given a potential ψ, no matter whether PP or UP, a *parametrized nested perturbation of* ψ is denoted as $\overline{\underline{\psi}}_\varepsilon$. For each $\varepsilon \geq 0$, $\overline{\underline{\psi}}_\varepsilon$ is an interval-valued potential. We require that: (i) $\varepsilon \leq \varepsilon' \Rightarrow \overline{\underline{\psi}}_\varepsilon^* \subseteq \overline{\underline{\psi}}_{\varepsilon'}^*$; and (ii) $\overline{\underline{\psi}}_{\varepsilon=0}^* = \psi$.

Let us describe some practical ways to implement nested perturbations as in Definition 13. For UPs, we perform a *rectangular* perturbation symmetrical with respect to the original sharp values. In other words, if $\psi(X_I)$ is an UP over the set of variables X_I, then $\overline{\underline{\psi}}_\epsilon(X_I)$ is an IPP such that, for each $x_I \in \Omega_{X_I}$,

$$\underline{\psi}_\varepsilon(x_I) := \psi(x_I) - \varepsilon, \tag{28}$$

$$\overline{\psi}_\varepsilon(x_I) := \psi(x_I) + \varepsilon. \tag{29}$$

Rectangular perturbations cannot be applied to PPs, because of the normalization and nonnegativity constraints. We consider instead nested perturbations in the form of ϵ-contaminations. Given an (unconditional) PP ϕ over X_I, the perturbed IPP $\overline{\underline{\phi}}_\epsilon$ is such that, for each $x_I \in \Omega_{X_I}$,

$$\underline{\phi}_\varepsilon(x_I) := (1 - \varepsilon) \cdot \phi(x_I), \tag{30}$$

$$\overline{\phi}_\varepsilon(x_I) := (1 - \varepsilon) \cdot \phi(x_I) + \varepsilon. \tag{31}$$

Such perturbation is only defined for $0 \leq \varepsilon \leq 1$ and, for $\varepsilon = 1$ the extension of the PP coincide with the so-called *vacuous* set, including any possible PP specification over X_I. We can similarly perturb a conditional PP $\phi(X_I|X_J)$, by applying the above procedure for each $x_J \in \Omega_{X_J}$.

Finally, let us introduce the notion of *critical level of perturbation* ε^*. We define ε^* as the maximum value of ϵ such that all the optimal policies of the corresponding IID (obtained according to Definition 12) return single decisions. The value of ε^* can be obtained with a bracketing over the parameter ε by running the IID evaluation algorithms described in Section 4 for different perturbation levels. Alternatively, we can also characterize the robustness of the model by computing the *failure level of perturbation* ε^{**}, which is intended as the minimum value of ϵ such that all the optimal policies of the corresponding IID are *vacuous*, i.e., all the decisions are returned. The perturbation can be simultaneously applied to all the potentials in the IID or restricted to a specific IPPs or IUPs. In the latter case it is possible to determine which one of the potentials of an ID has a higher impact on the MEU. This gives important information about the parameters deserving a more careful elicitation. A demonstrative example is reported here below.
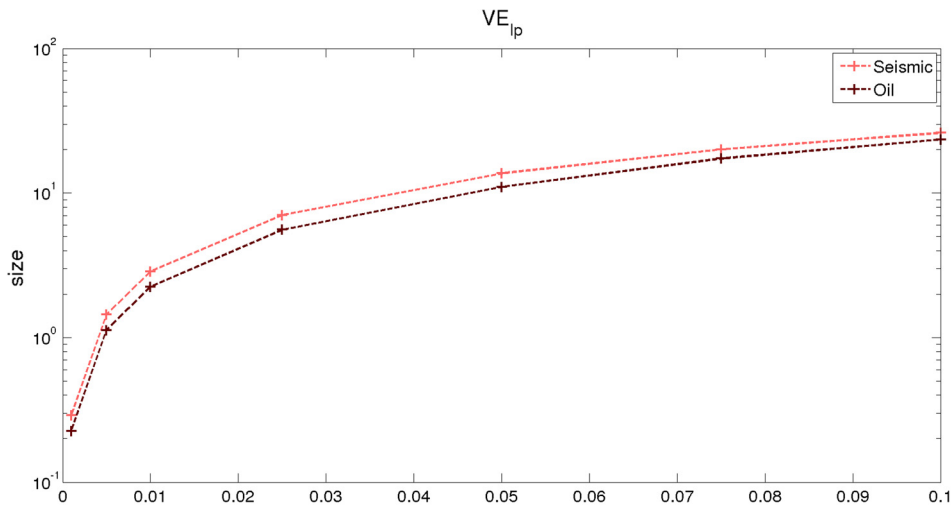
Fig. 4. Size of the interval-valued MEU as a function of the perturbation level ϵ.

Table 1
Number of chance, decision and utility nodes for the benchmark IIDs.

Name of the ID	$|\mathbf{X}|$	$|\mathbf{D}|$	$|\mathbf{U}|$
Appendicitis	4	1	1
Thinkbox	5	2	4
Oil	2	2	2
Oil Split Costs	2	2	3
NHL	17	3	1
Jaundice	21	2	1
Threat of Entry	3	9	1
Comp. Assym.	3	5	1
Chest Clinic	8	2	2

Example 12 *(Sensitivity analysis of the oil wildcatter's problem).* Consider the ID in Example 1. To evaluate the corresponding IID obtained by perturbation of this model the VE algorithm with linear programs is considered. We perturb the PPs associated to S (Seismic) and to O (Oil), i.e. $\phi(S|O, T)$ and $\phi(O)$. Fig. 4 depicts the size of the interval-valued MEU for increasing level of perturbation ϵ of these two potentials. The result is clear: perturbing $\phi(S|O, T)$ has a stronger effect than perturbing $\phi(O)$. Accordingly, we might conclude that the PP of S deserves a more careful quantification than that of O. Similar results are obtained by computing the critical perturbation levels ($\varepsilon^* = .0082$ for S and $\varepsilon^* = .0089$ for O) and the failure perturbation levels ($\varepsilon^{**} = .3749$ for S and $\varepsilon^{**} = .7499$ for O). Similar values and conclusions are obtained with the AR algorithm.

6. Empirical validation

For an empirical validation of the VE and AR algorithm for IIDs proposed in Section 4, we consider a benchmark of nine IDs modeling real decision tasks. Table 1 details the number of nodes of each type for these models. These IDs are transformed in IIDs by a perturbation of the original parameters based on the procedure described in Section 5. Besides the three algorithms proposed in Section 4, we also consider the generalization of the AR to IIDs as proposed by Fertig and Breese [13]. We denote as VE_{lp} our VE scheme based on linear programming, as VE_{outer} the faster version proposed in Section 4.2, as AR_{lp} our AR scheme (Section 4.3), and as AR_{fb} the algorithm of Fertig and Breese.

Fig. 5 shows the running times and the relative durations when compared with those of the precise counterparts (VE or AR for IDs). As expected, the two simplest approaches (AR_{fb} and VE_{outer}) roughly take the double of the time required by the precise evaluations (both upper and lower bounds are computed). Methods using linear programs (AR_{lp} and VE_{lp}) are slower due to the time required by the linear solver. In particular, the evaluation might be demanding if there are chance variables with many states, this being the case of NHL (which has a chance variable with twelve states). If we compare the interval versions of AR against VE, we see that the differences are typically small for small ID/IIDs, while with large models
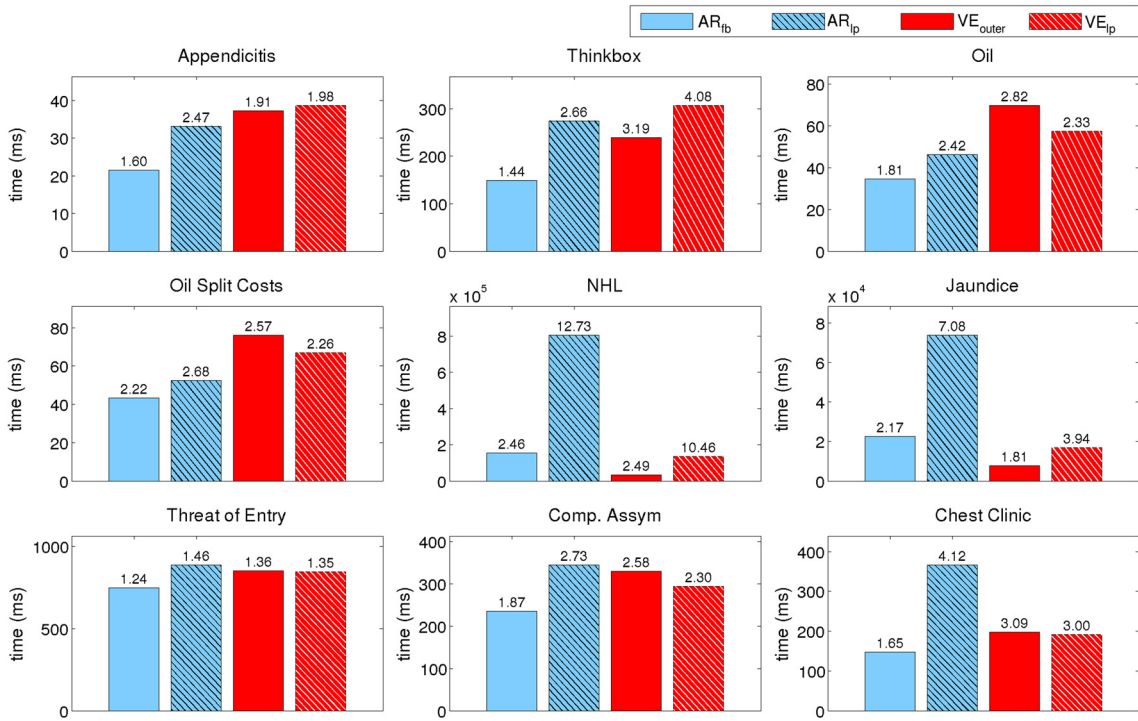
Fig. 5. Absolute (y-axis) and relative (numbers over the bars) running times for the IIDs in Table 1.

such as NHL or Jaundice AR might be very slow. In fact the reversal of and arc might introduce very large potentials, this being a very well-known issue even with standard IDs. All these algorithms have been implemented in Java.[4]

We also analyze the size of the interval-valued MEU as a function of the size of the intervals in the initial potentials (parametrized by the perturbation level ϵ). The results obtained with precise utilities are depicted in Fig. 6. As expected, the results based on the linear programming (AR_{lp} and VE_{lp}) are the most informative ones for all the IIDs. AR_{fb} is much less accurate with *Comp. Assym.* and *Threat of Entry*. This might be due to the high number of decisions: a weakness of AR_{fb} is that the maximization of an IUP is done by taking the highest upper bounds and the lowest upper bounds (instead of the highest lower bounds). VE_{outer} is also generally inaccurate and should be regarded as the algorithm of choice only if very severe constraints are posed on the running time. Similar results, with a sharp specification of the IPPs and intervals only in the IUPs are reported in Fig. 7. The two VE methods, which differs only in the treatment of the IPPs, produce the same results. In general, when only utilities are imprecise AR_{lp} offers the best results. Finally, by comparing the y-scales in Figs. 6 and 7, it can be observed that the imprecision in the IPPs seems to have a stronger effect than that in the IUPs on the size of the interval MEU.

7. Conclusions and future work

We have extended the ID formalism to support an interval-valued specification of the potentials. The corresponding models, called IIDs, have a direct sensitivity-analysis interpretation: an IID is equivalent to a collection of precise IDs whose potentials are consistent with the constraints induced by the intervals. Consequently, the set-valued (so-called *credal*) optimal policies of an IID include the single-valued optimal policies of the consistent IDs, as well as the interval-valued MEU of an IID contains all the MEU of the consistent IDs. Moreover, we extended to IIDs the classical *variable elimination* and *arc reversal* evaluation schemes for IDs. These two extensions are achieved by local optimization tasks, reduced to linear programs. For VE, a faster but less accurate procedure, that does not require linear programming, is also proposed. The latter approach introduces an additional outer approximation similar to that characterizing the generalization of the AR algorithm proposed by Fertig and Breese [13,12,6]. All these algorithms have the same asymptotic complexities of their classical, sharp-valued, counterparts. The empirical analysis showed that the approximations we introduce to keep the same complexity as with IDs did not compromise the informativeness of the inferences. In particular, the new methods based on linear programming are clearly more accurate that the algorithm of Fertig and Breese. Finally, we also proposed a possible application of IIDs to practical sensitivity analysis in precise IDs. Computing the maximal level of perturbation, no matter whether local or global, might allow to decide which are the potential/variables deserving a more careful elicitation process.

---

[4] All the software material used for the experiments presented in this section is freely available at the address http://leo.ugr.es/rcabanas/intervalids/.
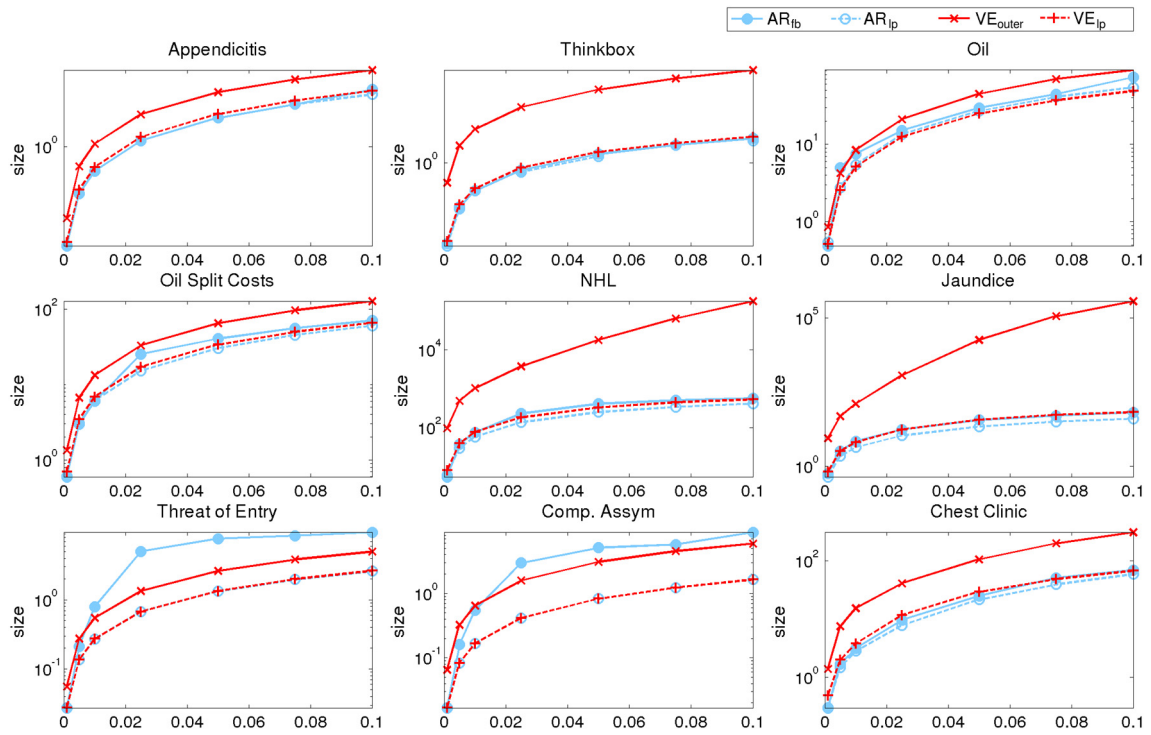
Fig. 6. Size of interval-valued MEU as a function of the perturbation level ϵ of the IPPs.
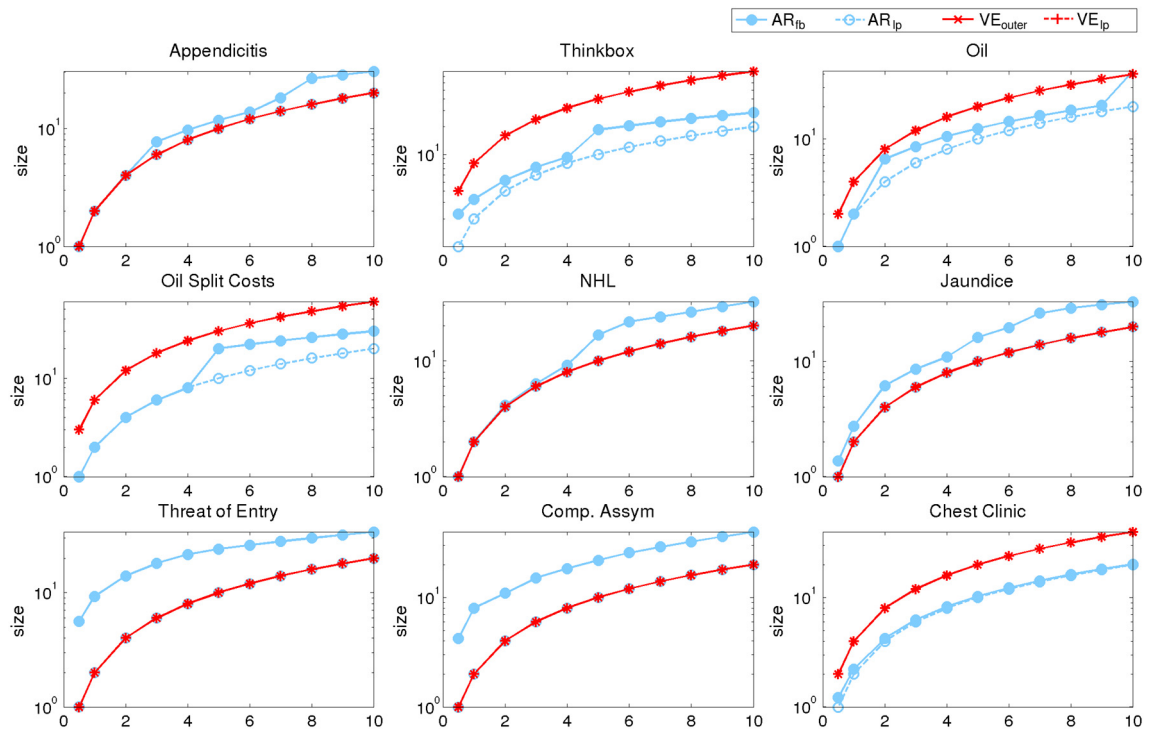

Fig. 7. Size of interval-valued MEU as a function of the perturbation level ϵ of the IUPs.

As a future work we intend to extend this formalism to more general imprecise frameworks, e.g., credal sets represented by extreme points or generic linear constraints. This should affect the computational complexity of evaluation process, thus making necessary the development of specific approximate algorithms. We also intend to apply some ideas proposed in a recent paper [30] about the relations between (sets of) probabilities and (sets of) utilities, to reduce inference in IIDs to inference in credal networks [1,2]. We also intend to extensively test the procedure we proposed for sensitivity analysis in practical IDs and compare it against the methods proposed so far with the same goal.

Acknowledgements

References

[1] A. Antonucci, C. de Campos, M. Zaffalon, Probabilistic graphical models, in: T. Augustin, F. Coolen, G. de Cooman, M. Troffaes (Eds.), Introduction to Imprecise Probabilities, Wiley, 2014, pp. 207–229, Chap. 9.
[2] A. Antonucci, C. de Campos, M. Zaffalon, D. Huber, Approximate credal network updating by linear programming with applications to decision making, Int. J. Approx. Reason. 58 (2014) 25–38.
[3] M.A. Artaso Landa, An empirical comparison of Influence Diagrams algorithms, Master's thesis, UNED, Escuela Técnica Superior de Ingeniería Informática, 2014.
[4] S. Benferhat, S. Smaoui, Hybrid possibilistic networks, Int. J. Approx. Reason. 44 (3) (2007) 224–243.
[5] H.L. Bodlaender, A linear time algorithm for finding tree-decompositions of small treewidth, in: Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing, ACM, 1993, pp. 226–234.
[6] J.S. Breese, K.W. Fertig, Decision making with interval influence diagrams, in: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, Elsevier Science Inc., 1990, pp. 467–480.
[7] C.J. Butz, J. Chen, K. Konkel, P. Lingras, A comparative study of variable elimination and arc reversal in Bayesian network inference, in: FLAIRS Conference, 2009.
[8] R. Cabañas, A. Antonucci, A. Cano, M. Gómez-Olmedo, Variable elimination for interval-valued influence diagrams, in: Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 13th European Conference, Proceedings, ECSQARU 2015, Compiègne, France, July 15–17, 2015, in: LNAI, vol. 9161, Springer, 2015, pp. 541–551.
[9] F.G. Cozman, Credal networks, Artif. Intell. 120 (2000) 199–233.
[10] L.M. de Campos, J.F. Huete, S. Moral, Probability intervals: a tool for uncertain reasoning, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 2 (02) (1994) 167–196.
[11] E. Fagiuoli, M. Zaffalon, Decisions under uncertainty with credal influence diagrams, Tech. rep. 51-98, IDSIA, 1998 (unpublished).
[12] K.W. Fertig, J.S. Breese, Interval influence diagrams, in: Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence, North-Holland Publishing Co., 1990, pp. 149–162.
[13] K.W. Fertig, J.S. Breese, Probability intervals over influence diagrams, IEEE Trans. Pattern Anal. Mach. Intell. 15 (3) (1993) 280–286.
[14] L. Garcia, R. Sabbadin, Complexity results and algorithms for possibilistic influence diagrams, Artif. Intell. 172 (8–9) (2008) 1018–1044.
[15] R.A. Howard, J.E. Matheson, Influence diagram retrospective, Decis. Anal. 2 (3) (2005) 144–147.
[16] N. Huntley, M.C.M. Troffaes, Normal form backward induction for decision trees with coherent lower previsions, Ann. Oper. Res. 195 (1) (2012) 111–134.
[17] G. Jeantet, O. Spanjaard, Computing rank dependent utility in graphical models for sequential decision problems, Artif. Intell. 175 (7) (2011) 1366–1389.
[18] F. Jensen, T. Nielsen, Bayesian Networks and Decision Graphs, Springer Verlag, 2007.
[19] D. Kikuti, F.G. Cozman, C.P. de Campos, Partially ordered preferences in decision trees: computing strategies with imprecision in probabilities, in: IJCAI Workshop on Advances in Preference Handling, 2005, pp. 118–123.
[20] U. Kjaerulff, Triangulation of graphs – algorithms giving small total state space, Research report R-90-09, Department of Mathematics and Computer Science, Aalborg University, Denmark, 1990.
[21] D. Koller, N. Friedman, Probabilistic Graphical Models: Principles and Techniques, MIT Press, 2009.
[22] T. Nielsen, F. Jensen, Sensitivity analysis in influence diagrams, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 33 (2) (2003) 223–234.
[23] H. Raiffa, Decision Analysis: Introductory Lectures on Choices Under Uncertainty, Addison-Wesley, 1968.
[24] N. Robertson, P.D. Seymour, Graph minors. IV. Tree-width and well-quasi-ordering, J. Comb. Theory, Ser. B 48 (2) (1990) 227–254.
[25] R. Shachter, Evaluating influence diagrams, Oper. Res. (1986) 871–882.
[26] P.P. Shenoy, Valuation-based systems for Bayesian decision analysis, Oper. Res. 40 (3) (1992) 463–484.
[27] M.C.M. Troffaes, Decision making under uncertainty using imprecise probabilities, Int. J. Approx. Reason. 45 (1) (2007) 17–29.
[28] H. Xu, P. Smets, Reasoning in evidential networks with conditional belief functions, Int. J. Approx. Reason. 14 (2–3) (1996) 155–185.
[29] M. Zaffalon, The naive credal classifier, J. Stat. Plan. Inference 105 (1) (2002) 5–21.
[30] M. Zaffalon, E. Miranda, Desirability and the birth of incomplete preferences, CoRR, arXiv:1506.00529, http://arxiv.org/abs/1506.00529.
[31] N. Zhang, D. Poole, Exploiting causal independence in Bayesian network inference, J. Artif. Intell. Res. 5 (1996) 301–328.
[32] L. Zhou, W. Liu, L. Wang, Influence diagram model with interval-valued utilities, in: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, IEEE Computer Society, 2009, pp. 601–605.