Interval-based possibility theory : Conditioning and probability/possibility transformations

THÈSE

en vue de l'obtention du

Doctorat de l'Université d'Artois

(Spécialité Informatique)

par

Amélie Levray

Rapporteurs :	Gabriele Kern-Isberner	Professeur des Universités, TU Dortmund
	Gabriella Pasi	Professeur des Universités, IR Lab Milano
Examinateurs :	Hélène Fargier	Directeur de Recherche, IRIT Toulouse
	Eric Lefevre	Professeur des Universités, LGI2A Béthune
	Philippe Leray	Professeur des Universités, LS2N Nantes
Encadrants :	Salem Benferhat	Professeur des Universités, Université d'Artois
	Karim Tabia	Maître de Conférences, Université d'Artois

CENTRE DE RECHERCHE EN INFORMATIQUE DE LENS – CNRS UMR 8188
Université d'Artois, rue Jean Souvraz, S.P. 18 F-62307, Lens Cedex France
Secrétariat : Tél.: +33 (0)3 21 79 17 23 – Fax : +33 (0)3 21 79 17 70
http://www.cril.fr

Mise en page avec memcril (B. Mazure, CRIL) et thloria (D. Roegel, LORIA).

Remerciements

This manuscript concludes the last three years of work I have done alongside Salem Benferhat and Karim Tabia. I would like to thank them for what they have brought to me and to this thesis. Their kindness and their advices were valuable. I have loved working with them. The liberty and guidance they gave me have allowed me to grow both personally and professionally.

A special thanks to all the members of the jury namely, Gabriele Kern-Isberner and Gabriella Pasi for agreeing to review my thesis and for their kind words on their report. I also thank Hélène Fargier, Eric Lefevre and Philippe Leray for being part of my defense jury.

To my friends, family and colleagues :

> Je remercie tous les membres du CRIL d'avoir fait de ces 3 années de thèse une aventure. En particulier, Virginie, Thomas, Jérôme et Nicolas.
>
> A tous mes amis, anciens camarades qui m'ont soutenu et avec qui j'ai pu décompressé, Déborah et Clément notamment.
>
> Enfin, je remercie de tout coeur toute ma famille qui m'a soutenu dans mes choix d'études et pendant ma thèse mais surtout pour avoir supporté mes extravagances après les semaines intensives au laboratoire.

Contents

Part III Analysis of probability-possibility transformations and MAP queries

Appendix

List of Figures

List of Tables

Introduction

Representing and reasoning under uncertainty is still an important and hot topic in the Artificial Intelligence community. Indeed, many real world problems need to handle uncertain, and more generally, imperfect information while requiring efficient inference and reasoning tools in order to draw conclusions, and make decisions, etc. To meet such requirements, many settings have been proposed to model different types of uncertainty ranging from classical probability and possibility theories to their generalizations such as imprecise probability theory and interval-based possibility theory to name a few.

Clearly, what is needed is i) flexible, expressive and compact representations and ii) efficient inference and reasoning machineries. For the first part regarding the quality of the representations, one can assert that imprecise probabilistic graphical models (commonly called credal networks) and probabilistic logics [FHM90, Luk99, Ada66] are two of the most flexible and compact representations of uncertain knowledge. However, inference and reasoning with such models is usually too computationally expensive [MdCBA14, dCC05] to be used in real applications, often involving large sets of variables. This is on the one hand. On the other one, flexible possibilistic representations such as interval-based possibilistic logic and interval-based possibilistic networks share the same advantages in terms of flexibility and compactness, and more interestingly they could show better results in reasoning and inference tasks since the possibilistic setting is qualitative and makes query answering simpler in many situations. For instance, the use of idempotent operators in the possibilistic setting like the minimum and maximum may lead to speeding up many tasks while answering some types of queries. This is the starting point of this thesis: Take advantage of flexible and compact possibilistic representations in order to reason with uncertain knowledge. Many issues have to be solved to reach such objectives. More precisely,

— Reasoning and inference in interval-based possibilistic settings have not been proposed. For instance, there is no study on conditioning uncertain information in interval-based possibilistic bases and networks. Our first works were devoted to developing a conditioning machinery for interval-based possibilistic logic. Such a machinery along with encoding interval-based possibilistic networks in the form of interval-based possibilistic bases will allow to update uncertain information when new pieces of evidence become available. Our contributions regarding these issues are presented in Chapters 4, 5 and 6.

— Given the fact that agents and experts in real world problems are more used to probabilistic representations and given the fact that empirical data is naturally captured by means of frequencies and probabilities, probabilistic representations are very natural. The question then is how to perform efficient inference and reasoning tasks given the very high computational complexity of inference in expressive probabilistic formalisms such as credal networks. Our idea is to transform such probabilistic representations into suitable possibilistic ones and answer queries using the possibilistic representations. Of course, transformations may result in some information loss but they may allow to benefit from the efficiency of the reasoning and inference machinery that will be developed for the interval-based possibilistic representations. Then there remains to find "good" transformations allowing to encode the probabilistic representations in the interval-based possibilistic setting and empirically evaluate such transformations. These issues are addressed in Chapters 7 and 8.

In this thesis, we are particularly interested in the possibilistic frameworks. Possibility theory [DP15, Zad99] is a well-known uncertainty theory. This framework was coined by L. Zadeh [Zad99] and it is developed by several researchers (*e.g.* Dubois and Prade [DP12], Yager [Yag82] and Borgelt and Kruse [BK03]). Possibility theory is based on a pair of dual measures allowing to evaluate the knowledge/ignorance relative to the event in hand. Among the main concepts of this framework are the ones of possibility distributions and possibilistic logic knowledge bases. Possibility theory and possibilistic logic [CSNC15, Dub14, Lan00, DP11, DGM11] are uncertainty frameworks particularly suited for representing and reasoning with uncertain, incomplete, prioritized and qualitative information. Uncertainty is syntactically represented by a set of weighted formulas of the form $K = \{(\varphi_i, \alpha_i) : i = 1, .., n\}$ where φ_i's are propositional formulas and α_i's are real numbers belonging to $[0, 1]$. The pair (φ_i, α_i) means that φ_i is certain (or important) to at least a degree α_i. An inference machinery has been proposed in [Lan00] to derive plausible conclusions from a possibilistic knowledge base, which needs $\log_2(m)$ calls to the satisfiability test of a set of propositional clauses (SAT), where m is the number of different levels used in K. Uncertainty is also represented at the semantic level by associating a possibility degree with each possible world (or interpretation).

In the literature, many extensions have been proposed for possibilistic logic to deal for instance with imprecise certainty degrees [BHLR11], symbolic certainty weights [BP05, CDT15], multi-agent beliefs [BDKP13], temporal and uncertain information [DLP91], uncertain conditional events [CP16, CPV14], generalized possibilistic logic [CSNC15, DP11, DPS12], justified beliefs [FL15], etc.

Interval-based possibility theory

Interval-based uncertainty representations extend the underlying uncertainty settings in order to encode uncertainty by means of intervals of possible degrees instead of single values. Such extensions allow more flexible representations especially to deal with poor information, imprecise or ill-known beliefs, confidence intervals and multi-source information [Dub06, NK14]. Such representations are very widely used in some applications such as sensitivity analysis. Interval-based possibilistic logic [BHLR11] extends the standard possibilistic logic setting to allow intervals of possible degrees instead of single values attached to the formulas of the knowledge base.

Conditioning [FKRS12, Ker04] is an important task for updating the current uncertain information when a new sure piece of information is received. A conditioning operator is designed to satisfy some desirable properties such as giving priority to the new information and ensuring minimal change while transforming an initial distribution into a conditional one. Conditioning in standard (single-valued) possibility theory has been addressed in many works [His78, LMDCM95, DP06, Fon97, DP97a, BTM99, dC, Hsi94, BDCPT13]. There are two major definitions of a possibility theory: min-based (or qualitative) possibility theory and product-based (or quantitative) possibility theory. At the semantic level, these two theories share the same definitions, including the concepts of possibility distributions, necessity measures, possibility measures and the definition of normalization conditions. However, they differ in the way they define possibilistic conditioning. The first one is called min-based conditioning [His78, DP90] (or qualitative-based conditioning) which is appropriate in situations where only the ordering between events is important. In this case, the unit interval $[0, 1]$ is viewed as an ordinal scale where only the minimum and the maximum operations are used for propagating uncertainty degrees. The second definition of conditioning is called product-based conditioning (or quantitative-based conditioning) where the unit interval is used in a general sense. In this case, the product operation can also be used in the propagation of uncertainty degrees.

Interval-based possibilistic logic defined in [BHLR11] is only specified for static situations and no form of conditioning has been proposed for updating the current knowledge and beliefs. In this thesis, we tackle both definitions of conditioning in an interval-based possibilistic setting. More precisely, we address two important issues. The first one is whether one can extend and increase the expressive power of standard possibilistic logic, by representing imprecision regarding uncertainty associated with formulas, without increasing the computational complexity of the reasoning process. The second important issue concerns foundations of conditioning in the interval-based setting. Some concerns arise such as what natural properties an interval-based conditioning should satisfy in a possibilistic setting. Another extension of possibility possibilistic framework can be proposed. For instance, a new extension of possibilistic logic where the weights associated with formulas are in the form of sets of uncertainty degrees. A set of certainty degrees associated with a formula may represent the reliability levels of different sources that support the formula. This can be provided along with a conditioning method. Among the other extensions, symbolic possibilistic logic [BP05, CDT15] deals with a special type of uncertainty where the available uncertain information is in the form of partial knowledge on the relative certainty degrees (symbolic weights) associated with formulas. In [BDKP13], a multiple agent extension of possibilistic logic is proposed. This extension associates sets of agents to sets of possibilistic logic formulas and aims to reason on the individual and mutual beliefs of the agents. Note that no form of conditioning the whole knowledge is proposed for this setting. Clearly, many of the qualitative extensions of possibilistic logic mentioned in this section could benefit from our conditioning operators as far as they can be encoded as set-valued possibilistic bases.

Probability-possibility transformations

A possibilistic formalism as we have said is a natural alternative to represent uncertain information. However, in many cases, information might not always be expressed using possibility degrees. Indeed, considering the different uncertain theories available to represent knowledge, a natural question is how to deal with these different theories at the same time. Transformations in that respect is a natural solution. The idea of transformation is to express the beliefs or uncertain information defined in some uncertainty formalism into another formalism. The aim is to minimize information loss in order to be able to infer the same conclusions in the new setting. In this thesis, we particularly focus on transformations from probability to possibility distributions. The early works involving probability and possibility theories were devoted to establishing connections between these two frameworks [Zad99, KG93]. These works are mostly interested in finding desirable properties to satisfy and then proposing transformations that guarantee the satisfaction of these properties. An example of such desirable properties is the consistency principle used to preserve as much information as possible. Probability-possibility transformations are useful in many ways. For instance, see [BCD07] for an example of propagating probabilistic (stochastic) and possibilistic information in risk analysis. Another motivation is the fact that probabilities are more suitable in a frequentist setting, but this requires a large number of data. And when data is not available in sufficient quantities then the possibilistic setting can fill this lack as in [MD06]. Another motivation for developing probability-possibility transformations is to use existing tools (e.g. algorithms and software) developed in one setting rather than developing everything from scratch. In this thesis, we analyze probability-possibility transformations with respect to reasoning tasks (such as marginalization and conditioning) and also within graphical models (with a particular focus on *MPE* and *MAP* queries).

MAP (**M**aximum **A** **P**osteriori) inference in probabilistic graphical models is a problem of great interest and has been investigated for years [Kwi14, KBvdG10, MdCC15, Pea89, SD03]. Thus, there exists a variety of methods and algorithms to compute the configuration of query variables with the highest probability given some observed variables. However, Bayesian networks, which are the most widely used probabilistic graphical models, might seem unfit for representing some kinds of information such as the knowledge of a group of experts, or incomplete knowledge. This is why more general frameworks are needed for allowing more flexibility especially regarding the model parameters. Credal networks [Coz00] have been designed to generalize Bayesian networks and offer more expressiveness as they represent uncertain information by means of credal sets instead of single probability values. The problem when reasoning with such general and expressive models is that they entail higher computational complexity. Methods and algorithms to compute *MAP* inference in credal networks exist and give good results in terms of accuracy [MdCBA13]. However, these methods are not very efficient in terms of computational complexity especially when dealing with problems having many variables. A question is how to improve computational complexity of *MAP* inference in credal networks. Transformations from imprecise probability theory to possibility theory is an idea. An experimental study would support the idea. This thesis provides answers to these issues.

Organization of the thesis and contributions

The first part of this thesis (**State of the art**) is composed of three chapters. Chapter 1 presents main concepts of probability theory [Kol60], possibility theory [Zad99], imprecise probability theory [Wal07], OCF (Ordinal conditional functions)[BHK14, EK14, KT12] and belief functions [Dem67, Sha76]. The second chapter concerns compact representations of uncertainty distributions (namely probability and possibility) by means of graphical models and knowledge bases. Chapter 3 gives a review on existing procedures that transform (im)precise probability distributions to possibility distributions [DFMP04, KG93, DPS93, MSMR06].

The second part (**Conditioning in interval-based and set-based possibilistic frameworks**) is dedicated to the proposal of a conditioning operator in the interval-based possibility and set-valued possibility frameworks. This part is divided into three chapters.

Chapter 4, entitled **Quantitative conditioning in interval-based possibilistic setting**, tackles conditioning in quantitative or product-based interval-based possibilistic setting. The proposed conditioning operator is based on the notion of compatible possibility distributions (resp. knowledge bases). It first gives a reminder on interval-based possibility theory. We characterize expected properties that a compatible conditioning should satisfy. Then we propose a natural definition of conditioning an interval-based possibility distribution with a new evidence. This definition is safe since it takes into account all the compatible distributions. We show that applying product-based conditioning leads to an interval-based possibility distribution. We provide the exact computations of lower and upper endpoints of intervals associated with each interpretation of the conditioned interval-based possibility distribution. Lastly, we propose a syntactic counterpart of conditioning over interval-based possibilistic bases. The proposed conditioning does not induce extra computational costs. Conditioning an interval-based possibilistic knowledge base has the same complexity as conditioning a standard possibilistic knowledge base.

Results of this chapter have been published in [BLTK15, BLT15b].

Chapter 5, entitled **Qualitative conditioning in an interval-based possibilistic setting**, tackles compatible-based conditioning in a qualitative (or min-based) interval-based possibility distribution. We first propose three natural postulates for an interval-based conditioning. We show that any interval-based conditioning satisfying these postulates is necessarily based on applying $\min$-based conditioning on each compatible standard possibility distribution. The second contribution consists in providing the exact lower and upper endpoints of $\min$-based conditioning an interval-based distribution and a proposal of efficient procedures to compute the lower and upper endpoints of the conditional interval-based possibility distribution. The third contribution concerns syntactic computations of conditioning where interval-based possibility distributions are compactly represented by interval-based knowledge bases. We again show that qualitative interval-based conditioning has the same computational complexity as the standard $\min$-based conditioning.

Results of this chapter are to appear shortly in FSS - Fuzzy Sets and Systems journal (a minor revision is required).

Chapter 6, entitled **Set-valued possibilistic framework : Definitions and conditioning**, focuses on a new extension of possibilistic logic where the weights associated with formulas are in the form of sets of uncertainty degrees. The first contribution of this paper concerns then the definition of a set-valued possibility theory which generalizes both standard possibility theory and interval-based possibility theory [BHLR11]. The second contribution deals with conditioning in a set-valued possibility theory setting. We again propose three natural postulates for a set-valued conditioning. These three postulates are the counterparts of the ones used for analyzing qualitative interval-based possibilistic conditioning. We also show that any set-valued conditioning satisfying these postulates is necessarily based on applying $\min$-based conditioning on each compatible standard possibility distribution. We provide the exact set of possibility degrees associated with $\min$-based conditioning a set-valued distribution. The last contribution concerns efficient and syntactic computations of conditioning set-valued knowledge bases.

Results of this chapter have been published in [BLTK16].

Chapter 7, entitled **Property analysis of probability-possibility transformations**, deals with probability-possibility transformations with respect to reasoning tasks and graphical models. In this work, we are interested in analyzing properties of probability-possibility transformations such as preserving marginalization, preserving conditioning and preserving independence relations. We analyze these properties when the available information is encoded by means of distributions or in the form of graphical models. We show that there is no transformation from the probabilistic into the possibilistic setting that guarantees most of the reasoning tasks dealt with in this work. For instance, regarding preserving marginalization, we show that no transformation can preserve the relative order of arbitrary events even if it preserves the relative order of interpretations.

Results of this chapter have been published in [BLT15c, BLT15d].

In Chapter 8, entitled **Approximation of Map Inference in Credal Networks**, we provide a new and efficient method for *MAP* inference in credal networks based on imprecise probability-possibility transformations. The main contributions of this work consist in proposing and analyzing a probability-possibility transformation allowing us to turn a credal network into a possibilistic network. We use these transformations to propose an approximate approach for MAP inference in credal networks. We experimentally evaluate and compare our approach to both existing exact and approximate approaches for *MAP* inference in credal networks. The benefits of our approach are reducing the computational time of *MAP* inference while ensuring narrower answer sets. An important contribution of this chapter concerns the analysis of *MAP* inference complexity in possibilistic networks. The work of this chapter is highlighted by an application to learning possibilistic networks with an application to classification problems.

Results of this chapter have been published in [BLT15a, HLLT17, HLLT16, BLT17b, BLT17a].

The **Conclusion** chapter contains a general review of the main contributions. It also highlights some future works which are directly linked to our contributions.

In the **Appendix**, one can find additional background notions on graphical models.

List of publications

[BLT15d] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *Transformations probabilistes-possibilistes: conditionnement, inférence et modèles graphiques*. In JIAF 2015 - Neuvième Journées d'Intelligence Artificielle Fondamentale, 29 June-3 July 2015, Rennes, France, 2015.

[BLT15c] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *On the Analysis of Probability-Possibility Transformations: Changing Operations and Graphical Models*. In ECSQARU 2015, Compiegne, France, July 15-17, 2015. Proceedings, 2015.

[BLTK15] Salem BENFERHAT, Amélie LEVRAY, Karim TABIA, and Vladik KREINOVICH. *Compatible-Based Conditioning in Interval-Based Possibilistic Logic*. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2777-2783, 2015.

[BLT15a] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *Probability-Possibility Transformations: Application to Credal Networks*. In SUM 2015, Québec City, Canada, pages 203–219. Springer, 2015.

[BLT15b] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *Conditionnement en Logique Possibiliste à Intervalles*. In LFA 2015 - 24ème Conférence sur la Logique Floue et ses Applications, 5-6 November 2015, Poitiers, France, 2015.

[HLLT17] Maroua HADDAD, Philippe LERAY, Amélie LEVRAY, and Karim TABIA. *Possibilistic networks parameter learning: Preliminary empirical comparison*. In JFRB 2016 - 8ème Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes, 27 June-1 July 2016, Clermont-Ferrand, France, 2016.

[BLTK16] Salem BENFERHAT, Amélie LEVRAY, Karim TABIA, and Vladik KREINOVICH. *Set-Valued Conditioning in a Possibility Theory Setting*. In Proceedings of the 22nd European Conference on Artificial Intelligence, ECAI 2016, The Hague, The Netherlands, 29 August-2 September 2016, pages 604–612. IOS Press, 2016.

[HLLT16] Maroua HADDAD, Philippe LERAY, Amélie LEVRAY, and Karim TABIA. *Learning the Parameters of Possibilistic Networks from Data: Empirical Comparison*. In FLAIRS Conference, pages 736–741. AAAI Press, 2017.

[BLT17b] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *Approximation de l'inférence MAP via les transformations probabilistes-possibilistes*. In JIAF 2017 - Onzième Journées d'Intelligence Artificielle Fondamentale, 3-7 July 2017, Caen, France, 2017.

[BLT17a] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. *Approximating MAP inference in credal networks using probability-possibility transformations*. In ICTAI 2017 - 29th International Conference on Tools with Artificial Intelligence, 2017, Boston, MA, USA

Part I

State of the art

Chapter 1

Uncertainty theories

When it comes to knowledge representation, a major concern is how to handle and represent uncertain information. In Artificial intelligence, since the last part of XXth century, the interest for knowledge representation tends to grow, especially the representation of information tainted with imprecision and uncertainty. This chapter gives a brief overview of some well-known uncertainty frameworks such as probability theory [Kol60, Jay03, Nea12], possibility theory [DP88, Zad99, DP98, Coo97], imprecise probability [Wal07, Wal00, Wei00, DP05], and belief functions [Dem67, Sha76, She94].

Contents

1.1 Introduction

Referring to [DP09] the term *information* refers to any collection of symbols or signs produced either through the observation of natural or artificial phenomena or by cognitive human activity with a view to help an agent understand the world or the current situation, making decisions, or communicating with other human or artificial agents.

In general, a piece of information can be generic or singular. *Generic* information refers to a collection of situations, for instance it can be a physical law, or information derived from repeated observations, or even a piece of commonsense knowledge. A *Singular* information refers to a particular situation considered as true, such as results of an observation, of tests or yet measures. It can also come from a testimony. The distinction between generic and singular information is important when considering problems of conditioning or revision of uncertain information [KH15].

When dealing with uncertainty, a distinction between so-called *objective* information resulting from sensor measurements and *subjective* information typically uttered by individuals (e.g. testimonies) is often made. Namely, a *subjective* information is individual and susceptible to change given one's personality, whereas *objective* information does not depend upon personal factors in the judgment made.

We can distinguish uncertain information by different types whether it is incomplete information, imprecise information, fuzzy information or uncertain information.

We define incomplete information as information that does not allow an agent to answer precisely a question in a given context. In this respect, incompleteness and imprecision are linked, as we interpret imprecision as a form of incompleteness, meaning that an imprecise answer only leads to incomplete information.

In the case of uncertain information, the agent is unsure of the true value of the information, but can qualify the uncertainty. This can be done with a numerical or symbolic label (e.g. linguistic). For instance, consider the statements:

— the probability that the meeting takes more than two hours is 0.7

— it is very possible that it will rain tomorrow

— it is not absolutely certain that Jim comes to the shop tomorrow

To sum up, information entails uncertainty and to model and reason with this uncertainty, it is necessary to have frameworks adapted for handling these different types of uncertain information. For that purpose, we rely on uncertainty theories. In this context, there exists quite a few, as far as we are concerned, we will present only probability theory and possibility theory; the extension of probability theory known as imprecise probability and to be more relevant we will discuss a few others.

Notations

In order to represent the epistemic state of an agent, we need to introduce some notations that are used in this chapter.

— $V = \{X_1, X_2, .., X_n\}$ denotes a set of variables

— x_i denotes a state of the variable X_i

— D_X denotes the domain of X, *i.e.* $x_i \in D_X$, in this thesis, domains are finite and discrete.

— Ω denotes the universe of discourse. When variables are considered, $\Omega = D_{X_1} \times D_{X_2} \times .. \times D_{X_n}$ which is the Cartesian product of all domains of variables in V

— ω or ω_i denotes an interpretation or configuration of Ω. It is an affectation of all variables

— Greek letters, such as ϕ, ψ (or ϕ_i, ψ_i), are used to denote a subset of interpretations, $\phi \subseteq \Omega, \psi \subseteq \Omega$

— p and P respectively denote a probability function and a probability measure respectively

— we use equivalently x_1, x_2, x_3 and $x_1x_2x_3$ to denote the configuration of the variables $X_1 = x_1, X_2 = x_2$ and $X_3 = x_3$

— $\perp$ described the notion of independence, for example $A \perp B$ means that A and B are independent

— π, Π and N represent the possibility function, the possibility measure and the necessity measure respectively

— $|_*$ and $|_m$ denote the two different possibilistic conditioning operators using the product $*$ or the minimum m

— for I an interval $\underline{I}$ and $\overline{I}$ represent the lower and upper bound of I

1.2 Probability theory

We commonly use the word "probability" on a daily basis as a degree of confidence that an event of an uncertain nature will occur. For example, the weather report might say "there is a low probability of light rain in the afternoon." Probability theory [Kol60, Jay03, Nea12] is the mathematical study of phenomenon characterized by uncertainty. It was inspired by chance games in the 17^{th} century. Probability theory is a well-known and widely used uncertainty framework. One of the building blocks of this setting is the one of probability distribution p assigning a probability degree to each elementary state of the world. Probability theory is ruled by Kolmogorov's axioms [Kol60] and have two main interpretations (namely, the frequentist and subjective interpretations).

In this section, we introduce the following background materials in probability theory, which will be used in the remainder of this thesis:

— probability distributions.

— probability measure.

— dependence and independence.

— Bayes' theorem.

1.2.1 Probability distribution

To measure uncertainty, we start with a given *space* of possible outcomes, denoted by Ω. For example, if we consider the age of an agent, we might set $\Omega = \{1, 2, 3, ..., 110\}$. The set Ω is also referred as the *universe of discourse*. In addition, we define an event as a set of possible outcomes. Formally, each event ϕ is a subset of Ω. In our agent's age example, the event $\{6\}$ represents the case where the agent is 6 years old, and the event $\{1, ..., 17\}$ represents the case of a minor agent.

Definition 1.1 (Probability distribution)**.** Let Ω be the universe of discourse, a probability distribution p maps to each state ω_i, a degree within the interval $[0, 1]$.

$$\begin{array}{rcl} p : \Omega & \rightarrow & [0,1] \\ \forall \omega_i \in \Omega,\ \omega_i & \rightarrow & p(\omega_i) \in [0,1]. \end{array} \tag{1.1}$$

When the universe of discourse Ω refers to a set of discrete variables V, $p(x_1 x_2 ... x_n)$ denotes the *probability mass function* of the variables in V, that is,

$$p(x_1 ... x_n) = p(X_1 = x_1, ..., X_n = x_n). \tag{1.2}$$

Note that a variable is denoted by an uppercase letter and its possible values are denoted by the corresponding lowercase letter. For example, if X_i is a binary variable, then x_i can be either 1 or 0.

One of the main concepts of probability theory is the concept of probability measure. The aim of a probability measure is to assign to every subset of Ω a real value measuring the degree of uncertainty about its occurrence.

Definition 1.2 (Probability measure)**.** A probability measure $P(\phi)$ maps any event $\phi \subseteq \Omega$ to a degree (real) in the interval $[0, 1]$ which reflects the odds of ϕ to realize. $P(\phi)$ is define as follows:

$$P(\phi) = \sum_{\omega_i \in \phi} p(\omega_i) \tag{1.3}$$

Example 1.1. On this example we give two different probability distributions, one over a set of configurations $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ (see Table 1.1). And another one over a set of binary variables $V = \{A, B\}$ with $D_A = \{a, \overline{a}\}$ and $D_B = \{b, \overline{b}\}$ (see Table 1.2). Thus, the universe of discourse for this table is given by $\Omega = \{ab, a\overline{b}, \overline{a}b, \overline{a}\overline{b}\}$.

On this example, we might want to compute the degree associated to the event $\phi = \{\omega_1, \omega_2\}$ and the degree associated to the event $\psi = \{ab, a\overline{b}\}$ which also refers to the degree of the event "A is true" ($A = a$). To do so, it is enough to sum up the probability degrees of ω_1 and ω_2.

ω_i	$p(\omega_i)$
ω_1	0.2
ω_2	0.5
ω_3	0.15
ω_4	0.15

Table 1.1 – Example of a probability distribution over a set of states

AB	$p(AB)$
ab	0.6
$a\bar{b}$	0.3
$\bar{a}b$	0.1
$\bar{a}\bar{b}$	0.0

Table 1.2 – Example of a probability distribution over two binary variables A and B

Kolmogorov's axioms

Probability theory is ruled by axioms, known as Kolmogorov axioms [Kol60]:

Axiom 1 (Non-negativity)**.**

$$\phi \subseteq \Omega, 0 \leq P(\phi) \leq 1 \tag{1.4}$$

An event belonging to the universe of discourse can be in the worst case impossible (with a probability degree of 0) or sure in the best of cases (with a probability degree of 1).

Axiom 2 (Normalization)**.**

$$P(\Omega) = 1 \tag{1.5}$$

The probability mass is entirely distributed to the events of the universe of discourse.

Axiom 3 ((Finite) Additivity)**.**

$$P(\phi \cup \psi) = P(\phi) + P(\psi) \text{ if } \phi \text{ and } \psi \text{ are disjoint sets} \tag{1.6}$$

This axiom is an aggregation formula that can be used to compute the probability of a union of disjoints subsets. It states that the uncertainty of a given subset is the sum of the uncertainties of its disjoints parts.

From the above axioms, many interesting properties [CGH96] of a probability measure can be derived. For example:

- **Property 1 (Boundary):** $P(\emptyset) = 0$.
- **Property 2 (Monotonicity):** If $\phi \subseteq \psi \subseteq \Omega$, then $P(\phi) \leq P(\psi)$.
- **Property 3 (Inclusion-Exclusion):** Given any pair of subsets ϕ and ψ of Ω, the following equality always holds:

$$P(\phi \cup \psi) = P(\phi) + P(\psi) - P(\phi \cap \psi) \tag{1.7}$$

Property 1 states that the evidence associated with a complete lack of information is defined to be zero. Property 2 shows that the evidence of the membership of an element in a set must be at least as great as the evidence that the element belongs to any of its subsets. In other words, the certainty of an element belonging to a given set ϕ must not decrease when adding elements to ϕ. Property 3 states that the probabilities of the sets ϕ, ψ, $\phi \cup \psi$ and $\phi \cap \psi$ are not independent; they are related by Equation (1.7).

Intuitively, the probability $P(\phi)$ of an event ϕ quantifies the degree of confidence that a state from ϕ will occur. If $P(\phi) = 1$, we are certain that one of the outcomes in ϕ occurs, and if $P(\phi) = 0$, we consider all of them impossible. Other probability values represent options that lie between these two extremes.

This description, however does not provide an answer to what the numbers mean. There are two common interpretations for probabilities.

— **Frequentist interpretation** where probabilities are related to frequencies. We consider the probability of an event as the fraction of times the event occurs if we repeat the experiment indefinitely. Finite frequentism remains the dominant view of probability in statistics, and in the sciences more generally.

— **Subjective probability** is an alternative interpretation and it views probabilities as *subjective* degrees of belief. Under this interpretation, the statement $P(\phi) = 0.4$ represents a subjective statement about one's own degree of belief that the event ϕ will come about. Thus, the statement "the probability of having a math test tomorrow is 50 percent" tells us that in the opinion of the speaker, the chances of a test or no test tomorrow are the same.

1.2.2 Probabilistic reasoning: marginalization and conditioning

Let $V = \{X_1, X_2, .., X_n\}$ be a set of discrete random variables and $\{x_1, .., x_n\}$ be a set of their possible realizations or instantiations. Let p be a probability distribution, then given this distribution one can apply some reasoning rules in order to infer information. These rules are defined in the following.

Once we define a distribution over the universe of discourse, we can consider a distribution over events (or a set of variables, or a unique variable). This distribution is referred to as *marginal probability distribution* and is given by Definition 1.3.

Definition 1.3 (Marginalization rule)**.** Let $\{X_1, \ldots, X_k\} \subseteq V$,

$$p(X_1..X_k) = \sum_{X_{k+1}..X_n} (p(X_1X_2..X_n)) \tag{1.8}$$

Marginalization is an extension to variables of the additivity axiom.

Remark. A marginal distribution is a probability distribution, therefore it satisfies the three axioms from Section 1.2.1.

Knowledge about the occurrence of an event can modify the probabilities of other events. For example, the probability of rain tomorrow afternoon can change after watching the weather report on television. Thus, each time new information becomes available, the probabilities of events may change. This leads to the concept of *conditional probability*.

Definition 1.4 (Conditioning)**.** Let ϕ and ψ be two events. Then the conditional probability of ϕ given ψ is computed following:

$$P(\phi|\psi) = \frac{P(\phi \cap \psi)}{P(\psi)} \tag{1.9}$$

We deduce from Definition 1.4, the product rule that allows to compute any intersection of events.

Definition 1.5 (Product rule)**.** Let ϕ and ψ be two events.

$$P(\phi \cap \psi) = P(\phi) * P(\psi|\phi) = P(\psi) * P(\phi|\psi) \tag{1.10}$$

The chain rule allows to compute any configuration of a distribution over a set of random variables using only probabilistic conditioning rule.

Definition 1.6 (Chain rule)**.** Let $V = \{X_1, \ldots, X_n\}$

$$P(X_1X_2..X_n) = P(X_1) * P(X_2|X_1) * .. * P(X_n|X_1..X_{n-1}) \tag{1.11}$$

Definition 1.7 (Bayes rule)**.** $\forall \phi \subseteq \Omega,\ \forall \psi \subseteq \Omega,$

$$P(\phi|\psi) = \frac{P(\phi) * P(\psi|\phi)}{P(\psi)} \tag{1.12}$$

Bayes rule allows to express $P(\phi|\psi)$ given $P(\psi|\phi)$ (invert the probability). This allows, for example, to compute probability measure of causes if the effects are known and the probability degrees of these effects given the causes.

Conditioning with uncertain evidence

There are two types of evidence that one may encounter: hard evidence and soft evidence. *Hard evidence* is information known to the true, which is also the type of evidence we have considered previously in the definition of conditioning (Definition 1.4). *Soft evidence* or uncertain evidence, on the other hand, is not certain: we may get an unreliable testimony that event ϕ occurred, which may increase our belief in ϕ but not to the point where we would consider it certain. Conditioning with uncertain information, therefore, differs from conditioning with hard evidence. We need to take into account the uncertainty about the new information. For example, the confidence we have in the weather report might not be complete, and after watching the report our belief changes but not with a certain measure so we must adjust our judgment.

Jeffrey's conditioning rule [Jef65] allows to revise a probability distribution p into p' given a third probability distribution depicting the new piece of uncertain information. Note that Bayes conditioning is a special case of Jeffrey's rule when $\alpha = 1$. This method involves:

— **A way to specify uncertain evidence**: let $\phi_1, ..., \phi_n$ be an exhaustive and mutually exclusive set of events where the uncertainty is described as a set of couples $\alpha = \{(\phi_i, \alpha_i), i = 1, ..., n\}$ which means that after the revision, the a posteriori probability of each event ϕ_i must be equal to α_i. Namely,

$$\forall (\phi_i, \alpha_i) \in \alpha,\ P'(\phi_i) = \alpha_i \tag{1.13}$$

— **Minimal change principle**: the minimal change principle says that the probabilistic conditioning of any event $\psi \subseteq \Omega$ given an uncertain event ϕ_i stays the same in the initial and revised probability distribution. More formally,

$$\forall \phi_i \subseteq \Omega,\ \forall \psi \subseteq \Omega,\ P'(\psi \mid \phi_i) = P(\psi \mid \phi_i) \tag{1.14}$$

The underlying interpretation of the revision involved in Equation (1.14) is that the conditioned probability distribution p' does not change the conditional probability degree of any event ψ given the uncertain event ϕ_i.

Equations (1.13) and (1.14) allow to condition a probability distribution p' given α. This is defined as:

$$\forall \psi \subseteq \Omega,\ P'(\psi) = \sum_{\phi_i} \alpha * \frac{P(\psi, \phi_i)}{P(\phi_i)} \tag{1.15}$$

Remark. The conditioned probability distribution p' obtained using Jeffrey's rule is unique and always exists [CD05, BTS10].

1.2.3 Probabilistic independence

Definition 1.8 (Independence of two events)**.** Let ϕ and ψ be two events included in Ω. Then ϕ is said to be independent of ψ, denoted $\phi \perp \psi$, if and only if

$$p(\phi|\psi) = p(\phi). \tag{1.16}$$

Otherwise ϕ is said to be dependent on ψ, denoted $\phi \not\perp \psi$.

Definition 1.9 (Independence of two variables)**.** Let X and Y be two variables of the set of random variables V. Then X is said to be independent of Y, denoted $X \perp Y$, if and only if

$$\forall x \in X \text{ and } y \in Y,\ \ p(x|y) = p(x). \tag{1.17}$$

Otherwise X is said to be dependent on Y, denoted $X \not\perp Y$.

Equation (1.17) means that if X is independent of Y, then our knowledge of Y does not affect our knowledge about X. Also, if X is independent of Y, we can then combine Equation (1.9) and (1.17) and obtain $\dfrac{p(x,y)}{p(y)} = p(x)$, which implies

$$p(x,y) = p(x)p(y). \tag{1.18}$$

Equation (1.18) says that if X is independent of Y, then the joint probability distribution of X and Y is equal to the product of their marginals.

The concept of dependence and independence of two random variables can be extended to the case of more than two random variables as follows:

Definition 1.10 (Independence of a set of variables)**.** The random variables $\{X_1, X_2, ..., X_m\}$ are said to be independent if and only if

$$\forall x_1 \in X_1, ..., x_m \in X_m,\ \ p(x_1, ..., x_m) = \prod_{i=1}^{m} p(x_i). \tag{1.19}$$

Otherwise they are said to be dependent.

In other words, $\{X_1, ..., X_m\}$ are said to be independent if and only if their probability distribution is equal to the product of their marginal probability distributions. Note that Equation (1.19) is a generalization of Equation (1.18).

The concepts of dependence and independence deal with two subsets of variables. Now we turn to a generalization of the concept of independence when more than two sets of variables are involved with the definition of conditional independence.

Definition 1.11 (Conditional independence and dependence)**.** Let X, Y and Z be three disjoint sets of variables, then X is said to be conditionally independent of Y given Z, denoted $X \perp Y|Z$, if and only if

$$\forall x \in X, y \in Y \text{ and } z \in Z,\ \ p(x|z,y) = p(x|z). \tag{1.20}$$

Otherwise X and Y are said to be conditionally dependent given Z, denoted $X \not\perp Y|Z$.

This means that once Z is known, knowing Y can no longer influence the probability of X.

Properties of conditional independence

There are five main properties of conditional independence [Daw79, Pea09].

— $X \perp Y|Z$ if and only if $Y \perp X|Z$ (Symmetry)

— $X \perp Y \cup W|Z$ if $X \perp Y|Z$ and $X \perp W|Z$ (Decomposition)

— $X \perp Y \cup W|Z$ if $X \perp W|Z \cup Y$ (Weak union)

— $X \perp Y|Z$ and $X \perp W|Z \cup Y$ if $X \perp W \cup Y|Z$ (Contraction)

— $X \perp Y|Z \cup W$ and $X \perp W|Z \cup Y$ if $X \perp W \cup Y|Z$ (Intersection)

Independence and conditional independence are two fundamental concepts of probability theory as it is the base of Bayesian networks [Pea88] which are compact representations based on probability theory.

1.3 Possibility theory

Possibility theory [DP88, Zad99, DP98, Coo97] is an uncertainty theory dedicated to deal with incomplete and qualitative information. It is largely comparable to probability theory because it is based on sets. It differs from the latter since it uses a pair of dual set functions (possibility and necessity measures) instead of only one. Moreover, it is maxitive and makes sense on ordinal structures. The name "Possibility theory" was coined by Zadeh [Zad99].

1.3.1 Possibility distribution

The basic building blocks of possibility theory were first described in [Zad99, DP80, DP12]. More recent accounts are in [DP98, DP93].

Definition 1.12 (Possibility distribution)**.** Let Ω be the universe of discourse, a possibility distribution maps to each state $\omega_i \in \Omega$, a degree within the interval $[0, 1]$.

$$\begin{aligned} \pi : \Omega &\rightarrow [0,1] \\ \forall \omega_i \in \Omega,\ \omega_i &\rightarrow \pi(\omega_i) \in [0,1]. \end{aligned} \tag{1.21}$$

The function π represents the knowledge of an agent where a distinction is made between what is plausible from what is less plausible, what is surprising from what is expected. It can represent extreme forms of knowledge with:

— $\pi(\omega) = 0$ which mean that state ω is rejected or impossible;

— $\pi(\omega) = 1$ which means that state ω is totally possible (plausible).

Note that possibility degrees are interpreted either *qualitatively* (in $\min$-based possibility theory) where only the "ordering" of the values is important, or *quantitatively* (in product-based possibility theory) where the possibilistic scale $[0, 1]$ is quantitative as in probability theory [DP98].

Contrary to probability theory where the uncertainty of an event ϕ is identified by the uncertainty of the complementary event (namely, $P(\phi) = 1 - P(\overline{\phi})$), in possibility theory given a possibility distribution π over Ω, we distinguish two dual measures that are the possibility measure and the necessity measure.

Possibility measure

The possibility measure represents the compatibility degree of the event ϕ with the encoded beliefs in π. In other words, this measure evaluates the consistency degree of the event ϕ with the information depicted in π. More formally,

Definition 1.13 (possibility measure)**.** The possibility measure $\Pi(\phi)$ is defined by:

$$\Pi(\phi) = \max_{\omega \in \Omega}\{\pi(\omega) : \omega \in \phi\}. \tag{1.22}$$

This possibility measure is characterized by:

— $\Pi(\phi) = 1$ and $\Pi(\overline{\phi}) = 0$ means that ϕ is necessarily true,

— $\Pi(\phi) = 1$ and $\Pi(\overline{\phi}) \in]0, 1[$ means that ϕ is plausible at a certain degree (ϕ is more plausible then $\overline{\phi}$),

— $\Pi(\phi) = 1$ and $\Pi(\overline{\phi}) = 1$ depicts total ignorance (both ϕ and $\overline{\phi}$ are possible),

— $\Pi(\phi) > \Pi(\psi)$ means that ϕ is more plausible than ψ.

Definition of necessity measure

The necessity measure is the dual measure of the possibility measure. It represents the certainty degree of ϕ given the beliefs expressed in π. Formally,

Definition 1.14 (necessity measure)**.** The necessity measure $N(\phi)$ is defined by:

$$N(\phi) = 1 - \Pi(\overline{\phi}) = \min_{\omega \notin \phi}(1 - \pi(\omega)) \tag{1.23}$$

Note that $N(\phi) > 0 \Rightarrow \Pi(\phi) = 1$ which means that an event is only necessary if it is completely possible. This property ensures the following inequality $N(\phi) \leq \Pi(\phi)$. In the same way, this measure is characterized by:

— $N(\phi) = 1$ and $N(\overline{\phi}) = 0$ means that ϕ is necessarily true,

— $N(\phi) \in]0, 1[$ and $N(\overline{\phi}) = 0$ means that ϕ is plausible at a certain degree,

— $N(\phi) = 0$ and $N(\overline{\phi}) = 0$ depicts total ignorance.

Axioms ruling possibility theory

Axiom 4 (Non-negativity)**.**

$$\forall \phi \subseteq \Omega, 0 \leq \Pi(\phi) \leq 1 \tag{1.24}$$

An event belonging to the universe of discourse can be in the worst case impossible (with a possibility degree of 0) or totally possible in the best of case (with a possibility degree of 1).

Axiom 5 (Normalization)**.** If Ω is exhaustive then at least one of the elements of Ω should be the actual world, so that

$$\exists \omega \in \Omega,\ \pi(\omega) = 1. \tag{1.25}$$

Distinct states may simultaneously have a degree of possibility equal to 1. Thus, a possibility distribution mapping to each state of Ω the possibility degree 1 reflects total ignorance.

The next axiom is what makes the very big difference with probability theory since this latter is additive while possibility theory is maxitive.

Axiom 6 (Maxitivity)**.**

$$\Pi(\phi \cup \psi) = \max(\Pi(\phi), \Pi(\psi)) \tag{1.26}$$

From these axioms, the following properties can be stated:

— $\max(\Pi(\phi), \Pi(\overline{\phi})) = 1$,

— $\Pi(\phi \cap \psi) \leq \min(\Pi(\phi), \Pi(\psi))$,

— $\min(N(\phi), N(\overline{\phi})) = 0$ is the only relation that links ϕ and $\overline{\phi}$,

— $N(\phi \vee \psi) \geq \max(N(\phi), N(\psi))$,

— $N(\phi \wedge \psi) = \min(N(\phi), N(\psi))$.

1.3.2 Possibilistic reasoning: marginalization and conditioning

As in the probabilistic setting, for possibility theory we also define marginalization and conditioning.

Definition 1.15 (Marginalization rule)**.** Let $\{X_1, \ldots, X_k\} \subseteq V$,

$$\pi(X_1..X_k) = \max_{X_{k+1}..X_n} (\pi(X_1 X_2..X_n)) \tag{1.27}$$

Remark. A marginal distribution is also a possibility distribution, therefore it satisfies the three axioms from Section 1.3.1.

Conditioning is an important belief change operation concerned with updating the current beliefs encoded by a probability distribution or a possibility distribution when a completely sure event ϕ (evidence) is available. While there are several similarities between the quantitative possibilistic and the probabilistic frameworks (conditioning is defined in the same way), the qualitative one is significantly different.

Quantitative conditioning

In the quantitative setting, the product-based conditioning (also known as Dempster rule of conditioning) is defined as follows: given a possibility distribution π, and a new evidence $\phi \subseteq \Omega$ ($\phi \neq \emptyset$) the conditional distribution $\pi(.|\phi)$ is obtained using (we assume here that $\Pi(\phi) > 0$):

$$\pi(\omega_i|_* \phi) = \begin{cases} \frac{\pi(\omega_i)}{\Pi(\phi)} & \text{if } \omega_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \tag{1.28}$$

Qualitative conditioning

Conditioning in the qualitative setting is defined as follows [His78]:

$$\pi(\omega_i|_m \phi) = \begin{cases} 1 & \text{if } \pi(\omega_i) = \Pi(\phi) \text{ and } \omega_i \in \phi; \\ \pi(\omega_i) & \text{if } \pi(\omega_i) < \Pi(\phi) \text{ and } \omega_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \tag{1.29}$$

Note that the two definitions of possibilistic conditioning satisfy the condition: $\forall \omega \in \phi$, $\pi(\omega) = \pi(\omega|\phi) \otimes \Pi(\phi)$ where $\otimes$ is either the product or min-based operator.

Different extensions of these two definitions have been proposed. For instance, in [BDCPT13] the authors dealt with syntactic hybrid conditioning of standard (point-wise) possibilistic knowledge bases with uncertain inputs.

Conditioning with uncertain information

Aside conditioning with sure piece of information, we also can condition with uncertain information. As in probability with Jeffrey's conditioning, Dubois and Prade [DP97b, BTS11] have addressed conditioning with uncertain information for the possibilistic setting. Given π a possibility distribution encoding the initial knowledge and $\Pi'(\phi)$ the uncertainty lying on the event ϕ, the uncertainty is denoted by (ϕ, α) such that $\Pi'(\phi) = \alpha$. Then the possibility distribution π' revised according to Dubois and Prade's rule [DP97b] must satisfy the following conditions:

1. $\forall \phi,\ \Pi'(\phi) = \alpha$
2. $\forall \phi \subseteq \Omega,\ \forall \psi \subseteq \Omega,\ \Pi'(\psi \mid \phi) = \Pi(\psi \mid \phi)$

Conditioning based on Jeffrey's rule in the quantitative setting was formalized by Dubois and Prade [DP97b] as follow:

Definition 1.16 (Quantitative conditioning with uncertain information)**.** Let π be a possibility distribution and $\phi_1, ..., \phi_n$ an exhaustive and mutually exclusive set of events where the uncertainty is of the form $\Pi'(\phi_i) = \alpha_i$. The revised possibility degree of an event $\psi \subseteq \Omega$ is computed using the formula:

$$\forall \psi \subseteq \Omega,\ \Pi'(\psi) = \max_{\phi_i}(\alpha_i * \frac{\Pi(\psi, \phi_i)}{\Pi(\phi_i)}) \tag{1.30}$$

According to Definition 1.16, the revised degree of a world $\omega_j \in \Omega$ is computed using the following Lemma 1.1.

Lemma 1.1.

$$\forall \omega_j \in \phi_i,\ \pi'(\omega_j) = \alpha_i * \frac{\pi(\omega_j)}{\Pi(\phi_i)}. \tag{1.31}$$

It is important to note that Definition 1.16 satisfies the axioms defined for possibility theory in particular the normalization condition.

Remark. The conditioned possibility distribution π' obtained using Jeffrey's rule is unique and always exists [BTS11].

Belief revision in qualitative possibility theory can be done using Jeffrey's rule of Definition 1.17 that has been also proposed by Dubois and Prade [DP97b].

Definition 1.17 (Qualitative conditioning with uncertain information)**.** Let π be a possibility distribution and $\phi_1, ..., \phi_n$ an exhaustive and mutually exclusive set of events where the uncertainty is of the form $\Pi'(\phi_i) = \alpha_i$. The revised possibility degree of an event $\psi \subseteq \Omega$ is computed using the formula:

$$\forall \psi \subseteq \Omega,\ \Pi'(\psi) = \max_{\phi_i}(\min(\Pi(\psi \mid \phi_i), \alpha_i)) \tag{1.32}$$

We can easily verify that Π' of Definition 1.17 satisfies the 3 axioms defining a possibility measure.

According to Definition 1.17, the conditioned possibility degree of a world $\omega_j \in \Omega$ is given by Lemma 1.2.

Lemma 1.2.

$$\forall \omega_j \in \phi_i,\ \pi'(\omega_j) = \begin{cases} \alpha_i & \text{if } \omega_j \geq \alpha_i \text{ or } \pi(\omega_j) = \Pi(\phi_i); \\ \pi(\omega_j) & \text{otherwise.} \end{cases} \tag{1.33}$$

Remark. Contrary to product-based conditioning, the conditioned possibility distribution π' obtained using qualitative Jeffrey's rule does not always exist [BTS11].

1.3.3 Possibilistic independence

Possibilistic independence relationship is based on possibilistic conditioning. Indeed, consider X, Y and Z three disjoint sets of variables, then X is said to be independent of Y given Z, if for any configuration z of D_Z, the conditional possibility degree of any configuration $x \in D_X$ remains the same for any configuration of $y \in D_Y$. More formally:

$$\forall x, y, z,\ \Pi(x|y, z) = \Pi(x|z). \tag{1.34}$$

Note that the definition is the same as in probability theory. The only difference lies in the definition of $\Pi(x|y, z)$ in the possibilistic setting. We have two different ways of conditioning; hence we have two definitions of independence:

Definition 1.18 (Conditional independence based on the minimum)**.** Initially defined as a non-interactivity relation in [Zad79], this relation is obtained using the conditioning based on the minimum (in the qualitative possibilistic setting) of Equation(1.29) and is as follows:

$$\forall x, y, z,\ \Pi(x|y, z) = \min(\Pi(x|z), \Pi(y|z)). \tag{1.35}$$

Definition 1.19 (Conditional independence based on the product)**.** The independence is based on the conditioning rule defined in the quantitative possibilistic setting. The relation is described as follows:

$$\forall x, y, z,\ \Pi(x|y, z) = \Pi(x|z) * \Pi(y|z). \tag{1.36}$$

Example 1.2. Let A, B and C be three binary variables. One can easily check in Table 1.3 that A and B are independent given C. More formally, that $\forall a_i \in D_A$, $\forall b_j \in D_B$, $\forall c_k \in D_C$ we have: $\pi(a_i|b_j c_k) = \pi(a_i|c_k)$.

ABC	$\pi_*(ABC)$	$\pi_m(ABC)$
$a_1b_1c_1$	.4	.4
$a_1b_1c_2$	.6	.6
$a_1b_2c_1$	.12	.3
$a_1b_2c_2$	1	1
$a_2b_1c_1$	.3	.3
$a_2b_1c_2$	.162	.3
$a_2b_2c_1$	.09	.3
$a_2b_2c_2$	.27	.3

Table 1.3 – Possibility distribution of two conditionally independent variables A and B given C

1.4 Interval-based probability theory

Probability and possibility theory are among the two most popular theories to represent and reason with uncertainty. However, it is difficult for an agent to provide precise and reliable crisp belief degrees. This has led researchers to develop alternative and flexible formalisms for representing and managing ill-known beliefs. Imprecise probability generalizes standard probability to encode ill-know beliefs where these latter can be encoded by means of sub-intervals of $[0, 1]$. We use real number based intervals $I = [\alpha, \beta] \subseteq [0, 1]$ to encode the uncertainty associated with worlds. We denote by $\mathcal{I}$ the set of all closed intervals over $[0, 1]$. If I is an interval, then we denote by $\overline{I}$ and $\underline{I}$ its upper and lower endpoints respectively. When all I's associated with interpretations are singletons (namely $\overline{I} = \underline{I}$), we refer to standard (or point-wise) distributions.

1.4.1 Credal sets

In real life we have to deal with incomplete knowledge and ill-known beliefs. Such imperfections are due to the lack of time, resources, patience from the agent, or the lack of confidence to provide exact probability values. Imperfections can also come from dealing with a group of disagreeing experts, each specifying a particular probability [Lev80]. It is therefore difficult to meet all the assumptions of a standard probabilistic model. Imprecise probability theory [Wal07, Wal00, Wei00, DP05] generalizes probability theory to encode imprecise and ill-known information.

A key notion in imprecise probability theory is the one of credal set.

Definition 1.20 (Credal set)**.** A credal set is a convex set of probability distributions.

Intuitively, if K is a convex set of probability measures, then *mixing* any two distributions p_1 and p_2 from K will result in a distribution p belonging to K. Mixing here means linearly combining a set of distributions p_1 ... p_k as follows: $p = \sum_{i=1}^{k}(\alpha_i * p_i)$ where $\sum_{i=1}^{k}\alpha_i = 1$. A credal set is often interpreted as a set of imprecise beliefs in the sense that the true uncertainty model (probability measure) is in this set but there is no way to determine it exactly due to lack of knowledge. In order to characterize a credal set, one can use a (finite [1]) set of extreme points [MD13] (edges of the polytope representing the credal set), probability intervals or linear constraints.

Interval-based probability distributions are a very natural and common way to specify imprecise information. In an interval-based probability distribution IP, every interpretation $\omega_i \in \Omega$ is associated with a probability interval $IP(\omega_i) = [\alpha_i, \beta_i]$ where $\underline{IP}(\omega_i) = \alpha_i$ (resp. $\overline{IP}(\omega_i) = \beta_i$) denotes the lower (resp. upper) bound of the probability of ω_i.

Definition 1.21 (Interval-based probability distribution)**.** Let Ω be the set of possible worlds. An interval-based probability distribution IP is a function that maps every interpretation $\omega_i \in \Omega$ to a closed interval $[\alpha_i, \beta_i] \subseteq [0, 1]$.

Example 1.3. Let A be a variable with a domain $D_A = \{a_1, a_2, a_3\}$. Table 1.4 provides an example of interval-based probability distributions.

A	$IP(A)$
a_1	$[0, .4]$
a_2	$[.1, .55]$
a_3	$[.1, .65]$

Table 1.4 – Example of an interval-based probability distribution

An interval-based probability distribution should satisfy the following constraints in order to ensure that the underlying credal set is not empty and every lower/upper probability bound is reachable.

$$\sum_{\omega_i \in \Omega} \underline{IP}(\omega_i) \leq 1 \leq \sum_{\omega_i \in \Omega} \overline{IP}(\omega_i) \tag{1.37}$$

$$\forall \omega_i \in \Omega, \underline{IP}(\omega_i) + \sum_{\omega_{j \neq i} \in \Omega} \overline{IP}(\omega_j) \geq 1 \text{ and } \overline{IP}(\omega_i) + \sum_{\omega_{j \neq i} \in \Omega} \underline{IP}(\omega_j) \leq 1 \tag{1.38}$$

In order to give a formal semantics for interval-based probability distributions, let us first define the concept of compatible probability distribution.

1. It is important to note that the number of extreme points can reach $N!$ where N is the number of interpretations [Wal07].

Definition 1.22 (Compatible probability distribution)**.** A probability distribution p over Ω is said compatible with IP if and only if $\forall \omega_i \in \Omega$, $p(\omega_i) \in IP(\omega_i)$.

Note that while a standard probability distribution p induces a complete order over the set of possible worlds Ω, an interval-based probability distribution IP may induce a partial order since some interpretations may be incomparable in case of overlapping intervals. In this thesis, a credal set K_i associated with a variable A_i having an interval-based probability distribution IP denotes the closed convex set of (standard) probability distributions p that are compatible with IP.

Definition 1.23 (Credal set associated with an Interval-based probability distribution)**.** Let $IP(X_i)$ be and interval-based probability distribution for the discrete variable X_i having the domain D_{X_i}. The credal set K_i associated with IP_i is the set of probability distributions p such that:

$$\begin{cases} p : \forall x_i \in D_i, p(x_i) \in IP(x_i) \\ \sum_{x_i \in D_i} p(x_i) = 1 \end{cases} \tag{1.39}$$

1.4.2 Reasoning with credal sets

Marginalization and conditioning are defined as follows:

Let $K(X_1..X_n)$ be a credal set over the set of variables $V = \{X_1..X_n\}$. Let V_1 and V_2 be two disjoint subsets of V such that $V_1 \cup V_2 = V$. Then,

$$K(V_1) = CH(\{\sum_{V_2} p(V_1, V_2) \text{ with } p(V_1, V_2) \in K(X_1..X_n)\}) \tag{1.40}$$

where CH is the convex hull operator.

As for conditioning, let ϕ be an evidence, then

$$K(X_1..X_n|\phi) = CH(\{p(X_1..X_n|\phi) \text{ with } p(X_1..X_n) \in K(X_1..X_n) \text{ and } p(\phi) > 0\}) \tag{1.41}$$

As the semantics used here to represent an imprecise probability distribution is the one of compatible probability distributions, we can equally redefine the concept of marginalization and conditioning using the set of compatible distributions when the credal set is represented by an interval-based probability distribution.

$$IP(V_1) = [\min_{p \in \mathcal{C}(IP)} p(V_1), \max_{p \in \mathcal{C}(IP)} p(V_1)] \tag{1.42}$$

As for conditioning, let ϕ be an evidence, then

$$IP(X_1, ..., X_n|\phi) = [\min_{p \in \mathcal{C}(IP)} p(X_1, ..., X_n|\phi), \max_{p \in \mathcal{C}(IP)} p(X_1, ..., X_n|\phi)] \tag{1.43}$$

In the rest of this thesis we will always consider credal sets as imprecise (interval-based) probability distribution.

Another generalization of probability and possibility theories is the one of belief functions.

1.5 Belief functions theory

Evidence theory (also known as Dempster-Shafer theory) has been initiated by Dempster [Dem67]. It has been then developed by Shafer [Sha76] with a more completed mathematical formalism and a proposal of a belief functions theory as a general setting to represent uncertainty. It allows to take into account both aleatory uncertainty and epistemic uncertainty into one formalism. It has many interpretations such as upper and lower probability, for example.

Let us consider a variable X which has for possible values $\{x_1, ..., x_n\}$. The set of values gives the universe of discourse: $\Omega = \{x_1, ..., x_n\}$. The set of singletons and disjunctive sets composed of values of X form the power-set 2^Ω, defined by:

$$2^\Omega = \{\emptyset, x_1, ..., x_n, \{x_1, x_2\}, ..., \{x_1, ..., x_n\}\} \tag{1.44}$$

The different elements of 2^Ω allows to represent the set of possible uncertainty situations.

There are several equivalent representations for quantifying beliefs within the belief functions framework. The three main representations are mass functions denoted by *m*, belief functions denoted by *bel* and plausibility functions denoted by *pl*.

1.5.1 Mass function

A *mass function* allows to encode our knowledge and other measures like belief function and plausibility function are derived from this assignment. Formally, a mass function *m* is a mapping $m : 2^\Omega \to [0, 1]$ assigning a mass value to each event $\phi \subseteq \Omega$ of the frame of discernment 2^Ω such that:

$$\sum_{\phi \in 2^\Omega} m(\phi) = 1. \tag{1.45}$$

An additional constraint requires that a mass function must not assign a positive value to the empty set.

$$m(\emptyset) = 0 \tag{1.46}$$

A mass function satisfying this property is called normalized. $m(\phi)$ is interpreted as the part of the belief that supports ϕ. In general, the mass function can be seen as a probability function defined on the power set

1.5.2 Belief and plausibility functions

The total amount of belief committed to a hypothesis $\phi \in 2^\Omega$, including all subsets $\psi \subseteq \phi$, is denoted by $bel(\phi)$. The function $bel : 2^\Omega \to [0, 1]$ is called a belief function. It can be directly computed from a mass function m.

$$\forall \phi \in 2^\Omega, \;\; bel(\phi) = \sum_{\psi \subseteq \phi} m(\psi). \tag{1.47}$$

If m is normalized, then as consequence of Equation (1.47)

$$bel(\Omega) = 1. \tag{1.48}$$

The plausibility $pl(\phi)$ is the amount of belief not strictly committed to $\overline{\phi}$ of ϕ. It therefore expresses how *plausible* a hypothesis ϕ is, *i.e.* how much belief mass potentially supports ϕ. On a formal level, a *plausibility function pl* : $2^\Omega \to [0, 1]$ is defined as

$$\forall \phi \in 2^\Omega, \;\; pl(\phi) = 1 - bel(\overline{\phi}). \tag{1.49}$$

The plausibility $pl(\phi)$ can be computed from a mass function m in the following way:

$$\forall \phi \in 2^\Omega, \;\; pl(\phi) = \sum_{\psi \subseteq \Omega, \; \psi \cap \phi \neq \emptyset} m(\psi). \tag{1.50}$$

1.5.3 Links to probability and possibility theories

Links can be established between the previous theories such as probability and possibility theory and belief functions. Indeed, when the mass function is a probabilistic mass function, it means that the beliefs over each elementary worlds are known precisely. Also, a belief function bel is sometimes interpreted as defining a "lower bound" for an unknown probability function P. Therefore, the real value of the occurrence of a subset $\phi \in 2^{\Omega}$ is surrounded on the lower part by $bel(\phi)$:

$$\forall \phi \subseteq \Omega, \ \ bel(\phi) \leq P(\phi). \tag{1.51}$$

In this case, bel and P are called compatible.

Whereas bel can be viewed as a lower bound for an unknown probability function P under a lower and upper probability interpretation, the plausibility can be viewed as an upper bound. For a normalized plausibility function pl, a compatible probability function P must satisfy the property

$$\forall \phi \subseteq \Omega, \ \ pl(\phi) \geq P(\phi). \tag{1.52}$$

De facto, the frame surrounding the real value of $P(\phi)$ is defined by:

$$bel(\phi) \leq P(\phi) \leq pl(\phi) \tag{1.53}$$

In case where ϕ is a singleton, this equation replaces $\leq$ by an equality.

bel and pl can therefore be seen as the upper and lower bound of an imprecise probability distribution.

Between possibility theory and belief function theory links can also be established, in belief function setting, if the focal elements $\phi_1, \phi_2, ..., \phi_n$ are nested (*i.e.* $\phi_1 \subseteq \phi_2 \subseteq ... \subseteq \phi_n$), then the belief function *bel* is called *consonant belief function* and for all $\phi, \psi \in 2^{\Omega}$, we have:

$$bel(\phi \cap \psi) = \min(bel(\phi), bel(\psi)) \tag{1.54}$$

and

$$pl(\phi \cup \psi) = \max(pl(\phi), pl(\psi)) \tag{1.55}$$

In this case *belief functions* are *necessity measures* and *possibility measures i.e.* $bel = N$ and $pl = \Pi$.

There exist a lot more general representations to encode uncertain information. For instance, we have *random sets, generalized P-boxes, clouds* as other means of representing probabilistic information. A random set is a set-valued mapping from a probability space to a set V [Dem67]. In [DDC07], the authors use mass functions [Sha76] to represent random sets. A p-box [FKG+03] (which literally means "probability box") is defined using a pair of cumulative distributions $[\underline{F}, \overline{F}]$ that defines a probability family. Modeling using p-box is generally used to deal with problems where the form of the probability distribution is known but where parameters such as then mean or the standard deviation are imprecise. This theory therefore, allows to deal together aleatory uncertainty and epistemic uncertainty. A P-box can be built given known distribution and appears to be a natural choice for parametric model with imprecise parameters. The intervals formed by the pair of distributions characterize the incomplete nature of the knowledge. A P-box can also be built from experts opinions [CGF12]. This theory of P-box will be used in one transformation from imprecise probabilities to possibilities in Chapter 3. Clouds are defined by interval-valued fuzzy sets that can also be linked to possibility distributions [DP05].

1.6 Ordinal conditional functions (OCF)

In this section, we present ordinal conditional functions (OCF) to represent epistemic states [Spo14, TCD15, Wil95, Wil94a, BHK14, EK14, KT12].

An OCF distribution can be simply viewed as a function that assigns to each interpretation ω of Ω an integer denoted by $k(\omega)$. $k(\omega)$ represents the degree of surprise of having ω as being the real world. $k(\omega) = 0$ means that nothing prevents ω for being the real world. $k(\omega) = 1$ means that ω is somewhat surprising to be the real world. $k(\omega) = +\infty$ simply means that it is impossible for ω to be the real world.

Example 1.4. Let a and b be two propositional symbols. Table 1.5 gives an example of an epistemic state represented by an OCF distribution k:

ω	$k(\omega)$
ab	4
$\neg ab$	1
$a\neg b$	1
$\neg a\neg b$	0

Table 1.5 – An example of an OCF distribution

From Table 1.5, the most normal state of world is the one where both a and b are false. The most surprising world (with a degree of surprise 4) is the one where both a and b are true.

From an OCF distribution k, one can induce a degree of surprise over events ϕ, simply denoted by $k(\phi)$ and defined by:

$$k(\phi) \;=\; \min\{k(\omega) : \omega \in \Omega, \omega \in \phi\}. \tag{1.56}$$

For example, from Table 1.5 we have $k(\neg a \vee \neg b) = \min(k(\neg a\neg b), k(\neg ab), k(a\neg b)) = 0$ while $k(a \vee b) = 1$.

In this framework, several works have been proposed for revising OCF distributions. For instance, in [Wil95, Wil94a] a general form of changing OCF distributions, called transmutations [Wil94b], has been proposed. In [Ker01, KE14] a revision of OCF distributions with a set of conditionals has also been proposed. And also a so-called multiple iterated belief c-revision proposed in [KH15, FKRS12, Ker04] for revising an OCF distribution with a consistent set of propositional formulas.

In practice, an OCF distribution k cannot be provided over a set of interpretations Ω (except if the number of propositional variables is small). A compact representation may be provided using for instance the concept of OCF networks [KE13, BT10, GP13, DG94, EK14] or the concept of weighted propositional knowledge bases.

Link to possibility theory

It is easy to see (cf. [DP13]) that the set function N_k defined by $N_k(\phi) = 1 - e^{-k(\neg\phi)}$ is a necessity measure, with values in a subset of the unit interval. Moreover because $k(\phi) \in \mathbb{N}$, $N_k(\phi) < 1, \forall\phi \neq \Omega$. The set $\{k(\omega)|\omega \in \Omega\}$ is the counterpart of a possibility distribution π on Ω, such that $\Pi(\phi) = \max\{\pi(\omega)|\omega \in \phi\}$.

Conclusion

Standard probability theory is the most used framework to represent uncertainty but it is always not suitable to represent every type of uncertain information. Therefore, alternative frameworks such that possibility theory, imprecise probability theory, belief functions theory or ordinal conditional functions have been proposed. Some of these formalism can be seen as extensions of probability theory such as imprecise probability. Others are devoted to a totally different kind of uncertainty. In general, theories can be linked to one another, we will see later examples of such links between probability and possibility.

Uncertainty formalism have also practical problems as they entail higher storage space and higher computational time complexity. Graphical models and logical-based models have been proposed to tackled such complexity problems. This is the goal of the next chapter.

Chapter 2

Compact uncertainty representations

Dealing with a lot of information and variables can induce higher space and time complexity. To efficiently represent and reason with uncertain information, we use compact representations. In the previous chapter, we introduced uncertain information and the various means to express them using uncertainty frameworks. Though representing information with distributions is not suitable as soon as the number of variables or the universe of discourse is too large. Clearly, representing real and complex situations implying a large amount of information requires compact representations.

Graphical models in that sense offer a nice and efficient alternative to represent and store uncertain information. Indeed, graphical models are *interpretable* as one can see the variables distinctively and see the dependence relations (e.g. relation of cause and effect). They are *modular*, indeed one can elicit and infer using the graph structure of the model. Lastly, they are *compact* as the number of entries (stored in local tables) can be significantly smaller than in the joint distribution. This solution allows to reduce space complexity but also in terms of reasoning offers a mean of representation that allows to efficiently reason with uncertain information.

Another compact representation often used to represent information is knowledge bases. The notion of uncertainty was introduced by adding weight to logical formulas expressing the degrees of certainty of the formulas.

In this chapter, we discuss the two different representations of uncertain information, by means of graphical models and by means of knowledge bases.

Contents

2.1 Graphical models

Graphs are important tools for building probabilistic and other models used in artificial intelligence. There exist two main types of graphs that have been used in uncertainty-based graphical models: *undirected* and *directed* models. In this thesis we only focus on *directed* graph, and we present their most important properties [2]. In the following, we get into details on the networks used to represent uncertain information discussed in the previous chapter. The syntactic definition of a graph is the same. The difference lies in the associated semantics.

In the rest of this chapter, the used notations are as follows:

2. Basics notions and definitions on graph theory can be found in Appendix A

— we denote respectively a Bayesian network, a possibilistic network, a credal network and an interval-based possibilistic network by $\mathcal{BN}$, $\mathcal{PN}$, $\mathcal{CN}$ and IPN,

— $\Theta_p, \Theta_\pi, \Theta_{IP}$ and $\Theta_{I\Pi}$ denote respectively the set of local distributions associated to $\mathcal{BN}, \mathcal{PN}, \mathcal{CN}$ and IPN.

— in a knowledge base, we define $\mathcal{L}$ as the language composed of propositional formulas.

2.1.1 Bayesian networks

The term *Bayesian networks* [Pea88] was coined by Judea Pearl [Pea89]. They are sometimes referred to as *probabilistic networks*, *belief networks*, *causal networks* or *DAG models*. A Bayesian network highlights three aspects [Pea89, Bay63]: the information represented in these networks; the inference machinery; and how they can support both causal and evidential reasoning.

Definition 2.1 (Bayesian network)**.** A *Bayesian network* $\mathcal{BN} = (G, \Theta_p)$ is defined by:

— *Graphical component*: a DAG $G = \{V, E\}$ where each node of the graph represents a variable of interest and the edges describe the (in)dependence relation among the variables.

— *Numerical component*: a set of local probability distributions Θ_p, for each variable X_i given its parents $par(X_i)$.

Local probability distributions must satisfy the following *normalization* conditions:

— If $par(X_i) = \emptyset$ (X_i is a root node), then the marginal distribution associated to X_i has to satisfy:

$$\sum_{x_i \in D_{X_i}} P(x_i) = 1 \tag{2.1}$$

— If $par(X_i) \neq \emptyset$ then the conditional distributions associated to X_i have to satisfy:

$$\forall u_i \in D_{par(X_i)}, \sum_{x_i \in D_{X_i}} P(x_i|u_i) = 1 \tag{2.2}$$

where $D_{par(X_i)}$ represents the domain associated to the parents of X_i.

Example 2.1. The following figure illustrates a Bayesian network $\mathcal{BN} = (V, E)$, with $V = \{A, B, C\}$ with domains respectively being $D_A = \{a_1, a_2\}$, $D_B = \{b_1, b_2\}$ et $D_C = \{c_1, c_2\}$.

A Bayesian network compactly encodes a joint distribution that can computed using the chain rule:

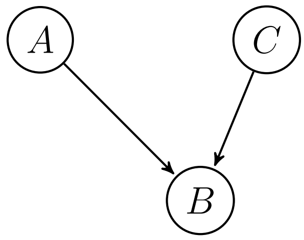
Definition 2.2 (Chain rule)**.** Given $X_1, ..., X_n$ a set of random variables. A Bayesian network over the set of variables $V = \{X_1, ..., X_n\}$ allows to encode a probability distribution factorized as follows:

$$P(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i|par(X_i)) \tag{2.3}$$

Remark. If all the local tables are normalized then the joint distribution is also normalized.

Example 2.2. The joint probability distribution encoded by the Bayesian network of Figure 2.1, computed using Equation (2.3), is given by Table 2.1.

A	$p(A)$
a_1	0.1
a_2	0.9



C	$p(C)$
c_1	0.6
c_2	0.4

B	A	C	$p(B \mid AC)$
b_1	a_1	c_1	0.2
b_2	a_1	c_1	0.8
b_1	a_1	c_2	0.7
b_2	a_1	c_2	0.3
b_1	a_2	c_1	0.5
b_2	a_2	c_1	0.5
b_1	a_2	c_2	0.4
b_2	a_2	c_2	0.6

Figure 2.1 – Example of a Bayesian network

A	B	C	$p(A, B, C)$
a_2	b_2	c_2	0.216
a_2	b_2	c_1	0.27
a_2	b_1	c_2	0.144
a_2	b_1	c_1	0.27
a_1	b_2	c_2	0.012
a_1	b_2	c_1	0.048
a_1	b_1	c_2	0.028
a_1	b_1	c_1	0.012

Table 2.1 – Joint probability distribution of the Bayesian network given by Figure 2.1

2.1.2 Possibilistic networks

As Bayesian networks are the main compact graphical representation for probability theory, possibilistic networks [BK02] are graphical representations of possibility theory.

Definition 2.3 (Possibilistic network)**.** A possibilistic network $\mathcal{PN} = (G, \Theta_\pi)$ is defined by two component, a graphical one and a numerical one.

- the *graphical component* is a directed acyclic graph $G = (V, E)$ on the set of variables V where the set of edges E represent the (in)dependence relationships between variables.
- the *numerical component* is a set of local possibility distributions $\Theta_\pi = \{\pi(X_i|par(X_i)), \forall X_i \in V\}$ of variables X_i given its parents $par(X_i)$.

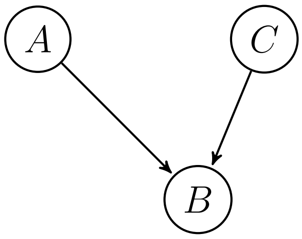
In the same way as Bayesian networks, the conditional distributions must satisfy normalizing condition.

$$\forall u_j \in D_{par(X_i)}, \quad \max_{x_i \in D_{X_i}} \pi(x_i|u_j)) = 1 \tag{2.4}$$

Example 2.3. Let us consider 3 variables A, B and C having as domains respectively $D_A = \{a_1, a_2\}$, $D_B = \{b_1, b_2\}$ and $D_C = \{c_1, c_2\}$. An example of possibilistic network is illustrated by Figure 2.2.

We distinguish two types of possibilistic networks: quantitative possibilistic networks corresponding to the numerical interpretation of the possibilistic scale and qualitative possibilistic networks corresponding to the ordinal interpretation. It is known that quantitative possibilistic networks are close to

A	$\pi(A)$
a_1	0.5
a_2	1



C	$\pi(C)$
c_1	1
c_2	0.1

B	A	C	$\pi(B \mid AC)$
b_1	a_1	c_1	0.8
b_2	a_1	c_1	1
b_1	a_1	c_2	1
b_2	a_1	c_2	0.7
b_1	a_2	c_1	1
b_2	a_2	c_1	0.4
b_1	a_2	c_2	0.6
b_2	a_2	c_2	1

Figure 2.2 – Example of a possibilistic network

the Bayesian networks since they share the same characteristics (essentially, the product operator) with theoretical and practical results almost identical.

Therefore, two definitions of the chain rule depending whether we use the quantitative scale or the qualitative scale.

Definition 2.4 (Quantitative chain rule)**.** The joint possibility distribution π associated to a quantitative possibilistic network on a set of variables $V = \{X_1, ..., X_n\}$ is computed in the same way as for Bayesian network using the product operator.

$$\pi(X_1, ..., X_n) = \prod_{i=1}^{n} \pi(X_i|par(X_i)) \tag{2.5}$$

where $\pi(X_i|par(X_i))$ is the local possibility distribution associated to the local variable X_i in the context of its parents.

Definition 2.5 (Qualitative chain rule)**.** The joint possibility distribution π associated to a qualitative possibilistic network on a set of variables $V = \{X_1, ..., X_n\}$ is computed using the $\min$ operator. Thus, the chain rule is given by:

$$\pi(X_1, ..., X_n) = \min_{i=1,...,n} \pi(X_i|par(X_i)) \tag{2.6}$$

Example 2.4. Given the possibilistic network of Figure 2.2, the joint possibility distribution defined by $\pi(A, B, C) = \min(\pi(A), \pi(B \mid AC), \pi(C))$ is given by Table 2.3. The joint possibility distribution defined by $\pi(A, B, C) = \pi(A) * \pi(B \mid AC) * \pi(C)$ is given by Table 2.2.

2.1.3 Credal networks

A credal network [Coz05, Coz00] is a graphical model that associates 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.

Definition 2.6 (Credal network)**.** A credal network $\mathcal{CN} = (G, \Theta_{Ip})$ is seen as an extension of a Bayesian network is also composed of:

A	B	C	$\pi(A,B,C)$
a_2	b_2	c_2	0.1
a_2	b_2	c_1	0.4
a_2	b_1	c_2	0.06
a_2	b_1	c_1	1
a_1	b_2	c_2	0.035
a_1	b_2	c_1	0.5
a_1	b_1	c_2	0.05
a_1	b_1	c_1	0.4

Table 2.2 – Possibility joint distribution of Figure 2.2 in the quantitative possibilistic setting

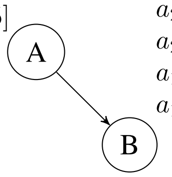
A	B	C	$\pi(A,B,C)$
a_2	b_2	c_2	0.1
a_2	b_2	c_1	0.4
a_2	b_1	c_2	0.1
a_2	b_1	c_1	1
a_1	b_2	c_2	0.1
a_1	b_2	c_1	0.5
a_1	b_1	c_2	0.1
a_1	b_1	c_1	0.5

Table 2.3 – Possibility joint distribution of Figure 2.2 in the qualitative possibilistic setting

- a *graphical component*: a DAG G with nodes representing variables and edges describing independence relationships
- a *numerical component*: a set of conditional credal sets, which consists in our case in imprecise probability distributions (intervals).

Example 2.5. Figure 2.3 gives an example of a credal network $\mathcal{CN}$ on two boolean variables A and B.

A	$Ip(A)$
a_2	$[.5, .9]$
a_1	$[.1, .5]$



A	B	$Ip(B\|A)$
a_2	b_2	$[.36, .4]$
a_2	b_1	$[.35, .35]$
a_1	b_2	$[.25, .25]$
a_1	b_1	$[0, .04]$

Figure 2.3 – Example of a credal network

The semantics associated to a credal network is a family of Bayesian networks, therefore we define the concept of compatible Bayesian network:

Definition 2.7 (Compatible Bayesian network)**.** A Bayesian network $\mathcal{BN} = (G, \Theta_p)$ is compatible with $\mathcal{CN} = (G, \Theta_{Ip})$ has the same graph structure and for each local probability distribution $p \in \Theta_p$ satisfy

$$\forall x_i \in D_{X_i|par(X_i)},\ p(x_i|par(x_i)) \in Ip(x_i|par(x_i)) \tag{2.7}$$

According to this semantics, a credal network $\mathcal{CN}$ encodes a set of joint probability distributions, called extensions and denoted $ext(\mathcal{CN})$, where each joint distribution $p \in ext(\mathcal{CN})$ is encoded by a compatible Bayesian network. A credal network $\mathcal{CN}$ induces a strong extension defined as follows:

Definition 2.8 (Strong extension)**.** The strong extension $ext_S(\mathcal{CN})$ associated with a credal network $\mathcal{CN}$ is the convex hull of all the joint probability distributions $ext(\mathcal{CN})$ encoded by the Bayesian networks $\mathcal{BN}$ compatible with $\mathcal{CN}$. Namely,

$$ext_S(\mathcal{CN}) = CH(\{P_{\mathcal{BN}} : \ \mathcal{BN} \text{ is compatible with } \mathcal{CN}\}). \tag{2.8}$$

Given an extension $ext(\mathcal{CN})$, one can compute an interval-based joint probability distribution as follows:

$$\underline{P}(a_1a_2..a_n) = \min_{p \in ext(\mathcal{CN})}(p(a_1a_2..a_n)) \tag{2.9}$$

$$\overline{P}(a_1a_2..a_n) = \max_{p \in ext(\mathcal{CN})}(p(a_1a_2..a_n)) \tag{2.10}$$

In Equations (2.9) and (2.10), $p(a_1a_2..a_n)$ is computed with the well-known chain rule in Bayesian networks (see Equation (2.3)). Note that the vertices of $ext(\mathcal{CN})$ can be obtained by considering only the set of vertices of the local credal sets K_i associated with the variables [Coz00].

This section presents all graphical models used in this thesis and how to compute the joint distribution or set of distributions given the used semantic. In the next section we present another kind of compact representations.

2.2 Possibilistic knowledge bases

In this section, we are particularly interested in possibilistic logic, a different compact representation often used to encode a possibility distribution. As possibilistic logic is a logic-based framework, some brief refresher on propositional logic is given in Appendix A.

Possibilistic logic [Ben10, DP04, Lan00] allows to handle uncertain information based on possibility theory. Beliefs are expressed by means of uncertain propositional formulas. The uncertainty is quantified by associating to each formula a degree of certainty corresponding to the minimal threshold of necessity degree. As we have seen in Chapter 1, possibility theory can be interpreted two ways, one quantitatively and the other one is qualitatively according to the scale used to represent the uncertainty. Therefore, an agent can express his beliefs using either contexts using quantitative possibilistic logic or qualitative possibilistic logic.

2.2.1 Possibilistic knowledge bases: syntax

In possibilistic logic, at a syntactic level information is encoded in a possibilistic knowledge base which is a finite set of weighted formulas defined as

$$K = \{(\varphi_i, \alpha_i) : i = 1, ..., n\}, \tag{2.11}$$

where the pair (φ_i, α_i) is a possibilistic formula with φ_i is an element of $\mathcal{L}$ and $\alpha_i \in [0, 1]$ is a valuation of φ_i representing the minimal necessity degree $N(\varphi_i)$. The necessity degree α_i represents the degree of certainty of φ_i of the knowledge base K encoding the beliefs of an agent upon the real state of the world, or its degree of priority if K encodes the preferences of an agent. Indeed, the formula φ_i is said to be more certain (resp. more prioritized) than ϑ if the weight associated to φ is greater than the one associated to ϑ.

2.2.2 From possibilistic knowledge bases to possibility distributions

A possibilistic knowledge base is one of well-known compact representations of a possibility distribution. Given a possibilistic base K, we can generate a unique possibility distribution where interpretations ω satisfying all propositional formulas in K have the highest possible degree $\pi(\omega) = 1$ (since they are fully consistent), whereas the others are pre-ordered with respect to the highest formulas they falsify. More formally: $\forall \omega \in \Omega$,

$$\pi_{K_\pi}(\omega) = \begin{cases} 1 & \text{if } \forall(\varphi, \alpha) \in K,\ \omega \models \varphi \\ 1 - \max\{\alpha_i : (\varphi_i, \alpha_i) \in K,\ \omega \not\models \varphi_i\} & \text{otherwise.} \end{cases} \tag{2.12}$$

The following lemma states that 'zero-weighted' formulas can be added or removed from possibilistic knowledge bases without changing the joint distributions.

Lemma 2.1. Let K be a possibilistic knowledge base K such that $(\delta, 0) \in K$. Let $K' = K \setminus \{(\delta, 0)\}$. Then $\forall \omega \in \Omega$, $\pi_K(\omega) = \pi_{K'}(\omega)$.

This lemma can be easily shown since if a formula δ has a certainty degree equal to 0, then if there is an interpretation ω that falsifies only the formula δ then, according to Equation (2.12), the possibility degree associated to ω will be $1 - 0 = 1$.

In the next subsection, we present a transformation of a possibilistic network into a possibilistic knowledge base.

2.2.3 From possibilistic networks to possibilistic knowledge bases

We consider that two representation formats are semantically equivalent when they encode the same joint distributions. In the possibilistic setting, transformation from possibilistic networks to possibilistic knowledge bases have been proposed. In [BDGP02], the authors proposed two definitions to translate a possibilistic graph into a base. Indeed, the two possibilistic networks are considered using either the qualitative chain rule or the quantitative chain rule, this leads to two definitions. The authors also discussed the converse translation.

Starting point

The possibilistic knowledge base associated with a possibilistic network results from a combination of elementary bases. These elementary bases are associated to each local distribution of the network. They are composed of weighted formulas where the weight refers to the conditional possibility degrees of the node variable.

Let $\{X_1, ..., X_n\}$ be the set of variables of the possibilistic network $\mathcal{PN}$. We describe $\mathcal{PN}$ as a set of triples such as $\mathcal{PN} = \{(x_i, par(x_i), \alpha_i) : \alpha_i = \Pi(x_i | par(x_i)) \neq 1$ is an element of the graph$\}$, where x_i is a value of the variable X_i and $par(x_i)$ is a configuration of the parents of X_i. Note that we restrict to degrees different from 1 since only those are used in the computation of the joint distributions. Indeed, 1 is a neutral element with respect to both minimum and product operator.

Example 2.6. Let us consider the following possibilistic network given by Figure 2.4 over 3 binary variables. As said in the above the starting point is to translate the network in the form of a set of triples.

Given the possibilistic network of Figure 2.4, the set of triples is: $\mathcal{PN}_A = \{(a_2, \emptyset, .45)\}, \mathcal{PN}_B = \{(b_2, \emptyset, .78)\}, \mathcal{PN}_C = \{(c_1, a_2b_1, .33), (c_1, a_2b_2, .1), (c_2, a_1b_2, .6)\}$.

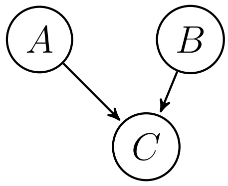
We associate to each triple $(x, par(x), \alpha)$ a possibilistic formula $(\neg x \vee \neg par(x), 1 - \alpha)$. And we can retrieve the possibility distribution $\pi_{x,par(x)}$ of the value x given a configuration of $par(x_i)$ with

$$\pi_{x,par(x)}(\omega) = \begin{cases} 1 & \text{if } \omega \models \neg x \vee \neg par(x), \\ \alpha & \text{otherwise (namely if } \omega \models x \wedge par(x)). \end{cases} \tag{2.13}$$

Example 2.7. With the previous example and the translation obtained from $\mathcal{PN}$, we associate to each triple the possibilistic formula, which gives us the set of weigthed formulas: $K_A = \{(a_1, .55)\}$, $K_B = \{(b_1, .22)\}$ and $K_C = \{(c_2 \vee a_1 \vee b_2, .67), (c_2 \vee a_1 \vee b_1, .9), (c_1 \vee a_2 \vee b_1, .4)\}$.

Now that we have obtained for each variable of the graph the associated possibilistic knowledge base, we can consider how to combine these knowledge bases. There are two methods given the possibilistic scale used.

A	$\pi(A)$
a_1	1
a_2	.45



B	$\pi(B)$
b_1	1
b_2	.78

C	A	B	$\pi(C \mid AB)$
c_1	a_1	b_1	1
c_1	a_1	b_2	1
c_1	a_2	b_1	.33
c_1	a_2	b_2	.1
c_2	a_1	b_1	1
c_2	a_1	b_2	.6
c_2	a_2	b_1	1
c_2	a_2	b_2	1

Figure 2.4 – Possibilistic network to be transformed in a knowledge base

Logical encoding of $\mathcal{PN}_m$

We first consider qualitative possibilistic networks where the conditioning rule use the minimum operation.

Definition 2.9 (Combination of two possibilistic knowledge bases)**.** Let Σ_1 and Σ_2 be two possibilistic knowledge bases. Let π_1 and π_2 be the two joint possibility distributions associated with Σ_1 and Σ_2 respectively. Let π_m be the combination of π_1 and π_2 using the minimum operation. Then, $\forall \omega,\ \pi_m(\omega) = \min(\pi_1(\omega), \pi_2(\omega))$. And the resulting base corresponding to π_m is: $\Sigma_m = \Sigma_1 \cup \Sigma_2$.

This leads us to the following Definition 2.10 that states formally how to compute the knowledge base associated to a qualitative possibilistic network.

Definition 2.10 (Possibilistic knowledge base associated to $\mathcal{PN}_m$)**.** The possibilistic knowledge base associated with $\mathcal{PN}_m = \{(x_i, par(x_i), \alpha_i) : \alpha_i = \Pi(x_i|par(x_i)) \neq 1\}$ is given by:

$$\Sigma_{\mathcal{PN}_m} = \{(\neg x_i \vee \neg par(x_i), 1 - \alpha_i) : (x_i, par(x_i), \alpha_i) \in \mathcal{PN}_m\}. \tag{2.14}$$

This is an important result that implies that we can apply possibilistic logic reasoning on possibilistic networks.

Example 2.8. Following the different knowledge bases of variables A, B and C given in the previous example by $K_A = \{(a_1, .55)\}$, $K_B = \{(b_1, .22)\}$ and $K_C = \{(c_2 \vee a_1 \vee b_2, .67), (c_2 \vee a_1 \vee b_1, .9), (c_1 \vee a_2 \vee b_1, .4)\}$, we can combine them using Definition 2.10 and it results in:

$$K_{\mathcal{PN}_m} = \{(a_1, .55), (b_1, .22), (c_2 \vee a_1 \vee b_2, .67), (c_2 \vee a_1 \vee b_1, .9), (c_1 \vee a_2 \vee b_1, .4)\}$$

.

Logical encoding of $\mathcal{PN}_*$

In the same way that we translate a qualitative possibilistic network into a possibilistic knowledge base, it is also possible to encode possibilistic networks in the quantitative setting based on the product operator.

The following combination operator that we present is commutative and associative, therefore we can apply the combination in any order.

Definition 2.11 (Combination of two possibilistic bases with the product)**.** Let $\Sigma_1 = \{(\phi_i, \alpha_i) : i \in I\}$ and $\Sigma_2 = \{(\psi_j, \beta_j) : j \in J\}$. Let π_1 and π_2 be the two joint possibility distributions associated with Σ_1 and Σ_2 respectively. Let π_* be the combination of π_1 and π_2 using the product operation. The resulting base associated to π_* is

$$\mathcal{C}_*(\Sigma_1, \Sigma_2) = \Sigma_1 \cup \Sigma_2 \cup \{(\phi_i \vee \psi_i, \alpha_i + \beta_i - \alpha_i * \beta_j) : i \in I,\ j \in J,\ \phi_i \vee \psi_i \neq \top\}. \tag{2.15}$$

Then the definition of the translation of a possibilistic network into a possibilistic knowledge base is given by:

Definition 2.12 (Possibilistic knowledge base associated to $\mathcal{PN}_*$)**.** The knowledge base associated with the graph $\mathcal{PN}_*$ is the combination by $\mathcal{C}_*$ of the knowledge bases associated to each node of the graph.

Example 2.9. Let us pick up Example 2.6, with the encoding of the network given by $K_A = \{(a_1, .55)\}$, $K_B = \{(b_1, .22)\}$ and $K_C = \{(c_2 \vee a_1 \vee b_2, .67), (c_2 \vee a_1 \vee b_1, .9), (c_1 \vee a_2 \vee b_1, .4)\}$. The approach here is different from the qualitative setting, we first combine K_A and K_B and obtain $K_{AB} = \{(a_1, .55), (b_1, .22), (a_1 \vee b_1, .649)\}$. Then we only have to combine the new knowledge base K_{AB} with K_C. Thus,

$$K_{ABC} = K_{AB} \cup K_C \cup \{(c_2 \vee a_1 \vee b_2, .8515), (c_1 \vee a_2 \vee b_1, .532), (c_2 \vee a_1 \vee b_1, .9649)\}$$

There exists also the converse translation [BDGP02] but since it is not needed to understand our contributions, we will not present it here.

2.3 Reasoning with uncertain information

We address in this section the problem that can arise in real-world applications, showing how each can be solved by the models defined above. There are at least four general types of queries that can be posed with respect to graphical models. Which type of query to use in a particular situation is not always trivial and some of the queries are guaranteed to be equivalent under certain conditions. We define formally these query types in the following. As most of the work done with regards to querying in graphical models has been done in the Bayesian networks, we will illustrate the following using a Bayesian network.

2.3.1 Querying graphical models

In general, answering queries on graphical models is about taking into account an evidence where one may want to compute the degree of a given event or compute the degree of a variable given an event and so on. Therefore, we denote E the set of variables called *evidence variables* and Q the set of query variables. And an evidence is denoted by the event $\phi \in D_E$ where the domain associated to E is the cartesian product of the domains of every variable in E.

Probability of evidence

One of the simplest queries is to ask for the degree or interval-degree of some variable instantiation e. For example, in the Asia network (Figure A.3 in appendix A [3]) we may be interested in knowing the probability that the patient has a positive x-ray but no dyspnoea, $P(X = yes, D = no)$. This can be computed easily by tools for modeling and reasoning with graphical models like JavaBayes [Coz01].

3. Indeed, for more information and details on queries in graphical models, see Appendix A

The query $P(\phi)$ is known as a *probability-of-evidence* query, although it refers to a very special type of evidence corresponding to the instantiation of some variables.

There are other types of evidence beyond variable instantiations. In fact, any propositional sentence can be used to specify evidence. For example, we may want to know the probability that the patient has either a positive x-ray or a dyspnoea, $X = yes \vee D = yes$. In general, the existing tools do not provide direct support for computing the probability of arbitrary pieces of evidence but such probabilities can be computed indirectly using some methods and tricks like the "auxiliary node" trick.

Prior and posterior marginals

If probability-of-evidence queries are one of the simplest, then *posterior-marginal queries* are one of the most common. The difference between *prior* and *posterior marginals* is that a prior marginal is a marginal distribution given no evidence. And the posterior marginal distribution is computed given some evidence ϕ,

$$p(x_1, ..., x_m|\phi) = \sum_{x_{m+1},\cdots,x_n} p(x_1, ..., x_n|\phi). \tag{2.16}$$

Computing posterior marginals comes down to summing probability-of-evidence.

The complexity of such query in the general case is proven to be NP-Hard [Coo90] in Bayesian networks.

Most probable explanations

We now turn to another class of queries: computing the *most probable explanation* (MPE). The goal here is to identify the most probable instantiation of network variables given some evidence. Specially, if $X_1, ..., X_n$ are all the network variables and if ϕ is the given evidence, the goal then is to identify an instantiation $x_1, ..., x_n$ for which the probability $p(x_1, ..., x_n|\phi)$ is maximal. Such full instantiation $x_1, ..., x_n$ will be called a *most probable explanation* given evidence ϕ. More formally, the answer is

$$argmax_{x_1,..,x_n \in D_\Omega}(p(x_1, ..., x_n|\phi)) \tag{2.17}$$

Maximum a posteriori hypothesis (MAP)

The MPE query is a special case of a more general class of queries for finding the most probable instantiation of a subset of network variables given some evidence. Specially, suppose that the set of all network variables is V and let Q be a subset of these variables. Given some evidence ϕ, our goal is then to find an instantiation q of variables Q for which the probability $p(q|\phi)$ is maximal. Any instantiation q that satisfies the previous property is known as a *maximum a posteriori hypothesis* (MAP). Clearly, MPE is a special case of MAP when the MAP variables include all network variables. One reason why a distinction is made between MAP and MPE is that MPE is much easier to compute. More formally, the answer to a MAP query is

$$argmax_{q \in D_Q}(p(q|\phi)) \tag{2.18}$$

2.3.2 Reasoning in probabilistic graphical models

We have just presented the four types of queries that are mainly used when reasoning with uncertain information. One of the simplest methods for inference in graphical models such as Bayesian networks is based on the principle of variable elimination. A process in which we successively remove variables from the graph while maintaining its ability to answer queries of interest. Another algorithm widely used in Bayesian networks is the junction tree algorithm [Jen96]. These algorithms are detailed in Appendix A.3 for more information.

As said previously, inference in probabilistic graphical models generally consists in computing the probability of an event. In credal networks, this equivalently comes down to computing lower or upper probabilities of an event. For **M**aximum **A** **P**osteriori (MAP), given an assignment ϕ of evidence variables E, the objective is to compute the most probable instantiation q of the query variables Q. In general, $Q \cap E = \emptyset$.

More formally, the MAP inference problem in credal networks comes down to compute:

$$argmax_{q \in D_Q}(IP(q \mid \phi)) \tag{2.19}$$

where $argmax$ denotes a decision criterion allowing to choose the set of "most probable configurations" of query variables. In the following, we will give some of the most used decision criteria that can be used for answering MAP queries in credal networks.

We need decision criteria to answer MAP queries in credal networks due to the representation by means of intervals. A natural criterion is the one of *Interval-dominance* (used for instance in [ACZ10] for classification, decision tasks, etc.) which refers to non-dominated instantiations of query variables.

Definition 2.13 (*Interval-dominance*)**.** An instantiation q_i of query variables $\mathcal{Q}$ dominates another instantiation q_j if and only if $\underline{IP}(q_i|\phi) > \overline{IP}(q_j|\phi)$ where ϕ is an instantiation of observation variables $\mathcal{O}$.

This criterion is not enough informative and often of little use in practice. Indeed, this criterion often results in large amounts of outcomes also called answer set (too many query variable instances are not dominated), making it difficult to make decisions for instance in classification where only one class (outcome) should be returned. The other possible criteria are the well-known criteria *Maximax*, *Maximin*, and *Hurwicz*. These three criteria are commonly used in decision making under uncertainty since the early 1950's. The *Maximax* criterion can be viewed as an optimistic criterion. It examines the maximum payoffs of alternatives and chooses the alternative whose outcome is the best. Definition 2.14 gives a formal definition of *Maximax* criterion for the imprecise probability setting.

Definition 2.14 (*Maximax* criterion)**.** An instantiation q_i of query variables $\mathcal{Q}$ is a result of MAP inference if and only if $\overline{IP}(q_i|\phi) \geq \max\{1 - \sum_{q_j \neq q_k} \underline{IP}(q_j|\phi), \forall q_k\}$, where ϕ is an instantiation of observed variables $\mathcal{O}$.

The *Maximin* criterion also known as the Wald's *Maximin* criterion is a pessimistic criterion. It suggests that the decision maker examines only the minimum payoffs of alternatives and chooses the alternative whose outcome is the least worst.

Definition 2.15 (*Maximin* criterion)**.** An instantiation q_i of query variables $\mathcal{Q}$ is a result of MAP inference if and only if $\underline{IP}(q_i|\phi) \geq \max\{1 - \sum_{q_j \neq q_k} \overline{IP}(q_j|\phi), \forall q_k\}$, where ϕ is an instantiation of observed variables $\mathcal{O}$.

The last criterion we review is the well-known *Hurwicz* criterion which attempts to find a trade-off between the extremes, posed by the optimistic and pessimistic criteria, by assigning a certain weight, a to optimism and the balance $1 - a$ to pessimism. This index reflects the decision maker attitude towards risk taking. A cautious decision maker will set $a = 1$ which reduces the *Hurwicz* criterion to the *Maximin* criterion. An adventurous decision maker will set $a = 0$ which reduces the *Hurwicz* criterion to the *Maximax* criterion.

Definition 2.16 formally defines the *Hurwicz's* criterion with imprecise probabilities using the coefficient $a = 0.5$.

Definition 2.16 (*Hurwicz's* criterion)**.** Let q_i be an instantiation of query variables $\mathcal{Q}$ with o an instantiation of observed variables of $\mathcal{O}$, $a = \{0.5 * (1 - \sum_{q_i \neq q_k} \overline{IP}(q_i|\phi)) + 0.5 * (1 - \sum_{q_i \neq q_k} \underline{IP}(q_i|\phi)), \forall q_k\}$. Then q_i is a result of *MAP* inference if and only if $a = \max_{\forall q_j} \{0.5 * (1 - \sum_{q_j \neq q_k} \overline{IP}(q_j|\phi)) + 0.5 * (1 - \sum_{q_j \neq q_k} \underline{IP}(q_j|\phi)), \forall q_k\}$.

Example 2.10. Let us show an example of these criteria over the following imprecise distribution (Table 2.4).

ω	$IP(\omega)$
ω_1	$[.25; .3]$
ω_2	$[.27; .32]$
ω_3	$[.26; .33]$
ω_4	$[.07; .12]$

Table 2.4 – Example of an imprecise distribution

On this example, clearly the only world that can be excluded from the results with the interval-dominance criterion is ω_4. The outcomes that can be obtained using the different criteria are listed in the following items.

— *Interval-dominance*: Answer set = $\{\omega_1, \omega_2, \omega_3\}$,

— *Maximax*: Answer set = $\{\omega_3\}$,

— *Maximin*: Answer set = $\{\omega_2\}$,

— *Hurwicz*: Answer set = $\{\omega_2, \omega_3\}$.

As shown in this example, one of the main problems of *MAP* inference in credal networks is that the number of outcomes may be very large especially when the interval-dominance criterion is used. The second big problem is the one of computational complexity of *MAP* inference in credal networks.

Imprecise probability theory as we described is a necessity regarding the lack of information that we can face in some applications.But dealing with such expressive frameworks requires a lot more effort for representation and reasoning purpose. This is particularly due to the number of compatible credal networks that we have to consider.

The following table summarizes complexity results of inference in Bayesian and credal networks [MdCBA14, dCC05].

	Query	Polytree	Bounded treewidth	Multiply-connected
Bayesian Networks	Pr	Polynomial	Polynomial	PP-Complete
	MPE	Polynomial	Polynomial	NP-Complete
	MAP	NP-Complete	NP-Complete	NP^{PP}-Complete
Credal Networks	Pr	NP-Complete	NP-Complete	NP^{PP}-Complete
	MPE	Polynomial	Polynomial	NP-Complete
	MAP	Σ_2^P-Complete	Σ_2^P-Complete	NP^{PP}-Hard

This table shows in particular the very high complexity of *MAP* inference in $\mathcal{CN}$s in the general case. In practice, the size of networks is often large. This motivates approximate inference approaches and in this thesis, we provide a kind of approximate inference method for *MAP* in $\mathcal{CN}$s by transforming the credal network into a possibilistic one $\mathcal{PN}$.

There exist multiple algorithms to perform the different tasks such as the elimination variable or junction tree, but in some frameworks such as credal networks or possibilistic networks there is not a lot of implementations of the algorithm. Nevertheless, some implementation as GL2U or JavaBayes provides some features for credal network and we actually base an experimental study on this platform to test our approach, this contribution is given in Chapter 8.

2.3.3 Reasoning in possibilistic knowledge bases

An important notion that plays a central role in the inference process in possibilistic knowledge bases is the one of strict α-cut. Let α be a positive real number. A strict α-cut, denoted by K_α, is a set of propositional formulas defined by $K_\alpha = \{\varphi : (\varphi, \beta) \in K \text{ and } \beta > \alpha\}$. The strict α-cut is useful to measure the inconsistency degree of K denoted by $\mathit{Inc}(K)$ and defined by:

$$\mathit{Inc}(K) = \begin{cases} 0 & \text{if } K_0 \text{ is consistent} \\ \max\{\alpha : K_\alpha \text{ is inconsistent}\} & \text{otherwise} \end{cases} \tag{2.20}$$

If $\mathit{Inc}(K) = 0$ then K is said to be completely consistent. If a possibilistic base K is partially inconsistent, then $\mathit{Inc}(K)$ can be seen as a threshold below which every formula is considered as not enough entrenched to be taken into account in the inference process. More precisely, we define the notion of core of knowledge base as composed of formulas with a certainty degree greater than $\mathit{Inc}(K)$, namely:

$$\mathit{Core}(K) = K_{\mathit{Inc}(K)} = \{\varphi : (\varphi, \alpha) \in K \text{ and } \alpha > \mathit{Inc}(K)\}$$

A formula φ is a consequence of a possibilistic base K, denoted by $K \vdash_\pi \varphi$, if and only if $\mathit{Core}(K) \vdash \varphi$, *i.e.*, if the formula φ is a classical consequence of the core of K.

Given the possibility distribution induced from the possibilistic knowledge base, the following completeness and soundness result holds:

$$K \vdash_\pi \varphi \text{ if and only if } \pi_K \models_\pi \varphi. \tag{2.21}$$

Example 2.11. We here give an example to illustrate the previous definitions. Let $K = \{(a \vee b, .7), (\neg a, .6), (\neg b, .2)\}$ be a possibilistic knowledge base.

In this case, $\mathit{Inc}(K) = .2$ and thus $\mathit{Core}(K) = \{a \vee b, \neg a\}$ and, for instance, $K \vdash_\pi \neg a \wedge b$.

From the semantic point of view, the possibility distribution is the following:

$$\begin{array}{lclclcl} \pi_K(\neg a \neg b) & = & .3 & \quad & \pi_K(\neg a b) & = & .8 \\ \pi_K(a \neg b) & = & .4 & \quad & \pi_K(ab) & = & .4 \end{array} \tag{2.22}$$

and then $\Pi_K(\neg a \wedge b) = .8 > \Pi_K(a \vee \neg b) = .4$ which implies $\pi_K \models_\pi \neg a \vee b$.

The concept of α-cut can be used to provide the syntactic counterpart of conditioning a possibilistic knowledge base with a propositional formula:

Definition 2.17. Let K be a possibilistic knowledge and ϕ be a sure piece of information. The result of conditioning K by ϕ, denoted K_ϕ is defined as follows:

$$\begin{aligned} K_\phi = & \{(\phi, 1)\} \cup \\ & \{(\varphi, \alpha) : (\varphi, \alpha) \in K \text{ and } K_{\geq \alpha} \wedge \phi \text{ is consistent.}\} \end{aligned}$$

Namely, K_ϕ is obtained by considering ϕ with a certainty degree '1', plus weighted formulas (φ, α) of K such that their α-cut is consistent with ϕ. It can be checked that:

$$\forall \omega \in \Omega,\ \pi_{K_\phi}(\omega) = \pi_K(\omega |_\otimes \phi),$$

where π_K and π_{K_ϕ} are given using Equation (2.12) and $\pi_K(. |_\otimes \phi)$ is obtained using the min-based conditioning of Equation (1.29) or the product-based conditioning of Equation (1.28).

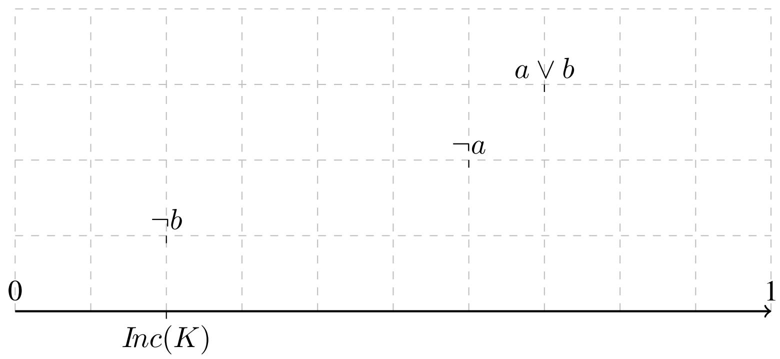


Figure 2.5 – Necessity degrees associated with the formulas of K

Conclusion

This chapter presented compact representations mainly used to encode probability distributions, possibility distributions, or interval-based extension of probability distributions. Indeed, we presented the main concepts defining a network and how to define their associated semantics (given by probability or possibility distribution when dealing with Bayesian or possibilistic networks). We presented possibilistic logic and provided the method to compute the associated possibility distribution. In this chapter, we also presented the inference process in graphical models and some complexity results in Bayesian and credal networks. Regarding *MAP* inference in possibility theory, it has not been established how complex it is to reason with information. One of our goals will be to tackle this problem; this will be addressed in Chapter 8.

Overall, graphical models and knowledge bases are the two mains compact representations used to encode beliefs. We have given the connections and bridges that have been made between probability, possibility and imprecise probability. In next chapter, we will study these bridges by presenting the existing transformations in the literature between probability and possibility but also between imprecise probability and possibility theories.

Chapter 3

From (im)precise probability to possibility distributions

In many situations, the available information may be encoded in different formalisms. One intuitive use of probability-possibility transformation would be for unifying the encoding of belief expressed with different theories. Moreover, transforming possibility measure into probability measure or reverse can be useful in any situation dealing with uncertain knowledge (e.g. statistical data). More precisely,

— *From probability to possibility*: transforming a probability measure into a possibility measure can be useful when the source of information is weak, or when computing with possibilities is less hard than computing with probabilities. Possibility theory is a mathematical theory that deals with different types of uncertainty and is therefore an alternative to probability theory. Possibility theory is dedicated to handling incomplete information. Hence, analyzing with ignorance is easily handle in possibility, or if one wants to reason qualitatively.

  When the amount of data is low, for example in the frequentist setting where it is necessary to have a large amount of data, it is not very sound to lie on probability theory. For example, Masson and Denoeux [MD06] describe how to use probability-possibility transformations when working with small empirical data sets.

— *From possibility to probability*: Possibility theory has been developed as a necessity to work with uncertain and incomplete information. In this respect, there exist a lot of platforms and tools in general to reason in probability theory. But in possibility theory, there does not exist as much platforms and so having a transformation that preserves all of the information can allow us to use these tools instead of creating or adapting into possibilistic context.

  It is also interesting to have such transformations in the case of decision making [Sme89] and multi-source information.

Contents

3.1 From probability distributions to possibility distributions

3.1.1 Principles and properties

Transformation procedures from probability to possibility have been studied in the past [DFMP04, KG93, DPS93, MSMR06]. Many researchers have worked on principles that have to be satisfied to ensure that information is properly preserved. These works are mainly restricted to the proposal of principles and of transformations satisfying those. For example, when having information about an event are encoded with both probability and possibility measures then these descriptions should be coherent. Which suggest consistency principle. Other principles are proposed to guarantee other desired properties.

Zadeh consistency principle

The first principle that transformations tried to satisfy is the consistency principle due to Zadeh [Zad99]: Zadeh [Zad99] measures the consistency between a probability and possibility distribution as follows:

$$C_z(\pi, p) = \sum_{i=1..n} \pi(\omega_i) * p(\omega_i) \tag{3.1}$$

where p and π are a probability and a possibility distributions respectively over a set of n worlds. It intuitively captures the fact that *"A high degree of possibility does not imply a high degree of probability, and a low degree of probability does not imply a low degree of possibility"*. The computed consistency degree is questionable [DFMP04, KG93] in the sense that two resulted possibility distributions can have the same consistency degree but do not contain the same amount of information. Indeed, the degree of consistency is all the more close to 1 given that $\forall \omega_i \in \Omega$, if $p(\omega_i) > 0$ then $\pi(\omega_i) = 1$. Multiple different distributions can be found but *does having the same degree of consistency mean that none of them is better than another, which criterion use to distinct them?*

Example 3.1. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$, p be a probability distribution and π_1, π_2 two possibility distributions of p. π_1 represents total ignorance, whereas π_2 only represents partial ignorance.

	p	π_1	π_2
ω_1	0.2	1	1
ω_2	0.1	1	1
ω_3	0.3	1	1
ω_4	0.4	1	1
ω_5	0	1	0

Table 3.1 – Example of two distributions having the same consistency degree according to Zadeh criterion

Through distributions of Table 3.1, we note that $C_z(\pi_1, p) = 1$ and $C_z(\pi_2, p) = 1$, yet here, distribution π_2 is clearly better than π_1 which represents total ignorance.

Klir consistency principle

Concept of consistency have been redefined by Klir [KG93]. By assuming that the worlds of Ω are ordered such as $p_i > 0$ and $p_i \geq p_{i+1}$, $\forall\ i = \{1..n\}$. A transformation should be based on these postulates:

- the *"scaling" postulate* forces each value π_i to be a function of $\frac{p_i}{p_1}$ (where $p_1 \geq ... \geq p_n$).
- the *"invariance of uncertainty" postulate* according to which p and π must have the same amount of uncertainty.
- the *consistency condition* which establish that what is probable must be possible, then π can be seen as an upper bound of p.

Dubois et Prade [DP80] have given an example to show that the *"scaling"* postulate of Klir can sometimes leads to violate the consistency principle. The second postulate is also questionable since it assumes that probability measure and possibility measure are commensurable.

Dubois and Prade consistency principle

Dubois and Prade [DFMP04] defined three postulates allowing to define the optimal transformation [DFMP04] which always exist and it is unique.

— *Consistency condition* states that for each event (*i.e.* a set of worlds) $\phi \subseteq \Omega$, $P(\phi) \leq \Pi(\phi)$. Here, the obtained possibility distribution should dominate the probability distribution.

— *Preference preservation*: $\forall(\omega_1, \omega_2) \in \Omega^2$, $p(\omega_1) > p(\omega_2)$ if and only if $\pi(\omega_1) > \pi(\omega_2)$ and $p(\omega_1) = p(\omega_2)$ if and only if $\pi(\omega_1) = \pi(\omega_2)$. Intuitively, if two worlds are ordered in a given way in p, then π should preserve the same order.

— *Maximum specificity principle*: This principle requires to search for the most specific possibility distribution that satisfies the two above conditions.

The concept of specificity of possibility distributions allows to select a possibility distribution given the quantity of information encoded. The definition of specificity relation is given as follow:

Definition 3.1 (Specificity)**.** Let π_1 and π_2 be two possibility distributions over Ω, then π_2 is said to be more specific than π_1 if and only if: $\forall\omega \in \Omega, \pi_2(\omega) \leq \pi_1(\omega)$.

3.1.2 Transformation rules

Many probability-possibility transformations have been proposed in the literature. We cite the *Optimal transformation* (OT) [DFMP04], *Klir transformation* (KT) [KG93], *Symmetric transformation* (ST) [DPS93], and *Variable transformation* (VT) [MSMR06]. The optimal transformation (OT) guarantees the most specific possibility distribution that satisfies Dubois and Prade's consistency principle.

Klir transformation

Following his definition of consistency principle, Klir defined two probability-possibility transformations, under two scales. For these transformations, it is assumed that elements of Ω are ordered in such way that: $\forall i = 1..n$, $p(\omega_i) > 0$, $p(\omega_i) \geq p(\omega_{i+1})$ with $p(\omega_{n+1}) = 0$.

— The *ratio-scale*: called normalized transformation and defined by:

$$\pi_i = \frac{p_i}{p_1} \tag{3.2}$$

— The *log-interval scale* : defined by :

$$\pi_i = (\frac{p_i}{p_1})^\alpha \tag{3.3}$$

where α is a parameter that belongs to the open interval $]0, 1[$.

Klir's transformation satisfies his consistency principle. But it does not satisfy Dubois and Prade consistency condition. This is illustrated by the following counter example.

Example 3.2. Let us have $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ such as each degree assigned at each world is illustrated in Table 3.2.

If we take $\phi = \{\omega_2, \omega_3, \omega_4\}$, then we have $P(\phi) = 0.3$ whereas $\Pi(\phi) = 0.142$. Here, instead of having $P(\phi) < \Pi(\phi)$, we have $P(\phi) > \Pi(\phi)$.

ω_i	$p(\omega_i)$	$\pi_{KT}(\omega_i)$
ω_1	0.7	1
ω_2	0.1	0.142
ω_3	0.1	0.142
ω_4	0.1	0.142

Table 3.2 – Example of violation of Dubois and Prade consistency principle

Optimal transformation.

Proposed by Dubois and Prade [DFMP04], it is defined as follows:

$$\pi(\omega_i) = \sum_{j/p(\omega_j) \leq p(\omega_i)} p(\omega_j) \tag{3.4}$$

OT is said to be optimal [DPS93] in the sense that it guarantees the most specific possibility distribution satisfying Dubois and Prade's consistency principle.

Symmetric transformation

After the *Optimal transformation*, the same authors in [DPS93] proposed a *Symmetric transformation*, defined as follows :

$$\pi(\omega_i) = \sum_{j=1}^{n} \min(p(\omega_i), p(\omega_j)) \tag{3.5}$$

ST is easier to compute but it is less optimal in the sense that it is less specific than OT.

Variable transformation

It is proposed by Mouchaweh et al. [MSMR06] and it is defined as follows: assume that the elements of Ω are ordered in such way that : $\forall i = 1..n$, $p_i > 0$, $p_i \geq p_{i+1}$ and $\pi_i > 0$, $\pi_i \geq \pi_{i+1}$ with $p_{n+1} = 0$ and $\pi_{n+1} = 0$, then :

$$\pi_i = (\frac{p_i}{p_1})^{k.(1-p_i)} \tag{3.6}$$

k is a constant which must be guaranteed by the following consistency condition:

$$\forall \omega \in \Omega : \; \pi(\omega) \geq p(\omega) \tag{3.7}$$

This condition is a particular case of the consistency principle of Dubois and Prade defined in Subsection 3.1.1. Indeed, condition of Equation (3.7) is the discrete case of the principle, i.e the distribution only contains singletons. To guarantee the consistency principle, then the constant k has to take its value in the following range:

$$0 \leq k \leq \frac{log(p_n)}{(1-p_n).log(\frac{p_n}{p_1})} \tag{3.8}$$

When the value of k is equal to its maximum: $k_{max} = \frac{log(p_n)}{(1-p_n).log(\frac{p_n}{p_1})}$, the possibility degree π_n, computed according to VT is equal to p_n. If the value of k increases beyond k_{max} then the value of π_n will become smaller than the one of p_n, which means that VT will not satisfy the consistency principle of Equation (3.7). Indeed, VT transforms a probability distribution in a non-lineary way, which means that VT adds more information to the bigger probability degrees than to the smaller ones. It is due to the power $k.(1-p_i)$ in Equation (3.6), smaller than the smallest of probability degrees. The difference between Klir transformation and VT is that the Klir transformation has a constant power α belonging to the unit interval $]0,1[$ in order to preserve the uncertainty, whereas the power $k.(1-p_i)$ in VT makes it more specific.

Bouguelid [Bou07] proposed VT_i, as an improvement of VT, in order to make it as specific as OT. Then, a parameter k_i is associated to each π_i. Formally, $\forall\ i = \{1..n\}$,

$$\pi_i = (\frac{p_i}{p_1})^{k_i.(1-p_i)} \tag{3.9}$$

where k_i belongs to the interval: $0 \leq k_i \leq \frac{log(p_i+p_{i+1}+...+p_n)}{(1-p_n).log(\frac{p_n}{p_1})}$, $\forall\ i = \{2..n\}$.

We will see later on that all these transformations suffer a lot of problems. Let us see an example of these transformations on a probability distribution with the following example.

Example 3.3. Let us consider two variables A and B with their domains respectively $D_A = \{a_1, a_2\}$ and $D_B = \{b_1, b_2, b_3\}$, the associated probability distribution is depicted in Table 3.3.

A	B	$p(A,B)$
a_1	b_1	0.36
a_1	b_2	0.18
a_1	b_3	0.06
a_2	b_1	0.2
a_2	b_2	0.12
a_2	b_3	0.08

Table 3.3 – Probability distribution of Example 3.3

Let us apply the transformation procedures OT, KT, ST, VT, and VT_i over the distribution of Table 3.3. We get Table 3.4.

A	B	Π_{KT}	Π_{OT,VT_i}	Π_{ST}	Π_{VT}
a_1	b_1	1	1	1	1
a_1	b_2	0.5	0.44	0.8	0.38
a_1	b_3	0.16	0.06	0.36	0.06
a_2	b_1	0.55	0.64	0.84	0.45
a_2	b_2	0.33	0.26	0.62	0.19
a_2	b_3	0.22	0.14	0.46	0.09

Table 3.4 – Possibility distribution obtained by the different transformations

Table 3.5 summarizes characteristics of KT, OT, ST, VT and VT_i. For each transformation, it is precised if it deals with the discrete case (D) and the continuous case (C) and if it satisfies the consistency principle (Cs), the preservation of preference (PP) and the maximum of specificity (MS). Clearly, OT and VT_i are the most interesting rules in the discrete case to transform from probability to possibility.

TR	$p \to \pi$	$\pi \to p$	Properties				
			D	C	Cs	PP	MS
KT	$\times$	$\times$	$\times$	$\times$		$\times$	
OT	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$
ST	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	
VT	$\times$		$\times$			$\times$	
VT_i	$\times$		$\times$		$\times$	$\times$	$\times$

Table 3.5 – Summary of transformations' properties

3.1.3 Converse transformation: from possibility to probability

In this thesis, we are basically trying to get around using probability theory or its extensions such as imprecise probability theory to express uncertain information. But in some cases it might be useful to have transformation from possibility to probability. Therefore, this small part addressed the reverse transformation. As we said previously, when going from a probabilistic representation to a possibilistic representation, some information is lost. The converse transformation, on the contrary, adds information to some possibilistic incomplete knowledge. First let us review the consistency principles such transformations must fulfill.

The principles involved in possibility-probability transformations are similar to the probability-possibility transformation as they are complementary. Thus, the first principle known as the probability-possibility consistency principle is the same as Dubois and Prade consistency principle (see Subsection 3.1.1).

The second principle is the preference preservation: a possibility distribution π induces a preference order on Ω, such that $\pi(\omega_i) > \pi(\omega_j)$ means that the world ω_i is preferred to ω_j. A transformation should therefore satisfy:

$$\pi(\omega_i) > \pi(\omega_j) \Rightarrow p(\omega_i) > p(\omega_j). \tag{3.10}$$

Which results in a unique condition for probability-possibility transformation:

$$\pi(\omega_i) > \pi(\omega_j) \Leftrightarrow p(\omega_i) > p(\omega_j). \tag{3.11}$$

One principle devoted to possibility-probability transformation is the one of insufficient reason. This principle claims that if all we know about ω is that ω belongs to a set ϕ, then we can assume that the maximal uncertainty about ω can be described by a uniform probability distribution over ϕ. Yager [Yag81] suggested a procedure that satisfies this principle and it is stated as follow:

Given a possibility distribution π, we apply this principle twice:

— on the unit interval: select α at random in $(0, 1]$ and consider $\phi_\alpha = \{\omega | \pi(\omega) \geq \alpha\}$

— on the selected level-cut ϕ_α: select ω at random in ϕ_α.

This means that if π can be described by a finite set of level-cuts $\phi_1, ..., \phi_n$ corresponding to $\pi_1 = 1 > \pi_2 > ... > \pi_n > \pi_{n+1} = 0$, the selection process is guided by the density function:

$$p(\omega) = \sum_{i=1,\ldots,n} \frac{\pi_i - \pi_{i+1}}{|\phi_i|} \mu_{\phi_i}(\omega), \ \forall \omega \tag{3.12}$$

This procedure actually corresponds to a transformation already proposed by Dubois and Prade [DP83, DP92] and discussed in the belief function setting by Smets [Sme89].

3.2 From interval-based probability distributions to possibilistic distributions

Probability theory, which is the most widely used theory, might seem unfit for representing some kinds of information such as the knowledge of a group of experts, or incomplete knowledge. This is why more general frameworks are needed for allowing more flexibility. Imprecise probability theory [DP05, Wei00, Wal00] in that respect has been designed to generalize probability theory and offers more expressiveness as they represent uncertain information by means of credal sets instead of single probability values. The problem when reasoning with such general and expressive models is that they entail higher computational complexity. This is why transforming to possibility theory can be useful.

When transforming uncertain information expressed by means of probability intervals to a possibility distribution, there is to the best of our knowledge only one work [MD06] where the authors learn possibility distributions from empirical data by transforming confidence intervals to possibility distributions.

3.2.1 Masson and Denoeux transformation (MD)

The starting point of this transformation is to consider an imprecise probability distribution as a means of encoding a partial order $\mathcal{M}$ over Ω. Indeed, contrary to precise probability distributions which encode complete order relations over Ω, interval-based ones encode partial orders in the form $\omega_i <_{IP} \omega_j$ in case where $u_i < l_j$. Let $\mathcal{M}$ be the partial order encoded by an imprecise probability distribution IP and let $\mathcal{C}$ be the set of linear extensions (complete orders) that are compatible with the partial order $\mathcal{M}$. The transformation proposed in [MD06] proceeds as follows:

— For every linear extension $\mathcal{C}_l \in \mathcal{C}$ and for each $\omega_i \in \Omega$, compute:

$$\pi^{C_l}(\omega_i) = \max_{p_1..p_n}\Big(\sum_{p_j \leq p_i} p_i\Big) \tag{3.13}$$

subject to the following constraints (in order to explore only compatible probability distributions satisfying the current linear extension C_l):

$$\begin{cases} p_i \in [l_i, u_i] \\ \sum_{i=1..n} p_i = 1 \\ p_1..p_n \text{ satisfy the linear extension } C_l \end{cases}$$

— Build the distribution π that dominates all the distributions π^{C_l} as follows: $\forall \omega_i \in \Omega$,

$$\pi(\omega_i) = \max_{\mathcal{C}_l \in \mathcal{C}}(\pi^{C_l}(\omega_i)) \tag{3.14}$$

The motivation of using Equation (3.14) is to guarantee that the obtained possibility distribution π dominates the probability intervals IP. This transformation tries on one hand to preserve the order of interpretations induced by IP and the dominance principle requiring that $\forall \phi \subseteq \Omega$, $P(\phi) \leq \Pi(\phi)$ on the other hand.

This transformation comes down to the Optimal transformation OT when intervals of degrees are singletons.

Example 3.4. Let us consider the following example in which we have an imprecise probability distribution over 4 worlds as described in Table 3.6 [4]. The first thing is to compute the set of possible linear

ω	$IP(\omega)$
ω_1	[.10, .28]
ω_2	[.34, .56]
ω_3	[.25, .46]
ω_4	[0, .08]

Table 3.6 – Imprecise probability distribution to be transformed by MD

extensions. Then for each linear extension, compute using a linear program solving on the Equation (3.13).

We have three possible linear extensions:

— $\omega_4 < \omega_1 < \omega_3 < \omega_2$

— $\omega_4 < \omega_1 < \omega_2 < \omega_3$

— $\omega_4 < \omega_3 < \omega_1 < \omega_2$

After computing each π^{C_l}, we obtain the following table.

ω	ω_1	ω_2	ω_3	ω_4
$\pi(\omega)$	.64	1	1	.08

There are two main drawbacks with the transformation of Equations (3.13) and (3.14):

— The first issue is about the computational complexity of such transformation. Applied directly, this latter can consider in the worst case $N!$ linear extensions where N is the number of possible worlds. The authors proposed in [MD06] an algorithm allowing to achieve some improvements during this transformation but it is still very costly when one considers variables having domains exceeding a dozen values, which is common in many applications.

— The second concern lies in the fact that this transformation does not guarantee that the obtained distribution is optimal is terms of specificity. Indeed, it was shown in [DDC07] that the transformation of Equation (3.14) results in a loss of information as it is not the most specific one dominating the considered imprecise probability distribution.

3.2.2 Cumulative distribution based transformation (CD)

The second transformation, called *CD* stands for *Cumulative Distribution*, is related to upper and lower cumulative distributions. In [DDC08], the authors discussed the connection that one can make between generalized p-box and possibility distributions and gave a representation of a p-box by two possibility distributions. Given a set of probability intervals and an ordering relation $\leq_{\mathcal{C}_I}$ on a linear extension $\mathcal{C}_l$ between elements ω_i, we can easily build a generalized p-box [DDC08], $[\underline{F}, \overline{F}]$ defined by two cumulative distributions $\underline{F}$ and $\overline{F}$. Given the consecutive sets $A_i = \{\omega_i, \forall \omega_i \in \Omega$ and such that $\omega_i \leq_{\mathcal{C}_I} \omega_j$ if and only if $i < j\}$, lower and upper generalized cumulative distributions corresponding to Ω are, respectively:

$$\underline{F}(\omega_i) = \underline{P}(A_i) = \max(\sum_{\omega_i \in A_i} l_j, 1 - \sum_{\omega_i \notin A_i} u_j)$$

$$\overline{F}(\omega_i) = \overline{P}(A_i) = \min(\sum_{\omega_i \in A_i} u_j, 1 - \sum_{\omega_j \notin A_i} l_j)$$

4. The interval values associated with worlds are randomly chosen.

From this two cumulative distributions, we can compute two possibility distributions $\pi_{\overline{F}}$ and $\pi_{\underline{F}}$ where:

$$\pi_{\underline{F}}(\omega_i) = 1 - \max\{\underline{F}(\omega_j) < \underline{F}(\omega_i) : j = 0..n\} \tag{3.15}$$

$$\pi_{\overline{F}}(\omega_i) = \overline{F}(\omega_i) \tag{3.16}$$

Example 3.5. By keeping the same table as the previous example (Table 3.6). If we consider the following linear extension : $\omega_4 < \omega_1 < \omega_3 < \omega_2$ then the resulting transformed possibility distribution by *CD* is given by:

ω	ω_1	ω_2	ω_3	ω_4
$\pi(\omega)$	.36	1	.66	.08

Conclusion

This chapter proposed a brief synthetic report mainly on transformation from probability to possibility theory. We also provided some examples from imprecise probability theory to possibility theory. Other transformations have been implemented in the continuous case rather than the finite case. Dubois and Prade [DPS93] proposed an approach so that the transformation of a continuous probability distribution into a specific possibility distribution allows the upper limits of an event probability to be computed.

Yamada [Yam01] proposed a transformation based on the evidence theory. He reviewed three new ideas based on evidence theory: one considers possibility on an ordinal scale, his work shows that the preservation of order principle gives a unique ordinal structure of possibilities. For the numerical scale, he investigated two methods and showed that one of them based on the equidistribution leads to a unique transformation satisfying consistency and order preservation principles. This transformation turned out to be the one proposed in [DP83].

We will see in Chapters 7 and 8 a deep analysis of the behavior of probability-possibility transformations with respect to reasoning tasks and the properties of transformations in graphical models. But first, let us focus on interval-based possibility distributions.

Part II

Conditioning in interval-based and set-based possibilistic frameworks

Chapter 4

Quantitative conditioning in interval-based possibilistic setting

Interval-based possibilistic logic is a flexible setting extending standard possibilistic logic such that each logical expression is associated with a sub-interval of $[0, 1]$. This chapter proposes our first contribution of this thesis. It first focuses on the fundamental issue of conditioning in the interval-based possibilistic setting. The first part of the chapter provides a refresher on interval-based possibilistic frameworks. The second part of the chapter proposes a set of natural properties that an interval-based conditioning operator should satisfy. We then give a natural and safe definition for conditioning an interval-based possibility distribution. This definition is based on applying standard product-based conditioning on the set of all associated compatible possibility distributions. In the third part of this chapter, Subsection 4.3.1 analyzes the obtained posterior distributions and provide a precise characterization of lower and upper endpoints of the intervals associated with interpretations which refers to the semantic conditioning. We also provide in Subsection 4.3.2 an equivalent syntactic computation of interval-based conditioning when interval-based distributions are compactly encoded by means of interval-based possibilistic knowledge bases. We show that interval-based conditioning is achieved without extra computational cost comparing to conditioning standard possibilistic knowledge bases.

Contents

4.1 Introduction to the interval-based possibilistic setting

The aim of this section is to present a more general framework of possibility theory, where uncertainty is not encoded with a single necessity value but by means of an interval [BHLR11]. This framework allows the introduction of an imprecision on degrees associated with beliefs. We use real number based intervals $I = [\alpha, \beta] \subseteq [0, 1]$ to encode the uncertainty associated with formulas. We denote by $\mathcal{I}$ the set of all closed intervals over $[0, 1]$. If I is an interval, then we denote by $\overline{I}$ and $\underline{I}$ its upper and lower endpoints respectively. When all I's associated with interpretations (resp. formulas) are singletons (namely $\overline{I} = \underline{I}$), we refer to standard (or point-wise) distributions (resp. standard possibilistic bases).

Given a set of intervals $I_i = [\alpha_i, \beta_i]$, where α_i, resp. β_i describes the lower, resp. the upper bound of I_i. We define the following operations:

— Max of intervals: $\mathcal{M}\{I_1, \ldots, I_n\} = [\max\{\alpha_1, \ldots, \alpha_n\}, \max\{\beta_1, \ldots, \beta_n\}]$

— Reverse of an interval: $1 \ominus I_i = [1 - \beta_i, 1 - \alpha_i]$

— Comparing intervals: $I_i = [\alpha_i, \beta_i] \lhd I_j = [\alpha_j, \beta_j]$ if and only if $\beta_i < \alpha_j$

By convention, $\mathcal{M}\{\emptyset\} = [0, 0]$.

4.1.1 Interval-based possibility theory

Interval-based and compatible possibility distributions

An interval-based possibility distribution, denoted by $I\pi$, is a function from Ω to $\mathcal{I}$. $I\pi(\omega) = I$ means that the possibility degree of ω is one of the elements of I. $I\pi$ only induces a partial pre-ordering between interpretations defined by $\omega < \omega'$ (ω' is preferred to ω) if and only if $I\pi(\omega) \lhd I\pi(\omega')$. Since $\lhd$ is a partial pre-order, an interval-based possibility distribution only induces a partial pre-order on interpretations. We also interpret an interval-based possibility distribution as a family of compatible standard possibility distributions defined by:

Definition 4.1 (Compatible possibility distribution)**.** Let $I\pi$ be an interval-based possibility distribution. A possibility distribution π is said to be compatible with $I\pi$ if and only if $\forall \omega \in \Omega,\ \pi(\omega) \in I\pi(\omega)$.

Of course, compatible distributions are not unique. We denote by $\mathcal{C}(I\pi)$ the set of all compatible possibility distributions with $I\pi$.

An interval-based possibility distribution is said to be weakly normalized if there exists ω such that $\overline{I\pi}(\omega) = 1$. Weak normalization reflects the existence of a normalized pointwise possibility distribution π which is compatible with $I\pi$.

Necessity and possibility measures

A natural way to define the counterparts of possibility and necessity measures associated with an event ϕ from an interval-based possibility distribution is to use the set of all compatible distributions, namely:

Definition 4.2. Let $I\pi$ be an interval-based possibility distribution and let ϕ be an event, then:

$$I\Pi(\phi) = [\min_{\pi \in \mathcal{C}(I\pi)} \Pi(\phi), \max_{\pi \in \mathcal{C}(I\pi)} \Pi(\phi)], \text{ and}$$

$$IN(\phi) = [\min_{\pi \in \mathcal{C}(I\pi)} N(\phi), \max_{\pi \in \mathcal{C}(I\pi)} N(\phi)].$$

Definition 4.2 is safe since it relies on all the compatible distributions as opposed to a possible approach when only an arbitrary set of compatible distributions is used.

In [BHLR11], the authors showed that in the particular case where intervals in a possibility distribution are only consisting of singletons, then the approach recovers the standard definitions of possibility measures.

4.1.2 Interval-based possibilistic bases

We now present the syntactic representation of interval-based possibilistic logic. We generalize the notion of a possibilistic base to an interval-based possibilistic knowledge base as follows.

Definition 4.3 (Interval-based possibilistic base)**.** An interval-based possibilistic base, denoted by IK, is a multi-set of formulas associated with intervals:

$$IK = \{(\varphi, I), \varphi \in \mathcal{L} \text{ and } I \in \mathcal{I}\} \tag{4.1}$$

The intuitive interpretation of (φ, I) is that the necessity degree of φ is one of the elements of $I = [\alpha, \beta]$.

An interval-based possibilistic base IK can be viewed as a family of standard possibilistic bases called compatible bases.

Definition 4.4 (Compatible possibilistic base)**.** A possibilistic base K is said to be compatible with an interval-based possibilistic base IK if and only if K is obtained from IK by replacing each interval-based formula $(\phi, [\alpha, \beta])$ by a standard possibilistic formula (ϕ, δ) with $\alpha \leq \delta \leq \beta$.

We also denote by $\mathcal{C}(IK)$ the infinite set of all compatible possibilistic bases associated with an interval-based possibilistic base IK. Contrarily to standard possibilistic bases, we do allow α to be equal to 0. The presence of the degree '0' means that there exists a compatible knowledge base where the formula φ is not explicitly supported.

Given an interval-based possibilistic base IK, we define two particular compatible possibilistic bases $\underline{IK}$ and $\overline{IK}$ by selecting either lower endpoints of intervals (pessimistic point of view) or upper endpoints of intervals (optimistic point of view):

1. $\underline{IK} = \{(\varphi, \alpha) : (\varphi, [\alpha, \beta]) \in IK\}$
2. $\overline{IK} = \{(\varphi, \beta) : (\varphi, [\alpha, \beta]) \in IK\}$

From interval-based possibilistic bases to interval-based possibility distributions

As in standard possibilistic logic, an interval-based knowledge base IK is also a compact representation of an interval-based possibility distribution $I\pi_{IK}$. The interval-based possibility distribution can be equivalently obtained using: i) an extension of the definition of π_K given by Equation (2.12) to deal with intervals, ii) the set of possibility distributions associated with compatible bases, and iii) the two particular compatible bases $\underline{IK}$ and $\overline{IK}$. This is summarized by Definition 4.5.

Definition 4.5 (Interval-based possibility distribution)**.** Let IK be an interval-based possibilistic base, then:

$$I\pi_{IK}(\omega) = [\min_{K \in \mathcal{C}(IK)} \pi_K(\omega), \max_{K \in \mathcal{C}(IK)} \pi_K(\omega)]$$

where π_K is a standard possibilistic distribution associated with the compatible base K.

4.2 Conditioning interval-based possibility distributions: Properties and definitions

Our first contribution concerns conditioning with interval-based possibility distributions. Interval-based possibilistic logic described until now is only specified for static situations and no form of conditioning has been proposed for updating the current knowledge and beliefs.

4.2.1 Properties of interval-based conditioning

Conditioning and belief change are important tasks for designing intelligent systems. Conditioning is concerned with updating the current beliefs when a new sure piece of information becomes available. In the possibilistic setting, given a possibilistic knowledge base K or a possibility distribution π and a new evidence ϕ, conditioning allows to update the old beliefs, encoded by π or K, with ϕ. There are several definitions of the possibilistic conditioning [DP97a, DP06, Fon97, His78, LMDCM95]. Conditioning operators are designed to satisfy some properties such as giving priority to the new information and performing minimal change.

This section gives natural properties that a conditioning operation should satisfy when interval-based possibility distributions are used. Let us first fix the values of $I\pi(.|\phi)$ for degenerate possibility distributions $I\pi$ when $\overline{I\Pi}(\phi) = 0$ or $\underline{I\Pi}(\phi) = 0$.

If $\overline{I\Pi}(\phi) = 0$ then by convention, as in standard possibility distributions, $\forall\omega \in \Omega$, $I\pi(\omega|\phi) = [1, 1]$. Similarly, if $\underline{I\Pi}(\phi) = 0$ (but $\overline{I\Pi}(\phi) > 0$) then $\forall\omega \in \Omega$,

$$I\pi(\omega|\phi) = \begin{cases} [0,0] & \text{if } I\pi(\omega) = [0,0] \text{ and } \omega \not\models \phi; \\ [0,1] & \text{otherwise.} \end{cases} \tag{4.2}$$

In the rest of this paper, we assume that $I\pi$ is not degenerate with respect to ϕ. Namely, we assume that $\underline{I\Pi}(\phi) > 0$ and hence for each compatible possibility distribution π, it is assumed that ϕ is somewhat possible in π (namely, $\Pi(\phi)$ >0).

Besides, in this section, we assume that $I\pi$ is normalized, namely there exists an interpretation ω such that $\overline{I\pi}(\omega) = 1$. Hence, $\mathcal{C}(I\pi)$ refers to the set of pointwise normalized possibility distributions.

In an interval-based setting, a conditioning operator " | " should satisfy the following suitable properties:

- (**IC1**) $I\pi(.|\phi)$ should be an interval-based distribution.
- (**IC2**) $\forall\omega \in \Omega$, if $\omega \not\models \phi$ then $I\pi(\omega|\phi) = [0, 0]$.
- (**IC3**) $\exists\omega \in \Omega$ such that $\omega \models \phi$ and $\overline{I\pi}(\omega|\phi) = 1$.
- (**IC4**) If $\forall\omega \not\models \phi$, $I\pi(\omega) = [0, 0]$ then $I\pi(.|\phi) = I\pi$.
- (**IC5**) $\forall\omega \in \Omega$, if $\omega \models \phi$ and $I\pi(\omega) = [0, 0]$ then $I\pi(\omega|\phi) = [0, 0]$.
- (**IC6**) $\forall\omega \models \phi$ and $\forall\omega' \models \phi$, if $\overline{I\pi}(\omega) < \underline{I\pi}(\omega')$ then $\overline{I\pi}(\omega|\phi) < \underline{I\pi}(\omega'|\phi)$.
- (**IC7**) $\forall\omega \models \phi$, $\forall\omega' \models \phi$, if $I\pi(\omega) = I\pi(\omega')$ then $I\pi(\omega|\phi) = I\pi(\omega'|\phi)$.

Property **IC1** simply states that the result of applying conditioning over an interval-based possibility distribution should result in an interval-based possibility distribution. Property **IC2** requires that when the new sure piece of information ϕ is observed then any interpretation that is a counter-model of ϕ should be completely impossible. Property **IC3** states that there exists at least a compatible possibility distribution π' of $I\pi(.|\phi)$ where $\Pi'(\phi) = 1$. Property **IC4** states that if ϕ is already fully accepted (namely, all counter-models of ϕ are already impossible) then $I\pi(.|\phi)$ should be identical to $I\pi$. Property **IC5** states that impossible interpretations (even if they are models of ϕ) remain impossible after conditioning. Properties **IC6** and **IC7** express a minimal change principle. **IC6** states that the strict relative ordering between models of ϕ should be preserved after conditioning. **IC7** states that equal models of ϕ should remain equal after conditioning.

4.2.2 Definitions and property-based analysis

This section provides a natural and safe definition of conditioning an interval-based possibility distribution using the set of compatible possibility distributions. More precisely, conditioning an interval-based possibility distribution $I\pi$ comes down to apply standard product-based conditioning on the set of all compatible possibility distributions $\mathcal{C}(I\pi)$ associated with $I\pi$. Namely,

Definition 4.6. The compatible-based conditioned interval-based possibility distribution is defined on the set of all compatible distributions such that: $\forall\omega \in \Omega$, $I\pi(\omega|_*\phi)$ is given by the set of all conditioned compatible distributions $\pi(\omega|_*\phi)$, where $|_*$ is given by Equation (1.28)

Definition 4.6 is illustrated by Figure 4.1.

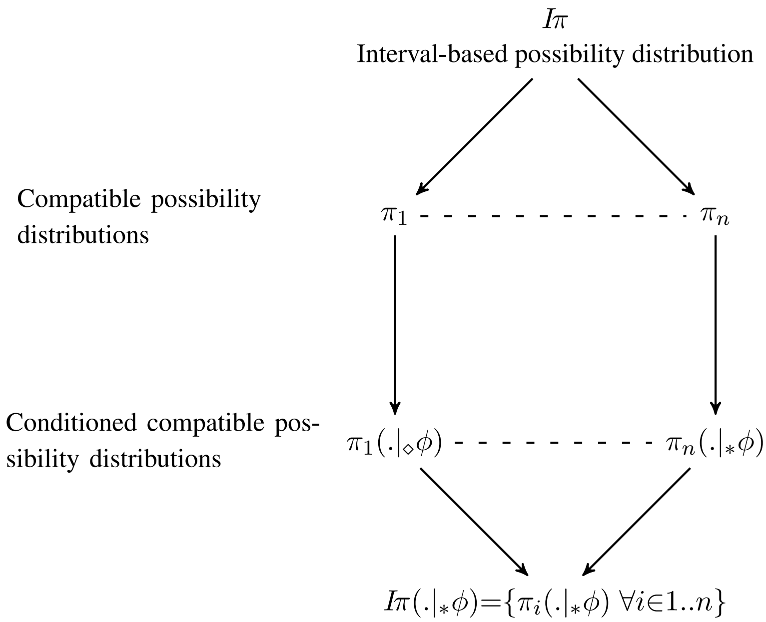


Figure 4.1 – Compatible-based conditioning

Conditioning according to Definition 4.6 is safe since it relies on all the compatible distributions as opposed to a possible approach when only an arbitrary set of compatible distributions is used.
The first important issue with compatible-based conditioning of Definition 4.6 is whether conditioning an interval-based distribution $I\pi$ with an evidence ϕ gives an interval-based distribution, namely whether the first property (**IC1**) is satisfied or not.

Proposition 4.1. *Let $I\pi$ be an interval-based distribution. Let ϕ be the new evidence and $|_*$ be the standard product-based conditioning given by Equation* (1.28). *Then $\forall\omega \in \Omega$,*

$$I\pi(\omega|_*\phi) = [\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|_*\phi)), \max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|_*\phi))]$$

is an interval.

Proof. Recall that it is assumed in the whole chapter that $\underline{I\Pi}(\phi) > 0$. Let us show that $\forall\omega \models \phi$, $I\pi(\omega|_*\phi)$ is indeed an interval. Assume that there exist two numbers α and β such that:

— $\alpha < \beta$,

— $\alpha \in I\pi(\omega|_*\phi)$, $\beta \in I\pi(\omega|_*\phi)$, and

— $\forall\gamma$ such that $\alpha < \gamma < \beta$ we have $\gamma \notin I\pi(\omega|_*\phi)$.

The assumption $\alpha \in I\pi(\omega|_*\phi)$ means that there exists a compatible possibility distribution π such that $\alpha = \dfrac{\pi(\omega)}{\Pi(\phi)}$. Similarly, $\beta \in I\pi(\omega|_*\phi)$ means that there exists another compatible possibility distribution π' such that $\beta = \dfrac{\pi'(\omega)}{\Pi'(\phi)}$. There are different cases to consider:

— $\pi(\omega) < \overline{I\pi}(\omega)$. Since $\alpha < \beta$ then trivially there exists ε such that $\alpha = \dfrac{\pi(\omega)}{\Pi(\phi)} < \dfrac{\pi(\omega)+\varepsilon}{\Pi(\phi)} < \beta$. Therefore it is enough to define a new compatible possibility distribution π'' such that $\pi''(\omega) = \pi(\omega) + \varepsilon$ and $\forall \omega'$, $\pi''(\omega') = \pi(\omega')$. Clearly, π'' is compatible and $\pi''(\omega|_*\phi) = \dfrac{\pi(\omega)+\varepsilon}{\Pi(\phi)} \in I\pi(\omega|_*\phi)$.

— $\pi'(\omega) > \underline{I\pi}(\omega)$. Since $\alpha < \beta$ then trivially there exists ε such that $\alpha < \dfrac{\pi'(\omega)-\varepsilon}{\Pi'(\phi)} < \beta = \dfrac{\pi'(\omega)}{\Pi'(\phi)}$. Therefore it is enough to define a new compatible possibility distribution π'' such that $\pi''(\omega) = \pi'(\omega) - \varepsilon$ and $\forall \omega'$, $\pi''(\omega') = \pi'(\omega')$. Clearly, π'' is compatible and $\pi''(\omega|_*\phi) = \dfrac{\pi'(\omega)-\varepsilon}{\Pi'(\phi)} \in I\pi(\omega|_*\phi)$.

— $\pi(\omega) = \overline{I\pi}(\omega)$ and $\pi'(\omega) = \underline{I\pi}(\omega)$. In this case we have $\pi(\omega) \geq \pi'(\omega)$. Since $\alpha = \dfrac{\pi(\omega)}{\Pi(\phi)} < \beta = \dfrac{\pi'(\omega)}{\Pi'(\phi)}$, we also have $\Pi(\phi) > \Pi'(\phi)$. Let ω_1 be an interpretation such that $\omega_1 \models \phi$ and $\pi(\omega_1) = \Pi(\phi) = \gamma$. Note first that $\gamma > \underline{I\pi}(\omega_1)$. Indeed, if $\gamma = \underline{I\pi}(\omega_1)$, then this means that $\forall \pi''$, π'' compatible with $I\pi$, we have $\pi''(\omega_1) \geq \gamma$. In particular, $\pi'(\omega_1) \geq \gamma = \Pi(\phi)$ which is impossible since $\Pi'(\phi) \geq \pi'(\omega_1)$ (recall that $\omega_1 \models \phi$), this means that $\Pi'(\phi) \geq \Pi(\phi)$ which contradicts the fact that $\Pi(\phi) \geq \Pi'(\phi)$. Note also that $\pi(\omega) \neq \Pi(\phi)$, otherwise $\alpha = \dfrac{\pi(\omega)}{\Pi(\phi)} = 1$ which is impossible since $\alpha < \beta$. Since $\gamma > \underline{I\pi}(\omega_1)$, let us define a new possibility distribution π'' from π as follows: $\forall \omega^* \in \Omega$,

$$\pi''(\omega^*) = \begin{cases} \pi(\omega^*) - \epsilon & \text{if } \pi(\omega^*) = \Pi(\phi) \text{ and } \omega^* \models \phi \\ \pi(\omega^*) & \text{otherwise} \end{cases}$$

where ϵ is a very small positive number. Clearly, π'' is compatible, since π is compatible and $\pi''(\omega^*) \in I\pi(\omega^*)$ for all $\omega^* \in \Omega$. Then one can easily check that:

$$\alpha = \frac{\pi(\omega)}{\Pi(\phi)} < \frac{\pi''(\omega)}{\Pi''(\phi)} = \frac{\pi(\omega)}{\Pi(\phi)-\epsilon} < \frac{\pi'(\omega)}{\Pi'(\phi)} = \beta.$$

□

In the rest of the chapter, we only use $I\pi(.|\phi)$ and $\pi(.|\phi)$ instead of $I\pi(.|_*\phi)$ and $\pi(.|_*\phi)$ to avoid heavy notations. The following proposition states that the compatible-based conditioning given in Definition 4.6 satisfies properties **IC1**-**IC7**.

Proposition 4.2. *Let $I\pi$ be a normalized interval-based possibility distribution. Let ϕ be the new evidence such that $\underline{I\Pi}(\phi) > 0$. Then the updated interval-based possibility distribution computed according to Definition 4.6 satisfies properties* **IC1**-**IC7**.

Proof. Let ϕ be the new evidence such that $\underline{I\Pi}(\phi) > 0$.

— For postulate **IC1**, the proof is the one of Proposition 4.1. Indeed, using the product-based conditioning of Equation (1.28), the obtained distribution $I\pi(.|\phi)$ using the compatible-based conditioning of Definition 4.6 is an interval-based one.

— For postulate **IC2**, following the compatible-based conditioning of Definition 4.6, $\forall\omega \not\models \phi$, $I\pi(\omega|\phi) = [\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi)), \max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi))] = [0,0]$, since $\pi(\omega|\phi) = 0$ for all $\pi \in \mathcal{C}(I\pi)$ and for all ω counter models of ϕ.

— For postulate **IC3**, recall that in $I\pi$, the new evidence is assumed to be somewhat possible, namely, $\underline{I\Pi}(\phi) > 0$. Then, $\forall\pi \in \mathcal{C}(I\pi)$, $\exists\omega^* \models \phi$ such that $\pi(\omega^*|\phi) = 1$. Hence, $\exists\omega^* \models \phi$ such that $\max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega^*|\phi))=\overline{I\pi}(\omega^*|\phi) = 1$.

— For postulate **IC4** stating that if ϕ is already fully accepted (namely, all counter-models of ϕ are already impossible) then $I\pi(.|\phi)$ should be identical to $I\pi$. ϕ is already fully accepted if and only if $\forall\omega \not\models \phi$, $I\pi(\omega) = [0,0]$ (namely, $\forall\pi \in \mathcal{C}(I\pi)$ and $\forall\omega \not\models \phi$, $\pi(\omega) = 0$). Two cases have to be considered:

  — *Case 1:* For counter-models of ϕ, according to Postulate **IC2**, $\forall\omega \not\models \phi$, $I\pi(\omega|\phi) = [0,0]$. Then, $\forall\omega \not\models \phi$, $I\pi(\omega) = I\pi(\omega|\phi) = [0,0]$.

  — *Case 2:* For models of ϕ, since ϕ is already fully accepted then $\forall\pi \in \mathcal{C}(I\pi)$, $\Pi(\phi) = 1$ since each compatible distribution π is normalized and $\forall\omega \not\models \phi$, $\pi(\omega) = 0$. Now, since $\forall\pi \in \mathcal{C}(I\pi)$, $\Pi(\phi) = 1$ then $\forall\pi \in \mathcal{C}(I\pi)$ and $\forall\omega \models \phi$, $\pi(\omega|\phi) = \frac{\pi(\omega)}{\Pi(\phi)} = \pi(\omega)$.

  Then from *Case 1* and *Case 2*, we conclude that if $\forall\omega \not\models \phi$, $I\pi(\omega) = [0,0]$ then $\forall\omega \in \Omega$, $I\pi(\omega) = I\pi(\omega|\phi)$.

— For postulate **IC5**, recall that $\underline{I\Pi}(\phi) > 0$ and that $I\pi(\omega|\phi) = [\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi)), \max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi))]$. Hence, if $I\pi(\omega) = [0,0]$ then $\forall\pi \in \mathcal{C}(I\pi)$, $\pi(\omega) = 0$ and $\pi(\omega|\phi)$=0. As a consequence, $[\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi)), \max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi))] = [0,0]$.

— For postulate **IC6**, let $\omega \models \phi$ and $\omega' \models \phi$ be two interpretations such that $\overline{I\pi}(\omega) < \underline{I\pi}(\omega')$. Then, on the one hand we have $\max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega)) < \min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega'))$ (following the definition of upper and lower endpoints). This means that $\forall\pi \in \mathcal{C}(I\pi)$, we have $\pi(\omega) < \pi'(\omega)$. On the other hand, $\forall\pi \in \mathcal{C}(I\pi)$, if $\pi(\omega) < \pi(\omega')$ then $\pi(\omega|\phi) < \pi(\omega'|\phi)$. We conclude that $\max_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi)) < \min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega'|\phi))$. Said in other words, if $\overline{I\pi}(\omega) < \underline{I\pi}(\omega')$ then $\overline{I\pi}(\omega|\phi) < \underline{I\pi}(\omega'|\phi)$.

— For postulate **IC7**, the proof is similar to the one of postulate **IC6**. Indeed, if $\omega \models \phi$ and $\omega' \models \phi$ are two interpretations such that $I\pi(\omega) = I\pi(\omega')$. Then, $\overline{I\pi}(\omega) = \overline{I\pi}(\omega')$ and $\underline{I\pi}(\omega) = \underline{I\pi}(\omega')$. This also means that $\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega)) = \min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega'))$. Now, if $\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega)) = \min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega'))$, then $\min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega|\phi)) = \min_{\pi\in\mathcal{C}(I\pi)}(\pi(\omega'|\phi))$ allowing to state that $I\pi(\omega|\phi) = I\pi(\omega'|\phi)$. The same result holds for the upper endpoint using $\max$ instead of $\min$.

□

4.3 Conditioning interval-based possibility distributions: Computations

4.3.1 Computing lower and upper endpoints of $I\pi(.|\phi)$

The objective now is to determine the lower and upper endpoints of $I\pi(.|\phi)$. Let us start with a particular case of interval-based distributions $I\pi$ where in each compatible distribution $\pi \in \mathcal{C}(I\pi)$, ϕ is accepted (namely, $\Pi(\phi) > \Pi(\overline{\phi})$). In this case, the computation of $I\pi(.|\phi)$ is immediate:

Proposition 4.3. *Let $I\pi$ be an interval-based possibility distribution and ϕ be a propositional formula such that $\overline{I\Pi}(\phi) = 1$ and $\overline{I\Pi}(\overline{\phi}) < 1$. Then*

— *If there is only one interpretation $\omega^* \in \Omega$ such that $\omega^* \models \phi$ and $\overline{I\pi}(\omega^*) = 1$ then $\forall \omega \in \Omega$,*

$$I\pi(\omega|\phi) = \begin{cases} [1,1] & \text{if } \omega = \omega^* \\ I\pi(\omega) & \text{if } \omega \neq \omega^* \text{ and } \omega \models \phi \\ [0,0] & \text{otherwise.} \end{cases} \tag{4.3}$$

— *Otherwise, $\forall \omega \in \Omega$,*

$$I\pi(\omega|\phi) = \begin{cases} I\pi(\omega) & \text{if } \omega \models \phi \\ [0,0] & \text{otherwise } (\omega \not\models \phi) \end{cases} \tag{4.4}$$

Proof. The proof of this proposition can be obtained from the proof of postulate **IC2** for counter-models of ϕ.

For models of ϕ, since $\overline{I\Pi}(\phi) = 1$ then ϕ is already accepted in each $\pi \in \mathcal{C}(I\pi)$ (namely, $\Pi(\phi) = 1$ for all $\pi' \in \mathcal{C}(I\pi)$) and $\forall \omega \models \phi$, $I\pi(\omega|\phi) = I\pi(\omega)$. □

Example 4.1. Let $I\pi$ be the interval-based distribution depicted in Table 4.1. In this example, the new

$\omega \in \Omega$	$I\pi(\omega)$
ab	$[1,1]$
$a\bar{b}$	$[.3,.6]$
$\bar{a}b$	$[.1,.4]$
$\bar{a}\bar{b}$	$[.3,.6]$

$\omega \in \Omega$	$I\pi(\omega\|\phi)$
ab	$[1,1]$
$a\bar{b}$	$[.3,.6]$
$\bar{a}b$	$[0,0]$
$\bar{a}\bar{b}$	$[0,0]$

Table 4.1 – Interval-based possibility distribution of Example 4.1.

evidence $\phi = a$ is already accepted. Following Proposition 4.3, the resulted distribution $I\pi(\omega|\phi)$ is also given in Table 4.1.

We now consider the complex case where $\overline{I\Pi}(\bar{\phi}) = 1$, namely there exists at least a compatible possibility distribution π where ϕ is not accepted. Recall that by Equation (1.28), we have $\forall \omega \in \phi$, $\pi(\omega|\phi) = \dfrac{\pi(\omega)}{\Pi(\phi)}$. Therefore, intuitively to get, for instance, the lower endpoint $\underline{I\pi}(\omega|\phi)$, it is enough to select a compatible distribution π that provides the smallest value for $\pi(\omega)$ (namely, if possible $\pi(\omega) = \underline{I\pi}(\omega)$) and the largest value for $\Pi(\phi)$ (namely, if possible $\Pi(\phi) = \overline{I\Pi}(\phi)$). Hence, in the following and without surprise we will constally find the expression $\dfrac{\underline{I\pi}(\omega)}{\overline{I\Pi}(\phi)}$ in the definition of $I\pi(\omega|\phi)$. The following two propositions give these bounds depending whether there exist a unique interpretation or several interpretations having their upper endpoints equal to $\overline{I\Pi}(\phi)$.

Proposition 4.4. *Let $I\pi$ be an interval-based distribution such that $\overline{I\Pi}(\neg\phi) = 1$. If there exist more than one model of ϕ having their upper endpoints equal to $\overline{I\Pi}(\phi)$, then $\forall \omega \in \Omega$:*

$$I\pi(\omega|\phi) = \begin{cases} \left[\dfrac{\underline{I\pi}(\omega)}{\overline{I\Pi}(\phi)}, \min\left(1, \dfrac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)}\right)\right] & \text{if } \omega \models \phi \\ [0,0] & \text{otherwise} \end{cases} \tag{4.5}$$

Proof. Let $\omega \in \Omega$ be an interpretation and $I\pi(.|\phi)$ be the conditioned interval-based distribution $I\pi$ with ϕ.

— The lower endpoint $\underline{I\pi}(\omega|\phi)$ equals $\dfrac{\underline{I\pi}(\omega)}{\overline{I\Pi}(\phi)}$. Indeed, this possibility degree $\dfrac{\underline{I\pi}(\omega)}{\overline{I\Pi}(\phi)}$ exists and is obtained by considering a compatible possibility distribution π where $\Pi(\phi) = \overline{I\Pi}(\phi)$ (recall that

$\pi(\omega) \leq \Pi(\phi)$). Besides, since for each compatible possibility distribution π' we have $\Pi'(\phi) \leq \overline{I\Pi}(\phi)$ and $\pi'(\omega) \geq \underline{I\pi}(\omega)$ then $\pi'(\omega|\phi) = \frac{\pi'(\omega)}{\Pi'(\phi)} \geq \frac{\underline{I\pi}(\omega)}{\overline{I\Pi}(\phi)}$.

— Similarly, the upper endpoint $\overline{I\pi}(\omega|\phi)$ is equal to $\min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)})$. Again, this possibility degree $\min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)})$ exists and is obtained, by considering a compatible possibility π where $\Pi(\phi) = \underline{I\Pi}(\phi)$ and $\pi(\omega) = \min(\underline{I\Pi}(\phi), \overline{I\pi}(\omega))$. Such compatible possibility distribution exists. Let us show that for every compatible possibility distribution π', we have $\pi'(\omega|\phi) \leq \min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)})$.
Two cases are to be considered:
  — $\overline{I\pi}(\omega) < \underline{I\Pi}(\phi)$. For every compatible possibility distribution π', we have $\pi'(\omega) \leq \overline{I\pi}(\omega)$. Hence $\pi'(\omega) \leq \min(\underline{I\Pi}(\phi), \overline{I\pi}(\omega))$ since $\overline{I\pi}(\omega) < \underline{I\Pi}(\phi)$. Besides, by definition $\Pi'(\phi) \geq \underline{I\Pi}(\phi)$. Therefore,
  $\pi'(\omega|\phi) = \frac{\pi'(\omega)}{\Pi'(\phi)} \leq \frac{\min(\underline{I\Pi}(\phi), \overline{I\pi}(\omega))}{\underline{I\Pi}(\phi)} = \min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)})$.
  — If $\overline{I\pi}(\omega) \geq \underline{I\Pi}(\phi)$ then trivially: $\pi'(\omega|\phi) = \frac{\pi'(\omega)}{\Pi'(\phi)} \leq \min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)})$ since $\min(1, \frac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)}) = 1$.

$\square$

Example 4.2. Let $I\pi$ be the normalized interval-based distribution of Table 4.2. Let $\phi = \overline{a}$ be the new evidence. In this example, we face the situation where we have more than one interpretation where $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi) = .6$. Hence, according to Proposition 4.4, the resulted distribution is given in Table 4.2.

$\omega \in \Omega$	$I\pi(\omega)$
ab	$[1, 1]$
$a\overline{b}$	$[.1, .4]$
$\overline{a}b$	$[.3, .6]$
$\overline{a}\overline{b}$	$[.3, .6]$

$\omega \in \Omega$	$I\pi(\omega\|\phi)$
ab	$[0, 0]$
$a\overline{b}$	$[0, 0]$
$\overline{a}b$	$[1/2, 1]$
$\overline{a}\overline{b}$	$[1/2, 1]$

Table 4.2 – Interval-based possibility distribution of Example 4.2.

For the purpose of the next proposition, we define the notion of *secondbest* in an interval-based possibility distribution.

Definition 4.7 (Secondbest)**.** Let $I\pi$ be an interval-based possibilistic distribution. Then,

$$secondbest(I\pi) = \max\{\overline{I\pi}(\omega') : \omega' \in \phi \text{ and } \overline{I\pi}(\omega') \neq \overline{I\Pi}(\phi)\}.$$

Note that this definition can be extended to interval-based possibilistic knowledge bases.

Example 4.3. Let us take the interval-based possibility distribution $I\pi(\omega)$ of Table 4.2. Then for $\phi = ab, a\overline{b}, \overline{a}b$ we have $secondbest(I\pi) = .6$. Indeed, the three interpretations that belongs to ϕ have respectively the upper bound equals to 1, $.4$ and $.6$. So the second best degree is $.6$.

The next proposition concerns the particular situation where there exists exactly one interpretation ω^*, model of ϕ, such that $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi)$. In this case, comparing to Proposition 4.4, only the lower endpoint of the interpretation ω^* will differ. More precisely:

Proposition 4.5. *Let $I\pi$ be an interval-based possibility distribution such that $\overline{I\Pi}(\overline{\phi}) = 1$. Assume that there exists exactly one interpretation ω^*, model of ϕ, such that $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi)$. Then $\forall \omega \in \Omega$,*

— *If $\omega \neq \omega^*$ then $I\pi(\omega|\phi)$ is the same as the one given in Proposition 4.4, namely:*

$$I\pi(\omega|\phi) = \begin{cases} \left[\dfrac{\underline{I\pi}(\omega)}{\overline{\overline{I\Pi}}(\phi)}, \min\left(1, \dfrac{\overline{I\pi}(\omega)}{\underline{I\Pi}(\phi)}\right)\right] & \text{if } \omega \models \phi \\ [0,0] & \text{otherwise} \end{cases} \tag{4.6}$$

— *If $\omega = \omega^*$, let $secondbest(I\pi)$ defined by Definition 4.7. Then:*

$$I\pi(\omega|\phi) = \begin{cases} [1,1] & \text{if } secondbest(I\pi) = 0 \\ \left[\min(1, \dfrac{\underline{I\pi}(\omega)}{secondbest(I\pi)}), 1\right] & \text{otherwise} \end{cases} \tag{4.7}$$

Proof. In the situation where there exists exactly one interpretation ω^* such that $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi)$. Then first $\forall \omega' \neq \omega^*$, we have :

$$I\pi(\omega'|\phi) = \left[\frac{\underline{I\pi}(\omega')}{\overline{\overline{I\Pi}}(\phi)}, \min\left(1, \frac{\overline{I\pi}(\omega')}{\underline{I\Pi}(\phi)}\right)\right]$$

The proof for this case is exactly the same as the one given in Proposition 4.4.
Now regarding the interpretation ω^*, there are two cases to consider :

— if $secondbest(I\pi) = 0$ then this means that $\forall \omega' \neq \omega^*$, $I\pi(\omega') = [0,0]$. Hence, for each compatible possibility distribution π we have $\pi(\omega^*|\phi) = 1$ and $\forall \omega' \neq \omega^*$, $\pi(\omega'|\phi)$ =0. Hence, $I\pi(\omega^*|\phi) = [1,1]$ and $\forall \omega' \neq \omega^*$, $I\pi(\omega'|\phi) = [0,0]$.

— if $secondbest(I\pi) > 0$ then this means that $\exists \omega' \neq \omega^*$, such that $I\pi(\omega') \neq [0,0]$. In this case,

$$I\pi(\omega^*|\phi) = \left[\min(\frac{\underline{I\pi}(\omega^*)}{secondbest(I\pi)}, 1), 1\right].$$

The upper endpoint (namely 1) is obtained by considering a compatible possibility distribution π where $\pi(\omega^*) = \overline{I\Pi}(\phi)$. The lower endpoint possibility degree exists and is obtained by considering another compatible possibility distribution π' where $\pi'(\omega^*) = \min(secondbest(I\pi), \underline{I\pi}(\omega^*))$ and for some ω', $\pi'(\omega') = secondbest(I\pi)$ (such ω' exists by assumption that $secondbest(I\pi) \neq 0$). Two cases are to be considered:

— If $secondbest(I\pi) > \underline{I\pi}(\omega^*)$, then one can check that for each compatible possibility distribution π" we have $\pi"(\omega^*) \geq \underline{I\pi}(\omega^*)$ (by definition of lower endpoints) and $\Pi"(\phi) \leq secondbest(I\pi)$. Therefore, $\pi''(\omega^*|\phi) = \dfrac{\pi"(\omega^*)}{\Pi"(\phi)} \geq \dfrac{\underline{I\pi}(\omega^*)}{secondbest(I\pi)}$. Therefore, $\underline{I\pi}(\omega^*|\phi) = \min_{\pi' \in \mathcal{C}(I\pi)}(\pi'(\omega^*|\phi)) = \dfrac{\underline{I\pi}(\omega)}{secondbest(I\pi)}$

— If $secondbest(I\pi) \leq \underline{I\pi}(\omega^*)$, then trivially, for each compatible possibility distribution π", $\pi''(\omega^*|\phi) \geq \min(\dfrac{\underline{I\pi}(\omega^*)}{secondbest(I\pi)}, 1)$ since $\min(\dfrac{\underline{I\pi}(\omega^*)}{secondbest(I\pi)}, 1) = 1$.

□

Example 4.4. Let $I\pi$ be the normalized interval-based distribution of Table 4.3. Let $\phi = \overline{a}$ be the new evidence. In this example, we face the situation where we have exactly one interpretation where $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi) = .6$. Hence, according to Proposition 4.5, $secondbest(I\pi) = .4$.

$\omega \in \Omega$	$I\pi(\omega)$
ab	$[1, 1]$
$a\overline{b}$	$[.3, .6]$
$\overline{a}b$	$[.1, .4]$
$\overline{a}\overline{b}$	$[.3, .6]$

$\omega \in \Omega$	$I\pi(\omega\vert\phi)$
ab	$[0, 0]$
$a\overline{b}$	$[0, 0]$
$\overline{a}b$	$[.1/.6, 1]$
$\overline{a}\overline{b}$	$[.3/.4, 1]$

Table 4.3 – Example of conditioning an interval-based possibility distribution using Proposition 4.5.

The nice feature of the proposed conditioning is that it extends the one used in standard possibility theory: namely when all intervals I, associated with interpretations, are singletons, then $\forall\omega \in \Omega$, $I\pi(\omega|\phi)$= $[\pi(\omega|\phi)$,$\pi(\omega|\phi)]$ where π is the unique compatible distribution associated with $I\pi$.

We now provide the syntactic counterpart of the compatible-based conditioning.

4.3.2 Syntactic characterization of compatible-based conditioning

Given an interval-based knowledge base IK and a new evidence ϕ, conditioning at the syntactic level comes down to altering IK into IK_ϕ such that the induced posterior interval-based possibility distribution $I\pi_{IK_\phi}$ equals the posterior interval-based possibility distribution $I\pi_{IK}(.|\phi)$ obtained by conditioning $I\pi_{IK}$ with ϕ as illustrated in Figure 4.2.

$$\begin{tikzcd}
IK \arrow[rr, "\text{Definition 4.5}"] \arrow[d, "\substack{\text{Syntactic}\\\text{conditioning}\\\text{with } \phi}"'] & & I\pi_{IK} \arrow[d, "\substack{\text{Conditioning}\\\text{on } \phi \text{ using}\\\text{Propositions 4.4 and 4.5}}"] \\
IK_\phi \arrow[r, "\text{Definition 4.5}"] & I\pi_{IK_\phi} \arrow[r, equal, "?"] & I\pi_{IK}(.|\phi)
\end{tikzcd}$$

Figure 4.2 – Equivalence of semantic and syntactic conditionings.

The aim of this section is then to compute a new interval-based knowledge base, denoted for the sake of simplicity by IK_ϕ, such that:

$$\forall\omega\in\Omega,\ I\pi_{IK}(\omega|\phi) = I\pi_{IK_\phi}(\omega), \tag{4.8}$$

where $I\pi_{IK_\phi}$ is the interval-based distribution associated with IK_ϕ using Definition 4.5, and $I\pi_{IK}(.|\phi)$ is the result of conditioning $I\pi_{IK}$ using the compatible-based conditioning presented in the previous section (Propositions 4.4 and 4.5).

To achieve this aim, we need to provide methods that directly operate on the interval-based knowledge base IK:

— to check whether $\overline{I\Pi}_{IK}(\phi) = 0$ (resp. $\underline{I\Pi}_{IK}(\phi) = 0$) or not,

— to check whether $\overline{I\Pi}_{IK}(\neg\phi) = 1$ or not,

— to compute $\underline{I\Pi}_{IK}(\phi)$ and $\overline{I\Pi}_{IK}(\phi)$,

— to compute $secondbest(I\pi_{IK})$,

— to check whether there exists a unique interpretation ω^* such that $\overline{I\pi}(\omega^*) = \overline{I\Pi}(\phi)$, and lastly

— to compute IK_ϕ.

Checking whether $\overline{I\Pi}_{IK}(\phi) = 0$ (resp. $\underline{I\Pi}_{IK}(\phi) = 0$) or not

Recall that an interval-based possibility distribution where $\overline{I\Pi}_{IK}(\phi) = 0$ expresses a very strong conflict with the evidence ϕ. Namely, IK strongly contradicts the formula ϕ.

Proposition 4.6. *Let IK be an interval-based possibilistic base and $I\pi_{IK}$ be its associated interval-based distribution. Then,*

i) $\overline{I\Pi}_{IK}(\phi) = 0$ *if and only if* $\{\varphi : (\varphi, I) \in IK \text{ and } I = [1,1]\} \cup \{\phi\}$ *is inconsistent. In this case,* $IK_\phi = \emptyset$.

ii) $\underline{I\Pi}_{IK}(\phi) = 0$ *if and only if* $\{\varphi : (\varphi, I) \in IK \text{ and } \overline{I} = 1\} \cup \{\phi\}$ *is inconsistent. In this case,* $IK_\phi = \{(\phi, [1,1]), (\neg\phi, [0,1])\}$.

Proof. — *Proof of item i)*: Since $\{\varphi : (\varphi, I) \in IK \text{ and } I = [1,1]\} \cup \{\phi\}$ is inconsistent then each interpretation ω falsifies at least one formula with an interval-based weight equal to [1,1]. Using Definition 4.5, we conclude that $\forall \omega \in \Omega$, ω model of ϕ, $I\pi_{IK}(\omega) = [0,0]$. Hence, $I\pi_{IK}$ is a degenerate interval-based possibility distribution. Similarly, if $\{\varphi : (\varphi, I) \in IK \text{ and } I = [1,1]\} \cup \{\phi\}$ is consistent then it accepts a model ω such that the maximal lower endpoint of intervals associated with formulas falsified by ω is strictly less than 1, hence $\overline{I\pi}_{IK}(\omega) > 0$. Therefore, $\overline{I\Pi}_{IK}(\phi) \neq 0$.

Lastly, one can easily check that the possibility distribution associated with $IK_\phi = \phi$ is: $\forall \omega \in \Omega$, $I\pi_{IK_\phi}(\omega) = [1,1]$, which is exactly the result of conditioning interval-based possibility distribution with $\overline{I\Pi}(\phi) = 0$.

— *Proof of item ii)*: Remember that IK is consistent, namely there exists $\omega^* \in \Omega$ such that $\overline{I\pi}(\omega^*) = 1$. Since $\{\varphi : (\varphi, I) \in IK \text{ and } \overline{I} = 1\} \cup \{\phi\}$ is inconsistent then each interpretation ω, model of ϕ, falsifies at least one formula with an upper endpoint weight equal to 1. Using Definition 4.5, we conclude that $\forall \omega \in \Omega$, ω model of ϕ, $I\pi_{IK}(\omega) = [0, \beta]$. Hence, $\underline{I\Pi}_{IK}(\phi) = 0$. Similarly, if $\{\varphi : (\varphi, I) \in IK \text{ and } \overline{I} = 1\} \cup \{\phi\}$ is consistent then it accepts a model ω such that the maximal upper endpoint of intervals associated with formulas falsified by ω is strictly less than 1, hence $\underline{I\pi}_{IK}(\omega) > 0$. Therefore, $\underline{I\Pi}_{IK}(\phi) \neq 0$. Again, one can easily check that the possibility distribution associated with $IK_\phi = \{(\phi, [1,1]),\ (\neg\phi, [0,1])\}$ is $\forall \omega \in \Omega$,

$$I\pi_{IK_\phi}(\omega) = \begin{cases} [0,1] & \text{if } \omega \models \phi \\ [0,0] & \text{otherwise} \end{cases}$$

. This is the result of conditioning $I\pi$ by ϕ when $\underline{I\Pi}(\phi) = 0$.

□

Example 4.5. Let $IK = \{(\neg a, [1,1]), (a \vee \neg b, [.4, .6])\}$ be an interval-based possibilistic knowledge base. The associated interval-based possibility distribution is given in Table 5.2. Let $\phi = a$ be the new evidence.

In this example, $\overline{I\Pi}_{IK}(\phi) = 0$ since $\{\varphi : (\varphi, I) \in IK \text{ and } I = [1,1]\} \cup \{\phi\} = \{\neg a\} \cup \{a\}$ is inconsistent. Hence, $IK_\phi = \emptyset$.

In the following, we assume that IK is such that ϕ is somewhat possible. Namely in its associated interval-based possibility distribution $I\pi_{IK}$, we have $\underline{I\Pi}_{IK}(\phi) > 0$.

Checking whether $\overline{I\Pi}_{IK}(\neg\phi) \neq 1$ or not

Proposition 4.7 shows how to syntactically check if ϕ is accepted or not, namely whether $\overline{I\Pi}_{IK}(\neg\phi) = 1$ or not.

$\omega \in \Omega$	$I\pi_{IK}(\omega)$
ab	$[0,0]$
$a\neg b$	$[0,0]$
$\neg ab$	$[.4,.6]$
$\neg a\neg b$	$[1,1]$

$\omega \in \Omega$	$I\pi_{IK}(\omega\|\phi)$
ab	$[1,1]$
$a\neg b$	$[1,1]$
$\neg ab$	$[1,1]$
$\neg a\neg b$	$[1,1]$

Table 4.4 – Interval-based possibility distribution induced by the interval-based possibilistic base of Example 4.5.

Proposition 4.7. *Let IK be an interval-based possibilistic base and $I\pi_{IK}$ be its associated possibility distribution. Then: $\overline{I\Pi}_{IK}(\neg\phi) \neq 1$ if and only if $\{\varphi : (\varphi, I) \in IK$ and $\underline{I} > 0\} \cup \{\neg\phi\}$ is inconsistent. In this case: $IK_\phi = IK \cup \{(\phi, [1,1])\}$.*

Namely, $\overline{I\Pi}(\neg\phi) \neq 1$ if the set of somewhat certain formulas of K (namely having their lower endpoints degrees greater than 0) is inconsistent with $\neg\phi$.

Proof. Assume that $\{\varphi : (\varphi, I) \in IK$ and $\underline{I} > 0\} \cup \{\neg\phi\}$ is inconsistent. Then each interpretation ω, model of $\neg\phi$, falsifies at least one formula with a lower endpoint weight strictly greater than 0. Using Definition 4.5, we conclude that $\forall\omega \in \Omega$, if $\omega \models \neg\phi$ then $\overline{I\pi}_{IK}(\omega) < 1$. Hence $\overline{I\Pi}_{IK}(\neg\phi) \neq 1$. Similarly, if $\{\varphi : (\varphi, I) \in IK$ and $\underline{I} > 0\} \cup \{\neg\phi\}$ is consistent then it accepts a model ω that satisfies all formulas (φ, I) with $\underline{I} > 0$. Hence, $\overline{I\pi}_{IK}(\omega) = 1$ and ω is a model of $\neg\phi$. Hence $\overline{I\Pi}(\neg\phi) = 1$. Now, it is easy to check that the joint interval-based possibility distribution associated with $IK_\phi = IK \cup \{(\phi, [1,1])\}$ is:

$$\forall\omega \in \Omega, I\pi_{IK_\phi} = \begin{cases} I\pi(\omega) & \text{if } \omega \models \phi \\ [0,0] & \text{otherwise} \end{cases}$$

which is exactly the one given in Proposition 4.3. □

Example 4.6. Let $IK = \{(\neg a, [0,.3]), (a \vee \neg b, [.4,1]), (a \wedge b, [.3,.6])\}$ be an interval-based possibilistic knowledge base. The associated interval-based possibility distribution is given in Table 4.5. Let $\phi = a$ be the new evidence. In this example, $\overline{I\Pi}_{IK}(\phi) \neq 1$ since $\{\varphi : (\varphi, I) \in IK$ and $\underline{I} > 1\} \cup \{\neg\phi\} =$

$\omega \in \Omega$	$I\pi_{IK}(\omega)$
ab	$[.7,.1]$
$a\neg b$	$[.4,.7]$
$\neg ab$	$[0,.6]$
$\neg a\neg b$	$[.4,.7]$

$\omega \in \Omega$	$I\pi_{IK_\phi}(\omega)$
ab	$[.7,1]$
$a\neg b$	$[.4,1]$
$\neg ab$	$[0,0]$
$\neg a\neg b$	$[0,0]$

Table 4.5 – Interval-based possibility distribution induced by the interval-based possibilistic base (IK and IK_ϕ) of Example 4.6.

$\{a \vee \neg b, a \wedge b, \neg a\}$ is inconsistent. Hence, $IK_\phi = IK \cup \{(\phi, [1,1])\}$. The associated distribution of IK_ϕ is given in Table 4.5.

Computing $\underline{I\Pi}_{IK}(\phi)$ and $\overline{I\Pi}_{IK}(\phi)$

The computation of $\underline{I\Pi}_{IK}(\phi)$ and $\overline{I\Pi}_{IK}(\phi)$ comes down to computing the inconsistency degrees of two particular standard possibilistic knowledge bases (considering only lower and upper endpoints of intervals associated with formulas) as it is stated by the following proposition:

Proposition 4.8. *Let IK be an interval-based knowledge base. Let $\underline{IK} = \{(\varphi, \underline{I}) : (\varphi, I) \in IK\}$ and $\overline{IK} = \{(\varphi, \overline{I}) : (\varphi, I) \in IK\}$. Then:*

$\underline{I\Pi}_{IK}(\phi) = 1 - Inc(\overline{IK} \cup \{(\phi, 1)\})$ *and*

$\overline{I\Pi}_{IK}(\phi) = 1 - Inc(\underline{IK} \cup \{(\phi, 1)\})$.

In Proposition 4.8, $Inc(K)$ is the inconsistency degree of a standard possibilistic knowledge base K given by Equation (2.20).

Proof. Let us start with the computation of $\underline{I\Pi}_{IK}(\phi)$. Recall that:

$$\underline{I\Pi}_{IK}(\phi) = \max\{\underline{I\pi}_{IK}(\omega) : \omega \in \Omega \text{ and } \omega \models \phi\}$$

and $I\pi_{IK}(\omega)$ is defined as follows:

$$I\pi_{IK}(\omega) = \begin{cases} [1,1] & \text{if } \forall(\varphi, I) \in IK,\ \omega \models \varphi \\ [1 - \max\{\overline{I} : (\varphi_i, I) \in K, \omega \not\models \varphi_i\}, 1 - \max\{\underline{I} : (\varphi_i, I) \in K, \omega \not\models \varphi_i\}] & \text{otherwise.} \end{cases}$$

If there exists ω, model of ϕ, such that $\forall(\varphi, I) \in IK,\ \omega \models \varphi$ then $I\Pi_{IK}(\phi) = [1, 1]$. This means that $IK^* = \{\varphi : (\varphi, I) \in IK\} \cup \{\phi\}$ is consistent. From the definition of inconsistency of a base, $Inc(\overline{IK}) = 0$ hence $\underline{I\Pi}_{IK}(\phi) = 1$.

Assume now that $IK^* = \{\varphi : (\varphi, I) \in IK\} \cup \{(\phi, [1, 1])\}$ is inconsistent. Then, by definition:

$$\begin{aligned} \underline{I\Pi}_{IK}(\phi) &= \max\{\underline{I\pi}_{IK}(\omega) : \omega \models \phi\} \\ &= \max\{1 - \max\{\overline{I} : (\varphi, I) \in IK,\ \omega \not\models \varphi\} : \omega \models \phi\} \\ &= 1 - \min\{\max\{\overline{I} : (\varphi, I) \in IK,\ \omega \not\models \varphi\} : \omega \models \phi\} \end{aligned}$$

Let us denote $\alpha = \min\{\max\{\overline{I} : (\varphi, I) \in IK,\ \omega \not\models \varphi\} : \omega \models \phi\}$
This means that:

$$\forall\omega \models \phi, \max\{\overline{I} : (\varphi, I) \in IK,\ \omega \not\models \varphi\} \geq \alpha \tag{4.9}$$

and

$$\exists\omega' \models \phi, \max\{\overline{I} : (\varphi, I) \in IK,\ \omega' \not\models \varphi\} = \alpha \tag{4.10}$$

The last equation means that:

$$\{\varphi : (\varphi, I) \in IK, \text{and } \overline{I} > \alpha\} \cup \{(\phi, 1)\} \text{ is consistent,}$$

and Equation (4.9) means that:

$$\{\varphi : (\varphi, I) \in IK, \text{and } \overline{I} \geq \alpha\} \cup \{(\phi, 1)\} \text{ is inconsistent.}$$

The degree α corresponds simply to the definition of inconsistency degree of the standard possibilistic knowledge bases obtained from IK by considering only the upper endpoints of intervals associated with formulas and ϕ with a degree 1 of certainty. Hence, we have: $\underline{I\Pi}_{IK}(\phi) = 1 - Inc(\overline{IK} \cup \{(\phi, 1)\})$.

The computation of $\overline{I\Pi}_{IK}(\phi)$ follows similar steps as the ones used in computing $\underline{I\Pi}_{IK}(\phi)$. □

Example 4.7. Let $IK = \{(a \wedge b, [.4, .7]), (a \vee \neg b, [.6, .9])\}$ be a interval-based possibilistic base and let $\phi = \neg a$ be a new evidence in hand. The associated distribution $I\pi_{IK}$ of IK is depicted in Table 4.6. From Table 4.6, we can compute $I\Pi_{IK}(\phi)$. Indeed, $I\Pi_{IK}(\phi) = [.3, .6]$. From the knowledge base IK, let us compute $\overline{I\Pi}(\phi) = 1 - Inc(\underline{IK} \cup \{(\phi, 1)\})$, $Inc(\underline{IK} \cup \{(\phi, 1)\}) = .4$ then $1 - .4 = .6$. The same goes for computing $\underline{I\Pi}(\phi)$.

ω	$I\pi_{IK}(\omega)$
ab	$[1,1]$
$a\neg b$	$[.3,.6]$
$\neg ab$	$[.1,.4]$
$\neg a\neg b$	$[.3,.6]$

Table 4.6 – Interval-based possibility distribution induced by IK of Example 4.7.

Checking the uniqueness of models of ϕ having upper endpoints equal to $\overline{I\Pi}_{IK}(\phi)$

We now need to show how to syntactically check whether, there exists a unique interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$. If an interpretation ω, model of ϕ, is such that $\overline{I\pi}_{IK}(\omega) = \overline{I\Pi}_{IK}(\phi)$ then ω is a model of $\Phi = \{\varphi : (\varphi, I) \in IK$ and $\underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\} \cup \{\phi\}$. Besides, if for some $\omega' \neq \omega$, $\overline{I\pi}_{IK}(\omega') < \overline{I\Pi}_{IK}(\phi)$ then this means that ω' falsifies at least one formula from $\Phi \cup \{\phi\}$. Additionally, assume that there exists a unique model ω^* of ϕ such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$. We are interested to know whether $\forall\omega' \neq \omega^*$, $I\pi(\omega') = [0,0]$. It is enough to check that all formulas in $\{\varphi : (\varphi, I) \in IK$and $\underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\}$ have their associated interval I equal to [1,1]. The main results of this subsection are summarized in the following proposition:

Proposition 4.9. *Let IK be an interval-based knowledge base. Let $I\pi_{IK}$ be its associated possibility distribution. Let $\Phi = \{\varphi$: $(\varphi, I) \in IK$ and $\underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\} \cup \{\phi\}$. Then:*

— $\Phi \cup \{\phi\}$ *admits a unique model if and only if there exists a unique interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$.*

— $\Phi \cup \{\phi\}$ *admits a unique model and each formula in Φ has $[1,1]$ as certainty-based interval weight if and only if there exists ω^* model of ϕ such that $\overline{I\pi}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ and $\forall\omega' \neq \omega^*$, $I\pi(\omega') = [0,0]$.*

Proof. — Proof of the first item: Let us show the proofs in both the directions. But, first recall that: $\overline{I\Pi}(\phi) = 1 - Inc(\underline{IK} \cup \{(\phi,1)\})$ from Proposition 4.8. Namely,

i) If $\Phi \cup \{\phi\}$ admits a unique model ω^* then $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$.
Since ω^* is the unique model of $\Phi \cup \{\phi\}$, then it falsifies all formulas of Φ. Recall that, all the formulas of Φ have, in IK, their lower endpoints greater than $Inc(\underline{IK} \cup \{(\phi,1)\})$. Then, ω^* falsifies formulas of IK, such that their endpoints are equal to $Inc(\underline{IK} \cup \{(\phi,1)\})$. Consequently $\overline{I\pi}_{IK}(\omega^*) = 1 - Inc(\underline{IK} \cup \{(\phi,1)\}) = \overline{I\Pi}_{IK}(\phi)$. Since ω^* is the unique model of $\Phi \cup \{\phi\}$ then $\forall\omega \in \Omega$, $\omega \neq \omega^*$, ω model of ϕ, ω falsifies at least a formula from Φ. Since all formulas in Φ have their endpoint greater than $Inc(\underline{IK} \cup \{(\phi,1)\})$ then $\overline{I\pi}(\omega) < 1 - Inc(\underline{IK} \cup \{(\phi,1)\})$. Hence, there exists exactly one model ω^* of ϕ such that $\pi(\omega^*) = \Pi(\phi)$.

ii) If there is a unique interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ then ω^* is the unique model of $\Phi \cup \{\phi\}$. Assume that there exists a unique interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$. According to Definition 4.5, $\overline{I\pi}_{IK}(\omega^*) = 1 - \max\{\underline{I}, (\varphi, I) \in IK$ and $\omega^* \nvDash \varphi\}$. Now, since $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ and $\overline{I\Pi}_{IK}(\phi)$ =1-$Inc(\underline{IK} \cup \{(\phi,1)\})$ (Proposition 4.8) then $\max(\underline{I}, (\varphi, I) \in IK$ and $\omega^* \nvDash \varphi) = Inc(\underline{IK} \cup \{(\phi,1)\})$. This means that ω^* does not falsify any formula from Φ given that formulas of Φ are associated in $\underline{IK}$ with weights strictly greater than $Inc(\underline{IK} \cup \{(\phi,1)\})$. Then ω^* is a model of ϕ and Φ, consequently, it is a model of $\Phi \cup \{\phi\}$ meaning that $\Phi \cup \{\phi\}$ admits a model. But is ω^* the unique model of $\Phi \cup \{\phi\}$? Assume that there exists another interpretation $\omega' \neq \omega^*$ such that $\overline{I\pi}_{IK}(\omega') < \overline{I\pi}_{IK}(\omega^*)$ and ω' is a model of $\Phi \cup \{\phi\}$. ω' is a model of

$\Phi \cup \{\phi\}$ if and only if ω' does not falsify any formula in $\Phi \cup \{\phi\}$. In this case, $\overline{I\pi}_{IK}(\omega') \geq 1 - Inc(\underline{IK} \cup \{(\phi, 1)\}) = \overline{I\Pi}_{IK}(\phi)$, which is a contradiction.

From items i) and ii), we conclude that $\Phi \cup \{\phi\}$ admits a unique model if and only if there exists a unique interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$.

— Proof of the second item: $\Phi \cup \{\phi\}$ *admits a unique model and each formula in* Φ *has* $[1, 1]$ *as certainty-based interval weight if and only if there exists* ω^* *model of* ϕ *such that* $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ *and* $\forall \omega' \neq \omega^*$, $I\pi_{IK}(\omega') = [0, 0]$.

a) Assume that $\Phi \cup \{\phi\}$ admits a unique model ω^* and each formula in Φ is associated with the certainty interval $[1, 1]$. This means that any another interpretation $\omega' \neq \omega^*$ falsifies at least one formula from $\Phi \cup \{\phi\}$, consequently, the interval-based possibility interval-degrees of ω' will be associated with the interval $[0, 0]$.

b) The converse can be stated as follows: Assume that there exists a unique ω^* model of ϕ such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ and $\forall \omega' \neq \omega^*$, $I\pi_{IK}(\omega') = [0, 0]$. This means that $\forall \omega' \neq \omega^*$, ω' falsifies at least a formula with the interval $[1, 1]$. Then, $\Phi \cup \{\phi\}$ is satisfiable and admits only ω^* as a model. Hence, $\Phi \cup \{\phi\}$ admits a unique model and all its formulas are associated with the certainty interval $[1, 1]$.

□

Example 4.8 (First item of Proposition 4.9)**.** Let $IK = \{(a \wedge b, [.4, .7]), (a \vee \neg b, [.6, .9])\}$ be an interval-based possibilistic base and let $\phi = \neg a$ be a new evidence in hand. The associated distribution $I\pi_{IK}$ of IK is already depicted in Table 4.6. In this example, we consider $\Phi = \{a \vee \neg b\} \cup \{\neg a\}$, which admits a unique model $\omega^* = \neg a \neg b$. From Proposition 4.9, we have indeed $\neg a \neg b$, model of ϕ, being the only interpretation having his upper endpoint equal to $\overline{I\Pi}_{IK}(\phi)$.

Example 4.9 (Second item of Proposition 4.9)**.** Let $IK = \{(a \wedge b, [.4, .7]), (a \vee \neg b, [1, 1])\}$ be a interval-based possibilistic base and let $\phi = \neg a$ be a new evidence in hand. The associated distribution $I\pi_{IK}$ of IK is depicted in Table 4.7. Here, we consider $\Phi = \{a \vee \neg b\} \cup \{\neg a\}$, which admits a unique

ω	$I\pi_{IK}(\omega)$
ab	$[1, 1]$
$a\neg b$	$[.3, .6]$
$\neg ab$	$[0, 0]$
$\neg a \neg b$	$[.3, .6]$

Table 4.7 – Interval-based possibility distribution induced by IK of Example 4.9.

model $\omega^* = \neg a \neg b$ and each formula of Φ meaning $\psi = a \vee \neg b$ has a weight equal to $[1, 1]$. From Proposition 4.9, we have $\neg a \neg b$, model of ϕ, being the unique interpretation having his upper endpoint equal to $\overline{I\Pi}_{IK}(\phi)$ and $\forall \omega' \models \phi$, $\omega' \neq \omega^*$, we have $I\pi(\omega') = [0, 0]$.

Computing $secondbest(IK)$

Recall that $\underline{IK} = \{(\varphi, \underline{I}) : (\varphi, I) \in IK\}$ and that $secondbest(IK)$ is only computed in the situation where there exists exactly one interpretation ω^*, model of ϕ, such that $\overline{I\Pi}(\phi) = \overline{I\pi}(\omega^*)$. In order to easily define $secondbest(I\pi_{IK})$, we first let $\mathcal{D} = \{\alpha_1, \ldots, \alpha_n\}$ to be the different degrees present in

$\underline{IK}$, with $\alpha_1 > \ldots > \alpha_n$. Then we define $(A_{\alpha_1}, A_{\alpha_2}, \ldots, A_{\alpha_n})$ as the WOP (well ordered partition) associated with $\underline{IK}$, obtained by letting:

$$A_{\alpha_i} = \{(\psi, \beta) : (\psi, \beta) \in \underline{IK} \text{ and } \beta = \alpha_i\}. \tag{4.11}$$

Namely, A_{α_i} is the subset of $\underline{IK}$ composed of all weighted formulas having a certainty degree equal to α_i. Then:

Proposition 4.10. *Assume that there exists exactly one interpretation ω^*, model of ϕ, such that $\overline{I\Pi}_{IK_\phi}(\phi) = \overline{I\pi}_{IK_\phi}(\omega^*)$. Let $(A_{\alpha_1}, A_{\alpha_2}, \ldots, A_{\alpha_n})$ be the WOP associated with $\underline{IK}$, where A_{α_i}'s are given by Equation (4.11). Define $secondbest(IK) = 1 - \min\{\alpha_i : \alpha_i > Inc(\underline{IK} \cup \{(\phi, 1)\})$ and A_{α_i} is a non-tautological formula $\}$. Then $secondbest(IK) = secondbest(I\pi_{IK})$.*

Proof. To see the proof, first recall that $\overline{I\Pi}_{IK}(\phi) = 1 - Inc(\underline{IK} \cup \{(\phi, 1)\})$ (Proposition 4.8). Recall also that there exists exactly one interpretation ω^*, model of ϕ, such that $\overline{I\pi}_{IK_\phi}(\omega) = \overline{I\Pi}_{IK_\phi}(\phi)$. All interpretations ω, different from ω^*, falsify at least one formula from IK having their lower endpoint greater than $Inc(\underline{IK} \cup \{(\phi, 1)\})$. Therefore, $secondbest(I\pi)$ is obtained by considered the smallest α_i such that A_{α_i} is a non-tautological formula (otherwise such formula is always satisfied) and where $\alpha_i > Inc(\underline{IK} \cup \{(\phi, 1)\})$. □

Example 4.10. Let $IK = \{(a \wedge b, [.4, .7]), (a \vee \neg b, [.6, .9])\}$ be a interval-based possibilistic base and let $\phi = \neg a$ be a new evidence in hand. The associated distribution $I\pi_{IK}$ of IK is depicted in the Example 4.7 - Table 4.6. In this example, ϕ admits a unique model $\omega^* = \neg a \neg b$ such that $\overline{I\Pi}_{IK_\phi}(\phi) = \overline{I\pi}_{IK_\phi}(\omega^*)$. From Proposition 4.10, the associated WOP with $\underline{IK}$ is given by $(A_{0.4}, A_{0.6})$ where $A_{0.4} = \{(a \wedge b, 0.4)\}$ and $A_{0.6} = \{(a \vee \neg b, 0.6)\}$. Let's compute $Inc(\underline{IK} \cup \{(\phi, 1)\}) = 0.4$ then $secondbest(IK) = 1 - 0.6 = 0.4$. Then we have $secondbest(IK) = secondbest(I\pi_{IK})$ ($secondbest(I\pi_{IK})$ is computed in Example 4.4).

Computing IK_ϕ

We are now ready to give the syntactic computation of IK_ϕ when $\overline{I\Pi}_{IK}(\neg\phi) = 1$. In order to simplify the notations, we now denote:

i) $\overline{\alpha}_I = 1 - \dfrac{1 - \overline{I}}{1 - Inc(\underline{IK} \cup \{(\phi, 1)\})}$

ii) $\underline{\alpha}_I = 1 - \dfrac{1 - \underline{I}}{1 - Inc(\overline{IK} \cup \{(\phi, 1)\})}$

iii) $2\alpha_I = 1 - \dfrac{1 - \overline{I}}{secondbest(IK)}$

iv) $\Phi = \{\varphi\text{: } (\varphi, I) \in IK \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\}$

The two following propositions provide the syntactic computation of IK_ϕ depending whether $\Phi \cup \{\phi\}$ admits more than one model or not:

Proposition 4.11 (General case: $\Phi \cup \{\phi\}$ has more than one model)**.** *Assume that $\Phi \cup \{\phi\}$ has strictly more than one model. Then:*

$$IK_\phi = \{(\phi, [1, 1])\} \cup \{(\varphi, [\max(0, \underline{\alpha}_I), \overline{\alpha}_I]) \colon (\varphi, I) \in IK, \text{ and } \underline{I} \geq Inc(\underline{IK} \cup \{(\phi, 1)\})\} \tag{4.12}$$

Proof. In order now to prove the proposition we have to show that $\forall\omega \in \Omega$, $I\pi_{IK_\phi} = I\pi_{IK}(.|\phi)$. First note that if an interpretation ω is not a model of ϕ, then by definition we have:

$$I\pi_{IK}(\omega) = I\pi_{IK}(\omega|\phi) = [0,0]$$

This is explained by the presence of $(\phi, [1,1])$ in IK_ϕ. Recall first that Φ has strictly more than one model which is equivalent to say that there exists more than one interpretation, model of ϕ, having its upper endpoint equal to $\overline{I\Pi}_{IK}(\phi)$. Let ω be an interpretation which is a model of ϕ. Then by Definition 4.5 we have two cases:

— ω is a model of Φ then by Definition 4.5 we have:

$$I\pi_{IK_\phi}(\omega) = [1,1].$$

Besides, from Proposition 4.1, we have:

$$I\pi_{IK}(\omega|\phi) = \left[\frac{\underline{I\pi}_{IK}(\omega)}{\overline{I\Pi}_{IK}(\phi)}, \min\left(1, \frac{\overline{I\pi}_{IK}(\omega)}{\underline{I\Pi}_{IK}(\phi)}\right)\right]$$

Since ω is a model of Φ we have: $\overline{I\pi}_{IK}(\omega) = 1 - Inc(\underline{IK} \cup \{(\phi,1)\})$. From, Proposition 4.8, we have $\overline{I\Pi}_{IK}(\phi) = 1 - Inc(\underline{IK} \cup \{(\phi,1)\})$. Hence, $\overline{I\pi}_{IK}(\omega) = \overline{I\Pi}_{IK}(\phi)$ and

$$\frac{\underline{I\pi}_{IK}(\omega)}{\overline{I\Pi}_{IK}(\phi)} = 1$$

Besides, since trivially, $\overline{I\Pi}_{IK}(\phi) \geq \underline{I\Pi}_{IK}(\phi)$, we also have:

$$\min\left(1, \frac{\overline{I\pi}_{IK}(\omega)}{\underline{I\Pi}_{IK}(\phi)}\right) = 1$$

Therefore,

$$I\pi_{IK}(\omega|\phi) = [1,1] = I\pi_{IK_\phi}(\omega).$$

— ω is a not a model of Φ then:

$$\begin{aligned}
\underline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\left\{1 - \frac{1-\overline{I}}{1 - Inc(\underline{IK} \cup \{(\phi,1)\})} : (\varphi, I) \in IK,\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\right\}. \\
&= 1 - \max\left\{1 - \frac{1-\overline{I}}{\overline{I\Pi}_{IK}(\phi)} : (\varphi, I) \in IK,\ \omega \not\models \varphi, \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\right\} \\
&= \min\left\{\frac{1-\overline{I}}{\overline{I\Pi}_{IK}(\phi)} : (\varphi, I) \in IK,\ \omega \not\models \varphi, \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\right\} \\
&= \frac{\min\left\{1-\overline{I} : (\varphi, I) \in IK\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi,1)\})\right\}}{\overline{I\Pi}_{IK}(\phi)} \\
&= \frac{\min\left\{1-\overline{I} : (\varphi, I) \in IK \text{ and } \omega \not\models \varphi\right\}}{\overline{I\Pi}_{IK}(\phi)} \\
&= \frac{1 - \max\left\{\overline{I} : (\varphi, I) \in IK \text{ and } \omega \not\models \varphi\right\}}{\overline{I\Pi}_{IK}(\phi)} \\
&= \frac{\underline{I\pi}_{IK}(\omega)}{\overline{I\Pi}_{IK}(\phi)}
\end{aligned}$$

$$\begin{aligned}
\overline{I\pi}_{IK_\phi}(\omega) &= 1-\max\left\{\max\left(0, 1-\frac{1-\underline{I}}{1-Inc(\overline{IK}\cup\{(\phi,1)\})}\right) : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and }\right.\\
&\quad \left.\underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\right\}.\\
&= 1-\max\left\{\max\left(0, 1-\frac{1-\underline{I}}{\underline{I\Pi}_{IK}(\phi))}\right) : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and } \underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\right\}.\\
&= \min\left\{1-\max\left(0, 1-\frac{1-\underline{I}}{\underline{I\Pi}_{IK}(\phi)}\right) : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and } \underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\right\}.\\
&= \min\left\{\min\left(1, \frac{1-\underline{I}}{\underline{I\Pi}_{IK}(\phi)}\right) : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and } \underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\right\}.\\
&= \min\left(1, \min\left\{\frac{1-\underline{I}}{\underline{I\Pi}_{IK}(\phi)} : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and } \underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\right\}\right).\\
&= \min\left(1, \frac{\min\{1-\underline{I} : (\varphi, I)\in IK,\ \omega\not\models\varphi \text{ and } \underline{I} > Inc(\underline{IK}\cup\{(\phi,1)\})\}}{\underline{I\Pi}_{IK}(\phi)}\right).\\
&= \min\left(1, \frac{\min\{1-\underline{I} : (\varphi, I)\in IK \text{ and } \omega\not\models\varphi\}}{\underline{I\Pi}_{IK}(\phi)}\right).\\
&= \min\left(1, \frac{1-\max\{\underline{I} : (\varphi, I)\in IK \text{ and } \omega\not\models\varphi\}}{\underline{I\Pi}_{IK}(\phi)}\right).\\
&= \min\left(1, \frac{\overline{I\pi}_{IK}(\omega)}{\underline{I\Pi}_{IK}(\phi)}\right).
\end{aligned}$$

Clearly, the obtained lower and upper endpoints are the same as the ones given in Proposition 4.1. Hence, we indeed have:

$$\forall\omega\in\Omega, I\pi_{IK_\phi}(\omega) = I\pi_{IK}(\omega|\phi).$$

□

Example 4.11. Let $IK = \{(a\wedge b, [.4, .7]), (\neg a\vee b, [.6, .9])\}$ be an interval-based possibilistic base. The interval-based possibility distribution $I\pi_{IK}$ corresponding to IK according to Definition 4.5 is given in Table 4.6.

Let us consider the new evidence being $\phi = \neg a$. From this example, $\Phi = \{\neg a\vee b\}$ and $\Phi\cup\{\phi\}$ has exactly two models. We face the case of Proposition 4.11. Therefore, $IK_\phi = \{(\neg a, [1, 1]), (a\wedge b, [0, 1/2]), (\neg a\vee b, [0, 5/6])\}$. Computing $I\pi_{IK_\phi}$ according to Definition 4.5, gives exactly the same distribution as the one of Example 4.2 when conditioned on $\phi = \neg a$ using Proposition 4.1.

Proposition 4.12 (Particular case: $\Phi\cup\{\phi\}$ has exactly one model)**.** *Assume that $\Phi\cup\{\phi\}$ admits a unique model.*

1. *If each formula in Φ has an interval equal to $[1, 1]$, then: $IK_\phi = \{(\varphi, [1, 1]) : (\varphi, [1, 1])\in IK$ and $Inc(\underline{IK}\cup\{\phi, 1\}) < 1\}\cup\{(\phi, [1, 1])\}$.*
2. *If there exists a formula in Φ with a certainty interval different from $[1, 1]$. Then: $IK_\phi = \{(\phi, [1, 1])\}\cup\{(\varphi, [\max(0, \underline{\alpha}_I), \overline{\alpha}_I]) : (\varphi, I)\in IK$, and $\underline{I} > Inc(\underline{IK}\cup\{(\phi, 1)\})\}\cup\{(\varphi, [0, \max(0, 2\alpha_I)]) : (\varphi, I)\in IK$, and $\underline{I} = Inc(\underline{IK}\cup\{(\phi, 1)\}) > 0\}$.*

Note that *Item 1* corresponds to the case where $secondbest(IK) = 0$.

Proof. In the same way we prove Proposition 4.11, here we aim to show that $\forall\omega \in \Omega$, $I\pi_{IK_\phi} = I\pi_{IK}(.|\phi)$.

1. *Proof of Item 1:* The fact that Φ admits a unique model and each formula in Φ has a certainty interval equal to [1,1] means that:

$$\forall\omega \in \Omega, I\pi_{IK}(\omega) = \begin{cases} [\underline{I\pi}(\omega), \overline{I\pi}(\omega)] & \text{if } \omega \text{ is model of } \Phi \\ [0,0] & \text{otherwise.} \end{cases}$$

   where $\underline{I\pi}(\omega) = 1 - Inc(\overline{IK})$ and $\overline{I\pi}(\omega) = 1 - Inc(\underline{IK})$.
   Let ω^* be the model of Φ. Then by Definition 4.5 we have

$$I\pi_{IK_\phi}(\omega^*) = [1, 1].$$

   Since Φ admits exactly one model, then each interpretation ω', which is different from ω, falsifies at least one formula from $IK_\phi = \{(\varphi, [1, 1]) : \varphi \in \Phi\}$. Hence, again by Definition 4.5 we have:

$$\forall\omega' \in \Omega, I\pi_{IK_\phi}(\omega') = [0, 0].$$

   Clearly, we have: $\forall\omega \in \Omega, I\pi_{IK}(\omega|\phi) = I\pi_{IK_\phi}(\omega)$.

2. *Proof of Item 2:* Note first when $Inc(\underline{IK} \cup \{(\phi, 1)\}) > 0$, then each interpretation falsifies at least one formula from IK_ϕ. Let ω^* be the model of Φ. If $Inc(\underline{IK} \cup \{(\phi, 1)\}) > 0$ then ω^* falsifies a formula from

$$\left\{(\varphi, \left[0, \max\left(0, 1 - \frac{1-\overline{I}}{secondbest(IK)}\right)\right] : (\varphi, I) \in IK, \text{and } \underline{I} = Inc(\underline{IK} \cup \{(\phi, 1)\}) > 0\right\}.$$

   Using Definition 4.5 we get: $\overline{I\pi}_{IK_\phi}(\omega^*) = 1$ and $\underline{I\pi}_{IK_\phi}(\omega^*) = 1 - \max\left(\left(1 - \frac{1-\overline{I}}{secondbest(I\pi_{IK})}\right), 0\right)$.
   When $Inc(\underline{IK} \cup \{(\phi, 1)\}) = 0$ then trivially $\overline{I\pi}_{IK_\phi}(\omega^*) = 1$ and $\underline{I\pi}_{IK_\phi}(\omega^*) = 1$.

$$\begin{aligned}
\underline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\left\{\max\left(0, 1 - \frac{1-\overline{I}}{secondbest(IK)}\right) : (\varphi, I) \in IK, \omega \nvDash \varphi \text{ and } \underline{I} = Inc(\underline{IK} \cup \{(\phi, 1)\})\right\}. \\
&= 1 - \max\left\{\max\left(0, 1 - \frac{1-\overline{I}}{secondbest(I\pi_{IK})}\right) : (\varphi, I) \in IK, \omega \nvDash \varphi \text{ and } \underline{I} = Inc(\underline{IK} \cup \{(\phi, 1)\})\right\}. \\
&= \min\left\{1 - \max\left(0, 1 - \frac{1-\overline{I}}{secondbest(I\pi_{IK})}\right) : (\varphi, I) \in IK,\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\right\}. \\
&= \min\left\{\min\left(1, \frac{1-\overline{I}}{secondbest(I\pi_{IK})}\right) : (\varphi, I) \in IK,\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\right\}. \\
&= \min\left(1, \min\left\{\frac{1-\overline{I}}{secondbest(I\pi_{IK})} : (\varphi, I) \in IK,\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\right\}\right). \\
&= \min\left(1, \frac{\min\{1 - \overline{I} : (\varphi, I) \in IK,\ \omega \not\models \varphi \text{ and } \underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\}}{secondbest(I\pi_{IK})}\right). \\
&= \min\left(1, \frac{\min\{1 - \overline{I} : (\varphi, I) \in IK \text{ and } \omega \not\models \varphi\}}{secondbest(I\pi_{IK})}\right).
\end{aligned}$$

$$= \min\left(1, \frac{1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \omega \not\models \varphi\}}{secondbest(I\pi_{IK})}\right).$$

$$= \min\left(1, \frac{\underline{I\pi_{IK}}(\omega)}{secondbest(I\pi_{IK})}\right).$$

For counter-models ω' of Φ, the proof is exactly the same as the one presented in the previous proposition (Proposition 4.11).

□

Example 4.12 (First item of Proposition 4.12)**.** Let us consider Example 4.9 with the new evidence being $\phi = \neg a$. In this example, we consider $\Phi = \{a \vee \neg b\} \cup \{\neg a\}$, which admits a unique model $\omega^* = \neg a \neg b$ and each formula of Φ, namely $\psi = a \vee \neg b$, has a weight equal to $[1, 1]$. From Proposition 4.12, 1^{st} item, we have $IK_\phi = \{(\neg a, [1, 1]), (a \vee \neg b, [1, 1])\}$. The computation of $I\pi_{IK_\phi}$ according to Definition 4.5 is given in Table 4.11.

ω	$I\pi_{IK_\phi}(\omega)$
ab	$[0, 0]$
$a\neg b$	$[0, 0]$
$\neg ab$	$[0, 0]$
$\neg a \neg b$	$[1, 1]$

Table 4.11 – Interval-based possibility distribution induced by IK_ϕ of Example 4.12.

Example 4.13 (Second item of Proposition 4.12)**.**
Let $IK = \{(a \wedge b, [.4, .7]), (a \vee \neg b, [.6, .9])\}$ be an interval-based possibilistic base. The interval-based possibility distribution $I\pi_{IK}$ corresponding to IK according to Definition 4.5 is given in Table 4.6.

Let us consider the new evidence $\phi = \neg a$. From this example, $\Phi = \{a \vee \neg b\}$ and $\Phi \cup \{\phi\}$ has exactly one model. We face the case of Proposition 4.12, 2^{nd} item. Hence, $IK_\phi = \{(\neg a, [1, 1]), (a \wedge b, [0, .1/.4]), (a \vee \neg b, [0, .5/.6])\}$. Computing $I\pi_{IK_\phi}$ according to Definition 4.5, gives exactly the same distribution as the one of Example 4.2 when conditioned on $\phi = \neg a$ using Proposition 4.5.

Algorithm 1 summarizes the main steps for computing IK_ϕ. When formulas in IK are in a clausal form then computing the conditioning of an interval-based possibilistic base has the same complexity as the one of conditioning standard possibilistic knowledge bases (namely, when I's are singletons). Indeed, for standard possibilistic knowledge bases K the hardest task consists in computing $Inc(K)$ which can be achieved in time in $\mathcal{O}(\log_2(m).\text{SAT})$ where SAT is a satisfiability test of a set of propositional clauses and m is the number of different weights in K. For an interval-based knowledge base, the main (hard) tasks in computing IK_ϕ are:

— The computation of $Inc(\underline{IK} \cup \{(\phi,1)\})$ and $Inc(\overline{IK} \cup \{(\phi, 1)\})$. This is done in $\mathcal{O}(\log_2(m).\text{SAT})$ where SAT is a satisfiability test of a set of propositional clauses and m is the number of different weights in $\underline{IK}$ and $\overline{IK}$,

— The test whether the sub-bases A or B are consistent or not. This needs only one call to a SAT solver.

Algorithm 1 Syntactic counterpart of conditioning

Require: An interval-based logic base $I\!K$ and a new evidence ϕ
Ensure: A new interval-based possibilistic base $I\!K_\phi$ such that $\forall\omega \in \Omega$, $I\pi_{I\!K_\phi}(\omega) = I\pi_{I\!K}(\omega|\phi)$.
Let $A = \{\varphi\colon (\varphi, I) \in I\!K$ and $I{=}[1,1]\} \cup \{\phi\}$
Let $B = \{\varphi\colon (\varphi, I) \in I\!K$ and $\overline{I}{=}1\} \cup \{\phi\}$
if A is inconsistent **then**
  $I\!K_\phi = \emptyset$ (*Prop. 4.6*).
else if B is inconsistent **then**
  $I\!K_\phi = \{(\phi, [1,1]), (\neg\phi, [0,1])\}$ (*Prop. 4.6*).
else if $\{\varphi : (\varphi, I) \in I\!K\} \cup \{\neg\phi\}$ is inconsistent **then**
  $I\!K_\phi = I\!K \cup \{(\phi, [1,1])\}$ (*Prop. 4.7*).
else if $\Phi \cup \{\phi\}$ admits a unique model **then**
  if each formula φ in Φ has a certainty interval equal to $[1,1]$ in $I\!K_\phi$ **then**
    $I\!K_\phi = \{(\varphi, [1,1]) : (\varphi, [1,1]) \in I\!K$ and $Inc(\underline{I\!K}) < 1\} \cup \{(\phi, [1,1])\}$ (*Prop. 4.12*).
  else
    $I\!K_\phi = \{(\phi, [1,1])\} \cup \{(\varphi, [\max(0, \underline{\alpha_I}), \overline{\alpha_I}]) : (\varphi, I) \in I\!K$, and $\underline{I} > Inc(\underline{I\!K} \cup \{(\phi,1)\})\} \cup \{(\varphi, [0, max(0, 2\alpha_I)]) : (\varphi, I) \in I\!K$, and $\underline{I} = Inc(\underline{I\!K} \cup \{(\phi,1)\}) > 0\}$ (*Prop. 4.12*).
  end if
else
  $I\!K_\phi = \{(\phi, [1,1])\} \cup \{(\varphi, [\max(0, \underline{\alpha_I}), \overline{\alpha_I}]) : (\varphi, I) \in I\!K$, and $\underline{I} \geq Inc(\underline{I\!K} \cup \{(\phi,1)\})\}$ (*Prop. 4.11*).
end if

— The computation of $secondbest(I\pi) = 1 - \min\{\alpha_i : \alpha_i > Inc(\underline{I\!K} \cup \{(\phi,1)\})$ and A_{α_i} is a non-tautological formula$\}$ (see Proposition 4.10). This needs: i) the computation of $Inc(\underline{I\!K} \cup \{(\phi,1)\})$, done again in $\mathcal{O}(\log_2(m).\text{SAT})$, and ii) checking for the lowest α_i such that A_{α_i} is a non-tautological formula, which is done in linear time (w.r.t the number of clauses in $I\!K$).

— Lastly, checking whether $\Phi = \{\varphi : (\varphi, I) \in I\!K$, and $\underline{I} > Inc(\underline{I\!K} \cup \{(\phi,1)\})\} \cup \{\phi\}$ admits a unique model. This can be done using two calls to a SAT solver. Indeed, checking whether there exists a unique interpretation ω^* such that $\overline{I\pi}_{I\!K}(\omega^*) = \overline{I\Pi}_{I\!K}(\phi)$ comes down to checking whether the formula $\Phi \cup \{\phi\}$ has a unique model. If this formula is under the clausal form, then this problem is the one of Unique-SAT. This can be done by launching two calls to a SAT solver: the first call is applied to the formula Φ. When it returns a model ω (recall that $\Phi \cup \{\phi\}$ is consistent), then a second call to a SAT solver with the formula $\Phi \wedge \neg\omega$ is performed (where $\neg\omega$ is a clause composed of the disjunction of literals that are not true in ω). If a SAT solver declares that the extended formula has no model, then we conclude that there exists a unique interpretation ω^* such that $\overline{I\pi}_{I\!K}(\omega^*) = \overline{I\Pi}_{I\!K}(\phi)$. Otherwise the formula $\Phi \cup \{\phi\}$ has at least two models.

To summarize, the overall complexity of computing $I\!K_\phi$ is:

Proposition 4.13. *Computing $I\!K_\phi$ is $\mathcal{O}(\log_2(m).SAT)$ where SAT is a satisfiability test of a set propositional clauses and m is the number of different weights in $\overline{I\!K}$ and $\underline{I\!K}$.*

Proposition 5.2 shows that the syntactic computation of conditioning an interval-based possibilistic base has exactly the same computational complexity of computing product-based conditioning of standard possibilistic knowledge bases.

Proof. For an interval-based knowledge base, the main (hard) tasks in computing IK_ϕ using Algorithm 3:

- The test whether the sub-bases A or B are consistent or not. This needs only one call to a SAT solver.
- Checking whether $\Phi = \{\varphi\colon (\varphi, I) \in IK$, and $\underline{I} > Inc(\underline{IK} \cup \{(\phi, 1)\})\} \cup \{\phi\}$ admits a unique model. This can be done using two calls to a SAT solver. Indeed, checking whether there exists a unique interpretation ω^* such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$ comes down to checking whether the formula $\Phi \cup \{\phi\}$ has a unique model. If this formula is under the clausal form, then this problem is the one of Unique-SAT. This can be done by launching two calls to a SAT solver: the first call is applied to the formula Φ. When it returns a model ω (recall that $\Phi \cup \{\phi\}$ is consistent), then a second call to a SAT solver with the formula $\Phi \wedge \neg\omega$ is performed (where $\neg\omega$ is a clause composed of the disjunction of literals that are not true in ω). If a SAT solver declares that the extended formula has no model, then we conclude that there exists a unique interpretation ω^* such that $\overline{I\pi}_{IK}(\omega^*) = \overline{I\Pi}_{IK}(\phi)$. Otherwise the formula $\Phi \cup \{\phi\}$ has at least two models.
- The computation of $Inc(\underline{IK} \cup \{(\phi,1)\})$ and $Inc(\overline{IK} \cup \{(\phi, 1)\})$. This is done in $\mathcal{O}(\log_2(m)$.SAT) where SAT is a satisfiability test of a set of propositional clauses and m is the number of different weights in $\underline{IK}$ and $\overline{IK}$,
- Lastly, the computation of $secondbest(I\pi) = 1 - \min\{\alpha_i : \alpha_i > Inc(\underline{IK} \cup \{(\phi,1)\})$ and A_{α_i} is a non-tautological formula$\}$ (see Proposition 4.10). This needs: i) the computation of $Inc(\underline{IK} \cup \{(\phi,1)\})$, done again in $\mathcal{O}(\log_2(m)$.SAT), and ii) the checking for the lowest α_i such that A_{α_i} is a non-tautological formula, which is done in linear time (w.r.t the number of clauses in IK).

□

4.4 Concluding remarks

Interval-based possibilistic logic offers an expressive and a powerful framework for representing and reasoning with uncertain information. This setting was only specified for static situations and no form of conditioning has been proposed for updating the knowledge and the beliefs. In this chapter, we proposed a set of natural postulates that a conditioning operator should satisfy. We have thus defined a conditioning operator based on compatibles. We showed that conditioning can be handled in a natural and safe way and without extra computational cost. We showed that applying product-based conditioning on the set compatible possible distributions gives an interval-based possibility distribution. We provided the exact computations of lower and upper endpoints of intervals associated with each interpretation of the conditioned interval-based possibility distributions. Lastly, we provided a syntactic counterpart of the compatible-based conditioning that does not imply extra computational cost.

Chapter 5

Qualitative conditioning in an interval-based possibilistic setting

Chapter 4 addressed the issue of conditioning with product-based conditioning definition. This chapter deals with conditioning uncertain information in a qualitative or $\min$-based interval-valued possibilistic setting. The first important contribution concerns a set of three natural postulates for conditioning interval-based possibility distributions. We show that any interval-based conditioning satisfying these three postulates is necessarily based on the set of compatible standard possibility distributions. The second contribution consists in a proposal of efficient procedures to compute the lower and upper endpoints of the conditional interval-based possibility distribution while the third important contribution provides a syntactic counterpart of conditioning interval-based possibility distributions in case where these latter are compactly encoded in the form of possibilistic knowledge bases.

Contents

5.1 Qualitative conditioning in possibility theory: a reminder

Recall that conditioning in the qualitative possibilistic setting is done using the following equation:

$$\pi(\omega_i|_m\phi) = \begin{cases} 1 & \text{if } \pi(\omega_i) = \Pi(\phi) \text{ and } \omega_i \in \phi; \\ \pi(\omega_i) & \text{if } \pi(\omega_i) < \Pi(\phi) \text{ and } \omega_i \in \phi; \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Before presenting our interval-based extension to the $\min$-based possibilistic conditioning, let us first focus on some natural properties that an interval-based conditioning should satisfy in a possibilistic setting.

5.2 Three natural requirements for the interval-based conditioning

Before presenting the three natural properties for $\min$-based conditioning, let us first show that $\min$-based conditioning does not satisfy **IC1**.

5.2.1 $\min$-based conditioning does not satisfy IC1

From the previous chapter, we proposed suitable properties that a conditioning operator should satisfy. The first important issue with compatible-based conditioning of Definition 4.6 using $\min$-based rule is that conditioning an interval-based distribution $I\pi$ with an evidence ϕ does not guarantee to result in an interval-based distribution, namely the first property (**IC1**) (*cf.* Subsection 4.2.1) is not always satisfied.

Indeed, let $|_m$ be the conditioning operator given by Equation (5.1). Then, there exists an interval-based possibility distribution (see Example 5.1), a propositional formula ϕ, and an interpretation ω such that $I\pi(\omega|_m\phi)$ is not an interval.

Counter-example 5.1. Let $I\pi$ be the normalized interval-based distribution of Table 5.1. Let $\phi = a$ be the new evidence. The compatible-based conditioned distribution $I\pi(.|_m\phi)$ is obtained by conditioning with $|_m$ of Equation (5.1) on the set of all compatible possibility distributions.

$\omega \in \Omega$	$I\pi(\omega)$
ab	$[.7, .9]$
$a\bar{b}$	$[.4, .7]$
$\bar{a}b$	$[.8, 1]$
$\bar{a}\bar{b}$	$[.4, .7]$

$\omega \in \Omega$	$I\pi(\omega|_m\phi)$
ab	$[1, 1]$
$a\bar{b}$	$[.4, .7]\cup\{1\}$
$\bar{a}b$	$[0, 0]$
$\bar{a}\bar{b}$	$[0, 0]$

Table 5.1 – Counter-example for Observation 1.

From Table 5.1, $I\pi(a\bar{b}|_m\phi)$ is not an interval. Indeed, one can check that for every compatible distribution π of $I\pi$, such that $\pi(a\bar{b}) \in [.4, .7[$ we have $\pi(a\bar{b}|_m\phi) \in [.4, .7[$ (since $\pi(ab) \geq .7$). Now, for compatible distributions where $\pi(a\bar{b}) = .7$ we have either $\pi(a\bar{b}|_m\phi) = .7$ (if $\pi(ab) > .7$) or $\pi(a\bar{b}|_m\phi) = 1$ (if $\pi(ab) = .7$). Hence, $\pi(a\bar{b}|_m\phi) = [.4, .7]\cup\{1\}$ which is not an interval.

This result motivates us to propose a new set of properties dedicated to qualitative conditioning.

The first natural requirement concerns the *degenerate* case, namely when each interval $I\pi(\omega)$ contains exactly one single degree $\pi(\omega)$. The result of the new conditioning procedure should coincide with the result $\pi(.|_m\phi)$ of the original conditioning procedure (Equation (1.29)). For each possibility distribution π, by $[\pi, \pi]$ we denote its interval-valued representation, *i.e.* an interval-valued possibility distribution for which, for every $\omega \in \Omega$, we have $I\pi(\omega) = [\pi(\omega),\pi(\omega)]$. In these terms, the above requirement takes the following form:

P1. For all π, $\phi \subseteq \Omega$ and $\omega \in \Omega$, $([\pi, \pi])(\omega|\phi) = [\pi(\omega|_m\phi), \pi(\omega|_m\phi)]$.
In other terms, let π be any possibility distribution and $I\pi$ such that $\forall\omega$, $I\pi(\omega) = [\pi(\omega), \pi(\omega)]$. Then $\forall\phi$, $I\pi(\omega|\phi) = [\pi(\omega|_m\phi), \pi(\omega|_m\phi)]$.

The second requirement is related to the fact that we do not know the exact values $\pi(\omega)$ since we only have partial information about them. In principle, if we can get some additional information about these values, then this would lead, in general, to narrower intervals (indeed, the width of an interval captures the ignorance regarding the exact value of $\pi(\omega)$). Let us define the concepts of specificity between interval-based possibility distribution:

Definition 5.1. Let $I\pi$ and $I\pi'$ be two interval-based possibility distributions. Then $I\pi$ is said to be more specific than $I\pi'$, denoted $I\pi \subseteq I\pi'$, if $I\pi(\omega) \subseteq I\pi'(\omega)$ holds for all $\omega \in \Omega$

It is reasonable to require that if we have new information about the original values $\pi(\omega)$, this should help us also to narrow down the corresponding values of conditional distributions:

P2. If $I\pi$ is more specific than $I\pi'$ (namely, $I\pi \subseteq I\pi'$) then $I\pi(.|\phi)$ is more specific than $I\pi'(.|\phi)$ (namely, $I\pi(.|\phi) \subseteq I\pi'(.|\phi)$).

It is obvious that postulates **P1** and **P2** are not sufficient to fully characterize the new extension. For example, we can take $([\pi, \pi])(.|\phi) = [\pi(.|_m\phi), \pi(.|_m\phi)]$ for degenerate interval-valued possibility distributions and $I\pi(\omega|\phi) = [0, 1]$ for all other $I\pi$. To avoid such extensions, it is reasonable to impose the following minimality condition:

P3. There exist no operation $I\pi(.|_1\phi)$ that satisfies both properties **P1–P2** and for which:

— $I\pi(\omega|_1\phi) \subseteq I\pi(\omega|\phi)$ for all $I\pi$, ω, and ϕ,

— $I\pi(\omega|_1\phi) \neq I\pi(\omega|\phi)$ for some $I\pi$, ω, and ϕ.

The following defines the concept of interval closure of a set of uncertainty degrees

Definition 5.2 (Interval closure)**.** Let A be a set of degrees between 0 and 1. We define the interval closure of A, denoted by $\mathrm{IntCl}(A)$, as the smallest (narrowest) interval that contains all the elements of A.

Example 5.1. Assume that A is a set defined as follows: $A = [.8, .9] \cup \{1\}$ then the closure of A is $\mathrm{IntCl}(A) = [.8, 1]$. Clearly, the interval $[.8, 1]$ is the narrowest sub-interval of $[0, 1]$ containing all the values of A.

The following theorem provides our first main result where we show that there is only one interval-based conditioning satisfying **P1-P3** and where the interval conditional possibility degree $I\pi(\omega|\phi)$ is defined as the interval closure of the set of all $\pi(.|_m\phi)$, where π is compatible with $I\pi$.

Theorem 5.1. *There exists exactly one interval-based conditioning, denoted by $I\pi(.|_m\phi)$, that satisfies the properties* **P1–P3***, and which is defined by:* $\forall\omega \in \Omega$,

$$I\pi(\omega|_m\phi) = \mathrm{IntCl}(\{\pi(\omega|_m\phi) : \pi \in \mathcal{C}(I\pi)\}) \tag{5.2}$$

where IntCl *is the interval closure given in Definition 5.2.*

Proof.
$1°$. We need to prove:

— that this closure $I\pi(.|_m\phi)$ satisfies the properties **P1–P3**, and

— that every operation $I\pi(.|\phi)$ that satisfies the properties **P1–P3** coincides with the interval closure of $I\pi(.|_m\phi)$.

$2°$. One can easily check that $I\pi(.|_m\phi)$ satisfies the properties **P1–P2**.

$3°$. Let us now prove that if an operation $I\pi(.|\phi)$ satisfies the properties **P1–P2**, then for every $I\pi$ and ϕ, we have $I\pi(.|_m\phi) \subseteq I\pi(.|\phi)$.

Then, for every distribution $\pi \in \mathcal{C}(I\pi)$, we have $([\pi, \pi]) \subseteq I\pi$ and thus, due to the postulate **P2**, we have $([\pi, \pi])(.|\phi) \subseteq I\pi(.|\phi)$. By the property **P1**, we have $([\pi, \pi])(\omega|\phi) = [\pi(\omega|\phi), \pi(\omega|\phi)]$. Thus, the above inclusion means that $\pi(.|\phi) \in I\pi(.|\phi)$.

The interval $I\pi(\omega|\phi)$ therefore contains all the values $\pi(\omega|\phi)$ corresponding to all possible $\pi \in \mathcal{C}(I\pi)$:

$$\{\pi(\omega|\phi) : \pi \in \mathcal{C}(I\pi)\} \subseteq I\pi(\omega|\phi). \tag{5.3}$$

Since the set $I\pi(\omega|\phi)$ is an interval, it therefore contains, with the set $\{\pi(\omega|\phi) : \pi \in \mathcal{C}(I\pi)\}$, its interval closure, *i.e.* the set $I\pi(\omega|_m\phi)$. Thus, we conclude that $I\pi(\omega|_m\phi) \subseteq I\pi(\omega|\phi)$ for all ω.

The statement is proven.

4°. We can now prove that $I\pi(.|_m\phi)$ also satisfies the property **P3**.

Indeed, if there is some other operation $|_1$ that satisfies **P1** and **P2**, and for which $I\pi(\omega|_1\phi) \subseteq I\pi(\omega|_m\phi)$ for all ω, then, since we have already proven the opposite enclosure in Part 3 of this proof, we conclude that $I\pi(\omega|_1\phi) = I\pi(\omega|_m\phi)$ for all ω, so indeed no narrower conditioning operation is possible.

5°. To complete the proof, let us show that if some $I\pi(.|\phi)$ satisfies the properties **P1**–**P3**, then it coincides with $I\pi(.|_m\phi)$.

Indeed, by Part 3 of this proof, we have $I\pi(\omega|_m\phi) \subseteq I\pi(\omega|\phi)$ for all ω. If we had $I\pi(\omega|_m\phi) \neq I\pi(\omega|\phi)$ for some ω and ϕ, this would contradict the minimality property **P3**. Thus, indeed, $I\pi(.|_m\phi) = I\pi(.|\phi)$. Uniqueness is proven, and so is for the proposition. □

We can now go one step beyond Theorem 5.1 and provide the exact bounds of intervals associated with $I\pi(.|_m\phi)$.

5.3 Computing lower and upper endpoints of conditional interval-based possibility distributions

The aim of this section is to compute the lower and upper endpoints of the conditional interval-based possibility distribution.

Proposition 5.1. *Let $I\pi$ be an interval-based distribution. Then the interval-based conditional distribution, satisfying **P1–P3**, is described by $I\pi(\omega|_m\phi) = [\underline{I\pi}(\omega|_m\phi), \overline{I\pi}(\omega|_m\phi)]$, such that $\forall\omega \in \Omega$:*

$$\underline{I\pi}(\omega|_m\phi) = \begin{cases} 0 & \text{if } \omega \notin \phi \\ 1 & \text{if } \forall\omega' \neq \omega,\ \omega' \in \phi \text{ and } \underline{I\pi}(\omega) \geq \overline{I\pi}(\omega') \\ \underline{I\pi}(\omega) & \text{otherwise} \end{cases} \tag{5.4}$$

and

$$\overline{I\pi}(\omega|_m\phi) = \begin{cases} 0 & \text{if } \omega \notin \phi \\ 1 & \text{if } \overline{I\pi}(\omega) \geq \underline{I\Pi}(\phi), \\ \overline{I\pi}(\omega) & \text{otherwise} \end{cases} \tag{5.5}$$

Let us briefly comment Proposition 5.1. Let $\omega \in \Omega$ be an interpretation. First, for $\omega \notin \phi$, whatever the considered compatible possibility distribution π, we have $\pi(\omega|_m\phi) = 0$. Hence, $I\pi(\omega|_m\phi) = [0, 0]$. Assume now that $\omega \in \phi$ and $\forall\omega' \in \phi$, $\underline{I\pi}(\omega) \geq \overline{I\pi}(\omega')$. This means that whatever is the considered compatible possibility distribution π, we have $\pi(\omega) \geq \max\{\pi(\omega)\colon \omega \in \phi\} = \Pi(\phi)$. Hence, $\pi(\omega|_m\phi) = 1$ and $I\pi(\omega|_m\phi) = [1, 1]$. Now, the last case for determining lower endpoint concerns the case where $\exists\omega' \in \phi$ such that $\underline{I\pi}(\omega) < \overline{I\pi}(\omega')$. This means that there exists a compatible possibility distribution π such that $\pi(\omega) = \underline{I\pi}(\omega) < \Pi(\phi)$, hence $\pi(\omega|_m\phi) = \underline{I\pi}(\omega)$ which is the smallest possible value. Similar reasoning goes for upper endpoints.

Example 5.2. Let $I\pi$ of Table 5.2 (left side table) be the interval-based possibility distribution that we want to condition with the new piece of information $\phi = \bar{c}$. min-based conditional distribution $I\pi(.|_m\phi)$ given in Table 5.2 (right side table) is obtained using either Proposition 5.1 or Theorem 5.1. For instance, for $\omega = ab\bar{c}$, whatever the considered compatible possibility distribution π, we have $\pi(ab\bar{c}|\phi)$ between .1 and 1. Thus, the interval closure of $I\pi(ab\bar{c}|_m\phi) = [.1, 1]$.

$\omega \in \Omega$	$I\pi(\omega)$	$\omega \in \Omega$	$I\pi(\omega\|\phi)$
abc	[1, 1]	abc	[0, 0]
$a\bar{b}c$	[.4,.6]	$a\bar{b}c$	[0, 0]
$\bar{a}bc$	[.3,.6]	$\bar{a}bc$	[0, 0]
$\bar{a}\bar{b}c$	[.3,.6]	$\bar{a}\bar{b}c$	[0, 0]
$ab\bar{c}$	[.1,.7]	$ab\bar{c}$	[.1, 1]
$a\bar{b}\bar{c}$	[.4,.6]	$a\bar{b}\bar{c}$	[.4, 1]
$\bar{a}b\bar{c}$	[.1,.6]	$\bar{a}b\bar{c}$	[.1, 1]
$\bar{a}\bar{b}\bar{c}$	[.3,.6]	$\bar{a}\bar{b}\bar{c}$	[.3, 1]

Table 5.2 – Interval-based distribution $I\pi$ and its conditioned distribution $I\pi(.|\phi)$

5.4 Syntactic computations of interval-based conditioning

We now provide the syntactic counterpart of the interval-based conditioning presented above. Given an interval-based possibilistic knowledge base IK and a new evidence ϕ, our aim is to compute the conditional base IK_ϕ corresponding to conditioning the information encoded in IK with ϕ. As illustrated in Figure 5.1, the aim of this subsection is therefore to propose the syntactic characterization of conditioning such that:

$$\forall\omega\in\Omega,\ I\pi_{IK}(\omega|_m\phi) = I\pi_{IK_\phi}(\omega), \tag{5.6}$$

where $I\pi_{IK_\phi}$ is the interval-based distribution associated with IK_ϕ, and $I\pi_{IK}(.|_m\phi)$ is the result of conditioning $I\pi_{IK}$ using the conditioning operator presented in the previous subsection (Proposition 5.1 or Theorem 5.1), and $[\phi]$ is the set of models of ϕ.

$$\begin{tikzcd}
IK \arrow[rr, "\text{Definition 4.5}"] \arrow[d, "\text{Aim of this section}"'] & & I\pi_{IK} \arrow[d, "\text{Theorem 5.1}"] \\
IK_\phi \arrow[r, "\text{Definition 4.5}"] & I\pi_{IK_\phi} \arrow[r, equal, "?"] & I\pi_{IK}(.|\phi)
\end{tikzcd}$$

Figure 5.1 – Conditioning interval-valued possibilistic information at the semantic and syntactic levels.

We first need to introduce some notations.

— $\alpha = Inc(\overline{IK} \cup \{(\phi, 1)\})$ and $\beta = Inc(\underline{IK} \cup \{(\phi, 1)\})$, Intuitively, α and β compute inconsistency degree intervals resulting from assuming that ϕ is fully true. This offers a characterization of $\overline{I\Pi}(\phi) = 1 - \beta$ and $\underline{I\Pi}(\phi) = 1 - \alpha$.

— Let ω^* be a model of $\{\psi : (\psi, I) \in IK \text{ and } \underline{I} > \beta\} \cup \{\phi\}$. Let $IK_{\neg\omega^*} = IK \cup \{(\neg\omega^*, [1,1])\}$ be a base obtained by adding the negation of ω^*, then we compute $\gamma = Inc(\underline{IK}_{\neg\omega^*} \cup \{(\phi, 1)\})$. Models ω of $\{\psi : (\psi, I) \in IK \text{ and } \underline{I} > \beta\}$ are exactly those having $\overline{I\pi}(\omega) = \overline{I\Pi}(\phi)$. γ computes the second best value of models of ϕ (since a model ω^* is excluded from IK) which is very useful for characterizing $\underline{I\pi}(\omega|_m\phi)$.

With the help of these notations, we are now ready to present the third contribution of this chapter.

Theorem 5.2. *Let IK be an interval-based knowledge base. Let $I\pi_{IK}$ be its associated possibility distribution. Let $IK_\phi = \{(\phi, [1,1])\} \cup \{(\varphi, I) : (\varphi, I) \in IK, \text{ and } \underline{I} > \alpha\} \cup \{(\varphi, [0, \overline{I}]) : (\varphi, I) \in IK, \text{ and } \underline{I} < \alpha \text{ and } \overline{I} > \gamma\}$. Then:*

$$\forall\omega, I\pi_{IK}(\omega|_m\phi) = I\pi_{IK_\phi}(\omega);$$

where $I\pi_{IK}(.|_m\phi)$ is the result of applying min*-based interval conditioning on $I\pi_{IK}$ (see Proposition 5.1 and Theorem 5.1), and $I\pi_{IK_\phi}$ is the interval-based distribution associated with IK_ϕ using the Definition 4.5.*

The knowledge base IK_ϕ resulting from conditioning IK with ϕ is composed of three parts:

— The first consists in adding ϕ as a fully certain information, $\{(\phi, [1,1])\}$. From Definition 4.5, all worlds that are outside ϕ (not satisfying ϕ) are excluded. This is in accordance with Proposition 5.1.

— The second part, $\{(\varphi, I) : (\varphi, I) \in IK, \text{ and } \underline{I} > \alpha\}$, contains a subbase of IK where the intervals are unchanged. This encodes the third item of definition of $\underline{I\pi}(\omega)$ and $\overline{I\pi}(\omega)$ in Proposition 5.1 (recall that $1 - \alpha = \underline{I\Pi}(\phi)$).

— The last part encodes exactly the situation where some possibility degrees (in Proposition 5.1) are shifted up to 1. This is reflected in possibilistic knowledge bases by shifting down some certainty degree to 0.

Proof. In order now to prove the theorem we have to show that $\forall\omega \in \Omega$, $I\pi_{IK_\phi} = I\pi_{IK}(.|_m[\phi])$. First note that if an interpretation ω is not a model of ϕ, then by definition we have:

$$I\pi_{IK_\phi}(\omega) = I\pi_{IK}(\omega|_m[\phi]) = [0, 0]$$

This is explained by the presence of $(\phi, [1,1])$ in IK_ϕ.
Now, for $\omega \models \phi$, we have two distinct cases:

— The case where ω falsifies a formula from: $\{(\varphi, I) : (\varphi, I) \in IK, \text{ and } \underline{I} > \alpha\}$ then:

$$\begin{aligned}
\underline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > \alpha\}.\\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > Inc(\overline{IK} \cup \{(\phi, 1)\})\}\\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > 1 - \underline{I\Pi}([\phi])\}\\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } 1 - \underline{I} < \underline{I\Pi}([\phi])\}\\
&= \underline{I\pi}(\omega) \quad \text{ if } \overline{I\pi}(\omega) < \underline{I\Pi}([\phi])\\
&= \underline{I\pi}_{IK}(\omega|_m[\phi]).
\end{aligned}$$

$$\begin{aligned}
\overline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\{\underline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > \alpha\}.\\
&= 1 - \max\{\underline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > Inc(\overline{IK} \cup \{(\phi, 1)\})\}\\
&= 1 - \max\{\underline{I} : (\varphi, I) \in IK \text{ and } \underline{I} > 1 - \underline{I\Pi}([\phi])\}\\
&= 1 - \max\{\underline{I} : (\varphi, I) \in IK \text{ and } 1 - \underline{I} < \underline{I\Pi}([\phi])\}\\
&= \overline{I\pi}(\omega) \quad \text{ if } \overline{I\pi}(\omega) < \underline{I\Pi}([\phi])\\
&= \overline{I\pi}_{IK}(\omega|_m[\phi]).
\end{aligned}$$

— The case where ω falsifies a formula from: $\{(\varphi, [0, \overline{I}]) : (\varphi, I) \in IK, \text{ and } \underline{I} \leq \alpha \text{ and } \overline{I} > \gamma\}$ then:

As γ computes the second best value of models of ϕ, for this proof, we use $secondbest(IK)$ to determine γ and $secondbest(IK) = 1 - secondbest(I\pi_{IK})$.

$$\begin{aligned}
\underline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} \leq \alpha \text{ and } \overline{I} > \gamma\} \\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} \leq Inc(\overline{IK} \cup \{(\phi, 1)\}) \text{ and } \overline{I} > secondbest(IK)\} \\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } \underline{I} \leq 1 - \underline{I\Pi}([\phi]) \text{ and } \overline{I} > 1 - secondbest(I\pi_{IK})\} \\
&= 1 - \max\{\overline{I} : (\varphi, I) \in IK \text{ and } 1 - \underline{I} \geq \underline{I\Pi}([\phi]) \text{ and } 1 - \overline{I} < secondbest(I\pi_{IK})\} \\
&= \underline{I\pi}(\omega) \quad \text{ if } \overline{I\pi}(\omega) \geq \underline{I\Pi}([\phi]) \text{ and } \underline{I\pi}(\omega) < secondbest(I\pi_{IK}) \\
&= \underline{I\pi}_{IK}(\omega|_m[\phi]).
\end{aligned}$$

$$\begin{aligned}
\overline{I\pi}_{IK_\phi}(\omega) &= 1 - \max\{0 : (\varphi, I) \in IK \text{ and } \underline{I} \leq \alpha \text{ and } \overline{I} > \gamma\}. \\
&= 1 - \max\{0 : (\varphi, I) \in IK \text{ and } \underline{I} \leq Inc(\overline{IK} \cup \{(\phi, 1)\}) \text{ and } \overline{I} > secondbest(IK)\} \\
&= 1 - \max\{0 : (\varphi, I) \in IK \text{ and } \underline{I} \leq 1 - \underline{I\Pi}([\phi]) \text{ and } \overline{I} > 1 - secondbest(I\pi_{IK})\} \\
&= 1 - \max\{0 : (\varphi, I) \in IK \text{ and } 1 - \underline{I} \geq \underline{I\Pi}([\phi]) \text{ and } 1 - \overline{I} > secondbest(I\pi_{IK})\} \\
&= 1 \quad \text{ if } \underline{I\pi}(\omega) \geq \underline{I\Pi}([\phi]) \text{ and } \underline{I\pi}(\omega) < secondbest(I\pi_{IK}) \\
&= \overline{I\pi}_{IK}(\omega|_m[\phi]).
\end{aligned}$$

□

Let us see an example to illustrate Theorem 5.2.

Example 5.3. Let IK be an interval-based possibilistic knowledge base such that

$$IK = \{(a \wedge b, [.4, .6]), (a, [0, .7]), (c \vee \neg b, [.3, .9])\}.$$

The associated interval-based possibility distribution π_{IK} (using Definition 4.5) is the same as the one given in Table 5.2. Let $\phi = \neg c$ (and $\phi = [\phi]$ the set of models of $\neg c$) be the new evidence. For the computation of IK_ϕ, let us first compute the values of α, β and γ. Then, we have: $\alpha = Inc(\{(a \wedge b, .6), (a, .7), (c \vee \neg b, .9), (\neg c, 1)\}) = .6$, $\beta = Inc(\{(a \wedge b, .4), (a, 0), (c \vee \neg b, .3), (\neg c, 1)\}) = .3$ and $\gamma = .4$.

Hence, according to Theorem 5.2, the result of conditioning IK by ϕ is given by: $IK_\phi = \{(a \wedge b, [0, .6]), (a, [0, .7]), (c \vee \neg b, [0, .9]), (\neg c, [1, 1])\}$. And if we compare with Example 5.2, where the distribution $I\pi(.|_m\phi)$ is conditioned according to Proposition 5.1 then the associated interval-based distribution to IK_ϕ is exactly the same. Hence, Theorem 5.2 indeed provides a compact encoding of the conditioning procedure.

The following proposition gives the computational complexity of conditioning an interval-based possibilistic knowledge base IK according to Theorem 5.2.

Proposition 5.2. *Let IK be an interval-based possibilistic knowledge base and ϕ be the new evidence. Let IK_ϕ be interval-based possibilistic knowledge base computed according to Theorem 5.2. Then IK_ϕ have the same size as IK and computing IK_ϕ is in $O(\log_2(m).SAT)$ where SAT is a satisfiability test of a set propositional clauses and m is the number of different weights in $\overline{IK}$ and $\underline{IK}$.*

Clearly, once the parameters α, β, γ are computed, computing IK_ϕ from $\{IK, \phi, \alpha, \beta, \gamma\}$ is straightforward and it is done in linear time. Indeed, computing α, β, γ mainly comes down to compute the inconsistency degrees of $\overline{IK}$ and $\underline{IK}$. This needs $\log_2(m)$ calls to a SAT solver exactly as in standard possibilistic logic [Lan00]. Hence, the syntactic counterpart of conditioning an interval-based possibilistic base has exactly the same computational complexity as computing the min-based conditioning of a standard possibilistic base.

5.5 From interval-based possibilistic networks to interval-based possibilistic knowledge bases

In Chapter 2, we have presented how to translate a possibilistic network into a possibility knowledge base. In this section, we adapt the same transformation to the interval-based possibility setting. Especially, this is interesting in order to exploit the results of conditioning an interval-based possibilistic knowledge base. Indeed, the previous section showed that using compatible-based conditioning over an interval-based possibilistic knowledge base is done without inducing extra computational cost.

5.5.1 Interval-based possibilistic network

With regards to interval-based possibilistic networks [BLT14a, BLT14b], it is seen as an extension of possibilistic networks. And as for credal networks, interval-based possibilistic networks can be seen as a family of possibilistic networks. More precisely, interval-based possibilistic networks allow to compactly encode families of joint possibility distributions. Intervals offer more flexibility to represent and to handle incomparable events. Thus, in the following we define an interval-based possibilistic network and a compatible possibilistic network.

Definition 5.3 (Interval-based possibilistic network)**.** An *Interval-based possibilistic network* $IPN = (G, \Theta_{I\pi})$ compactly encodes an interval-based possibility distribution and is composed of

— a *graphical component*: a DAG describing independence relationships between variables in V

— a *numerical component*: a set of local interval-based possibility distributions $\Theta_{I\pi}$

Note that in case where all the parameters of the network are singletons (pointwise-based possibilities), then the network is a standard network.

Definition 5.4 (Compatible possibilistic network)**.** A possibilistic network $\mathcal{PN}$ is compatible with an interval-based possibilistic network IPN if it shares the same graphical structure and for each local distribution:

$$\forall x_i \in D_{X_i|par(X_i)},\ \pi(x_i|par(x_i)) \in I\pi(x_i|par(x_i)) \tag{5.7}$$

In [BLT14a], the authors proposed two different semantics of coherent interval-based possibilistic network, one based on compatible models and one based on extending the chain rule. We present the one based on compatible models, more details can be found in [BLT14a].

Semantics based on compatible models

Coherent interval-based possibilistic network IPN can be semantically seen as a family of compatible possibilistic (pointwise) networks. Each of these compatible networks therefore encodes a joint possibility distribution that we call c-model (for compatible model).

Definition 5.5 (Compatible models)**.** A possibility distribution π is a compatible model (c-model) for an interval-based possibilistic network $IPN = (G, \Theta_{IPN})$ if there exist a standard network $\mathcal{PN} = (G, \Theta_\pi)$ compatible with IPN such that:

$$\forall x_1, ..., x_n \in \Omega,\ \pi(x_1...x_n) = \otimes_{i=1}^{n}(I\pi(x_i|par(x_i))) \tag{5.8}$$

where $\otimes$ denotes either the product or the $\min$ operator given the scale we are using.

Definition 5.5 formally states that the possibility distribution π are c-models if they are associated with a compatible possibilistic network $\mathcal{PN}$ using the $\min$-based chain rule of Equation (2.6) or the product-based chain rule of Equation (2.5). The set of all c-models is denoted by $\mathcal{C}$(c-models)

5.5.2 Qualitative encoding of an interval-based possibilistic network

In [BDGP02], Benferhat et al. have investigated the question of transforming possibilistic logic bases into possibilistic causal networks and conversely. In their paper, the authors have provided a way of encoding a possibilistic network into a possibilistic knowledge base in both possibilistic scale (quantitative and qualitative). And as it has been done in the standard setting, we use, for sake of simplicity, a set of triples to represent an interval-based possibilistic network. Given the set of variables, derived from the network IPN, $V = \{X_1, .., X_n\}$, then $IPN_{A_i} = \{(x_i, par(x_i), \iota_i) :$ where $\iota_i = I\pi(x_i|par(x_i)) \neq [1,1]$ is an element of the graph $\}$ where x_i is an instance of the variable X_i and $par(x_i)$ is an element of the Cartesian product of the domain of the variable's parents of X_i.

Therefore, we can define the set of weighted formulas associated to each triple by the following definition.

Definition 5.6. For a variable X_i, given a set of triple $IPN_{X_i} = \{(x_i, par(x_i), \iota_i) : x_i \in D_{X_i}, par(x_i) \in D_{par(X_i)}\}$ associated to X_i, the knowledge base IK_{X_i} is composed with the weighted formulas:

$$IK_{X_i} = \{(\neg x_i \vee \neg par(x_i), 1 \ominus \iota_i) : (x_i, par(x_i), \iota_i) \in IPN_{X_i}\} \tag{5.9}$$

Note that the operator $\ominus$ is the reverse of an interval, defined in [BHLR11] and recalled in 4.1 by:

$$1 \ominus \iota = [1 - \overline{\iota}, 1 - \underline{\iota}] \tag{5.10}$$

From this set of formulas, we can easily retrieve the associated distribution:

$$I\pi_{x_i, par(x_i)}(\omega) = \begin{cases} [1,1] & \text{if } \omega \models \neg x_i \vee \neg par(x_i), \\ \iota_i & \text{otherwise.} \end{cases} \tag{5.11}$$

This process is done step by step. The first one is building for each variable of the interval-based possibilistic network IPN, the associate knowledge base using Definition 5.6. Then we need to combine one by one each knowledge base. To do so, we have to define a combination operator $\mathcal{C}^m$. In the qualitative possibilistic setting, the process is easy and arises from the min-based chain rule. Indeed, the value of the local interval-based possibility distribution induced by the triple $(x_i, par(x_i), \iota)$ coincides with the (local) conditional interval-based possibility value $I\pi(x_i|par(x_i))$. Thus, the combination operation in the qualitative setting is described in the following definition.

Definition 5.7. Let IK_1 and IK_2 be two interval-based possibilistic knowledge bases associated with two different variables. Let $I\pi_1$ and $I\pi_2$ be the associated joint interval-based possibility distributions of IK_1 and IK_2 respectively. Let $\mathcal{C}^m$ be the combination operator, given by the min operator, of $I\pi_1$ and $I\pi_2$. Then the resulting interval-based possibilistic knowledge base is given by:

$$\mathcal{C}^m(IK_1, IK_2) = IK_1 \cup IK_2 \tag{5.12}$$

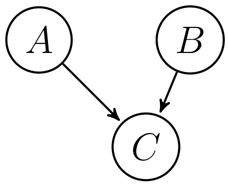
In the following example, we illustrate all of the notions introduced so far. Given an interval-based possibilistic network that we transpose in the form of triples, we give the associate weighted formulas and the resulting interval-based possibilistic knowledge base.

Example 5.4. Figure 5.2 represents an interval-based possibilistic network over a set of 3 boolean variables $V = \{A, B, C\}$, with their domains respectively $D_A = \{a_1, a_2\}$, $D_B = \{b_1, b_2\}$ and $D_C = \{c_1, c_2\}$.

The set of triples associated to IPN is given by:

$$\{(a_1, \emptyset, [.2, 1]), (a_2, \emptyset, [.3, .6])\} \cup \{(b_1, \emptyset, [.3, .8]), (b_2, \emptyset, [.5, 1])\} \cup$$

A	$I\pi(A)$
a_1	$[.2, 1]$
a_2	$[.3, .6]$



B	$I\pi(B)$
b_1	$[.3, .8]$
b_2	$[.5, 1]$

C	A	B	$I\pi(C \mid AB)$
c_1	a_1	b_1	$[.3, .7]$
c_1	a_1	b_2	$[.6, .8]$
c_1	a_2	b_1	$[.4, .5]$
c_1	a_2	b_2	$[.1, .2]$
c_2	a_1	b_1	$[.1, 1]$
c_2	a_1	b_2	$[1, 1]$
c_2	a_2	b_1	$[.6, 1]$
c_2	a_2	b_2	$[.5, 1]$

Figure 5.2 – Example of an interval-based possibilistic network

$A\ B\ C$	$I\pi_{IPN}(ABC)$
$a_1\ b_1\ c_1$	$[.2, .7]$
$a_1\ b_1\ c_2$	$[.1, .8]$
$a_1\ b_2\ c_1$	$[.2, .8]$
$a_1\ b_2\ c_2$	$[.2, 1]$
$a_2\ b_1\ c_1$	$[.3, .5]$
$a_2\ b_1\ c_2$	$[.3, .6]$
$a_2\ b_2\ c_1$	$[.1, .2]$
$a_2\ b_2\ c_2$	$[.3, .6]$

$A\ B\ C$	$I\pi_{IK}(ABC)$
$a_1\ b_1\ c_1$	$[.2, .7]$
$a_1\ b_1\ c_2$	$[.1, .8]$
$a_1\ b_2\ c_1$	$[.2, .8]$
$a_1\ b_2\ c_2$	$[.2, 1]$
$a_2\ b_1\ c_1$	$[.3, .5]$
$a_2\ b_1\ c_2$	$[.3, .6]$
$a_2\ b_2\ c_1$	$[.1, .2]$
$a_2\ b_2\ c_2$	$[.3, .6]$

Table 5.3 – Joint distribution induced by the network IPN and joint distribution induced by the knowledge base IK_{ABC}.

$\{(c_1, a_1b_1, [.3, .7]), (c_1, a_1b_2, [.6, .8]), (c_1, a_2b_1, [.4, .5]), (c_1, a_2b_2, [.1, .2]), (c_2, a_1b_1, [.1, 1]),$
$(c_2, a_2b_1, [.6, 1]), (c_2, a_2b_2, [.5, 1])\}$

For each nodes of the network, we compute the interval-based possibilistic knowledge base associated:

— $IK_A = \{(a_2, [0, .8]), (a_1, [.4, .7])\}$

— $IK_B = \{(b_2, [.2, .7]), (b_1, [0, .5])\}$

— $IK_C = \{(c_2 \vee a_2 \vee b_2, [.3, .7]), (c_2 \vee a_2 \vee b_1, [.2, .4]), (c_2 \vee a_1 \vee b_2, [.5, .6]), (c_2 \vee a_1 \vee b_1, [.8, .9]), (c_1 \vee a_2 \vee b_2, [0, .9]), (c_1 \vee a_1 \vee b_2, [0, .4]), (c_1 \vee a_1 \vee b_1, [0, .5])\}$

Then the knowledge base associated to the interval-based possibilistic network IPN is given by: $IK_{ABC} = IK_A \cup IK_B \cup IK_C$ The interval-based possibilistic knowledge base contains 11 weighted formulas, which might seem a lot considering that the number of states of Ω is only 8. Indeed, some of this formulas are not useful to compute the joint interval-based distribution, but it is more complicated to evaluate which formulas are irrelevant without considering all cases. For instance, a formula can be subsumed by the existence of two others, this problem of eliminating non-necessary formulas from the knowledge base is a complex problem which can be related to the problem of redundancy in logic [Lib05].

An important matter that needs to be investigated is the equivalence of the semantics of those representations. Let us then, compute the two interval-based possibility distributions induced by the two models.

On this example, we are able to state that the interval-based possibility distributions are equivalent. This is the main concern of the next subsection.

5.5.3 Semantical equivalence of the network and the translate knowledge base

In this subsection, we present some results on the interval-based possibilistic encoding. We made two assumptions on this translation. We first need to assure that the knowledge base is an interval-based one. The next proposition states that the transformation of an interval-based possibilistic network in the qualitative setting using the previous definition results in an interval-based possibilistic knowledge base.

Proposition 5.3. *Let IPN be an interval-based possibilistic network then IK_{IPN} is an interval-based possibilistic distribution.*

The proof is straightforward. Since the result of the transformation is the union of interval-based possibilistic knowledge bases and that none of the bases shares the same formulas.

Next we are interested in proving that the proposed transformation between an interval-based possibilistic network and an interval-based possibilistic knowledge base results in the same joint interval-based possibility distributions. Meaning that the two formalisms are semantically equivalent.

The following proposition states that the two induced distributions are equivalent as they assign the same interval degrees for each state of the world.

Proposition 5.4. *Given the joint distribution $I\pi_{IPN}$ associated to IPN, given the joint distribution $I\pi_{IKB}$ associated to the interval-based possibilistic knowledge base IKB_{IPN} computed using Equation (4.5). Then:*

$$\forall \omega \in \Omega,\ I\pi_{IPN}(\omega) = I\pi_{IKB}(\omega)$$

Proof. We need to prove that: $\forall \omega_i \in \Omega$, $I\pi_{IPN}(\omega) = I\pi_{IK}(\omega)$. More precisely:

$$\underline{I\pi_{IPN}}(\omega) = \underline{I\pi_{IK}}(\omega) \text{ et } \overline{I\pi_{IPN}}(\omega) = \overline{I\pi_{IK}}(\omega). \tag{5.13}$$

— First:

$$\begin{aligned}
\underline{I\pi_{IK}}(\omega) &= 1 - \max\{\overline{\alpha} : (\varphi, \alpha) \in IK, \omega \nvDash \varphi\} \\
&= \min\{1 - \overline{\alpha} : (\varphi, \alpha) \in IK, \omega \nvDash \varphi\} \\
&= \min\{1 - \overline{\alpha} : (\neg x_i \vee \neg par(x_i), \alpha) \in IK, \omega \nvDash \neg x_i \vee \neg par(x_i)\} && (*) \\
&= \min\{\underline{\alpha} : (\neg x_i \vee \neg par(x_i), \alpha) \in IK, \omega \nvDash \neg x_i \vee \neg par(x_i)\} \\
&= \min\{\underline{I\pi_{x_i, par(x_i)}}(\omega) : (\neg x_i \vee \neg par(x_i), \alpha) \in IK, \omega \nvDash \neg x_i \vee \neg par(x_i)\} && (**) \\
&= \underline{I\pi_{IPN}}(\omega)
\end{aligned}$$

  — (*) given by the definition of the combination operator $\mathcal{C}^m$ and

  — (**) given by the definition of *IPN* in the forms of a set of triple (Equation (5.11)).

— Then the other bound can be proven using the same reasoning.

□

The previous result is an important one that allows us to design efficient inference machinery for the interval-based possibilistic networks. When it comes to reason with belief graphical models we mostly are interested in two inference tasks: finding the most probable assignment (*MAP* inference) and computing marginal distributions. In the interval-based possibilistic knowledge base, we have shown that computing IK_ϕ is in $O(\log_2(m).SAT)$ where SAT is a satisfiability test of a set propositional clauses and m is the number of different weights in $\overline{IK}$ and $\underline{IK}$ which is the exact same complexity as in standard possibilistic logic. Moreover, the combination operator $\mathcal{C}^m$ proposed here is an extension to interval-based degree of the one in [BDGP02], so the amount of different weighted in $\overline{IK}$ is the same that in any K compatible. This leads us to investigate the complexity of the proposed translation

Proposition 5.5. *The translation from IPN to IK using the combination operator $\mathcal{C}^m$ is done in linear time in the number of variables.*

The above proposition states that the complexity is linear in the number of variables. The proof of that statement is explained by the fact that the union of the knowledge bases is straightforward. And to transform each local distribution into a knowledge base, we need to consider every interpretations in the distribution, this is done in one loop.

5.6 Concluding remarks

In the previous results on quantitative conditioning, we have proposed a set of seven postulates **IC1-IC7** for product-based conditioning. The question is how to relate our postulates **P1-P3** to **IC1-IC7**? Of course, the postulates **IC1-IC7** use the product-based operator while **P1-P3** use the min-based conditioning. Now, if **P1** is replaced by **P'1** stating that:

$$\textbf{P'1}\ \forall \pi, \phi \subseteq \Omega \text{ and } \omega \in \Omega, ([\pi, \pi])(\omega|\phi) = [\pi(\omega|_{*}\phi), \pi(\omega|_{*}\phi)]. \tag{5.14}$$

Then we can show that an interval-based conditioning that satisfies **P'1**,**P2**,**P3** necessarily satisfies **IC1-IC7** but the converse is false.

This chapter addressed the issue of conditioning in a qualitative interval-based possibilistic setting. Four main contributions were presented:

i) A set of three natural postulates **P1-P3** ensuring that any interval-based conditioning satisfying these three postulates is necessarily based on min-based conditioning the set of compatible standard possibility distributions. The first postulate **P1** aims to recover the standard min-based conditioning in case where all the intervals contain singleton values (all lower endpoints coincide with upper endpoints). The second postulate **P2** captures a kind of specificity regarding conditioning interval-based sets of beliefs while the third postulate **P3** aims to ensure a minimality condition.

ii) Efficient procedures to compute the lower and upper endpoints of the conditional interval-based possibility distribution. Such procedures exclude any state of the world that is inconsistent with the new evidence in hand and perform some kind of normalization based on the concept of compatible possibility distribution without generating the whole set of compatible distributions.

iii) A syntactic counterpart of conditioning interval-based possibilistic bases. This counterpart performs some tests and does some modifications on the formulas of the original knowledge base such that the new evidence is integrated with a certainty degree of 1. This ensures the same result as if the knowledge base were conditioned at the semantic level.

iv) An encoding of interval-based possibilistic networks into interval-based possibilistic knowledge bases. This procedure provides the semantical equivalence of the two representations. This allows to perform efficiently conditioning in interval-based possibilistic networks.

Interestingly enough, the syntactic counterpart of min-based conditioning has also the same complexity as conditioning standard possibilistic knowledge bases. More precisely, conditioning an interval-based possibilistic knowledge base does not require extra computational cost compared with conditioning a standard possibilistic base.

Chapter 6

Set-valued possibilistic framework : Definitions and conditioning

Chapters 4 and 5 presented conditioning interval-based possibility distributions and knowledge bases. This chapter deals with conditioning uncertain information where the weights associated with formulas are in the form of sets of uncertainty degrees. The first part of the chapter studies set-valued possibility theory where we provide a characterization of set-valued possibilistic logic bases and set-valued possibility distributions by means of the concepts of compatible possibilistic logic bases and compatible possibility distributions respectively. The second part addresses conditioning set-valued possibility distributions. We first adapt the set of three natural postulates proposed in Chapter 5 for conditioning set-valued possibility distributions. We then show that any set-valued conditioning satisfying these three postulates is necessarily based on conditioning the set of compatible standard possibility distributions. The last part of the chapter shows how one can efficiently compute set-valued conditioning over possibilistic knowledge bases.

Contents

6.1 Set-valued possibility theory and set-valued possibilistic logic

Let us first start with a short example to motivate our extension.

Example 6.1. Suppose we are interested in the amenities and facilities of a hotel in Paris to organize a conference. For this, we posted a question on a specialized Internet platform. To simplify, the question was about the presence of a large conference room in the hotel (represented by the variable c) and if the hotel has a great restaurant (represented by a the variable r) to host the gala dinner. We also asked people to specify how certain of the answers they are, using a unit scale $[0, 1]$. Assume that we got three answers of three people: p_1 is a former hotel employee, the second, p_2, is an employee of the Paris tourism office and the third, $p3$, is a client of the hotel. The certainty levels of these people with respect to different scenarios [5] are summarized as follows:

5. In this example, the scenario cr means that the hotel has a conference room and has a great restaurant while the scenario $c\neg r$ means that the hotel has a conference room but does not have a great restaurant .

Table 6.1 – Example of multiple sources information

	p_1	p_2	p_3
cr	1	1	1
$\neg cr$	1	1	1
$c\neg r$	.3	.2	.4
$\neg c\neg r$	.4	.4	.4

Table 6.2 – Set-valued distribution corresponding to the multiple source information of Table 6.1.

	$I\pi$
cr	$\{1\}$
$\neg cr$	$\{1\}$
$c\neg r$	$\{.2, .3, .4\}$
$\neg c\neg r$	$\{.4\}$

In this example, the confidence degrees provided by the responders can be viewed as possibility degrees. Now, suppose that we got hundreds or thousands of answers or suppose that there is a large number of variables, then it will be interesting to find a compact way to encode the obtained answers and more importantly to reason with them (answer any query of interest and update the available information when new sure information is obtained). Set-valued possibility theory is especially tailored to this type of information.

Let us now introduce the concept of set-valued possibility distribution.

6.1.1 Set-valued possibility distributions

In the set-valued possibilistic setting, the available knowledge is encoded by a set-valued possibility distribution $I\pi$ where each state ω is associated with a finite set $I\pi(\omega)$ of possible values of possibility degrees $\pi(\omega)$.

If S is a set, then we denote by $\overline{S}$ and $\underline{S}$ the maximum and minimum values of S respectively. When all S's associated with interpretations (or formulas) are singletons (meaning that $\overline{S} = \underline{S}$), we refer to standard distributions (resp. standard possibilistic bases). Here, $\underline{I\pi}(\omega)$ (resp. $\overline{I\pi}(\omega)$) denotes the minimum (resp. maximum) of the possibility degrees of ω.

Clearly, set-valued possibility theory is also an extension of interval-based possibility theory [BHLR11], where the set is denoted as an interval of possible values. Therefore, we now consider sets of degrees and we define a set-valued possibility distribution as follows:

Definition 6.1 (Set-valued possibility distribution)**.** *A set-valued possibility distribution* $I\pi$ is a mapping $I\pi : \Omega \to \mathcal{S}$ from the universe of discourse Ω to the set $\mathcal{S}$ of all sub-sets included in the interval $[0, 1]$, with the normalization property requiring that $\max_{\omega\in\Omega} \overline{I\pi}(\omega) = 1$.

The information corresponding to Example 6.1 could be compactly encoded as follows:

Example 6.2. **(Example 6.1 cont'd.)** Let us represent the available knowledge from Example 6.1 as a set-valued possibility distribution given in Table 6.2.

As in an interval-based possibility theory [BHLR11], we also interpret a set-valued possibility distribution as a family of compatible standard possibility distributions defined by:

Table 6.3 – Example of set-valued possibility distribution $I\pi$, compatible possibility distributions π_1 and π_2 and a non compatible one π_3.

$\omega \in \Omega$	$I\pi$
cr	$\{1\}$
$\neg cr$	$\{1\}$
$c\neg r$	$\{.2, .3, .4\}$
$\neg c\neg r$	$\{.4\}$

$\omega \in \Omega$	π_1	π_2	π_3
cr	1	1	.4
$\neg cr$	1	1	1
$c\neg r$	.3	.4	.2
$\neg c\neg r$	.4	.4	.4

Definition 6.2. Let $I\pi$ be a set-valued possibility distribution. A normalized possibility distribution π is said to be compatible with $I\pi$ if and only if $\forall \omega \in \Omega,\ \pi(\omega) \in I\pi(\omega)$.

As shown in Example 6.3, compatible distributions are not unique. We denote by $\mathcal{C}(I\pi)$ the set of all possibility distributions compatible with $I\pi$.

Example 6.3. Let $I\pi$ be a set-valued possibility distribution described in Table 6.3.

Then following Definition 6.2, the possibility distributions π_1 and π_2 (from Table 6.3) are compatible with $I\pi$.

However, π_3 is not compatible with $I\pi$ since $\pi_3(cr) = .4 \notin I\pi(cr) = \{1\}$.

Let us now see how to generalize standard possibilistic logic into a set-valued setting.

6.1.2 Set-valued possibilistic logic

Contrary to standard possibilistic logic where the uncertainty is described with single values, set-valued possibilistic logic uses sets.

The syntactic representation of set-valued possibilistic logic generalizes the notion of a possibilistic base to a set-valued possibilistic knowledge base as follows:

Definition 6.3. A set-valued possibilistic knowledge base, denoted by IK, is a set of propositional formulas associated with sets:

$$IK = \{(\varphi, S), \varphi \in \mathcal{L} \text{ and } S \text{ is a set of degrees in } [0,1]\}$$

In Definition 6.3, $\varphi \in \mathcal{L}$ denotes again a formula of a propositional language $\mathcal{L}$.

A set-valued possibilistic base IK can be viewed as a family of standard possibilistic bases called compatible bases. More formally:

Definition 6.4 (Compatible possibilistic base)**.** A possibilistic base K is said to be compatible with a set-valued possibilistic base IK if and only if K is obtained from IK by replacing each set-valued formula (φ, S) by a standard possibilistic formula (φ, α) with $\alpha \in S$.

In other words, each compatible possibilistic base is such that $K = \{(\varphi, \alpha) : (\varphi, S) \in IK$ and $\alpha \in S\}$.

We also denote by $\mathcal{C}(IK)$ the finite set of all compatible possibilistic bases associated with a set-valued possibilistic base IK.

Example 6.4. In the following, we will use this set-valued possibilistic knowledge base to illustrate our propositions. Let IK be a set-valued possibilistic knowledge base such that:

$$IK = \{(\neg c \vee r, \{.4, .7, .8\}), (r, \{.6\})\}.$$

An example of a compatible possibilistic knowledge base is:

$$K = \{(\neg c \vee r, .4), (r, .6)\}.$$

As in standard possibilistic logic, a set-valued knowledge base IK is also a compact representation of a set-valued possibility distribution $I\pi_{IK}$.

6.1.3 From set-valued possibilistic bases to set-valued possibility distributions

Let us go one step further with the contribution on how to compute the set-valued possibility distribution from a set-valued base.

Let $IK = \{(\varphi_i, S_i)$: i=1, ..., $n\}$ be a set-valued possibilistic knowledge base. A natural way to define a set-valued possibility distribution, associated with IK and denoted by $I\pi_{IK}$, is to consider all standard possibility distributions associated with each compatible knowledge base. Namely:

Definition 6.5. Let IK be a set-valued possibilistic knowledge base. The set-valued possibility distribution $I\pi_{IK}$ associated with IK is defined by:

$$\forall \omega \in \Omega,\ I\pi_{IK}(\omega) = \{\pi_K(\omega) : K \in \mathcal{C}(IK)\}.$$

Recall that $\mathcal{C}(IK)$ is the set of compatible knowledge bases (given in Definition 6.4) and π_K is given by Equation (2.12).
Similar to the single valued possibilistic logic setting, we can get rid of some formulas of a set-valued knowledge base without any information loss. More precisely, we can ignore any formula of IK attached with only one certainty degree equal to zero, as stated in the following lemma.

Lemma 6.1. Let IK be a set-valued possibilistic base such that $(\delta, \{0\}) \in IK$. Let IK'=$IK \setminus \{(\delta, \{0\})\}$. Then $\forall \omega \in \Omega$, $I\pi_{IK}(\omega)$=$I\pi_{IK'}(\omega)$.

Lemma 6.1 is again useful for establishing proofs of some propositions. The idea behind this lemma stands in the definition of compatible bases and Lemma 2.1. Indeed, in the case where IK is such that $(\delta, \{0\}) \in IK$, then in every compatible base K, we have $(\delta, 0) \in K$, therefore, as stated in Lemma 2.1, the weighted formula $(\delta, 0)$ can be ignored from K without changing its associated distributions, and this can be generalized to the set-valued formula $(\delta, \{0\})$.

Let us now characterize $I\pi_{IK}$. The following proposition provides the conditions under which the highest possibility degree '1' belongs to $I\pi_{IK}(\omega)$:

Proposition 6.1. *Let IK be a set-valued possibilistic knowledge base. Let ω be an interpretation. Then:*

$$1 \in I\pi_{IK}(\omega) \textit{ if and only if } \omega \vDash \bigwedge\{\varphi : (\varphi, S) \in IK \textit{ and } \underline{S} > 0\}$$

Namely, $1 \in I\pi_{IK}(\omega)$ if and only if ω satisfies all formulas having a strictly positive certainty degree.

Proof. Recall that $1 \in I\pi_{IK}(\omega)$ means that there exists a compatible possibilistic base $K \in \mathcal{C}(IK)$ such that $\pi_K(\omega) = 1$. Now, formulas of K having a certainty degree equal to '0' can be removed, thanks to Lemma 2.1, without changing π_K. The fact that $\pi_K(\omega) = 1$ implies that ω is a model of $\{\varphi : (\varphi, \alpha) \in K, \alpha > 0\}$. This also means that ω is also a model of $\{\varphi, (\varphi, S) \in IK,\ \underline{S} > 0\}$.

Let us now show the converse. Assume that ω is a model of $\{\varphi, (\varphi, S) \in IK,\ \underline{S} > 0\}$. Let K be a compatible possibilistic knowledge base obtained from IK by replacing each set-valued S by its lower bound $\underline{S}$. Clearly, $\{\varphi : (\varphi, \underline{S}) \in K\}$ is satisfied by ω. Hence, $1 \in I\pi_{IK}(\omega)$. □

Example 6.5. **(Example 6.4 cont'd)** Let us continue with the knowledge base from Example 6.4. Recall that $IK = \{(\neg c \vee r, \{.4, .7, .8\}), (r, \{.6\})\}$. Following Proposition 6.1, interpretations cr and $\neg cr$ will have among their possibility degrees the degree 1 (namely $1 \in I\pi_{IK}(cr)$ and $1 \in I\pi_{IK}(\neg cr)$) since these interpretations are models of all the formulas of IK attached only to strictly positive degrees.

We now study under which conditions a possibility degree $(1-\alpha)$ belongs to $I\pi_{IK}(\omega)$, with $\alpha \in [0,1]$. Clearly, if $(1-\alpha) \in I\pi(\omega)$ then there exists a compatible base K such that $\pi_K(\omega)$=$1-\alpha$. Hence, there exists $(\varphi,\alpha) \in K$ such that $\omega \not\models \varphi$. Then there exists $(\varphi, S) \in IK$ such that $\omega \not\models \varphi$ and $\alpha \in S$.

To determine the possible values of $I\pi_{IK}(\omega)$, it is enough to browse all certainty degrees associated with formulas of IK falsified by ω and check whether their inverse will belong or not to $I\pi_{IK}(\omega)$.

This is precisely specified by the following proposition:

Proposition 6.2. *Let ω be an interpretation. Let $A = \bigcup\{S : (\varphi, S) \in IK,\ \omega \not\models \varphi\}$. Let $a \in A \cup \{0\}$. Then,*

$$(1-a) \in I\pi_{IK}(\omega) \textit{ if and only if } \omega \models \{\varphi : (\varphi, S) \in IK,\ \underline{S} > a\}$$

Proof. Proposition 6.2 recovers Proposition 6.1 in case where $a = 0$. Hence, we only focus on the case $a > 0$. To see the proof, assume that $a > 0$ and $(1-a) \in I\pi_{IK}(\omega)$. This means that there exists a compatible possibilistic knowledge base $K \in \mathcal{C}(IK)$, such that $\pi_K(\omega) = 1 - a$.

This means that $\{\varphi : (\varphi, b), b > a\}$ is consistent and satisfied by ω. Since $\{\varphi : (\varphi, S), \underline{S} > a\} \subseteq \{\varphi : (\varphi, b), b > a\}$, this also means that $\{\varphi : (\varphi, S),\ \underline{S} > a\}$ is consistent and satisfied by ω.

Let us show the converse. Assume that $\omega \models \{\varphi : (\varphi, S),\ \underline{S} > a\} \wedge \omega$. Clearly, if $A = \emptyset$ (namely, $a = 0$) or $A = \{0\}$ then whatever is the compatible base K, ω will satisfy each formula in K, hence $\pi_K(\omega) = 1$, and $(1-a) \in I\pi_{IK}(\omega)$. Assume that $a \in A$ and $a > 0$. Let (φ_1, S_1) be a formula of IK such that $a \in S_1$ and $\omega \not\models \varphi_1$. Let K be a compatible base defined by:

$$K = \{(\varphi, \underline{S}) : (\varphi, S) \in IK,\ \varphi \neq \varphi_1\} \cup \{(\varphi_1, a)\}.$$

Namely, K is obtained from IK by replacing S by $\underline{S}$ for each formula in IK, except for φ_1 where a is used instead of $\underline{S}$. It is easy to see that K is compatible with IK, namely $K \in \mathcal{C}(IK)$. It is also easy to see that $\pi_K(\omega) = 1-a$, since $\{\varphi : (\varphi, b) \in K, b > a\}$ is satisfied by ω, $\{\varphi : (\varphi, b) \in K, b > a\} \cup \{(\varphi_1, a)\}$ is falsified by ω. Therefore $(1-a) \in I\pi_{IK}(\omega)$. □

Let us continue our example, and illustrate Proposition 6.2.

Example 6.6. **(Example 6.4 cont'd)** We need to check which degrees belong to $I\pi_{IK}(\omega)$. For each interpretation, we first compute $A = \bigcup\{S : (\varphi, S) \in IK,\ \omega \not\models \varphi\}$. For instance, let us consider $\omega = c\neg r$ then $A = \{.4, .7, .8, .6\}$. Now, let us analyse each value a of $A \cup \{0\}$,

— For $a = 0$, $c\neg r \not\models \{\neg c \vee r, r\}$, then $1 \notin I\pi_{IK}(c\neg r)$;

— For $a = .4$, $c\neg r \not\models \{r\}$, then $.6 \notin I\pi_{IK}(c\neg r)$;

— For $a = .7$, $\emptyset \wedge c\neg r$ is consistent, then $.3 \in I\pi_{IK}(c\neg r)$;

— For $a = .8$, $\emptyset \wedge c\neg r$ is consistent, then $.2 \in I\pi_{IK}(c\neg r)$

— Finally, for $a = .6$, $\emptyset \wedge c\neg r$ is consistent, then $.4 \in I\pi_{IK}(c\neg r)$.

Then we can conclude that $I\pi_{IK}(c\neg r) = \{.2, .3, .4\}$.

Let us take another interpretation, for instance ω=$\neg c\neg r$. Then $A = \{.6\}$ and for each $a \in A \cup \{0\}$,

— For $a = 0$, $\neg c\neg r \not\models \{\neg c \vee r, r\}$, then $1 \notin I\pi_{IK}(\neg c\neg r)$;

— And for $a = .6$, $\emptyset \wedge \neg c \neg r$ is consistent, then $.4 \in I\pi_{IK}(\neg c \neg r)$.
We can conclude that $I\pi_{IK}(\neg c \neg r) = \{.4\}$.
The whole distribution is exactly the one given in Example 6.2.

Let us now deal with the issue of conditioning a set-valued possibilistic base. The following section extends $\min$-based conditioning to set-valued possibility distributions.

6.2 Conditioning set-valued possibilistic information

Before providing our extension of $\min$-based conditioning to the set-valued setting, let us first focus on the natural properties that a set-valued conditioning operator should fulfill.

6.2.1 Three natural requirements for the set-valued conditioning

The first natural requirement (called recovering standard conditioning) is that in the *degenerate* case, namely when each set $I\pi(\omega)$ contains exactly one single degree $\pi(\omega)$, the result of the new conditioning procedure should coincide with the result $\pi(.|_m\phi)$ of the original conditioning procedure. For each possibility distribution π, by $\{\pi(\omega)\}$ we denote its set-valued representation, *i.e.*, a set-valued possibility distribution for which, for every $\omega \in \Omega$, we have $I\pi(\omega) = \{\pi(\omega)\}$. In these terms, the above requirement takes the following form:

S1. If for every $\omega \in \Omega$, we have $I\pi(\omega) = \{\pi(\omega)\}$, then $I\pi(\omega|\phi) = \{\pi(\omega|_m\phi)\}$ for all ω and ϕ.

The second requirement (called specificity) is related to the fact that we do not know the precise values $S\pi(\omega)$ since we only have partial information about them. In principle, if we can get some additional information about these values, then this would lead, in general, to narrower sets (indeed, the cardinality of a set captures the ignorance regarding the exact value of $\pi(\omega)$). Let us define the concepts of specificity between set-valued possibility distribution:

Definition 6.6. Let $I\pi$ and $I\pi'$ be two set-valued possibility distributions. Then $I\pi$ is said to be more specific than $I\pi'$, denoted $I\pi \subseteq I\pi'$, if $I\pi(\omega) \subseteq I\pi'(\omega)$ holds for all $\omega \in \Omega$.

S2. If $I\pi(\omega) \subseteq I\pi'(\omega)$ for all ω, then $I\pi(\omega|\phi) \subseteq I\pi'(\omega|\phi)$ for all ω.

Of course, these two postulates are not sufficient. For example, we can take $S\pi(.|\phi) = \{\pi(.|_m\phi)\}$ for degenerate set-valued possibility distributions and $I\pi(\omega|\phi) = [0, 1]$ for any other set-valued distribution $I\pi$. To avoid such extensions, it is reasonable to impose the following minimality condition:

S3. There does not exist a conditioning operation '$|_1$' that satisfies both properties **S1**–**S2** and for which:

— $I\pi(\omega|_1\phi) \subseteq I\pi(\omega|\phi)$ for all $I\pi$, ω, and ϕ,

— $I\pi(\omega|_1\phi) \neq I\pi(\omega|\phi)$ for some $I\pi$, ω, and ϕ.

S3 is called minimality condition. The following theorem provides one of our main results where we show that there is only one set-valued conditioning satisfying **S1-S3** and where the set conditional possibility degree $I\pi(\omega|\phi)$ is defined as the closure of the set of all $\pi(.|_m\phi)$, where π is compatible with $I\pi$.

Theorem 6.1. *There exists exactly one set-valued conditioning, also denoted by $I\pi(.|\phi)$ for sake of simplicity, that satisfies the properties* **S1**–**S3**, *and which is defined by:* $\forall\omega \in \Omega$,

$$I\pi(\omega|\phi) = \{\pi(\omega|_m\phi) : \pi \in \mathcal{C}(I\pi)\} \tag{6.1}$$

where $|_m$ is the $\min$*-based conditioning given in Equation (1.29).*

Proof. 1°. Let us denote the corresponding set-based conditioning by $I\pi(.|\phi)$. We need to prove:

— that this closure $I\pi(.|\phi)$ satisfies the properties **S1**–**S3**, and

— that every operation $I\pi(.|_1\phi)$ that satisfies the properties **S1**–**S3** coincides with the set-conditioning $I\pi(.|\phi)$.

2°. One can easily see that the operation $I\pi(.|\phi)$ satisfies the properties **S1**–**S2**.

3°. Let us now prove that if an operation $I\pi(.|_1\phi)$ satisfies the properties **S1**–**S2**, then for every $I\pi$ and ϕ, we have $I\pi(.|\phi) \subseteq I\pi(.|_1\phi)$.

Then, for every distribution $\pi \in \mathcal{C}(I\pi)$, we have $\{\pi\} \subseteq I\pi$ and thus, due to the postulate **S2**, we have $\{\pi\}(.|_1\phi) \subseteq I\pi(.|\phi)$. By the property **S1**, we have $\{\pi\}(\omega|_1\phi) = \{\pi(\omega|_m\phi)\}$. Thus, the above inclusion means that $\pi(.|_m\phi) \in I\pi(.|_1\phi)$.

The set $I\pi(\omega|_1\phi)$ therefore contains all the values $\pi(\omega|_m\phi)$ corresponding to all possible $\pi \in \mathcal{C}(I\pi)$:

$$\{\pi(\omega|_m\phi) : \pi \in \mathcal{C}(I\pi)\} \subseteq I\pi(\omega|_1\phi). \tag{6.2}$$

Thus, we conclude that $I\pi(\omega|\phi) \subseteq I\pi(\omega|_1\phi)$ for all ω.

The statement is proven.

4°. We can now prove that $I\pi(.|\phi)$ also satisfies the property **S3**.

Indeed, if there is some other operation $|_1$ that satisfies **S1** and **S2**, and for which $I\pi(\omega|_1\phi) \subseteq I\pi(\omega|\phi)$ for all ω, then, since we have already proven the opposite inclusion in Part 3 of this proof, we conclude that $I\pi(\omega|_1\phi) = I\pi(\omega|\phi)$ for all ω, so indeed no narrower conditioning operation is possible.

5°. To complete the proof, let us show that if some $I\pi(.|_1\phi)$ satisfies the properties **S1**–**S3**, then it coincides with $I\pi(.|\phi)$.

Indeed, by Part 3 of this proof, we have $I\pi(\omega|\phi) \subseteq I\pi(\omega|_1\phi)$ for all ω. If we had $I\pi(\omega|\phi) \neq I\pi(\omega|_1\phi)$ for some ω and ϕ, this would contradict the minimality property **S3**. Thus, indeed, $I\pi(.|\phi) = I\pi(.|_1\phi)$. Uniqueness is proven, and so is for the theorem. □

6.2.2 Analyzing set-based conditioning

Now, we can go one step beyond Theorem 6.1 and provide the exact contents of the conditioned set $I\pi(.|_m\phi)$. Let us first start with the following lemma which delimits the set of possible values associated with models of ϕ after the conditioning operation.

Lemma 6.2. Let $I\pi$ be a set-valued possibility distribution. Let $\phi \subseteq \Omega$. Then $\forall \omega \in \Omega$,

— If $\omega \nvDash \phi$, $I\pi(\omega|\phi) = \{0\}$,

— And if $\omega \vDash \phi$, $I\pi(\omega|\phi) \subseteq I\pi(\omega) \cup \{1\}$.

The proof of this lemma is immediate. Indeed, if π is a standard possibility distribution, then by definition $\pi(\omega|_m\phi)$ is either equal to $\pi(\omega)$ or to 1 for models of ϕ. Hence, the only admissible values for $I\pi(\omega|\phi)$ are those in $I\pi(\omega)$ and the value 1. For counter-models of ϕ (namely, $\omega \nvDash \phi$), then clearly $I\pi(\omega|\phi) = \{0\}$ since $\pi(\omega|_m\phi) = 0$ for each compatible distributions π.

Given this lemma, we need to answer two questions:

— Under which conditions does the fully possibility degree 1 belong to $I\pi(\omega|\phi)$?

— Under which conditions will a given possibility degree $a \in I\pi(\omega)$ still belong to $I\pi(\omega|\phi)$?

The answer to these questions is given in the following proposition:

Proposition 6.3. *Let $I\pi$ be a set-valued possibility distribution. Let $\phi \subseteq \Omega$.*

Table 6.4 – Set-valued distribution $S\pi$ of Example 6.2 conditioned by ϕ=$\neg r$.

	$I\pi(.\|\phi)$
cr	$\{0\}$
$\neg cr$	$\{0\}$
$c\neg r$	$\{.2, .3, 1\}$
$\neg c\neg r$	$\{1\}$

i) $1 \in I\pi(\omega|\phi)$ *if and only if* $\forall\omega' \neq \omega$, $\overline{I\pi}(\omega) \geq \underline{I\pi}(\omega')$.

ii) Let $a \in I\pi(\omega)$ *(with* $a \neq 1$*). Then* $a \in I\pi(\omega|\phi)$ *if and only if* $\exists\omega' \neq \omega$, $\overline{I\pi}(\omega') > a$.

Proof. For item (i) assume that $1 \in I\pi(\omega|\phi)$. This means that there exists a compatible distribution π of $I\pi$ such that $\pi(\omega|_m\phi) = 1$. This also means that $\forall\omega' \neq \omega$, $\pi(\omega) \geq \pi(\omega')$. Since, $\overline{I\pi}(\omega) \geq \pi(\omega)$, and $\pi(\omega') \geq \underline{I\pi}(\omega')$, hence we have $\forall\omega' \neq \omega$, $\overline{I\pi}(\omega) \geq \underline{I\pi}(\omega')$. For the converse, assume that $\forall\omega'$, $\overline{I\pi}(\omega) \geq \underline{I\pi}(\omega')$. Let π be a compatible distribution such that $\pi(\omega) = \overline{I\pi}(\omega)$ and $\forall\omega' \neq \omega$, $\pi(\omega') = \underline{I\pi}(\omega)$. Clearly, $\forall\omega' \neq \omega$, $\pi(\omega) > \pi(\omega')$. Hence $\pi(\omega|_m\phi) = 1$ and $1 \in I\pi(\omega|\phi)$.

For item (ii), let $a \in I\pi(\omega)$ where $a \neq 1$. Assume that $\exists\omega' \neq \omega$, such that $\overline{I\pi}(\omega') > a$. Consider a compatible distribution π where $\pi(\omega')$=$\overline{I\pi}(\omega')$ and $\pi(\omega)$=a. Then clearly, $\pi(\omega_m|\phi)$=$a \in I\pi(\omega|\phi)$. For the converse, assume that $a \in I\pi(\omega|\phi)$ and $a \neq 1$. This means that there exists a compatible distribution π such that $\pi(\omega|_m\phi)$=$a < 1$. Hence, $\exists\omega'$, $\pi(\omega)$=$a < \pi(\omega')$. Since $\pi(\omega') \leq \overline{I\pi}(\omega')$ this means that $\overline{I\pi}(\omega') > a$.

□

Example 6.7. In this example, we deal with conditioning a set-valued possibility distribution. Therefore, let us continue Example 6.2 and assume that the manager of the hotel tells us that the restaurant of the hotel has closed down definitively a few weeks ago. Then we need to condition with the new piece of information $\phi = \neg r$. Let us run the conditioning operation step by step. For every interpretation model of ϕ,

— For $\omega = c\neg r$,

 i) since, with $\omega' = \neg c\neg r$, $.4 \geq .4$, then $1 \in I\pi(c\neg r|\neg r)$;

 ii) For $a = .2$, since, $\overline{I\pi}(\neg c\neg r)$=$.4 > .2$, then $.2 \in I\pi(c\neg r|\neg r)$.
 For $a = .3$, since, $\overline{I\pi}(\neg c\neg r)$=$.4 > .2$, then $.3 \in I\pi(c\neg r|\neg r)$.
 For $a = .4$, since, $\overline{I\pi}(\neg c\neg r)$=$.4 \not> .4$, then $.4 \notin I\pi(c\neg r|\neg r)$.

— For the interpretation ω=$\neg c\neg r$, we follow the same computation steps.

— For counter-models of $\neg r$, we have $I\pi(\omega|\phi) = \{0\}$.

Given the distribution in Table 6.2, we sum up the result of conditioning this distribution in Table 6.4.

6.3 A syntactic counterpart of set-valued conditioning

Let us first consider again conditioning a standard possibilistic knowledge base K and rewrite the result of conditioning K. Recall that $K_{\geq a} = \{\varphi : (\varphi, \alpha) \in K \text{ and } \alpha \geq a\}$ be a set of propositional formulas from K having a weight greater or equal to a. Then, the result of conditioning K by ϕ, denoted by K_ϕ, given by Definition 2.17 can be rewritten as:

$$\begin{aligned} K_\phi &= \{(\phi, 1)\} \\ &\cup \{(\varphi, \alpha) : (\varphi, \alpha) \in K_{\geq\alpha} \wedge \phi \text{ is consistent }\} \\ &\cup \{(\varphi, 0) : (\varphi, \alpha) \in K_{\geq\alpha} \wedge \phi \text{ is inconsistent }\}. \end{aligned}$$

The only difference with Definition 2.17 is that '0' weighted formulas have been added. This has no influence thanks to Lemma 2.1. Namely, K_ϕ is obtained from K by adding ϕ with a fully certainty degree and ignore some formulas from K. By ignoring some formulas, we mean the certainty degrees of these formulas are set to '0'.

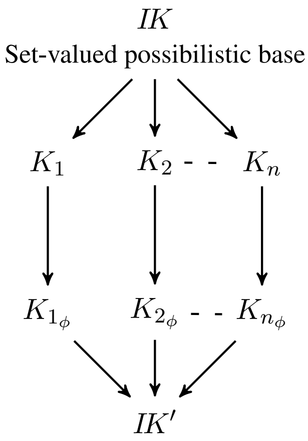


Figure 6.1 – Compatible-based conditioning

The aim of this section is to provide syntactic computation of set-valued conditioning when set-valued possibility distributions are compactly represented by set-valued possibilistic knowledge bases. As illustrated in Figure 6.1, the input is an initial set-valued knowledge base IK and a formula ϕ. The output is a new set-valued knowledge base IK' that results from conditioning the set of all compatible bases of IK with ϕ. This new set-valued knowledge base IK' is obtained by considering the set of all compatible possibilistic knowledge bases, $K_i \in \mathcal{C}(IK)$. More precisely, it is done in three steps:

— First, from IK we generate the set of compatible bases $K_1, K_2, ..., K_n$

— then, we condition each compatible base K_i with ϕ. The result is K_{i_ϕ} and obtained using Definition 2.17.

— Lastly, we define IK' by associating with each formula φ of IK the set of degrees present in at least one conditioned K_{i_ϕ}.

Namely: $IK' = \{(\varphi, S) : S = \bigcup\{\alpha_k : (\varphi, \alpha_k) \in K_\phi, K \in \mathcal{C}(IK)\}\}$.

Hence, a naive algorithm for computing IK' is given.

Clearly, this algorithm is not satisfactory since the number of compatible bases may be exponential.

Our aim is then to equivalently compute IK' without exploiting the set of all compatible possibilistic knowledge bases.

It is easy to show that $\forall \omega \in \Omega$, $\pi_{K'}(\omega) = \pi_K(\omega|\phi)$. Now, in the set-valued setting, conditioning IK comes down first to apply standard conditioning on each compatible base then gathering all certainty degrees. Clearly, IK' is obtained from IK by ignoring some weight. The conditions under which a weight should be ignored is given by the following proposition:

Proposition 6.4. *Let IK be a set-valued knowledge base, ϕ be a propositional formula. Let $(\gamma, S) \in IK$ and $a \in S$. Let S' be the new set associated with γ in IK'. Then:*

$$a \in S' \text{ if and only if } \phi \wedge \{\varphi : (\varphi, S) \in IK, \underline{S} \geq a\} \wedge \gamma \text{ is consistent.}$$

Proof. The proof is as follows. Assume that $a \in S'$. This means that there exists a compatible base K such that $(\gamma, a) \in K'$. Since $\{\varphi : (\varphi, \alpha) \in K'\}$ is consistent, and $(\gamma, a) \in K'$ and $(\phi, 1) \in K'$ then

Algorithm 2 Naive computation of IK'

Require: IK: a set-valued knowledge base
  ϕ: a propositional formula
Ensure: IK': the result of conditioning IK with ϕ
  $IK' \longleftarrow \{(\phi, 1)\}$
  for $(\gamma, S) \in IK$ **do**
    $S' \longleftarrow \emptyset$
    for K compatible with IK **do**
      Compute K_ϕ
      $S' \longleftarrow S' \cup \{\alpha : (\gamma, \alpha) \in K_\phi\}$
    end for
    $IK' \longleftarrow IK' \cup \{(\gamma, S')\}$
  end for
  return IK'

trivially $\phi \wedge \gamma \wedge \{\varphi : (\varphi, b) \in K'\}$ is consistent. Hence, $\phi \wedge \gamma \wedge \{\varphi : (\varphi, b) \in K',\ b \geq a\}$ is consistent and $\phi \wedge \gamma \wedge \{\varphi : (\varphi, S) \in IK,\ \underline{S} \geq a\}$ is consistent.

Now, assume that $\phi \wedge \gamma \wedge \{\varphi : (\varphi, S) \in IK,\ \underline{S} \geq a\}$ is consistent. Let K be a compatible base, where each (φ, S) such that $\varphi \neq \gamma$ is replaced by $(\varphi, \underline{S})$ and (γ, S) is replaced by (γ, a). Clearly, K is a compatible. Besides, $(\gamma, a) \in K'$ since $K_{\geq a} \wedge \phi$ is consistent. Hence, $a \in S'$. □

Based on the above propositions, we propose an algorithm (Algorithm 3) to compute the result of conditioning IK with ϕ. It consists in browsing all the degrees of IK and checking whether each degree should be replaced by 0 or not.

Algorithm 3 Syntactic set-valued conditioning

Require: IK: a set-valued knowledge base
  ϕ: a propositional formula
Ensure: IK': the result of conditioning IK with ϕ
  $IK' \longleftarrow \{(\phi, 1)\}$
  for $(\gamma, S) \in IK$ **do**
    $S' \longleftarrow \emptyset$
    for $a \in S$ **do**
      if (#) $\phi \wedge \gamma \wedge \{\varphi : (\varphi, S) \in IK, \underline{S} \geq a\}$ is consistent **then**
        $S' \longleftarrow S' \cup \{a\}$
      else
        $S' \longleftarrow S' \cup \{0\}$
      end if
      $IK' \longleftarrow IK' \cup \{(\gamma, S')\}$
    end for
  end for
  return IK'

In Algorithm 3, the costly task is checking consistency of the statement marked by (#). Hence, the complexity of computing IK' is $O(|IK| * n * SAT)$ where n is the number of different certainty levels in IK (namely, $n = |\bigcup\{S : (\varphi, S) \in IK\}|$). This is stated in the following proposition.

Proposition 6.5. *Let IK be a set-valued possibilistic knowledge base and ϕ be the new evidence. Let*

Table 6.5 – Set-valued distribution corresponding to set-valued knowledge base IK'.

	$I\pi_{IK'}$
cr	$\{0\}$
$\neg cr$	$\{0\}$
$c\neg r$	$\{.2, .3, 1\}$
$\neg c\neg r$	$\{1\}$

IK' be a set-valued possibilistic knowledge base computed using Algorithm 3. Then computing IK_ϕ is in $O(|IK| * n * SAT)$ where SAT is a satisfiability test of a set propositional clauses and n is the number of different weights in IK.

Example 6.8. Let us illustrate Algorithm 3. To do so, we continue Example 6.4 where $IK = \{(\neg c \vee r, \{.4, .7, .8\}), (r, \{.6\})\}$ and with the new information $\phi = \neg r$. For each pair (φ, S),

- First let us take $(\neg c \vee r, \{.4, .7, .8\})$ then:
  - For $a = .4$, $\{r, \neg c \vee r\} \wedge \{\neg r\} \wedge \{\neg c \vee r\}$ is not consistent then, $0 \in S'$;
  - For $a = .7$, $\emptyset \wedge \{\neg r\} \wedge \{\neg c \vee r\}$ is consistent then, $.7 \in S'$;
  - We use the same reasoning for $a = .8$, then, $.8 \in S'$.
- Now for the second pair $(r, \{.6\})\}$ we have:
  - For $a = .6$, $\{r\} \wedge \{\neg r\} \wedge \{r\}$ is not consistent so $0 \in S'$;

The new base is $IK' = \{(\neg r, \{1\}), (\neg c \vee r, \{0, .7, .8\}), (r, \{0\})\}$. Thanks to Lemma 6.1, we can exclude the pair $(r, \{0\})$, this is our new base: $IK' = \{(\neg r, \{1\}), (\neg c \vee r, \{0, .7, .8\})\}$. The corresponding set-valued possibility distribution according Definition 6.5 is given in Table 6.5.

6.4 Conclusion

This chapter dealt with representing and reasoning with qualitative information in a possibilistic setting and it provided three main contributions:

- The first one is a new extension of possibilistic logic called set-valued possibilistic logic particularly suited for reasoning with qualitative and multiple source information. We provided a natural semantics in terms of compatible possibilistic bases and compatible possibility distributions.
- The second main contribution deals with a generalization of the well-known min-based or qualitative conditioning to the new set-valued setting. The chapter proposes three natural postulates ensuring that any set-valued conditioning satisfying these three postulates is necessarily based on the set of compatible standard possibility distributions.
- The third main contribution concerns the syntactic characterization of set-valued conditioning. Efficient procedures are proposed to compute the exact set-valued possibility distributions and their syntactic counterparts. Interestingly enough, the proposed setting generalizes standard possibilistic and conditioning does not require extra computational cost with respect to the standard single valued possibilistic setting. We provide an algorithm which does not generate explicitly the set of all compatible possibilistic knowledge bases.

Note that the idea of compatible-based conditioning in the interval-based possibilistic setting is somehow similar to conditioning in credal sets [ACdCT14, Lev80] and credal networks [Coz00] where the concept of convex set refers to the set of compatible probability distributions composing the credal set.

Regarding the computational cost, conditioning in credal sets is done on the set of extreme points (edges of the polytope representing the credal set) but their number can reach $N!$ where N is the number of interpretations [Wal07]. In this chapter, our set-valued conditioning operator has a complexity close to the one of standard possibilistic knowledge base.

Part III

Analysis of probability-possibility transformations and MAP queries

Chapter 7

Property analysis of probability-possibility transformations

The last three chapters made it clear that possibilistic frameworks have flexible and expressive compact representations and have efficient inference machinery. In order to benefit from such machinery we need to study, with respect to reasoning tasks, transformations from probabilistic settings to the possibilistic one. In this chapter, we focus on probability-possibility transformations in the context of changing operations and graphical models. Existing works mainly propose probability-possibility transformations satisfying some desirable properties. The analysis of the behavior of these transformations with respect to changing operations (such as conditioning and marginalization) have not been addressed. This chapter analyses the commutativity of probability-possibility transformations with respect to some reasoning tasks such as marginalization and conditioning. Another crucial issue addressed in this chapter is the one of probability-possibility transformations in the context of graphical models.

Contents

7.1 Properties analyzed for probability-possibility transformations

The purpose of this chapter is to study the commutativity of transformations on reasoning tasks such that conditioning and marginalization. Sudkamp [Sud92] was first to study this question but his focus was limited to the case where the resulted distributions are *identical*. He showed that there is no transformation satisfying commutativity of transformations with respect to operations like conditioning and marginalization. In this chapter, we only focus on the preservation of the ordering between interpretations. Such study is important especially for handling *MPE* and *MAP* queries for which it is not required to have real certainty degrees of interpretations but only the ordering between interpretations. Indeed,

MAP and MPE queries only care for interpretations having the maximum degree (given a particular configuration, like an evidence). We denote by TR the ordering between interpretations such that $TR(p)$ is the descending order of interpretations given the degrees in p.

$$\begin{tikzcd}
p \arrow[rr, "\text{Changing Operation}"] \arrow[d] & & p' \arrow[d] \\
\pi = TR(p) \arrow[r, "\text{Changing Operation}"] & \pi'' \arrow[r, leftrightarrow, "TR?"'] & \pi' = TR(p')
\end{tikzcd}$$

Figure 7.1 – Preserving the ordering between interpretations

We consider operations on distributions as depicted on Figure 7.1. Let p be a probability distribution. On one hand, one may obtain a possibility distribution by first applying a changing operation on p (which leads to p') then applying a probability-possibility transformation on p' (which leads to π'). On the other hand, we may obtain a possibility distribution by first transforming the probability distribution p (with $TR(p)$) and then applying the corresponding changing operation which leads to a possibility distribution (π''). Our objective is to compare these distributions and see if they encode the same ordering between every interpretations $\omega \in \Omega$.

Let us consider TR a probability-possibility transformation procedure that preserves Dubois and Prade principles(such as the ones presented in Chapter 3 Section 3.1.2: OT, KT, ST and VT). We try to answer the following questions:
Normalization condition: Does OT (resp. KT, ST and VT) transformation gives a normalized possibility distribution? Namely, do we have

$$\Pi_{OT}(\Omega) = 1?$$

$$\Pi_{KT}(\Omega) = 1?$$

$$\Pi_{ST}(\Omega) = 1?$$

$$\Pi_{VT}(\Omega) = 1?$$

Plausibility ordering between events: Does TR preserve the plausibility ordering between events? Namely, let ϕ and ψ be two events in Ω such that $P(\phi) < P(\psi)$, then do we have:

$$\Pi_{TR}(\phi) < \Pi_{TR}(\psi) \ ?$$

Marginalization task: Is the ordering between interpretations in a marginalized probability distribution the same as the one in a marginalized transformed possibility distribution. More precisely, let p be a probability distribution over A and B, let p_A be the marginalized probability distribution over A and let a_1 and a_2 be two worlds of D_A such that $p_A(a_1) < p_A(a_2)$. Then we have $\Pi_{A-TR}(a_1) < \Pi_{A-TR}(a_2)$ but do we have:

$$\Pi_{TR}(a_1) < \Pi_{TR}(a_2)$$

where π_{TR} is the transformed possibility distribution of p?

Conditioning task: Is the ordering between interpretations in a conditioned probability distribution the same as the one in a conditioned transformed possibility distribution. More precisely, let p be a probability distribution and let ϕ be an event. Let ω_1 and ω_2 be two interpretations of Ω such that $P(\omega_1|\phi) < P(\omega_2|\phi)$. Then do we have:

$$\Pi_{TR}(\omega_1|\phi) < \Pi_{TR}(\omega_2|\phi)$$

where π_{TR} is the possibility distribution associated with p using TR?

Independence relationships: Let p be a probability distribution over a set of variables $V = \{X_1, ..., X_n\}$ such that $X_i \perp X_j$ (with $i \neq j$). Let π_{TR} be the possibility distribution transformed by TR. Then do we have:

$$\forall x_i \in D_{X_i},\ x_j \in D_{X_j},\ \ \Pi_{TR}(x_i|x_j) = \Pi_{TR}(x_i)$$

MPE and MAP queries: Let p be a probability distribution. Is the result of a *MPE* (resp. *MAP*) query in p the same as the result in π_{TR}?

Namely, let ω^* be the result of a *MPE* query in p and let ω^{TR} be the result of the *MPE* query in π_{TR} then is $\omega^* = \omega^{TR}$? In the same way, let ω_* be the result of a *MAP* query in p and let ω_{TR} be the result of the *MAP* query in π_{TR} then is $\omega_* = \omega_{TR}$?

The following sections provide answers to each of these questions.

7.2 Preserving normalization

In Chapter 3, we outlined existing works on probability-possibility transformations. It is important to see if each proposed probability-possibility transformation operation gives a normalized possibility distribution. This section shows that it is the case for each probability-possibility transformation.

7.2.1 OT transformation

Recall that OT (the optimal transformation) satisfies all the principles defined in Section 3.1.1. We show that OT also satisfies the normalization principle. Namely, any normalized probability distribution will give a normalized possibility distribution using OT transformation procedure.

Proposition 7.1. *Let p be a probability distribution over Ω. Let π_{OT} be the possibility distribution resulting from the transformation of p with OT. If p is normalized then π_{OT} is normalized.*

Proof. Let p be a normalized probability distribution, then we have:

$$\sum_{\omega_i \in \Omega} p(\omega_i) = 1$$

Let π_{OT} be the possibility distribution transformed by:

$$\pi_{OT}(\omega_i) = \sum_{\omega_j / p(\omega_j) \leq p(\omega_i)} p(\omega_j)$$

Let us consider $\omega_{\max}$ as the configuration having the highest probability degree according to p (*i.e.* $\forall\, \omega_i \in \Omega,\ p(\omega_i) \leq p(\omega_{\max})$). Then,

$$\pi_{OT}(\omega_{\max}) = \sum_{\omega_j / p(\omega_j) \leq p(\omega_{\max})} p(\omega_j) = \sum_{\omega_i \in \Omega} p(\omega_i) = 1$$

The possibility distribution π_{OT} is indeed always normalized. □

We can state the same results for KT, ST and VT respectively.

7.2.2 KT, ST and VT transformations

Proposition 7.2. *Let p be a probability distribution over Ω. Let π_{KT} (resp. π_{ST}, π_{VT}) be the possibility distribution resulting from the transformation of p with KT (resp. ST, VT). If p is normalized then π_{KT} (resp. π_{ST}, π_{VT}) is normalized.*

Proof. Let us prove for each transformation starting with KT.

— Let p be a probability distribution over Ω ordered such as $p_i > 0,\ p_i \geqslant p_{i+1}$ with $p_{n+1} = 0$.

$$\pi_1 = \frac{p_1}{p_1} = 1$$

The transformed distribution KT is always normalized.

— In the case where $TR = ST$. Let us recall that the transformation ST is defined by: $\pi_i = \sum_{j=1}^{n} \min(p_i, p_j)$. Let us note the highest probability $p_{\max}$, we obtain:

$$\pi_{\max} = \sum_{j=1}^{n} \min(p_{\max}, p_j) = \sum_{j=1}^{n} p_j = 1$$

Therefore, π_{ST} is always normalized.

— In the last case, $TR = VT$. Let us assume that the elements in Ω are ordered such as: $\forall\ i = \{1..n\},\ p_i > 0,\ p_i \geq p_{i+1}$ with $p_{n+1} = 0$, and:

$$\pi_i = (\frac{p_i}{p_1})^{k.(1-p_i)}$$

where k is a constant belonging to: $0 \leq k \leq \frac{log(p_n)}{(1-p_n).log(\frac{p_n}{p_1})}$.

The transformation also gives us a normalized possibility distribution, indeed:

$$\pi_1 = (\frac{p_1}{p_1})^{k.(1-p_i)} = 1^{k.(1-p_i)} = 1$$

□

7.3 Preserving plausibility ordering between events

One of the principle of Dubois and Prade requires that the order of interpretations must be preserved but nothing is said regarding arbitrary events (sets of interpretations). A natural question is therefore to see if the ordering between arbitrary events is also preserved.

The following example shows that OT transformation does not preserve the plausibility ordering between events.

Example 7.1. Let us consider a probability distribution over 3 worlds $\Omega = \{\omega_1, \omega_2, \omega_3\}$. The possibility distribution of p obtained using OT transformation is depicted in Table 7.1 along with the probability distribution p and the probability and possibility measures of all events $\phi \subseteq \Omega$.

We notice that while the ordering between interpretations is preserved, the plausibility ordering between events is not guaranteed to be preserved:

$$P(\{\omega_1, \omega_2\}) > P(\{\omega_1, \omega_3\}) \text{ whereas } \Pi(\{\omega_1, \omega_2\}) = \Pi(\{\omega_1, \omega_3\}) \tag{7.1}$$

ϕ	$P(\phi)$	$\Pi(\phi)$
$\{\omega_1\}$	0.6	1
$\{\omega_2\}$	0.3	0.4
$\{\omega_3\}$	0.1	0.1
$\{\omega_1, \omega_2\}$	0.9	1
$\{\omega_1, \omega_3\}$	0.7	1
$\{\omega_2, \omega_3\}$	0.4	0.4
$\{\omega_1, \omega_2, \omega_3\}$	1	1

Table 7.1 – Probability and possibility measure of events 2^Ω

Example 7.1 shows that the plausibility ordering between events is not preserved by the OT transformation. The same counter-example can be used to prove that the same goes for KT, ST and VT transformations.

One question that stands out is to know if there exists a probability-possibility transformation operation that preserves the principle of consistence, preserves the plausibility ordering between interpretations and preserves the plausibility ordering between events. Proposition 7.3 shows that it is impossible to obtain such probability-possibility transformation operation.

Proposition 7.3. *Let TR be a probability-possibility transformation. Then there exists a probability distribution p and $\pi = TR(p)$ where there exist $\phi \subseteq \Omega$, $\psi \subseteq \Omega$, with $\phi \neq \psi$ such that*

$$P(\phi) < P(\psi) \not\Rightarrow \Pi(\phi) < \Pi(\psi)$$

Since in probability theory we use the additivity axiom to compute $P(\phi)$ and in possibility theory we use the maxitivity axiom then the strict ordering between events is not always preserved. Example 7.2 provides a counter-example. Note that this result is valid in both the quantitative and the qualitative settings.

Example 7.2. Let p be a probability distribution on $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\pi = TR(p)$ and let α, β and γ be probabilities such that $\begin{cases} \alpha > \beta > \gamma \\ \alpha + \beta + \gamma & = 1 \end{cases}$

ω_i	$p(\omega_i)$	$\pi(\omega_i)$
ω_1	α	1
ω_2	β	β_1
ω_3	γ	γ_1

Table 7.2 – Example of probability-possibility transformation where the plausibility ordering between events is not preserved.

In this example, ω_1 is the most probable interpretation. Let ϕ and ψ be two events such that $\phi = \{\omega_1\}$ and $\psi = \{\omega_1, \omega_2\}$. Since ω_1 is included in the two events then $\Pi(\phi) = \Pi(\psi) = 1$. But $P(\phi) < P(\psi)$ because $P(\phi) = \alpha$ and $P(\psi) = \alpha + \beta$.

Now assume that $\beta + \gamma > \alpha$ and $\phi = \{\omega_1\}$ and $\psi = \{\omega_2, \omega_3\}$ then $P(\phi) < P(\psi)$ while $\Pi(\phi) > \Pi(\psi)$.

This example illustrates that if we have a strict order between two events in the probability distribution, the possibility measures of these events in the possibility distribution can be equal. It also shows that we can have $P(\phi) < P(\psi)$ but $\Pi(\phi) > \Pi(\psi)$. Indeed, the following proposition characterizes such a situation.

Proposition 7.4. *Let TR be a probability-possibility transformation. Let p be a probability distribution such as $p(\omega_{\max}) < \sum_{\omega_i \neq \omega_{\max}} p(\omega_i)$. Let π_{TR} be the possibility distribution obtained by transforming p by TR procedure and $\phi = \Omega \setminus \{\omega_{max}\}$. Then*

$$P(\phi) > p(\omega_{\max}) \Rightarrow \Pi_{TR}(\phi) < \pi_{TR}(\omega_{\max}) \tag{7.2}$$

Proof. Let p be a probability distribution over $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Assume that this distribution is such as:

$$\begin{cases} p(\omega_1) > p(\omega_2) > p(\omega_3) \\ p(\omega_1) < p(\omega_2) + p(\omega_3) \end{cases}$$

Table 7.3 depicts p and its transformed possibility distribution π_{TR}.

ω_i	$p(\omega_i)$	$\pi_{TR}(\omega_i)$
ω_1	0.4	1
ω_2	0.35	α
ω_3	0.25	β

Table 7.3 – Probability distribution where $p(\omega_1) < p(\omega_2) + p(\omega_3)$ and its transformed possibility distribution

TR procedure transforms p into π_{TR} by preserving the order of interpretations, thus $1 > \alpha > \beta$. Considering the events $\phi = \{\omega_2, \omega_3\}$ and $\psi = \{\omega_1\}$, we have $\Pi(\phi) < \Pi(\psi)$ and $P(\phi) > P(\psi)$. □

We therefore distinguish two different cases:

i) $\omega_{max} > \sum_{\omega_i \in \Omega,\ \omega_i \neq \omega_{max}} p(\omega_i)$ (*i.e.* $\omega_{max} > 0.5$), we lose the strict order.

ii) $\omega_{max} < 0.5$ then it exists at least one event ϕ such as $P(\phi) > P(\{\omega_{max}\})$ but $\Pi(\phi) < \Pi(\{\omega_{max}\})$

As said previously, this result is due to the additivity axiom against the maxitivity axiom. We are now interested in preserving the ordering between interpretations during the operation of marginalization. We can expect the result to be the same as this one since both probability and possibility marginalization operations are based on the same axioms than for the computations of $P(\phi)$ and $\Pi(\phi)$.

7.4 Preserving the ordering between interpretations after marginalization

This subsection analyzes the preservation of the ordering between interpretations during marginalization procedure. Let us first illustrate with OT transformation. Let p be a probability distribution over 3 binary variables A, B and C.

Table 7.4 represents the probability distribution and its transformed possibility distribution by OT, and Table 7.5 represents the marginalized distribution of Table 7.4 on variable C.

The transformation by OT of the distribution $p(A, B)$ of Table 7.5 is depicted by Table 7.6.

Note that the ordering between interpretations is not preserved in this example. Indeed, the operation of marginalization on C comes down to compute the probability measure of the events: $\{abc, ab\bar{c}\}$, $\{a\bar{b}c, a\bar{b}\bar{c}\}$, $\{\bar{a}bc, \bar{a}b\bar{c}\}$, $\{\bar{a}\bar{b}c, \bar{a}\bar{b}\bar{c}\}$. And we have seen with Proposition 7.3, that the order of events is not necessarily preserved. We can conclude that OT transformation does not preserve the ordering between interpretations after applying marginalization operation. This result applies also for KT, ST and VT transformations.

A	B	C	$p(A,B,C)$	$\pi_{OT}(A,B,C)$
a	b	c	0.09	0.292
a	b	$\overline{c}$	0.09	0.292
a	$\overline{b}$	c	0.21	0.712
a	$\overline{b}$	$\overline{c}$	0.21	0.712
$\overline{a}$	b	c	0.288	1
$\overline{a}$	b	$\overline{c}$	0.032	0.04
$\overline{a}$	$\overline{b}$	c	0.072	0.112
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.008	0.008

Table 7.4 – Probability distribution and possibility distribution transformed by OT

A	B	$p(A,B)$	$\pi_{OT}(A,B)$
a	b	0.18	0.292
a	$\overline{b}$	0.42	0.712
$\overline{a}$	b	0.32	1
$\overline{a}$	$\overline{b}$	0.08	0.112

Table 7.5 – Marginalization of Table 7.4 on variable C

A	B	$\pi_{OT}(A,B)$
a	b	0.26
a	$\overline{b}$	1
$\overline{a}$	b	0.58
$\overline{a}$	$\overline{b}$	0.08

Table 7.6 – Transformation of $p(A,B)$ of Table 7.5 by OT

Proposition 7.5. *Let TR be a probability-possibility transformation operation (or function). Then there exists a probability distribution p over $V = \{X_1, .., X_n\}$ and $\pi = TR(p)$. Let p' be the marginal distribution of p on the set $V' \subseteq \{X_1, .., X_n\}$ with its domain $D_{V'}$, π'' is the marginal distribution of π on X' and $\pi' = TR(p')$. Then there exist two values x_i, $x_j \in D_{V'}$ such that $i \neq j$,*

$$\pi''(x_i) < \pi''(x_j) \textit{ holds but } \pi'(x_i) < \pi'(x_j) \textit{ does not hold.}$$

It is easy to show that the two resulting distributions do not encode the same ordering, as it is shown in the following counter-example.

Counter-example 7.1. The following table gives an example of probability distribution and the resulting possibility distribution using TR transformation.

A	B	$p(A,B)$	$\pi(A,B)$
a	b	0.4	1
a	$\overline{b}$	0.1	α_3
$\overline{a}$	b	0.3	α_1
$\overline{a}$	$\overline{b}$	0.2	α_2

In this example $1 > \alpha_1 > \alpha_2 > \alpha_3$ (this is given by the preference preservation principle). Let π' be the possibility distribution obtained by applying marginalization then transformation TR and let π'' be the possibility distribution obtained by transformation TR and then application of marginalization operation. Then:

— $p(a) = 0.5$ and $p(\overline{a}) = 0.5$

— $\pi'(a) = 1$ and $\pi'(\overline{a}) = 1$

— $\pi''(a) = 1$ and $\pi''(\overline{a}) = \max(\alpha_1, \alpha_2)$

Consequently $\pi'(a) = \pi'(\overline{a})$ unlike $\pi''(a) > \pi''(\overline{a})$, which means that transforming using TR does not preserve the order. $\square$

The general observation about marginalization is that there is no transformation that can preserve the ordering between interpretations.

Given a new piece of information to be taken into account, the question is: Does conditioning before the transformation or after change the ordering in the resulted possibility distributions. More precisely, in the next section, we compare the ordering between interpretations of the conditioned distributions. This is done using both conditioning rules (quantitative and qualitative) but also under soft evidence (Jeffrey's conditioning rules).

7.5 Preserving the ordering between interpretations after a conditioning operation

Let ϕ be an evidence. Let p be a probability distribution. Let p' be the probability distribution p conditioned with ϕ. Let $\pi = TR(p)$ and $\pi' = TR(p')$ the associate possibility distributions of p and p' using TR transformation. Let π'' be the possibility distribution π conditioned with ϕ. In this section, we answer the following question: *Does π' encode the same ordering between interpretations than π''?* More formally, is the following equation valid?

$$\rhd(\pi') = \rhd(\pi'') \tag{7.3}$$

7.5.1 Quantitative possibilistic setting

Recall that conditioning in probability and conditioning in quantitative possibilistic setting with a certain piece of information $\phi \subseteq \Omega$, is done the same way. Since TR transformation preserves the order of interpretations, we can assume that the order in the conditioned probability distribution and the one after transformation will be the same.

Proposition 7.6. *Let $\phi \subseteq \Omega$ be an observation. Let p be a probability distribution and $\pi = TR(p)$. Let π'' be a distribution obtained by conditioning π (using the quantitative conditioning rule) with ϕ. Let π' be the distribution obtained by conditioning p (in the probabilistic setting) with ϕ then transforming using TR transformation.*

$$\forall \omega_i, \omega_j \in \Omega,\ \pi''(\omega_i \mid \phi) < \pi''(\omega_j \mid \phi) \Rightarrow \pi'(\omega_i \mid_p \phi) < \pi'(\omega_j \mid_p \phi) \tag{7.4}$$

The idea of this proposition is that if we first transform using a probability-possibility transformation TR then apply conditioning, then the ordering between interpretations $\omega_i \in \Omega$ will be the same than the one if we condition first and transform after.

Proof. Given the evidence $\phi \subseteq \Omega$. Let $\omega_i, \omega_j \in \Omega^2$ and $p(\omega_i) > p(\omega_j)$. If $\omega_i \in \phi$ and $\omega_j \in \Omega$, then

$$\frac{p(\omega_i)}{P(\phi)} > \frac{p(\omega_j)}{P(\phi)} \Rightarrow p(\omega_i \mid \phi) > p(\omega_j \mid \phi)$$

Since TR is a transformation that preserves the order of interpretations, we can state that $\pi'(\omega_i \mid \phi) > \pi'(\omega_j \mid \phi)$. On the other hand, we have $p(\omega_i) > p(\omega_j) \Rightarrow \pi(\omega_i) > \pi(\omega_j)$. And by conditioning with the quantitative rule with ϕ, we have:

$$\pi(\omega_i) > \pi(\omega_j) \Rightarrow \frac{\pi(\omega_i)}{\Pi(\phi)} > \frac{\pi(\omega_j)}{\Pi(\phi)} \Rightarrow \pi''(\omega_i \mid \phi) > \pi''(\omega_j \mid \phi)$$

The other cases $\omega_i \in \phi, \omega_j \notin \phi$ and $\omega_i \notin \phi, \omega_j \notin \phi$ are trivial. Indeed, by conditioning with ϕ, every interpretation that does not belong to the event ϕ are no longer taken into account. □

7.5.2 Qualitative possibilistic setting

Recall that qualitative conditioning is defined as:

$$\pi(\omega_i \mid_m \phi) = \begin{cases} 1 & \text{if } \pi(\omega_i) = \Pi(\phi) \text{ and } \omega_i \in \phi \\ \pi(\omega) & \text{if } \pi(\omega_i) < \Pi(\phi) \text{ and } \omega_i \in \phi \\ 0 & \text{otherwise.} \end{cases}$$

Qualitative conditioning can be seen as the following computation: $\min(\pi(\omega_i), \Pi(\phi))$ when the world ω_i belongs to the event ϕ. This conditioning also guarantees that the order is preserved. Therefore, Proposition 7.7 shows that the ordering encoded in the possibility distribution obtained from conditioning then transformation and the ordering encoded in the possibility distribution obtained from transformation then conditioning are the same.

Proposition 7.7. *Let $\phi \subseteq \Omega$ be an observation. Let p be a probability distribution and $\pi = TR(p)$. Let π'' be a distribution obtained by conditioning π (using the qualitative conditioning rule) with ϕ. Let π' be the distribution obtained by conditioning p (in the probabilistic setting) with ϕ then transforming using TR transformation.*

$$\forall \omega_i, \omega_j \in \Omega,\ \pi''(\omega_i \mid \phi) < \pi''(\omega_j \mid \phi) \Rightarrow \pi'(\omega_i \mid_m \phi) < \pi'(\omega_j \mid_m \phi) \tag{7.5}$$

Proof. Given the evidence $\phi \subseteq \Omega$. As in the quantitative setting, we discard every interpretation that does not belong to the event ϕ. Let $\omega_i, \omega_j \in \phi$ with $p(\omega_i) > p(\omega_j)$ then $p(\omega_i \mid \phi) > p(\omega_j \mid \phi)$. Since TR is a transformation that preserves the order, we state that $\pi'(\omega_i \mid \phi) > \pi'(\omega_j \mid \phi)$. On the other side, we have $p(\omega_i) > p(\omega_j) \Rightarrow \pi(\omega_i) > \pi(\omega_j)$. By conditioning with the qualitative rule with ϕ, we have:

$$\pi(\omega_i) > \pi(\omega_j) \Rightarrow \min(\pi(\omega_i), \Pi(\phi)) > \min(\pi(\omega_j), \Pi(\phi)) \Rightarrow \pi''(\omega_i \mid \phi) > \pi''(\omega_j \mid \phi)$$

Also, note that if $\omega_i \in \phi$ then $\pi(\omega_i) \leq \Pi(\phi)$. □

Since both conditioning rules (quantitative and qualitative) preserve the ordering between interpretations, and since probability-possibility transformations (seen in Chapter 3) satisfy Dubois and Prade principles, hence:

Corollary 7.1. Let $TR \in \{OT, KT, ST, VT\}$. Let π'' be a possibility distribution obtained with TR then conditioned with ϕ. Let π' be the possibility distribution obtained by conditioning (in the probabilistic setting) then transformed with TR.

$$\forall \phi \subseteq \Omega,\ argmax_{\omega_i \in \Omega}(\pi''(\omega_i \mid \phi)) = argmax_{\omega_i}(\pi'(\omega_i \mid \phi)) \tag{7.6}$$

7.5.3 Preserving conditioning under uncertain information

Let us focus now on conditioning with uncertain information. We need to compare the distribution built by conditioning the transformed possibility distribution and the distribution transformed from the conditioned probability one. This has never been investigated before. Therefore, we are interested in two problems, the first one is about finding if the two distributions are equivalent. The second issue is to know if the order of interpretations is the same in both conditioned distributions. Let us first start with an example.

A	B	$p(A,B)$
a	b	0.4
a	$\bar{b}$	0.1
$\bar{a}$	b	0.2
$\bar{a}$	$\bar{b}$	0.3

Table 7.7 – Probability distribution on A and B

A	B	$p'(A,B)$
a	b	0.56
a	$\bar{b}$	0.14
$\bar{a}$	b	0.12
$\bar{a}$	$\bar{b}$	0.18

Table 7.8 – Revised probability distribution of p

Example 7.3. Given two variables A and B, with domains $D_A = \{a, \bar{a}\}$ and $D_B = \{b, \bar{b}\}$. Let p (Table 7.7) be a probability distribution over A and B.

From Table 7.7, we have $P(\{A = a\}) = 0.5$. Now, let us consider the new information $P'(\{A = a\}) = 0.7$. Then the revised distribution of p with ϕ is given by Table 7.8. Now, let us transform these distributions with OT, this gives Tables 7.7 and 7.8.

A	B	$\pi_{OT}(A,B)$
a	b	1
a	$\bar{b}$	0.1
$\bar{a}$	b	0.3
$\bar{a}$	$\bar{b}$	0.6

Table 7.9 – Transformed possibility distribution of p with OT

A	B	$\pi'(A,B)$
a	b	1
a	$\bar{b}$	0.26
$\bar{a}$	b	0.12
$\bar{a}$	$\bar{b}$	0.44

Table 7.10 – Transformed possibility distribution of p' with OT

In order to compare the revised distributions, we need to revise in the possibilistic setting (here we have chosen the quantitative conditioning rule). In the revised Table 7.11, the conditioning operation is done using the possibility distribution $\Pi'(\{A = a\}) = 1$ and $\Pi'(\{A = \bar{a}\}) = 0.3$ (*i.e.* the transformed distribution by OT of the probability distribution used to condition in the probabilistic setting).

A	B	$\pi''(A,B)$
a	b	1
a	$\bar{b}$	0.1
$\bar{a}$	b	0.15
$\bar{a}$	$\bar{b}$	0.3

Table 7.11 – Possibility distribution π_{OT} revised using Equation (1.30)

From Tables 7.10 and 7.11 we note that $\pi''(a\bar{b}) > \pi''(\bar{a}b)$ whereas in $\pi'(a\bar{b}) < \pi'(\bar{a}b)$. We conclude that when a new information is not precise, then using OT transformation procedure does not guarantee that the ordering between interpretations will be the same.

More precisely, let OT be the optimal probability-possibility transformation defined in Subsection 3.1.2. Then there exists a probability distribution p and $\pi = OT(p)$ such that there exist two values x_i, $x_j \in \Omega$, with $i \neq j$, and

$$\pi''(x_i) < \pi''(x_j) \text{ holds but } \pi'(x_i) < \pi'(x_j) \text{ does not hold.}$$

Where π'' is a distribution obtained with OT then revised (using Jeffrey's rule of conditioning of Equation (1.30)). And π' is the distribution revised then transformed with OT.

We have shown for OT transformation which is the optimal transformation. We now investigate the case of any transformation TR satisfying Dubois and Prade principles. More precisely, let TR be a probability-possibility transformation operation (or function). Then there exists a probability distribution p and $\pi = TR(p)$ such that there exist two values $x_i, x_j \in \Omega$ with $i \neq j$, and

$$\pi''(x_i|\phi) < \pi''(x_j|\phi) \text{ holds but } \pi'(x_i|\phi) < \pi'(x_j|\phi) \text{ does not hold.}$$

Where p' is the probability distribution revised with the imperfect information ϕ. And $\pi' = TR(p')$ is the possibility distribution transformed by TR, and π'' is a possibility distribution π revised with ϕ.

Proof. The aim is, here, to prove that conditioning with uncertain information then transforming gives neither the same distributions, nor encodes the same ordering between interpretations than transforming first and then revising. The above proposition holds in both quantitative and qualitative case. First, let us tackle the quantitative case.

Consider the following probability distribution of Table 7.12, with $P(a) = \alpha_1 + \alpha_2$ and $P(\overline{a}) = \alpha_2 + \alpha_2$ and $\alpha_1 > \alpha_2$, We want to revise our distribution $p(A, B)$ by taking into account that $p(A)$ now depicts total ignorance. The distribution $p'(A, B)$ of Table 7.12 defines the conditioning of $p(A, B)$ by the equiprobable distribution $P'(a) = 0.5$ and $P'(\overline{a}) = 0.5$.

A	B	$p(A,B)$	$p'(A,B)$
a	b	α_1	$\frac{\alpha_1}{\alpha_1+\alpha_2} \times 0.5$
a	$\overline{b}$	α_2	$\frac{\alpha_2}{\alpha_1+\alpha_2} \times 0.5$
$\overline{a}$	b	α_2	0.25
$\overline{a}$	$\overline{b}$	α_2	0.25

Table 7.12 – Probability distribution and its revised with uncertain information

A	B	$\pi_{TR}(A,B)$
a	b	1
a	$\overline{b}$	β
$\overline{a}$	b	β
$\overline{a}$	$\overline{b}$	β

Table 7.13 – Transformed distribution of Table 7.12 by TR

Table 7.14 depicts the transformed distribution of Table 7.12 and Table 7.15 depicts the revised distribution of $\pi(A, B)$ of Table 7.13. We revise Table 7.13 using the possibility distribution $\Pi'(a) = 1$ and $\Pi'(\overline{a}) = 1$. From Table 7.14 we have $\gamma_1 < \gamma_2$.

A	B	$\pi'(A,B)$
a	b	1
a	$\overline{b}$	γ_1
$\overline{a}$	b	γ_2
$\overline{a}$	$\overline{b}$	γ_2

Table 7.14 – Probability distribution $p'(A, B)$ of Table 7.12 transformed by any transformation TR

A	B	$\pi''(A,B)$
a	b	1
a	$\overline{b}$	β
$\overline{a}$	b	1
$\overline{a}$	$\overline{b}$	1

Table 7.15 – Possibility distribution of Table 7.13 conditioned using (1.30)

Without necessarily knowing the ordering between interpretations represented by unknown variables, we can still notice that the order is not preserved given that in Table 7.15, three interpretations are at the same level ($\pi'(ab) = \pi'(\overline{a}b) = \pi'(\overline{a}\overline{b}) = 1$). And in Table 7.14, only one interpretation gives a possibility degree of 1.

Let us show the same for the qualitative setting, meaning that instead of conditioning with Equation (1.30), we revise with Equation (1.32), which gives us Table 7.16. Let us note that it is exactly the

A	B	$\pi'(A,B)$
a	b	1
a	$\bar{b}$	β
$\bar{a}$	b	1
$\bar{a}$	$\bar{b}$	1

Table 7.16 – Possibility distribution of Table 7.13 revised using Equation (1.32)

same table as the one we have revised in the quantitative setting (Table 7.15), which allows us to derive the same conclusions as in the quantitative setting.

To summarize, conditioning with uncertain information (in quantitative and qualitative setting) does not guarantee to preserve the ordering between interpretations and therefore does not give the same distributions. □

So far, we analyzed the preservation of the ordering between interpretations with respect to marginalization and conditioning. This two different operations are indeed necessary to compute certain types of queries like *MPE* and *MAP* queries. We now investigate the consequences of our first results on the preservation of the results of *MPE* and *MAP* queries.

7.6 Preserving *MPE* and *MAP* queries

The results expected when dealing with *MPE* queries are not degrees but configurations having a certainty degree. Therefore, the question we need to answer is whether the result of an *MPE* query derived from a probability distribution is the same as the one derived from the transformed possibility distribution.

Proposition 7.8. *Let p be a probability distribution and π_{TR} be a transformation preserving the order of interpretations. Then*

$$\forall\, \phi \subseteq \Omega,\; argmax_{\omega_i \in \Omega}(p(\omega_i|\phi)) = \; argmax_{\omega_i \in \Omega}(\pi(\omega_i|\phi)) \tag{7.7}$$

Proof. The proof of Proposition 7.8 is quite trivial. Let us consider the order over $\Omega = \{\omega_1, ..., \omega_n\}$ as the following: $p(\omega_1) > p(\omega_2) > ... > p(\omega_n)$. If the transformation TR preserves the ordering between interpretations, then: $\pi(\omega_1) > \pi(\omega_2) > ... > \pi(\omega_n)$.

Computing $argmax_{\omega_i \in \Omega}(p(\omega_i|\phi))$, comes down to computing the configuration having the highest probability among the configurations belonging to the event ϕ. Let $\omega_j = argmax_{\omega_i \in \Omega}(p(\omega_i, \phi))$ then

$$\forall\, \omega_i \in \phi, \text{ such as } \omega_j \neq \omega_i,\; p(\omega_j) > p(\omega_i)$$

Since the order of interpretations is preserved then,

$$\forall\, \omega_i \in \phi, \text{ such as } \omega_j \neq \omega_i,\; \pi_{TR}(\omega_j) > \pi_{TR}(\omega_i)$$

Thus, $\omega_j = argmax_{\omega_i \in \Omega}(\pi_{TR}(\omega_i, \phi))$. The answer of the *MPE* query is the same in both distributions. □

As the probability-possibility transformations OT, KT, ST and VT preserve the ordering between interpretations, we have the following corollary.

Corollary 7.2. Let $TR \in \{OT, KT, ST, VT\}$. Let p be a probability distribution and π_{TR} be the possibility distribution resulting using TR.

$$\forall\, \phi \subseteq \Omega,\; argmax_{\omega_i \in \Omega}(p(\omega_i, \phi)) =\; argmax_{\omega_i \in \Omega}(\pi_{TR}(\omega_i, \phi))$$

Let us consider *MAP* queries that search for the most plausible configuration of a subset of variables given an evidence.

Given the following universe of discourse $\Omega = D_{X_1} *...* D_{X_n}$ (with the variables $V = \{X_1, ..., X_n\}$), let $V' \subseteq V$ be a subset of variables.

Proposition 7.9. *Let TR be a probability-possibility distribution. There exists a probability distribution p and $\pi_{TR} = TR(p)$ such that there exists ϕ a sure piece of information and*

$$argmax_{x' \in D_{V'}}(p(x' \mid \phi)) \neq argmax_{x' \in D_{V'}}(\pi_{TR}(x' \mid \phi)) \tag{7.8}$$

Proof. The proof of this proposition is easy. Using Proposition 7.5, since marginalization does not always preserve the ordering between interpretations, then it is not guaranteed that the interpretation having the maximum degree given the evidence will be the same in the probability distribution and in the possibility distribution. □

We have seen that in the case where the evidence is sure then the ordering between interpretations is preserved during conditioning, however marginalization cannot ensure the same ordering. Therefore, the following corollary states that:

Corollary 7.3. Let $TR \in \{OT, KT, ST, VT\}$. There exists p a probability distribution and π_{TR} is the possibility distribution transformed with TR such that there exists ϕ be a sure new piece of information and

$$argmax_{x' \in D_{V'}}(p(x' \mid \phi)) \neq argmax_{x' \in D_{V'}}(\pi_{TR}(x' \mid \phi))$$

In the case where the observation is no longer a sure piece of information, and that we deal with Jeffrey rule of conditioning, we have seen that the ordering between interpretation is not always preserved, adding to it the result of the ordering between interpretations with respect to marginalization, we have:

Corollary 7.4. Let $TR \in \{OT, KT, ST, VT\}$. There exists p a probability distribution and $\pi_{TR} = TR(p)$ such that there exists ϕ a uncertain piece of information and

$$argmax_{x' \in D_{V'}}(p(x' \mid \phi)) \neq argmax_{x' \in D_{V'}}(\pi_{TR}(x' \mid \phi))$$

This first part of the chapter analyzes the commutativity of reasoning tasks with respect to transformations from the literature but also for any probability-possibility transformation preserving Dubois and Prade consistency principles. In particular, we were interested in the ordering between interpretations. In the following, we analyze properties preserved by transformations in graphical models. Indeed, graphical models compactly encode distributions using the notion of independence relations and we first study the preservation of independence relations in distributions. The second part concerns the preservation of the ordering between interpretations in the joint distribution before and after the transformation of the network.

7.7 Preserving independence relations

The first sections of this chapter clearly showed that transformations may induce a loss of information when moving from a probability distribution to a possibility distribution. We are now interested in independence relations. This notion is important as it plays a key role in the building of graphical models. As a first step, let us focus on existing probability-possibility transformation procedures (OT, KT, ST and VT).

7.7.1 OT transformation

Let I_p (resp. I_π) denotes the set of independence relations described in p (resp. π). Let OT be the probability-possibility distribution. Then there exists p a probability distribution and $\pi_{OT} = OT(p)$ such that there exists an independence relation I and

$$I \in I_p \text{ but } I \notin I_\pi \tag{7.9}$$

The following counter-example shows that OT transformation is not guaranteed to preserve independence relations.

Counter-example 7.2. From Table 7.17, we have $B \perp C|A$. Tables 7.18 and 7.19 depict $\Pi_{OT}(B|AC)$ and $\Pi_{OT}(B|A)$. We need to verify that π describes also $B \perp C|A$.

A	B	C	$p(A,B,C)$	$\pi_{OT}(A,B,C)$
a	b	c	0.09	0.292
a	b	$\overline{c}$	0.09	0.292
a	$\overline{b}$	c	0.21	0.712
a	$\overline{b}$	$\overline{c}$	0.21	0.712
$\overline{a}$	b	c	0.288	1
$\overline{a}$	b	$\overline{c}$	0.032	0.04
$\overline{a}$	$\overline{b}$	c	0.072	0.112
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.008	0.008

Table 7.17 – Probability distribution and its transformed possibility distribution by OT

A	B	C	$\pi_{OT}(B\|_p AC)$	$\pi_{OT}(B\|_m AC)$
a	b	c	0.41	0.292
a	b	$\overline{c}$	0.41	0.292
a	$\overline{b}$	c	1	1
a	$\overline{b}$	$\overline{c}$	1	1
$\overline{a}$	b	c	1	1
$\overline{a}$	b	$\overline{c}$	1	1
$\overline{a}$	$\overline{b}$	c	0.112	0.112
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.2	0.008

Table 7.18 – Possibility distribution of B given AC

A	B	$\pi_{OT}(B\|_p A)$	$\pi_{OT}(B\|_m A)$
a	b	0.41	0.292
a	$\overline{b}$	1	1
$\overline{a}$	b	1	1
$\overline{a}$	$\overline{b}$	0.112	0.112

Table 7.19 – Possibility distribution of B given A

Tables 7.18 and 7.19 show that B and C are no longer independent given A (nor using the quantitative conditioning, nor the quantitative conditioning). Indeed, in the quantitative setting:

$$\left.\begin{array}{r}\pi(\bar{b}|_p\overline{ac}) = 0.2\\ \pi(\bar{b}|_p\bar{a}) = 0.112\end{array}\right\} \text{therefore } \pi(\bar{b}|_p\overline{ac}) \neq \pi(\bar{b}|_p\bar{a})$$

And in the qualitative setting:

$$\left.\begin{array}{r}\pi(\bar{b}|_m\overline{ac}) = 0.008\\ \pi(\bar{b}|_m\bar{a}) = 0.112\end{array}\right\} \text{therefore } \pi(\bar{b}|_m\overline{ac}) \neq \pi(\bar{b}|_m\bar{a})$$

Therefore $\pi(\bar{b}|\overline{ac}) \neq \pi(\bar{b}|\bar{a})$, meaning that $B \not\perp C \mid A$. □

We now provide the same analysis for KT, ST and VT.

7.7.2 KT, ST and VT transformations

Let I_p (resp. I_π) denotes the set of independence relations described in p (resp. π). Let $TR \in \{KT, ST, VT\}$ be the probability-possibility distribution. Then there exists p a probability distribution and $\pi_{TR} = TR(p)$ with $TR \in \{KT, ST, VT\}$ such that there exists an independence relation I and

$$I \in I_p \text{ but } I \notin I_\pi \tag{7.10}$$

As for OT, we show that there exists a probability distribution p such that transforming using KT may induce the loss of some independence relations.

Counter-example 7.3. Table 7.20 depicts the probability distribution and its transformation by KT.

A	B	C	$p(A,B,C)$	$\pi_{KT}(A,B,C)$
a	b	c	0.054	0.2
a	b	$\bar{c}$	0.032	0.12
a	$\bar{b}$	c	0.006	0.02
a	$\bar{b}$	$\bar{c}$	0.008	0.03
$\bar{a}$	b	c	0.27	1
$\bar{a}$	b	$\bar{c}$	0.144	0.53
$\bar{a}$	$\bar{b}$	c	0.27	1
$\bar{a}$	$\bar{b}$	$\bar{c}$	0.216	0.8

Table 7.20 – Probability distribution and its transformed possibility distribution by KT

The independence relation described in the above table is: $A \perp C$. Using Tables 7.21 and 7.22 we can clearly see that in π_{KT} we do not have $A \perp C$.

Indeed, in the quantitative setting we have:

$$\left.\begin{array}{r}\pi(a|_pc) = 0.148\\ \pi(a) = 0.2\end{array}\right\} \text{therefore } \pi(a|_pc) \neq \pi(a)$$

And in the qualitative setting:

$$\left.\begin{array}{r}\pi(a|_mc) = 0.12\\ \pi(a) = 0.2\end{array}\right\} \text{therefore } \pi(a|_mc) \neq \pi(a)$$

Therefore $\pi(a|_mc) \neq \pi(a)$ and $A \not\perp C$. □

A	C	$\pi_{KT}(A\|_p C)$	$\pi_{KT}(A\|_m C)$
a	c	0.2	0.2
a	$\overline{c}$	0.148	0.12
$\overline{a}$	c	1	1
$\overline{a}$	$\overline{c}$	1	1

Table 7.21 – Possibility distribution of A given C

A	$\pi_{KT}(A)$
a	0.2
$\overline{a}$	1

Table 7.22 – Possibility distribution of A

In the following counter-example, we show that ST do not necessarily preserve all of the independence relations.

Counter-example 7.4. In the probability distribution of Table 7.23, $B \perp C | A$. Let us check that by transforming with ST, we preserve the independence relation.

A	B	C	$p(A,B,C)$	$\pi_{ST}(A,B,C)$
a	b	c	0.09	0.562
a	b	$\overline{c}$	0.09	0.562
a	$\overline{b}$	c	0.21	0.922
a	$\overline{b}$	$\overline{c}$	0.21	0.922
$\overline{a}$	b	c	0.288	1
$\overline{a}$	b	$\overline{c}$	0.032	0.232
$\overline{a}$	$\overline{b}$	c	0.072	0.472
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.008	0.064

Table 7.23 – Probability distribution and its transformed possibility distribution by ST

The computations in the quantitative setting show that:

$$\left.\begin{array}{r}\pi(\overline{b}|_p\overline{a}\overline{c}) = 0.28 \\ \pi(\overline{b}|_p\overline{a}) = 0.472\end{array}\right\} \text{therefore } \pi(\overline{b}|_p\overline{a}\overline{c}) \neq \pi(\overline{b}|_p\overline{a})$$

And in the qualitative case:

$$\left.\begin{array}{r}\pi(\overline{b}|_m\overline{a}\overline{c}) = 0.064 \\ \pi(\overline{b}|_m\overline{a}) = 0.472\end{array}\right\} \text{therefore } \pi(\overline{b}|_m\overline{a}\overline{c}) \neq \pi(\overline{b}|_m\overline{a})$$

Therefore $\pi(\overline{b}|\overline{a}\overline{c}) \neq \pi(\overline{b}|\overline{a})$, and $B \not\perp C \mid A$. □

Lastly, we transform with VT transformation and check if the independence relations are preserved in the resulting possibility distribution.

Counter-example 7.5. Let A, B and C be 3 boolean variables. The set Ω is the Cartesian product of their domains. The constant k used in VT transformation is defined as [MSMR06]:

$$0 \leq k \leq \frac{log(p_n)}{(1-p_n).log(\frac{p_n}{p_1})}$$

The constant k should be chosen between 0 and $\frac{log(p_n)}{(1-p_n).log(\frac{p_n}{p_1})} \approx 1.35$, here we have chosen for the counter-example $k = 1$ in order to simplify the computations.

A	B	C	$p(A,B,C)$	$\pi_{VT}(A,B,C)$
a	b	c	0.09	0.35
a	b	$\overline{c}$	0.09	0.35
a	$\overline{b}$	c	0.21	0.78
a	$\overline{b}$	$\overline{c}$	0.21	0.78
$\overline{a}$	b	c	0.288	1
$\overline{a}$	b	$\overline{c}$	0.032	0.12
$\overline{a}$	$\overline{b}$	c	0.072	0.28
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.008	0.03

Table 7.24 – Probability distribution and its transformed possibility distribution by VT

It is important to note that it does not exist a k that ensures the preservation of all independence relations. Table 7.24 entails the following independence relation (in the probabilistic setting): $B \perp C|A$. Let us verify that after the transformation, we retrieve the same independence, in both quantitative and qualitative case. Therefore we compute $\Pi_{VT}(B|AC)$ and $\Pi_{VT}(B|A)$.

In the quantitative setting:

$$\left.\begin{array}{r}\pi(\overline{b}|_p\overline{a}\overline{c}) = 0.24\\ \pi(\overline{b}|_p\overline{a}) = 0.28\end{array}\right\} \text{therefore } \pi(\overline{b}|_p\overline{a}\overline{c}) \neq \pi(\overline{b}|_p\overline{a})$$

And in the qualitative setting:

$$\left.\begin{array}{r}\pi(\overline{b}|_m\overline{a}\overline{c}) = 0.03\\ \pi(\overline{b}|_m\overline{a}) = 0.28\end{array}\right\} \text{therefore } \pi(\overline{b}|_m\overline{a}\overline{c}) \neq \pi(\overline{b}|_m\overline{a})$$

Thus $\pi(\overline{b}|\overline{a}\overline{c}) \neq \pi(\overline{b}|\overline{a})$, and $B \not\perp C \mid A$. □

We have shown that transformations OT, KT, ST and VT are not guaranteed to preserve all of the independence relations. *Can it exist a transformation that preserves all of the independence relations by going from p to π ?* The following section deals with these questions.

7.7.3 Preserving independence relations in the general case

We have shown in Section 7.3 that the plausibility ordering between events is not always preserved using any probability-possibility transformation. We are now interested in preserving independence relations.

Preserving independence relations in the qualitative case

We first study this matter in the qualitative case.

Proposition 7.10. *Let TR be a probability-possibility transformation. There exists p a probability distribution and $\pi_{TR} = TR(p)$ where one can have three events ϕ, ψ and $\alpha \subseteq \Omega$,*

$$P(\phi \mid \psi \cup \alpha) = P(\phi \mid \alpha) \textit{ holds but } \Pi_{TR}(\phi \mid_{min} \psi \cup \alpha) = \Pi_{TR}(\phi \mid_{min} \alpha) \textit{ does not hold} \tag{7.11}$$

Proof. Let a probability distribution p describing A and B as independent. And given the following order $1 > \alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > 0$ with $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$.

Since the transformation preserves the ordering between interpretations, we have $1 > \beta_2 > \beta_3 > \beta_4$. To retrieve the independence between A and B, we should have $\Pi(a|_{min}b) = \Pi(a)$ *i.e.* $\beta_4 = \beta_3$ yet $\beta_4 < \beta_3$ so A and B are not independent in π_{TR}. □

A	B	$p(A,B)$	$\pi(A,B)$
a	b	α_4	$TR(\alpha_4) = \beta_4$
a	$\bar{b}$	α_3	$TR(\alpha_3) = \beta_3$
$\bar{a}$	b	α_2	$TR(\alpha_2) = \beta_2$
$\bar{a}$	$\bar{b}$	α_1	$TR(\alpha_1) = 1$

Table 7.25 – Probability distribution and its transformed possibility distribution

In general, for the qualitative setting, we just proved that no probability-possibility transformation can ensure that all of the independence relations between events described in the probabilistic setting are preserved in the possibilistic setting. However, we are able to characterize particular cases where the independence relations will be preserved.

Indeed, we have noticed that when interpretations having the maximum probability degree belong to the intersection of events, the independence relation is preserved. This is illustrated by Figure 7.2.

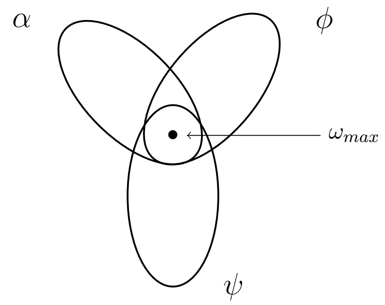


Figure 7.2 – Representation of the events ϕ, ψ and α where $\omega_{max} \in \phi \cap \psi \cap \alpha$

This allows us to write Proposition 7.11.

Proposition 7.11. *Let p be a probability distribution and $\pi_{TR} = TR(p)$. Let ϕ, ψ and $\alpha \subseteq \Omega$ be three events such that $\phi \perp \psi \mid \alpha$ is true in p. Let ω_{max} be the most probable world in $\phi \cup \psi \cup \alpha$.*

$$\text{If } \omega_{max} \in \alpha \cap \psi \cap \phi \ \text{ then } \Pi_{TR}(\phi \mid \psi \cup \alpha) = \Pi_{TR}(\phi \mid \alpha) \tag{7.12}$$

Proof. Let p be a probability distribution. Let Ω be the universe of discourse and given the 3 events ϕ, ψ and $\alpha \subseteq \Omega$.

Given that $\phi \perp \psi \mid \alpha$ in p, and that π_{TR} is obtained by transforming p using TR which preserves the order of interpretations, we already showed that the following equality is not always verified.

$$\Pi(\phi \mid \psi\alpha) = \Pi(\phi \mid \alpha)$$

However, let ω_{max} be the most probable (and therefore the most possible) in $\phi \cup \psi \cup \alpha$, then if $\omega_{max} \in \phi \cap \psi \cap \alpha$:

$$\Pi(\phi \mid \psi\alpha) = \Pi(\phi \mid \alpha) \ \text{ since } \min(\pi(\omega_{max}), \pi(\omega_{max})) = \min(\pi(\omega_{max}), \pi(\omega_{max}))$$

In this particular case, in both quantitative and qualitative possibilistic settings, we can state that the independence relations between events are preserved given any transformation. □

Consequences on the preservation of independence relations between variables

We have just shown, in the qualitative case in particular, that the independence relations between events are not always preserved. The following Corollary 7.5 therefore states that the same result can be applied to variables.

Corollary 7.5. Since the independence relations between events are not necessarily preserved, then the independence relations between variables is not always preserved.

Proof sketch. Checking that two (or more) variables are independent comes down to check the independence relations of events where an event is the set of values of the variable's domain. Since the independence relations of events is not always preserved, then the independence relations between variables is not always preserved either. □

Our work on preserving independence relations between variables showed that when transforming from probability to possibility we are not guaranteed to keep the independence relations. But we noticed that probability-possibility transformations can also create independence relations. Let us see an example with KT transformation.

Example 7.4. Table 7.26 depicts a probability distribution and its transformation by KT.

A	B	C	$p(A,B,C)$	$\pi_{KT}(A,B,C)$
a	b	c	0.012	0.036
a	b	$\bar{c}$	0.096	0.286
a	$\bar{b}$	c	0.048	0.143
a	$\bar{b}$	$\bar{c}$	0.144	0.429
$\bar{a}$	b	c	0.07	0.208
$\bar{a}$	b	$\bar{c}$	0.224	0.667
$\bar{a}$	$\bar{b}$	c	0.07	0.208
$\bar{a}$	$\bar{b}$	$\bar{c}$	0.336	1

Table 7.26 – Probability distribution and its transformation by KT

From p we have the independence relation $A \perp C$. Considering Tables 7.27 and 7.28 we can clearly see the loss of the independence relation between A and C. But we also find a new independence relation in the transformed distribution π_{KT}, indeed we have $A \perp B$.

A	B	$\pi_{KT}(A\|_p B)$
a	b	0.429
a	$\bar{b}$	0.429
$\bar{a}$	b	1
$\bar{a}$	$\bar{b}$	1

Table 7.27 – Possibility distribution of A given B

A	$\pi_{KT}(A)$
a	0.429
$\bar{a}$	1

Table 7.28 – Possibility distribution of A

These results on preserving independence relations are easily overcome if one applies transformations to graphical models. This is the purpose of the next section.

7.8 Transforming a Bayesian network into a possibilistic network

This section first recalls how to transform a graphical model from one setting to another. Then we study the ordering between interpretations in the joint distributions.

7.8.1 Definition of graphical transformation

The definition of a transformation in graphical models is quite simple and natural. The goal is to keep the same structure of the graph and to transform the local distributions. Hence, graphical models transformation is defined as:

Definition 7.1 (Graphical models transformation)**.** Let $\mathcal{N}$ a network with G is graph structure, then $\mathcal{N}_{TR}$ is a transformed network by TR composed of:

— a graphical component which is the same graph structure G as $\mathcal{N}$

— a numerical component which is the set of local distributions of $\mathcal{N}$ transformed by TR

For instance, we can adapt this definition to probability-possibility transformation or even from imprecise probability to possibility transformation by defining transformation from Bayesian networks to possibilistic networks or respectively from credal networks to possibilistic networks. In this chapter we focus on transformation from Bayesian networks to possibilistic networks.

A direct result from graphical models transformations is that it ensure the preservation of independence relations between variables. Let us denote $I_{\mathcal{BN}}$ the set of independence relations in a Bayesian network $\mathcal{BN}$ and $I_{\mathcal{PN}_{TR}}$ the set of independence relations of the possibilistic network transformed by TR.

Proposition 7.12. *Let $\mathcal{BN}$ be a Bayesian network and $\mathcal{PN}_{TR}$ be the possibilistic network transformed by TR. Then,*

$$\forall I \in I_{\mathcal{BN}}, \;\; I \in I_{\mathcal{PN}_{TR}} \tag{7.13}$$

This is easily proved since the graph structure is the same and thus encodes the same independence relations. Another advantage of graphical models transformations is computationally it is less consuming to transform a set of local tables than a whole joint distribution.

The problem now is that there is no guarantee that the order of interpretations and events is preserved in the obtained possibilistic network and its underlying joint distribution. Figure 7.3 illustrates the issue preserving the ordering between the interpretations in the joint distributions when transforming a Bayesian network into a possibilistic one.

For the sake of clarity we will denote π'' the joint distribution from the Bayesian network $\mathcal{BN}$ transformed by TR, and π'_* (resp. π'_m) the joint distribution of the possibilistic network $\mathcal{PN}_{TR}$ transformed from the Bayesian network obtained using the quantitative (resp. qualitative) chain rule.

We start by checking with the existing transformation OT.

Example 7.5. Let us consider the Bayesian network over 3 boolean variables of Figure 7.4 with the associated local distributions transformed by OT transformation.

Using the possibilistic chain rules in both setting, the joint distributions, of Figure 7.4, are depicted in Table 7.29.

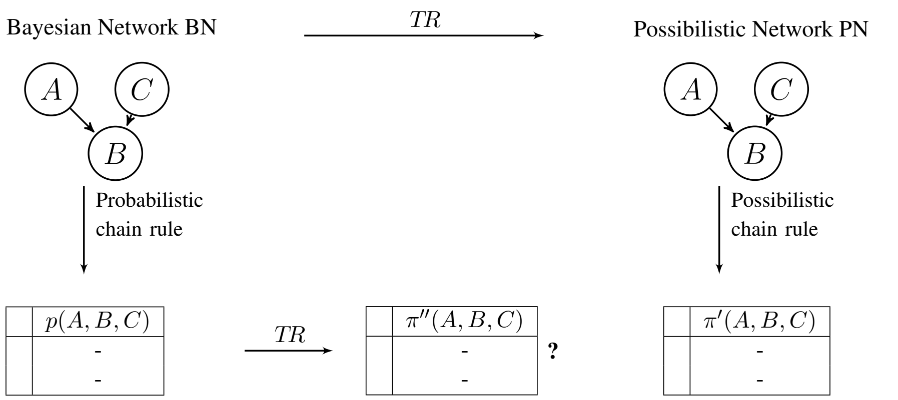
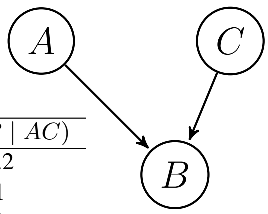


Figure 7.3 – Belief graphical models transformation

A	$p(A)$	$\pi_{OT}(A)$
a	0.1	0.1
$\overline{a}$	0.9	1



C	$p(C)$	$\pi_{OT}(C)$
c	0.6	1
$\overline{c}$	0.4	0.4

B	A	C	$p(B \mid AC)$	$\pi_{OT}(B \mid AC)$
b	a	c	0.2	0.2
$\overline{b}$	a	c	0.8	1
b	a	$\overline{c}$	0.7	1
$\overline{b}$	a	$\overline{c}$	0.3	0.3
b	$\overline{a}$	c	0.5	1
$\overline{b}$	$\overline{a}$	c	0.5	1
b	$\overline{a}$	$\overline{c}$	0.4	0.4
$\overline{b}$	$\overline{a}$	$\overline{c}$	0.6	1

Figure 7.4 – Bayesian network and its associated possibilistic network using OT

A	B	C	$p(A, B, C)$	$\pi''(A, B, C)$	$\pi'_*(A, B, C)$	$\pi'_m(A, B, C)$
$\overline{a}$	$\overline{b}$	$\overline{c}$	0.216	0.46	0.4	0.4
$\overline{a}$	$\overline{b}$	c	0.27	1	1	1
$\overline{a}$	b	$\overline{c}$	0.144	0.244	0.16	0.4
$\overline{a}$	b	c	0.27	1	1	1
a	$\overline{b}$	$\overline{c}$	0.012	0.024	0.012	0.1
a	$\overline{b}$	c	0.048	0.1	0.1	0.1
a	b	$\overline{c}$	0.028	0.052	0.04	0.1
a	b	c	0.012	0.024	0.02	0.1

Table 7.29 – Comparison of joint distributions derived from Figure 7.4

One problem here is that the order of interpretation is modified. For instance, we have $\pi''(a\bar{b}\bar{c}) = \pi''(abc)$ whereas in the distribution π'_* computed from the chain rule (2.5), $\pi'_*(a\bar{b}\bar{c}) < \pi'_*(abc)$. In the qualitative setting, we have $\pi''(a\bar{b}c) > \pi''(ab\bar{c})$ whereas $\pi'_m(a\bar{b}c) = \pi'_m(ab\bar{c})$.

We now consider TR a probability-possibility transformation that satisfies consistency principle of Dubois and Prade. We are interested in the ordering induced in the joint distributions by the Bayesian network and by the possibilistic network transformed using TR procedure. First, let us deal with the qualitative case.

7.8.2 Preserving joint distributions in the qualitative case

Let us now check if the ordering between interpretations induced by $p_{\mathcal{BN}}$ (the joint distribution encoded by the Bayesian network $\mathcal{BN}$) is preserved in the obtained joint possibility distribution $\pi_{\mathcal{PN}}$ (the joint distribution encoded by the possibilistic network $\mathcal{PN}$). Proposition 7.14 answers this question.

Proposition 7.13. *Let TR be a probability-possibility transformation. Then there exist a Bayesian network $\mathcal{BN}$, $\omega_1 \in \Omega$ and $\omega_2 \in \Omega$ where:*

$$\pi''(\omega_1) < \pi''(\omega_2) \text{ does not imply } \pi'(\omega_1) < \pi'(\omega_2) \tag{7.14}$$

with i) $\pi''(\omega) = TR(p(\omega))$ and p is the joint distribution induced by $\mathcal{BN}$ and ii) π' is the joint distribution induced by $\mathcal{PN}$ using Definition 7.1.

Proof. Given a Bayesian over two variables A and B. Here, $A \perp B$ and A and B have the same distribution (here, $\alpha_1 > \alpha_2$).

A	$p(A)$
a	α_1
$\bar{a}$	α_2

B	$p(B)$
b	α_1
$\bar{b}$	α_2

The joint distribution $p(A, B)$ of the Bayesian network $\mathcal{BN}$ is given by the following table.

A	B	$p(A,B)$	$\pi''(A,B)$
a	b	α_1^2	γ_1
a	$\bar{b}$	$\alpha_1 * \alpha_2$	γ_2
$\bar{a}$	b	$\alpha_1 * \alpha_2$	γ_2
$\bar{a}$	$\bar{b}$	α_2^2	γ_3

A	B	$\pi'_m(A,B)$
a	b	1
a	$\bar{b}$	β
$\bar{a}$	b	β
$\bar{a}$	$\bar{b}$	β

Table 7.30 – The joint distributions derived from the Bayesian network and the associated possibilistic network using TR

Clearly, $\pi''(ab) > \pi''(a\bar{b}) = \pi''(\bar{a}b) > \pi''(\bar{a}\bar{b})$. And from π'_m, $\pi'_m(ab) > \pi'_m(a\bar{b}) = \pi'_m(\bar{a}b) = \pi'_m(\bar{a}\bar{b})$.

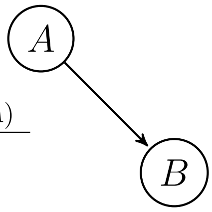
On this example, whatever the probability-possibility transformation used, we may lose the strict order in the joint distribution after the transformation of the Bayesian network. □

This proof states that the strict order is not guaranteed to be preserved. The question now is to know *given an order over π'' such as $\pi''(\omega_i) > \pi''(\omega_j)$, can we have $\pi'_m(\omega_i) < \pi'_m(\omega_j)$?* Let us take a look at a probability-possibility transformation from the literature with ST transformation.

Example 7.6. Figure 7.5 represents a Bayesian network and its transformation with ST, then the two joint distributions induced from the network are described in Table 7.31.

Table 7.32 is the transformed possibility distribution of $p(A, B)$. At first sight, we notice that the two distributions are not identical. But in Table 7.31, we have $\pi_{RP-min}(a_1 b_2) < \pi_{RP-min}(a_2 b_2)$ and in Table 7.32, we have $\pi_{RB-ST}(a_1 b_2) > \pi_{RB-ST}(a_2 b_2)$. Thus, the order of these two interpretations is reverse. We observe the same phenomenon with VT transformation using the same Bayesian network.

A	$p(A)$	$\pi_{ST}(A)$
a_1	0.6	1
a_2	0.4	0.8



A	B	$p(B \mid A)$	$\pi_{ST}(B \mid A)$
a_1	b_1	0.6	1
a_1	b_2	0.3	0.7
a_1	b_3	0.1	0.3
a_2	b_1	0.5	1
a_2	b_2	0.3	0.8
a_2	b_3	0.2	0.6

Figure 7.5 – Bayesian network and its transformed possibilistic network with ST

A	B	$p(A,B)$	$\pi'_m(A,B)$
a_1	b_1	0.36	1
a_1	b_2	0.18	0.7
a_1	b_3	0.06	0.3
a_2	b_1	0.2	0.8
a_2	b_2	0.12	0.8
a_2	b_3	0.08	0.6

Table 7.31 – Probability and possibility distributions induced from Figure 7.5

A	B	$\pi''(A,B)$
a_1	b_1	1
a_1	b_2	0.8
a_1	b_3	0.36
a_2	b_1	0.84
a_2	b_2	0.62
a_2	b_3	0.46

Table 7.32 – Possibility distribution transformed with ST

The following proposition generalizes the above observation to any probability-possibility transformation TR.

Proposition 7.14. *Let TR be a probability-possibility transformation. Then there exists a Bayesian network $\mathcal{BN}$ such $\exists \omega_1 \in \Omega$, $\exists \omega_2 \in \Omega$ where:*

$$\pi''(\omega_1) < \pi''(\omega_2) \textit{ does not imply } \pi'_m(\omega_1) < \pi'_m(\omega_2)$$

where: i) $\pi''(\omega) = TR(p(\omega))$ and p is the joint distribution induced by $\mathcal{BN}$ and ii) π'_m is the joint distribution induced by $\mathcal{PN}$.

The following counter-example proves Proposition 7.14 in the qualitative setting.

Counter-example 7.6. Let $\mathcal{BN}$ be the Bayesian network of Figure 7.6 over two disconnected variables A and B. Note that the probability distribution $p(A)$ in $\mathcal{BN}$ is a permutation [6] of the probability distribution $p(B)$. Hence, the transformation of $p(A)$ and $p(B)$ by TR gives $\pi(A)$ and $\pi(B)$ where $\pi(B)$ is also a permutation of $\pi(A)$. In this example, since TR is assumed to preserve the order of interpretations, we have 1> $\alpha_1 > \alpha_2 > \alpha_3$.

The probability and possibility degrees of interpretations a_1b_1 and a_2b_2 are

— $p(a_1b_1) = 0.4 * 0.15 = 0.06$

— $p(a_2b_2) = 0.2 * 0.2 = 0.04$ then, $p(a_1b_1) > p(a_2b_2)$ and $\pi''(a_1b_1) > \pi''(a_2b_2)$ (a)

— $\pi(a_1b_1) = \alpha_3$

6. The permutation property of probability-possibility transformations is discussed in [Sud92].

A	$p(A)$	$\pi(A)$
a_1	0.4	1
a_2	0.2	α_2
a_3	0.25	α_1
a_4	0.15	α_3

B	$p(B)$	$\pi(B)$
b_1	0.15	α_3
b_2	0.2	α_2
b_3	0.25	α_1
b_4	0.4	1

Figure 7.6 – Bayesian-possibilistic network transformation

— $\pi(a_2 b_2) = \alpha_2$ then, $\pi'_m(a_1 b_1) < \pi'_m(a_2 b_2)$ (b)

From (a) and (b) one can see that the relative order of interpretations is reversed whatever is the used transformation in the ordinal setting. In the same way, in the quantitative setting, the relative order of interpretations can not be preserved by any transformation. □

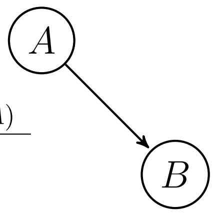
We proved that any probability-possibility transformation TR, that preserves consistency principles, when used in graphical models does not ensure to give the same ordering between interpretations in the joint distributions. These results are valid for the qualitative possibilistic setting. We are now interested in the quantitative possibilistic setting.

7.8.3 Preserving joint distributions in the quantitative case

In the quantitative case, we show that probability-possibility transformations defined in the Chapter 3 does not, as in the qualitative case, always preserved the ordering between interpretations when transforming from a Bayesian network to a possibilistic one. For instance, let us illustrate with OT transformation.

Example 7.7. Let us take the following network (Figure 7.5), and its transformation by OT.

A	$p(A)$	$\pi_{OT}(A)$
a_1	0.6	1
a_2	0.4	0.4



A	B	$p(B \mid A)$	$\pi_{OT}(B \mid A)$
a_1	b_1	0.6	1
a_1	b_2	0.3	0.4
a_1	b_3	0.1	0.1
a_2	b_1	0.5	1
a_2	b_2	0.3	0.5
a_2	b_3	0.2	0.2

Figure 7.7 – Bayesian network with local distributions transformed by OT

As in the qualitative case, we notice that the order between two interpretations is inverted. Indeed, in Table 7.33 we have $\pi_{RP-*}(a_2 b_3) < \pi_{RP-*}(a_1 b_3)$ but in Table 7.34, we actually have $\pi_{RB-OT}(a_2 b_3) > \pi_{RB-OT}(a_1 b_3)$.

In the qualitative possibilistic setting, we generalized this result to any probability-possibility transformation. However, due to the characteristics of quantitative possibilistic chain rule (using the product operator) it is difficult to prove that it does not always preserve the ordering between interpretations in the joint distributions. This can be explained by the incomparability of the results of two products of real

A	B	$p(A,B)$	$\pi'_*(A,B)$
a_1	b_1	0.36	1
a_1	b_2	0.18	0.4
a_1	b_3	0.06	0.1
a_2	b_1	0.2	0.4
a_2	b_2	0.12	0.2
a_2	b_3	0.08	0.08

Table 7.33 – Joint distributions induced from the network of Figure 7.7

A	B	$\pi_{RB-OT}(A,B)$
a_1	b_1	1
a_1	b_2	0.44
a_1	b_3	0.06
a_2	b_1	0.64
a_2	b_2	0.26
a_2	b_3	0.14

Table 7.34 – Possibility distribution transformed by OT

numbers. Meaning that when we have two probability degrees denoted by β_1 and β_2 and two possibility degrees α_1 and α_2 with $\beta_1 < \beta_2$ and $\alpha_1 < \alpha_2$ then one cannot state for sure that

$$\alpha_1 * \alpha_2 < \beta_1 * \beta_2 \text{ nor } \alpha_1 * \alpha_2 > \beta_1 * \beta_2$$

7.9 Conclusion

This chapter dealt with analysis properties of probability-possibility transformations.We showed that probability-possibility transformations that preserve consistency principles of Dubois and Prade preserve the normalization condition, preserve also the ordering between interpretations after conditioning. Therefore, TR procedure preserves the results of MPE queries. Considering marginalization operation, we showed that there is no transformation that can guarantee the preservation of the plausibility ordering between arbitrary events.

An interesting question is therefore whether there exist particular probability distributions p such that the transformation operation TR preserves the relative ordering between interpretations after marginalization. A first natural idea is uniform probability distributions. Any transformation TR should preserve normalization which results in an uniform possibility distribution (where each state is associated to the possibility's degree of 1). Consequently, any event will have a possibility degree of 1, meaning that there will not be a reversal in the order of interpretation on marginals distributions for example. Another kind of probability distributions is called "atomic bond system" [Sno96] or big-stepped [DFP04, BDP99] or lexicographic probability distributions p defined by: $\forall \omega_i \in \Omega$, $p(\omega_i) > \sum\{p(\omega_j) : \omega_j \in \Omega \text{ and } p(\omega_j) < p(\omega_i)\}$. Clearly, if p is a big-stepped distribution then the transformation TR preserves the ordering between events after marginalisation. Note however that for both particular cases (uniform and big-stepped distribution) the ordering between non-elementary events is not preserved.

It is known and this chapter enhanced that probability-possibility transformations suffer from loss of information as we move from an additive framework to a qualitative or semi-qualitative framework. But this does not mean we can do useful tasks with transformations. Next chapter uses probability-possibility transformations to study MAP inference in credal networks. MAP infence based on sets of probabilities is known for their high computational complexity in comparison to Bayesian or possibilistic networks. We provide approximation methods for MAP inference, using probability-possibility transformations, with a better computational cost.

Chapter 8

Approximation of Map Inference in Credal Networks

This chapter focuses on belief graphical models and provides an efficient approximation of *MAP* inference in credal networks using probability-possibility transformations. We first recall two transformations from credal networks to possibilistic ones that are suitable for *MAP* inference in credal networks. Then we provide experimental studies that compare our approach with both standard exact and approximate *MAP* inference in credal networks. This chapter also provides an analysis of *MAP* inference complexity using possibilistic networks and the results definitely open new perspectives for *MAP* inference in credal networks.

In the second part of this chapter, we apply imprecise probability to possibility transformation to learning process. Indeed, we compare learning possibilistic network from imperfect or imprecise information to learning a credal network that we transform into a possibilistic one. We also conduct an experimental study to evaluate the predictive power of possibilistic classifiers given the two approaches.

Contents

8.1 Brief recall on imprecise probability and possibility transformation

In this chapter, we used the two transformations from imprecise probability to possibility presented in Chapter 3, Section 3.2. Here is a quick reminder:

— the first one *MD* is from Masson and Denoeux [MD06], is given by:

$$\pi(\omega_i) = \max_{\mathcal{C}_l \in \mathcal{C}} (\pi^{C_l}(\omega_i))$$

where π^{C_l} is given by Equation (3.13) in Chapter 3.

— the second one *CD* based on cumulative distributions is given by:

$$\pi_{\underline{F}}(\omega_i) = 1 - \max\{\underline{F}(\omega_j) < \underline{F}(\omega_i) : j = 0..n\}$$

$$\pi_{\overline{F}}(\omega_i) = \overline{F}(\omega_i)$$

These two equations $\pi_{\underline{F}}$ and $\pi_{\overline{F}}$ are written as they have been defined in [DDC08] and in Equations (3.15) and (3.16). But as for the use of $\pi_{\underline{F}}$, we will simply consider as a possibility distribution, the lower generalized cumulative distribution $\pi_{\underline{F}}(\omega_i) = \underline{F}(\omega_i)$ and we normalize it.

The first concern we study is as for the previous chapter, the commutativity of these transformations.

8.2 Commutativity of marginalization and conditioning

This section checks whether the interval-based probability-possibility transformations are commutative with respect to two major change operations that are marginalization and conditioning. Namely, the question dealt with here is: Given an imprecise probability distribution IP, do we get exactly the same results when i) we first transform IP into a possibility distribution π then apply the change operation in the possibilistic setting and when ii) we first apply the change operation in the interval-based setting then transform the result into a possibility distribution.

Let us first recall that TR denotes an interval-based probability-possibility transformation satisfying the following principles:

— *Dominance:* The possibility distribution π obtained from the IPD IP by TR dominates every probability distribution p compatible with IP, namely $\forall\phi \subseteq \Omega$, $\pi(\phi) \geq p(\phi)$.

— *Order preservation:* Given two interpretations $\omega_i \in \Omega$ and $\omega_j \in \Omega$, $\pi(\omega_i) < \pi(\omega_j)$ if and only if $\overline{IP}(\omega_i) < \underline{IP}(\omega_j)$.

Regarding the commutativity of transformations with respect to change operations like marginalization and conditioning used to answer MAP queries, since probability distributions are special cases of imprecise probability distributions, it can be expected that for the commutativity issue, the transformations exhibit the same properties.

8.2.1 Commutativity with respect to marginalization

Proposition 8.1 provides the answer for marginalization :

Proposition 8.1. *Let TR be an interval-based probability-possibility transformation. Then there exists an imprecise probability distribution IP, two events $\phi \subseteq \Omega$, $\psi \subseteq \Omega$ with $\phi \neq \psi$, and $\pi = TR(IP)$ such that $\overline{IP}(\phi) < \underline{IP}(\psi)$ but $\Pi(\phi) > \Pi(\psi)$.*

Proposition 8.1 asserts that no interval-based probability-possibility transformation can guarantee the preservation of the order of events as shown in the following example.

Example 8.1. Let IP be an imprecise probability distribution of Table 8.1 where $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $\pi = TR(IP)$.

In this example, α_1, α_2 and α_3 are possibility degrees such that $1 > \alpha_1 > \alpha_2 > \alpha_3$ in order to satisfy the preference preservation principle.

Now, let ϕ and ψ be two events such that $\phi = \{\omega_1\}$ and $\psi = \{\omega_2, \omega_3\}$. We have $\Pi(\phi) = 1 > \Pi(\psi) = \max(\alpha_1, \alpha_2)$ while $\overline{IP}(\phi) = .4 < \underline{IP}(\psi) = .6$.

ω_i	$IP(\omega_i)$	$\pi(\omega_i)$
ω_1	$[.36, .4]$	1
ω_2	$[.35, .35]$	α_1
ω_3	$[.25, .25]$	α_2
ω_4	$[0, .04]$	α_3

Table 8.1 – Example showing the loss of the order of events.

As shown in Example 8.1, the strict order of events is not preserved by TR because of the different behavior of the additivity axiom in the probabilistic setting and the maxitivity axiom of the possibilistic setting used by the marginalization operation.

As a consequence of Proposition 8.1, we have the following Lemma:

Lemma 8.1. Let TR be an interval-based probability-possibility transformation. Then there exists an imprecise probability distribution IP over $\Omega = \{\omega_1, \omega_2, .., \omega_n\}$ and a partition $\Omega' = \{W_1, W_2 ... W_k\}$ of Ω with $k < n$. Let $\pi = TR(IP)$, IP' is obtained by marginalizing IP on Ω' according to Equation (1.40) and π' is obtained by marginalizing π on Ω' in the possibilistic setting. Then there may exist an event $W_i \in \Omega'$ such that

$$\pi(W_i) \neq \pi'(W_i). \tag{8.1}$$

Proof sketch. The proof follows from Proposition 8.1 since if the order of events is not preserved then the underlying marginalized distributions must be different. □

8.2.2 Commutativity with respect to conditioning

Let us now check the commutativity issue with respect to conditioning. For standard probability distributions, we have the following finding [BLT15a]:

Proposition 8.2. *Let p be a probability distribution over Ω and let $\phi \subseteq \Omega$ be an evidence. Let TR be a probability-possibility transformation, p' be a probability distribution obtained by conditioning p by ϕ, $\pi'' = TR(p')$ and π' is the possibility distribution obtained by conditioning $\pi = TR(p)$ by ϕ. Then, $\forall \omega_i, \omega_j \in \Omega$,*

$$\pi'(\omega_i) < \pi'(\omega_j) \text{ if and only if } \pi''(\omega_i) < \pi''(\omega_j).$$

Note that Proposition 8.2 is valid in both the product and the min-based possibilistic settings and it states that the order of interpretations is not affected by the order of applying the transformation and the conditioning operation. For imprecise probability distributions, the following proposition states that the partial order encoded by IP after conditioning is preserved in the (complete) order induced by π after conditioning on the same evidence.

Proposition 8.3. *Let IP be an imprecise probability distribution over Ω and let $\phi \subseteq \Omega$ be an evidence. Let TR be an interval-based probability-possibility transformation, $IP' = IP(.|\phi)$ be a posterior probability distribution obtained by conditioning IP by ϕ, $\pi'' = TR(IP')$ and $\pi' = \pi(.|\phi)$ is the possibility distribution obtained by conditioning $\pi = TR(IP)$ by ϕ. Then,*

$$\forall \omega_i, \omega_j \in \Omega,\ \pi'(\omega_i) < \pi'(\omega_j) \text{ if and only if } \pi''(\omega_i) < \pi''(\omega_j).$$

Proof sketch. The idea of the proof is that since conditioning in both the probabilistic and possibilistic settings consists in discarding the worlds that are not models of the evidence ϕ (by assigning them a 0 probability/possibility degree) then renormalizing the obtained distribution. Hence, the order of interpretations that are models of ϕ is not affected by the order of application of transformation/conditioning operations. □

8.3 Experimental studies on MAP inference

In this section, we give the results of our experimental studies where we have used different criteria to assess MAP queries accuracy in credal networks.

8.3.1 Experimentation setup

Before giving a detailed record of what we have implemented for the experimental study, let us recall that there exists no platform or implemented algorithm that can compute MAP inference in possibilistic networks. Furthermore, there is also no platform that computes MAP inference in credal networks. Yet, there exist packages to perform some inference tasks. Precisely, those packages return the probability degrees or intervals of a variable given an evidence. We implemented:

- the algorithm to transform a credal network into a possibilistic network,
- the algorithm to perform inference in possibilistic networks,
- the algorithm to compute MAP results from the results of the inference algorithm in credal networks and possibilistic networks. Precisely, using the criteria on the imprecise probability distribution and for possibility distribution, it is done by capturing the world with the maximum degree, which if the distribution is normalized, is 1.

As stated above, the first part was to develop the transformation of a credal network into a possibilistic network. This program computes for every linear extension the Equation (3.13) using the linear programming solver GLPK, and returns a new BIF file (adapted to possibilistic networks purposes). This is for the first transformation of Masson and Denoeux, for the *CM* transformation, we simply program Equations (3.15) and (3.16) using one linear extension instead of the whole set. The one we are choosing for this transformation is driven by the format of the file used to represent the credal network. Indeed, JavaBayes uses a format called BIF (*Bayesian Interchange Format*) that stores the credal set of a variable as a set of extreme points. We can easily translate the set of extreme points to an imprecise probability distribution by setting the lower endpoints of an interpretation as the minimum of the set of extreme points for this interpretation and setting the upper endpoints as the maximum. In our case, to choose the linear extension, we simply take the order given by the average of all the extreme points. It is a natural assumption. Both of the obtained distributions are possibility distributions surrounding the imprecise probability distribution [DDC07].

The second part of the implementation was to develop a program that solves the inference problem in possibilistic networks. We used JavaBayes as a base to develop this part especially since probability theory and quantitative possibility theory have a lot in common. Thus, the part that uses the chain rule does not change (it is for both the product-based chain rule), and the only modification refers to the computation of the possibility degree of two events. In probability, it is done using the additivity axiom whereas in possibility theory it is done using the maxitivity axiom.

To program the MAP answering from the distribution of the requested variables, we need first to choose which criteria to use. Therefore, this part of the program returns all interpretations that have the maximum degree according to the used criterion.

The rest of the script is described with the following Algorithm 4.

Algorithm 4 Script of the experimental study

Require: A BIF File representing a credal network $\mathcal{CN}$
Ensure: Precision of the MAP Inference
  $\mathcal{PN}_{MD} = MD(\mathcal{CN})$
  $\mathcal{PN}_{CD} = CD(\mathcal{CN})$
  for $i = 1$ **to** $i = n$ **do**
    Create the new file altered with a new variable resulting in the conjunction of the i variables (if $i \neq 1$)
    Apply JavaBayes on $IP_{jb} = TR(\mathcal{CN})$
    Apply GL2U on $IP_{gl} = TR(\mathcal{CN})$
    Apply JavaPoss on $\mathcal{PN}_{MD}$ and $\mathcal{PN}_{CD}$ results in π_{MD} and π_{CD}
    $S_{gl-id} = \mathit{Interval\text{-}dominance}(IP_{gl})$
    $S_{jb-id} = \mathit{Interval\text{-}dominance}(IP_{jb})$
    $S_{\max} = \mathit{Maximax}(IP_{jb})$
    $S_{\min} = \mathit{Maximin}(IP_{jb})$
    $S_{hur} = \mathit{Hurwicz}(IP_{jb})$
    $S_{MD} = MAP_{Poss}(\pi_{MD})$
    $S_{CD} = MAP_{Poss}(\pi_{CD})$
  end for

8.3.2 Evaluation criteria

The benchmarks used in the current work are presented in Table 8.2.

Networks	Topology	#Nodes	$\max$ |domain|
Alarm	Multiply-connected	37	4
Insurance	Multiply-connected	27	5
Poly	Polytree	10	4
Multi	Multiply-connected	6	4

Table 8.2 – Credal networks used in the experimentations.

In order to compare the results of *MAP* inference in credal networks and their possibilistic counterparts, each query Q is submitted to a credal network $\mathcal{CN}$ (using JavaBayes) then to the corresponding possibilistic network $\mathcal{PN}$ obtained from $\mathcal{CN}$. And in the same way, Q is submitted to a credal network through JavaBayes and to the same credal network using GL2U. The results are compared through the accuracy criterion defined as follows:

$$accuracy(Q_1, Q_2...Q_n) = \frac{1}{n} \sum_{i:1..n} \frac{|CN_{MAP}(Q_i) \cap PN_{MAP}(Q_i)|}{|CN_{MAP}(Q_i) \cup PN_{MAP}(Q_i)|}, \tag{8.2}$$

where $CN_{MAP}(Q_i)$ (resp. $PN_{MAP}(Q_i)$) denotes the results of the query Q_i submitted to the network $\mathcal{CN}$ (resp. $\mathcal{PN}$). This criterion evaluates the agreement between the results of $\mathcal{CN}$ to the *MAP* queries and the ones of $\mathcal{PN}$.

Thus, the experiment provides:

- Accuracy rates compared to the exact algorithm implemented in JavaBayes software:
  - The accuracy of the approximate inference algorithm used in *GL2U* software.

  - The accuracy of *MAP* queries using possibilistic networks obtained by transforming the credal network using *MD* and the accuracy using *CD* transformation.
- Inclusion rates: we compute the proportion of results returned by approximate algorithm that are included in the exact one.
- Accuracy and inclusion rates for each algorithm using the four criterion on the results of the exact approach (JavaBayes).
- Proportion of results: we compare the number of outputs to the number of possible outcomes in order to highlight the confusion level of the algorithm.
- Results on different numbers of query variables: we vary the number of query variables between 1 to 5. For each case, we tested around 200 networks.

8.3.3 Results

This subsection can be divided into two types of results, quantitative ones and qualitative ones.

Quantitative results

This represents the significant part of this experiment in terms of the size of the networks handled and the size of queries. One of the main objectives of this experiment was to show that our approach could considerably reduce the time complexity of *MAP* inference and this is what we expose in Table 8.3. Indeed, this table shows the number of files answered by all of the approaches. And especially, we can notice that *GL*, in terms of number of query variables, can not handle more that 3 variables. Therefore, this approach is very limited. On the contrary, our approach based on the transformations is always better in terms of the number of networks answered and also does not differ from 1 to 5 or more query variables. For the number of query variables 5, and especially for the type of network *multi*, we

# query vars	1			2			3			4			5		
Algorithm	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*	*MD*	*CD*	*GL*
Alarm	187	187	**143**	149	149	149	77	77	68	63	63	**0**	43	43	**0**
Insurance	180	180	164	152	152	152	116	116	**52**	55	55	**0**	_	_	_
Poly	200	200	**140**	200	200	190	200	200	180	200	200	**0**	180	180	**0**
Multi	200	200	**110**	200	200	200	200	200	**120**	200	200	**0**	_	_	_

Table 8.3 – Number of files answered by algorithms.

don't have any results due to the number of variables in the network (we have 6 variables, of which 5 are requested). Since we force at least one observation.

Qualitative results

We have shown that our approach outperforms the other approximate approach in terms of rapidity of execution. So a natural question is about the actual quality of the results. How good are the results? To answer this question, we provide in the table 8.4 some information about the number of outputs returned over the number of possible outcomes. But also the percentage of configurations returned that are included in the exact set of answers where the exact set of answers is given by JavaBayes [7].

In Table 8.4, there are three different results that support our method:

7. Those depicts the results over *poly* networks and with 1 query variable

Criterion	MD	CD	GL	
.794	.685	.36	.88	$\%\ answers/all$
Inter-dom	.967	1	.891	% Inclusion
.36	.685	.36	.88	$\%\ answers/all$
Maximax	.546	.74	.629	% Inclusion

Table 8.4 – Proportion of returned answers over all possible outcomes vs Proportion of included answer sets

i) When using *Interval-dominance* criterion, the number of configurations returned by JavaBayes as the result of MAP inference is around 80% of possible responses. This number clearly shows a lot of confusion and in order to make decisions, one can ask about the relevance of all these results. On the other hand, with *Maximax* criterion we observe, with a proportion of .36, more pruned results. And note that in this case, the method using CD transformation gives a similar number of configurations.

ii) Regarding the transformation MD and the preservation of information, if we look at the proportion of returned answers combined to the proportion of included answers, we can see that MD is the transformation that will preserve the information the better. As recalled in 3.2.1, this transformation comes down to the optimal transformation when considering standard probability measures instead of an interval. These results hold when considering *Interval-dominance* criterion, indeed by transforming with MD, we also increase the imprecision so having a high number of results allows to keep good results with MD in terms of included answers and without returning too much of the possible configurations. Comparing the proportion of returned answers for MD transformation, we notice that this transformation gives the closest network in respect with the credal network.

iii) The table finally shows that the approximate approach GL generally gives sets of results larger than the exact approach. And even more, as the number of requested variables increases, GL tends to return all possible outcomes.

In the following, we expose through graphics the accuracy of each method MD, CD and GL. The axis x must be read as A# for Alarm file and P# for Poly file with # is the number of requested variables. We present the results of two types of files, polytrees and multiply-connected networks, and with three different criteria, *Interval-dominance*, *Maximax* and *Hurwicz*. Indeed, we forget in this part *Maximin* criterion due to the similarity in the accuracy with *Maximax* and *Hurwicz* criteria.

On Figure 8.1, the approximate method GL gives better results for both types of networks. Except beyond 4 query variables where it can no longer return answers. This problem can be explained by the fact that the variables are chosen randomly and it can affect the difficulty of the MAP inference algorithm implemented. So from this graphic, we can conclude that using *interval-dominance* criterion will favor the GL method when having small networks with small amount of query variables. The results for GL are easily explained and echoed the previous results presented in Table 8.4. Indeed, by the fact that this method returns around 88% of the configurations, it is more likely to be in the 79% of the results returned by the exact method. So this method, if decision is not the end goal, can be chosen.

As for the approximate approach using MD transformation, if we correlate the accuracy results observed in the graphics, with the proportion given in Table 8.4, then MD is slightly better than GL. Indeed, by returning less configurations than the exact approach and having a better proportion of included answers, it balances the accuracy rate which is still better that CD. As well, this approach is not bounded by the size of the network nor the size of the query variable set.

CD in that case should be used when MAP inference is done in order to make a decision, for example,

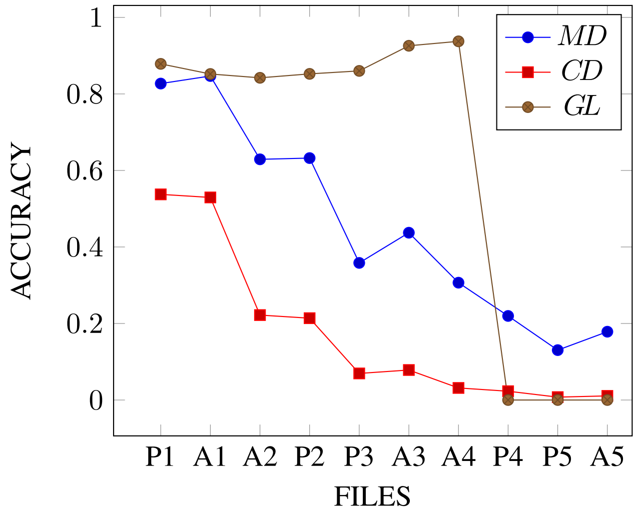


Figure 8.1 – Comparison between *MD*, *CD* and *GL* using *Interval-dominance* criterion

in case of classification [HLLT17] since it returns less answers than the exact approach but still all of these answers are included in the set of answers of the exact approach.

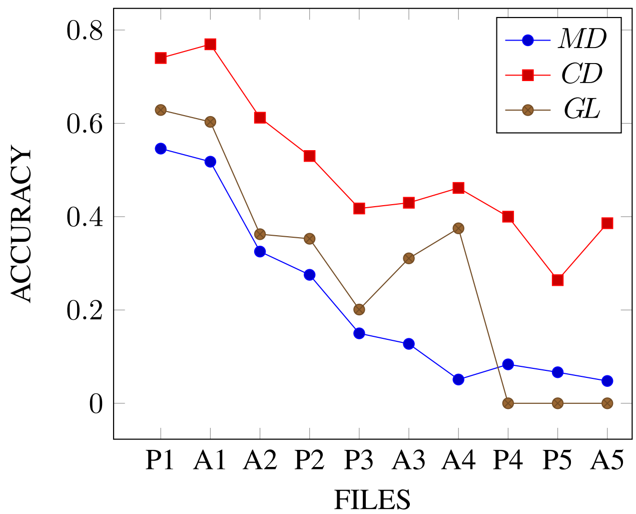


Figure 8.2 – Comparison between *MD*, *CD* and *GL* using *Maximax* criterion

Now, considering *Maximax* criterion which we chose to prune the results from the exact method, we observe on Figure 8.2 that *CD* gives the best results in terms of accuracy but also in terms of inclusion (cf. Table 8.4). Still, the accuracy decreases when the number of requested variables increases. This criterion being more restrictive, it obviously leads to more imprecise results.

In order to not favor optimistic or pessimistic evaluation, we also conducted our experiment using *Hurwicz* criterion with the 0.5 degree associated to each evaluation. In terms of results (Figure 8.3), they are more or less the same as *Maximax* criterion. This is why, in the last graphic (Figure 8.4), we compare those 3 criteria with *CD* method.

What we can see from Figure 8.4 is that the three criteria mostly behave the same way. We can con-

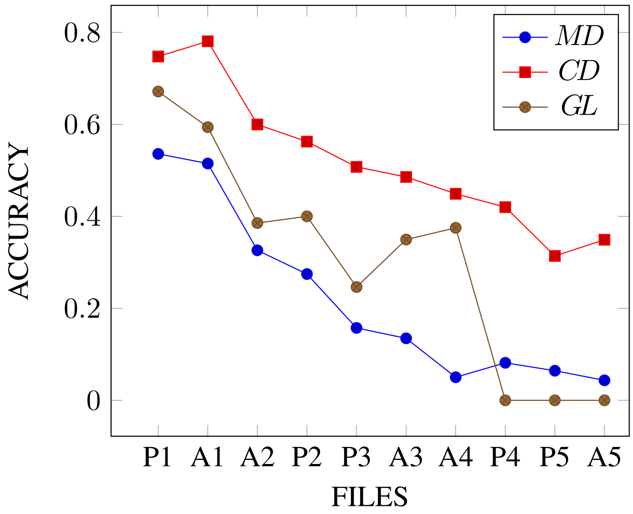


Figure 8.3 – Comparison between *MD*, *CD* and *GL* using *Hurwicz* criterion

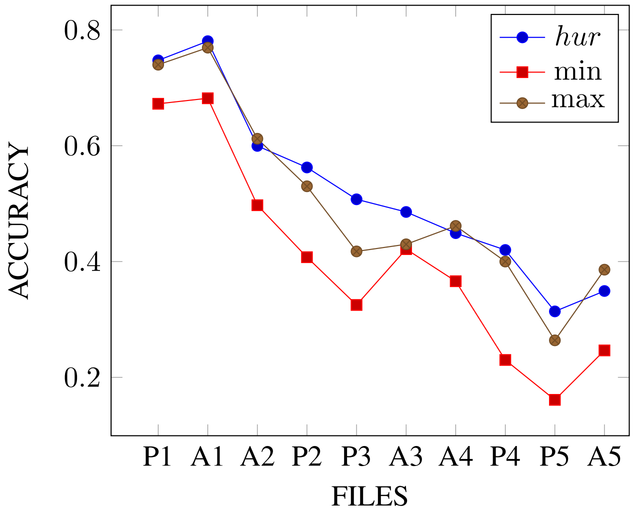


Figure 8.4 – Comparison between *Maximax*, *Maximin*, *Hurwicz* criteria for *CD* method

clude, from this, that from those three criteria, one should choose *Hurwicz* criterion, and if we would like to favor an optimistic (resp. pessimistic) evaluation, we could increase the degree of *Hurwicz* criterion (resp. decrease) on the first part of the degree α (see Definition 2.16).

Overall, the three approximate algorithms show the same behavior towards the number of requested variables, the accuracy rates all decrease as the number of variables increases. There are a numerous and various criteria to be taken into account in *MAP* inference. Variables, whether it is evidence variables or query variables are chosen randomly, which in some cases may make the process more difficult. We can now provide an overview of what to use in what context.

- When there is a need to take into account any possible configuration that can be a result of *MAP* inference, it is better to use *GL* with *Interval-dominance* criterion or *MD* if you have large network and a lot of query variables.

- ☐ When there is a need to make a decision with the results and thus, need a more pruned set, it is preferable to use *CD* method with *Hurwicz* criterion where where it is possible to play with the degree α to favor an optimistic or a pessimistic approach.

This section shows empirically that possibilistic networks ensure an interesting trade-off in terms of accuracy and computational time. This led us to start investigating the issue of computational complexity in possibilistic networks. The following section provides some preliminary findings.

8.4 On the complexity of inference in possibilistic networks

There is no systematic study of complexity issues for inference in possibilistic networks and most of the works assume that the same complexity results in Bayesian networks still hold in the possibilistic setting. This section briefly shows that inference in possibilistic networks is less costly that in Bayesian networks.

8.4.1 Notations and preliminaries

We will use the following notations:

— $\mathcal{PN}$ a possibilistic network over a set of variables $V = \{V_1, V_2, .., V_n\}$.

— $X_1, .., X_n$ is a set of boolean variables. x_i denotes the truth assignment $X_i = true$ while $\overline{x_i}$ the truth assignment $X_i = false$.

— $\psi(X_1, X_2, .., X_n)$ is a 3-CNF formula encoding a 3-SAT satisfiability problem. A CNF is conjunction of clauses where a clause is a disjunction of literals and a literal is either a boolean variable X_i or its negation $\overline{X_i}$.

— SAT (Satisfiability problem): Given a CNF, the question is to know whether there is a truth assignment x of X making ψ true. If the CNF of the SAT problem contains at least three literals per clause, then the problem is known to be NP-Complete [Coo]. This is the class of problems needing to search a solution in a search space exponential in the size of the problem (here, the number of boolean variables) but it is easy to check if a given instantiation x is a solution.

Recall that in the literature, there are typically three types of queries that one would need when reasoning with belief graphical models. In possibilistic networks, these queries can be stated as follows:

— Computing the possibility degree of an event (Po): This problem consists in computing the degree $\Pi(q|\phi)$ for an event q of interest given an evidence ϕ.

— Computing the most plausible explanation (*MPE*). Given an evidence ϕ of some variables O, the objective is to compute the most plausible instantiation q of **all the remaining (unobserved) variables** Q. Note that here $O \cup Q = V$ and $Q \cap O = \emptyset$.

— Computing the maximum a posteriori (*MAP*). Given some evidence ϕ, the objective is to compute the most plausible instantiation q of the query variables Q. In *MAP* queries, $Q \cap O = \emptyset$. Note that when Q and O span over all variables, the problem is known as the most plausible explanation (*MPE*).

8.4.2 Complexity analysis

Let us start with analyzing *MPE* queries. The decision problem corresponding to such queries can be formally stated as follows:

Definition 8.1. Let $\mathcal{PN}$ be a multiply-connected possibilistic network and ϕ be an evidence. Let **D-**MPE be the decision problem: Is there an instantiation q of **all non observed variables** Q such that $\Pi(q, \phi) > t$? with $t \in [0, 1]$.

Recall that in MPE queries, $Q = X \setminus E$. Intuitively, the decision problem for computing the most plausible configuration q given the evidence ϕ comes down to answering whether the degree $\Pi(q, \phi)$ is greater than a rational number t.

Theorem 8.1. **D-**MPE *is NP-complete.*

Proof. We need to prove the membership and the hardness of the complexity:

— **Membership** (**D-**MPE **is in** NP): Given an instance q, it is easy to check if $\Pi(q, \phi) > t$. Indeed, $x = (q, \phi)$ is a complete instantiation of the network variables, hence the possibility degree $\Pi(q, \phi)$ is computed in polynomial time (more precisely, in linear time) in the size of the network (number of variables) using the chain rule (namely, $\Pi(X_1, .., X_n) = \prod_{i=1}^{n} \pi(X_i|par(X_i))$).

— **Hardness**: This is done by reducing the 3-SAT problem to **D-**MPE in possibilistic networks (similar to reducing the 3-SAT problem to **D-**MPE problem in Bayesian networks [Dar09]).
Given a 3-CNF $\psi(X_1, X_2, .., X_n)$ over a set of n boolean variables, the query is whether there exists a truth assignment $x = (x_1, x_2, .., x_n)$ making ψ true. The reduction from 3-CNF to a possibilistic network PN_ψ is done as follows:

  — If ψ involves a unique variable X_i then PN_ψ is composed of only one node X_i with $\pi(x_i) = 1$ and $\pi(\overline{x_i}) = 1$ (attaching a possibility degree of 1 to both x_i and $\overline{x_i}$ aims to allow both them to be fully possible).

  — If ψ is in the form $\neg\mu$ then add a boolean variable X_ψ as a child of X_μ with ($\pi(x_\psi|x_\mu) = 0$ et $\pi(\overline{x_\psi}|x_\mu) = 1$).

  — If ψ is in the form $\mu \vee \varphi$ then add a boolean variable X_ψ as a child of variables X_μ and X_φ with ($\pi(x_\psi|x_\mu, x_\varphi) = 1$, $\pi(x_\psi|x_\mu, \overline{x_\varphi}) = 1$, $\pi(x_\psi|\overline{x_\mu}, x_\varphi) = 1$ and $\pi(x_\psi|\overline{x_\mu}, \overline{x_\varphi}) = 0$.

  — If ψ is in the form $\mu \wedge \varphi$ then add a boolean variable X_ψ as a child of variables X_μ and X_φ with ($\pi(x_\psi|x_\mu, x_\varphi) = 1$, $\pi(x_\psi|x_\mu, \overline{x_\varphi}) = 0$, $\pi(x_\psi|\overline{x_\mu}, x_\varphi) = 0$ and $\pi(x_\psi|\overline{x_\mu}, \overline{x_\varphi}) = 0$.

The main idea of this reduction is to introduce three kinds of nodes: i) nodes for the boolean variables $X_1..X_n$ of the 3-CNF formula ψ, ii) nodes for encoding the logical gates that can be found in a CNF formula, namely (negation, disjunction and conjunction) and iii) a final node representing the whole 3-CNF formula ψ. It is clear that the size of the network is polynomial in the number of variables and clauses of the 3-CNF formula.
Now, let ψ be a 3-CNF formula and PN_ψ be the possibilistic network built from ψ. Let L_ψ be the only leaf of this network. Then

$$\Pi(x_1, .., x_n, l_\psi = 1) = \begin{cases} 1 & \text{if and only if } x_1, .., x_n \models \psi \\ 0 & \text{if and only if } x_1, .., x_n \models \neg\psi \end{cases} \tag{8.3}$$

It is clear that by construction if a full truth assignment $(x_1, .., x_n, l_\psi = 1)$ of the possibilistic network has a possibility degree of 1 if and only if $(x_1, .., x_n)$ is a model of ψ. Finding a full assignment of nodes X_1,..X_n ensuring $\Pi(x_1, .., x_n, l_\psi = 1) = 1$ comes down to solving the 3-SAT problem which is NP-complete. □

The MPE problem was firstly shown to be NP-complete in multiply-connected Bayesian networks in [Shi94].

In order to show the complexity of MAP queries, we need first to show the one of computing the possibility degree of an event of interest q given an evidence ϕ.

Definition 8.2. Let PN be a multiply-connected possibilistic network and ϕ be an evidence and q an instantiation of non observed variables Q. **D-Po** is the decision problem: Does the statement $\Pi(q, \phi) > t$ hold? (with $t \in [0, 1]$).

Contrary to Bayesian networks where computing $P(q, \phi)$ from a Bayesian network is a counting problem (marginalization computations need product and summation operations) and it is PP-Complete [Rot96], the decision problem corresponding to computing $\Pi(q, \phi)$ from a possibilistic network comes down to answer the statement: *Is there a full instantiation $(x_1, .., x_n)$ of the network variables $(X_1, .., X_n)$ that is compatible with q and ϕ and such that $\Pi(x_1, .., x_n) > t$.* Indeed, $\Pi(q, \phi) > t$ means that $\max_{x \in \Omega \cap q \cap \phi} \Pi(x) > t$. This is due to the fact possibility theory is maxitive while probability theory is additive. The maxitivity property makes the problem of computing a possibility degree less computationally expensive than computing a probability in Bayesian networks.
Clearly **D-Po** is a search problem and the same reduction form 3-SAT to $D - MPE$ can be adapted for **D-Po**. Hence the following corollary:

Corollary 8.1. **D-Po** is NP-complete.

The proof is the same as for Theorem 8.1.
Same reasoning holds for *MAP* queries. Searching for the most plausible instantiation of query variables Q given an evidence ϕ comes down to a search problem. More formally,

Definition 8.3. Let PN be a multiply-connected possibilistic network and ϕ be an evidence. **D-***MAP* be the decision problem: Is there an instantiation q of non observed variables Q such that $\Pi(q, \phi) > t$? (where $t \in [0, 1]$).

Intuitively, the decision problem for computing the a posteriori most plausible configuration q given the evidence ϕ is answering whether the $\Pi(q, \phi)$ is greater that a rational number t. Please recall that in *MAP* queries, $Q \subseteq X \setminus E$. As for *MPE* queries, answering *MAP* queries in possibilistic networks comes down to answer whether the statement: is there an instantiation $(x_1, .., x_n)$ of the network variables $(X_1, .., X_n)$ that is compatible with q and e and such that $\Pi(x_1, .., x_n) > t$. As a consequence, **D-***MAP* has the same complexity as **D-Po** and **D-***MPE* , namely NP-complete as stated by the following corollary:

Theorem 8.2. **D-***MAP is NP-complete.*

The proof is also the same as for Theorem 8.1. In Bayesian networks, this problem is shown to be NP^{PP}-complete [PD04].

To conclude on the complexity of *MAP* inference in possibilistic networks, given a multiply-connected network, we reduce the complexity from NP^{PP}-complete to NP-complete.
We provided a new approach to perform *MAP* inference in credal network using possibilistic network. Classification in this respect is an application of *MAP* inference and we are interested in a comparison between learning a possibilistic network from uncertain data and learning a credal network then transform it into a possibilistic one.

8.5 Learning possibilistic network parameters from imprecise data: application of *MAP* inference

In this section, we are interested in learning belief graphical models, the evaluation is generally done by comparing reference networks with the learned ones. Reference networks are graphical models that are either chosen by an expert or randomly generated. From the reference model, a dataset is generated following the distribution encoded by the reference model. This dataset is then used to learn models using the approach to be evaluated. The problem then comes down to compare the learn model with the reference one. A comparison may take into account only the joint measures encoded by the learned and the reference models. In addition, one may want also to take into account the structure of the learned and reference models.

Given that we are only interested in comparing possibilistic networks with same structure, there is no need to consider the graphical component in our comparisons. One simple but costly way of comparing the reference network with the learned one is to compare only the joint distribution encoded by the reference model with the learned model distribution. An example of similarity measure for possibility distributions is information affinity [JAE+07]. However the size of the distribution may be very huge (it fact, it is exponential in the number of variables of the network) making it impossible to compare joint possibility distributions. We propose a heuristic method that compares the networks local distributions locally and aggregates the results to provide an overall similarity score of two possibilistic networks.

8.5.1 Similarity of two possibility distributions

Many measures were proposed for assessing the similarity between two possibility distributions π_1 and π_2 over the same universe of discourse Ω. Among such measures, information affinity [JAE+07], is defined as follows:

$$InfoAff(\pi_1, \pi_2) = 1 - \frac{d(\pi_1, \pi_2) + Inc(\pi_1, \pi_2)}{2} \tag{8.4}$$

where $d(\pi_1, \pi_2)$ represents the mean Manhattan distance between possibility distributions π_1 and π_2 and it is defined as follows:

$$d(\pi_1, \pi_2) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} |\pi_1(\omega_i) - \pi_2(\omega_i)|.$$

As for $Inc(\pi_1, \pi_2)$, it is a measure of inconsistency and it assesses the conflict degree between π_1 and π_2. Namely,

$$Inc(\pi_1, \pi_2) = 1 - \max_{\omega_i \in \Omega}(\pi_1(\omega_i) \wedge \pi_2(\omega_i)) \text{ where } \pi_1(\omega_i) \wedge \pi_2(\omega_i) \tag{8.5}$$

denotes a combination operation of two possibility distributions. In [JAE+07] , the $\min$ operator is used in a qualitative setting. In a quantitative setting, a product operator can be used as well.

Example 8.2. Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and let π_1 and π_2 be two possibility distributions such that $\pi_1 = (1, .8, .4, 0)$ and $\pi_2 = (.9, 1, .2, .2)$. Then $InfoAff(\pi_1, \pi_2) = 1 - \frac{d(\pi_1,\pi_2)+Inc(\pi_1,\pi_2)}{2} = 1 - \frac{(0.175+0.1)}{2} = 0.8625$.

The measure of Equation (8.4) satisfies the following natural properties:

— **(P1) Non-negativity**: $InfoAff(\pi_1, \pi_2) \geq 0$.

— **(P2) Symmetry**: $InfoAff(\pi_1, \pi_2) = InfoAff(\pi_2, \pi_1)$.

— **(P3) Upper bound and Non-degeneracy**: $InfoAff(\pi_1, \pi_2)$ is maximal if and only if π_1 and π_2 are identical. Namely, $InfoAff(\pi_1, \pi_2) = 1$ if and only if $\forall \omega \in \Omega$, $\pi_1(\omega) = \pi_2(\omega)$.

— **(P4) Lower bound**: $InfoAff(\pi_1, \pi_2)$ is minimal if and only if π_1 and π_2 contain maximally contradictory possibility distributions. Namely, $InfoAff(\pi_1, \pi_2) = 0$ if and only if

i) $\forall \omega \in \Omega$, $\pi_1(\omega) \in \{0, 1\}$ and $\pi_2(\omega) \in \{0, 1\}$, and

ii) $\pi_1(\omega) = 1 - \pi_2(\omega)$

— **(P5) Inclusion**: If π_1, π_2 and π_3 are three possibility distributions over the same universe of discourse Ω and $\forall\omega \in \Omega$, $\pi_1(\omega) \leq \pi_2(\omega) \leq \pi_3(\omega)$ then $\mathit{InfoAff}(\pi_1, \pi_2) \geq \mathit{InfoAff}(\pi_1, \pi_3)$.

— **(P6) Permutation**: This property states that permuting the degrees or indexes of possibility distributions should result in the same information affinity. Formally, $\mathit{InfoAff}(\pi_1, \pi_2) = \mathit{InfoAff}(\sigma(\pi_1), \sigma(\pi_2))$ where π_1, π_2 are two possibility distributions over Ω and $\sigma(\pi)$ is a permutation [8] of elements of π.

Unfortunately, the affinity measure of Equation (8.4) applies only on possibility distributions and cannot be directly applied for assessing the similarity of two possibilistic networks.

8.5.2 Similarity of two possibilistic networks

To assess the similarity of two possibilistic networks G_1 and G_2 having the same structure (same DAG), it may be relevant to compare every local possibility distribution π_1^i in the network G_1 with π_2^i, namely its corresponding distribution in G_2. This can be done for instance using an aggregation function that takes into account all the local distributions and returns a global similarity score between $G1$ and $G2$.

$$\mathit{GrInfoAff}(G_1, G_2) = \mathit{Agg}_{i=1..m}(\mathit{InfoAff}(\pi_1^i, \pi_2^i)) \tag{8.6}$$

To the best of our knowledge, there is no decomposable similarity measure over possibilistic networks. As examples of aggregation functions, one can use the *minimum*, *maximum*, *mean*, *weignted mean*, *sum*, *product*, etc. In order to study the properties of similarity measures of Equation (8.6), let us first rephrase properties **(P1)-(P6)** in case where the possibility distributions π_1 and π_2 are compactly encoded by means of networks G_1 and G_2.

— **(GP1) Non-negativity**: $\mathit{GrInfoAff}(G_1, G_2) \geq 0$.

— **(GP2) Symmetry**: $\mathit{GrInfoAff}(G_1, G_2) = \mathit{GrInfoAff}(G_2, G_1)$.

— **(GP3) Upper bound and Non-degeneracy**: $\mathit{GrInfoAff}(G_1, G_2)$ is maximal if and only if the joint possibility distributions π_{G_1} and π_{G_2} encoded respectively by G_1 and G_2 are identical. Namely, $\mathit{GrInfoAff}(G_1, G_2) = 1$ if and only if $\forall i = 1..n$, $\forall x_i \in D_i$, $\pi_1(x_1x_2..x_n) = \pi_2(x_1x_2..x_n)$. This property only requires that the two joint possibility distributions encoded by G_1 and G_2 are identical to give a maximal similarity score.

— **(GP4) Lower bound**: $\mathit{GrInfoAff}(G_1, G_2)$ is minimal if and only if the joint distributions π_{G_1} and π_{G_2} contain maximally contradictory possibility distributions. Namely, $\mathit{GrInfoAff}(G_1, G_2) = 0$ if and only if

i) $\forall i = 1..n$, $\forall x_i \in D_i$, $\pi_{G_1}(x_1x_2..x_n) \in \{0, 1\}$ and $\pi_{G_2}(x_1x_2..x_n) \in \{0, 1\}$, and

ii) $\pi_{G_1}(x_1x_2..x_n) = 1 - \pi_{G_2}(x_1x_2..x_n)$

— **(GP5) Inclusion**: If π_{G_1}, π_{G_2} and π_{G_3} are three possibility distributions encoded respectively by three possibilistic networks G_1, G_2 and G_3 such that $\forall x_i \in D_i$, $\pi_{G_1}(x_1x_2..x_n) \leq \pi_{G_2}(x_1x_2..x_n) \leq \pi_{G_3}(x_1x_2..x_n)$ then $\mathit{GrInfoAff}(G_1, G_2) \geq \mathit{GrInfoAff}(G_1, G_3)$.

— **(GP6) Permutation**: This property states that permuting the degrees or indexes of joint possibility distributions should result in the same $\mathit{GrInfoAff}$. Formally, $\mathit{GrInfoAff}(\pi_{G_1}, \pi_{G_2}) = \mathit{GrInfoAff}(\sigma(\pi_{G_1}), \sigma(\pi_{G_2}))$ where $\sigma(\pi_{G_i})$ is a permutation of the degrees or indexes of the joint possibility distribution π_{G_i}.

8. For example, let $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and $\pi_1 = (1, .7, 0)$ and $\pi_2 = (.6, 1, .2)$ and let $\sigma(\pi_1) = (0, .7, 1)$ and $\sigma(\pi_2) = (.2, 1, .6)$. Then it is clear that $\mathit{InfoAff}(\pi_1, \pi_2) = \mathit{InfoAff}(\sigma(\pi_1), \sigma(\pi_2))$.

The following proposition provides for each aggregation function among max, min, *sum*, *mean* and *product* the set of properties defined above that are satisfied by the affinity measure based on such an aggregation function.

Proposition 8.4. *Let G_1 and G_2 be two possibilistic networks defined over the same set of variables $V = \{X_1, .., X_n\}$ and sharing the same DAG. Then $GrInfoAff$ satisfies the properties given in Table 8.5 depending on the used aggregation function.*

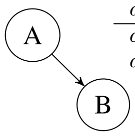
	Maximum	Minimum	Sum	Mean	Product
Non-negativity (GP1)	✓	✓	✓	✓	✓
Symmetry (GP2)	✓	✓	✓	✓	✓
Upper bound (GP3)	✓	✓	✗	✓	✓
Lower bound (GP4)	✓	✗	✓	✓	✗
Inclusion (GP5)	✓	✗	✗	✗	✗
Permutation (GP6)	✓	✓	✓	✓	✓

Table 8.5 – Properties satisifed by some aggregation functions

Proof sketch. While it is obvious that all the aggregation functions of Table 8.5 make the similarity measure $GrInfoAff$ non-negative and symmetric, the satisfaction of the other properties depend on the used aggregation function.
For the *Upper bound property (GP3)*, all the aggregation functions except the Sum satisfy this property since two possibilistic networks with same structure must have the same local distributions. If the affinity measure $GrInfoAff$ were based on the Sum aggregation function, then this bound will depend on the number of variables, their domains and the structure of the network (actually, it will depend on the number of conditional tables in the possibilistic network).
For the *Lower bound property (GP4)*, it is clear that it is enough for two local tables in G_1 and G_2 to have their $InfoAff$ equal to zero (in case of strong conflict between these tables) to force $GrInfoAff$ using the Minimum or the Product aggregation functions to equal zero as shown in the following example.

A	π_{G_1}(A)
a_1	1
a_2	.2

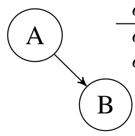
A	B	π_{G_1}(B|A)
a_1	b_1	1
a_1	b_2	1
a_2	b_1	0
a_2	b_2	1



A	B	$\pi_{G_1}(AB)$
a_1	b_1	1
a_1	b_2	1
a_2	b_1	0
a_2	b_2	.2

A	π_{G_2}(A)
a_1	1
a_2	.2

A	B	π_{G_2}(B|A)
a_1	b_1	1
a_1	b_2	1
a_2	b_1	1
a_2	b_2	0



A	B	$\pi_{G_2}(AB)$
a_1	b_1	1
a_1	b_2	1
a_2	b_1	.2
a_2	b_2	0

Figure 8.5 – Example of two possibilistic networks G_1 and G_2 and their joint distributions π_{G_1} and π_{G_2}.

For the inclusion property (GP5) , only the Maximum aggregation satisfies this property. It is easy to find a counter example with three possibilistic networks G_1, G_2 and G_3 encoding three joint distributions such $\pi_{G_1}(x_1..x_n) \leq \pi_{G_2}(x_1..x_n) \leq \pi_{G_3}(x_1..x_n)$ for any variables instance $x_1..x_n$ but

$GrInfoAff(G_1, G_2) < GrInfoAff(G_1, G_3)$.
For the permutation property (GP6), all the aggregation functions satisfy this property since permuting the joint distributions π_{G_1} and π_{G_2} encoded by the two networks G_1 and G_2 respectively comes down to permute some local tables. Now, since the affinity measure $InfoAff$ applied to local distributions satisfies the permutation property, then this will change the result of the aggregation. □

In our first series of experiments, we used the Mean aggregation function since it satisfies most of the properties and it outputs a score taking into account all the local scores of local tables.

8.6 Experimental studies: learning from imprecise data

In this section we provide two series of experiments to compare two approaches that we present in the following for learning the parameters of possibilistic networks. The first series of experiments is carried out on synthetic imprecise data while the second one is done on real datasets with missing values used in supervised classification.

8.6.1 Approaches to learn the parameters of possibilistic networks

Learning the parameters of a possibilistic network is the problem of assessing the entries of local possibility tables $\pi(X_i|par(X_i))$ for each variable X_i given a structure G and a dataset $\mathcal{D}$. The structure here is assumed to be given (*e.g.* when learning naive classifiers, the structure is fixed in advance by assumption) or learnt automatically. There are basically two ways to learn the parameters [HLA15]: i) Transformation-based approach (TA for short) and ii) Possibilistic-based approach (PA for short). Note that the authors in [SP15] propose a possibilistic-based method for learning the structure of a Bayesian network.

Transformation-based approach

In this work, our contribution consists in using transformations from a credal network into a possibilistic network. This means that from the data we build a credal network then transform. Learning a credal network from imprecise data is quite easy and natural. The transformation used in this experiment is the *cumulative distributions* (*CD*).

Given a DAG, let $\mathcal{D}$ be a dataset which has the following format:

x_1	y_2	z_2
$\{x_1, x_2\}$	$\{y_1, y_3\}$	z_2
x_3	y_2	$\{z_1, z_2\}$

where each line corresponds to an entry of the data. Given the structure of the DAG G depicting the dependence relations we want to compute the lower and upper endpoints of the interval for each local distributions. Let us consider first the case of a variable that has no parents. In this case, for a variable X_i, we simply count the number of occurrences $X_i = x_j$ in the dataset and then normalize by the number of entries of the dataset. To distinguish the lower from upper bound, we identify in the dataset which line contains x_j as a unique value. Meaning that to compute the number of occurrences of the lower bound, we count only entry where x_j is the unique value for the variable X_i, in the format given above if we take variable X and its value x_1, it only appears one time alone. To compute the number of occurrences for the upper bound, we count any line where x_j appears. For instance, x_1 appears twice. This gives us, $\underline{IP}(x_1) = 1/3$ and $\overline{IP}(x_1) = 2/3$. This means that x_1 appears at least $1/3$ times and at most $2/3$ times. More formally, given a DAG and a dataset $\mathcal{D}$. Let us denote by N_{ij} the number of occurrences of variable $X_i = x_j$, U_{ij} the number of occurrences $X_i = x_j$ precise and N_l the number of lines in the dataset. Then for a variable X_i,

$$\underline{IP}(x_j) = \frac{U_{ij}}{N_l} \text{ and } \overline{IP}(x_j) = \frac{N_{ij}}{N_l} \tag{8.7}$$

Now, let us consider the case where the variable X_i has parents that we denote by the subset $\{X_j\}$. In this case, we now compute for each configuration of the parents, the number of occurrences of $X_i = x_k$ in the dataset and then normalize by the number of lines that contains the configuration of the parents. In the same way we distinguish the lower and upper endpoints of the interval by separating the precise value $X_i = x_k$ to the imprecise value $x_k \in X_i$. Note that we not only consider configurations of the parents that are precise but all of the lines where the configuration appears. More formally, let us denote by N_{ijk} the number of occurrences $X_i = x_k$ given the configuration j of the parents of X_i. Let U_{ijk} be the number of precise occurrences of $X_i = x_k$ such that $j \in D_{par(X_i)}$. Finally, we denote by N_{lj} the number of entries in the dataset for the configuration j, then for a variable X_i we have $\forall j \in D_{par(X_i)}$

$$\underline{IP}(x_k|j) = \frac{U_{ijk}}{N_{lj}} \tag{8.8}$$

$$\overline{IP}(x_k|j) = \frac{N_{ijk}}{N_{lj}} \tag{8.9}$$

Example 8.3. Given the following dataset in Table 8.6 where A and B are two variables having for domains $D_A = \{a_1, a_2, a_3, a_4\}$ and $D_B = \{b_1, b_2\}$.

A	B
a_1	b_1
a_2	b_1
a_3	$?$
$\{a_3, a_4\}$	b_2
a_1	b_1

Table 8.6 – Imprecise dataset of variable A and B

In this example, when having a subset like $\{a_3, a_4\}$ is how the notion of imprecision is expressed. In the same way, $'?'$ expressed total ignorance, meaning that it can be either b_1 or b_2, it is equivalent to having the subset $\{b_1, b_2\}$.

We obtain the following imprecise probability distribution and its transformed possibility distribution.

The linear extension used in this example is: $a_4 < a_2 < a_3 < a_1$ for A and $b_2 < b_1$ for B.

Possibilistic-based approach

One view of possibility theory is to consider a possibility distribution π on a variable X_i as a *contour function* of a random set [S+76] pertaining to D_i, the domain of X_i. A random set in D_i is a random variable which takes its values on subsets of D_i. More formally, let D_i be a finite domain. A basic probability assignment or mass function is a mapping $m : 2^{D_i} \longmapsto [0, 1]$ such that $\sum_{x_i \subseteq D_i}(m(x_i)) = 1$ and $m(\emptyset) = 0$. A set $x_i \subseteq D_i$ such that $m(x_i) > 0$ is called a focal set.

A	$IP(A)$	$\pi_{TA}(A)$
a_1	$[2/5, 2/5]$	1
a_2	$[1/5, 1/5]$	.4
a_3	$[1/5, 2/5]$	.6
a_4	$[0/5, 1/5]$	.2

Table 8.7 – Imprecise probability distribution and its transformed possibility over variable A issued from the dataset

B	$IP(B)$	$\pi_{TA}(B)$
b_1	$[3/5, 4/5]$	1
b_2	$[1/5, 2/5]$	.4

Table 8.8 – Imprecise probability distribution and its transformed possibility over variable B issued from the dataset

The possibility degree of an event x_i is the probability of the possibility of the event *i.e.* the probability of the disjunction of all events (focal sets) x'_i in which this event is included [BSK09]:

$$\pi(x_i) = \sum_{x'_i | x_i \cap x'_i \neq \emptyset} m(x'_i) \tag{8.10}$$

A random set is said to be *consistent* if there is at least one element x_i contained in all focal sets x'_i and the possibility distribution induced by a consistent random set is, thereby, normalized. Exploring this link between possibility theory and random sets theory has been extensively studied, in particular, in learning tasks, we cite for instance [BSK09, Jos97]. In what follows, we present obtained results *i.e.* the possibilistic-likelihood-based parameters algorithm.

Given a DAG and an imprecision degree S_i, let $\mathcal{D}_{ij} = \{d_{ij}^{(l)}\}$ be a dataset relative to a variable X_i, $d_{ij}^{(l)} \in D_{ij}$ (resp. $d_{ij}^{(l)} \subseteq D_{ij}$) if data are precise (resp. imprecise). The number of occurrences $X_i = x_{ik}$ such that $par(X_i) = j$, denoted by N_{ijk}, is the number of times $X_i = x_{ik}$ such that $par(X_i) = j$ appears in $\mathcal{D}_{ij} : N_{ijk} = \text{card}(\{l \text{ such that } X_i = x_{ik} \text{ such that } par(X_i) = j \in d_{ij}^{(l)}\})$.

$$\pi(X = X_{ik} | \hat{par}(X_i) = j) = \frac{N_{ijk}}{\sum_{k=1}^{r_i} N_{ijk}} * S_i \tag{8.11}$$

where q_i is $\text{card}(par(X_i))$, $r_i = \text{card}(D_i)$ and S_i corresponds to the imprecision degree relative to a variable X_i. To obtain normalized possibility distributions, we divide each obtained distribution by its maximum. It is evident that this operation eliminates S_i. However, we could assign to each value of X_i an imprecision degree which could be either set by an expert or inferred from the dataset to learn from.

8.6.2 Assessing the similarity of possibilistic networks

Experimentation setup

In this experiment, given a dataset $\mathcal{D}$ and a network structure (DAG) $\mathcal{S}$, we compare learning a possibilistic network parameters using two approaches, TA and PA. We denote by G^{TA} (resp. G^{PA}) the possibilistic network having the structure $\mathcal{S}$ and its parameters are learned over the dataset $\mathcal{D}$ using the transformation-based approach TA based on the p-box transformation (resp. the possibilistic-based approach PA).

We first generated a set of possibilistic networks with different features (number of variables, number of parents per variable, rate of imprecise data, etc.). For each possibilistic network G, we generate datasets according to G. More precisely, for each possibilistic network G (characterized by its number of variables denoted $\#$ *variables*, the mean number of parents per node denoted μ *variables* and the mean domain size of variables μ *domain*), we generate many datasets (with different sizes). Regarding the dataset generation process, it consists in generating an imprecise dataset representative of its possibility distribution. The sampling process constructs a database of N (predefined) observations by instantiating all variables with respect to their possibility distributions using the α-*cut* notion expressed as follows:

$$\alpha - \text{cut}_{X_i} = \{x_i \in D_i \text{ such that } \pi(x_i) \geq \alpha\} \tag{8.12}$$

where α is randomly generated from $[0, 1]$. Obviously, variables are most easily processed with respect to a topological order, since this ensures that all parents are instantiated. Instantiating a parentless variable corresponds to computing its α-*cut*. Instantiating a conditioned variable X_i such that $par(X_i = X)$ corresponds to computing the α-*cut* of $\pi(X_i|par(X_i) = X)$ computed as follows:

$$\pi(X_i|par(X_i) = X) = \max_{x_i \in X}(\pi(X_i|x_i), \pi(x_i)) \tag{8.13}$$

Table 8.9 gives the details on the generated possibilistic networks and the corresponding datasets.

Name	# variables	μ parents	μ domain	# datasets
Net10	10	1.6	3.9	9
Net20	20	2.65	3.41	8
Net30	30	2.76	3.48	7

Table 8.9 – Datasets properties used in experiments 1.

Results

Table 8.10 gives the results of computing the similarity on each dataset $\mathcal{D}_i$, the possibilistic network G_i^{TA} (resp. G_i^{PA}) learned using the TA (resp. PA) approach with the reference network G_i used to generate D_i. The results of Table 8.10 show that on the one hand the learned possibilistic networks

Dataset	TA	PA
Net10	0.63	0.86
Net20	0.64	0.86
Net30	0.67	0.86

Table 8.10 – Results of experiments 1.

using the TA approach are close to the reference ones. Namely, they have rather a good similarity with the reference possibilistic networks used to generate the datasets. Moreover, the obtained similarity scores do not seem to be affected by the number of variables, variable domains size, etc. Regarding the possibilistic networks learned using the PA approach, their similarity scores are slightly better, but this is expected as the datasets generation process and the PA approach have the same view of possibility degrees. Such results also rise the issue of similarity measures on possibilistic networks which is still an open issue.

8.6.3 Predictive power of possibilistic classifiers

In this section, we evaluate the predictive power of credal network classifiers [CZ08], naive Bayes classifiers and naive possibilistic classifiers [BT12]. More precisely, we compare on many datasets the classification efficiency of naive credal classifier (NCC for short) and the corresponding possibilistic classifiers obtained either using the possibilistic-based approach (PNC_{PA}) or using the transformation-based approach (PNC_{TA}). Moreover, we compare our results to naive Bayes classifier (NBC) as a baseline.

Classification using belief graphical models is a special kind of inference: given an observation, it is required to determine the class label of the observed instance among a predefined set of class labels. In classification problems, one node represents the class variable C while the remaining ones are attributes $V = \{X_1, X_2, .., X_n\}$ that may be observable. Given an observation denoted $(x_1, x_2, .., x_n)$ of X, the candidate class c is predicted by possibilistic classifiers as follows:

$$c = argmax_{c_k \in D_C}(\Pi(c_k|x_1x_2..x_n)), \tag{8.14}$$

where the term $\Pi(c_k|x_1x_2..x_n)$ denotes the conditional possibility degree of having c_k the actual class given the observation x_1x_2,..,x_n.

A naive possibilistic (resp. Bayes) network classifier is a simple form of possibilistic (resp. Bayes) classifier. It assumes that attributes are independent in the context of the class node. Hence, the only dependencies allowed in naive networks are from the class node C to each attribute X_i. Learning a naive classifier in our context comes down to learning the local tables (namely the table of C and a conditional table for of each X_i in the context of C) from data since the structure is fixed in advance.

Experimentation setup

To evaluate the NCC classifier, we use measures used in [CZ08]:

— *Determinacy (Det):* It is the percentage of predictions outputting a unique (precise) class label.

— *Single-Accuracy (SiAcc):* It denotes the percentage of correct classifications when the predictions of NCC are precise.

— *Set-Accuracy (SetAcc)*: It is the proportion of imprecise predictions containing the right class label.

A 10-fold cross validation is used in this experiment.

Benchmarks

The experimental study is carried out on the following datasets where some data values are missing (here, missing data is assumed to be not missing at random). The first four datasets of Table 8.11 are real datasets used in the literature for evaluating classifiers with missing data [9]. The remaining ones are collected from different sources.

Name	# instances	# variables	# classes	% missing
breast	286	9	2	4 %
housevotes	435	16	2	24 %
mushroom	8124	22	2	31 %
post-operative	90	8	3	3 %
audiology	226	70	24	98%
sick	3772	30	2	20%
primary-tumor	339	18	21	46%
kr-vs-kp	3196	37	2	0 %
soybean	683	36	19	18%
crx	690	16	2	2%

Table 8.11 – Datasets used in our experiments.

9. http://sci2s.ugr.es/keel/missing.php

Results

Table 8.12 gives the results of evaluating the NCC classifier on the datasets of Table 8.11.

Dataset	Det	SiAcc	SetAcc
breast	92.43 %	74.08 %	100 %
housevotes	99.52 %	90.26 %	100 %
mushroom	96.10 %	99.56 %	100 %
post-operative	49.67 %	67.57 %	84.36 %
audiology	7.76%	99.55%	99.03%
sick	98.93%	97.54%	100%
primary-tumor	13.59%	77.11%	63.37%
kr-vs-kp	99.18%	88.16%	100%
soybean	47.38%	92.56%	97.85%
crx	94.01%	86.34%	100%

Table 8.12 – Results of NCC classifier on datasets of Table 8.11.

Table 8.12 shows good single accuracy rates with high determinacy rates except for the *post-operative*, *audiology* and *primary-tumor* datasets. Typically, it's on small datasets with many classes where the NCC is not efficient.

Table 8.13 gives the results of evaluating the NBC (Naive Bayes Classifier), PNC_{TA} and PNC_{PA} classifiers on the datasets of Table 8.11. Results of Table 8.13 show that classifiers NBC, PNC_{PA}

	% of correct classifications		
Dataset	NBC	PNC_{PA}	PNC_{TA}
breast	**72.88%**	72.73 %	70.27%
housevotes	**90.11 %**	89.19 %	58.71 %
mushroom	**95.73 %**	77.35 %	85.34 %
post-operative	68.11 %	67.78 %	**71.11%**
audiology	**72.79%**	55.90%	11.54%
sick	**96.97%**	95.53%	94.41%
primary-tumor	**49.54%**	28.42%	43.42%
kr-vs-kp	**87.82%**	85.86%	86.89%
soybean	**92.66%**	80.46%	75.51 %
crx	85.38%	85.80%	**91.01%**

Table 8.13 – Results of the NBC, PNC_S and PNC_T classifiers on the datasets of Table 8.11.

and PNC_{TA} have most of the time comparable results in terms of correct classification rates on some datasets but they show real performances on some other datasets. This is also valid for the results of the NCC classifier. Now, comparing PNC_{PA} and PNC_{TA}, this latter achieves better results on two datasets while the former has better classification rates on the two other datasets. It is not obvious what really makes a given approach better, a thorough analysis of the properties of the datasets is needed to help understanding such results.

8.7 Concluding remarks

The main objective of this chapter was to evaluate empirically *MAP* inference in credal networks using probability-possibility transformation. We provided a new approach to perform *MAP* inference in credal networks and this by transforming a credal network into a possibilistic one. We carried out experiments to compare our approach to an approximate approach for *MAP* inference in credal networks (*GL*). The benefits of such an approach are reducing the inference time of *MAP* inference while ensuring narrower answer sets. Experimental results showed that, first, using the approximate algorithm (*GL*) on credal networks was not computationally interesting due to the limits it has shown when the number of query variables increases. Then, when using criterion like *Hurwicz*, *CD* algorithm performed quite efficiently on numerous networks and numerous query variables. One thing that we have not been mentioning so far, is the complexity of our transformation *MD* and *CD*, this is to be taken into account when choosing an approach. And in this matter, *CD* is quite a direct translation and does not imply a high complexity, contrary to *MD* transformation. This supports even more the choice of *CD* that gives a good alternative to approximate *MAP* inference in credal networks.

The second part of the chapter tackles the problem of learning the parameters of a possibilistic network, we compare two methods for assessing the parameters of a possibilistic network given a structure and a dataset. The first series of experiments in our comparison mainly showed that the possibilistic-based method learns slightly better and more information in terms of information affinity than the method based on the probability-possibility transformation. This is not really surprising since the data was generated according to the possibility distributions of the reference networks. This also confirms that there is inevitably some information loss when transforming probability distributions into possibilistic ones [DFMP04, BLT15a]. Regarding the second series of experiments, given the results of *MAP* inference using transformed possibilistic networks, we evaluate the predictive power of classifiers. One important result is that the classifiers based on possibilistic networks have comparable efficiency with naive Bayes and credal classifiers. On the other hand, the possibilistic classifiers where the parameters have been learned with two different approaches have basically comparable results. Overall, such results are preliminary but encouraging, a further comparative study on a large number of benchmarks and problems (classification and inference in general) using naive and non naive models, will be needed to really compare the two approaches.

Conclusion

The works presented in this thesis contribute to the development of efficient formalisms to handle uncertain information. The first part of this thesis presented the foundations of various frameworks used to represent uncertainty and their compact representations by means of graphical models and knowledge bases. It also presented the different bridges existing between the uncertainty frameworks that are probability theory and possibility theory.

The second and third parts of this thesis presented our contributions. In the first part of the contributions, we dealt with conditioning in an interval-based possibilistic framework and set-valued possibilistic framework. The purpose was to develop a conditioning machinery for interval-based possibilistic logic. Conditioning in a standard possibilistic setting differs whether we consider a qualitative or quantitative scale. Our works dealt with both definitions of possibilistic conditioning. This led us to investigate a new extension of possibilistic logic, defined as set-valued possibilistic logic, and its conditioning machinery in the qualitative possibilistic setting. These results, especially in terms of complexity, led us to study transformations, more precisely from probability to possibility theories. The second part of our contributions dealt with probability-possibility transformation procedures. Indeed, we analyzed properties of reasoning tasks such as conditioning and marginalization. And to answer the second concern presented in the introduction, we tackled transformations from imprecise probability theory to possibility theory with a particular interest in *MAP* inference.

In more details, in Chapter 4, we first proposed natural properties (**IC1-IC7**) that any conditioning operator in an interval-based possibilistic setting should satisfy. Our first result showed that quantitative conditioning in an interval-based possibilistic setting satisfied all of our proposed postulates. We computed lower and upper bounds of the conditioned interval-based possibilistic distributions and provided the counterpart of conditioning in interval-based possibilistic logic. The surprising and interesting result was that reasoning with the set of all compatible possibilistic bases is not more expensive than reasoning from standard possibilistic bases. Hence, we extend standard possibilistic logic framework without extra computational cost.

However given the postulates **IC1-IC7**, we showed that qualitative conditioning (using $\min$-based operator) does not satisfy **IC1** meaning that the result of conditioning an interval-based possibility distribution is not guaranteed to give an interval-based possibility distribution. From there, we investigated new properties (**P1-P3**) for $\min$-based conditioning in an interval-based possibilistic setting. We proved that the only conditioning satisfying these three postulates was using the notion of Interval-closure definition on the set of conditioned compatible possibility distributions. We provided the efficient procedures to compute lower and upper bounds of the conditioned interval-based possibility distribution as well as the syntactic counterpart of qualitative conditioning. We showed that conditioning in interval-based possibilistic logic, like quantitative conditioning, had the same complexity as conditioning in standard possibilistic logic.

Using Interval-closure on all compatible distributions (resp. knowledge bases) can induce a larger imprecision in the result. To counter this issue, we proposed a new setting to represent possibility degrees named of **set-valued possibility theory**. This framework generalizes possibility theory by encoding the available knowledge using sets of possibility degrees. Clearly, set-valued possibility theory is also an extension of interval-based possibility theory [BHLR11], where the set is denoted as an interval of possible values. The two settings view a knowledge base (resp. possibility distribution) as a family of compatible bases (resp. distributions). Of course, intervals are particular sets. In this contribution, we defined the foundations of set-valued possibilistic setting (syntax and semantic aspects) and then we proposed a set of postulates (**S1-S3**) that characterizes a qualitative conditioning operator. These three postulates **S1**, **S2**, and **S3** guarantee the uniqueness of the conditioning operation. Lastly, we gave the procedures to compute the exact conditioned set-valued possibility distributions and the syntactic counterpart in set-valued possibilistic logic. As for conditioning in an interval-based possibilistic setting, we showed that conditioning in a set-valued possibilistic setting does not induce extra computational cost compared to conditioning in standard possibilistic setting.

In Chapter 7, the purpose was to study probability-possibility transformations with respect to reasoning tasks. We analyzed different properties such as the preservation of plausibility ordering between interpretations (resp. events) during a probability-possibility transformation, as well as the preservation of such order when applying reasoning tasks (e.g. conditioning, marginalization). In this work we have only considered probability-possibility transformations that satisfy the consistency principle defined by Dubois and Prade [DFMP04]. We showed that when one transforms a probability distribution into a possibility distribution then apply conditioning, the order between interpretations is preserved. We also analyzed imprecise probability to possibility transformations with respect to reasoning tasks (conditioning and marginalization). The last contribution concerned a study on MAP inference in credal networks. We highlighted new and interesting complexity results on MAP inference in the possibilistic setting. We showed that inference in possibilistic networks is less costly than in Bayesian networks. We also applied our results on MAP inference to classification by learning a credal network using probability-possibility transformations.

Among the open questions that can be addressed in future works is conditioning in another form of compact representations like interval-based possibilistic networks [BLT14a, BLT14b]. A first glimpse of a method to do so is discussed in Chapter 5 where we proposed a translation from interval-based possibilistic networks to interval-based possibilistic knowledge bases in the qualitative setting. We will tackle in details this type of transformation operation in both qualitative and quantitative settings.

Another open question concerns probability-possibility transformations. Indeed, we analyzed from probability theory to possibility theory. A future work will tackle properties-based analysis of transformations from possibility to probability. These transformations can also be applied to graphical models and knowledge bases. Indeed, numerous platforms exit for inference machinery in probabilistic setting and if possibility-probability transformations show good results on information preservation, then we could use these platforms to reason in possibility theory. Our works on transformations from probability to possibility were limited to distributions and graphical models. It could be interesting to apply probability-possibility transformations to knowledge bases and analyze the preservation of the ordering between interpretations with respect to reasoning tasks.

In this thesis we only experiments MAP inference in a quantitative possibilistic setting (using the product operator), a comparison between the two possibilistic settings (qualitative vs. quantitative) should be considered in future works. We plan to investigate new algorithms for *MAP* inference in possibilistic networks. We argued that contrary to the probabilistic setting, *MAP* inference in possibilistic networks have better computational complexity than belief probabilistic networks. This will definitely open new perspectives for approximate *MAP* inference in Bayesian credal networks.

Appendix

Appendix A

Background notions

A.1 Graphical models

In this section we introduce some elementary concepts of graph theory that are needed in the understanding of the graphical models used to represent uncertain information.

A.1.1 Basic definitions of graphs

Assume $V = \{X_1, X_2, ..., X_n\}$ is the set of variables. The set V can be graphically represented by a set of *nodes*, or *vertices*, one node for each element of V. These nodes can be linked by arcs which are referred to as *edges*. The set of all links is denoted by $E = \{(X_i, X_j)|X_i \text{ and } X_j \text{ are linked }\}$ where (X_i, X_j) is used to denote the link between X_i and X_j. The sets V and E define a graph. We first give an example, then a more formal definition of a graph.

Example A.4. Let us consider a graph G on the set of variables $V = \{X_1, X_2, X_3\}$ and the following set of edges $E = \{(X_1, X_2), (X_3, X_2)\}$.

Definition A.4 (Graph or network)**.** A graph $G = (V, E)$ is defined by two sets V and E, where V is a finite set of nodes $V = \{X_1, X_2, ..., X_n\}$ and E is a set of edges, that is, a subset of ordered pairs of distinct nodes.

The words "graph" and "network" are used synonymously in this thesis. The links of a graph can be *directed* or *undirected*, depending on whether or not the direction of the link matters. We only focus on directed graphs and therefore we discuss the definition and some of their characteristics.

Definition A.5 (Directed link)**.** Let $G = (V, E)$ be a graph. When $(X_i, X_j) \in E$ and $(X_j, X_i) \notin E$, the link (X_i, X_j) is called a directed link. A directed link between nodes X_i and X_j is denoted by $X_i \to X_j$ where X_i is called parent and X_j is called child.

Definition A.6 (Directed graph)**.** A graph in which all the links are directed is called a directed graph.

Example A.5. In this example, we illustrate the graph defined in Example A.4 as a directed graph on Figure A.1.

Definition A.7 (Parents and children)**.** When there is a directed link $X_i \to X_j$ from X_i to X_j, then X_i is said to be a parent of X_j, and X_j is said to be a child of X_i.

The set of all parents of a given node X_i is denoted by $par(X_i)$. For instance, in Example A.5, more precisely on Figure A.1, the parents of node X_2 is given by $par(X_2) = \{X_1, X_3\}$.

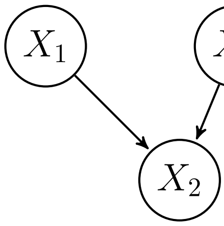


Figure A.1 – Example of directed network

Directed graphs

Definition A.8 (Connected directed graphs)**.** A directed graph is said to be *connected* if each node is connected to another by at least one link. Otherwise, it is said to be *disconnected*

Definition A.9 (Trees and multiply-connected directed graphs)**.** A connected directed graph is said to be a *tree* if between every pair of nodes there exists a unique path [10]. Otherwise, it is said to be *multiply-connected*.

There also exist two different types of trees in directed graphs depending on the number of arrows pointing to the same node.

Definition A.10 (Cyclic and acyclic graphs)**.** A directed graph is said to be *cyclic* if it contains at least one cycle [11]. Otherwise, if it is called a *directed acyclic graph* (*DAG*).

Directed acyclic graphs play an important role as they are used as a basis for building the most known uncertainty networks as *Bayesian networks*, possibilistic networks or even credal networks.

Definition A.11 (Simple trees and polytrees)**.** A directed tree is called a *simple tree* if every node has at most one parent. Otherwise, it is called a *polytree*.

Example A.6. In the following figure we show an example of two different directed graphs, a polytree and a multi-connected tree.

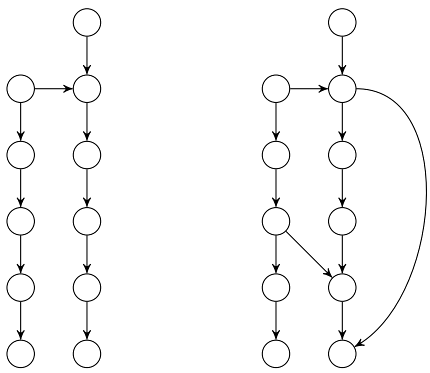


Figure A.2 – On the left: polytree, and on the right: multiply-connected tree

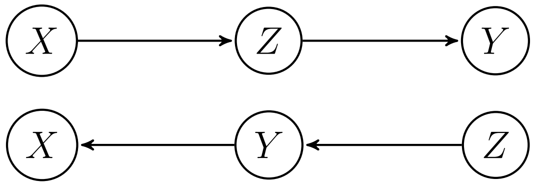
10. A path is finite sequence of edges which connect a sequence of nodes with the edges all directed in the same direction.
11. where a *cycle* is defined as a closed directed path in a graph.

A.1.2 Graph separation

Some properties in graphs allow to characterize the flow of the information in a Bayesian network. This is called *D-separation*, it answers the question: *How an observation of one or many variables affects the belief we have on other variables?* and it is based on the concept of independence.

Let us consider two subset of random variables X and Y and another subset of random variables Z then X and Y are d-separated by Z if and only if Z blocks every paths between X and Y. There can exist different types of connexions between these subsets of variables.

— serial connexion



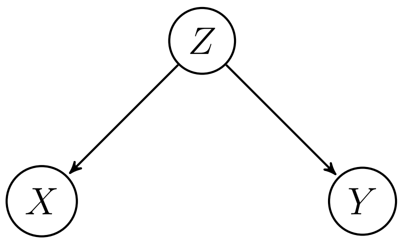
The flow of information between X and Y is blocked by Z.

— $X \not\perp Y$ (X and Y are not independent)

— $X \perp Y|Z$ (if Z is known then X and Y are independent)

— $Y \perp X|Z$ (conversely, Y and X are independent given Z)

— divergent connexion



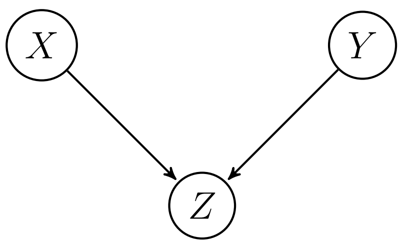
The flow of information between X and Y is blocked by Z.

— $X \not\perp Y$ (X and Y are not independent)

— $X \perp Y|Z$ (if Z is known then X and Y are independent)

— $Y \perp X|Z$ (if Z is known then Y and X are independent)

— convergent connexion



The flow of information between X and Y is not blocked by Z.

— $X \perp Y$ (X and Y are independent)

— $X \not\perp Y|Z$ (if Z is known then X and Y are not independent)

— $Y \not\perp X|Z$ (conversely, knowing Z, Y and X are not independent)

A.2 Querying graphical models

A.2.1 Probability of evidence

One of the simplest queries is to ask for the degree or interval-degree of some variable instantiation e. For example [12], in the Asia network (Figure A.3) we may be interested in knowing the probability that the patient has a positive x-ray but no dyspnoea, $P(X = yes, D = no)$. This can be computed easily by tools such as JavaBayes, leading to a probability of about 3.96%. The variables $E = \{X, D\}$ are called *evidence variables* in this case and the query $P(e)$ is known as a *probability-of-evidence* query, although it refers to a very special type of evidence corresponding to the instantiation of some variables.

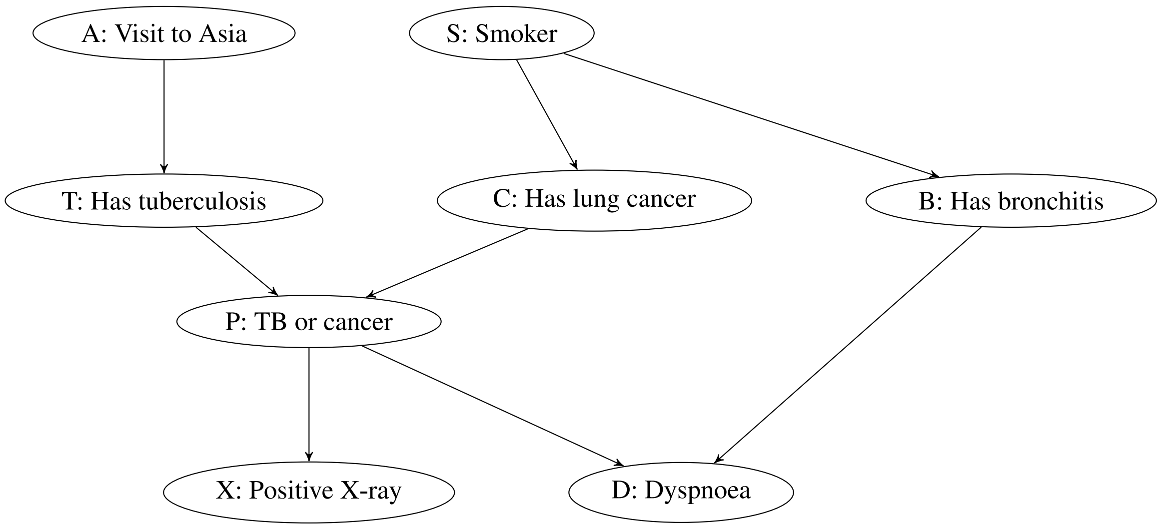


Figure A.3 – Asia network where each variable is boolean "yes" or "no"

There are other types of evidence beyond variable instantiations. In fact, any propositional sentence can be used to specify evidence. For example, we may want to know the probability that the patient has either a positive x-ray or a dyspnoea, $X = yes \vee D = yes$. Tools in general do not provide direct support for computing the probability of arbitrary pieces of evidence but such probabilities can be computed indirectly using the following method.

We can add an auxiliary node E to the network, declare nodes X and D as the parents of E, and then adopt the following conditional probability distribution on E (see Table A.1).

X	D	E	$p(e\|x, d)$
yes	yes	yes	1
yes	no	yes	1
no	yes	yes	1
no	no	yes	0

Table A.1 – Conditional probability distribution of E given X and D

Given this conditional probability distribution, the event $E = yes$ is then equivalent to $X = yes \vee D = yes$ and hence, we can compute the probability of the latter by computing the probability of the

12. Note that the examples taken in this section come from the book "Modeling and reasoning with Bayesian network" [Dar09]

former.

This method, known as *auxiliary-node* method, is practical only when the number of evidence variables is small enough, as the conditional probability distribution size grows exponentially in the number of these variables. However, this type of conditional probability distribution is quite special as it only contains probability degree equal to 0 or 1. When a conditional probability distribution satisfies this property, we say that it is *deterministic*. We also refer to the corresponding node as a *deterministic node*.

A.2.2 Prior and posterior marginals

If probability-of-evidence queries are one of the simplest, then *posterior-marginal queries* are one of the most common. The difference between *prior* and *posterior marginals* is that a prior marginal is a marginal distribution given no evidence. And the posterior marginal distribution is computed given some evidence e,

$$p(x_1, ..., x_m|e) = \sum_{x_{m+1},\ldots,x_n} p(x_1, ..., x_n|e). \tag{A.1}$$

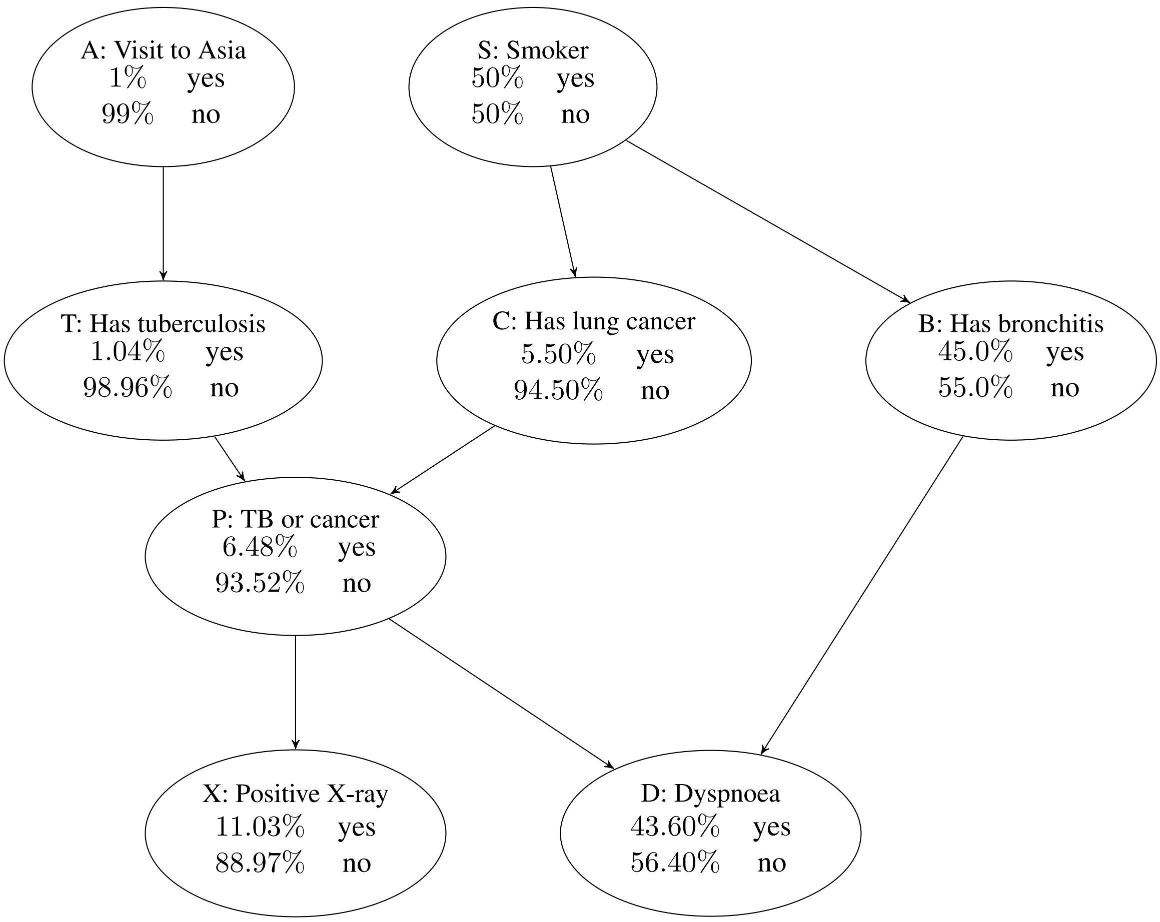


Figure A.4 – Prior marginals in the Asia network

Figure A.4 depicts a network where the prior marginals are shown for every variable in the network. Figure A.5 depicts also the Asia network but where posterior marginals are shown for every variable given that the patient has a positive x-ray but no dyspnoea, $(e : X = yes,\ D = no)$.

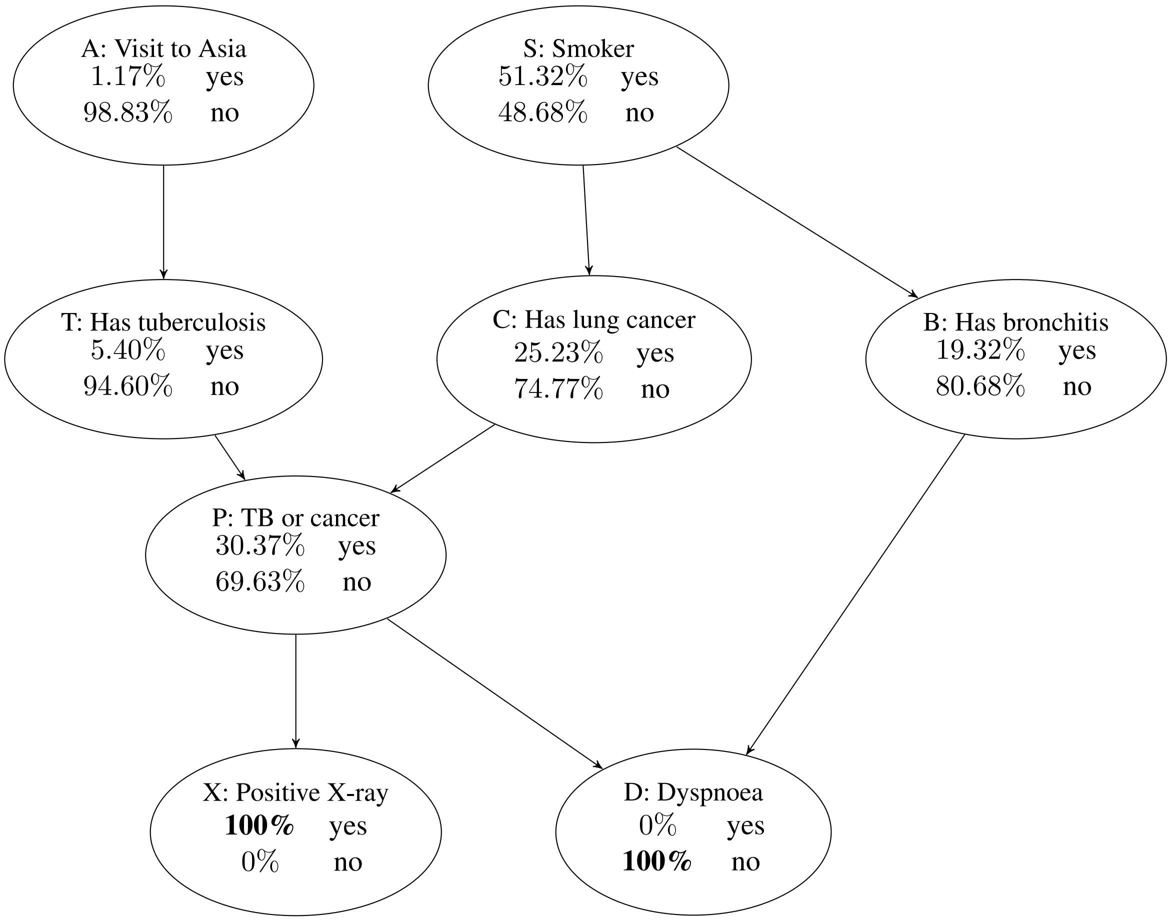


Figure A.5 – Posterior marginals in the Asia network given a positive x-ray and no dyspnoea

A.2.3 Most probable explanations

We now turn to another class of queries: computing the *most probable explanation* (MPE). The goal here is to identify the most probable instantiation of network variables given some evidence. Specially, if $X_1, ..., X_n$ are all the network variables and if e is the given evidence, the goal then is to identify an instantiation $x_1, ..., x_n$ for which the probability $p(x_1, ..., x_n|e)$ is maximal. Such an instantiation $x_1, ..., x_n$ will be called a *most probable explanation* given evidence e.

Consider the Asia network still with the evidence $e : X = yes,\ D = yes$ (a patient with positive x-ray and dyspnoea), if we compute the MPE queries given e, then the result corresponds to a patient that made no visit to Asia, is a smoker, and has lung cancer and bronchitis but no tuberculosis.

It is important to note here that an MPE cannot be obtained directly from posterior marginals. That is, if $x_1, ..., x_n$ is an instantiation obtained by choosing each value x_i so as to maximize the probability $p(x_i|e)$, then $x_1, ..., x_n$ is not necessarily an MPE. Consider the posterior marginals in Figure A.5 as an example. If we choose for each variable the value with maximal probability, we get an explanation in which the patient is a smoker:

$$\omega : A = no,\ S = yes,\ T = no,\ C = no,\ B = no,\ P = no,\ X = yes,\ D = no.$$

This instantiation has a probability of
$approx 20.03\%$ given the evidence $e : X = yes,\ D = no$. However, the most probable explanation

given e is the one in which the patient is not a smoker:

$$\omega : A = no,\ S = no,\ T = no,\ C = no,\ B = no,\ P = no,\ X = yes,\ D = no.$$

This instantiation has a probability of $\approx 38.57\%$ given evidence $e : X = yes,\ D = no.$

A.2.4 Maximum a posteriori hypothesis (MAP)

The MPE query is a special case of a more general class of queries for finding the most probable instantiation of a subset of network variables. Specially, suppose that the set of all network variables is V and let Q be a subset of these variables. Given some evidence e, our goal is then to find an instantiation q of variables Q for which the probability $p(q|e)$ is maximal. Any instantiation q that satisfies the previous property is known as a *maximum a posteriori hypothesis* (MAP). Moreover, the variables in Q are known as MAP variables or query variables. Clearly, MPE is a special case of MAP when the MAP variables include all network variables. One reason why a distinction is made between MAP and MPE is that MPE is much easier to compute algorithmically. An issue that we will address in the following.

Using the same Asia network from Figure A.3, let us consider a patient with a positive x-ray and no dyspnoea, so the evidence is $X = yes,\ D = no$. The MAP variables are $Q = \{A, S\}$, so we want to know the most likely instantiation of these variables given the evidence. Given the evidence e, the result of MAP query is

$$A = no,\ S = yes.$$

This instantiation have a probability degree of $\approx 50.74\%$ given the evidence.

A common method for approximating MAP is to compute an MPE and then return the values it assigns to MAP variables. We say in this case that we are *projecting* the MPE on MAP variables. However, we stress that this is only an approximation scheme as it may return an instantiation of the MAP variables that is not maximally probable. Consider again the MPE example where the MPE instantiation was:

$$\omega : A = no,\ S = no,\ T = no,\ C = no,\ B = no,\ P = no,\ X = yes,\ D = no.$$

under $X = yes,\ D = no$. Projecting this MPE on the variables $Q = \{A, S\}$, we get the instantiation

$$A = no,\ S = no,$$

which has a probability $\approx 48.09\%$ given the evidence. This instantiation is clearly not a MAP as we found a more probable instantiation earlier, that is, $A = no,\ S = yes$ with a probability of about 50.74%.

There is a relatively general class of situations in which the solution to a MAP query can be obtained immediately from an MPE solution by projecting it on the MAP variables. To formally define this class of situations, let E be the evidence variables, Q be the MAP variables, and R be all other network variables. The condition is that there is at most one instantiation r of variables R that is compatible with any particular instantiations q and e of variables Q and e, respectively. More formally, if $p(q, e) > 0$, then $p(q, e, r) \neq 0$ for exactly one instantiation r.

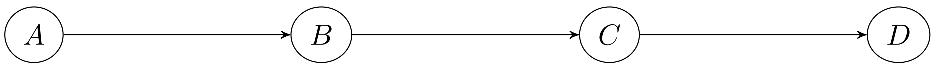
A.3 Different algorithms for inference in Bayesian networks

A.3.1 Efficient inference in trees - variable elimination

Let us illustrate the process of variable elimination through an Bayesian example. Let us consider the Bayesian network in Figure A.7 with $D_A = \{a, \overline{a}\}$, $D_B = \{b, \overline{b}\}$, $D_C = \{c, \overline{c}\}$ and $D_D = \{d, \overline{d}\}$

and suppose that we are interested in computing the degree of $D = d$, meaning computing $P(d)$.

$$\begin{array}{rcl} P(d) & = & \sum_{A,B,C}(P(ABCd)) \\ & = & \sum_{A,B,C}(P(d|C) * P(C|B) * P(B|A) * P(A)) \\ & = & \sum_{B,C}(P(d|C) * P(C|B) * \sum_{A} P(B|A) * P(A)) \\ & = & \sum_{B,C}(P(d|C) * P(C|B) * P(B)) \\ & = & \sum_{C}(P(d|C) * \sum_{B} P(C|B) * P(B)) \\ & = & \sum_{C}(P(d|C) * P(C)) \\ & = & P(d) \end{array} \quad . \tag{A.2}$$



A	$p(A)$
a	.4
$\overline{a}$	.6

A	B	$p(B\|A)$
a	b	.2
a	$\overline{b}$	.8
$\overline{a}$	b	.5
$\overline{a}$	$\overline{b}$	.5

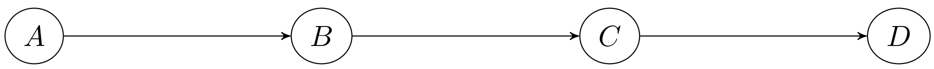
B	C	$p(C\|B)$
b	c	.6
b	$\overline{c}$	.4
$\overline{b}$	c	.2
$\overline{b}$	$\overline{c}$	.8

C	D	$p(D\|C)$
c	d	.2
c	$\overline{d}$	.8
$\overline{c}$	d	.6
$\overline{c}$	$\overline{d}$	.4

Figure A.6 – Network on which to perform variable elimination

What is the probability of the event $\phi = \{\overline{d}\}$ (*i.e.* $P(\overline{d})$)? The first step is to merge the table of variable A with the table of variable $B|A$ to remove the variable A.

$$f_B(B) = \sum_{A} P(B|A) * P(A).$$



A	$p(A)$
a	.4
$\overline{a}$	.6

A	B	$p(B\|A)$
a	b	$.2 * .4 = .08$
a	$\overline{b}$	$.8 * .4 = .32$
$\overline{a}$	b	$.5 * .6 = .3$
$\overline{a}$	$\overline{b}$	$.5 * .6 = .3$

B	C	$p(C\|B)$
b	c	.6
b	$\overline{c}$	.4
$\overline{b}$	c	.2
$\overline{b}$	$\overline{c}$	.8

C	D	$p(D\|C)$
c	d	.2
c	$\overline{d}$	.8
$\overline{c}$	d	.6
$\overline{c}$	$\overline{d}$	.4

Figure A.7 – Elimination of variable A

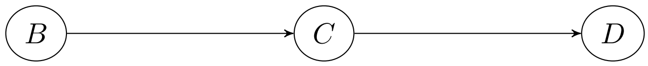
The second step consists in merging the distribution of variable B with the distribution of C to eliminate variable B.

$$f_C(C) = \sum_{B} P(C|B) * P(B).$$

We keep eleminating variables in the way until we are left with our variable of interest D. Indeed, we eliminate C by combining tables of variables C and D.

$$f_D(D) = \sum_{C} P(D|C) * P(C).$$

Table A.2 depicts the final distribution on D after the elimination of the other variables, and for the result of our query $D = \overline{d}$, then $P(\overline{d}) = .54$.



B	$p(B)$
b	.38
$\bar{b}$	.62

B	C	$p(C\|B)$
b	c	$.6 * .38 = .228$
b	$\bar{c}$	$.4 * .38 = .152$
$\bar{b}$	c	$.2 * .62 = .124$
$\bar{b}$	$\bar{c}$	$.8 * .62 = .496$

C	D	$p(D\|C)$
c	d	.2
c	$\bar{d}$	.8
$\bar{c}$	d	.6
$\bar{c}$	$\bar{d}$	.4

Figure A.8 – Elimination of variable B

D	$p(D)$
d	.46
$\bar{d}$	.54

Table A.2 – Distribution of D after elimination of A, B and C

One could have simply constructed the joint distribution of the network and marginalized on the variables of interest to obtain the marginal distribution as requested. The fact is doing so we would have computed unnecessary information. Variable elimination in this sense, in a slightly intelligent way to sum out variables from the joint without actually constructing its explicit representation.

In terms of complexity in Bayesian networks, we can distinguish different cases depending on the network structure:

— in trees: the complexity is linear in the number of variables

— in polytrees: the complexity of variable elimination is exponential in the tree width

— complexity of the best elimination order: in the general case we always need to fix an order for the variable elimination which is a NP-complete problem.

A.3.2 The junction tree algorithm

The junction tree algorithm is one of the most used algorithms and among the most efficient ones. It generalizes most of the inference algorithms existing in Bayesian networks. It was developed by Jensen, Lauritzen and Spiegelhalter [Jen96, LS90]. This algorithm is based on both graph theory and probability theory to find a efficient factorization of the probability distributions encoded by a Bayesian network. This algorithm translates the initial Bayesian network into a structure called junction tree which we use to answer queries. It is composed of 3 main steps:

— **step 1**: Construction

  — moralization of the initial graph

  — triangulation of the moral graph

  — construction of the junction tree

— **step 2**: Initialization

— **step 3**: Answering the queries

These steps are the same in both probabilistic and possibilistic settings. Let us expand a little theses steps:

Moralization: A moral graph is a graph where all the parents of a node are connected through a link.

Figure A.9 – Moralization step

For this step, it is enough to create first a non directed graph by discarding the edges then "marry" the parents of each nodes in the graph.

Triangulation: undirected graph is said to be triangulated if and only if for every cycle of length 4 or more, there exists a link between two non consecutive nodes of the cycle.

Triangulation consist in adding links until the graph is triangulated. A graph can have different triangulations which gives different set of cliques. A clique in that respect is a subset of variables forming a complete subgraph meaning that two given variables are always connected. Moreover, the triangulated graph can have, also, multiple junction tree. The problem of optimality for triangulation is NP-complete [ref].

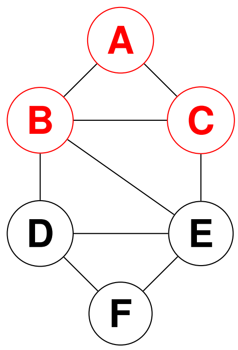


Figure A.10 – Triangulation step giving the cliques: $C_{ABC}, C_{BCE}, C_{BDE}, C_{DEF}$

Construction of the junction tree: This step consists in transforming the triangulated graph into a junction tree. Following the triangulation, the graph is composed of cliques denoted by C_i. Note that a junction tree built from a triangulated graph is not unique. For this step, it is enough to connect the identified cliques with the condition that every clique between two other cliques C_i and C_j must contain $C_i \cap C_j$. Once all adjacent cliques are identified, we insert a separator denoted by S_{ij}, between each pair of cliques C_i and C_j, containing the shared variables.

Once the junction tree build, the initialization step consists in quantifying the tree by transforming local distributions in joint distribution attached to cliques and separators.

Broadly speaking, the initialization procedure is resumed as:

— for each clique C_i, affect a uniform distribution: $\theta^1_{C_i} = 1$

— for each separator S_{ij}, affect a uniform distribution: $\theta^1_{S_{ij}} = 1$

— for each variable $X_k \in V$, choose a clique C_i containing $\{X_k\} \cup par(X_k)$ and update the capacities.

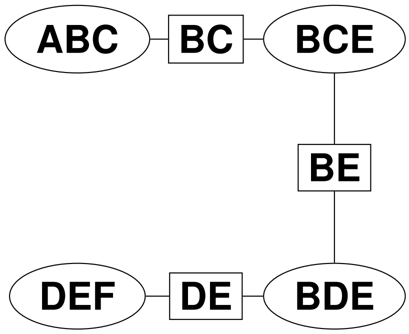


Figure A.11 – Junction tree associated to the graph of Figure A.9

Note that if a clique has too many nodes, the computation of the joint distribution can be costly.

Now comes the propagation step, before that let us introduce the notion of global coherence in junction tree. Let C_i and C_j two adjacent cliques in the junction tree and let S_{ij} be their separator. The cliques C_i and C_j are said to be stable or consistent if:

$$\forall s_{ij} \in D_{S_{ij}}, \sum_{C_i \backslash S_{ij}} \theta_{C_i}(c_i) = \theta_{S_{ij}} = \sum_{C_j \backslash S_{ij}} \theta_{C_j}(c_j) \quad (A.3)$$ [13]

If every link in a junction tree is coherent then the junction tree is said to be globally coherent. Once the junction tree initialized, global propagation ensures the global coherence through the message passing between the cliques. The first step is to arbitrary choose a *pivot* clique which will start the two phases process:

— collection phase: each clique sends a message to its adjacent cliques towards the pivot.

— distribution phase: each clique send a message in the opposite way of the pivot to its adjacent cliques starting from the pivot itself until reaching one of the end of the tree.

This is a brief description of the steps to follow to build the junction tree.

A.4 A brief refresher on propositional logic

Propositional logic is used as a tool for representing and reasoning about events. Let us take an example and consider the following situation that involves an alarm that can be triggered by burglaries, or an earthquake. We can express the event of having either a burglar or an earthquake. This is written using the following propositional formula:

$$\text{Burglary} \lor \text{Earthquake}.$$

In this situation, *Burglary* and *Earthquake* are called *propositional variables* and $\lor$ represents logical disjunction. With propositional logic we can make more complicated statement, such as:

$$\text{Burglary} \lor \text{Earthquake} \Rightarrow \text{Alarm},$$

where $\Rightarrow$ describes logical implication. Meaning that a burglary or an earthquake is guaranteed to turn on the alarm. Consider also the formula:

$$\neg\text{Burglary} \land \neg\text{Earthquake} \Rightarrow \neg\text{Alarm},$$

13. This is valid in the probabilistic setting, for the possibilistic setting, we would have the maximum operator instead of the summation for instance.

where $\neg$ represents the logical negation and $\wedge$ represent the logical conjunction. And this means that if there is no burglary nor earthquake, the alarm will not trigger.

Overall, propositional formulas are formed using a set of propositional variables, a, b, c. These variables are *boolean* variables. A formula φ can be of two forms, either given by an atomic formula of the form a_i, and interpreted as the variable a_i is *true*. The second form of formula that can be written in propositional logic is by using the logical connectives $\neg$, $\vee$ and $\wedge$. If φ and τ are formulas, then $\neg\varphi$, $\varphi \vee \tau$ and $\varphi \wedge \tau$ are formulas. These formulas composed the language $\mathcal{L}$.

Other connectives can be introduced, such as implication $\Rightarrow$ and equivalence $\Leftrightarrow$. These connectives can be expressed using the three previous ones, in particular the formula $\varphi \Rightarrow \tau$ can be written as $\neg\varphi \vee \tau$. Also, $\varphi \Leftrightarrow \tau$ can be written as $(\varphi \Rightarrow \tau) \vee (\tau \Rightarrow \varphi)$. With this preliminaries we can define a knowledge base.

Definition A.12 (Knowledge base)**.** A propositional *knowledge base*, denoted K is a set of propositional formulas $\varphi_1, ..., \varphi_n$, that is interpreted as a conjunction $\varphi_1 \wedge \varphi_2 \wedge ... \wedge \varphi_n$.

Here are some properties:

— $\varphi \wedge \neg\varphi$ is a contradiction (will never hold).

— $\varphi \vee \neg\varphi$ is a tautology (always holds).

— φ and $\varphi \Rightarrow \tau$ implies τ.

— $\varphi \vee \tau$ is equivalent to $\tau \vee \varphi$.

The semantic associated to a propositional knowledge base is the one of models.

Definition A.13 (Model of a formula)**.** Given a formula φ, a model ω is a mapping of all propositional variables to a truth value such as $\omega \models \varphi$. We say that ω satisfies (or entails) the event φ.

Now we say that a world ω is a *model* of a knowledge base K if it is a model of all the formulas in K. A *world* is the equivalent of a configuration which means that the value of each propositional variable is known. This definition stands in the same line as the definition of world we have seen for uncertain theories in the previous chapter.

The set of worlds that satisfy φ is denoted by:

$$Mods(\varphi) = \{\omega : \omega \models \varphi\}.$$

Bibliography

[ACdCT14] Thomas AUGUSTIN, Frank COOLEN, Gert de COOMAN, and Matthias TROFFAES, editors. Introduction to Imprecise Probabilities. Wiley, 2014.
Cited page(s) 95

[ACZ10] Alessandro ANTONUCCI, Giorgio CORANI, and Marco ZAFFALON. « Bayesian Networks with Imprecise Probabilities: Theory and Application to Classification ». Technical report, IDSIA - Dalle Molle Institute for Artificial Intelligence, 2010.
Cited page(s) 36

[Ada66] Ernest W. ADAMS. Probability and the Logic of Conditionals. In Jaakko HINTIKKA and Patrick SUPPES, editors, Aspects of Inductive Logic, volume 43 of Studies in Logic and the Foundations of Mathematics, pages 265 – 316. Elsevier, 1966.
Cited page(s) 1

[Bay63] Thomas BAYES. « An essay towards solving a problem in the doctrine of chances ». Phil. Trans. of the Royal Soc. of London, 53:370–418, 1763.
Cited page(s) 27

[BCD07] Cédric BAUDRIT, Ines COUSO, and Didier DUBOIS. « Joint propagation of probability and possibility in risk analysis: Towards a formal framework ». International Journal of Approximate Reasoning, 45(1):82 – 105, 2007.
Cited page(s) 3

[BDCPT13] Salem BENFERHAT, Célia DA COSTA PEREIRA, and Andrea TETTAMANZI. « Syntactic Computation of Hybrid Possibilistic Conditioning under Uncertain Inputs ». In IJCAI, Page 6822, Beijing, China, August 2013. AAAI.
Cited page(s) 2, 17

[BDGP02] Salem BENFERHAT, Didier DUBOIS, Laurent GARCIA, and Henri PRADE. « On the transformation between possibilistic logic bases and possibilistic causal networks ». International Journal of Approximate Reasoning, 29(2):135 – 173, 2002.
Cited page(s) 32, 34, 81, 83

[BDKP13] Asma BELHADI, Didier DUBOIS, Faiza KHELLAF-HANED, and Henri PRADE. « Multiple agent possibilistic logic ». Journal of Applied Non-Classical Logics, 23(4):299–320, 2013.
Cited page(s) 2, 3

[BDP99] Salem BENFERHAT, Didier DUBOIS, and Henry PRADE. « Possibilistic and Standard Probabilistic Semantics of Conditional Knowledge Bases ». J. Log. Comput., 9(6):873–895, 1999.
Cited page(s) 122

[Ben10] Salem BENFERHAT. « Graphical and Logical-Based Representations of Uncertain Information in a Possibility Theory Framework ». In Amol DESHPANDE and Anthony

HUNTER, editors, Scalable Uncertainty Management - 4th International Conference, SUM 2010, Toulouse, France, September 27-29, 2010. Proceedings, volume 6379 of Lecture Notes in Computer Science, pages 3–6. Springer, 2010.

Cited page(s) 31

[BHK14] Christoph BEIERLE, Rita HERMSEN, and Gabriele KERN-ISBERNER. « Observations on the Minimality of Ranking Functions for Qualitative Conditional Knowledge Bases and Their Computation ». In William EBERLE and Chutima BOONTHUM-DENECKE, editors, Proceedings of the Twenty-Seventh International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014, Pensacola Beach, Florida, May 21-23, 2014. AAAI Press, 2014.

Cited page(s) 4, 24

[BHLR11] Salem BENFERHAT, Julien HUÉ, Sylvain LAGRUE, and Julien ROSSIT. « Interval-Based Possibilistic Logic ». In IJCAI 2011, Proceedings of the 22nd International Joint Conference on Artificial Intelligence, Barcelona, Catalonia, Spain, July 16-22, 2011, pages 750–755, 2011.

Cited page(s) 2, 3, 5, 50, 51, 81, 86, 146

[BK02] Christian BORGELT and Rudolf KRUSE. Graphical models - methods for data analysis and mining. Wiley, 2002.

Cited page(s) 28

[BK03] Christian BORGELT and Rudolf KRUSE. « Learning possibilistic graphical models from data ». Fuzzy Systems, IEEE Transactions on, 11(2):159–172, Apr 2003.

Cited page(s) 2

[BLT14a] Salem BENFERHAT, Sylvain LAGRUE, and Karim TABIA. « Analysis of interval-based possibilistic networks ». In Torsten SCHAUB, Gerhard FRIEDRICH, and Barry O'SULLIVAN, editors, ECAI 2014 - 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic - Including Prestigious Applications of Intelligent Systems (PAIS 2014), volume 263 of Frontiers in Artificial Intelligence and Applications, pages 963–964. IOS Press, 2014.

Cited page(s) 80, 146

[BLT14b] Salem BENFERHAT, Sylvain LAGRUE, and Karim TABIA. « Interval-Based Possibilistic Networks ». In Umberto STRACCIA and Andrea CALÌ, editors, Scalable Uncertainty Management - 8th International Conference, SUM 2014, Oxford, UK, September 15-17, 2014. Proceedings, volume 8720 of Lecture Notes in Computer Science, pages 37–50. Springer, 2014.

Cited page(s) 80, 146

[BLT15a] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « Probability-Possibility Transformations: Application to Credal Networks ». In SUM 2015, Québec City, Canada, pages 203–219. Springer, 2015.

Cited page(s) 5, 6, 125, 144

[BLT15b] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « Conditionnement en Logique Possibiliste à Intervalles ». In LFA 2015 - 24ème Conférence sur la Logique Floue et ses Applications, 5-6 November 2015, Poitiers, France, 2015.

Cited page(s) 4, 6

[BLT15c] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « On the Analysis of Probability-Possibility Transformations: Changing Operations and Graphical Models ».

In ECSQARU 2015, Compiegne, France, July 15-17, 2015. Proceedings, 2015.
Cited page(s) 5, 6

[BLT15d] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « Transformations probabilistes-possibilistes: conditionnement, inférence et modl̀es graphiques ». In JIAF 2015 - Neuvième Journées d'Intelligence Artificielle Fondamentale, 29 June-3 July 2015, Rennes, France, 2015.
Cited page(s) 5, 6

[BLT17a] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « Approximating MAP inference in credal networks using probability-possibility transformations ». In ICTAI 2017 - 29th International Conference on Tools with Artificial Intelligence, Boston, MA, USA, 2017. To appear.
Cited page(s) 5, 6

[BLT17b] Salem BENFERHAT, Amélie LEVRAY, and Karim TABIA. « Approximation de l'inférence MAP via les transformations probabilistes-possibilistes ». In JIAF 2017 - Onzième Journées d'Intelligence Artificielle Fondamentale, 3-7 July 2017, Caen, France, 2017.
Cited page(s) 5, 6

[BLTK15] Salem BENFERHAT, Amélie LEVRAY, Karim TABIA, and Vladik KREINOVICH. « Compatible-Based Conditioning in Interval-Based Possibilistic Logic ». In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2777–2783, 2015.
Cited page(s) 4, 6

[BLTK16] Salem BENFERHAT, Amélie LEVRAY, Karim TABIA, and Vladik KREINOVICH. « Set-Valued Conditioning in a Possibility Theory Setting ». In Gal A. KAMINKA, Maria FOX, Paolo BOUQUET, Eyke HÜLLERMEIER, Virginia DIGNUM, Frank DIGNUM, and Frank van HARMELEN, editors, ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016), volume 285 of Frontiers in Artificial Intelligence and Applications, pages 604–612. IOS Press, 2016.
Cited page(s) 5, 6

[Bou07] Mohamed Said BOUGUELID. Contribution à l'application de la reconnaissance des formes et la théorie des possibilités au diagnostic adaptatif et prédictif des systèmes dynamiques. Université de Champagne-Ardenne, Decembre 2007.
Cited page(s) 44

[BP05] Salem BENFERHAT and Henri PRADE. « Encoding formulas with partially constrained weights in a possibilistic-like many-sorted propositional logic ». In IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30-August 5, 2005, pages 1281–1286, 2005.
Cited page(s) 2, 3

[BSK09] Christian BORGELT, Matthias STEINBRECHER, and Rudolf R. KRUSE. Graphical models: representations for learning, reasoning and data mining, volume 704. Wiley, 2009.
Cited page(s) 140

[BT10] Salem BENFERHAT and Karim TABIA. « Belief Change in OCF-Based Networks in Presence of Sequences of Observations and Interventions: Application to Alert Correlation ». In Byoung-Tak ZHANG and Mehmet A. ORGUN, editors, PRICAI 2010:

Trends in Artificial Intelligence, 11th Pacific Rim International Conference on Artificial Intelligence, Daegu, Korea, August 30-September 2, 2010. Proceedings, volume 6230 of Lecture Notes in Computer Science, pages 14–26. Springer, 2010.

Cited page(s) 24

[BT12] Salem BENFERHAT and Karim TABIA. « Inference in possibilistic network classifiers under uncertain observations ». Annals of Mathematics and Artificial Intelligence, 64(2-3):269–309, 2012.

Cited page(s) 141

[BTM99] Bernard De BAETS, Elena TSIPORKOVA, and Radko MESIAR. « Conditioning in possibility theory with strict order norms ». Fuzzy Sets and Systems, 106(2):221 – 229, 1999.

Cited page(s) 2

[BTS10] Salem BENFERHAT, Karim TABIA, and Karima SEDKI. « On analysis of unicity of Jeffrey's rule of conditioning in a probabilistic framework ». In International Symposium on Artificial Intelligence and Mathematics, ISAIM 2010, Fort Lauderdale, Florida, USA, January 6-8, 2010, 2010.

Cited page(s) 13

[BTS11] Salem BENFERHAT, Karim TABIA, and Karima SEDKI. « Jeffrey's rule of conditioning in a possibilistic framework - An analysis of the existence and uniqueness of the solution ». Ann. Math. Artif. Intell., 61(3):185–202, 2011.

Cited page(s) 18

[CD05] Hei CHAN and Adnan DARWICHE. « On the revision of probabilistic beliefs using uncertain evidence. ». Artif. Intell., 163(1):67–90, 2005.

Cited page(s) 13

[CDT15] Claudette CAYROL, Didier DUBOIS, and Fayçal TOUAZI. « Symbolic Possibilistic Logic: Completeness and Inference Methods (regular paper) ». In Thierry DENOEUX and Sébastien DESTERCKE, editors, European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU), Compiègne, 15/07/2015-17/07/2015, number 9161 in LNAI, pages 485–495. Springer, juillet 2015.

Cited page(s) 2, 3

[CGF12] Giuseppe CURCURÙ, Giacomo Maria GALANTE, and Concetta Manuela La FATA. « Epistemic uncertainty in fault tree analysis approached by the evidence theory ». Journal of Loss Prevention in the Process Industries, 25(4):667 – 676, 2012.

Cited page(s) 23

[CGH96] Enrique CASTILLO, Jose M. GUTIERREZ, and Ali S. HADI. Expert Systems and Probabiistic Network Models. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1996.

Cited page(s) 11

[Coo] Stephen A. COOK. « The Complexity of Theorem-Proving Procedures ». In Proceedings of the 3rd Annual ACM Symposium on Theory of Computing, May 3-5, 1971, Shaker Heights, Ohio, USA, pages 151–158.

Cited page(s) 132

[Coo90] Gregory F. COOPER. « The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks ». Artif. Intell., 42(2-3):393–405, 1990.

Cited page(s) 35

[Coo97] Gert De COOMAN. « Possibility theory Part I: Measure-and integral-theoretics groundwork; Part II: Conditional possibility; Part III: Possibilistic independence ». International Journal of General Systems, 25:291–371, 1997.

Cited page(s) 8, 15

[Coz00] Fabio G. COZMAN. « Credal networks ». Artificial Intelligence, 120(2):199 – 233, 2000.

Cited page(s) 4, 29, 31, 95

[Coz01] Fábio Gagliardi COZMAN. « JavaBayes, Bayesian Networks in Java ». http://www.cs.cmu.edu/~javabayes/Home/, 1998 - 2001.

Cited page(s) 34

[Coz05] Fábio Gagliardi COZMAN. « Graphical models for imprecise probabilities ». Int. J. Approx. Reasoning, 39(2-3):167–184, 2005.

Cited page(s) 29

[CP16] Giulianella COLETTI and Davide PETTURITI. « Finitely maxitive conditional possibilities, Bayesian-like inference, disintegrability and conglomerability ». Fuzzy Sets and Systems, 284:31–55, 2016.

Cited page(s) 2

[CPV14] Giulianella COLETTI, Davide PETTURITI, and Barbara VANTAGGI. « Coherent T-conditional Possibility Envelopes and Nonmonotonic Reasoning ». In Information Processing and Management of Uncertainty in Knowledge-Based Systems - 15th International Conference, IPMU 2014, Montpellier, France, July 15-19, 2014, Proceedings, Part III, pages 446–455, 2014.

Cited page(s) 2

[CSNC15] Sofie De CLERCQ, Steven SCHOCKAERT, Ann NOWÉ, and Martine De COCK. « Multilateral Negotiation in Boolean Games with Incomplete Information Using Generalized Possibilistic Logic ». In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2890–2896, 2015.

Cited page(s) 2

[CZ08] Giorgio CORANI and Marco ZAFFALON. « Learning Reliable Classifiers From Small or Incomplete Data Sets: The Naive Credal Classifier 2 ». Journal of Machine Learning Research, 9:581–621, April 2008.

Cited page(s) 141, 142

[Dar09] Adnan DARWICHE. Modeling and Reasoning with Bayesian Networks. Cambridge University Press, 2009.

Cited page(s) 133, 154

[Daw79] A. Philip DAWID. « Conditional Independence in Statistical Theory ». Journal of the Royal Statistical Society. Series B (Methodological), 41(1):1–31, 1979.

Cited page(s) 15

[dC] Gert de COOMAN. « Integration and Conditioning in Numerical Possibility Theory ». Annals of Mathematics and Artificial Intelligence, 32(1):87–123.

Cited page(s) 2

[dCC05] Cassio Polpo de CAMPOS and Fábio Gagliardi COZMAN. « The Inferential Complexity of Bayesian and Credal Networks ». In Leslie Pack KAELBLING and Alessandro SAFFIOTTI, editors, IJCAI-05, Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence, Edinburgh, Scotland, UK, July 30 - August 5, 2005, pages

1313–1318. Professional Book Center, 2005.

Cited page(s) 1, 37

[DDC07] Sébastien DESTERCKE, Didier DUBOIS, and Eric CHOJNACKI. « Transforming Probability Intervals into Other Uncertainty Models. ». In EUSFLAT 2007 proc., volume 2, pages 367–373, Ostrava, Czech Republic, 2007.

Cited page(s) 23, 47, 126

[DDC08] Sébastien DESTERCKE, Didier DUBOIS, and Eric CHOJNACKI. « Unifying practical uncertainty representations - I: Generalized p-boxes ». Int. J. Approx. Reasoning, 49(3):649–663, 2008.

Cited page(s) 47, 124

[Dem67] Arthur P. DEMPSTER. « Upper and Lower Probabilities Induced by a Multivalued Mapping ». The Annals of Mathematical Statistics, 38:325–339, 1967.

Cited page(s) 4, 8, 21, 23

[DFMP04] Didier DUBOIS, Laurent FOULLOY, Gilles MAURIS, and Henri PRADE. « Probability-Possibility Transformations, Triangular Fuzzy Sets, and Probabilistic Inequalities ». Reliable Computing, 10(4):273–297, 2004.

Cited page(s) 4, 40, 41, 42, 43, 144, 146

[DFP04] Didier DUBOIS, Hélène FARGIER, and Henry PRADE. « Ordinal and Probabilistic Representations of Acceptance ». J. Artif. Int. Res., 22(1):23–56, July 2004.

Cited page(s) 122

[DG94] Adnan DARWICHE and Moisés GOLDSZMIDT. « On the Relation between Kappa Calculus and Probabilistic Reasoning ». In Ramon López de MÁNTARAS and David POOLE, editors, UAI '94: Proceedings of the Tenth Annual Conference on Uncertainty in Artificial Intelligence, Seattle, Washington, USA, July 29-31, 1994, pages 145–153. Morgan Kaufmann, 1994.

Cited page(s) 24

[DGM11] Pilar DELLUNDE, Lluis GODO, and Enrico MARCHIONI. « Extending possibilistic logic over Gödel logic ». Int. J. Approx. Reasoning, 52(1):63–75, 2011.

Cited page(s) 2

[DLP91] Didier DUBOIS, Jérôme LANG, and Henri PRADE. « Timed possibilistic logic ». Fundam. Inform., 15(3-4):211–234, 1991.

Cited page(s) 2

[DP80] Didier DUBOIS and Henry PRADE. Fuzzy Sets and Systems: Theory and Applications. New York: Academic Press, 1980.

Cited page(s) 15, 41

[DP83] Didier DUBOIS and Henri PRADE. « Unfair coins and necessity measures: a possibilistic interpretation of histograms ». Fuzzy Sets and Systems, 10(1):15–20, 1983. DP034.

Cited page(s) 45, 48

[DP88] Didier DUBOIS and Henri PRADE. Possibility Theory: An Approach to Computerized Processing of Uncertainty (traduction revue et augmentée de Théorie des Possibilités). Plenum Press, New York, 1988.

Cited page(s) 8, 15

[DP90] Didier DUBOIS and Henri PRADE. « The logical view of conditioning and its application to possibility and evidence theories ». Int. J. Approx. Reasoning, 4(1):23–46, 1990.

Cited page(s) 2

[DP92] Didier DUBOIS and Henri PRADE. « Evidence, knowledge, and belief functions ». International Journal of Approximate Reasoning, 6(3):295 – 319, 1992.
Cited page(s) 45

[DP93] Didier DUBOIS and Henry PRADE. « Fuzzy sets and probability: misunderstandings, bridges and gaps ». Fuzzy Systems, 1993.
Cited page(s) 15

[DP97a] Didier DUBOIS and Henri PRADE. « Bayesian conditioning in possibility theory ». Fuzzy Sets and Systems, 92(2):223 – 240, 1997. Fuzzy Measures and Integrals.
Cited page(s) 2, 53

[DP97b] Didier DUBOIS and Henri PRADE. « A synthetic view of belief revision with uncertain inputs in the framework of possibility theory. ». Int. J. Approx. Reasoning, 17(2-3):295–324, 1997.
Cited page(s) 18

[DP98] Didier DUBOIS and Henri PRADE. « Possibility Theory: Qualitative and Quantitative Aspects », pages 169–226. Springer Netherlands, Dordrecht, 1998.
Cited page(s) 8, 15

[DP04] Didier DUBOIS and Henri PRADE. « Possibilistic logic: a retrospective and prospective view ». Fuzzy Sets and Systems, 144:3–23, 2004.
Cited page(s) 31

[DP05] Didier DUBOIS and Henri PRADE. « Interval-valued Fuzzy Sets, Possibility Theory and Imprecise Probability ». In Eduard MONTSENY and Pilar SOBREVILLA, editors, Proceedings of the Joint 4th Conference of the European Society for Fuzzy Logic and Technology and the 11th Rencontres Francophones sur la Logique Floue et ses Applications, Barcelona, Spain, September 7-9, 2005, pages 314–319. Universidad Polytecnica de Catalunya, 2005.
Cited page(s) 8, 20, 23, 46

[DP06] Didier DUBOIS and Henri PRADE. Possibility Theory and its Applications: a Retrospective and Prospective view. In Giacomo DELLA RICCIA, Didier DUBOIS, Rudolf KRUSE, and Hanz-Joachim LENZ, editors, Decision Theory and Multi-Agent Planning, volume 482 of CISM International Centre for Mechanical Sciences, pages 89–109. Springer Vienna, 2006.
Cited page(s) 2, 53

[DP09] Didier DUBOIS and Henry PRADE. « Formal Representations of Uncertainty », Chapter 3. ISTE, London, UK., 2009.
Cited page(s) 8

[DP11] Didier DUBOIS and Henri PRADE. Generalized Possibilistic Logic. In Salem BENFERHAT and John GRANT, editors, Scalable Uncertainty Management, volume 6929 of Lecture Notes in Computer Science, pages 428–432. Springer Berlin Heidelberg, 2011.
Cited page(s) 2

[DP12] Didier DUBOIS and Henri PRADE. « Possibility Theory », pages 2240–2252. Springer New York, New York, NY, 2012.
Cited page(s) 2, 15

[DP13] Didier DUBOIS and Henri PRADE. « Updating with Belief Functions, Ordinal Conditioning Functions and Possibility Measures ». CoRR, abs/1304.1118, 2013.
Cited page(s) 24

[DP15] Didier DUBOIS and Henri PRADE. Possibility Theory and Its Applications: Where Do We Stand? In Janusz KACPRZYK and Witold PEDRYCZ, editors, Springer Handbook of Computational Intelligence, pages 31–60. Springer Berlin Heidelberg, 2015.
Cited page(s) 2

[DPS93] Didier DUBOIS, Henri PRADE, and Sandra SANDRI. On Possibility/Probability Transformations. In R. LOWEN and M. ROUBENS, editors, Fuzzy Logic, pages 103–112. Kluwer Academic Publishers, Dordrecht, 1993.
Cited page(s) 4, 40, 42, 43, 48

[DPS12] Didier DUBOIS, Henri PRADE, and Steven SCHOCKAERT. « Stable Models in Generalized Possibilistic Logic ». In Principles of Knowledge Representation and Reasoning: Proceedings of the Thirteenth International Conference, KR 2012, Rome, Italy, June 10-14, 2012, 2012.
Cited page(s) 2

[Dub06] Didier DUBOIS. « Possibility theory and statistical reasoning ». Computational statistics & data analysis, 51(1):47–69, 2006.
Cited page(s) 2

[Dub14] Didier DUBOIS. « On Various Ways of Tackling Incomplete Information in Statistics ». Int. J. Approx. Reasoning, 55(7):1570–1574, 2014.
Cited page(s) 2

[EK14] Christian EICHHORN and Gabriele KERN-ISBERNER. « LEG Networks for Ranking Functions ». In Eduardo FERMÉ and João LEITE, editors, Logics in Artificial Intelligence - 14th European Conference, JELIA 2014, Funchal, Madeira, Portugal, September 24-26, 2014. Proceedings, volume 8761 of Lecture Notes in Computer Science, pages 210–223. Springer, 2014.
Cited page(s) 4, 24

[FHM90] Ronald FAGIN, Joseph Y. HALPERN, and Nimrod MEGIDDO. « A logic for reasoning about probabilities ». Information and Computation, 87(1):78 – 128, 1990. Special Issue: Selections from 1988 IEEE Symposium on Logic in Computer Science.
Cited page(s) 1

[FKG+03] Scott FERSON, Vladik KREINOVICH, Lev GINZBURG, Davis S. MYERS, and Kari SENTZ. « Constructing Probability Boxes and Dempster-Shafer Structures ». 04 2003.
Cited page(s) 23

[FKRS12] Marcelo A. FALAPPA, Gabriele KERN-ISBERNER, Maurício D. Luís REIS, and Guillermo Ricardo SIMARI. « Prioritized and Non-prioritized Multiple Change on Belief Bases ». J. Philosophical Logic, 41(1):77–113, 2012.
Cited page(s) 2, 24

[FL15] Tuan-Fang FAN and Churn-Jung LIAU. « A Logic for Reasoning about Justified Uncertain Beliefs ». In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, pages 2948–2954, 2015.
Cited page(s) 2

[Fon97] Pascale FONCK. « A comparative study of possibilistic conditional independence and lack of interaction ». International Journal of Approximate Reasoning, 16(2):149–171, 1997.
Cited page(s) 2, 53

[GP13] Moisés GOLDSZMIDT and Judea PEARL. « Reasoning With Qualitative Probabilities Can Be Tractable ». CoRR, abs/1303.5406, 2013.

Cited page(s) 24

[His78] Ellen HISDAL. « Conditional possibilities independence and non interaction ». Fuzzy Sets and Systems, pages 283–297, 1978.

Cited page(s) 2, 17, 53

[HLA15] Maroua HADDAD, Philippe LERAY, and Nahla Ben AMOR. « Learning possibilistic networks from data: a survey ». In IFSA-EUSFLAT-15, Gijón, Spain., June 30, 2015., 2015.

Cited page(s) 138

[HLLT16] Maroua HADDAD, Philippe LERAY, Amélie LEVRAY, and Karim TABIA. « Possibilistic networks parameter learning: Preliminary empirical comparison ». In JFRB 2016 - 8ème Journées Francophones sur les Réseaux Bayésiens et les Modèles Graphiques Probabilistes, 27 June-1 July 2016, Clermont-Ferrand, France, 2016.

Cited page(s) 5, 6

[HLLT17] Maroua HADDAD, Philippe LERAY, Amélie LEVRAY, and Karim TABIA. « Learning the Parameters of Possibilistic Networks from Data: Empirical Comparison ». In FLAIRS Conference, pages 736–741. AAAI Press, 2017.

Cited page(s) 5, 6, 130

[Hsi94] Yen-Teh HSIA. « Possibilistic Conditioning and Propagation ». In Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence, UAI'94, pages 336–343, San Francisco, CA, USA, 1994. Morgan Kaufmann Publishers Inc.

Cited page(s) 2

[JAE+07] Ilyes JENHANI, Nahla Ben AMOR, Zied ELOUEDI, Salem BENFERHAT, and Khaled MELLOULI. « Information Affinity: A New Similarity Measure for Possibilistic Uncertain Information. ». In Khaled MELLOULI, editor, ECSQARU, volume 4724, pages 840–852. Springer, 2007.

Cited page(s) 135

[Jay03] Edwin Thompson JAYNES. Probability Theory: The Logic of Science. Cambridge University Press, 2003.

Cited page(s) 8, 9

[Jef65] Richard JEFFREY. « The Logic of Decision ». 1965.

Cited page(s) 13

[Jen96] Finn V. JENSEN. Introduction to Bayesian Networks. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1996.

Cited page(s) 36, 159

[Jos97] Cliff JOSLYN. « Measurement of possibilistic histograms from interval data ». International Journal Of General System, 26(1-2):9–33, 1997.

Cited page(s) 140

[KBvdG10] Johan KWISTHOUT, Hans L. BODLAENDER, and Linda C. van der GAAG. « The Necessity of Bounded Treewidth for Efficient Inference in Bayesian Networks ». In ECAI, volume 215 of Frontiers in Artificial Intelligence and Applications, pages 237–242. IOS Press, 2010.

Cited page(s) 4

[KE13] Gabriele KERN-ISBERNER and Christian EICHHORN. « Intensional Combination of Rankings for OCF-Networks ». In Chutima BOONTHUM-DENECKE and G. Michael YOUNGBLOOD, editors, Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2013, St. Pete Beach, Florida. May 22-24, 2013. AAAI Press, 2013.

Cited page(s) 24

[KE14] Gabriele KERN-ISBERNER and Christian EICHHORN. « Structural Inference from Conditional Knowledge Bases ». Studia Logica, 102(4):751–769, 2014.

Cited page(s) 24

[Ker01] Gabriele KERN-ISBERNER. Conditionals in Nonmonotonic Reasoning and Belief Revision - Considering Conditionals as Agents, volume 2087 of Lecture Notes in Computer Science. Springer, 2001.

Cited page(s) 24

[Ker04] Gabriele KERN-ISBERNER. « A Thorough Axiomatization of a Principle of Conditional Preservation in Belief Revision ». Ann. Math. Artif. Intell., 40(1-2):127–164, 2004.

Cited page(s) 2, 24

[KG93] Georges J. KLIR and James F. GEER. Information-Preserving Probability- Possibility Transformations: Recent Developments. In R. LOWEN and M. ROUBENS, editors, Fuzzy Logic, pages 417–428. Kluwer Academic Publishers, Dordrecht, 1993.

Cited page(s) 3, 4, 40, 41, 42

[KH15] Gabriele KERN-ISBERNER and Daniela HUVERMANN. « Multiple Iterated Belief Revision Without Independence ». In Ingrid RUSSELL and William EBERLE, editors, Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2015, Hollywood, Florida. May 18-20, 2015., pages 570–575. AAAI Press, 2015.

Cited page(s) 8, 24

[Kol60] Andrey N. KOLMOGOROV. Foundations of the Theory of Probability. Chelsea Pub Co, 2 edition, 1960.

Cited page(s) 4, 8, 9, 10, 11

[KT12] Gabriele KERN-ISBERNER and Matthias THIMM. « A Ranking Semantics for First-Order Conditionals ». In Luc De RAEDT, Christian BESSIÈRE, Didier DUBOIS, Patrick DOHERTY, Paolo FRASCONI, Fredrik HEINTZ, and Peter J. F. LUCAS, editors, ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track, Montpellier, France, August 27-31 , 2012, volume 242 of Frontiers in Artificial Intelligence and Applications, pages 456–461. IOS Press, 2012.

Cited page(s) 4, 24

[Kwi14] Johan KWISTHOUT. « Treewidth and the Computational Complexity of MAP Approximations ». In Probabilistic Graphical Models, volume 8754 of Lecture Notes in Computer Science, pages 271–285. Springer, 2014.

Cited page(s) 4

[Lan00] Jérôme LANG. Possibilistic logic: complexity and algorithms. In Handbook of Defeasible Reasoning and Uncertainty Management Systems, volume 5, pages 179–220. Kluwer Academic, 2000.

Cited page(s) 2, 31, 79

[Lev80] Isaac LEVI. The enterprise of knowledge : an essay on knowledge, credal probability, and chance. MIT Press Cambridge, Mass, 1980.

Cited page(s) 20, 95

[Lib05] Paolo LIBERATORE. « Redundancy in logic I: CNF propositional formulae ». Artif. Intell., 163(2):203–232, 2005.

Cited page(s) 82

[LMDCM95] Juan F. Huete LUIS M. DE CAMPOS and Serafin MORAL. « Possibilistic independence ». Proceedings of EUFIT 95, 1:69–73, 1995.

Cited page(s) 2, 53

[LS90] Steffen L. LAURITZEN and David J. SPIEGELHALTER. Readings in Uncertain Reasoning. Chapter Local Computations with Probabilities on Graphical Structures and Their Application to Expert Systems, pages 415–448. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

Cited page(s) 159

[Luk99] Thomas LUKASIEWICZ. « Probabilistic Deduction with Conditional Constraints over Basic Events ». J. Artif. Int. Res., 10(1):199–241, April 1999.

Cited page(s) 1

[MD06] Marie-Hélène MASSON and Thierry DENOEUX. « Inferring a possibility distribution from empirical data. ». Fuzzy Sets and Systems, 157(3):319–340, 2006.

Cited page(s) 3, 40, 46, 47, 123

[MD13] Enrique MIRANDA and Sébastien DESTERCKE. « Extreme Points of the Credal Sets Generated by Elementary Comparative Probabilities ». In Linda C. van der GAAG, editor, Symbolic and Quantitative Approaches to Reasoning with Uncertainty - 12th European Conference, ECSQARU 2013, Utrecht, The Netherlands, July 8-10, 2013. Proceedings, volume 7958 of Lecture Notes in Computer Science, pages 424–435. Springer, 2013.

Cited page(s) 20

[MdCBA13] Denis Deratani MAUÁ, Cassio Polpo de CAMPOS, Alessio BENAVOLI, and Alessandro ANTONUCCI. « On the Complexity of Strong and Epistemic Credal Networks ». In UAI. AUAI Press, 2013.

Cited page(s) 4

[MdCBA14] Denis Deratani MAUÁ, Cassio Polpo de CAMPOS, Alessio BENAVOLI, and Alessandro ANTONUCCI. « Probabilistic Inference in Credal Networks: New Complexity Results ». J. Artif. Intell. Res. (JAIR), 50:603–637, 2014.

Cited page(s) 1, 37

[MdCC15] Denis Deratani MAUÁ, Cassio Polpo de CAMPOS, and Fábio Gagliardi COZMAN. « The Complexity of MAP Inference in Bayesian Networks Specified Through Logical Languages ». In IJCAI, pages 889–895. AAAI Press, 2015.

Cited page(s) 4

[MSMR06] Patrice Billaudel MOAMAR SAYED MOUCHAWEH, Mohamed Said Bouguelid and Bernard RIERA. Variable probability-possibility transformation. pages 417–428. September 2006.

Cited page(s) 4, 40, 42, 43, 113

[Nea12] Richard E. NEAPOLITAN. Probabilistic Reasoning In Expert Systems: Theory and Algorithms. CreateSpace Independent Publishing Platform, 2012.

Cited page(s) 8, 9

[NK14] Hung T. NGUYEN and Vladik KREINOVICH. « How to fully represent expert information about imprecise properties in a computer system: random sets, fuzzy sets, and beyond: an overview ». International Journal of General Systems, 43(6):586–609, 2014.
Cited page(s) 2

[PD04] James D. PARK and Adnan DARWICHE. « Complexity Results and Approximation Strategies for MAP Explanations ». J. Artif. Int. Res., 21(1):101–133, February 2004.
Cited page(s) 134

[Pea88] Judea PEARL. « On logic and probability ». Computational Intelligence, 4:99–103, 1988.
Cited page(s) 15, 27

[Pea89] Judea PEARL. Probabilistic reasoning in intelligent systems - networks of plausible inference. Morgan Kaufmann series in representation and reasoning. Morgan Kaufmann, 1989.
Cited page(s) 4, 27

[Pea09] Judea PEARL. Causality: Models, Reasoning and Inference. Cambridge University Press, New York, NY, USA, 2nd edition, 2009.
Cited page(s) 15

[Rot96] Dan ROTH. « On the hardness of approximate reasoning ». Artificial Intelligence, 82(1):273 – 302, 1996.
Cited page(s) 134

[S+76] Glenn SHAFER and OTHERS. A mathematical theory of evidence, volume 1. Princeton university press Princeton, 1976.
Cited page(s) 139

[SD03] Solomon Eyal SHIMONY and Carmel DOMSHLAK. « Complexity of probabilistic reasoning in directed-path singly-connected Bayes networks ». Artif. Intell., 151(1-2):213–225, 2003.
Cited page(s) 4

[Sha76] Glenn SHAFER. A mathematical Theory of Evidence. Princeton University Press, 1976.
Cited page(s) 4, 8, 21, 23

[She94] Prakash P. SHENOY. Advances in the Dempster-Shafer Theory of Evidence. Chapter Using Dempster-Shafer's Belief-function Theory in Expert Systems, pages 395–414. John Wiley & Sons, Inc., New York, NY, USA, 1994.
Cited page(s) 8

[Shi94] Solomon Eyal SHIMONY. « Finding MAPs for Belief Networks is NP-Hard. ». Artif. Intell., 68(2):399–410, 1994.
Cited page(s) 133

[Sme89] Philippe SMETS. « Constructing the Pignistic Probability Function in a Context of Uncertainty ». In Max HENRION, Ross D. SHACHTER, Laveen N. KANAL, and John F. LEMMER, editors, UAI '89: Proceedings of the Fifth Annual Conference on Uncertainty in Artificial Intelligence, Windsor, Ontario, Canada, August 18-20, 1989, pages 29–40. North-Holland, 1989.
Cited page(s) 40, 45

[Sno96] Paul SNOW. « Standard probability distributions described by rational default entailment ». 1996.
Cited page(s) 122

[SP15] Mathieu SERRURIER and Henri PRADE. « Learning Structure of Bayesian Networks by Using Possibilistic Upper Entropy », pages 87–95. Springer International Publishing, Cham, 2015.

Cited page(s) 138

[Spo14] Wolfgang SPOHN. The Laws of Belief - Ranking Theory and Its Philosophical Applications. Oxford University Press, 2014.

Cited page(s) 24

[Sud92] Thomas SUDKAMP. « On probability-possibility transformations ». Fuzzy Sets and Systems, pages 73–81, 1992.

Cited page(s) 98, 120

[TCD15] Fayçal TOUAZI, Claudette CAYROL, and Didier DUBOIS. « Possibilistic reasoning with partially ordered beliefs ». J. Applied Logic, 13(4):770–798, 2015.

Cited page(s) 24

[Wal00] Peter WALLEY. « Towards a unified theory of imprecise probability ». Int. J. Approx. Reasoning, 24(2-3):125–148, 2000.

Cited page(s) 8, 20, 46

[Wal07] Anton WALLNER. « Extreme points of coherent probabilities in finite spaces ». International Journal of Approximate Reasoning, 44(3):339 – 357, 2007. Reasoning with Imprecise Probabilities.

Cited page(s) 4, 8, 20, 96

[Wei00] Kurt WEICHSELBERGER. « The theory of interval-probability as a unifying concept for uncertainty ». Int. J. Approx. Reasoning, 24(2-3):149–170, 2000.

Cited page(s) 8, 20, 46

[Wil94a] Mary-Anne WILLIAMS. « On the Logic of Theory Base Change ». In Craig MACNISH, David PEARCE, and Luís Moniz PEREIRA, editors, Logics in Artificial Intelligence, European Workshop, JELIA '94, York, UK, September 5-8, 1994, Proceedings, volume 838 of Lecture Notes in Computer Science, pages 86–105. Springer, 1994.

Cited page(s) 24

[Wil94b] Mary-Anne WILLIAMS. « Transmutations of Knowledge Systems ». In Jon DOYLE, Erik SANDEWALL, and Pietro TORASSO, editors, Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94). Bonn, Germany, May 24-27, 1994., pages 619–629. Morgan Kaufmann, 1994.

Cited page(s) 24

[Wil95] Mary-Anne WILLIAMS. « Iterated Theory Base Change: A Computational Model ». In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, IJCAI 95, Montréal Québec, Canada, August 20-25 1995, 2 Volumes, pages 1541–1549. Morgan Kaufmann, 1995.

Cited page(s) 24

[Yag81] Ronald R. YAGER. « A procedure for ordering fuzzy subsets of the unit interval ». Information Sciences, 24(2):143 – 161, 1981.

Cited page(s) 45

[Yag82] Ronald .R. YAGER. Fuzzy Set and Possibility Theory: Recent Developments. Pergamon Press, New York, 1982.

Cited page(s) 2

[Yam01] Koichi YAMADA. « Probability-Possibility Transformation Based On Evidence Theory ». pages 70–75, 2001.

Cited page(s) 48

[Zad79] Lofti A. ZADEH. « A theory of approximate reasoning ». In J.E. HAYES, D. MICHIE, and L.I MIKULICH, editors, Machine intelligence, pages 149–194. Elsevier, Amsterdam, 1979.

Cited page(s) 19

[Zad99] Lotfi A. ZADEH. « Fuzzy sets as a basis for a theory of possibility ». Fuzzy Sets and Systems, 100:9–34, 1999.

Cited page(s) 2, 3, 4, 8, 15, 41

Résumé

Cette thèse contribue au développement de formalismes efficaces pour représenter l'information incertaine. Les formalismes existants tels que la théorie des probabilités ou la théorie des possibilités sont parmi les cadres les plus connus et utilisés pour représenter ce type d'information. Différentes extensions (e.g. théorie des probabilités imprécises, théorie des possibilités à intervalles) ont été proposées pour traiter des informations incomplètes ou des connaissances mal-connues, ainsi que pour raisonner avec les connaissances d'un groupe d'experts. Les contributions de cette thèse sont divisées en deux parties. Dans la première partie, nous développons le conditionnement dans le cadre des possibilités à intervalles et dans le cadre des possibilités ensemblistes. Conditionner dans le cadre standard diffère que l'on considère l'échelle possibiliste qualitative ou quantitative. Notre travail traite les deux définitions du conditionnement possibiliste. Ce qui nous amène à étudier une nouvelle extension de la logique possibiliste, définie comme logique possibiliste ensembliste, et son opérateur de conditionnement dans le cadre possibiliste qualitatif. Ces résultats, plus spécialement en termes de complexité, nous amène à étudier les transformations, plus précisément des transformations du cadre probabiliste vers le cadre possibiliste. En effet, nous analysons des propriétés les tâches de raisonnement comme la marginalisation et le conditionnement. Nous nous attaquons aussi aux transformations des probabilités imprécises vers les possibilités avec un intérêt particulier pour l'inférence MAP.

Mots-clés: Cadre des possibilités à intervalles, conditionnement, raisonnement, transformation probabiliste-possibiliste, inference MAP.

Abstract

This thesis contributes to the development of efficient formalisms to handle uncertain information. Existing formalisms such as probability theory or possibility theory are among the most known and used settings to represent such information. Extensions and generalizations (*e.g.* imprecise probability theory, interval-based possibilistic theory) have been provided to handle uncertainty such as incomplete and ill-known knowledge and reasoning with the knowledge of a group of experts. We are particularly interested in reasoning tasks within these theories such as conditioning.

The contributions of this thesis are divided in two parts. In the first part, we tackle conditioning in interval-based possibilistic framework and set-valued possibilistic framework. The purpose is to develop a conditioning machinery for interval-based possibilistic logic. Conditioning in a standard possibilistic setting differs whether we consider a qualitative or quantitative scale. Our works deal with both definitions of possibilistic conditioning. This leads us to investigate a new extension of possibilistic logic, defined as set-valued possibilistic logic, and its conditioning machinery in the qualitative possibilistic setting. These results, especially in terms of complexity, lead us to study transformations, more precisely from probability to possibility theories. The second part of our contributions deals with probability-possibility transformation procedures. Indeed, we analyze properties of reasoning tasks such as conditioning and marginalization. We also tackle transformations from imprecise probability theory to possibility theory with a particular interest in MAP inference.

Keywords: Interval-based possibilistic setting, conditioning, reasoning tasks, probability-possibility transformation, MAP inference..