Graphical Models for Imprecise Probabilities

Fabio Gagliardi Cozman [a]

[a] *Decision Making Lab - Escola Politécnica - Universidade de São Paulo*
Av.Prof. Mello Moraes, 2231, São Paulo, SP - Brazil
fgcozman@usp.br

Abstract

This paper presents an overview of graphical models that can handle imprecision in probability values. The paper first reviews basic concepts and presents a brief historical account of the field. The main characteristics of the credal network model are then discussed, as this model has received considerable attention in the literature.

Key words: Imprecise probabilities, sets of probability distributions, graphical models, credal networks

1 Introduction

Geographic, biologic, economic, and many other kinds of relations are routinely depicted with graphs. Consequently it should not be surprising that graphs are also employed to represent interactions among random variables: for example, Bayesian networks and Markov random fields use graph-theoretic concepts to represent complex statistical situations. Perhaps the most profound contribution of graph-theoretic ("graphical") methods in probabilistic modeling has been a way of thinking that emphasizes locality of interactions as the key to compactness and efficiency. Graphs form a language; this language is visually pleasant and computationally efficient. What else could be asked for?

Researchers interested in non probabilistic calculi have not dismissed the success of graphical models. Imprecise probabilities and graphs have been married quite a few times, either because one wishes to extend the success of standard graphical models to the realm of imprecise probabilities, or because one thinks that standard graphical models are unrealistic unless they can handle imprecision in probability values. This paper offers an overview of graphical models aimed at imprecise probabilities, with the primary intent to be introductory and didactic — at the expense of formality and technical detail. Some historical perspective is provided in Section 3, but no comprehensive review is attempted. The strategy here is to focus on a particular type of model, credal networks, and to use this model to convey the central challenges and promises of graphical techniques.

2 Graphical models for precise probabilities

The purpose of this section is to fix key terminology and to indicate the scope of the paper. A *graph* is an object consisting of a set of *nodes* and a set of *edges* connecting nodes [7]. In this paper we focus on graphical models that have nodes/edges associated with statistical objects.

A *Bayesian network* consists of a directed acyclic graph where each node is associated with a random variable and with conditional probability distributions [79]. (Note that here we start using "node" and "variable" interchangeably.) Edges indicate direct dependency, and are often embodied with a causal interpretation: an edge from X to Y suggests that X somehow causes Y [80,95]. An *influence diagram* is similar to a Bayesian network, but is equipped with *decision* and *value* nodes; the purpose of an influence diagram is to represent sequential decision problems in compact form [24,63]. A *Markov random field* consists of an undirected graph where each clique (completely connected group of nodes) is associated with a non-negative function, called a *potential* [67]. Other models combine directed, undirected, bidirectional and dotted edges [27,28,70,80,95]. In *Markov Decision Processes*, each node represents a state, and the process can transit from a state to the next state by a number of paths (the edges) [11,82]. Typically these graphical models are used either to produce *inferences* (the computation of the posterior probability for one or more events) or to produce configurations of variables that maximize some appropriate quantity.

Central to all these graphical models are *Markov properties.* A Markov property relates graphical entities to probabilistic independence relations. For example, take the graph in Figure 1. What is the "Bayesian network" interpretation for this graph? The answer is given by the *Markov property for Bayesian networks*: the nondescendant nonparents of a node are independent of the node conditional on the node's parents. For instance, the graph in Figure 1 imposes independence of X and Z conditional on Y. Markov random fields, Markov Decision Processes, and other graphical models display different Markov properties.

In short, we have that: (1) Graphs provide a compact and efficient language to represent multivariate statistical models; and (2) The interpretation of a

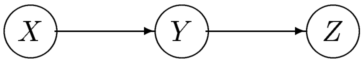


Fig. 1. A simple directed acyclic graph.

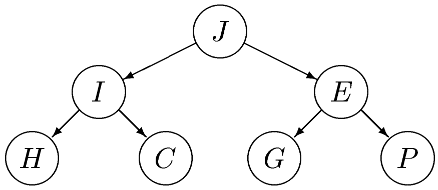
graphical model is given by a Markov property.

3 Graphical models for imprecise probabilities

Probabilities are often stated through *assessments* such as $P(A) \leq 1/2$, $P(B) = 3/10$, $P(A \cup B) \geq 4/5$, for events A and B. One of the recurring problems in probability theory is how to handle a collection of assessments that can be satisfied by more than one probability measure. The answer already articulated by Boole [10] is that the assessments imply probability intervals over events. For example, $P(A) \leq 1/2$ and $P(A \cup B) \geq 4/5$ imply $P(B) \in [3/10, 1]$. Such a formulation has been revisited and refined by many researchers, among which de Finetti [43] and other statisticians of the "de Finettian school" [13], and Hailperin [58] and Nilsson (who gave it the name *probabilistic logic*) [77]. The rules of probabilistic logic have often been depicted through graph fragments [47]; however these graph fragments have functioned only as visual representations. A similar situation has occurred in expert systems like MYCIN [94] or INFERNO [83], where rules requiring manipulation of imprecise beliefs have been often represented graphically, but have not inherited any semantics from the graphical forms.

The development of Bayesian networks during the eighties suggested new ways to combine uncertain reasoning with graphs. Algorithms for inference in polytree-shaped Bayesian networks [79] inspired a number of influential papers on *hierarchical hypothesis spaces*. To understand the idea, consider Figure 2, which presents a piece of medical knowledge discussed by Gordon and Shortliffe [55]. Each node in this figure represents an event that is decomposed into its children nodes. A degree of support can be attributed to any node, indicating how much that node is believed to be true. Gordon and Shortliffe proposed a representation of interval-valued degrees of support based on belief functions, and a mechanism for combining belief functions based on Dempster rule. Shafer, Shenoy and co-workers have developed message-passing algorithms for combination of belief functions in hierarchical hypotheses spaces [89–93], and the framework has been gradually extended in various directions [3,68].

Several other graphical models for imprecise probabilities surfaced around 1990. Fertig and Breese derived approximate inference algorithms for influence diagrams associated with lower bounds on probability values [12,46]. Van der Gaag started from a different mix: instead of directed acyclic graphs



J: Jaundice
I: Intrahepatic
E: Extrahepatic
H: Hepatitis
C: Cirrosis
G: Gall-stone
P: Pancreatic-cancer

Fig. 2. Graphical model discussed by Gordon and Shortliffe [55]; nodes represent events, and leaves are exhaustive and mutually exclusive.

and probability intervals, she adopted undirected models and general linear constraints on probability values [99] — the result of these choices is a linear programming algorithm that can efficiently produce inferences. A different scheme was proposed by Wellman: instead of probability values, one should use qualitative notions such as "occurrence of event A increases the probability of event B" [106]. The result is a *qualitative Bayesian network*; research on this topic remains strong since its inception [8]. Yet another proposal for representation of imprecise probabilistic knowledge is the "order-of-magnitude" approach, where probabilities are represented up to ordinal or infinitesimal values [49,96]. Qualitative and ordinal probabilities also received graphical formulations [40,85], and elicitation procedures that can handle both quantitative and qualitative assessments have also generated steady interest [44,86].

As a short digression, note that during the eighties and nineties many concepts of conditioning for probability intervals and 2-monotone capacities were formulated and discussed in the literature [22,60]. Quite a few of those concepts faced technical and semantic difficulties, and this situation probably contributed to delays in the theory of graphical models for imprecise probability. Clearly, it is difficult to construct a graphical model when the underlying uncertainty calculus cannot properly handle conditioning.

The beginning of the nineties witnessed the first publications explicitly combining general sets of probability measures and directed acyclic graphs [20,97]. At that time the field of robust Bayesian statistics was actively using sets of probability measures to represent perturbations in statistical models [5]. Difficulties that plagued probability intervals and 2-monotone capacities were found not to apply to sets of probability measures, and a rather complete theory of imprecise probabilities, that extensively employed sets of probability measures, was published by Walley in 1991 [102]. These developments led to a marriage between sets of probabilities and directed acyclic graphs that has been strong ever since. The next section discusses the theory of directed acyclic graphs associated with sets of probability measures — structures often referred to as *credal networks*.

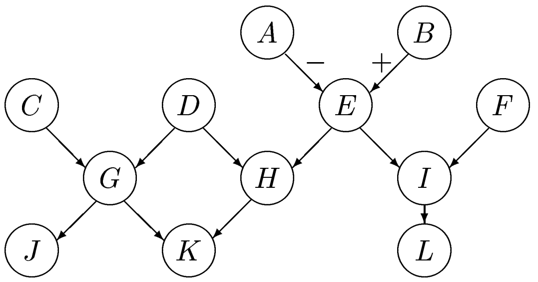
A few additional research efforts deserve mention in this brief historical account, at the risk of missing some relevant contributions. Chrisman [23] has presented a quite original model for undirected graphs associated with probability intervals. Lukasiewicz [73], Thone et al. [98] and Luo et al. [74] have presented graphical models that extend probabilistic logic. Several of those algorithmic developments are discussed in an overview paper by Cano and Moral [18]; their detailed review is quite complementary to the present paper. Relatively little attention has been given to graphical models that incorporate decisions and imprecise probabilities — however there has been recent effort by Danielson et al. [39] to process decision trees and influence diagrams associated with linear constraints on probability values (the DecideIT program has been produced in the course of that research). A related recent development is the construction of classification trees with sets of probabilities [1].

Finally, the class of *imprecise Markov Decision Processes* should be mentioned. An imprecise Markov Decision Process is obtained when the probabilistic requirements on Markov Decision Processes are relaxed: the transition from current to next state is modeled by a set of probability measures or by probability intervals. Work on imprecise Markov Decision Processes started in the seventies [87] and has been revisited a few times since then [9,52,61,107,108]. Up to now there have not been "graphical imprecise Markov Decision Processes" in the literature.

4 Credal networks

A credal network is a graphical model that associates nodes and variables with sets of probability measures. An informal way to convey the content of a credal network is to think about it as a *representation for a set of Bayesian networks over a fixed set of variables.* Note that there is no commitment as to whether one of these Bayesian networks is the "correct" one.

The most obvious motivation for credal networks is to have them as "relaxed" Bayesian networks. In a Bayesian network, the Markov property implies that we must specify a (unique) probability distribution for every variable conditional on any configuration of the variable's parents. This may be a difficult process for several reasons. Existing beliefs may be incomplete or vague, or there may be no resources to gather/process enough information so as to reach a precise probability assessment. Even if experimental data are available, one may not be comfortable with point estimates and may select probability intervals as estimates. It may also be the case that a group of individuals is responsible for specifying probability values, and these individuals cannot agree on precise probability values. Hence we may want to specify a set of probability distributions for every variable conditional on the variable's parents. When we do so, we obtain a credal network.



$$
\begin{aligned}
&P(H = h|D = d, E = e) = 0.2\\
&P(H = h|D = d, E = \neg e) = 0.9\\
&P(H = h|D = \neg d, E = e) = 0.8\\
&P(H = h|D = \neg d, E = \neg e) = 0.8\\
&P(J = j|G = g) = 0.1\\
&P(J = j|G = \neg g) = 0.9\\
&P(K = k|G = g, H = h) = 0.3\\
&P(K = k|G = g, H = \neg h) = 0.8\\
&P(K = k|G = \neg g, H = h) = 0.2\\
&P(K = k|G = \neg g, H = \neg h) = 0.9
\end{aligned}
$$

Fig. 3. An example credal network.

More than just "relaxed" Bayesian networks, credal networks offer a knowledge representation tool. Most people do use probability intervals and qualitative relationships in their dealings with uncertainty; few people can assess probability values up to their third decimal place. Moreover, most people can handle disagreeing sources of probabilistic assessments, even when such a mix does not lead to a single probability measure. People can handle imprecise probabilities; a flexible and general knowledge representation tool for artificial intelligence applications should do just as much.

It is interesting at this point to present examples of credal networks, leaving a more detailed definition of concepts to Section 5. Two artificial examples are discussed in the remainder of this section, so as to illustrate the basic elements and the representational power of credal networks. Readers interested in real applications may consult the work of Antonucci et al [4] for a complex credal network constructed both from expert opinions and data, and the work of Zaffalon et al [115] for a credal network constructed from data and used for classification in a medical scenario.

Consider first the graph in Figure 3. All variables are Boolean; a variable X has values x and $\neg x$. Suppose the network in Figure 3 was created by several experts, reflecting a multitude of views and beliefs.

An expert was hired to establish the probabilities for variables A, B and E. The expert first declared that A was "probable" while B was "between improbable and impossible." Using Renooij's verbal scale [86] as guidance, these verbal statements were translated to $P(A = a) \in [0.75, 0.85]$ and $P(B = b) \in [0.0, 0.15]$. The expert then applied the conventions of qualitative networks to $p(E|A, B)$ [106], as indicated by the plus and minus signs in Figure 3. That is, the expert indicated that

$$
\begin{aligned}
P(E = e|A = a, B = b) &\leq P(E = e|A = \neg a, B = b)\,,\\
P(E = e|A = a, B = \neg b) &\leq P(E = e|A = \neg a, B = \neg b)\,,\\
P(E = e|A = a, B = b) &\geq P(E = e|A = a, B = \neg b)\,,\\
P(E = e|A = \neg a, B = b) &\geq P(E = e|A = \neg a, B = \neg b)\,.
\end{aligned}
$$

The expert could also state the precise assessment $P(E = e|A = a, B = b) = 0.4$. The largest set of functions $P(E|A, B)$ that satisfy these qualitative and numeric assessments has seven vertices. Each vertex is specified by a triplet containing values $P(E = e|A = a, B = \neg b)$, $P(E = e|A = \neg a, B = b)$, and $P(E = e|A = \neg a, B = \neg b)$; the seven vertices are given by the following triplets: $[0, 0.4, 0]$; $[0, 1, 0]$; $[0.4, 1, 0.4]$; $[0, 0.4, 0.4]$; $[0.4, 0.4, 0.4]$; $[0, 1, 1]$; $[0.4, 1, 1]$.

A second expert was hired to examine variables F, I and L. The expert assessed $P(F = f) = 0.2$. The expert then took a Noisy-OR function [79] to model $p(I|E, F)$, with "link" probabilities $P(I = i|E = e, F = \neg f) = 0.9$ and $P(I = i|E = \neg e, F = f) = 0.8$. The expert decided to have a "leak" probability, but could not assess its value precisely, and adopted an *interval* leak probability of [0.1,0.2]. To assess $p(L|I)$, the expert consulted a database with experiments, but she was unsure about priors for the estimates, and took an Imprecise Dirichlet Model [6,104,113] over them, producing interval estimates for $P(L = l|I = i)$ and $P(L = l|I = \neg i)$. The expert obtained $P(L = l|I = i) \in [0.5, 0.6]$ and $P(L = l|I = \neg i) \in [0.4, 0.5]$.

A group of three experts was then hired to model the remaining variables. The experts used a large database to obtain precise estimates for $P(H|D, E)$, $P(J|G)$, and $P(K|G, H)$, shown in Figure 3. No data was available for C, D, and G. After much discussion, the experts produced assessments $P(C = c) \geq 0.4$, $P(C = \neg c) \geq 0.5$, $P(D = d) \geq 0.2$, and $P(D = \neg d) \geq 0.7$. The experts did not agree at all on $P(G|C, D)$; indicating first, second and third opinions as vectors (all in the same order), the opinions of the experts were:

$$\begin{aligned}
P(G = g|C = c, D = d) &= [0.2, 0.3, 0.4], \\
P(G = g|C = c, D = \neg d) &= [0.7, 0.6, 0.5], \\
P(G = g|C = \neg c, D = d) &= [1, 1, 1], \\
P(G = g|C = \neg c, D = \neg d) &= [0.8, 0.9, 0.8].
\end{aligned}$$

The experts recommended that, for every possible combination of probability values within the elicited bounds and sets, the joint distribution should be produced using the Markov condition in the graph of Figure 3. Consider a few inferences with this network (an *inference* here is the computation of a tight interval containing all possible values for the probability of an event).[1] For example, $P(D = d|A = \neg a, F = f, K = k) \in [0.17, 0.45]$. Note that inferences do not assume more information than available in the model, but they do yield

[1] Inferences were computed with the JavaBayes system version 0.347, available under GPL from http://www.pmr.poli.usp.br/ltd/Software/javabayes. This system contains a naive enumeration algorithm for manipulation of sets of probabilities.

valuable information — we can say that $P(D = d|A = \neg a, F = f, K = k)$ is smaller than 0.5 if we need to make a decision concerning this value.

The important point in this example is that the credal network summarizes a large variety of assessments, translating different kinds of beliefs into a uniform and understandable language. Qualitative, verbal, empirical and subjective information are all organized into a single structure.

Consider a second example, taken from Cozman et al. [34]. The example is based on Jaeger's version of the "Holmes problem," a situation that mixes first-order logic constructs with probabilities [66]. The story is this. If a person v lives in LA, then she may (probabilistically) sound the alarm, depending on whether there is a burglary and whether there is an earthquake. If v does not live in LA, then she may (probabilistically) sound the alarm in case there is a burglary. Here v is a universally quantified variable in a domain $\mathcal{V}$, and the relations `alarm`(v), `lives-in`(v,LA), `burglary`(v) and `earthquake`(LA) describe v's situation. To simplify the notation, denote `alarm`(v) by A_v, `lives-in`(v,LA) by L_v, `burglary`(v) by B_v and `earthquake`(LA) by E. For each relation Y, either y or $\neg y$ holds.

Jaeger presents a model for the "Holmes problem" that is based on Bayesian networks [66]. Jaeger uses the first-order description of the "Holmes problem" to build a Bayesian network for any given domain — that is, given a domain, Jaeger's method produces a Bayesian network. Take a domain $\mathcal{V}_H$ containing G and H and such that l_G holds; in this case Jaeger's method constructs the graph in Figure 4. To do so, Jaeger's method assumes that

(1) all probability values are precisely known;
(2) $P(a_v|l_v, B_v, E)$ is a Noisy-OR function of B_v and E;
(3) $P(a_v|\neg l_v, B_v, E)$ is independent of E.

This strategy is clearly attractive; however it fails if there is imprecision in probability values, or if there is disagreement about how to define the distribution $p(A_v|L_v, B_v, E)$.[2] In case these assumptions fail, a credal network can be used without difficulties.

Take Figure 4 and consider the following assessments, which attempt to translate the rather vague scenario of the "Holmes problem":

(1) $P(e) \in [0.01, 0.1]$.
(2) $P(b_v) \in [0.001, 0.01]$ for any v in the domain.
(3) $P(l_v) \in [0.05, 0.15]$ for any v in the domain.

[2] The automatic resort to Noisy-OR functions is somewhat artificial, and it is characteristic of methods that produce a single Bayesian network out of logical and probabilistic constructs [50,53,65,78,81].

(4) $P(a_v|l_v, b_v, \neg e) = 0.9$,
$P(a_v|l_v, \neg b_v, e) = 0.2$,
and $P(a_v|l_v, b_v, e) \geq 0.9$:
that is, alarm with burglary and earthquake is more probable than alarm with just burglary when v lives in LA.

(5) $P(a_v|l_v, \neg b_v, \neg e) \in [0.0, 0.1]$:
that is, there is a "leak" probability between 0 and 0.1 that the alarm sounds even with no burglary and no earthquake when v lives in LA.

(6) $P(a_v|\neg l_v, b_v, e) = 0.9$,
$P(a_v|\neg l_v, b_v, \neg e) = 0.9$,
$P(a_v|\neg l_v, \neg b_v, e) = 0.0$:
and $P(a_v|\neg l_v, \neg b_v, \neg e) = 0.0$:
that is, probabilities are precise and do not depend on E when v does not live in LA.

Suppose we take every possible combination of probability values within the indicated bounds, and produce joint distributions using the Markov condition in the graph of Figure 4. We obtain $P(a_H) \in [0.0001, 0.0253]$, $P(a_H|e) \in [0.0108, 0.0388]$, $P(a_G) \in [0.0029, 0.1179]$ and $P(a_G|e) \in [0.2007, 0.2080]$. Note that inferences produce rather small intervals; even though only a few assessments are used to build the credal network, the structural assumptions represented by the graph greatly constrain probability values.

In this example a credal network is used to mix logical statements with flexible probabilistic assessments. An alternative way to combine logical and probabilistic assessments would be to employ a probabilistic logic [47,58,59,72,77]. However, it would be important to state assessments of independence as well — without such assessments we cannot give meaning to the beautifully concise graphical representation in Figure 4. We would then face the difficulty that inference in general probabilistic logic with independence assessments is likely to be higher than the complexity of independence-free probabilistic logic — and the latter is already quite disheartening [72]. How could we then have a flexible and compact language to represent logical, probabilistic, and independence assessments? In short, *we need a language that is more flexible than traditional probability theory, but more structured than general probabilistic logic.* Credal networks offer one such balancing act.

5 The theory of credal sets and credal networks

Before we analyze in further detail the properties of credal networks, we should discuss a few definitions. A set of probability measures is called a *credal set* (from Levi's *credal states* [71]). There is some debate on whether credal sets

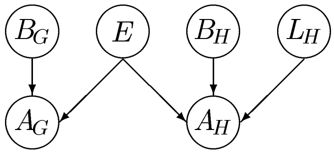


Fig. 4. The network for the "Holmes problem" with domain $V_H = \{G, H\}$ and assuming l_G holds.

should be closed or open [88,102], and whether credal sets should be convex [3] or not [69,71]. In this paper a credal set can be closed or open, convex or non-convex.

A credal set defined by probability distributions $p(X)$ is denoted by $K(X)$. Given a credal set $K(X)$ and a bounded function $f(X)$, the *upper* and *lower* expectations of $f(X)$ are defined respectively as $\overline{E}[f(X)] = \sup_{p(X)\in K(X)} E_p[f(X)]$ and $\underline{E}[f(X)] = \inf_{p(X)\in K(X)} E_p[f(X)]$, where $E_p[f(X)]$ indicates standard expectation. *Upper* and *lower probabilities* are defined similarly. A lower expectation can be viewed as an assessment of the form $E_p[f(X)] \geq \underline{E}[f(X)]$ (that is, a linear constraint on the space of probability distributions for X). A credal set K and its convex hull produce the *same* lower/upper expectations and lower/upper probabilities.

The most commonly adopted scheme for conditioning in credal sets is elementwise Bayes rule (that is, conditioning is obtained by applying Bayes rule to each element of a credal set). [4] Such an intuitive prescription, called the *generalized Bayes rule* by Walley [102,103], can be justified axiomatically in various ways [51,56,102]. Note that if $K(X)$ is convex, then $K(X|A)$ is convex as well [71].

There are two different ways to represent conditioning with respect to random variables. First, consider the collection of *separately specified* conditional credal sets $\{K(X|Y = y) : y \text{ is a value of } Y\}$. We denote this collection of credal sets by $K(X|Y)$. Second, consider the direct specification of a set of functions $p(X|Y)$; call such a set an *extensive* conditional credal set, and denote it by $L(X|Y)$. It should be clear that $K(X|Y)$ and $L(X|Y)$ are quite different objects; the first is a set of credal sets, and the second is a single set of functions. [5] In the first example in Section 4, $K(L|I)$ is separately specified

[3] A credal set is convex if, for any measures P_1 and P_2 in the set, the measure $\alpha P_1 + (1 - \alpha)P_2$ is in the set for any $\alpha \in [0, 1]$.

[4] Alternative schemes are discussed by Chrisman [22], Moral [75] and de Cooman and Zaffalon [42].

[5] A note on terminology: Moral and Cano use the term *conditioned to the values* to indicate separately specified credal sets (the latter term is used by Walley [102]

while $L(E|A, B)$ is extensively specified.

A credal network consists of a directed acyclic graph, where each node in the graph is associated with a random variable, and where each variable X is associated with conditional credal sets $K(X|\text{pa}(X))$ or $L(X|\text{pa}(X))$. The goal is to combine these "local" credal sets into a set of joint distributions satisfying a Markov condition on the graph. [6] To accomplish this, it is necessary to define: (1) what is the Markov condition on the graph; (2) how to combine the local credal sets. The remainder of this section discusses these points.

We start by discussing independence concepts in the context of credal sets. There are at least two possibilities (to simplify the discussion, we assume that every event has positive lower probability):

- First, we may require factorization for all distributions in the credal set (that is, $p(X, Y) = p(X)\,p(Y)$ for all distributions). Such a requirement implies non-convexity of the credal set. To remain agnostic with respect to convexity, we may require factorization just for the *vertices* of the credal set. We then say that X and Y are *strongly independent*. Strong independence is the most commonly adopted concept in the literature; variants of this concept were already implicit in Huber's work on frequentist robustness [64] and in the first proposals for credal networks [20,97].
- Second, we may require irrelevance of conditioning: We say that Y is *epistemically irrelevant* to X if $\underline{E}[f(X)|Y = y] = \underline{E}[f(X)]$ for any bounded $f(X)$ and any y. It turns out that epistemic irrelevance is not symmetric — Y may be epistemically irrelevant to X while X is not epistemically irrelevant to Y [102]. We say that X and Y are *epistemically independent* if X is epistemically irrelevant to Y and Y is epistemically irrelevant to X [102].

The literature contains a number of alternative definitions for independence [26,33,41]; even though some unification of concepts may be possible [32,76], it seems unlikely that a single concept will become prevalent in all applications of credal sets. Note that such concepts are often not equivalent; for example, strong independence implies epistemic independence, but not the converse.

There are several ways to define *conditional independence* as well [76]. Say that X and Y are *strongly independent conditional* on Z if the vertices of the credal set $K(X, Y|Z = z)$ factorize for every z. And say that X and Y are *epistemi-*

and Cozman [31]), and use *conditional set* and *conditional information* to indicate extensive conditional credal sets [76].

[6] Andersen and Hooker have discussed credal networks that combine local and non-local credal sets [2]; the discussion here is restricted to networks containing only local sets. The advantage of using only local sets is that it is always possible to satisfy the Markov condition discussed later; non-local sets may be inconsistent with the Markov condition.

cally independent conditional on Z if $\underline{E}[f(X)|Y = y, Z = z] = \underline{E}[f(X)|Z = z]$ for any bounded function $f(X)$ and any (y, z), and $\underline{E}[g(Y)|X = x, Z = z] = \underline{E}[g(Y)|Z = z]$ for any bounded function $g(Y)$ and any (x, z).

Suppose we have a directed acyclic graph, random variables $X_1, \ldots, X_n$, "local" credal sets (either separately or extensively specified), and we have settled on a concept of independence. It seems reasonable to assume that every variable X_i associated with the graph is independent of its nondescendant nonparents given its parents. This is the *Markov condition for credal networks* — note that the condition depends on the adopted concept of independence. We then have all ingredients of a credal network: a graph, variables, credal sets, and a Markov condition. Now we have to decide how to build a set of *joint* distributions over $X_1, \ldots, X_n$ out of these ingredients.

In general, there may be several sets of joint distributions that are consistent with a given collection of marginal and conditional credal sets. Any one of these sets of joint distributions is called an *extension* of the marginal and conditional credal sets. Usually one is interested in the largest possible extension for a given set of assessments (Walley refers to these extensions as "natural" ones [102]). For a credal network, we might consider the largest extension satisfying the Markov condition as the "natural semantics" for the network. So we have:

(1) the largest extension that satisfies the Markov condition with respect to strong independence — the *strong extension*;
(2) the largest extension that satisfies the Markov condition with respect to epistemic independence — the *epistemic extension*.

Other extensions could be generated for a credal network, but the strong and the epistemic extensions are the only ones that have received systematic attention so far in the literature. [7] Strong extensions were already implicit in the first proposals for credal networks [20,97] and have received considerable attention [2,21,30,45,109]. Comparatively few results are known concerning epistemic extensions [31,35].

Given a credal network with local separately specified credal sets $K(X_i|\mathrm{pa}(X_i))$, the strong extension of the network is the convex hull of the set containing all joint distributions that factorize as

$$\prod_i p(X_i|\mathrm{pa}(X_i)) \,, \qquad (1)$$

where the conditional distributions $p(X_i|\mathrm{pa}(X_i) = \pi_k)$ are selected from the

[7] A note on terminology: Couso et al. employ the term *independence in the selection* to refer to strong independence and reserve *strong independence* to the more specific case of strong extensions [26].

local credal sets $K(X_i|\mathrm{pa}(X_i) = \pi_k)$ [32]. If present, extensive conditional credal sets $L(X_i|\mathrm{pa}(X_i))$ can be used in Expression (1): Instead of selecting $p(X_i|\mathrm{pa}(X_i) = \pi_k)$ from $K(X_i|\mathrm{pa}(X_i) = \pi_k)$, one would then select the function $p(X_i|\mathrm{pa}(X_i))$ from the set of functions $L(X_i|\mathrm{pa}(X_i))$.

The examples discussed in Section 4 showed inferences computed with respect to strong extensions of the networks.

Strong extensions are appropriate in a variety of contexts. In particular, strong extensions are appropriate when one views a credal set as a black-box containing the "true" probability measure [54]. Under this "sensitivity analysis" interpretation, there is a Bayesian network hidden "inside" a credal network, and it makes sense to restrict our attention to joint distributions that satisfy the Markov property for standard stochastic independence. Strong extensions are also appropriate when several experts specifying a credal network disagree about probability values but agree that every acceptable joint distribution should satisfy standard stochastic independence (first example in Section 4). An appealing property of strong extensions is that they display the same separation properties of standard Bayesian networks (that is, *d-separation* holds in strong extensions) [31]. A final argument in favor of strong extensions is that they are not exclusively dependent on strong independence; it is possible to generate strong extensions from conditions involving only epistemic independence [32].

6 Inference and learning with strong extensions

Given a credal network, one may be interested in lower/upper probabilities or expectations, or one may be interested in particular values of variables that "dominate" other values according to some criterion. Often the term *inference* is used to indicate the computation of lower/upper probabilities in some extension. In this case, if X_q is a *query* variable and $\mathbf{X}_E$ represents a set of *observed* variables, the inference is the computation of tight bounds for $p(X_q|\mathbf{X}_E)$ for one or more values of X_q.

For inferences with strong extensions, it is known that the distributions that minimize/maximize $p(X_q|\mathbf{X}_E)$ belong to the set of vertices of the extension [45]. [8] The difficulty faced by inference algorithms is the potentially enormous number of vertices that a strong extension may have, even for small networks.

[8] This is valid when unobserved variables are assumed to be *missing at random*; if the "missingness mechanism" is entirely unknown, comparisons between probabilities can often be produced efficiently even for densely connected credal networks [42].

For example, take the network in Figure 1, and assume that X, Y and Z have four values each, with separately specified credal sets with four vertices each. There are then 4^9 potential vertices of the strong extension (that is, 4^9 different distributions following Expression (1)). The only credal networks that are amenable to efficient exact inferences are polytree-shaped networks with binary variables [45]. Other networks, even polytree-shaped ones, face tremendous computational challenges [36]. Exact inference algorithms typically examine potential vertices of the strong extension to produce the required lower/upper values [15,21,31,36,37]. Approximate inference algorithms can produce either *outer* or *inner* approximations: the former produce intervals that enclose the correct probability interval between lower and upper probabilities [19,38,57,97], while the latter produce intervals that are enclosed by the correct probability interval [2,15,16,30]. Some of these algorithms emphasize enumeration of vertices, while others resort to optimization techniques (as computation of lower/upper values for $p(X_q|\mathbf{X}_E)$ is equivalent to minimization/maximization of a fraction containing polynomials in probability values). Rather detailed overviews of inference algorithms for imprecise probabilities have been published by Cano and Moral [17,18].

Procedures that "learn" credal networks from data have been investigated in the last decade. A few scenarios have been explored in the literature. The simplest situation is to learn a credal network for a given graph, using a complete database of categorical variables and Imprecise Dirichlet priors [29,115]. Another scenario is to learn a credal network for a given graph, categorical variables, and missing data [84,110,112]. Finally, the most complex situation is to learn the graph *and* the local credal sets from complete or missing data [14,110,114] — there is a great lack of methods that learn general graphical structures directly from data. Most of the effort in learning credal networks has been so far directed to *credal classification* (that is, to use a credal network for classification); in fact, some of the most successful applications of credal networks have appeared in the context of credal classification [84,113]. Credal classification differs from Bayesian classification in that an observation may be labeled with a *set* of classes — all classes that are not dominated by any other classes with respect to posterior probability. Thus credal classification requires more than just computation of minima and maxima of probabilities; credal classification requires that undominated classes be identified through comparisons between classes. Classification is an important topic for application of credal networks, with a huge potential and still a great number of challenges.

7 Conclusion

The purpose of this paper was to present a broad overview of graphical models for imprecise probabilities and a more detailed discussion of credal networks. Such graphical models are quite flexible in their representational power and yet are quite compact and easy to construct. The theory of credal networks has received considerable attention and is now reaching a reasonable degree of maturity.

Given the breadth of the field, this overview is certain to have missed relevant work on graphical models and imprecise probabilities. It is hoped that such omissions are not many and can be forgiven. In any case, two topics should at least be mentioned: graphical models for possibility measures, and graphical models that handle zero lower probabilities. Both topics are important and raise a large number of conceptual and technical problems; they were omitted for lack of space. A vast literature on graphical models for possibility measures can be consulted [48]. With regard to zero lower probabilities [25,102,105], there has been recent work on this subject, and new concepts and algorithms have been proposed to handle such situations [100,101].

Many types of graphical models have been little explored, particularly models that deal with decision-making. There have also been few available software packages for manipulation of graphical models with imprecise probabilities. There is certainly no shortage of potential contributions to be made regarding graphical models for imprecise probabilities.

Acknowledgments

I would like to thank the editor Marco Zaffalon for many useful suggestions and for incredible patience when waiting for versions of this paper; in fact I regret that Marco could not be a co-author of this paper. Marco also suggested the informal way of thinking about credal networks discussed at the beginning of Section 4. I would also like to acknowledge valuable contributions by the reviewers and by José Carlos F. da Rocha.

This work was (partially) developed in collaboration with HP Brazil R&D; the author has been partially supported by CNPq through grant 3000183/98-4.

References

[1] J. Abellán and S. Moral. Maximum entropy in credal classification. *Third Int. Symp. on Imprecise Probabilities and Their Applications*, pages 1–15, Lugano, Switzerland, 2003. Carleton Scientific.

[2] K. A. Andersen and J. N. Hooker. Bayesian logic. *Decision Support Systems*, 11:191–210, 1994.

[3] B. Anrig, R. Bissig, R. Haenni, J. Kohlas, and N. Lehmann. Probabilistic argumentation systems: Introduction to assumption-based modeling with ABEL. Technical Report 99-1, University of Fribourg, 1999.

[4] A. Antonucci, Andrea Salvetti, and Marco Zaffalon Assessing debris flow hazard by credal nets. *Conference on Soft Methods in Probability and Statistics*, to appear, 2004.

[5] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer-Verlag, 1985.

[6] J.-M. Bernard. Implicative analyis for multivariate data using an imprecise Dirichlet model. *J. Statistical Planning and Inference*, 105:83–103, 2002.

[7] B. Bollobás. *Modern Graph Theory.* Springer-Verlag, 1998.

[8] J. H. Bolt, S. Renooij, and L. C. van der Gaag. Upgrading ambiguous signs in QPNs. *Conf. on Uncertainty in Artificial Intelligence*, pages 73–80, San Francisco, California, 2003. Morgan Kaufmann.

[9] B. Bonet and J. Pearl. Qualitative MDPs and POMDPs: An order-of-magnitude approximation. *Conf. on Uncertainty in Artificial Intelligence*, pages 61–68, San Francisco, California, 2002. Morgan Kaufmann.

[10] G. Boole. *The Laws of Thought.* Dover edition, 1958.

[11] C. Boutilier, S. Hanks, and T. Dean. Decision-theoretic planning: Structural assumptions and computational leverage. *J. of Artificial Intelligence Research*, 11:1–94, 1999.

[12] J. S. Breese and K. W. Fertig. Decision making with interval influence diagrams. *Uncertainty in Artificial Intelligence 6*, pages 467–478. Elsevier Science, North-Holland, 1991.

[13] G. Bruno and A. Gilio. Applicazione del metodo del simplesso al teorema fondamentale per le probabilità nella concezione soggettivistica. *Statistica*, 40:337–344, 1980.

[14] L. M. Campos and J. F. Huete. Learning non probabilistic belief networks. *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, volume 747 of *Lecture Notes in Computer Science*, pages 57–64. Springer-Verlag, 1993.

[15] A. Cano, J. E. Cano, and S. Moral. Convex sets of probabilities propagation by simulated annealing. *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 4–8, Paris, France, 1994.

[16] A. Cano and S. Moral. A genetic algorithm to approximate convex sets of probabilities. *Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 2:859–864, 1996.

[17] A. Cano and S. Moral. A review of propagation algorithms for imprecise probabilities. *First Int. Symp. on Imprecise Probabilities and Their Applications*, pages 51–60, Ghent, Belgium, 1999. Imprecise Probabilities Project, Universiteit Ghent.

[18] A. Cano and S. Moral. *Algorithms for imprecise probabilities*, In *Handbook of Defeasible Reasoning and Uncertainty Management Systems (vol. 5)*, pages 369–420. Kluwer Academic Publishers, 2000.

[19] A. Cano and S. Moral. Using probability trees to compute marginals with imprecise probabilities. *Int. J. of Approximate Reasoning*, 29:1–46, 2002.

[20] J. Cano, M. Delgado, and S. Moral. Propagation of uncertainty in dependence graphs. *European Conf. on Symbolic and Quantitative Approaches to Uncertainty*, pages 42–47, 1991.

[21] J. Cano, M. Delgado, and S. Moral. An axiomatic framework for propagating uncertainty in directed acyclic networks. *Int. J. of Approximate Reasoning*, 8:253–280, 1993.

[22] L. Chrisman. Incremental conditioning of lower and upper probabilities. *Int. J. of Approximate Reasoning*, 13(1):1–25, 1995.

[23] L. Chrisman. Propagation of 2-monotone lower probabilities on an undirected graph. *Conf. on Uncertainty in Artificial Intelligence*, pages 178–186, San Francisco, California, United States, 1996. Morgan Kaufmann.

[24] R. T. Clemen. *Making Hard Decisions: An Introduction to Decision Analysis.* Duxbury Press, 1997.

[25] G. Coletti and R. Scozzafava. Stochastic independence in a coherent setting. *Annals of Mathematics and Artificial Intelligence*, 35:151–176, 2002.

[26] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5:165–181, 2000.

[27] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Springer-Verlag, New York, 1999.

[28] D. R Cox and N. Wermuth. *Multivariate Dependencies: Models, Analysis, and Interpretation.* Chapman & Hall/CRC Press, Boca Raton, Florida, 1996.

[29] F. G. Cozman and L. Chrisman. Learning convex sets of probability from data. Technical Report CMU-RI-TR 97-25, Robotics Institute, Carnegie Mellon University, Pittsburgh, June 1997.

[30] F. G. Cozman. Robustness analysis of Bayesian networks with local convex sets of distributions. *Conf. on Uncertainty in Artificial Intelligence*, pages 108–115, San Francisco, 1997. Morgan Kaufmann.

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

[32] F. G. Cozman. Separation properties of sets of probabilities. *Conf. on Uncertainty in Artificial Intelligence*, pages 107–115, San Francisco, 2000. Morgan Kaufmann.

[33] F. G. Cozman. Computing lower expectations with Kuznetsov's independence condition. *Third Int. Symp. on Imprecise Probabilities and Their Applications*, pages 177–187, Lugano, Switzerland, 2003. Carleton Scientific.

[34] F. G. Cozman, C. P. de Campos, J. S. Ide and J. C. F. da Rocha. Propositional and relational Bayesian networks associated with imprecise and qualitative probabilistic assessments. *Conf. on Uncertainty in Artificial Intelligence*, to appear, 2004.

[35] F. G. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 112–121, Ithaca, New York, 2001.

[36] J. C. F. da Rocha and F. G. Cozman. Inference with separately specified sets of probabilities in credal networks. *Conf. on Uncertainty in Artificial Intelligence*, pages 430–437, San Francisco, California, 2002. Morgan Kaufmann.

[37] J. C. F. da Rocha and F. G. Cozman. Inference in credal networks with branch-and-bound algorithms. *Third Int. Symp. on Imprecise Probabilities and Their Applications*, pages 482–495. Carleton Scientific, 2003.

[38] J. C. F. da Rocha, F. G. Cozman and C. P. de Campos. Inference in polytrees with sets of probabilities. *Conf. on Uncertainty in Artificial Intelligence*, pages 217–224, San Francisco, California, United States, 2003. Morgan Kaufmann.

[39] M. Danielson, L. Ekenberg, J. Johansson, and A. Larsson. The DecideIT decision tool. *Third Int. Symp. on Imprecise Probabilities and Their Applications*, pages 204–217, Lugano, Switzerland, 2003. Carleton Scientific.

[40] A. Darwiche and M. Goldszmidt. On the relation between kappa calculus and probabilistic reasoning. *Conf. on Uncertainty in Artificial Intelligence*, pages 145–153, San Francisco, California, 1994. Morgan Kaufmann.

[41] L. de Campos and S. Moral. Independence concepts for convex sets of probabilities. *Conf. on Uncertainty in Artificial Intelligence*, pages 108–115, San Francisco, California, United States, 1995. Morgan Kaufmann.

[42] G. de Cooman and M. Zaffalon. Updating beliefs with incomplete observations. *Artificial Intelligence*, to appear.

[43] B. de Finetti. *Theory of probability, vol. 1-2.* Wiley, New York, 1974.

[44] M. J. Druzdzel and L. C. van der Gaag. Elicitation of probabilities for belief networks: Combining qualitative and quantitative information. *Conf. on Uncertainty in Artificial Intelligence*, pages 141–148, San Francisco, California, 1995. Morgan Kaufmann.

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

[46] K. W. Fertig and J. S. Breese. Interval influence diagrams. *Uncertainty in Artificial Intelligence 5*, pages 149–161. Elsevier Science Publishers, North-Holland, 1990.

[47] A. M. Frisch and P. Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69:93–122, 1994.

[48] J. Gebhardt and R. Kruse. Learning possibilistic networks from data. *Fifth Int. Workshop on Artificial Intelligence and Statistics*, pages 233–243, Fort Lauderdale, Florida, United States, 1995.

[49] H. Geffner. *Default Reasoning: Causal and Conditional Theories.* MIT Press, 1992.

[50] L. Getoor, N. Friedman, D. Koller, B. Taskar. Learning probabilistic models of relational structure. In *Int. Conf. on Machine Learning*, pp. 170–177, 2001.

[51] F. J. Giron and S. Rios. Quasi-Bayesian behaviour: A more realistic approach to decision making? *Bayesian Statistics*, pages 17–38. University Press, Valencia, Spain, 1980.

[52] R. Givan, S. M. Leach, and T. Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 122(1-2):71–109, 2000.

[53] S. Glesner, D. Koller. Constructing flexible dynamic belief networks from first-order probalistic knowledge bases. In *European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pp. 217–226, 1995.

[54] I. J. Good. *Good Thinking: The Foundations of Probability and its Applications.* University of Minnesota Press, Minneapolis, 1983.

[55] J. Gordon and E. H. Shortliffe. *The Dempster-Shafer theory of evidence*, In E. H. Shortliffe and B. G. Buchanan, *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project*, Chapter 13, pages 272–292. Addison-Wesley, 1985.

[56] A. J. Grove and J. Y. Halpern. Updating sets of probabilities. *Conf. on Uncertainty in Artificial Intelligence*, pages 173–182, San Francisco, California, 2002. Morgan Kaufmann.

[57] V. A. Ha, A. Doan, V. H. Vu, and P. Haddawy. Geometric foundations for interval-based probabilities. *Annals of Mathematics and Artificial Intelligence*, 24(1-4):1–21, 1998.

[58] T. Hailperin. *Sentential Probability Logic.* Lehigh University Press, Bethlehem, United States, 1996.

[59] R. Fagin, J. Y. Halpern, N. Megiddo. A logic for reasoning about probabilities. *Information and Computation*, 87:78–128, 1990.

[60] J. Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[61] D. Harmanec. A generalization of the concept of Markov decision process to imprecise probabilities. *First Int. Symp. on Imprecise Probabilities and Their Applications*, pages 175–182, Ghent, Belgium, 1999. Imprecise Probabilities Project.

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

[63] R. A. Howard and J. E. Matheson. Influence diagrams. *Readings on the Principles and Applications of Decision Analysis (vol. II)*, pages 719–762. Strategic Decisions Group, Menlo Park, California, 1984.

[64] P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101, 1964.

[65] M. Jaeger. Relational Bayesian networks. In *Conf. on Uncertainty in Artificial Intelligence*, pp. 266–273, San Francisco, 1997.

[66] M. Jaeger. The *Primula* system: user's guide. Available at http://www.cs.auc.dk/˜jaeger/Primula.

[67] R. Kindermann and J. L. Snell. *Markov Random Fields and Their Applications.* AMS, Providence, Rhode Island, 1980.

[68] J. Kohlas and P. P. Shenoy. *Computation in Valuation Algebras*, In *Handbook of Defeasible Reasoning and Uncertainty Management Systems (vol. 5)*, pages 5–40. Kluwer, Dordrecht, 2000.

[69] H. E. Kyburg and M. Pittarelli. Set-based Bayesianism. *IEEE Transactions on Systems, Man and Cybernetics A*, 26(3):324–339, 1996.

[70] S. L. Lauritzen and N. Wermuth. Graphical models for associations between variables, some of which are qualitative and some quantitative. *Annals of Statistics*, 17:31–57, 1989.

[71] I. Levi. *The Enterprise of Knowledge.* MIT Press, 1980.

[72] T. Lukasiewicz. Probabilistic logic programming. *European Conf. on Artificial Intelligence*, pp. 388-392, 1998.

[73] T. Lukasiewicz. Probabilistic deduction with conditional constraints over basic events. *J. of Artificial Intelligence Research*, 10:199–241, 1999.

[74] C. Luo, C. Yu, J. Lobo, G. Wang, and T. Pham. Computation of best bounds of probabilities from uncertain data. *Computational Intelligence*, 12(4):541–566, 1996.

[75] S. Moral. Calculating uncertainty intervals from conditional convex sets of probabilities. *Uncertainty in Artificial Intelligence 8*, pages 199–206, 1992.

[76] S. Moral and A. Cano. Strong conditional independence for credal sets. *Annals of Mathematics and Artificial Intelligence*, 35:295–321, 2002.

[77] N. J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

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

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

[80] J. Pearl. *Causality: models, reasoning, and inference.* Cambridge University Press, Cambridge, United Kingdom, 2000.

[81] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64:81–129, 1993.

[82] M. L. Puterman. *Markov Decision Processes.* John Wiley and Sons, 1994.

[83] J. R. Quinlan. INFERNO, a cautious approach to uncertain inference. *The Computer Journal*, 26(3):255–269, 1983.

[84] M. Ramoni and P. Sebastiani. Robust Bayes classifiers. *Artificial Intelligence*, 125(1-2):209–226, 2001.

[85] S. Renooij, S. Parsons, and P. Pardieck. Using kappas as indicators of strength in qualitative probabilistic networks. *European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 87–99. Springer Verlag, 2003.

[86] S. Renooij and C. L. M. Witteman. Talking probabilities: communicating probabilistic information with words and numbers. *Int. J. of Approximate Reasoning*, 22:169–194, 1999.

[87] J. K. Satia and R. E. Lave Jr. Markovian decision processes with uncertain transition probabilities. *Operations Research*, 21:728–740, 1970.

[88] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. A representation of partially ordered preferences. *Annals of Statistics*, 23(6):2168–2217, 1995.

[89] G. Shafer and R. Logan. Implementing Dempster's rule for hierarchical evidence. *Artificial Intelligence*, 33:271–298, 1987.

[90] G. Shafer, P. P. Shenoy, and K. Mellouli. Propagating belief functions in qualitative Markov trees. *Int. J. of Approximate Reasoning*, 1:349–400, 1987.

[91] P. P. Shenoy. A valuation-based language for expert systems. *Int. J. of Approximate Reasoning*, 3:383–411, 1989.

[92] P. P. Shenoy and G. Shafer. Propagating belief functions with local computations. *IEEE Expert*, 1(3):43–52, 1986.

[93] P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. *Uncertainty in Artificial Intelligence 4*, pages 169–198. Elsevier Science Publishers, North-Holland, 1990.

[94] E. H. Shortliffe and B. G. Buchanan. *Rule-based expert systems: the MYCIN experiments of the Stanford Heuristic Programming Project.* The Addison-Wesley series in artificial intelligence. Addison-Wesley, Reading, Mass., 1985.

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

[96] W. Spohn. *Ordinal conditional functions: a dynamic theory of epistemic states*, In W. L. Harper and B. Skyrms, *Causation in Decision: Belief, Change, and Statistics (vol. II)*, pages 105–134. Kluwer Academic Publishers, 1988.

[97] B. Tessem. Interval probability propagation. *Int. J. of Approximate Reasoning*, 7:95–120, 1992.

[98] H. Thone, U. Guntzer, and W. Kieβling. Increased robustness of Bayesian networks through probability intervals. *Int. J. of Approximate Reasoning*, 17:37–46, 1997.

[99] L. C. van der Gaag. Computing probability intervals under independency constraints. *Uncertainty in Artificial Intelligence 6*, pages 457–466. Elsevier Science, 1991.

[100] B. Vantaggi. Graphical models for conditional independence structures. *Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 332–341. Shaker, 2001.

[101] B. Vantaggi. Graphical representation of asymmetric graphoid structures. *Third Int. Symp. on Imprecise Probabilities and Their Applications*, pages 560–574. Carleton Scientific, 2003.

[102] P. Walley. *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London, 1991.

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

[104] P. Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal Royal Statistical Society B*, 58(1):3–57, 1996.

[105] K. Weichselberger. The theory of interval-probability as a unifying concept for uncertainty. *Int. J. of Approximate Reasoning*, 24(2-3):149–170, 2000.

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

[107] C. C. White III and H. K. El-Deib. Parameter imprecision in finite state, finite action dynamic programs. *Operations Research*, 34(1):120–129, 1986.

[108] C. C. White III and H. K. El-Deib. Markov decision processes with imprecise transition probabilities. *Operations Research*, 42(4):739–749, 1994.

[109] M. Zaffalon. *Inferenze e Decisioni in Condizioni di Incertezza con Modelli Grafici Orientati.* PhD thesis, Università di Milano, February 1997.

[110] M. Zaffalon. Robust discovery of tree-dependency structures. In *Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 394–403, Netherlands, 2001. Shaker.

[111] M. Zaffalon. Statistical inference of the naive credal classifier. In *Second Int. Symp. on Imprecise Probabilities and Their Applications*, pages 384–393, Netherlands, 2001. Shaker.

[112] M. Zaffalon. Exact credal treatment of missing data. *J. of Probabilistic Planning and Inference*, 105(1):105–122, 2002.

[113] M. Zaffalon. The naive credal classifier. *J. of Probabilistic Planning and Inference*, 105(1):5–21, 2002.

[114] M. Zaffalon and E. Fagiuoli. Tree-based credal networks for classification. *Reliable Computing*, 9(6):487–509, 2003.

[115] M. Zaffalon, K. Wesnes, and O. Petrini. Reliable diagnosis of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artificial Intelligence in Medicine*, 29(1-2):61–79, 2003.