



HAL
open science

Uncertainty representation and combination: new results with application to nuclear safety issues

Sébastien Destercke

► **To cite this version:**

Sébastien Destercke. Uncertainty representation and combination: new results with application to nuclear safety issues. Other [cs.OH]. Université Paul Sabatier - Toulouse III, 2008. English. NNT : 2008TOU30179 . tel-00355145

HAL Id: tel-00355145

<https://theses.hal.science/tel-00355145>

Submitted on 22 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

Présentée en vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse III - Paul Sabatier

Discipline : informatique

Soutenu par Sébastien Destercke

Le 29 Octobre 2008

Représentation et combinaison d'informations incertaines : nouveaux résultats avec applications aux études de sûreté nucléaires

Uncertainty representation and combination: new results with application to nuclear safety issues

Devant le jury composé de:

| | |
|--------------------------|--|
| Eric Chojnacki | Ingénieur de recherche, IRSN, Cadarache - <i>Co-encadrant</i> |
| Gert de Cooman | Professeur, Gent Universiteit - <i>Examineur</i> |
| Thierry Denoeux | Professeur, Université de technologie de Compiègne - <i>Rapporteur</i> |
| Didier Dubois | Directeur de recherche, CNRS, IRIT - <i>Directeur de recherche</i> |
| Philippe Fortemps | Professeur, Faculté Polytechnique de Mons - <i>Examineur</i> |
| Serafin Moral | Professeur, University of Granada - <i>Rapporteur</i> |
| Olivier Strauss | Maître de conférence, Université Montpellier 2 - <i>Rapporteur</i> |
| Monique Pontier | Professeur émérite, Université Paul Sabatier - <i>Président</i> |

École doctorale : Mathématiques, Informatique et Télécommunications de Toulouse - ED 475

Laboratoire d'accueil : Institut de Recherche en Informatique de Toulouse

Équipe d'accueil : Raisonnements Plausibles, Décision, Méthodes de preuve

To my father

Acknowledgements

This work has been funded by the Institut de Radioprotection et Sûreté Nucléaire (IRSN).

My first thanks go to my two supervisors, Didier Dubois and Eric Chojnacki, which have given me far more than a simple help to finish this work. Rather than a long list of personal qualities that each of them possesses, I want to say how much I've appreciated working with them, or simply talking with them, all along these three years.

Thanks to the members of my jury, which have had the courage to read most of this (long) thesis. For different reasons, the presence of each of them was important to me. Of course, some special thanks go to my three referees, Thierry Denoeux, Serafin Moral and Olivier Strauss, for their comments and the various discussions about my works (which, I hope, are only a start).

I also especially thank those people who have given me the opportunity to collaborate with them on scientific interests that we had in common. So thank Quique, Lev, Gert, Matthias, for the many good times spent together around scientific questions and beers.

There is an awfully long list of other persons I'd like to thank, be it for enjoyable moments spent together, scientific conversations, supports, friendship and I-do-not-know what else memorable slices of life. So, in a non exhaustive and disordered list, they are: Samir, Erik, Marie, Nathan, Fortunato, Filip, Jeremy, Nathalie, Kevin, Gero, Alessandro, Frank, Chaouki, Claire, Laurence, Diana, Pedro, Jean, Pierre-Guy, Jeffrey, Luigi, Ines, Margot, Herman, Thomas, H  l  ne, Rebecca, Ric, Chris, . . .

I would also like to thank Douglas Adams, Terry Pratchett, Dan Simmons, The Monty Pythons , The Cohen Brothers, Kevin Smith, Philip K. Dick, Manu Larcenet, Jiro Tanigushi, Frank Miller and many others simply for existing and doing what they're doing.

My final thanks go to my family, who have supported me, even if they still do not really understand what my job is about (except that it can involve, at some point, cows). So thank you mother, H  l  ne, Nicolas, and the woman who has the courage and strength of will to share my life: Julie.

Résumé

Ces dernières années ont vu apparaître de nombreux arguments convergents vers la conclusion que les probabilités classiques ne pouvaient pas rendre compte, à la fois dans leur calcul et leur représentation, de l'imprécision ou de l'incomplétude éventuellement présente dans l'information disponible concernant un système, une variable, un paramètre. Aussi des théories de l'incertain ayant pour but de prendre explicitement en compte cette imprécision ont émergé. Les trois principales de ces théories sont, de la plus à la moins générale: la théorie des probabilités imprécises, la théorie des ensembles aléatoires, la théorie des possibilités.

Avec elles sont également apparues de nouvelles difficultés et de nouvelles questions relatives à la représentation et au traitement des incertitudes: difficultés d'ordre pratique lors de la manipulation des informations, la prise en compte explicite de l'imprécision posant de nouveaux problèmes calculatoires ; questions sur l'interprétation de certaines notions (indépendance, conditionnement) pour lesquels il y avait un consensus assez fort dans le cadre classique des probabilités ; problèmes d'unifications dus au fait que les théories proposent des calculs, des solutions et des modes de traitement différents. En effet, en choisissant un cadre alternatif ou plus expressif pour représenter et traiter l'incertitude, des problèmes qui étaient auparavant "cachés" par le cadre relativement contraignant des probabilités classiques refont surface.

Dans ce travail, nous apportons des réponses partielles à ces problèmes, à la fois en essayant d'interpréter les différentes notions au sein de cadres unificateurs et en proposant des méthodes de manipulation pratiques. Nous nous intéressons principalement aux problèmes suivants:

- L'étude des représentations pratiques d'incertitudes. En particulier, nous situons des représentations récemment proposées (p-boxes, nuages) par rapport à des représentations plus anciennes. Cela nous permet de mettre à jour un nombre intéressant de relations, facilitant de futures manipulations pratiques.

- La combinaison d'informations provenant de sources multiples. En particulier, nous nous intéressons aux problèmes de la combinaison d'informations partiellement inconsistantes et de la prise en compte de dépendances entre sources. Nous nous intéressons aussi brièvement au problème de l'évaluation de la qualité de l'information transmise.
- La modélisation de la notion d'indépendance entre variables, cette notion étant essentielle lors de la combinaison de modèle marginaux d'incertitudes en modèles joints. Nous nous contentons de donner une vue générale de la problématique ainsi que quelques premiers résultats, vu que l'étude complète de ces notions nécessiterait un travail de recherche en soi.

Nous considérons aussi brièvement les problèmes de la prise de décision, et détaillons des applications pratiques mettant en oeuvre quelques unes des méthodes et représentations étudiées dans ce travail.

Abstract

In these last years, many arguments appeared, converging to the fact that classical probabilities cannot adequately handle or represent imprecision or incompleteness in the available information concerning a system, a variable or a parameter. Hence, alternative theories proposing to address and solve this issue have emerged. The three main such theories are, from the more to the less general: imprecise probability theory, random set theory, possibility theory.

With them also appeared new difficulties and questions related to the representation and treatment of uncertainty: difficulties regarding the practical handling of uncertainties, since explicitly modeling imprecision often means an higher computational complexity when treating the information; questions related to the interpretation of some notions (conditioning, independence) that almost met general consensus in classical probabilities; problems of unification due to the fact that uncertainty calculus and treatments are sometimes different between different theories and interpretations. Actually, by choosing a different or a more expressive framework to handle uncertainty, issues that were previously "hidden" by the somewhat restrictive setting of classical probability theory are no longer hidden in the new setting.

In this work, we bring some partial answers to above issues, first by trying to settle different problematics in unified settings, second by proposing practical methods allowing to handle uncertainty in an efficient way. We focus mainly on the following issue:

- The study of practical uncertainty representations. In particular, we situate more recent uncertainty representations (p-boxes and clouds) with respect to older uncertainty representations. This lead us to expose a number of interesting relations between representations, eventually leading to an easier practical handling of such representations.
- The combination of information coming from multiple sources. In particular, we look at the two problems of combining partially consistent information and of taking account of potential dependencies between information sources. We also address the issue of evaluating the quality of the delivered information by the use of past assessments

- Modeling and interpreting notions of independence between variables, these notions being essential in the construction of joint uncertainty models from marginal ones. Here, we simply give a general picture of the (many) notions existing in the uncertainty theories considered here, and propose some first results eventually leading to a unified frame. Indeed, a full study of the complex notion of independence would require a work of its own.

Finally, we briefly look at the problems of decision making, and give some details about two applications achieved during this work and using some of the methods exposed therein.

Contents

| | | |
|----------|--|-----------|
| 1 | Résumé Français de la thèse (French Summary of the thesis) | 1 |
| 1.1 | Introduction (Chapitre 2) | 1 |
| 1.2 | Représentations pratiques d’incertitude (Chapitre 3) | 6 |
| 1.2.1 | Mesures non-additives et représentations connues | 6 |
| 1.2.2 | P-boxes généralisées | 12 |
| 1.2.3 | Nuages | 14 |
| 1.3 | Traitement de sources multiples d’informations (Chapitre 4) | 16 |
| 1.3.1 | Opérations de fusion d’information basiques | 18 |
| 1.3.2 | Utilisation des sous-ensembles maximaux cohérents (SMC) | 23 |
| 1.4 | Incertaines et (In)dépendance (Chapitre 5) | 28 |
| 1.5 | Prise de décision dans l’incertain (Chapitre 6) | 33 |
| 1.6 | Applications illustratives (Chapitre 7) | 35 |
| 1.6.1 | Evaluation et synthèse d’informations appliquées à des codes de calculs nucléaires | 35 |
| 1.6.2 | Application de la méthode RaFu à un cas d’étude | 37 |
| 1.7 | Conclusions et perspectives (Chapitre 8) | 39 |
| 2 | Introduction | 41 |

| | | |
|----------|--|-----------|
| 2.1 | Reasoning under uncertainty (with quantitative models): a general view . . . | 42 |
| 2.2 | About the present work | 45 |
| 3 | Practical uncertainty representations | 49 |
| 3.1 | Non-additive measures and representations of uncertainty | 51 |
| 3.1.1 | Capacities and transformations of capacities | 52 |
| 3.1.2 | Practical representations in imprecise probability | 55 |
| 3.1.3 | Sketching a first summary of relationships | 63 |
| 3.2 | Introduction and study of generalised p-boxes | 65 |
| 3.2.1 | Definition of generalized p-boxes | 66 |
| 3.2.2 | Connecting generalized p-boxes with possibility distributions | 69 |
| 3.2.3 | Connecting Generalized p-boxes and random sets | 70 |
| 3.2.4 | Generalized p-boxes and imprecise probability assignments | 74 |
| 3.2.5 | Computing with generalized p-boxes: first results on propagation . . . | 79 |
| 3.3 | Clouds | 82 |
| 3.3.1 | Definition of clouds | 83 |
| 3.3.2 | Clouds in the setting of possibility theory | 85 |
| 3.3.3 | Generalized p-boxes as a special kind of clouds | 88 |
| 3.3.4 | The Nature of Non-comonotonic Clouds | 91 |
| 3.3.5 | Clouds and imprecise probability assignments | 96 |
| 3.4 | A word on continuous representations on the real line | 103 |
| 3.4.1 | Practical continuous representations on the real line | 103 |
| 3.4.2 | Continuous clouds on the real line | 105 |
| 3.4.3 | Thin continuous clouds | 108 |
| 3.5 | Combinations of uncertainty representations into higher order models | 109 |

| | | |
|----------|---|------------|
| 3.5.1 | A quick review of the literature | 110 |
| 3.5.2 | Fuzzy random variables | 112 |
| 3.6 | Conclusions and perspectives | 115 |
| 4 | Treating multiple sources of information | 119 |
| 4.1 | Basics of Information fusion in uncertainty | 121 |
| 4.1.1 | A classification of fusion operators | 121 |
| 4.1.2 | Mathematical properties of fusion operators | 122 |
| 4.1.3 | Basic fusion operators in uncertainty theories | 127 |
| 4.2 | Treating the conflict by adaptive rules using maximal coherent subsets | 147 |
| 4.2.1 | Maximal coherent subsets (MCS) rule: basic methodology | 149 |
| 4.2.2 | Level-wise MCS on the real line with possibility distributions | 152 |
| 4.3 | Towards a cautious conjunctive rule in random set theory | 166 |
| 4.4 | Assessing sources quality: a general framework | 172 |
| 4.4.1 | Rational requirements and general methodology | 172 |
| 4.4.2 | Evaluating sources in probability | 173 |
| 4.4.3 | Evaluating sources in uncertainty theories | 176 |
| 4.5 | Conclusions and perspectives | 181 |
| 5 | Independence and uncertainty | 185 |
| 5.1 | Finding our way in the jungle of independence: towards a taxonomy | 187 |
| 5.1.1 | Judgment of (ir)relevance: a classification | 187 |
| 5.1.2 | (Ir)relevance in imprecise probability theories: first steps towards a taxonomy | 191 |
| 5.2 | Relating irrelevance notions to event-trees: first results | 206 |
| 5.2.1 | Event-trees | 206 |

| | | |
|----------|---|------------|
| 5.2.2 | Probability trees | 207 |
| 5.2.3 | Imprecise probability trees | 209 |
| 5.2.4 | Forward irrelevance in event trees | 210 |
| 5.2.5 | Usefulness and meaningfulness of the result | 213 |
| 5.3 | A consonant approximation of consonant and independent random sets | 215 |
| 5.4 | Conclusions and perspectives | 220 |
| 6 | Decision Making | 223 |
| 6.1 | Decision making in uncertainty theories | 224 |
| 6.1.1 | Classical expected utility | 225 |
| 6.1.2 | Decision making in imprecise probability theory | 226 |
| 6.1.3 | Decision making in random set theory | 228 |
| 6.2 | Practical computations of lower/upper expectations: the case of p-boxes . . . | 229 |
| 6.2.1 | General problem statement and proposed solutions | 230 |
| 6.2.2 | Unimodal u_a | 232 |
| 6.2.3 | Many extrema | 235 |
| 6.3 | Decision in industrial risk analysis | 237 |
| 7 | Illustrative applications | 239 |
| 7.1 | Information evaluation and fusion applied to nuclear computer codes | 239 |
| 7.1.1 | Introduction to the problem | 240 |
| 7.1.2 | Modeling the information | 242 |
| 7.1.3 | Evaluating the sources | 243 |
| 7.1.4 | Merging the information supplied by the sources | 245 |
| 7.2 | Hybrid propagation with numerical accuracy | 248 |
| 7.2.1 | RaFu method: efficiency in numerical hybrid propagation | 248 |

| | | |
|----------|--|------------|
| 7.2.2 | Case-study | 257 |
| 7.2.3 | Modeling uncertainty sources | 257 |
| 7.2.4 | RaFu method application | 259 |
| 8 | Conclusions, perspectives and open problems | 263 |
| A | Uncertainty theories: a short introduction | 267 |
| A.1 | Probability theory: a short introduction | 267 |
| A.2 | Imprecise probability theory | 269 |
| A.3 | Random set theory | 271 |
| A.3.1 | Shafer's belief functions and Smet's TBM | 271 |
| A.3.2 | Dempster's random sets | 272 |
| A.3.3 | Random sets as credal sets | 273 |
| A.4 | Possibility theory | 274 |
| B | Some notions of order theory | 277 |
| C | Random sets: inclusion, least commitment principle, pignistic probability | 279 |
| C.1 | Inclusion relations between random sets | 279 |
| C.2 | Least-commitment principle (LCP) | 281 |
| C.3 | Pignistic probability $BetP$ | 281 |
| D | Proofs | 283 |
| D.1 | Proofs of Section 3.2 | 283 |
| D.2 | Proofs of Section 3.3 | 288 |
| D.3 | Proofs of Section 5.3 | 294 |
| D.4 | Proofs of Section 6.2 | 296 |

| | | |
|----------|--|------------|
| E | Remarks on Nested-Disjoint clouds | 301 |
| F | Generalized p-boxes on complete chain | 303 |
| F.1 | Characterization of generalized p-boxes | 304 |
| F.1.1 | Definition | 304 |
| F.1.2 | The Field \mathcal{H} | 305 |
| F.2 | The lower expectation of p-boxes: A Choquet Integral Representation | 306 |
| F.3 | Approximating lower expectation by limits of p-boxes | 309 |
| F.4 | Relating random sets with p-boxes on the unit interval | 311 |
| F.4.1 | Lower expectation on general functions | 313 |
| F.4.2 | Lower expectation on continuous functions | 314 |
| F.5 | Conclusions | 316 |
| G | (Ir)relevance statements, structural judgments and event trees | 317 |
| G.1 | Forward irrelevance, strong independence and repetition independence | 317 |
| G.2 | Towards a characterization of epistemic independence in event-trees | 319 |
| | Bibliography | 323 |
| | List of Figures | 341 |
| | List of Tables | 345 |

List of notations

General

| | |
|---|---|
| \mathbb{R} | The real line |
| X, Y, X_i | Variables |
| $\mathcal{X}, \mathcal{Y}, \mathcal{X}_i$ | Spaces on which variables assume values |
| x, y, x_i | Elements or values of a space |
| A, E, A_i, E_i | Subsets of a space |
| $ \mathcal{X} , A $ | Cardinality (length) of a space \mathcal{X} or of a subset (interval) A |
| $\wp(\mathcal{X})$ | Power set of \mathcal{X} |
| σ | Permutation |
| Σ_σ | Set of possible permutations |
| $\mathbf{1}_{(A)}$ | Indicator function of event A |
| $[[N]]$ | The set of positive natural numbers $\{1, \dots, N\}$ |
| \mathcal{H} | Convex Hull |
| I, I_i | Closed interval of real numbers |
| $\mathcal{L}(\mathcal{X})$ | Set of all real-valued bounded functions on \mathcal{X} |

Uncertainty representations

| | |
|----------------------------|---|
| μ | (additive) Capacity |
| $\underline{\mu}$ | Sub-additive capacity or lower confidence measure |
| $\bar{\mu}$ | Super-additive capacity or upper confidence measure |
| $\mathbb{P}_{\mathcal{X}}$ | Set of all probability distributions on \mathcal{X} |
| $p(p(x))$ | Probability mass/assignment (of element x) |

| | |
|---|---|
| $P (P(A))$ | Probability distribution (Probability measure of event (subset) A) |
| $E (E(f))$ | Expected value (of function f) with respect to P |
| \mathcal{P} | Credal set (closed convex set of probabilities) |
| ext \mathcal{P} | Set of extreme probability assignments of credal set \mathcal{P} |
| $\underline{P}, \bar{P} (\underline{P}(A), \bar{P}(A))$ | Lower/upper probabilities (of event A) |
| $\underline{E}, \bar{E} (\underline{E}(f), \bar{E}(f))$ | Lower/upper expectations (of function f) |
| π | Possibility distribution or cloud upper distribution |
| δ | Sufficiency distribution or cloud lower distribution |
| $[\pi, \delta]$ | Cloud |
| L | Set of imprecise probability assignments |
| $[\underline{F}, \bar{F}]$ | (Generalized) p-box |
| (m, \mathcal{F}) | Random set |
| m | Basic probability assignment |
| \mathcal{F} | Set of focal elements |
| $(m, \widetilde{\mathcal{F}})$ | Fuzzy random variable/Fuzzy belief structure |
| $\widetilde{\mathcal{F}}$ | Set of fuzzy focal elements |

Information fusion

| | |
|-----------|---|
| φ | General fusion operator |
| \top | T-norm |
| \perp | T-conorm |
| λ | Weight |
| \odot | Conjunctive combination rule of random sets |
| \otimes | Dempster combination rule |
| \oplus | Dubois and Yager's rule of combination |
| \wedge | Cautious merging rule of random sets |
| K | Maximal Coherent Subset |

*“Writing in English is the most ingenious torture
ever devised for sins committed in previous lives.”*

— James Joyce (1882–1941)

*“They were born into a world that was against them in a thousand little ways, and then
devoted most of their energies to making it worse”*

— Terry Pratchett (1948–?) and Neil Gaiman (1960–?)

Chapter 1

Résumé Français de la thèse (French Summary of the thesis)

If you cannot read French (or can also read English and prefer to skip this summary), please go to p.41.

1.1 Introduction (Chapitre 2)

Ce travail de thèse présente des résultats relatifs à la représentation et au traitement des incertitudes entourant la valeur que peut prendre une variable, cette incertitude pouvant provenir soit de la variabilité intrinsèque des phénomènes influençant la valeur de cette variable, soit de l'imprécision ou du manque de fiabilité des informations disponibles.

Ici, nous nous intéressons aux cas où l'incertitude est représentée par des modèles numériques qui ne sont ni des probabilités précises (parce que l'information disponible est trop pauvre), ni de simples ensembles de valeurs (parce que nous disposons d'informations permettant de savoir quels éléments sont plus à même d'être observés).

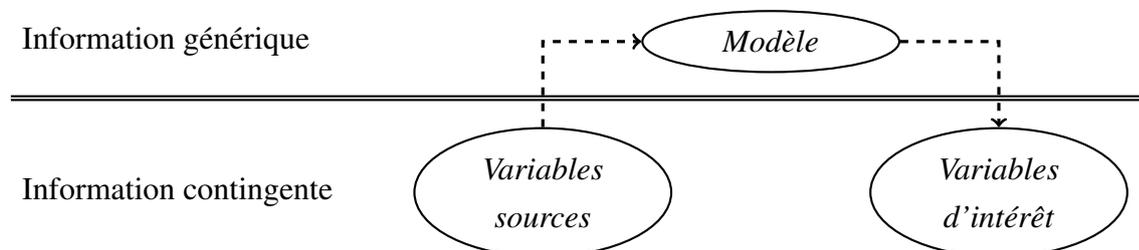
Afin de répondre à ce type de problème, différentes théories de l'incertain ont émergé ces dernières années. Il s'agit, entre autres, des théories des possibilités [85], des ensembles aléatoires [151] et des probabilités imprécises [203] (il s'agit là des trois théories principales sur lesquelles nous allons nous concentrer). Par rapport à ces théories, notre position est double:

- d'une part, nous attachons une grande importance aux aspects permettant d'unifier ces théories et le traitement de l'incertitude qui en découle.
- d'autre part, nous pensons que chacune de ces théories possède son propre intérêt, et est à même de répondre à des questions pour lesquelles d'autres théories apportent parfois des réponses insatisfaisantes. Dans ce sens, nous pensons que la question essentielle n'est pas de savoir de manière absolue quelle théorie est "meilleure" que les autres, mais plutôt de chercher à savoir quelle théorie s'adapte le mieux à une situation donnée.

A cet effet, nous nous attacherons, tout au long de ce travail, à souligner les points communs entre les différentes théories, tout en apportant des solutions pratiques, parfois particulières à l'une ou l'autre théorie, aux problèmes que peut poser le traitement des incertitudes. Nous nous concentrerons plus particulièrement sur les problèmes courants posés par les études de sûreté nucléaire.

Par incertitude, nous entendons donc les situations où l'information ne permet pas d'identifier de manière exacte l'état d'un système ou la valeur d'une variable. Par traitement, nous entendons la manipulation de l'information disponible de manière raisonnée, afin d'en déduire d'autres informations potentiellement utiles. Nous différencions également deux niveaux différents d'informations: un niveau générique, qui concerne les connaissances et modèles généraux; un niveau contingent, regroupant les informations propres à une situation particulière. Par exemple, un modèle physique d'écoulement de fluide constituera de l'information générique, tandis que la vitesse d'un fluide dans une expérience donnée sera de l'information contingente.

Afin de pouvoir décrire facilement les traitements auxquels nous nous intéresserons durant ce travail, nous considérons le modèle simplifié, donné par la figure suivante:



Traitement de l'incertitude: cadre général

Les variables sources sont celles à propos desquelles nous possédons de l'information. Le modèle décrit les liens (génériques) qui existent entre variables sources et variables d'intérêt,

et permet d'obtenir de l'information sur ces dernières à partir d'information sur les variables sources. Enfin, les variables d'intérêt sont les variables sur lesquels nous voulons obtenir une(des) information(s) utile(s) permettant de résoudre un problème donné.

A partir de cette figure, nous pouvons définir un certain nombre de problèmes relatifs au traitement de l'incertitude:

- **Modélisation:** construction d'un modèle générique à partir d'informations contingentes (observations particulières). Il s'agit d'un procédé de type inductif, souvent appelé apprentissage dans le domaine de l'intelligence artificielle (IA) ou inférence paramétrique dans le domaine des statistiques.
- **Inférence:** tirer des conclusions plausibles à partir d'informations disponibles. Il s'agit d'un procédé déductif et impersonnel, qui consiste ici à tirer des conclusions sur les variables d'intérêt à partir d'informations sur les variables sources, par le biais d'un modèle générique. En statistique, ce type d'opération est souvent associée au problème de prédiction. Les problèmes d'inférences comprennent:
 - **propagation directe à travers un modèle déterministe:** propager les incertitudes sur les entrées (variables sources) d'un modèle déterministe (i.e. fonction) pour estimer les incertitudes sur les sorties (variables d'intérêt). Il s'agit du type d'inférence le plus souvent fait en analyse de risque et en études de sûreté. Notons que cette inférence est monotone, dans le sens où plus l'incertitude sur les entrées est petite, plus celle sur les sorties l'est également.
 - **propagation inverse à travers un modèle déterministe:** similaire à la propagation directe, excepté que les variables sources sont maintenant les sorties, et qu'il faut estimer l'incertitude sur les entrées. La difficulté pour ce genre d'inférence est que le modèle est très rarement inversible, et que les dépendances entre les entrées sont généralement mal connues. Comme la propagation directe, cette opération est monotone.
 - **propagation/conditionnement sur un modèle stochastique:** à partir d'une observation sur les variables sources, inférer les valeurs plausibles des variables d'intérêt en propageant cette information à travers un modèle générique stochastique (i.e. chaîne de Markov, réseaux de Bayes). Dans ce cas, l'incertitude concerne le modèle, et non plus les variables. Ce type d'inférence se rencontre plus souvent dans le domaine de l'IA. Notons également que cette opération n'est pas forcément monotone, et qu'une information plus précise sur les variables sources

peut augmenter l'incertitude sur les variables d'intérêt. Ce phénomène est plus connu sous le nom de dilation.

- **Fusion d'informations:** action de synthétiser l'information provenant de plusieurs sources en un message simple et interprétable, tout en prenant en compte l'inconsistance entre les informations et les éventuelles dépendances entre les sources. La fusion d'information s'opère entre informations de même niveau de généralité.
- **Prise de décision:** action de déterminer l'ensemble des actions optimales à prendre dans une situation donnée et en fonction des informations disponibles. Au contraire de l'inférence, c'est un procédé personnel (le sens d'optimal peut dépendre de la personne prenant la décision) et qui a un impact sur le monde environnant une fois la décision prise. Néanmoins, inférence et prise de décisions, même si elles sont différentes, sont souvent liées par le fait que les résultats d'une inférence sont souvent utilisés pour prendre une décision.
- **Révision:** action de modifier nos croyances ou connaissances avec l'arrivée de nouvelles informations, pas forcément cohérentes avec les croyances ou connaissances initiales. De même que la fusion d'information, réviser se fait entre informations de même niveau de généralité.

Bien entendu, il est difficile de rendre compte avec cette figure et ces descriptions relativement simples de la complexité présente dans des applications réelles. En pratique, il peut être difficile de déterminer quelle est la meilleure réponse à apporter à un problème, ou encore quel est le niveau de généralité de tel ou tel type d'information. Néanmoins, de telles figures simplifiées peuvent servir de point de départ aux réflexions qui détermineront ensuite le traitement le plus adéquat à appliquer à une situation.

Dans ce travail, nous ne nous intéresserons qu'à certains des problèmes évoqués plus haut. Plus particulièrement, nous nous concentrerons sur des problèmes souvent rencontrés en études de sûreté ou en analyse de risques.

Le chapitre 3 s'intéresse au problème de représenter l'incertitude entourant la valeur d'une variable. Une attention toute particulière est réservée aux représentations simples et pratiques, qui sont les plus souvent utilisées dans les applications. En particulier, nous étudions les relations entre les représentations suivantes : distributions de possibilités, distributions imprécises de probabilités, p-boxes, nuages et ensembles aléatoires. Afin de faciliter leur comparaison, nous introduisons un modèle dit de p-boxes généralisées.

Le chapitre 4 concerne le cas où de multiples sources fournissent des informations à propos d'une même variable. Nous étudions d'abord comment cette information peut être synthétisée en un message simple, en donnant une attention particulière aux problèmes du traitement de l'inconsistance entre les informations et de la prise en compte de dépendances entre les sources. Dans la seconde partie du chapitre, nous discutons d'une méthode permettant d'évaluer la qualité de l'information fournie par les sources et donc, dans un certain sens, leur fiabilité.

Dans le chapitre 5, nous étudions les notions d'indépendance qui peuvent exister entre différentes variables. En effet, si dans les probabilités classiques toutes les notions d'indépendance sont formellement équivalentes à la définition de l'indépendance stochastique, indépendamment de leur interprétation, ce n'est plus vrai lorsque des modèles probabilistes imprécis sont utilisés. Dans ce dernier cas, il existe autant de définitions formelles que d'interprétations. Puisque la notion d'indépendance est centrale dans la construction de modèles joint d'incertitude à partir de modèles marginaux (une situation qui arrive souvent en analyse de risque), nous étudions et esquissons un premier cadre général dans lequel situer les différentes notions d'indépendances rencontrées en probabilité imprécise. La question de relier ces notions aux arbres d'événements est brièvement abordée.

Le chapitre 6 est consacré au problème de la prise de décision. Après un bref compte-rendu des différents critères étendant aux probabilités imprécises le critère classique de la maximisation de l'espérance mathématique, nous donnons quelques résultats pratiques relatifs aux calculs de ces espérances lorsque l'incertitude est modélisée par une p-box définie sur les réels.

Finalement, le chapitre 7 expose deux applications réalisées dans le cadre de la thèse au moyen du logiciel de traitement des incertitudes SUNSET développé par l'IRSN. La première concerne l'application des méthodes présentées au chapitre 4 aux résultats d'études d'incertitude réalisées sur des codes de calculs simulant la rupture d'un système de refroidissement dans un réacteur nucléaire. La seconde concerne l'application à un cas d'étude d'une méthode numérique de propagation des incertitudes, dénommée RaFu et actuellement utilisée par l'IRSN. Nous décrivons d'abord la méthode, puis les résultats obtenus sur le cas d'étude.

1.2 Représentations pratiques d'incertitude (Chapitre 3)

Dans ce chapitre, nous nous intéressons aux représentations simples et pratiques qui permettent de modéliser notre incertitude à propos de la valeur que peut prendre une variable X dans un espace \mathcal{X} . Parmi ces représentations simples, on trouve: les ensembles classiques, les distributions de probabilité [108] et de possibilités [85], les distributions imprécises de probabilités [42], les ensembles aléatoires [151], les boîtes de probabilités (communément appelées p-boxes) [104], les variables aléatoires floues [34] et, plus récemment, les nuages [159].

Ces représentations, du fait de leur simplicité, facilitent souvent la manipulation des incertitudes, notamment en termes calculatoires. Elles sont également utiles pour résumer des résultats complexes, ou pour éliciter¹ des informations. Néanmoins, afin de manipuler correctement ces représentations, il est nécessaire de les comparer et d'établir des liens entre elles, ces liens pouvant également montrer comment des outils de différentes théories peuvent être appliqués à une même représentation.

Débuter cette comparaison et établir de tels liens sont les objets de ce chapitre. Comme toutes les représentations étudiées ici peuvent s'interpréter comme des cas particuliers d'ensembles convexes de distributions de probabilités, nous utiliserons ce langage pour pouvoir relier et comparer les différentes représentations².

Nous nous intéressons plus particulièrement aux deux représentations plus récentes que sont les p-boxes et les nuages, et dont les liens avec les autres représentations pratiques d'incertitude ont été peu explorés jusqu'ici.

1.2.1 Mesures non-additives et représentations connues

Nous introduisons d'abord les outils mathématiques et représentations connus permettant de modéliser explicitement l'imprécision présente dans l'information. Contrairement aux probabilités classiques, où une seule mesure est utilisée, ces représentations modélisent l'incertitude au moyen de deux mesures conjuguées, l'une représentant l'idée de certitude, l'autre de plausibilité. L'importance de l'imprécision peut ensuite être mesurée par la différence entre ces deux mesures (les probabilités étant retrouvées lorsque les deux mesures coïncident).

¹On appelle élicitation la procédure qui consiste à demandé une évaluation de son incertitude à un expert

²notons néanmoins que l'interprétation en termes d'ensembles de probabilités n'est pas la seule possible, comme le montre l'Appendice A

1.2.1.1 Capacités

Les fonctions d'ensembles que sont les *capacités* [25] sont utiles pour représenter l'incertitude.

Definition 1.1. Etant donné un ensemble fini \mathcal{X} , une capacité sur \mathcal{X} est une fonction μ , définie sur les sous-ensembles $\wp(\mathcal{X})$ de \mathcal{X} , telle que:

- $\mu(\emptyset) = 0, \mu(\mathcal{X}) = 1$ (conditions aux bornes)
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ (monotonie)

Une capacité vérifiant

$$\forall A, B \subseteq \mathcal{X}, A \cap B = \emptyset, \mu(A \cup B) \geq \mu(A) + \mu(B) \quad (1.1)$$

est dite *super-additive*. La notion duale, appelée *sous-additivité*, est obtenue en renversant l'inégalité dans l'équation ci-dessus.

Etant donné une capacité μ sur \mathcal{X} , sa capacité conjuguée μ^c est définie par $\mu^c(E) = \mu(\mathcal{X}) - \mu(E^c) = 1 - \mu(E^c)$ pour tout ensemble $E \subset \mathcal{X}$ avec E^c le complément de E . Une capacité est dite additive si l'inégalité de l'équation (1.1) devient une égalité. Une capacité additive est sa propre conjuguée ($\mu = \mu^c$), et est une mesure de probabilité P .

Quand elles sont utilisées pour représenter l'incertitude, les valeurs d'une capacité mesurent le degré de confiance dans le fait qu'un événement va être observé. Dans ce cadre, les capacités super-additives modélisent l'idée de certitude (puisque $\mu(E) + \mu(E^c) \leq 1$), tandis que les sous-additives modélisent l'idée de plausibilité ($\mu(E) + \mu(E^c) \geq 1$).

Le noyau \mathcal{P}_μ d'une capacité super-additive μ définie sur \mathcal{X} est l'ensemble (convexe) des mesures de probabilités qui la domine:

$$\mathcal{P}_\mu = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall E \subseteq \mathcal{X}, \mu(E) \leq P(E)\}.$$

avec $\mathbb{P}_{\mathcal{X}}$ l'ensemble des mesures de probabilité définies sur \mathcal{X} . Par dualité, le noyau est également l'ensemble des mesures de probabilités dominées par la capacité conjuguée μ^c (sous-additive), ce qui veut dire que, par la suite, l'on se concentrera exclusivement sur l'une ou l'autre de ces capacités (typiquement, la super-additive). Notons que le noyau peut être vide, et un moyen de s'assurer qu'il ne l'est pas consiste à faire appel à des propriétés des capacités modélisant l'incertitude, telle la n -monotonie:

Definition 1.2. Une capacité μ super-additive sur \mathcal{X} est n -monotone, avec $n > 0$ et $n \in \mathbb{N}$, si et si seulement si pour toute collection $\mathcal{A} = \{A_i \subseteq \mathcal{X} \mid i \in \mathbb{N}, 0 < i \leq n\}$ d'événements A_i , l'équation

$$\mu\left(\bigcup_{A_i \in \mathcal{A}} A_i\right) \geq \sum_{I \subseteq \mathcal{A}} (-1)^{|I|+1} \mu\left(\bigcap_{A_i \in I} A_i\right)$$

est vérifiée.

Si une capacité est n -monotone, alors elle est assurée d'être $(n - 1)$ -monotone, mais pas forcément $(n + 1)$ -monotone. Une capacité est dite ∞ -monotone si elle est monotone pour tout n . Une condition suffisante (mais pas nécessaire) pour qu'une capacité ait un noyau non-vidé soit qu'elle soit 2-monotone. A partir d'une capacité, il est possible de définir de nombreuses transformations bijectives [115], notamment la transformée de Möbius:

Definition 1.3. Etant donné une capacité μ sur \mathcal{X} , sa transformée de Möbius est la fonction $m : \wp(X) \rightarrow \mathbb{R}$ des sous-ensemble de \mathcal{X} vers les réels, qui associe à chaque sous-ensemble E de \mathcal{X} la valeur

$$m(E) = \sum_{B \subseteq E} (-1)^{|E \setminus B|} \mu(B)$$

Notons que la fonction m est non-négative si et seulement si la capacité μ est ∞ -monotone [178, Ch.2.7]. Dans ce dernier cas, nous l'appelons distribution de masse. La transformée de Möbius d'une mesure de probabilité est sa distribution de probabilité (m est positive et non-nulle uniquement sur les singletons).

1.2.1.2 Ensembles de probabilités

Walley [203] considère une représentation encore plus générale de l'incertitude par des paires de bornes duales (inférieures/supérieures). Au lieu de se restreindre à des événements (sous-ensembles), il étend ses mesures à des bornes sur les espérances mathématiques de fonctions réelles et bornées de \mathcal{X} (les événement correspondant alors à des fonctions indicatrices). Il montre que ce langage est équivalent (en terme d'expressivité) à celui consistant à modéliser l'incertitude par des ensembles convexes de mesures de probabilités, dénotés ici \mathcal{P} [136].

Ce langage étant très général, nous nous en servons pour comparer les représentations pratiques que nous considérons, et nous adoptons la terminologie suivante:

Definition 1.4. Soit \mathbb{F}_1 et \mathbb{F}_2 deux cadres généraux de représentation d'incertitude, a et b deux instances particulières de ces cadres, et $\mathcal{P}_a, \mathcal{P}_b$ les ensembles de probabilités induits par a et b . Alors:

- Nous disons que le cadre \mathbb{F}_1 *généralise* cadre \mathbb{F}_2 si et seulement si pour tout $b \in \mathbb{F}_2$, $\exists a \in \mathbb{F}_1$ tel que $\mathcal{P}_a = \mathcal{P}_b$ (ou, également, que \mathbb{F}_2 est un cas particulier de \mathbb{F}_1).
- Nous disons que le cadre \mathbb{F}_1 et \mathbb{F}_2 sont *équivalents* si et seulement si pour tout $b \in \mathbb{F}_2$, $\exists a \in \mathbb{F}_1$ tel que $\mathcal{P}_a = \mathcal{P}_b$ et inversement.
- Nous disons que le cadre \mathbb{F}_2 est *représentable* par le cadre \mathbb{F}_1 si et seulement si pour tout $b \in \mathbb{F}_2$, il existe une collection $\{a_1, \dots, a_k | a_i \in \mathbb{F}_1\}$ telle que $\mathcal{P}_b = \mathcal{P}_{a_1} \cap \dots \cap \mathcal{P}_{a_k}$.
- Nous disons qu'un élément $a \in \mathbb{F}_1$ *approche extérieurement* (intérieurement) un élément $b \in \mathbb{F}_2$ si et seulement si $\mathcal{P}_b \subseteq \mathcal{P}_a$ ($\mathcal{P}_a \subseteq \mathcal{P}_b$).

Dans ce travail, nous pouvons nous restreindre aux ensembles induits par des bornes de probabilités sur les événements.

Nous définissons une probabilité inférieure \underline{P} comme une capacité super-additive. L'ensemble de probabilités $\mathcal{P}_{\underline{P}}$ lui correspondant est alors le noyau de cette capacité. Nous considérons ici des probabilités inférieures dites cohérentes, c'est-à-dire des probabilités inférieures qui sont les enveloppes de l'ensemble de probabilités induit (i.e. pour tout ensemble $A \subseteq \mathcal{X}$, nous avons $\underline{P}(A) = \min_{P \in \mathcal{P}_{\underline{P}}} (P(A))$). En d'autres termes, les bornes fournies pour modéliser l'incertitude sont atteintes par $\mathcal{P}_{\underline{P}}$ et sont optimales (i.e. elles ne peuvent être réduites sans réduire $\mathcal{P}_{\underline{P}}$). Néanmoins, ces ensembles, même s'ils constituent des cas particuliers de modèles plus généraux, peuvent rester difficile à manipuler du fait de leur complexité. Deux exemples de modèles moins généraux introduits par leurs auteurs avec l'intention de fournir des outils pratiques de manipulation d'incertitude sont les boîtes de probabilités (P-boxes) ainsi que les distributions imprécises de probabilités.

P-boxes Une p-box $[\underline{F}, \overline{F}]$ est définie comme une paire de distributions cumulées définies sur les réels, telles que $\underline{F} \leq \overline{F}$ (\underline{F} domine stochastiquement \overline{F}). Une p-box $[\underline{F}, \overline{F}]$ induit l'ensemble de probabilités $\mathcal{P}_{[\underline{F}, \overline{F}]} = \{P \in \mathbb{P}_{\mathbb{R}} | \forall r \in \mathbb{R}, \underline{F}(r) \leq P((-\infty, r]) \leq \overline{F}(r)\}$, et il est utile de noter qu'une p-box consiste à donner des bornes de confiance sur des intervalles emboîtés $(-\infty, r]$.

Contrairement aux distributions cumulées uniques [28], les p-boxes permettent aux experts d'exprimer leur opinion sur la valeur de percentiles de manière imprécise (en fournissant

des intervalles au lieu de valeurs uniques). Il existe également des outils numériques efficaces [209] permettant de manipuler les p-boxes pour faire rapidement des propagations conservatrices d'incertitude.

Distributions imprécises de probabilité Une distribution imprécise de probabilité consiste en un ensemble L d'intervalles $L = \{[l(x), u(x)] | x \in \mathcal{X}\}$ tels que $l(x) \leq p(x) \leq u(x)$ pour tout x . Cet ensemble L décrit notre connaissance imprécise sur les probabilités des singletons, et induit l'ensemble de probabilités $\mathcal{P}_L = \{P \in \mathbb{P}_{\mathcal{X}} | \forall x \in \mathcal{X}, l(x) \leq p(x) \leq u(x)\}$. Comme l'ont montré De Campos *et al.* [42], se restreindre à un tel ensemble de contraintes a de nombreux avantages calculatoires. C'est également une représentation très utile dans le cas de données multinomiales, ou pour la modélisation d'histogrammes imprécis.

1.2.1.3 Ensembles aléatoires

Ici, nous définissons un ensemble aléatoire (discret), noté (m, \mathcal{F}) , comme une fonction, appelée distribution de masse, $m : \wp(\mathcal{X}) \rightarrow [0, 1]$ des sous-ensembles de \mathcal{X} dans l'intervalle unité, non-négative et normée ($\sum_{E \subseteq \mathcal{X}} m(E) = 1$). Un sous-ensemble E ayant une masse positive est appelé ensemble focal, et nous notons \mathcal{F} l'ensemble des éléments focaux d'un ensemble aléatoire. A partir de cette fonction, Shafer [178] définit trois fonctions, respectivement de croyance, plausibilité et commonalité:

$$Bel(A) = \sum_{E, E \subseteq A} m(E) \quad (\text{Belief}).$$

$$Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E) \quad (\text{Plausibility}).$$

$$Q(A) = \sum_{E, E \supseteq A} m(E) \quad (\text{Commonality}).$$

La fonction de croyance Bel ainsi définie est une capacité ∞ -monotone, et m est sa transformée de Möbius. Nous pouvons définir l'ensemble de probabilité

$$\mathcal{P}_{(m, \mathcal{F})} = \{P \in \mathbb{P}_{\mathcal{X}} | \forall A \subseteq \mathcal{X}, Bel(A) \leq P(A) \leq Pl(A)\}$$

induit par un ensemble aléatoire.

Les ensembles aléatoires constituent donc des cas particuliers de probabilités inférieures. A l'inverse, ils généralisent les p-boxes [132]. Il n'existe pas de lien précis entre ensembles aléatoires et distributions imprécises de probabilité, dans le sens où l'un ne généralise pas

l'autre, et inversement. Cependant, de nombreux auteurs ont étudié comment une représentation pouvait être approchée par l'autre [135, 60, 42]

Un des intérêt applicatif majeur des ensembles aléatoires est qu'ils peuvent être vus comme des distributions de probabilité portant sur des ensembles, ce qui implique que les méthodes de simulations du type Monte-Carlo peuvent facilement leur être appliquées. Quand ils sont définis sur les réels, il est courant de restreindre les ensembles focaux à un nombre fini d'intervalles, ce qui permet d'étendre les résultats de l'analyse d'intervalles [152] aux intervalles aléatoires [91].

1.2.1.4 Distributions de possibilités

Une distribution de possibilité est une fonction $\pi : \mathcal{X} \rightarrow [0, 1]$ de l'espace \mathcal{X} dans l'intervalle unité, telle que $\pi(x) = 1$ pour au moins un élément de \mathcal{X} . A partir de cette distribution sont définies plusieurs mesures [79], parmi lesquelles les mesures de possibilité, nécessité et suffisance:

$$\begin{aligned}\Pi(A) &= \sup_{x \in A} \pi(x) \\ N(A) &= 1 - \Pi(A^c) \\ \Delta(A) &= \inf_{x \in A} \pi(x).\end{aligned}$$

Leur propriétés caractéristiques sont $N(A \cap B) = \min(N(A), N(B))$ et $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ pour toute paire d'événements A, B de \mathcal{X} . Etant donné un degré $\alpha \in [0, 1]$, les α -coupes strictes et régulières $A_{\bar{\alpha}}$ et A_{α} sont les ensembles $A_{\bar{\alpha}} = \{x \in \mathcal{X} \mid \pi(x) > \alpha\}$ et $A_{\alpha} = \{x \in \mathcal{X} \mid \pi(x) \geq \alpha\}$. Nous notons $\alpha_0 = 0 < \alpha_1 < \dots < \alpha_M = 1$ l'ensemble fini des valeurs distinctes prises par π .

La mesure de nécessité étant une capacité ∞ -monotone, une distribution de possibilité π constitue un cas particulier d'ensemble aléatoire, et définit l'ensemble aléatoire $(m, \mathcal{F})_{\pi}$ dont les ensembles focaux E_i de masse $m(E_i)$ sont, pour $i = 1, \dots, M$ [82]:

$$\begin{cases} E_i = \{x \in \mathcal{X} \mid \pi(x) \geq \alpha_i\} = A_{\alpha_i} \\ m(E_i) = \alpha_i - \alpha_{i-1} \end{cases} \quad (1.2)$$

Une distribution de possibilité π induit donc également un ensemble de probabilités \mathcal{P}_{π} tel que

$$\mathcal{P}_{\pi} = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall A \subseteq \mathcal{X}, N(A) \leq P(A) \leq \Pi(A)\}.$$

Les distributions de possibilité sont les modèles les plus simples pouvant modéliser explicitement l'imprécision. Cette simplicité fait qu'elles sont très faciles à manipuler, mais aussi peu expressives par rapport à d'autres modèles. Cependant, il existe de nombreux cas où elles sont utiles; Par exemple, le fait qu'elles puissent se voir comme des bornes de confiance inférieures d'ensembles emboîtés les rend très pratiques pour éliciter de l'information, ou pour représenter des intervalles de confiance centrés autour d'une valeur [11] (représentation très souvent utilisée en statistique).

1.2.2 P-boxes généralisées

Les représentations précédentes sont connues depuis un certain temps et ont donc déjà été étudiées par de nombreux auteurs. Nous proposons et étudions maintenant une nouvelle représentation consistant en une généralisation des p-boxes à des espaces \mathcal{X} discrets arbitraires. Il y a (au moins) deux bonnes raisons pour étudier une telle généralisation: premièrement, les p-boxes étant déjà très utiles lorsque définies sur l'espace des réels, il semble normal de vouloir les généraliser; deuxièmement, nous verrons que les p-boxes généralisées sont très utiles pour étudier les nuages proposés par Neumaier [159]. Ces derniers ayant été proposés récemment pour représenter l'imprécision présente dans l'information, il est nécessaire de les positionner par rapport à d'autres représentations, travail que nous réalisons ici.

Rappelons d'abord que deux fonctions f et f' sont comonotones si et seulement si pour toute paire d'éléments $x, y \in \mathcal{X}$, nous avons $f(x) < f(y) \Rightarrow f'(x) \leq f'(y)$. En d'autres termes, il existe une permutation σ de $\{1, 2, \dots, n\}$ telle que $f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \dots \geq f(x_{\sigma(n)})$ et $f'(x_{\sigma(1)}) \geq f'(x_{\sigma(2)}) \geq \dots \geq f'(x_{\sigma(n)})$. Nous définissons alors une p-box généralisée comme:

Definition 1.5. Une p-box généralisée $[\underline{F}, \overline{F}]$ sur un espace \mathcal{X} est une paire de fonctions comonotones $\underline{F}, \overline{F}, \underline{F} : \mathcal{X} \rightarrow [0, 1]$ et $\overline{F} : \mathcal{X} \rightarrow [0, 1]$ de \mathcal{X} vers $[0, 1]$ telles que \underline{F} est plus petite que \overline{F} (i.e. $\underline{F} \leq \overline{F}$) et il existe au moins un élément x de \mathcal{X} pour lequel $\overline{F}(x) = \underline{F}(x) = 1$.

Et, à partir d'une p-box généralisée $[\underline{F}, \overline{F}]$, nous pouvons définir un pré-ordre complet $\leq_{[\underline{F}, \overline{F}]}$ sur \mathcal{X} tel que $x \leq_{[\underline{F}, \overline{F}]} y$ si $\underline{F}(x) \leq \underline{F}(y)$ et $\overline{F}(x) \leq \overline{F}(y)$, grâce à la condition de comonotonité. Pour simplifier les notations, nous considérons que les éléments de \mathcal{X} sont indicés tels que $i < j$ implique $x_i \leq_{[\underline{F}, \overline{F}]} x_j$, et que $|\mathcal{X}| = n$. Nous définissons ensuite l'ensemble de probabilités induit par la p-box généralisée comme

$$\mathcal{P}_{[\underline{F}, \overline{F}]} = \{P \in \mathbb{P}_{\mathcal{X}} \mid i = 1, \dots, n, \underline{F}(x_i) \leq P(\{x_1, \dots, x_i\}) \leq \overline{F}(x_i)\}.$$

Notons que si \mathcal{X} est l'ensemble réel, et $\leq_{[\underline{F}, \overline{F}]}$ l'ordre naturel entre les nombres, nous retrouvons les p-boxes classiques.

La proposition suivante montre que l'incertitude décrite par les p-boxes généralisées peut être décrite par des paires de distributions de possibilités comonotones:

Proposition 1.1. *Toute p-box généralisée $[\underline{F}, \overline{F}]$ sur \mathcal{X} est représentable (voir Définition 1.4) par une paire de distributions de possibilités $\pi_{\overline{F}}, \pi_{\underline{F}}$, c'est-à-dire $\mathcal{P}_{[\underline{F}, \overline{F}]} = \mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$, avec:*

$$\pi_{\overline{F}}(x_i) = \beta_i \quad \text{et} \quad \pi_{\underline{F}}(x_i) = 1 - \max \{ \alpha_j \mid j = 0, \dots, i, \alpha_j < \alpha_i \}$$

pour $i = 1, \dots, n$, avec $\alpha_0 = 0$.

A l'inverse, toute distribution de possibilité peut-être vue comme une p-boxe généralisée réduite à sa seule distribution supérieure ou inférieure, ce qui veut dire que les distributions de possibilité sont des cas particuliers de p-boxes généralisées.

La proposition suivante indique que les p-boxes généralisées sont des cas particuliers d'ensembles aléatoires:

Proposition 1.2. *Les p-boxes généralisées sont des cas particuliers d'ensembles aléatoires, c'est-à-dire que pour toute p-box généralisée $[\underline{F}, \overline{F}]$ définie sur \mathcal{X} , il existe toujours un ensemble aléatoire $(m, \mathcal{F})_{[\underline{F}, \overline{F}]}$ tel que $\mathcal{P}_{[\underline{F}, \overline{F}]} = \mathcal{P}_{(m, \mathcal{F})_{[\underline{F}, \overline{F}]}}$.*

et, si nous notons $0 = \gamma_0 < \gamma_1 < \dots < \gamma_M = 1$ les valeurs distinctes prises par les fonctions $\underline{F}, \overline{F}$ de la p-box sur les éléments de \mathcal{X} , cet ensemble aléatoire est défini, pour $j = 1, \dots, M$, comme suit:

$$\begin{cases} E_j = \{x_i \in \mathcal{X} \mid (\pi_{\overline{F}}(x_i) \geq \gamma_j) \wedge (1 - \pi_{\underline{F}}(x_i) < \gamma_j)\} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (1.3)$$

Le lien entre p-boxes généralisées et distributions imprécises de probabilités est moins direct, puisqu'aucune des deux représentations ne généralise l'autre. Considérons d'abord un ensemble L et une indexation (arbitraire) des éléments de \mathcal{X} . Pour $i = 1, \dots, n$, notons $l(x_i) = l_i$ et $u(x_i) = u_i$. Une p-box généralisée approchant extérieurement L peut alors être

construite grâce aux équations suivantes:

$$\begin{aligned} \underline{F}'(x_i) = \underline{P}(A_i) = \alpha'_i &= \max\left(\sum_{x_i \in A_i} l_i, 1 - \sum_{x_i \notin A_i} u_i\right) \\ \overline{F}'(x_i) = \overline{P}(A_i) = \beta'_i &= \min\left(\sum_{x_i \in A_i} u_i, 1 - \sum_{x_i \notin A_i} l_i\right) \end{aligned} \quad (1.4)$$

avec $\underline{P}, \overline{P}$ les probabilités inférieures et supérieures induites par l'ensemble L . Chaque permutation des éléments de \mathcal{X} donne alors une p-box généralisée différente. Maintenant, considérons l'ensemble Σ_σ des permutations σ de \mathcal{X} et $[\underline{F}, \overline{F}]'_\sigma$ une p-box généralisée correspondant à une permutation particulière. La proposition suivante montre que les distributions imprécises de probabilités sont représentables par des p-boxes:

Proposition 1.3. *Soit un ensemble L décrivant une distribution imprécise de probabilité, et $[\underline{F}, \overline{F}]'_\sigma$ la p-box généralisée obtenue avec la permutation σ à partir de L et des équations (1.4). Alors, nous avons la relation suivante:*

$$\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{[\underline{F}, \overline{F}]'_\sigma} \quad (1.5)$$

ce qui nous permet de relier les distributions imprécises de probabilité aux p-boxes généralisées. Maintenant que nous avons positionné cette représentation par rapport aux autres, nous pouvons étudier les nuages de Neumaier [159], qui comme nous allons le voir ont de fortes connections avec les p-boxes généralisées.

1.2.3 Nuages

Definition 1.6. Un nuage est défini par une paire de distributions $\delta : \mathcal{X} \rightarrow [0, 1]$ et $\pi : \mathcal{X} \rightarrow [0, 1]$ de l'espace \mathcal{X} vers $[0, 1]$, telles que δ est inférieure à π (i.e. $\delta \leq \pi$), avec $\pi(x) = 1$ pour au moins un élément x dans \mathcal{X} , et $\delta(y) = 0$ pour au moins un élément y dans \mathcal{X} . δ et π sont respectivement les distributions inférieure et supérieure du nuage.

Notons que, d'un point de vue mathématique, ces nuages sont équivalents à des ensembles flous valués par intervalles assortis de contraintes aux bornes. Plus précisément, le nuage est mathématiquement équivalent à un ensemble flou dont la fonction d'appartenance à comme valeur, pour l'élément x , l'intervalle $[\delta(x), \pi(x)]$. Etudier les nuages en tant que représentation de l'incertitude sur X permet donc également d'apporter un nouvel éclairage sur les interprétations possible à donner à un ensemble flou valué par intervalles.

Neumaier [159] définit un ensemble de probabilités $\mathcal{P}_{[\pi, \delta]}$ correspondant à un nuage $[\pi, \delta]$ comme

$$\mathcal{P}_{[\pi, \delta]} = \{P \in \mathbb{P}_{\mathcal{X}} \mid P(\{x \in \mathcal{X} \mid \delta(x) \geq \alpha\}) \leq 1 - \alpha \leq P(\{x \in \mathcal{X} \mid \pi(x) > \alpha\})\}$$

.

Etant donné l'ensemble fini des M valeurs prises par les distributions δ et π sur \mathcal{X} , notées $0 = \gamma_0 < \gamma_1 < \dots < \gamma_M = 1$, les coupes strictes et régulières sont définies comme

$$B_{\overline{\gamma}_i} = \{x \in \mathcal{X} \mid \pi(x) > \gamma_i\} \text{ et } B_{\gamma_i} = \{x \in \mathcal{X} \mid \pi(x) \geq \gamma_i\} \quad (1.6)$$

pour la distribution supérieure π et

$$C_{\overline{\gamma}_i} = \{x \in \mathcal{X} \mid \delta(x) > \gamma_i\} \text{ et } C_{\gamma_i} = \{x \in \mathcal{X} \mid \delta(x) \geq \gamma_i\} \quad (1.7)$$

pour la distribution inférieure δ .

De même que pour les p-boxes généralisées, la proposition suivante montre que les nuages sont représentables par des paires de distributions de possibilité

Proposition 1.4. *Un nuage $[\pi, \delta]$ est représentable par une paire de distributions de possibilité $1 - \delta$ et π , c'est-à-dire:*

$$\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_{\pi} \cap \mathcal{P}_{1 - \delta}$$

La proposition suivante formalise plus en avant le lien existant entre nuages et p-boxes généralisées:

Proposition 1.5. *Soit $[\pi, \delta]$ un nuage sur \mathcal{X} . Alors, les trois assertions suivantes sont équivalentes:*

- (i) *Le nuage $[\pi, \delta]$ peut être encodé comme une p-box généralisée $[\underline{F}, \overline{F}]$ telle que $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_{[\underline{F}, \overline{F}]}$*
- (ii) *δ et π sont comonotones ($\delta(x) < \delta(y) \Rightarrow \pi(x) \leq \pi(y)$)*
- (iii) *les ensembles $\{B_{\overline{\gamma}_i}, C_{\gamma_j} \mid i, j = 0, \dots, M\}$ définis par les équations (1.6) et (1.7) forment une séquence d'ensembles emboîtés (i.e. ils sont complètement (pré)-ordonnés par la relation d'inclusion).*

Cette proposition indique que les p-boxes généralisées constituent des cas particuliers de nuages, puisqu'elles sont équivalentes aux nuages pour lesquels les distributions δ et π sont comonotones. A partir de maintenant, nous appellerons de tels nuages *comonotones*. Ce résultat indique entre autres choses que les nuages comonotones sont des cas particuliers d'ensembles aléatoires, et induisent donc des probabilités inférieures ∞ -monotones.

Nous montrons dans ce travail qu'il n'en va pas de même pour la plupart des nuages non-comonotones, qui induisent en général des probabilités inférieures qui ne sont pas 2-monotones (sans pour autant que l'ensemble de probabilité induit soit vide). Ces résultats indiquent que, d'un point de vue purement pratique, les nuages non-comonotones apparaissent comme moins intéressants que leur contre-partie comonotone.

A l'instar des p-boxes généralisées, il n'y a pas de lien direct entre nuages et distributions imprécises de probabilité. Il est cependant possible de reprendre les résultats concernant les p-boxes généralisées (ces dernières étant des cas particuliers de nuages), et notamment la Proposition 1.3. Il est également possible de reprendre et d'étendre d'autres transformations proposées pour approcher extérieurement des distributions imprécises de probabilité par des distributions de possibilité [141].

Notons également que la plupart des résultats obtenus ici et reliant les p-boxes généralisées et les nuages à d'autres représentations d'incertitude s'étendent facilement au cas de représentations continues définies sur les réels. Les résultats obtenus concernant les représentations pratiques d'incertitude sont résumés par la figure 1.1.

Deux autres problèmes qui sont brièvement considérés dans le chapitre concernent d'une part l'utilisation des modèles hiérarchiques de second ordre (Section 3.5), et plus particulièrement le cas des variables aléatoires floues [32, 217], d'autre part la propagation des p-boxes généralisées à travers un modèle déterministe (Section 3.2.5), ce qui nous permet, entre autre, de mettre en évidence l'utilité potentielle des relations exhibées dans le chapitre.

1.3 Traitement de sources multiples d'informations (Chapitre 4)

En pratique, lorsque la valeur d'une variable ou d'un paramètre X est mal connue, il arrive souvent que plusieurs sources (e.g. experts, capteurs, modèles physiques différents) fournissent des informations concernant cette variable ou ce paramètre. Dans cette situation, deux problèmes différents mais corrélés sont (i) la construction d'une représentation synthétique et interprétable, plus facile à manipuler que des informations éparses et (ii) l'évaluation de la

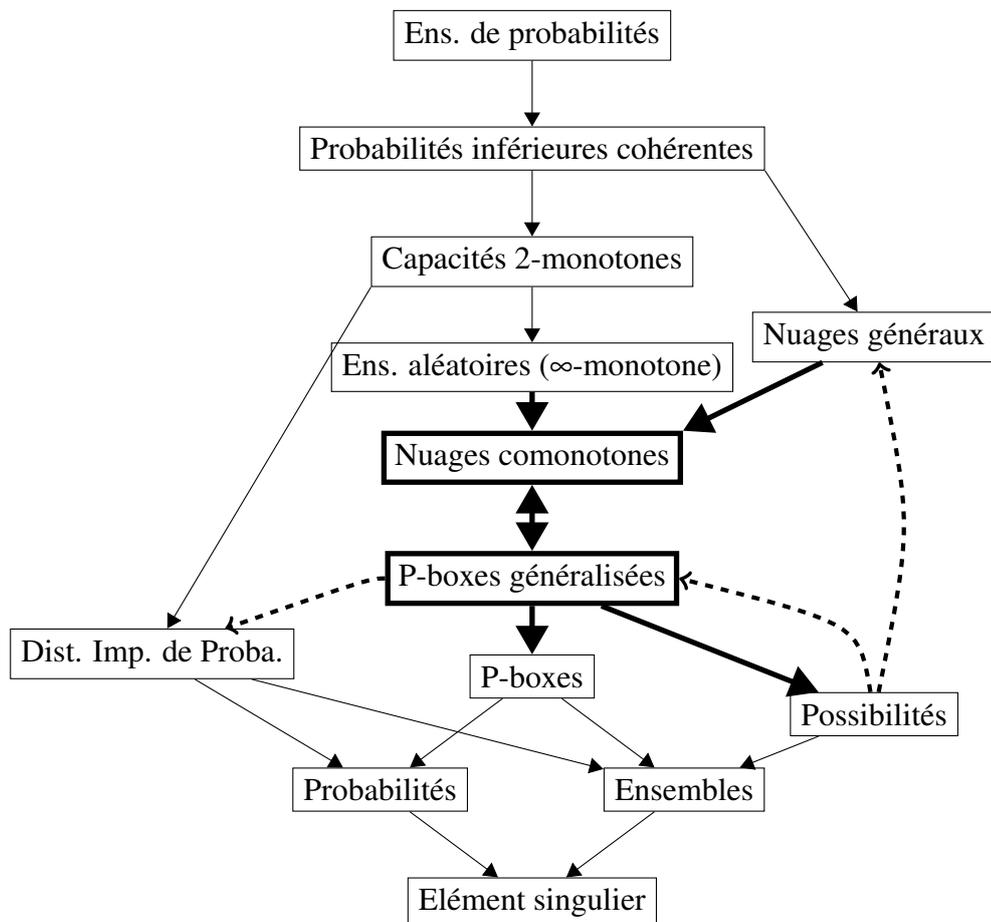


Figure 1.1: Relations entre représentations pratiques: résumé $A \longrightarrow B$: B est un cas particulier de A . $A \dashrightarrow B$: B est représentable par A

qualité de l'information fournie par les sources.

Dans ce chapitre, nous nous penchons sur chacun de ces deux problèmes. Concernant le premier, dénommé en général problème de fusion d'information, nous rappelons d'abord les méthodes de synthèse de base pour chacune des théories considérées ici, pour ensuite nous pencher plus spécifiquement sur les problèmes de traitement des inconsistances dans l'information et des dépendances pouvant exister entre les sources d'information. Pour résoudre le problème des inconsistances, nous proposons l'utilisation de la notion de sous-ensembles maximaux cohérents comme une solution générale et attractive à la fois d'un point de vue théorique et conceptuel. Concernant les dépendances, nous proposons l'utilisation d'une règle prudente basée sur la théorie des fonctions de croyances et du principe du moindre engagement.

1.3.1 Opérations de fusion d'information basiques

Soit \mathbb{F} un cadre de traitement des incertitudes (i.e., possibilités, ensembles aléatoires ou ensembles de probabilités), $\mathbb{F}_{\mathcal{X}}$ l'ensemble des représentations du cadre \mathbb{F} définie sur l'ensemble fini \mathcal{X} . L'information donnée par N sources étant modélisée par une représentation appartenant à \mathbb{F} , une opération de fusion φ est une fonction $\varphi : (\mathbb{F}_{\mathcal{X}})^N \rightarrow \mathbb{F}_{\mathcal{X}}$ qui résume les informations données par les sources en une représentation unique.

Supposons une notion d'inclusion, notée $\subset_{\mathbb{F}_{\mathcal{X}}}$, définie entre les éléments de $\mathbb{F}_{\mathcal{X}}$. Etant donné de l'information provenant de sources multiples et représentée par des modèles $a_i \in \mathbb{F}_{\mathcal{X}}$, $i = 1, \dots, N$, la fusion d'information peut suivre trois comportements principaux [18, 93]:

- **conjonctif**: un comportement conjonctif est le pendant d'une intersection d'ensembles. Le résultat $\varphi(a_1, \dots, a_N)$ d'une telle opération est tel que $\varphi(a_1, \dots, a_N) \subset_{\mathbb{F}_{\mathcal{X}}} a_i$ pour $i = 1, \dots, N$. Un opérateur conjonctif réduit donc l'incertitude globale, et fournit un résultat plus précis que chacune des sources prise séparément. Il suppose que toute les sources sont fiables, et peut fournir un résultat très peu fiable, voire vide, en cas d'inconsistances dans l'information fournie par les sources.
- **disjonctif**: un comportement disjonctif est le pendant d'une union d'ensembles. Le résultat $\varphi(a_1, \dots, a_N)$ d'une telle opération est tel que $\varphi(a_1, \dots, a_N) \supset_{\mathbb{F}_{\mathcal{X}}} a_i$ pour $i = 1, \dots, N$. Un opérateur disjonctif augmente donc l'incertitude globale, et fournit un résultat moins précis que chacune des sources prise séparément. Il fait la supposition qu'au moins une des sources est fiable. Le résultat d'une telle opération est générale-

ment très fiable, mais peut être très (trop) imprécis, ce qui réduit son utilité.

- **compromis**: le résultat d'un comportement de compromis se situe entre la disjonction et la conjonction. De tels comportements sont généralement utilisés quand les informations fournies par les sources sont partiellement inconsistantes. L'objectif d'un tel comportement est d'obtenir un résultat qui ait un bon équilibre entre informativité et fiabilité. Nous distinguons deux types de compromis:
 - **adaptatifs**: un comportement de compromis sera appelé adaptatif si le résultat dépend du contexte. Le but est de passer d'un comportement conjonctif à un comportement disjonctif au fur et à mesure que l'inconsistance entre les informations augmente. On retrouve alors la disjonction (conjonction) en cas d'inconsistance totale (consistance totale) entre les sources. Entre ces deux situations, le comportement est de compromis. Les méthodes utilisant les sous-ensembles maximaux cohérents, que nous considérons plus tard, en sont de bons représentants.
 - **non-adaptifs**: un comportement de compromis est *non-adaptatif* quand il se comporte toujours de la même manière, quelque soit le contexte. Les moyennes arithmétiques pondérées (ou combinaisons convexes) constituent un exemple typique et populaire de tels opérateurs, et sont de loin les opérateurs de fusion les plus utilisés en pratique.

Nous rappelons ensuite comment les opérateurs de base (conjonctions, disjonctions, combinaison convexe), ceux qui sont le plus souvent utilisés pour en construire de plus complexes, se déclinent dans chacune des théories considérées ici (probabilités imprécises, ensembles aléatoires, possibilités).

Probabilités imprécises soit N sources dont les informations concernant X sont modélisées par les ensembles de probabilités $\mathcal{P}_1, \dots, \mathcal{P}_N$. Les opérateurs principaux de fusion d'information se définissent alors comme suit [202]:

- **conjonction**: la conjonction $\mathcal{P}_{\cap(1:N)}$ des ensembles $\mathcal{P}_1, \dots, \mathcal{P}_N$ se définit comme

$$\mathcal{P}_{\cap(1:N)} = \bigcap_{i=1}^N \mathcal{P}_i$$

- disjonction: la disjonction $\mathcal{P}_{\cap(1:N)}$ des ensembles $\mathcal{P}_1, \dots, \mathcal{P}_N$ se définit comme

$$\mathcal{P}_{\cup(1:N)} = \bigcup_{i=1}^N \mathcal{P}_i$$

et, en pratique, c'est souvent l'enveloppe convexe $\mathcal{H}(\mathcal{P}_{\cup(1:N)})$ de $\mathcal{P}_{\cup(1:N)}$ qui est considérée, étant plus facile à manipuler et équivalente d'un point de vue comportemental [202].

- combinaison convexe: la combinaison convexe $\mathcal{P}_{\cap(1:N)}$ des ensembles $\mathcal{P}_1, \dots, \mathcal{P}_N$, auxquels sont associés les poids non-négatifs et de somme unitaire $\lambda_1, \dots, \lambda_N$ se définit comme

$$\mathcal{P}_{\Sigma(1:N)} = \sum_{i \in \llbracket N \rrbracket} \lambda_i \mathcal{P}_i$$

Ensembles aléatoires soit N sources dont les informations concernant X sont modélisées par les ensembles aléatoires $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$. A partir de ces N ensembles aléatoires, nous définissons une distribution de masse jointe $m_{(1:N)}$, définie sur le produit Cartésien \mathcal{X}^N comme une distribution ayant les distributions m_i , $i = 1, \dots, N$ pour marginales. c'est-à-dire, pour tout $j \in 1, \dots, N$ et $i_j \in \{1, \dots, |\wp(\mathcal{X})|\}$, $m_{(1:N)}$ est telle que:

$$m_{(1:N)}(\cdot \times E_{i_j}) := \sum_{\substack{i_1, \dots, i_{j-1} \in \{1, \dots, |\wp(\mathcal{X})|\} \\ i_{j+1}, \dots, i_N \in \{1, \dots, |\wp(\mathcal{X})|\}}} m_{(1:N)}(E_{i_1} \times \dots \times E_{i_j} \times \dots \times E_{i_N}) = m_j(E_{i_j}) \quad (1.8)$$

et $m_{(1:N)}$ ne reÁoit une masse positive que si $E_{i_j} \in \mathcal{F}_j$, l'ensemble des ensembles focaux de $(m, \mathcal{F})_j$, et ce pour $j = 1, \dots, N$. Les opérateurs principaux de fusion d'information se définissent alors comme suit:

- conjonction: une conjonction des ensembles aléatoires $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ se définit en deux étapes
 1. la construction d'une distribution jointe satisfaisant (1.8)
 2. allouer chaque masse jointe $m_{(1:N)}(\times_{j=1}^N E_j)$ à l'ensemble $\cap_{j=1}^N E_j$, avec $E_j \in \mathcal{F}_j$ for $j = 1, \dots, N$.

Notons qu'en prenant le produit des masses, on retrouve la règle bien connue de combinaison du Modèle des Croyances Transférables (i.e. la règle de combinaison de Dempster non-normalisée).

- disjonction: une disjonction des ensembles aléatoires $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ se définit en deux étapes
 1. la construction d'une distribution jointe satisfaisant (1.8)
 2. allouer chaque masse jointe $m_{(1:N)}(\times_{j=1}^N E_j)$ à l'ensemble $\cup_{j=1}^N E_j$, avec $E_j \in \mathcal{F}_j$ for $j = 1, \dots, N$.
- combinaison convexe: des ensembles aléatoires $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ se définit comme l'ensemble aléatoire $(m, \mathcal{F})_{\Sigma(1:N)}$ ayant une distribution de masse telle que, pour chaque ensemble $E \subseteq \mathcal{X}$:

$$m_{\Sigma(1:N)}(E) = \sum_{i=1}^N \lambda_i m_i(E). \quad (1.9)$$

Possibilités soit N sources dont les informations concernant X sont modélisées par les distributions π_1, \dots, π_N . Les opérateurs principaux de fusion d'information se définissent alors comme suit:

- conjonction: la conjonction $\pi_{\top(1:N)}$ des distributions π_1, \dots, π_N se définit pour tout $x \in \mathcal{X}$ comme

$$\pi_{\top(1:N)}(x) = \top_{i=1, \dots, N} \pi_i(x)$$

avec \top une norme triangulaire, couramment appelée t-norme³. Les plus souvent utilisées sont l'opération minimum, qui peut être associée à une hypothèse de dépendance entre sources (puisque c'est la seule t-norme idempotente), et le produit (hypothèse d'indépendance).

- disjonction: la disjonction $\pi_{\perp(1:N)}$ des distributions π_1, \dots, π_N se définit pour tout $x \in \mathcal{X}$ comme

$$\pi_{\perp(1:N)}(x) = \perp_{i=1, \dots, N} \pi_i(x)$$

avec \perp une conorme triangulaire ou t-conorme, qui sont les opérateurs duaux⁴ des t-normes. La t-conorme la plus souvent utilisée est le maximum (t-conorme la moins pénalisante).

- combinaison convexe: la combinaison convexe $\pi_{\Sigma(1:N)}$ des distributions π_1, \dots, π_N , auxquelles sont associés les poids non-négatifs et de somme unitaire $\lambda_1, \dots, \lambda_N$ pour tout $x \in \mathcal{X}$ se

³Une t-norme est une fonction $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ associative, commutative, non-décroissante en chaque membre et ayant 1 comme élément neutre

⁴Dans le sens où pour tout $(x, y) \in [0, 1]^2$, $\perp(x, y) = 1 - \top(1 - x, 1 - y)$

défini comme

$$\pi_{\Sigma(1:N)}(x) = \sum_{i=1}^N \lambda_i \pi_i(x)$$

Les propriétés et liens existants entre les opérateurs des différentes théories sont également étudiés. Pour chacun de ces opérateurs, nous proposons également des contreparties s'appliquant aux nuages de Neumaier, dont nous étudions les propriétés et les liens avec les opérations des autres théories.

Nuages soit N sources dont les informations concernant X sont modélisées par les nuages $[\pi, \delta]_1, \dots, [\pi, \delta]_N$. Alors, nous proposons les opérateurs principaux suivants:

- **conjonction:** nous définissons la conjonction $[\pi, \delta]_{\cap}$ des nuages $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ comme

$$[\pi, \delta]_{\cap(1:N)} = [\pi_{\cap(1:N)}, \delta_{\cap(1:N)}] = [\min_{i=1}^N(\pi_i), \max_{i=1}^N(\delta_i)] \quad (1.10)$$

.

- **disjonction:** nous définissons la disjonction $[\pi, \delta]_{\cup}$ des nuages $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ comme

$$[\pi, \delta]_{\cup(1:N)} = [\pi_{\cup(1:N)}, \delta_{\cup(1:N)}] = [\max_{i=1}^N(\pi_i), \min_{i=1}^N(\delta_i)] \quad (1.11)$$

.

- **combinaison convexe:** étant donné les poids de somme unitaire $\lambda_1, \dots, \lambda_N$ associés aux nuages $[\pi, \delta]_1, \dots, [\pi, \delta]_N$, nous définissons la combinaison convexe de nuages $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ comme

$$[\pi, \delta]_{\Sigma(1:N)} = [\pi_{\Sigma(1:N)}, \delta_{\Sigma(1:N)}] = \left[\sum_{i=1}^N \lambda_i \pi_i, \sum_{i=1}^N \lambda_i \delta_i \right] \quad (1.12)$$

Bien que ces opérateurs (conjonctions, disjonctions, combinaisons convexes) soient suffisants pour traiter les problèmes où les informations sont soit très inconsistantes, soit très consistantes entre elles, ils ne sont pas assez flexibles pour obtenir un modèle utile dans le cas où les informations sont partiellement inconsistantes. Dans ce dernier cas, nous proposons l'utilisation des sous-ensembles maximaux cohérents comme une réponse générale attractive à la fois d'un point de vue conceptuel et théorique.

1.3.2 Utilisation des sous-ensembles maximaux cohérents (SMC)

Les méthodes basées sur les SMC [173] consistent à utiliser une opération conjonctive au sein des sous-groupes de sources dont les informations sont consistantes, pour ensuite synthétiser les différents résultats par une opération de fusion disjonctive. Le résultat obtenu est donc proche d'une conjonction pure si les informations montrent une grande consistance, et proche d'une disjonction si les informations sont fortement inconsistantes. Cette approche permet de répondre de manière adaptative et simple au double objectif (parfois difficile à atteindre) qui consiste à gagner un maximum d'informativité tout en restant consistant avec l'information donnée par chaque source.

Nous revoyons d'abord comment les SMC peuvent s'appliquer de manière générale, avant d'étudier un cas plus précis s'appliquant à la théorie des possibilités.

1.3.2.1 Application générale

Probabilités imprécises L'information étant donnée par N ensembles de probabilités $\mathcal{P}_1, \dots, \mathcal{P}_N$, un sous ensemble⁵ $K \subset \llbracket N \rrbracket$ est maximal cohérent (dit SMC) si $\bigcap_{i \in K} \mathcal{P}_i \neq \emptyset$ et s'il est maximal avec cette propriété (tout ajout d'une source à K conduit à une conjonction vide). Le résultat de la fusion par SMC est alors:

$$\mathcal{P}_{MCS(1:N)} = \mathcal{H} \left(\bigcup_{j=1}^k \bigcap_{i \in K_j} \mathcal{P}_i \right) \quad (1.13)$$

avec k le nombre de SMC, K_j les SMC et \mathcal{H} l'enveloppe convexe de l'ensemble.

Ensembles aléatoires L'information étant donnée par N ensembles aléatoires $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, un ensemble résultant d'une fusion par SMC est construit en trois étapes

1. la construction d'une distribution jointe satisfaisant (1.8)
2. pour chaque masse jointe $m_{(1:N)}(\times_{j=1}^N E_j)$, $K \subset \llbracket N \rrbracket$ est un SMC si $\bigcap_{i \in K} E_i \neq \emptyset$ et si il est maximal avec cette propriété. Soit K_1, \dots, K_k les SMC pour cette masse jointe.
3. allouer la masse $m_{(1:N)}(\times_{i=1}^N E_i)$ à l'ensemble $\bigcup_{j=1}^k \bigcap_{i \in K_j} E_i$, avec $E_i \in \mathcal{F}_i$ pour $i = 1, \dots, N$.

⁵Rappelons que nous notons $\llbracket N \rrbracket := \{1, \dots, N\}$ l'ensemble des N premiers naturels

Possibilités l'information étant donnée par N distributions de possibilités π_1, \dots, π_N , un sous-ensemble $K \subset \llbracket N \rrbracket$ est un SMC si $\min_{i \in K} \pi_i \neq \emptyset$ et est maximum avec cette propriété. Dans ce cas, le résultat d'une fusion par SMC est

$$\pi_{MCS(1:N)} = \max_{j=1}^k \min_{i \in K_j} \pi_i. \quad (1.14)$$

Les règles de fusion ci-dessus ont l'avantage qu'elles permettent de traiter l'inconsistance de manière flexible, tout en requérant un minimum d'information (dans le cas des probabilités imprécises et distributions de possibilités, seuls les modèles représentant les informations fournies sont nécessaires). De plus, elles satisfont nombre de propriétés qui apparaissent comme désirables dans un processus de fusion d'information.

Si ces règles sont séduisantes d'un point de vue théorique et conceptuel, les appliquer peut poser des problèmes calculatoires importants, puisqu'extraire les sous-ensembles maximaux cohérents est, en général, un problème de complexité NP-complète [140]. Néanmoins, il existe des cadres où extraire les SMC ne présente pas une telle complexité, et c'est notamment le cas lorsque les ensembles sont des intervalles définis sur les réels. Pour cette raison, nous étudions une méthode SMC s'appliquant aux intervalles flous (distributions de possibilités dont les α -coupes sont des intervalles) et travaillant à α -coupes constantes. Cette méthode, qui étend l'équation (1.14) et est un cas particulier des SMC appliqués aux ensembles aléatoires, reste, globalement, de complexité linéaire par rapport au nombre de sources, et est donc facilement applicable.

1.3.2.2 SMC appliqué aux distributions de possibilité sur les réels.

Nous considérons donc le cas où X prend sa valeur sur l'espace des réels, et où l'information est modélisée par N distributions de possibilités π_i , $i = 1, \dots, N$. Cela peut être, par exemple, des experts donnant leurs avis en terme d'intervalles de confiance emboîtés.

Nous proposons donc d'utiliser la méthode MCS α -coupe par α -coupe (inspirée de travaux précédents [77]). Etant donné les distributions π_1, \dots, π_N , à chaque niveau $\alpha \in [0, 1]$ correspond une série de N intervalles $E_{i,\alpha}$, avec $E_{i,\alpha}$ l' α -coupe de la distribution π_i . Soit $K_{j,\alpha}$ les sous-ensembles maximaux d' α -coupes tels que $\bigcap_{i \in K_{j,\alpha}} E_{i,\alpha} \neq \emptyset$ (Pour chaque niveau, l'extraction de ces sous-intervalles est de complexité globalement linéaire). Nous définissons ensuite l'ensemble $E_{MCS,\alpha}$ comme l'union des intersections associées aux sous-ensembles

$K_{j,\alpha}$, c'est-à-dire

$$E_{MCS,\alpha} = \bigcup_{j=1,\dots,f(\alpha)} \bigcap_{i \in K_{j,\alpha}} E_{i,\alpha} \quad (1.15)$$

avec $f(\alpha)$ le nombre de sous-ensembles maximaux cohérents distincts au niveau α . Notons qu'en général, $E_{MCS,\alpha}$ est une union d'intervalles disjoints, et ne constituent pas une distribution de possibilité, puisque la relation $E_{MCS,\alpha} \supset E_{MCS,\beta}$ n'est pas vraie pour toute paire de valeur $\beta, \alpha \in [0, 1]$ telles que $\beta > \alpha$, et les ensembles ne sont donc pas emboîtés.

Néanmoins, en pratique, il y aura un nombre fini de $p + 1$ valeurs $0 = \beta_1 \leq \dots \leq \beta_p \leq \beta_{p+1} = 1$ telles que les ensembles $E_{MCS,\alpha}$ seront emboîtés pour toute valeur $\alpha \in (\beta_k, \beta_{k+1}]$, et cela pour $k = 1, \dots, p$. Etant donné les distributions π_1, \dots, π_N , ces valeurs sont assez faciles à extraire.

En appliquant l'équation (1.15) aux niveaux $\alpha \in (\beta_k, \beta_{k+1}]$, nous obtenons donc un ensemble flou F_k non-normalisé, dont le degré d'appartenance varie dans $(\beta_k, \beta_{k+1}]$ (les ensembles $E_{MCS,\alpha}$ étant emboîtés entre ces valeurs). Il est ensuite possible de renormaliser cet ensemble flou de manière proportionnelle pour obtenir un ensemble flou \tilde{F}_k et de lui affecter une masse $m(\tilde{F}_k) = \beta_{k+1} - \beta_k$. Le résultat est alors formellement équivalent à un ensemble flou aléatoire (ou encore à une fonction de croyance floue), c'est-à-dire un ensemble aléatoire où les éléments focaux sont des ensembles flous. Le procédé est illustré par la figure 1.2.

Nous proposons ensuite une série d'outils pratiques permettant de manipuler plus facilement l'information résumée contenue dans cette variable aléatoire floue. Nous proposons également des variantes permettant de prendre en compte, dans la procédure de fusion proposée ci-dessus, des informations supplémentaires (e.g. nombre de sources fiables, facteurs de fiabilité des sources, distance entre les informations). Entre-autres choses, nous proposons de prendre la distribution de possibilité π_c correspondant à la fonction de contour, c'est-à-dire,

$$\forall x \in \mathcal{X}, \quad \pi_c(x) = \sum_{i=1}^p m_i \nu_{\tilde{F}_i}(x), \quad (1.16)$$

comme résumé essentiel de l'information. Cette distribution est ensuite plus facile à manipuler, et d'un point de vue axiomatique satisfait la plupart des propriétés requises par d'autres auteurs.

Dans la suite du chapitre concernant la fusion d'information, nous étudions également de plus près la prise en compte de dépendances entre sources au sein de la théorie des ensembles aléatoires dans la fusion conjonctive (Section 4.3). Cette théorie semble particulièrement bien

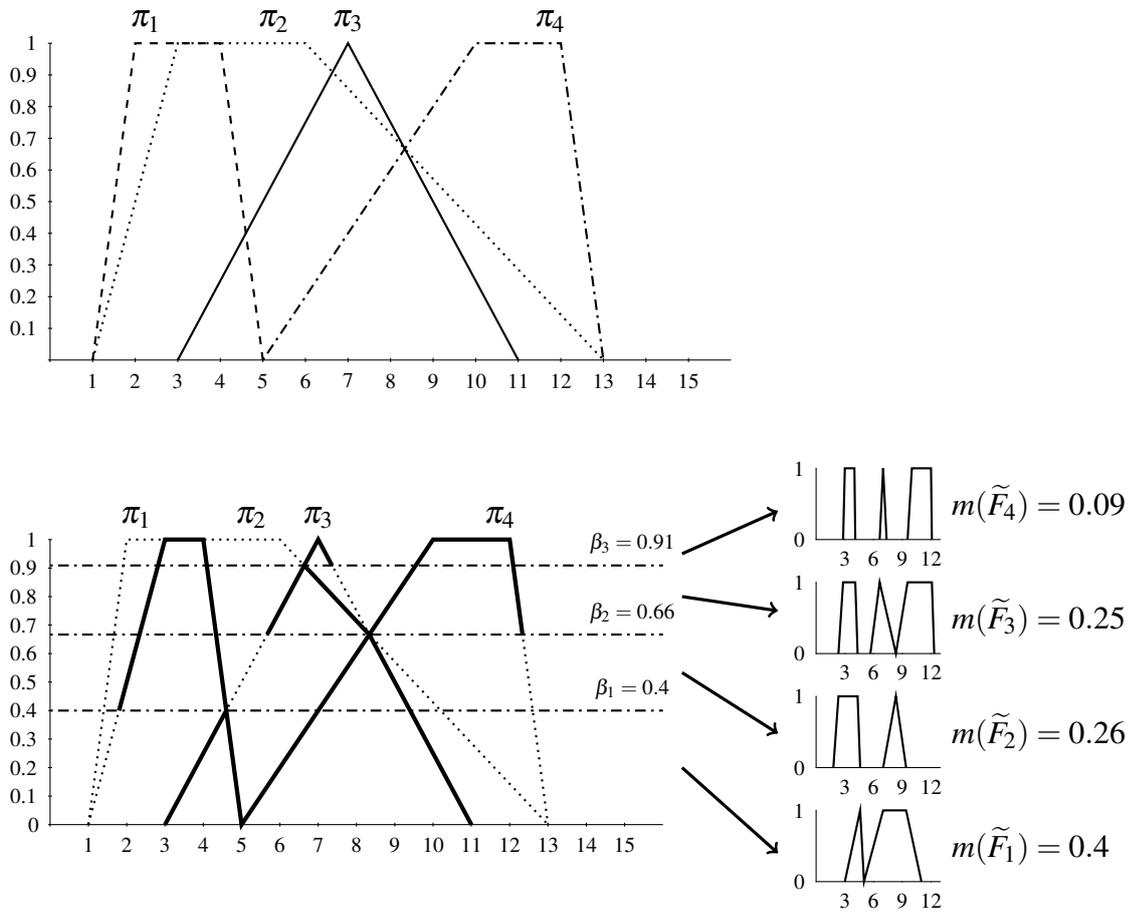


Figure 1.2: Sous-ensembles maximaux cohérents: illustration

adaptée pour prendre en compte de telles dépendances, qui peuvent s'exprimer par le biais de la construction de la distribution de masse jointe (première étape des différents opérateurs de fusion définis plus haut pour les ensembles aléatoires). En particulier, nous donnons quelques premiers résultats permettant d'utiliser le principe de moindre engagement pour fusionner prudemment plusieurs sources d'informations. Un des intérêts théorique (et pratique) de ces résultats est qu'ils indiquent que lorsque les ensembles aléatoires sont équivalents à des distributions de possibilités (i.e. éléments focaux emboîtés), nous retrouvons la fusion par t-norme minimum, c'est-à-dire le mode de fusion conjonctif le plus prudent en théorie des possibilités.

1.3.2.3 Evaluation de l'information

Comme nous l'avons vu plus haut, certains opérateurs de fusion (e.g. la moyenne arithmétique pondérée) requièrent d'affecter des poids aux différentes sources d'informations, ce qui revient à leur affecter des importances, ou une fiabilité différente à chacune. Déterminer ces poids de manière rationnelle n'est pas toujours chose aisée. De plus, dans des activités scientifiques telles que l'analyse de risque ou les études de sûreté, il est important que ces poids, s'ils doivent être déterminés, le soient de la manière la plus objective possible.

C'est pourquoi nous proposons une méthode générale (Section 4.4), inspirée de travaux précédents [28, 174] et applicable à l'ensemble des théories explorées ici, permettant d'évaluer la qualité de l'information délivrée par les différentes sources, en se basant sur des performances passées. Afin d'explicitier notre approche sur un exemple dégénéré, nous considérerons une source ayant donné pour information concernant la valeur de la variable X un sous-ensemble $A \subseteq \mathcal{X}$.

La méthode consiste à affecter à chaque source un score, basé sur un ensemble de variables témoins⁶ et sur deux critères qui sont:

- **Informativité:** mesure la précision de l'information fournie par une source pour une variable témoin donnée. Plus précise a été la source, plus élevée est son informativité. Selon le cadre de travail, nous proposons une mesure d'informativité qui généralise la valeur suivante:

$$Inf = 1 - \frac{|A|}{|\mathcal{X}|}$$

valeur qui reflète bien la précision de l'intervalle A par rapport à l'ignorance (l'ensemble \mathcal{X})

⁶une variable témoin est une variable pour laquelle la source a précédemment donné des informations imprécises, et dont la valeur réelle a été observée ultérieurement

- **Calibration:** mesure la cohérence entre l'information fournie par la source et les valeurs observées pour une variable témoin. En gardant à l'esprit que notre source a considéré que la valeur se trouvait dans l'ensemble A , nous différencions deux cas: le cas où la valeur précise prise par X est observée, et celui où l'observation est elle-même entourée d'incertitudes. Dans le cas où la valeur précise x^* est observée, alors la calibration est simplement la mesure de confiance supérieure que la source accordait à cette valeur (dans le cas d'un intervalle, 1 si $x^* \in A$, 0 sinon). Dans le cas où l'observation permet simplement de conclure $X \in B$, nous considérons la mesure de calibration

$$Cal = \frac{|A \cap B|}{|B|}$$

et nous proposons aux différents cadre considérés dans ce travail. Cette mesure d'inclusion de B dans A permet en effet de mesurer combien la source jugeait B plausible.

Une source de bonne qualité est donc une source recevant un haut score, c'est-à-dire une source informative et bien calibrée. Les formules proposées dans le manuscrit répondent à un certain nombre de critère rationnels:

1. les mesures doivent récompenser les sources à la fois informatives et bien calibrées,
2. la calibration ne devrait être influencée que par nos observations ou nos connaissances concernant les variables témoins,
3. les mesures devraient être comparables entre-elles, quelque soit la nature et le nombre de variables témoins (en d'autres termes, elles devraient avoir une métrique commune).

1.4 Incertitudes et (In)dépendance (Chapitre 5)

Nous nous intéressons ensuite aux notions d'indépendance entre variables. En effet, si la définition formelle d'indépendance stochastique en probabilité classique fait l'unanimité une fois que l'on s'affranchit du problème de l'interprétation, ce n'est plus le cas lorsque l'on considère des modèles imprécis. Dans ce dernier cas, à différentes interprétations correspondent différentes définitions formelles d'indépendance, et il est donc nécessaire de les étudier de plus près pour bien les maîtriser, et savoir quand chacune peut s'appliquer.

Nous proposons donc une première taxonomie qui nous permet de classer les types d'indépendance selon la nature des relations qu'elles décrivent, basée sur différents travaux [203, Ch.9], [144], [216], [6].

Cette taxonomie distingue parmi les notions d'indépendance celles qui sont:

- *Non-Informative (nInf.)* ou *Informative (Inf.)*: par non-informative, nous entendons ces notions d'indépendance qui traduisent l'absence de connaissance sur les relations entre variables, et qui suppose donc que toute relation est possible. Par informative, nous entendons les notions qui traduisent la connaissance d'une absence de relation, et qui permettent de modéliser cette absence.
- *Subjective (Sub.)* ou *Objective (Obj.)*: une notion d'indépendance est dite subjective si elle concerne nos croyances à propos des valeurs que peuvent prendre des variables, et objective si elle cherche à décrire des propriétés intrinsèques au phénomène observé
- *Symétrique (Sym)* or *Asymétrique (Asym.)*: Une notion est symétrique si dire que X_1 est indépendant de X_2 implique automatiquement que X_2 est indépendant de X_1 . Ce type d'indépendance rend souvent plus facile la construction de modèles d'incertitude joints à partir de modèles locaux ou marginaux, et permet donc de stocker l'information sous cette forme. Cette propriété est souvent appelée factorisation. Ces notions symétriques ne permettent pas de modéliser des notions d'indépendance qui concerne des relations asymétriques.

Il faut alors utiliser des notions d'indépendances asymétriques, certes moins pratiques mais qui permettent d'exprimer de telles notions. Une notion d'indépendance est asymétrique si dire que X_1 est indépendant de X_2 n'est pas équivalent à dire que X_2 est indépendant de X_1 . Ces relations asymétriques sont souvent de types évidentiels ou causals

- les relations d'indépendance causales expriment l'idée que deux variables ne sont pas causalement reliées. Par exemple, elles permettent d'exprimer qu'une habitude de vie n'est pas la (une) cause d'une maladie.
- les relations d'indépendance évidentielles expriment l'idée qu'apprendre la valeur d'une variable ne va pas changer nos croyances à propos d'une autre variable, ce qui n'implique pas qu'apprendre la valeur de cette dernière ne changera pas nos croyances à propos de la première.

Nous rappelons et classons ensuite les principales notions d'indépendances rencontrées dans les théories de l'incertain, c'est-à-dire les notions d'interactions inconnues; d'indépendance forte; d'indépendance de répétition; de non-pertinence et d'indépendance épistémique; d'indépendance d'ensembles aléatoires; de non-interaction possibilistes. Pour simplifier les définitions, nous considérons des relations entre deux variables X_1 et X_2 prenant leurs valeurs dans $\mathcal{X}_1, \mathcal{X}_2$.

Nous notons $\mathcal{X}_{(1:2)} = \mathcal{X}_1 \times \mathcal{X}_2$ le produit cartésien des deux espaces, et $X_{(1:2)}$ une variable prenant ses valeurs sur $\mathcal{X}_{(1:2)}$.

Definition 1.7. Soit deux ensembles marginaux de probabilités $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ représentant notre incertitude sur les variables X_1, X_2 prenant leurs valeurs dans $\mathcal{X}_1, \mathcal{X}_2$. La notion d'*interaction inconnue* entre X_1, X_2 est la donnée de l'ensemble de probabilités jointes $\mathcal{P}_{UI, X_{(1:2)}}$ tel que

$$\mathcal{P}_{UI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2} \in \mathcal{P}_{X_2}\}$$

avec P_{X_1}, P_{X_2} les probabilités marginales de $P_{X_{(1:2)}}$ sur les domaines $\mathcal{X}_1, \mathcal{X}_2$ et $\mathbb{P}_{\mathcal{X}}$ l'ensemble des probabilités sur \mathcal{X} .

La notion d'interaction exprime l'idée que nos informations ne nous permettent pas de connaître les relations qui peuvent lier X_1 et X_2 , et revient donc à toutes les considérer.

Definition 1.8. Soit deux ensembles marginaux de probabilités $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ représentant notre incertitude sur les variables X_1, X_2 prenant leurs valeurs dans $\mathcal{X}_1, \mathcal{X}_2$. La notion d'*indépendance forte* [33] entre X_1, X_2 est la donnée de l'ensemble de probabilités jointes $\mathcal{P}_{SI, X_{(1:2)}}$ tel que

$$\mathcal{P}_{SI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid P_{X_{(1:2)}} = P_{X_1} \otimes P_{X_2}, P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2} \in \mathcal{P}_{X_2}\}$$

avec \otimes le produit de mesure usuel.

La notion d'indépendance forte exprime l'idée que X_1 et X_2 prennent leurs valeurs suivant deux processus aléatoires stochastiquement indépendants.

Definition 1.9. Soit deux ensembles marginaux $\mathcal{P}_{X_1} = \mathcal{P}_{X_2} = \mathcal{P}_X$ identiques représentant notre incertitude sur deux variables X_1, X_2 prenant leurs valeurs dans $\mathcal{X}_1, \mathcal{X}_2$, avec $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$. La notion d'*indépendance de répétition* [33] entre X_1, X_2 est la donnée de l'ensemble de probabilités jointes $\mathcal{P}_{RI, X_{(1:2)}}$ tel que

$$\mathcal{P}_{RI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid P_{X_{(1:2)}} = P_X \otimes P_X, P_X \in \mathcal{P}_X\}$$

avec \otimes le produit de mesure usuel.

La notion d'indépendance de répétition revient à considérer que X_1 et X_2 suivent deux processus aléatoires identiques et indépendants. Cette notion correspond à la notion statistique usuelle d'échantillons indépendants et identiquement distribués.

Definition 1.10. Soit deux ensembles marginaux $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ représentant notre incertitude sur X_1, X_2 prenant leur valeur dans $\mathcal{X}_1, \mathcal{X}_2$. Alors, la notion de *non-pertinence épistémique* [33] de X_1 envers X_2 est la donnée de l'ensemble de probabilités jointes $\mathcal{P}_{EIrr_{1 \rightarrow 2}, X_{(1:2)}}$ tel que

$$\mathcal{P}_{EIrr_{1 \rightarrow 2}, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid \forall x_{(1:2)} \in \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(x_{(1:2)}) = P_{X_1}(x_1)P_{X_2}(x_2|x_1), \\ P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2}(\cdot|x_1) \in \mathcal{P}_{X_2}\}$$

avec $P_{X_2}(\cdot|x_1)$ les probabilités conditionnelles potentielles sur \mathcal{X}_2 étant donné x_1 .

La notion de non-pertinence épistémique exprime l'idée qu'apprendre la valeur d'une variable (ici X_1) ne changera pas nos croyances actuelles quand à la valeur possible de X_2 . C'est une notion asymétrique, et la notion symétrique correspondante (indépendance épistémique) est obtenue en utilisant la notion de non-pertinence dans les deux sens (i.e. X_1 non-pertinent envers X_2 , et inversement).

Definition 1.11. Soit $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ deux ensembles aléatoires représentant l'incertitude sur les variables X_1, X_2 prenant leur valeur dans $\mathcal{X}_1, \mathcal{X}_2$, et $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ les ensembles de probabilités induits. La notion d'*indépendance d'ensembles aléatoires* [57, 207] entre X_1, X_2 est la donnée de l'ensemble de probabilités jointes

$$\mathcal{P}_{RSI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid \forall A \subseteq \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(A) \leq \sum_{\substack{(E_{X_1} \times E_{X_2}) \cap A \neq \emptyset \\ E_{X_i} \in \mathcal{F}_{X_i}}} m_{X_1}(E_{X_1})m_{X_2}(E_{X_2})\}$$

avec $m_{X_i}(E_{X_i})$ la masse donnée à l'ensemble focal E_{X_i} dans $(m, \mathcal{F})_{X_i}$

Cette notion traduit l'indépendance entre les distributions de masses dans la théorie des ensembles aléatoires. Elle est plus difficile à interpréter au sein de la théorie des probabilités imprécises, mais dans ce dernier cas, elle peut servir d'approximation à d'autres notions d'indépendance, plus difficiles à manipuler. Une autre notion d'indépendance relative aux ensembles aléatoire, dénommée indépendance cognitive, est également brièvement décrite dans le chapitre, mais moins discutée, du fait de l'existence de peu de résultats la concernant.

Definition 1.12. Soit deux distributions marginales π_{X_1}, π_{X_2} représentant l'incertitude sur les variables X_1, X_2 prenant leur valeur dans $\mathcal{X}_1, \mathcal{X}_2$. Alors, la notion de *non-interaction possibiliste* [218] entre X_1, X_2 est la donnée de la distribution jointe $\pi_{PI, X_{(1:2)}}$ telle que, pour tout $x_{(1:2)} \in \mathcal{X}_{(1:2)}$

$$\pi_{PI, X_{(1:2)}}(x_{(1:2)}) = \min(\pi_{X_1}(x_1), \pi_{X_2}(x_2))$$

à laquelle peut ensuite être associée l'ensemble de probabilités jointes $\mathcal{P}_{PI, X_{(1:2)}}$.

| Notion | Inf./nInf. | Obj./Sub. | Sym/Asym. | Expressible in | | |
|------------------------|------------|-----------|-----------|----------------|----|---|
| | | | | IP | RS | P |
| Int. inconnue | nInf. | Sub. | Sym | ✓ | ✓ | × |
| Non-int. possibiliste | Inf./nInf. | Sub. | Sym | ✓ | ✓ | ✓ |
| Ind. cognitive | Inf. | Sub. | Asym. | ? | ✓ | × |
| Ind. d'ens. aléatoires | Inf. | Sub. | Sym | ✓ | ✓ | × |
| Non-pert. épistémique | Inf. | Sub. | Asym. | ✓ | ? | × |
| Ind. de Kuznetsov | Inf. | Sub. | Sym | ✓ | ? | × |
| Ind. forte | Inf. | Obj. | Sym | ✓ | ✓ | × |
| Ind. de répétition | Inf. | Obj. | Sym | ✓ | ? | × |

Table 1.1: Notions d'indépendance et de non-pertinence dans l'incertain: résumé (? : question à résoudre)

La non-interaction possibiliste peut traduire deux notions. D'une part, elle peut s'interpréter comme une hypothèse de dépendance totale entre niveaux de confiance, et d'autre part, on peut la voir comme la réduction de la notion d'interaction inconnue à un cadre purement possibiliste (la possibilité jointe correspondant à la non-interaction possibiliste est alors vue comme la "trace" partielle d'un jugement d'interaction inconnue).

Le tableau 1.1 résume comment les différentes notions évoquées plus haut se situent par rapport à notre classification. La notion d'indépendance développée par Kuznetsov [134, 36] et qui se base sur les bornes des espérances mathématiques y figure également, ainsi que l'indépendance cognitive [178, Ch.7.5], afin que le tableau soit complet. La figure 1.3, quant à elle, montre les relations d'inclusions qui existent entre les modèles joints résultant des différentes notions. A nouveau, nous utilisons le langage des ensembles de probabilités pour faciliter la comparaison.

Dans la suite, nous nous intéressons au problème d'interpréter la notion de non-pertinence épistémique par le biais d'arbres d'événements et de la théorie probabiliste développée par Shafer [179] autour de ces derniers. Comme l'ont montré des recherches récentes [47, 48], il existe en effet de forts liens entre ce cadre et la théorie des probabilités imprécises. Nous montrons donc que la notion de non-pertinence entre de multiples variables est équivalente à la notion d'indépendance dans des arbres d'événements particuliers, que nous appelons arbres standards (Section 5.2).

$$\begin{array}{ccc}
 \mathcal{P}_{SI,X(1:2)} \subseteq \mathcal{P}_{KI,X(1:2)} \subseteq \mathcal{P}_{EInd,X(1:2)} \subseteq \left\{ \begin{array}{l} \mathcal{P}_{EIrr,X_1 \rightarrow 2} \\ \mathcal{P}_{EIrr,X_2 \rightarrow 1} \end{array} \right\} \subseteq \mathcal{P}_{RSI,X(1:2)} \subseteq \mathcal{P}_{CI,X(1:2)} \subseteq \mathcal{P}_{UI,X(1:2)} \\
 \cup \qquad \cup \\
 \mathcal{P}_{RI,X(1:2)} \qquad \mathcal{P}_{PI,X(1:2)}
 \end{array}$$

Figure 1.3: Relations d'inclusion des modèles joints à partir de modèles marginaux $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$

Nous explorons ensuite comment certaines notions d'indépendances, plus facile à manipuler et calculatoirement plus avantageuses, peuvent en approcher d'autres. En particulier, nous étendons des résultats précédemment obtenus pour le cas bi-dimensionnel [89] à un cadre général (n dimensions, n étant un nombre quelconque), et permettant d'approcher de manière conservative l'indépendance entre ensembles aléatoires (dont la structure jointe présente une complexité croissant exponentiellement avec le nombre de dimensions) par la notion de non-interaction possibiliste (dont la complexité de structure jointe ne croit pas avec le nombre de dimensions). Nous discutons ensuite de l'utilité d'une telle approche dans un cadre pratique (Section 5.3).

1.5 Prise de décision dans l'incertain (Chapitre 6)

Dans ce chapitre, nous nous intéressons brièvement au problème de prise de décision dans l'incertain. Même si ce problème ne concerne pas à proprement parlé le traitement de l'incertitude, il est difficile de le dissocier totalement de tels traitements, puisque ces derniers sont (presque) toujours utilisés en vue de prendre une décision. En particulier, les études de risques et de sûreté sont souvent associées à des décisions recouvrant à la fois un aspect économique et humain.

La prise de décision consiste (du moins en traitement de l'incertitude) à choisir, parmi un ensemble fini \mathcal{A} de choix possibles, les actions optimales (du point de vue du décideur). Dans ce travail, nous considérons un cadre relativement restreint de prise de décision, puisque nous considérons:

- i qu'à chaque choix $a \in \mathcal{A}$ peut être associé un gain (utilité) de valeur réelle et précise, noté $u_a : \mathcal{X} \rightarrow \mathbb{R}$, et que $u_a(x)$ représente l'intérêt de choisir l'action a quand X prend la valeur

$x \in \mathcal{X}$.

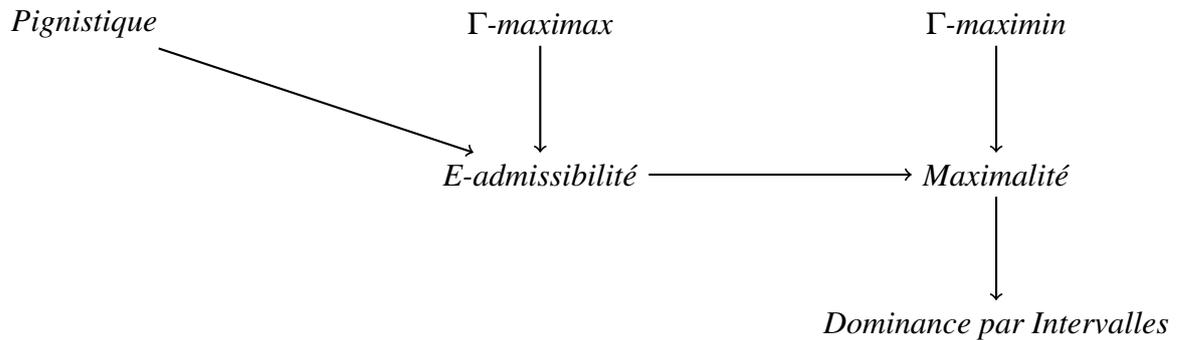
- ii que les choix ne peuvent pas être combinés entre-eux (i.e., pas de mélanges convexes de choix)
- iii que nous sommes dans un environnement statique. Nous ne considérons donc pas le problème de déterminer des séquences optimales de choix
- iv qu'un choix $a \in \mathcal{A}$ ne modifie pas l'incertitude sur X , i.e., nous supposons l'indépendance entre choix et état de la variable.

Lorsque l'incertitude sur X est modélisée par une probabilité précise, le choix optimal est souvent celui qui maximise l'espérance mathématique de l'utilité. Pourvu que l'incertitude soit représentable fidèlement par une distribution de probabilité, ce critère semble être un bon choix et a été justifié théoriquement par de nombreux auteurs. Dit autrement, il consiste à établir un (pré-)ordre complet entre les différents choix, construit à partir des espérances mathématiques.

Cependant l'information disponible ne permet pas toujours de représenter l'incertitude par une probabilité unique, et dans ce dernier cas, comme pour l'indépendance, il existe de nombreux moyens d'étendre le critère de maximisation de l'espérance mathématique au cadre des probabilités imprécises [195], ensembles aléatoires ou distributions de possibilité. Ces extensions (critères) suivent globalement deux principes [198] : soit elles cherchent à toujours établir un ordre complet entre les actions, en travaillant sur des probabilités particulières des ensembles ou sur les bornes d'espérances, soit elles relâchent la condition de complétude dans l'ordre induit, et autorise à déterminer un ensemble d'actions optimales non-comparables, plutôt qu'une seule.

Parmi les critères qui suivent la première voie se trouve le Γ -maximin [114], le Γ -maximax, le critère d'Hurwicz [122], la probabilité pignistique⁷ $BetP$ [187]. Parmi les critères suivant la seconde voie se trouvent la Maximalité, la dominance par Intervalles et l'E-admissibilité. La figure ci-dessous montre les implications qui existe entre ces critères ($A \rightarrow B$ indiquant qu'une action optimale au sens de A le sera aussi au sens de B).

⁷Néologisme dérivé du mot latin *pignus*, signifiant décision



Ces critères requièrent souvent de calculer des bornes d’espérances mathématiques sur des ensembles de probabilités. Si cela s’avère relativement aisé lorsque \mathcal{X} est un ensemble fini [198], il en va tout autrement lorsqu’il est continu. Nous étudions donc ensuite le cas particulier où \mathcal{X} est un sous-ensemble des réels, où l’incertitude est modélisée par une p-box $[E, \bar{F}]$ et où u_a est une fonction continue sur \mathcal{X} . Nous proposons des premiers résultats permettant d’obtenir les formules analytiques des fonctions cumulées qui vont permettre d’atteindre les bornes d’espérances.

1.6 Applications illustratives (Chapitre 7)

Nous donnons ensuite quelques détails relatifs à deux applications réalisées durant ce travail de thèse, au moyen du logiciel SUNSET développé à l’IRSN.

1.6.1 Evaluation et synthèse d’informations appliquées à des codes de calculs nucléaires

La première application concerne l’application des méthodes développées dans le chapitre 4 aux résultats d’analyses d’incertitude effectuées avec des codes de calculs nucléaires différents au cours d’un programme OCDE appelé BEMUSE [160]. Ce programme, regroupant 10 participants, avait pour but de comparer les méthodologies d’analyse d’incertitude en les appliquant à un cas d’accident de perte de réfrigérant du circuit primaire d’une centrale nucléaire par grosse brèche.

Ce type d’accident, qui provoque une diminution du flux de réfrigérant, est en effet critique. Une centrale nucléaire produisant de l’énergie et de la chaleur interne, c’est le rôle du système réfrigérant que de garder la température à un niveau acceptable. Quand une grosse

brèche dans ce circuit survient, le système est immédiatement arrêté d'urgence, cependant il est nécessaire, même après cet arrêt, de s'assurer que la température ne dépasse pas un niveau critique pouvant engendrer une catastrophe (e.g. fuite de radio-éléments). En effet, d'importantes quantités de chaleur peuvent encore être émises après arrêt du réacteur, du fait des réactions résiduelles.

Lors du projet BEMUSE, 10 participants ont donc appliqué leurs méthodes d'analyses d'incertitude sur un cas expérimental réalisé sur une installation de taille réduite. Chacun d'entre eux a eu à déterminer ses incertitudes sources, ses variables d'entrée et a pu utiliser des codes de calculs différents. Bien que les résultats de l'ensemble des participants aient tous été assez proches des valeurs expérimentales observées, ils présentaient néanmoins quelques différences, et étaient assez difficiles à comparer à première vue. Il nous a donc semblé utile d'appliquer aux résultats du programme les méthodes développées au chapitre 4. Nous nous sommes restreints aux théories des probabilités classiques et des possibilités, ce qui était suffisant pour mettre en exergue les points de convergence et de divergence entre les approches basées sur les probabilités classiques et celles utilisant d'autres théories de l'incertain, ainsi que les avantages présentés par chacune.

1.6.1.1 Evaluation des sources

Puisque les valeurs expérimentales d'un certain nombre de variables étaient disponibles, nous avons pu réaliser une évaluation de la qualité des informations obtenues par les différentes études d'incertitude.

D'un point de vue méthodologique, les résultats nous ont permis de montrer que les deux approches utilisées (probabiliste et possibiliste) conduisaient à des résultats similaires, du fait qu'elles étaient bâties sur les mêmes concepts, mais que des différences étaient cependant observées, ces différences étant dues aux différences de formalismes. En particulier, puisque peu de variables témoins étaient utilisées (4), nous avons pu mettre en exergue les problèmes rencontrés par la méthode probabiliste dans de tel cas (i.e. pouvoir discriminant diminué). Rappelons qu'en pratique, il est recommandé d'utiliser un minimum de 10 variables témoins pour utiliser l'approche probabiliste.

D'un point de vue plus pratique, les observations suivantes ont pu être faites:

- **rang par rapport au code utilisé** : lors de l'évaluation, il a pu être observé que le rang (la qualité) donnée à un participant était peu dépendant du code de calcul qu'avait utilisé par ce participant. Ce résultat montre que l'influence de l'utilisateur sur les résultats

produits par le code peut être très grande, et souligne l'importance de posséder une bonne expérience d'utilisation et de bonnes connaissances des processus en jeux.

- **validation/évaluation du code** : un aspect important encore matière à débat dans la modélisation de phénomènes physiques complexes est la manière de valider un code de calcul (i.e. de certifier que ses résultats sont conformes à l'expérience). Les résultats des méthodes d'évaluations étant significatifs par eux même, ils pourraient être utilisés dans les procédures de validation de codes.
- **validation des observations informelles** : les méthodes utilisées ont également permis de donner une base solide et rationnelle à des observations jusqu'ici faites de manière informelle. Elles sont donc également un moyen d'appuyer des conclusions et de les conforter.

1.6.1.2 Synthèse de l'information apportée par les sources

La synthèse de l'information nous a permis de mettre en évidence deux avantages des méthodes utilisées:

1. **Utilité des évaluations** : à la fois pour l'approche probabiliste et possibiliste, l'utilisation des poids déterminés par l'étape précédente (section 1.6.1.1) nous a permis d'améliorer le résultat de la synthèse d'information (à la fois en informativité et calibration), soit en les utilisant directement dans une combinaison convexe, soit en se restreignant aux informations fournies par des sous-groupes de "meilleures" sources.
2. **Quantification du conflit** : l'utilisation des méthodes de fusion possibiliste, et notamment de l'opération conjonctive, nous ont permis de fournir une valeur quantifiée et visuelle du conflit pouvant exister entre certains groupes de sources (utilisateurs du même code, ...), quantification jusqu'ici peu (ou pas) réalisée.

Les informations fournies par les sources étant assez consistantes entre-elles, nous n'avons pas cru bon appliquer les méthodes SMC aux données du cas BEMUSE.

1.6.2 Application de la méthode RaFu à un cas d'étude

La méthode de propagation hybride [9, 12] propose de différencier, parmi N variables X_1, \dots, X_N entourées d'incertitudes les variables dont la valeur est gouvernée par un aléa intrinsèque (in-

certitude aléatoire) des variables dont la valeur est fixe, mais mal connue du fait de manque d'informations (incertitude épistémique). Dans la méthode de propagation hybride, les premières sont modélisées par des probabilités et propagées au moyen de méthodes de simulation classiques (e.g. Monte-Carlo), tandis que les secondes sont modélisées par des distributions de possibilités et propagées au moyen du principe d'extension, qui suppose la non-interaction possibiliste entre les variables. Le résultat de cette propagation est une variable aléatoire floue qui est ensuite post-traitée en fonction du résultat voulu.

Deux des problèmes d'ordre pratique de la méthode hybride de propagation sont qu'elle requiert de nombreux calculs et ne fournit pas d'évaluation de l'erreur due à l'approximation numérique réalisée durant la propagation. Si ce problème est mineur lorsque le modèle à travers lequel on propage est simple et que beaucoup de calculs peuvent être réalisés sans coûts élevés, il ne l'est plus lors de propagations à travers des codes de calculs complexes, pour lesquels le nombre de simulations (calculs) à pouvoir réaliser est généralement limité, du fait de leur coût élevé, à la fois en temps et en argent.

Nous proposons donc une méthode numérique pratique, appelée RaFu et développée dans le logiciel SUNSET, qui reprend les bases théoriques de la propagation hybride tout en se proposant d'optimiser le nombre de simulations à réaliser (d'échantillons à considérer) pour atteindre un objectif donné. Elle se base sur le fait qu'en pratique, le décideur désirera rarement obtenir la variable aléatoire floue dans son entièreté, mais seulement quelques une de ses caractéristiques.

La méthode RaFu consiste donc à demander au décideur, avant d'effectuer la propagation, de spécifier la nature d'un triplet de paramètres $(\gamma_S, \gamma_E, \gamma_A)$ correspondant à la réponse désirée: γ_S correspond aux aspects statistiques de la réponse désirée; γ_E concerne les aspects épistémiques, c'est-à-dire concernant les distributions de possibilité; γ_A sert à spécifier la précision numérique qui veut être atteinte, et permet un contrôle de l'erreur numérique.

Nous appliquons ensuite la méthode RaFu à un cas d'étude simplifié afin de l'illustrer. Ce cas d'étude consiste à évaluer le niveau de couronnement d'un barrage afin de le dimensionner. Ce cas d'étude, bien que simple, nous permet de mettre en évidence l'effet de l'approximation numérique, et l'importance de prendre en compte cette erreur numérique dans les calculs.

1.7 Conclusions et perspectives (Chapitre 8)

Dans ce travail, nous avons étudié plusieurs aspects du traitement des incertitudes en présence d'imprécision, progressant à la fois vers une unification des différentes théories et vers des outils pratiques permettant de facilement manipuler l'information.

Du chapitre 3, dans lequel nous avons étudié les représentations pratiques d'incertitude, à la fois anciennes et plus récentes, nous pouvons conclure que les p-boxes généralisées (ou, de façon équivalente, les nuages comonotones) sont des représentations aux propriétés intéressantes qui permettent de faire le lien entre p-box classiques, possibilités et nuages. Leur interprétation en terme de bornes de confiance sur des intervalles emboîtés les rend également intéressantes du point de vue de l'élicitation experte.

Les perspectives données dans ce chapitre sont principalement l'étude de la manipulation pratique des p-boxes généralisées et des nuages (définir et étudier le conditionnement, la propagation, la fusion, la marginalisation, ...) ainsi que l'extension des résultats obtenus à des domaines de définition plus généraux. Ces perspectives ont été déjà abordées, bien que de façon incomplète, dans le présent travail. Des optiques de recherche moins directes mais qui pourraient s'avérer intéressantes consisteraient à explorer les connections qui pourraient exister entre ces représentations et la notion de bipolarité ou encore avec la théorie des ensembles flous valués par intervalles.

Le chapitre 4 se conclut avec quelques recommandations concernant l'utilisation des opérateurs de fusion d'informations au sein des théories de l'incertain: les opérateurs disjonctifs et conjonctifs devraient être réservés aux cas où les sources sont respectivement totalement inconsistantes et consistantes, et les opérateurs de compromis devraient être utilisés dans les situations intermédiaires. Concernant ces derniers, nous avons étudié de plus près les opérateurs s'appuyant sur la notion de sous-ensembles maximaux cohérents, qui nous apparaissent comme les mieux adaptés au traitement des informations partiellement consistantes, à la fois théoriquement et conceptuellement. Néanmoins, leur application peut poser quelques problèmes calculatoires, et il est nécessaire de chercher des méthodes efficaces permettant de les mettre en oeuvre, soit en utilisant des cadres simplifiés, soit en développant des heuristiques efficaces. En nous restreignant aux distributions de possibilités définies sur les réels, nous avons opté pour la première solution. Nous avons également commencé l'étude de la prise en compte de dépendances mal connues dans le cadre des ensembles aléatoires.

En ce qui concerne la méthode SMC appliquées aux distributions de possibilité, elle demande surtout à être validée à un niveau pratique, tout comme l'approche que nous avons

proposée pour évaluer les sources d'information à partir de performances passées. L'approche proposée pour prendre en compte les dépendances entre sources demande quant à elle à être étudiée de plus près, pour l'étayer d'un point de vue théorique et la rendre plus accessible d'un point de vue pratique.

Le chapitre 5 concernant l'indépendance a engendré plus de questions qu'il n'a fourni de réponses, la modélisation et l'interprétation de l'indépendance lorsque l'imprécision est prise en compte requérant un long travail de recherche. Néanmoins, nous avons apporté quelques éléments de réponse en proposant un début de taxonomie permettant de classifier les notions d'indépendance, et en débutant l'étude consistant à interpréter ces notions d'indépendance au travers d'arbres d'événements. Les perspectives incluent, entre autres choses, la clarification des points d'interrogations laissés dans le tableau 1.1, ainsi que la poursuite de l'étude des notions d'indépendance dans les arbres d'événements.

Dans les chapitres 6 et 7, nous avons surtout abordé des problèmes d'ordre pratique liés à l'utilisation de modèles représentant explicitement l'imprécision. Les perspectives pour ces deux chapitres consistent principalement à poursuivre l'effort d'analyse et à proposer des solutions plus générales ou plus efficaces aux problèmes posés, notamment en ce qui concerne le calcul de bornes d'espérances pour des représentations définies sur les réels (e.g. en étendant nos résultats à d'autres cas et représentations) et la propagation numérique d'incertitude à travers des modèles complexes (e.g. par l'utilisation de surfaces de réponses ou de techniques du type MCMC).

Chapter 2

Introduction

*“Everything should be made as simple as possible, but not **simpler**”*

— Albert Einstein (1879–1955)

This work presents results related to the treatment of uncertainty bearing on variables whose exact value is not perfectly known, this lack of knowledge being due either to the aleatory nature of some phenomena influencing this value or to a lack of precise and fully reliable information concerning this value. More precisely, we interest ourselves to the case where uncertainty is modeled by numerical (quantitative) representations, which are neither (precise) probability distributions (because we do not have sufficient information) nor sets (because we do have information about which elements of the space are more likely to be observed).

In recent decades, different uncertainty theories have emerged to address properly this kind of situations. In this work, we restrict ourselves to the three main such theories: possibility theory, random set theory and imprecise probability theory (see Appendix A). We could say that the position we have with respect to uncertainty treatment is somewhat dual, in the following sense:

- We attach a great importance to unification of uncertainty handling, in the sense that we think it essential to make bridges and to emphasize convergence points between different theories, rather than confining ourselves to one exclusive theory.
- We consider each uncertainty theory as potentially useful *per se*, as long as it is sufficiently theoretically justified. Indeed, we see the absolute statement that one is "better"

than the other as overrated: some theories are more general than others, some are more fitted to a given situation than others, some are more mature than others, some have interpretations better fitted to a particular situation or problem, some dispose of more convenient tools than others. To us, the main question is not "which is the best?" but rather "when, where, why and how each theory should be used?"

2.1 Reasoning under uncertainty (with quantitative models): a general view

By reasoning, we mean manipulating information in a sensible and, as far as possible, rational way, in order to derive plausible and useful conclusions.

By under uncertainty, we mean that the available information does not allow to perfectly know all component of considered systems, and that we are uncertain about their exact current state or value.

We also make a distinction between two levels of information: *generic information* corresponding to background knowledge and general beliefs about the world, and *contingent information* corresponding to information, belief or knowledge concerning a peculiar situation. For example, that *birds fly in general* is a generic information, while any information concerning my next-door neighbor bird is peculiar. A computer code or an analytical function modeling the evolution of the temperature in a nuclear reactor core during an accident are *generic information*, while the values observed during a particular accident are contingent. Algorithms encoded in a robot constitute its generic information, while information received by it through sensors or other devices is contingent.

We define the following general simplified frame, pictured in Figure 2.1 which will be instrumental to define problems considered in the sequel:

- *Source variables* are variables about which we have some information, that is we have an idea about the value they assume.
- *Variables of interest* are those variables on which we want to have information, because knowing their value will help in further decisions, but for which we do not have direct information.
- *A model* models generic information about the relationship existing between source variables and variables of interest. It allows to use information concerning source variables

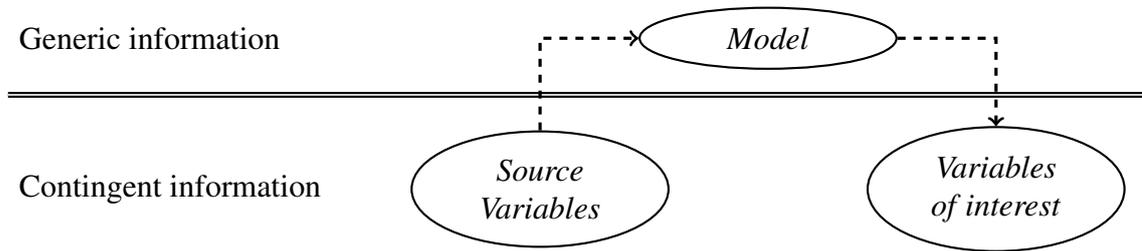


Figure 2.1: Uncertainty treatment: general frame

in order to have information about variables of interest

Depending on the situation, the uncertainty can bear on source variables (e.g. input variables in risk analysis) or on the models (e.g. Markov chains, Bayes networks). There can be multiple source variables, variables of interest or models. Also, this frame can be chained, in the sense that some variables that were of interest for one system can become source variables for another system (e.g. experts use source variables and internal models to provide knowledge on some variables of interest, which will become source variables in a subsequent propagation through another model).

The following problems, commonly encountered in uncertainty treatments, all fit in Figure 2.1:

- **Model choice/design:** the process of choosing and designing a model of the system on which further treatments are to be done. This step requires careful thinking, as many models can compete as good candidate for a particular system. For instance, fuzzy rule bases, neural networks or SVM can all be used as universal approximators of functions, and it is not always clear which one should be used in which situation. Statistical tests procedures can be seen as tools to check that a particular model is fitted to the considered system.
- **Model Identification/validation:** once the model is chosen, it remains to identify its features for the particular problem at hand. This process consists in using available knowledge in order to build or identify a model. It is principally an inductive process, since it mainly consists of using contingent knowledge to build a generic model within the chosen framework, which can then be used to make inferences. In AI, this process can be associated to learning, while in statistics it corresponds more to inductive or parametric inference, whose aim are to identify a generic model (e.g. parameters) from the data.

- **Inference:** inference is defined as the process of drawing (plausible) conclusions from premises or evidence. It is primarily a deductive, theoretical and rather impersonal process. In this work, we interpret inference as the act of drawing conclusions on variables of interest from observations on source variables, with the help of a (fixed) generic model. It thus consists of using generic information to draw plausible conclusions from contingent information. Note that, in classical statistics, what we consider here as inference is often called prediction. Typical inference processes encompass
 - **(direct) propagation through deterministic model:** propagating uncertainty on source variables through a deterministic model (e.g. a computer code, an analytical model, ...) to evaluate the uncertainty on variables of interest. By deterministic model, we mean a model such that to one precise input correspond one precise output. This kind of inference is the most usual in industrial risk analysis, where most of the time, models are analytical formulas or computer codes simulating complex physical phenomena. Such inference processes are monotonic with respect to uncertainty in our knowledge, in the sense that reducing uncertainty on source variables will reduce uncertainty on variables of interest once propagation is done.
 - **inverse propagation:** similar to direct propagation, except that this time, source variables on which information are available are the output of the (deterministic) model and interest variables are (unknown) parameters or inputs of the models, and the aim is to infer the most plausible values of these inputs or parameters. The difficulty is that most of the time the model is not invertible and dependencies between inputs, outputs and parameters are not known. Note that this inference process remains monotonic in the same sense as in above: the more we know about source variables, the less the resulting uncertainty on variables of interest.
 - **propagation through/conditioning on uncertain models:** Given some observations on source variables and a so-called stochastic model (i.e. Markov Chain, Bayes Net, Probability tree), infer plausible value of the variable of interest. This type of inference is more commonly encountered in Artificial Intelligence. It can be associated to the act of focusing our generic beliefs (or information) on a subclass corresponding to our observations. We assume here that the model is uncertain, but that our observations are not. In this case, it is well known that monotonicity with respect to uncertainty about singular information do not hold, since it can happen that a more precise observation can give us less decisive inferences. This phenomenon is often referred as dilation within uncertainty theories.

- **Information fusion:** If multiple sources provide information about the same source variables, variables of interest or potential models, information fusion consists in merging all these information items into a reliable and informative summary, while coping with possible dependencies between sources and inconsistencies in information. Information fusion only makes sense with information of an equivalent level of generality, i.e. merging contingent information with contingent information, and generic information with generic information.
- **Decision making:** the process of determining optimal actions, given the current evidence on variables of interest. Decision have consequences, in the sense that, once optimal actions are determined, applying one of them changes the environment. Decision is also more personal, since an action that is optimal for a subject¹ will not necessarily be so for another subject (e.g. employees and shareholders of same company do not generally share the same objectives). Determining an optimal act for a given subject is typically done by eliciting utilities or preferences from the subject on a set of different feasible actions. In our opinion, processes of inference and of decision making should be considered separately, since even if they're closely related (inference is often used to make decisions), their respective purposes are different.
- **Revision:** the process of revision consists in modifying our current knowledge or beliefs in minimal way, given the arrival of new information which are not necessarily coherent with our current knowledge or beliefs. As for information fusion, revising only makes sense with information having a similar level of generality.

Of course, our picture does not encompass all the complexity encountered in real applications nor the variety of frameworks dealing with uncertainty, and in practice distinguishing between different processes and between information levels of generality is not an easy task. . . Nevertheless, such a picture (and other similar representations) is a good starting point, and can serve as a useful guideline to answer the questions "when, where, why and how use particular tools?".

2.2 About the present work

This work studies some of the problems we have just mentioned, and more specifically those commonly encountered in industrial risk analysis or in safety studies. For each of these prob-

¹Subject is taken here in a wide sense: it can be an organization, a whole country, or a single person

lems, and in accordance with our view about uncertainty treatment, we...

- ...position the general problematic, make a partial review of solutions proposed by uncertainty theories, and as much as possible, recall or show the links existing between these solutions.
- ...propose methodologies bringing solution to the considered problem, some of them being set in the frame of a particular theory of uncertainty. Our first objective while developing these methodologies was to keep them tractable and easily applicable.

In Chapter 3, we study the problem of representing uncertainty about the value assumed by a variable X on a finite domain \mathcal{X} . We give special attention to practical uncertainty models allowing for an easier handling of uncertainty in applications. In particular, we study extensively the relations between the following models: possibility distributions, imprecise probability assignments, p-boxes, clouds and random sets. To be able to relate more efficiently these models, we introduce the notion of generalized p-boxes, which will be instrumental to relate possibility distributions, p-boxes and clouds together. Some attention is given to continuous models defined on the real line and to so-called hierarchical models, that is models defined on multiple levels.

Chapter 4 concerns the treatment of uncertainty when multiple sources all provide information about the value that a single variable X that may assume on a finite domain \mathcal{X} , this information being modeled by representations introduced in Chapter 3.

In the first part of this chapter, we study the means to summarize the information provided by the sources in a synthetic, operational and interpretable message. We give some special attention to two different problems encountered by such synthesis:

- dealing with inconsistencies present between the pieces of information provided by the different sources. We propose the use of the logical notion of maximally coherent subsets as a way to cope with such inconsistencies. We fully study an extension of this notion to possibility distributions defined on the real line, and propose practical tools to make the method easier to use in practice.
- dealing, in the frame of random set theory, with sources whose (in)dependencies are not well known. With this respect, we give first results eventually leading to a practical cautious merging rule.

The second part is devoted to the problem of evaluating the quality of the information delivered by the sources. Since most of the time every source do not have the same reliability, it is desirable, when possible, to know which ones are more reliables than others. Here, we are concerned with those cases in which sources have previously given information about variables whose true value is now known. For these cases, we propose a general methodology allowing to evaluate sources on the basis of their previous assessments, that aims at being as objective as possible in its evaluation.

Chapter 5 concentrates on the (many) notions of independence that can link multiple input variables X_1, \dots, X_N assuming values on finite domains $\mathcal{X}_1, \dots, \mathcal{X}_N$.

In classical probability theory, all notions of independence formally reduce to the stochastic independence, irrespectively of their interpretation. This is no longer the case when using imprecise probabilistic models, for which there are as many distinct formal definitions of independence as there are distinct interpretations. Since the notion of independence is central in the construction of joint uncertainty models from marginal ones (a situation that often happens in risk analysis), we review and attempt to browse a general picture of independence notions in imprecise probability theories. We then give first results indicating that event trees framework is a promising framework when it comes down to interpret and use independence assessments.

Chapter 6 then briefly addresses the problem of decision making under uncertainty. In this chapter, we quickly review different usual criteria used to determine optimal actions from a set of possible actions (we assume that actions can be associated to a utility function), and we then give some results about the practical computation of (lower and upper) expected utilities when uncertainty models are lower and upper cumulative distributions (i.e., p-boxes) defined on the real line. The peculiar problem of decision making in industrial risk analyses or safety studies is then adressed.

Finally, Chapter 7 shows two illustrative applications developed with the help of SUNSET, the software for uncertainty treatment developed at the Institut de Radioprotection et de Sûreté Nucléaire (IRSN). The first concerns the application of methods developed in Chapter 4 to results of uncertainty studies performed with nuclear computer codes simulating an accident in a nuclear reactor core. The second concerns a numerical propagation technique developed in the SUNSET software and called RaFu. The method is first described, before being applied to a case-study concerning the design of a dam.

Some concluding words are then provided in Chapter 8.

Chapter 3

Practical uncertainty representations

“Knowing ignorance is strength. Ignoring knowledge is sickness”

— Lao Tse (~ 500 B.C.)

Contents

| | | |
|------------|--|-----------|
| 3.1 | Non-additive measures and representations of uncertainty | 51 |
| 3.1.1 | Capacities and transformations of capacities | 52 |
| 3.1.1.1 | n-monotonicity | 53 |
| 3.1.1.2 | Möbius inverse | 54 |
| 3.1.2 | Practical representations in imprecise probability | 55 |
| 3.1.2.1 | Lower/upper probabilities | 56 |
| 3.1.2.2 | Probability boxes (p-boxes) | 57 |
| 3.1.2.3 | Imprecise probability assignments | 58 |
| 3.1.2.4 | Random (disjunctive) sets | 59 |
| 3.1.2.5 | Possibility distributions | 60 |
| 3.1.3 | Sketching a first summary of relationships | 63 |
| 3.1.3.1 | P-boxes in the landscape of uncertainty representations | 64 |
| 3.1.3.2 | Imprecise probability assignments in the landscape of un- certainty representations | 64 |
| 3.1.3.3 | Preliminary summary | 65 |
| 3.2 | Introduction and study of generalised p-boxes | 65 |
| 3.2.1 | Definition of generalized p-boxes | 66 |
| 3.2.2 | Connecting generalized p-boxes with possibility distributions | 69 |

| | | |
|------------|--|------------|
| 3.2.3 | Connecting Generalized p-boxes and random sets | 70 |
| 3.2.4 | Generalized p-boxes and imprecise probability assignments | 74 |
| 3.2.4.1 | Approximations between the two representations | 74 |
| 3.2.4.2 | Linking the two representations | 77 |
| 3.2.5 | Computing with generalized p-boxes: first results on propagation | 79 |
| 3.3 | Clouds | 82 |
| 3.3.1 | Definition of clouds | 83 |
| 3.3.2 | Clouds in the setting of possibility theory | 85 |
| 3.3.2.1 | General clouds and possibility distributions | 86 |
| 3.3.2.2 | Using possibility distributions to check non-emptiness of $\mathcal{P}_{[\pi, \delta]}$ | 86 |
| 3.3.3 | Generalized p-boxes as a special kind of clouds | 88 |
| 3.3.4 | The Nature of Non-comonotonic Clouds | 91 |
| 3.3.4.1 | Characterization | 92 |
| 3.3.4.2 | Outer approximation of a non-monotonic cloud | 94 |
| 3.3.4.3 | Inner approximation of a non-comonotonic cloud | 95 |
| 3.3.5 | Clouds and imprecise probability assignments | 96 |
| 3.3.5.1 | Exploiting probability-possibility transformations | 96 |
| 3.3.5.2 | Using generalized p-boxes | 101 |
| 3.4 | A word on continuous representations on the real line | 103 |
| 3.4.1 | Practical continuous representations on the real line | 103 |
| 3.4.2 | Continuous clouds on the real line | 105 |
| 3.4.3 | Thin continuous clouds | 108 |
| 3.5 | Combinations of uncertainty representations into higher order models | 109 |
| 3.5.1 | A quick review of the literature | 110 |
| 3.5.2 | Fuzzy random variables | 112 |
| 3.5.2.1 | Interpretation as a 1 st order model | 112 |
| 3.5.2.2 | Interpretations as a 2 nd order model | 114 |
| 3.6 | Conclusions and perspectives | 115 |

When we are not certain about the value assumed by a variable X in a space \mathcal{X} , there exist several practical representations that can model this uncertainty. Such simple representations include, but are not limited to: sets, probability distributions [108], possibility distributions [85], imprecise probability assignments [42], random sets [151], probability boxes (p-boxes for short) [104], random fuzzy variables [34] and, more recently, clouds [159].

Mathematically, all these representations can be interpreted as closed convex sets of (finitely) additive probabilities, and are therefore less general than this latter representation. Although

less generality implies less expressiveness, it also often allows for a more efficient handling of uncertainty. Simplified representations are thus of importance when we have to trade expressiveness (possibly losing some information) against computational efficiency. They are also instrumental in elicitation tasks and in the interpretation or representation of complex results. Moreover, in a number of cases, they will be sufficient to faithfully model the available information.

With such a bunch of simplified representations, it seems natural to study their links as well as to compare their respective expressive power. Such a study is the purpose of the present chapter, in which we explore the relationships between various representations. Laying bare relationships facilitates a unified handling and treatment of uncertainty, and suggests how tools used for one theory can eventually be useful in the setting of other theories.

The main contribution of this chapter is to propose a generalised version of p-boxes and to show that it constitutes the missing link between possibility distributions, usual p-boxes and clouds. We first present and briefly recall the basic settings to represent uncertainty, the representation studied in the sequel and the known links between them (Section 3.1). We then introduce and study a generalised version of p-boxes, subsequently used to link between possibility distributions, p-boxes and clouds (Section 3.2). We then explore the recent formalism of clouds and its link with other representations (Section 3.3). We also study the extension of some of our results to representations defined on the continuous real line (Section 3.4), before considering the combination of uncertainty representations into so-called hierarchical models (Section 3.5).

3.1 Non-additive measures and representations of uncertainty

As argued in Appendix A, single probability distributions, as uncertainty models, cannot adequately account for scarceness, imprecision or unreliability in the available information or knowledge.

Alternative representations and theories considered in this work (i.e., imprecise probability theory [203], random (disjunctive) sets [151], possibility theory [85]) have the potential to lay bare the existing imprecision or incompleteness in the information. This imprecision is expressed by the means of a pair of (conjugate) lower and upper confidence measures on events rather than by a single one.

In this section, we recall the main mathematical tools used to characterize these repre-

sentations, before reviewing the main practical numerical representation tools available as of to-date and the known links between them.

3.1.1 Capacities and transformations of capacities

Set-functions called *capacities* [25] are handy tools to represent uncertainty.

Definition 3.1 (Capacity). Given a finite space \mathcal{X} , a capacity on \mathcal{X} is a function μ , defined on the power set $\wp(\mathcal{X})$ of \mathcal{X} , such that:

- $\mu(\emptyset) = 0, \mu(\mathcal{X}) = 1$ (boundary conditions)
- $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ (monotonicity)

A capacity such that

$$\forall A, B \subseteq \mathcal{X}, A \cap B = \emptyset, \mu(A \cup B) \geq \mu(A) + \mu(B) \quad (3.1)$$

is said to be *super-additive*. The dual notion, called *sub-additivity*, is obtained by reversing the inequality in Equation (3.1).

A capacity is said to be *additive* if the inequality in Equation (3.1) is turned into an equality. An additive capacity is formally equivalent to a probability measure, denoted P . When \mathcal{X} is finite, a probability P can also be expressed by its probability distribution p defined on \mathcal{X} such that $p(x) = P(\{x\})$. Then $\forall x \in \mathcal{X}, p(x) \geq 0, \sum_{x \in \mathcal{X}} p(x) = 1$ and $P(A) = \sum_{x \in A} p(x)$. We note by $\mathbb{P}_{\mathcal{X}}$ the set of all probability distributions on \mathcal{X} .

Given a capacity μ on \mathcal{X} , its *conjugate* capacity μ^c is defined by $\mu^c(E) = \mu(\mathcal{X}) - \mu(E^c) = 1 - \mu(E^c)$ for any subset $E \subseteq \mathcal{X}$ with E^c the complement of E . We call *cautious* the super-additive capacities, since $\mu(E) + \mu(E^c) \leq 1$ for any subset $E \subseteq \mathcal{X}$, and *bold* the sub-additive capacities, since $\mu(E) + \mu(E^c) \geq 1$ for any subset $E \subseteq \mathcal{X}$. Note that additive capacities (i.e., probability measures) are both *cautious* and *bold*.

When used to represent and model uncertainty (which is the case in this work), the value of a capacity on a subset evaluates the degree of confidence in the corresponding event. Cautious capacities are tailored for modeling the idea of certainty. Bold capacities may account for the weaker notion of plausibility.

A probability measure P in $\mathbb{P}_{\mathcal{X}}$ is said to *dominate* a capacity μ on \mathcal{X} if and only if, we have $\mu(E) \leq P(E)$ for every subset $E \subseteq \mathcal{X}$. The *core* \mathcal{P}_{μ} of a capacity μ on \mathcal{X} is the (closed convex) set of probability measures dominating it, that is

$$\mathcal{P}_{\mu} = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall E \subseteq \mathcal{X}, \mu(E) \leq P(E)\}. \quad (3.2)$$

Note that the core of a cautious capacity can be empty, since cautiousness is a necessary but not sufficient condition for a capacity to have a non-empty core. Necessary and sufficient conditions for non-emptiness are provided by Walley [203, Ch.2], but checking these conditions often involve to check an high number of inequalities, making them hard to use in practice. An alternative to check the non-emptiness of the core is the use of specific characteristics of capacities, such as n -monotonicity.

3.1.1.1 n -monotonicity

Choquet [25] defines n -monotonicity as follows:

Definition 3.2 (n -monotonicity). A super-additive (cautious) capacity μ on \mathcal{X} is n -monotone, where $n > 0$ and $n \in \mathbb{N}$, if and only if for any set $\mathcal{A} = \{A_i \subseteq \mathcal{X} \mid i \in \mathbb{N}, 0 < i \leq n\}$ of events A_i , it holds that

$$\mu\left(\bigcup_{A_i \in \mathcal{A}} A_i\right) \geq \sum_{I \subseteq \mathcal{A}} (-1)^{|I|+1} \mu\left(\bigcap_{A_i \in I} A_i\right) \quad (3.3)$$

The conjugate capacity μ^c of a n -monotone capacity is called n -alternating. If a capacity is n -monotone, then it is also $(n-1)$ -monotone, but not necessarily $(n+1)$ -monotone. An ∞ -monotone capacity is a capacity that is n -monotone for every $n > 0$. On a finite space, a capacity is ∞ -monotone if it is n -monotone with $n = |\mathcal{X}|$. A n -monotone capacity or its dual are also often called Choquet capacities of order n .

The two particular cases of 2-monotone (also called convex) capacities and ∞ -monotone capacities have deserved special attention in the literature [22, 203, 145]. Indeed, 2-monotone capacities always have a non-empty core and ∞ -monotone capacities have interesting mathematical properties that greatly increase computational efficiency when manipulating them. Most of the representations studied in this chapter have such properties.

It must be noticed that Choquet's initial definition of n -monotonicity is very general and is not restricted to events, contrary to what Definition 3.2 could suggests. De Cooman *et al.* [51]

consider a generalization of Definition 3.2 on lattices of bounded real-valued functions on \mathcal{X} , and study yet a more generalized version of n -monotonicity in a subsequent work [52], essentially by dropping the normalization condition ($\mu(\mathcal{X}) = 1$) of Definition 3.1. Nevertheless, Definition 3.2 will be sufficient in most parts of this work (n -monotonicity on lattices of bounded real-valued functions is used in Appendix F to study p-boxes defined on totally ordered spaces).

3.1.1.2 Möbius inverse

Given a capacity μ on \mathcal{X} , one can obtain multiple equivalent representations by applying various (bijective) transformations to it [115]. Using such transformations can be of practical usefulness when manipulating capacities. One such transformation, useful in this work, is the Möbius inverse:

Definition 3.3 (Möbius inverse). Given a capacity μ on \mathcal{X} , its Möbius transform is a mapping $m : \wp(\mathcal{X}) \rightarrow \mathbb{R}$ from the power set of \mathcal{X} to the real line, which associates to any subset E of \mathcal{X} the value

$$m(E) = \sum_{B \subseteq E} (-1)^{|E \setminus B|} \mu(B)$$

And we have $\sum_{E \in \mathcal{X}} m(E) = 1$, $m(\emptyset) = 0$, due to the boundary conditions on capacities. Moreover, the following proposition holds:

Proposition 3.1. [22] *Let μ be a capacity on \mathcal{X} . Then, its Möbius transform m is non-negative if and only if μ is ∞ -monotone.*

Otherwise, there are some events E for which $m(E)$ is negative. The set-function m is actually the unique solution [178, Ch.2.7] to the set of 2^n equations

$$\forall A \subseteq \mathcal{X}, \mu(A) = \sum_{E \subseteq A} m(E),$$

given any capacity μ . The Möbius transform of a probability measure P coincides with its distribution p , assigning positive masses to singletons only.

Remark 3.1. Möbius inverse can be applied to any mapping $f : \wp(\mathcal{X}) \rightarrow \mathbb{R}$ such that to this mapping f is associated the mass function m_f taking, for any event $E \subseteq \mathcal{X}$, the value

$$m_f(E) = \sum_{B \subseteq E} (-1)^{|E \setminus B|} f(B)$$

and this transformation remains bijective, since for any event $E \subset \mathcal{X}$ we can retrieve $f(E)$ by computing

$$f(E) = \sum_{B \subseteq E} m_f(B)$$

3.1.2 Practical representations in imprecise probability

We begin this section by general considerations about credal sets, before introducing the uncertainty representations we're going to study and relate together.

In Walley's [203] theory of imprecise probabilities, uncertainty is represented by lower bounds given over real-valued function of \mathcal{X} (i.e. so-called lower previsions, see Appendix A for details). Such lower bounds have an expressive power equivalent to closed convex sets \mathcal{P} of (finitely additive) probability measures P , and constitute one of the most general existing uncertainty model (although not the most general [206]). Such sets are commonly called *credal sets* [136], and will be so in the present work.

As imprecise probability theory is very general, we can express all representations considered in this work in terms of credal sets, making the comparison between uncertainty representations easier. To clarify this comparison, we adopt the following terminology:

Definition 3.4 (Representations relations). Let \mathbb{F}_1 and \mathbb{F}_2 denote two uncertainty representation frameworks, a and b particular representatives of such frameworks, and $\mathcal{P}_a, \mathcal{P}_b$ the credal sets induced by these representatives a and b . Then:

- Framework \mathbb{F}_1 is said to *generalize* framework \mathbb{F}_2 if and only if for all $b \in \mathbb{F}_2$, $\exists a \in \mathbb{F}_1$ such that $\mathcal{P}_a = \mathcal{P}_b$ (we also say that \mathbb{F}_2 is a special case of \mathbb{F}_1).
- Frameworks \mathbb{F}_1 and \mathbb{F}_2 are said to be *equivalent* if and only if for all $b \in \mathbb{F}_2$, $\exists a \in \mathbb{F}_1$ such that $\mathcal{P}_a = \mathcal{P}_b$ and conversely.
- Framework \mathbb{F}_2 is said to be *representable* in terms of framework \mathbb{F}_1 if and only if for all $b \in \mathbb{F}_2$, there exists a subset $\{a_1, \dots, a_k | a_i \in \mathbb{F}_1\}$ such that $\mathcal{P}_b = \mathcal{P}_{a_1} \cap \dots \cap \mathcal{P}_{a_k}$
- A representative $a \in \mathbb{F}_1$ is said to *outer-approximate* (inner-approximate) a representative $b \in \mathbb{F}_2$ if and only if $\mathcal{P}_b \subseteq \mathcal{P}_a$ ($\mathcal{P}_a \subseteq \mathcal{P}_b$)

3.1.2.1 Lower/upper probabilities

In this paper, uncertainty described by lower probabilities (lower previsions assigned to events) are sufficient to our purpose. We define a *lower* probability \underline{P} as a super-additive capacity on \mathcal{X} . The conjugate capacity noted \bar{P} is the dual upper probability. This duality allows us to work only on the lower (or the upper) bound. The credal set $\mathcal{P}_{\underline{P}}$ induced by a lower probability \underline{P} is its core:

$$\mathcal{P}_{\underline{P}} = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall A \subseteq \mathcal{X}, P(A) \geq \underline{P}(A)\} \quad (3.4)$$

Conversely, given a credal set \mathcal{P} , its lower envelope P_* on events is defined for every event $A \subseteq \mathcal{X}$ as $P_*(A) = \min_{P \in \mathcal{P}} P(A)$. As a lower envelope is a super-additive capacity, it is a lower probability. The upper envelope $P^*(A) = \max_{P \in \mathcal{P}} P(A)$ is the conjugate of P_* . In this work, we consider so-called *coherent* lower probabilities \underline{P} , that is, lower probabilities that coincide with the lower envelopes of their core, i.e. for all events A of \mathcal{X} ,

$$\underline{P}(A) = \min_{P \in \mathcal{P}_{\underline{P}}} P(A).$$

In general, the credal set $\mathcal{P}_{\underline{P}}$ induced by the lower envelope \underline{P} of an original credal set \mathcal{P} is such that $\mathcal{P} \subseteq \mathcal{P}_{\underline{P}}$, since $\mathcal{P}_{\underline{P}}$ is a projection of \mathcal{P} on events. To characterize general credal sets, we need the more powerful language of lower bounds on expected values of bounded-real valued functions, which are enough to completely characterize any credal set \mathcal{P} (see Appendix A and Walley [203]). Nevertheless, as we will see, restricting to lower probabilities is sufficient in many practical cases.

Describing $\mathcal{P}_{\underline{P}}$ by the values of \underline{P} on every elements of the power set $\wp(\mathcal{X})$ can be very tedious and computationally expensive. Other means that can be useful to describe $\mathcal{P}_{\underline{P}}$ include:

- The set $\mathbf{ext}_{\mathcal{P}_{\underline{P}}}$ of extreme points of the convex set $\mathcal{P}_{\underline{P}}$ (See Walley [203, Ch.3.] for general considerations and, among others, Quaeghebeur and de Cooman [169], Wallner [208] for practical considerations)
- The set of constraints on sums of probability assignments on elements of \mathcal{X} :

$$\forall A \subseteq \mathcal{X}, \underline{P}(A) \leq \sum_{x \in A} p(x) \leq \bar{P}(A). \quad (3.5)$$

And we say that these constraints are *consistent* if the credal set $\mathcal{P}_{\underline{P}}$ is non-empty (i.e., there exist a solution to the set of constraints (3.5)), and that they are *tight* if \underline{P} is a coherent lower probability (i.e., bounds of constraints (3.5) cannot be reduced without ruling out some solutions). Although both these descriptions can have a complexity as high as storing every value of \underline{P} , they can be useful to illustrate some points. Most practical representations do not exhibit such a complexity. We now introduce those representations.

3.1.2.2 Probability boxes (p-boxes)

Recall that a cumulative distribution F_1 is said to stochastically dominate another cumulative distribution F_2 if only if F_1 is point-wise lower than F_2 : $F_1 \leq F_2$.

A probability box [104] (p-box for short) is defined as a pair $[\underline{F}, \overline{F}]$ of (discrete) cumulative distributions on \mathbb{R} , such that \underline{F} stochastically dominates \overline{F} ($\underline{F} \leq \overline{F}$). A p-box $[\underline{F}, \overline{F}]$ induces the following credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$:

$$\mathcal{P}_{[\underline{F}, \overline{F}]} = \{P \in \mathbb{P}_{\mathbb{R}} \mid \forall r \in \mathbb{R}, \underline{F}(r) \leq P((-\infty, r]) \leq \overline{F}(r)\} \quad (3.6)$$

It is useful to notice at this point that sets $(-\infty, r]$ are nested, thus $\mathcal{P}_{[\underline{F}, \overline{F}]}$ can be described by lower and upper bounds on a collection of nested sets (already mentioned by Kozine and Utkin [130]). This characteristic will be central in the study of generalized p-boxes. Constraints induced by a p-box are consistent and tight as soon as $\underline{F} \leq \overline{F}$.

Outer approximation Given a credal set \mathcal{P} defined on \mathbb{R} , it is always possible to extract the corresponding p-box by considering its lower envelope restricted to events of the type $(-\infty, r]$, letting $\underline{F}(r) = \underline{P}((-\infty, r])$, $\overline{F}(r) = \overline{P}((-\infty, r])$, with $\underline{P}, \overline{P}$ the lower and upper probabilities of \mathcal{P} . By definition, the credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$ induced by this p-box is an outer approximation of \mathcal{P} (i.e., $\mathcal{P} \subseteq \mathcal{P}_{[\underline{F}, \overline{F}]}$), and $\mathcal{P}_{[\underline{F}, \overline{F}]}$ is the tightest outer approximation of \mathcal{P} induced by a p-box.

Practical aspects Cumulative distributions are often used in elicitation processes to extract (precise) probabilistic knowledge from experts [28]: p-boxes can directly benefit from such methods and elicitation tools, with the advantages of allowing some imprecision in the representation (e.g., allowing experts to give imprecise percentiles). So-called probabilistic arithmetic [209] also provides a very efficient numerical framework for particular statistical calculations with p-boxes. Finally, p-boxes are sufficient to represent and summarize final results

when only the violation of a threshold has to be checked (a usual situation in risk and safety studies).

3.1.2.3 Imprecise probability assignments

Imprecise probability assignments are another simple uncertainty representation. An imprecise probability assignment on \mathcal{X} is defined as a set of lower and upper bounds on elements x of \mathcal{X} . It can be represented and identified by a set L of intervals $L = \{[l(x), u(x)] | x \in \mathcal{X}\}$ such that $l(x) \leq p(x) \leq u(x)$ for all $x \in \mathcal{X}$, with $p(x) = P(\{x\})$. Imprecise probability assignments are studied extensively by De Campos *et al.* [42], who call them probability intervals. An imprecise probability assignment L induces the credal set

$$\mathcal{P}_L = \{P \in \mathbb{P}_{\mathcal{X}} | \forall x \in \mathcal{X}, l(x) \leq p(x) \leq u(x)\} \quad (3.7)$$

\mathcal{P}_L is thus defined by a set of $|\mathcal{X}|$ constraints bearing only on probability assignments. De Campos *et al.* [42] have studied necessary and sufficient conditions for these constraints to be consistent and tight (they call it respectively non-emptiness and reachability). These conditions correspond, for all $x \in \mathcal{X}$, to:

$$\sum_{x \in \mathcal{X}} l(x) \leq 1 \leq \sum_{x \in \mathcal{X}} u(x) \quad \text{consistency (non-emptiness)} \quad (3.8)$$

$$u(x) + \sum_{y \in \mathcal{X} \setminus \{x\}} l(y) \leq 1 \text{ and } l(x) + \sum_{y \in \mathcal{X} \setminus \{x\}} u(y) \geq 1 \quad \text{tightness (reachability)} \quad (3.9)$$

and any set L of non-tight (but consistent) constraints can easily be transformed in a set L' of tight constraints, by letting $l'(x) = \inf_{P \in \mathcal{P}_L} (p(x))$ and $u'(x) = \sup_{P \in \mathcal{P}_L} (p(x))$. From now on, we will always consider consistent and tight sets L , since others have little interest.

Given a imprecise probability assignment L , coherent lower and upper probabilities induced by \mathcal{P}_L on all events $A \subset \mathcal{X}$ are easily calculated by the following expressions

$$\underline{P}(A) = \max\left(\sum_{x \in A} l(x), 1 - \sum_{x \in A^c} u(x)\right), \quad \bar{P}(A) = \min\left(\sum_{x \in A} u(x), 1 - \sum_{x \in A^c} l(x)\right). \quad (3.10)$$

De Campos *et al.* [42] have shown that these lower and upper probabilities are Choquet capacities of order 2.

Outer-approximation Similarly to p-boxes, given a credal set \mathcal{P} on \mathcal{X} , it is always possible to extract the corresponding imprecise probability assignment L by considering its lower envelope restricted to elements, letting $l(x) = \inf_{p \in \mathcal{P}}(p(x))$ and $u(x) = \sup_{p \in \mathcal{P}}(p(x))$ for all $x \in \mathcal{X}$. Anew, the induced credal set \mathcal{P}_L is an outer approximation of \mathcal{P} , and it is the tightest outer approximation induced by an imprecise probability assignment.

Practical aspects Imprecise probability assignments are very convenient tools to model or represent uncertainty on multinomial data, where they can express lower and upper confidence probability bounds. They are particularly fitted to the case where only a small size sample is available [141]. On the real line, discrete probability intervals correspond to imprecisely known histograms. Computational advantages offered by imprecise probability assignments have been discussed at length by De Campos *et al.* [42] (some of them will be recalled in subsequent chapters).

3.1.2.4 Random (disjunctive) sets

A discrete random set (see Appendix A for more details), noted (m, \mathcal{F}) , over a space \mathcal{X} is defined as a mapping $m : \wp(\mathcal{X}) \rightarrow [0, 1]$ from the power set of \mathcal{X} to the unit interval, with $\sum_{E \subseteq \mathcal{X}} m(E) = 1$ and $m(\emptyset) = 0$. m is often called a *basic probability assignment (bpa)*, and we will sometimes use this terminology in this work. A set E receiving a positive mass is called a *focal element*, and we note \mathcal{F} the set of focal elements. From this mass assignment, Shafer [178] defines three set functions, called *belief, plausibility and commonality functions*, such that for every event $A \subseteq \mathcal{X}$:

$$Bel(A) = \sum_{E, E \subseteq A} m(E) \quad (\text{Belief}).$$

$$Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E) \quad (\text{Plausibility}).$$

$$Q(A) = \sum_{E, E \supseteq A} m(E) \quad (\text{Commonality}).$$

It can be checked [178, Ch.2.3] that the belief function of a random set is an ∞ -monotone capacity, and that the associated mass assignment is its Möbius transform. Conversely, any ∞ -monotone capacity is induced by one and only one random set.

Links with previous representations The belief function induced by a random set (m, \mathcal{F}) being an ∞ -monotone capacity, it can be interpreted as a special case of a coherent lower probability. In this case, a random set (m, \mathcal{F}) induces the credal set

$$\mathcal{P}_{(m, \mathcal{F})} = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall A \subseteq \mathcal{X}, Bel(A) \leq P(A) \leq Pl(A)\} \quad (3.11)$$

We are not aware of any practical and general solution allowing to build, from a given credal set \mathcal{P} , a random set (m, \mathcal{F}) such that the associated credal set $\mathcal{P}_{(m, \mathcal{F})}$ is a tight outer approximation of \mathcal{P} (tight in the sense that any random set $(m, \mathcal{F})'$ such that $\mathcal{P}_{(m, \mathcal{F})}' \subseteq \mathcal{P}_{(m, \mathcal{F})}$ would no longer induce an outer approximation of \mathcal{P}). Due to the potential complexity of the random set representation, this problem is far from obvious in the general case. Solutions for particular cases can nevertheless be proposed.

Practical aspects In general, $|\wp(\mathcal{X})| - 2$ values are still needed to completely specify a random set, thus not necessarily reducing the complexity of the model representation with respect to capacities. However, belief functions used in practice are often defined by only a few positive focal elements, and do not exhibit such a complexity. Such simpler belief functions can result from expert judgments or from statistical experiments, $m(A)$ becoming the probability of an observation or testimony of the form $x \in A$.

As practical models of uncertainty, random sets have many advantages. First, as they can be seen as probability distributions over subsets of \mathcal{X} , they can be easily simulated by classical methods such as Monte-Carlo sampling, which is not the case for Choquet capacities that are not ∞ -monotone. On the real line, a discrete random set is often restricted to a finite collection of closed intervals with associated weights, and one can then easily extend results from interval analysis [152] to random intervals [91, 118].

3.1.2.5 Possibility distributions

Possibility distributions are the primary mathematical tools of possibility theory (see Appendix A). A possibility distribution is a mapping $\pi : \mathcal{X} \rightarrow [0, 1]$ from a space \mathcal{X} to the unit interval such that $\pi(x) = 1$ for at least one element x in \mathcal{X} .

From a possibility distribution π can be defined several set-functions [79], among which

the *possibility, necessity and sufficiency measures*:

$$\Pi(A) = \sup_{x \in A} \pi(x) \quad (\text{Possibility measures}). \quad (3.12)$$

$$N(A) = 1 - \Pi(A^c) \quad (\text{Necessity measures}). \quad (3.13)$$

$$\Delta(A) = \inf_{x \in A} \pi(x) \quad (\text{Sufficiency measures}). \quad (3.14)$$

Their characteristic properties are: $N(A \cap B) = \min(N(A), N(B))$ and $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ for any pair of events A, B of \mathcal{X} . A possibility measure is usually said to be *maxitive*.

Given a degree $\alpha \in [0, 1]$ the strong ($A_{\bar{\alpha}}$) and regular (A_{α}) α -cuts of a distribution π are subsets respectively defined as

$$A_{\bar{\alpha}} = \{x \in \mathcal{X} \mid \pi(x) > \alpha\} \quad (3.15)$$

$$A_{\alpha} = \{x \in \mathcal{X} \mid \pi(x) \geq \alpha\} \quad (3.16)$$

These α -cuts are nested, since if $\alpha > \beta$, then $A_{\alpha} \subseteq A_{\beta}$. On finite spaces, the set of values $\{\pi(x), x \in \mathcal{X}\}$ is of the form $\alpha_0 = 0 < \alpha_1 < \dots < \alpha_M = 1$, meaning that in this case there is only M distinct α -cuts.

Links with previous representations A necessity measure (resp. a possibility measure) can be viewed as a belief function (resp. a plausibility function), whose associated random set has nested focal elements (Already noticed by Shafer [178, Ch.10], who calls such random sets consonant). A possibility distribution π , defines a random set $(m, \mathcal{F})_{\pi}$ having, for $i = 1, \dots, M$, the following focal sets E_i with masses $m(E_i)$ [82]:

$$\begin{cases} E_i = \{x \in \mathcal{X} \mid \pi(x) \geq \alpha_i\} = A_{\alpha_i} \\ m(E_i) = \alpha_i - \alpha_{i-1} \end{cases} \quad (3.17)$$

In this nested situation, the same amount of information is contained in the mass function m and the possibility distribution $\pi(x) = Pl(\{x\})$. In this case, the plausibility, belief and commonality measures are respectively equivalent to the possibility, necessity and sufficiency measure of the associated possibility measure.

Since the necessity measure is a particular belief function it is also an ∞ -monotone capacity, hence a particular coherent lower probability. If the necessity measure is viewed as a

coherent lower probability, its possibility distribution induces the credal set

$$\mathcal{P}_\pi = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall A \subseteq \mathcal{X}, N(A) \leq P(A) \leq \Pi(A)\} \quad (3.18)$$

It is useful to recall here a result proved by Dubois *et al.* [78], and by Couso *et al.* [31] in a much more general setting, which links probabilities P that are in \mathcal{P}_π with constraints on α -cuts:

Proposition 3.2. *Given a possibility distribution π and the induced convex set \mathcal{P}_π , then $P \in \mathcal{P}_\pi$ if and only if we have, for all α in $(0, 1]$:*

$$1 - \alpha \leq P(\{x \in \mathcal{X} \mid \pi(x) > \alpha\})$$

This result means that the probabilities P in the credal set \mathcal{P}_π can also be described in terms of constraints on strong α -cuts of π (i.e. $1 - \alpha \leq P(A_\alpha)$). When \mathcal{X} is finite, this comes down to characterize \mathcal{P}_π with M constraints that are lower probabilities on nested sets.

Outer approximation As for p-boxes and imprecise probability assignments, given a credal set \mathcal{P} , it is relatively easy to extract a possibility distribution π from \mathcal{P} such that the induced credal set \mathcal{P}_π is an outer approximation of \mathcal{P} . The procedure to build such a distribution is given by Algorithm 1. This algorithm obviously depends on the rankings of elements of \mathcal{X} . There are $|\mathcal{X}|!$ ways of choosing this ranking, potentially resulting in $|\mathcal{X}|!$ possibility distributions. Thus, unlike the case of p-boxes and imprecise probability assignments, there is not a unique tightest distribution π extractable from \mathcal{P} such that $\mathcal{P} \subseteq \mathcal{P}_\pi$. Nevertheless, the possibility distribution π built through Algorithm 1 is the tightest such that $\mathcal{P} \subseteq \mathcal{P}_\pi$, **given a specific ranking** of elements of \mathcal{X} . Up to now, we are not aware of existing efficient methods to determine the ranking giving one of the most specific possibility distribution covering \mathcal{P} . We can nevertheless mention the work of Dubois and Prade [89], who consider the problem of building outer and inner consonant approximations (i.e., possibility distributions) to a given random set. In this work, outer approximations are built by an algorithm similar to Algorithm 1 (but restricted to random sets), and an algorithm to find the covering possibility distribution having the minimal expected cardinality¹ is given.

¹For a random set (m, \mathcal{F}) , the expected cardinality $|C|((m, \mathcal{F}))$ is $|C|((m, \mathcal{F})) = \sum_{E \in \mathcal{F}} |E| m(E)$

Algorithm 1: Extraction from \mathcal{P} of a possibility distribution π such that $\mathcal{P} \subseteq \mathcal{P}_\pi$

Input: Credal Set \mathcal{P} on \mathcal{X} with $|\mathcal{X}| = n$

Output: Possibility distribution π such that $\mathcal{P} \subseteq \mathcal{P}_\pi$

Take an (arbitrary) ranking $\{x_1, \dots, x_n\}$ of the n elements x of \mathcal{X}

for $i = 1, \dots, n$ **do**

 | Build sets $A_i = \{x_1, \dots, x_i\}$ forming a nested collection (i.e., $A_1 \subset \dots \subset A_n = \mathcal{X}$)

for $i = 1, \dots, n$ **do**

 | compute $\bar{P}(A_i)$

for $i = 1, \dots, n$ **do**

 | take $\pi(x_i) = \bar{P}(A_i)$

Practical aspects $|\mathcal{X}| - 1$ values are needed to fully assess a possibility distribution, which makes it the simplest numerical uncertainty representation explicitly coping with imprecise or incomplete knowledge. This simplicity makes this representation very easy to handle. This also implies less expressive power, in the sense that, for any event A , either $\Pi(A) = 1$ or $N(A) = 0$ (i.e. intervals $[N(A), \Pi(A)]$ are of the form $[0, \alpha]$ or $[\beta, 1]$). This means that, in several situations, possibility distributions will be insufficient to exactly reflect the available information.

Nevertheless, the expressive power of possibility distributions fits various practical situations. Indeed, they can be interpreted as a set of nested sets with different confidence degrees (the bigger the set, the highest the confidence degree). Moreover, a recent psychological study [170] shows that possibility distributions are convenient in elicitation procedures. On the real line [78], possibility distributions can model, for example, an expert opinion concerning the value of a badly known parameter by means of a finite collection of nested confidence intervals. Similarly, it is natural to view nested confidence intervals coming from statistics as a possibility distribution. Another practical case where uncertainty can be modeled by possibility distributions is the case of vague linguistic assessments concerning probabilities [45].

3.1.3 Sketching a first summary of relationships

Now that we have reviewed the main numerical simplified representations of uncertainty, it is time to sketch a first drawing of the relationships between them. In the next sections, we will complete this first summary, to finish with a figure encompassing generalised p-boxes and clouds.

To sketch this first summary, we must first say a word about how imprecise probability as-

signments and p-boxes relate to other models, that is random sets and possibility distributions.

3.1.3.1 P-boxes in the landscape of uncertainty representations

There is no direct relationship between p-boxes and possibility distributions (in the sense that none can be seen as a special case of the other). Baudrit and Dubois [11] study in detail the relation between the credal set \mathcal{P}_π induced by a possibility distribution π and the credal set $\mathcal{P}_{[\underline{F}, \overline{F}]_\pi}$ induced by the p-box $[\underline{F}, \overline{F}]_\pi$ extracted from \mathcal{P}_π .

Kriegler and Held [132] have recently shown that any p-box $[\underline{F}, \overline{F}]$ can be represented by an equivalent random set $(m, \mathcal{F})_{[\underline{F}, \overline{F}]}$, and provide an efficient algorithm to build such a random set. However, as noticed by Kriegler and Held [132] (and before that by Ferson *et al.* [103]), different random sets can induce the same p-box (i.e., random sets whose associated credal sets have the same projections on events of the type $(-\infty, r]$). This means that p-boxes are special cases of random sets.

3.1.3.2 Imprecise probability assignments in the landscape of uncertainty representations

There is no direct relationship between imprecise probability assignments, random sets and possibility distributions. Indeed, upper and lower probabilities induced by tight imprecise probability assignments are only ensured to be order 2 Choquet capacities (i.e., they are not necessarily order 3 Choquet capacities, although some of them are), while belief functions and necessity measures are ∞ -monotone capacities. In general, one can only approximate one representation by the other.

Transforming an imprecise probability assignments L into a possibility distribution π_L such that $\mathcal{P}_L \subseteq \mathcal{P}_{\pi_L}$ or, conversely, transforming a possibility distribution π into an imprecise probability assignments L_π such that $\mathcal{P}_\pi \subseteq \mathcal{P}_{L_\pi}$ can easily be done through the methods respectively described in Section 3.1.2.5 (Algorithm 1) and Section 3.1.2.3.

Similarly, it is simple to transform a random set (m, \mathcal{F}) into the tightest set $L_{(m, \mathcal{F})}$ of imprecise probability assignments such that $\mathcal{P}_{(m, \mathcal{F})} \subseteq \mathcal{P}_{L_{(m, \mathcal{F})}}$ (i.e. $L_{(m, \mathcal{F})}$ is an outer approximation of the random set). The method consists of taking for all $x \in \mathcal{X}$:

$$l(x) = Bel(x) \text{ and } u(x) = Pl(x) \quad (3.19)$$

and since belief and plausibility functions are the lower envelope of the induced credal set $\mathcal{P}_{(m, \mathcal{F})}$, we are sure that the so-built imprecise probability assignment L is tight.

The converse problem, i.e. to transform a set L of imprecise probability assignments into a random set is studied by Lemmer and Kyburg [135]. They concentrate on transforming the set L into a random set (m, \mathcal{F}) inner approximating L (i.e., $\mathcal{P}_{(m, \mathcal{F})} \subseteq \mathcal{P}_L$). On the contrary, Denoeux [60] extensively studies the problem of transforming a set L of probability intervals into a random set $(\overline{m}, \overline{\mathcal{F}})$ that is an outer approximation (i.e., $\mathcal{P}_L \subseteq \mathcal{P}_{(\overline{m}, \overline{\mathcal{F}})}$), providing efficient methods to achieve such a transformation.

3.1.3.3 Preliminary summary

The main relations existing between imprecise probabilities, lower/upper probabilities, random sets, imprecise probability assignments, p-boxes and possibility distributions, are pictured on Figure 3.1. From top to bottom, it goes from the more general, expressive and complex theories to the less general, less expressive but simpler representations. An arrow is directed from a general representation to a less general one.

To make the picture more complete, we add sets and single elements to it. A set \mathcal{X} represents the fact that all we know is that variable X will take its value in \mathcal{X} , and nothing more. In other word, except for the fact that it will be in the set \mathcal{X} , we are totally ignorant about which values are more likely to occur than others. Such a state of ignorance can be modeled by possibility distributions ($\pi(x) = 1$ if $x \in \mathcal{X}$, zero otherwise) and imprecise probability assignments ($l(x) = 0, u(x) = 1$ if $x \in \mathcal{X}$, $u(x) = l(x) = 0$ otherwise), and therefore by all the other uncertainty representations mentioned above, except for single probability distributions. More generally, it is modeled by the lower capacity that takes value 1 on \mathcal{X} , and zero on all other events. A single value x models a state of complete certainty, since we're sure that $X = x$. It is equivalent to a set reduced to this single value ($\mathcal{X} = \{x\}$), and can be modeled by the Dirac probability distribution $P(\{x\}) = p(x) = 1$.

3.2 Introduction and study of generalised p-boxes

As recalled in Section 3.1.2.2, p-boxes are useful and practical representations of uncertainty used in many applications. So far, they only make sense on the (discretized) real line and their definition requires the natural ordering of numbers. This is a bit restrictive, and since the model is already quite useful in this restrictive setting, extending the model to more general

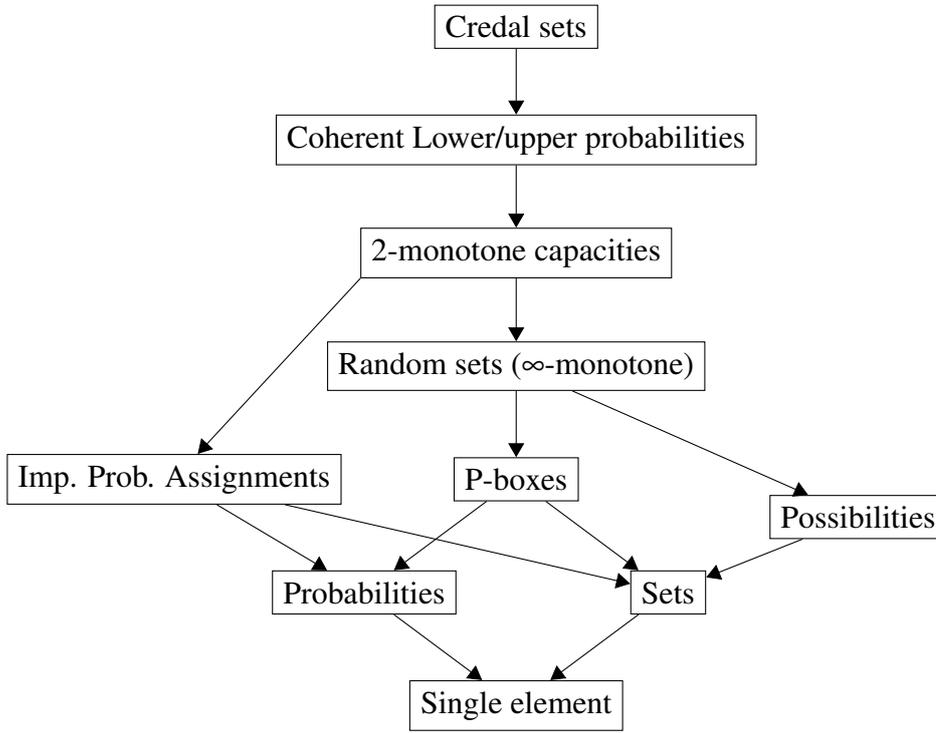


Figure 3.1: Representation relationships: summary $A \rightarrow B$: B is a special case of A

settings is potentially interesting. Moreover, as we will see, such extensions can give a better understanding of characteristic proper to such representations (e.g. the use of an implicit order on \mathcal{X}).

In this section, we study such an extension to arbitrary and completely pre-ordered finite spaces. We first define the extension of p-boxes to such spaces, before exploring its links with possibility distributions, random sets and imprecise probability assignments.

3.2.1 Definition of generalized p-boxes

First recall that two mappings f and f' from a finite ranked-set $\mathcal{X} = \{x_1, \dots, x_n\}$ to the real line \mathbb{R} are said to be *comonotonic* if there is a common permutation σ of $\{1, 2, \dots, n\}$ such that $f(x_{\sigma(1)}) \geq f(x_{\sigma(2)}) \geq \dots \geq f(x_{\sigma(n)})$ and $f'(x_{\sigma(1)}) \geq f'(x_{\sigma(2)}) \geq \dots \geq f'(x_{\sigma(n)})$. In other words, f and f' are comonotonic if and only if for any pair of elements $x, y \in \mathcal{X}$, we have $f(x) < f(y) \Rightarrow f'(x) \leq f'(y)$. We define a generalized p-box as follows:

Definition 3.5 (Generalized p-box). A generalized p-box $[\underline{F}, \overline{F}]$ over a finite space \mathcal{X} is a pair of comonotonic mappings $\underline{F}, \overline{F}, \underline{F} : \mathcal{X} \rightarrow [0, 1]$ and $\overline{F} : \mathcal{X} \rightarrow [0, 1]$ from \mathcal{X} to $[0, 1]$ such that

\underline{F} is pointwise lower than \overline{F} (i.e. $\underline{F} \leq \overline{F}$) and there is at least one element x in \mathcal{X} for which $\overline{F}(x) = \underline{F}(x) = 1$.

Since each distribution $\underline{F}, \overline{F}$ is fully specified by $|\mathcal{X}| - 1$ values, it follows that $2|\mathcal{X}| - 2$ values completely determine a generalized p-box. Note that, given a generalized p-box $[\underline{F}, \overline{F}]$, we can always define a *complete* pre-ordering $\leq_{[\underline{F}, \overline{F}]}$ on \mathcal{X} such that $x \leq_{[\underline{F}, \overline{F}]} y$ if $\underline{F}(x) \leq \underline{F}(y)$ and $\overline{F}(x) \leq \overline{F}(y)$, due to the comonotonicity condition. If \mathcal{X} is the (discretized) real line and if $\leq_{[\underline{F}, \overline{F}]}$ is the natural ordering of numbers, then we retrieve usual p-boxes, showing that Definition 3.5 is indeed a generalization of the usual notion of p-box. Potential useful cases encompassed by this generalization are multidimensional (discrete) models defined on \mathbb{R}^d with $d > 1$ (provided an appropriate pre-ordering on elements of \mathbb{R}^d is given).

To simplify notations, we will consider that, given a generalized p-box $[\underline{F}, \overline{F}]$, elements x of \mathcal{X} are indexed such that $i < j$ implies that $x_i \leq_{[\underline{F}, \overline{F}]} x_j$, and that $|\mathcal{X}| = n$. A $[\underline{F}, \overline{F}]$ -downset, denoted $(x]_{[\underline{F}, \overline{F}]}$, will be of the form $\{x_i \in \mathcal{X} | x_i \leq_{[\underline{F}, \overline{F}]} x\}$.

The credal set induced by a generalized p-box $[\underline{F}, \overline{F}]$ can now be defined as

$$\mathcal{P}_{[\underline{F}, \overline{F}]} = \{P \in \mathbb{P}_{\mathcal{X}} | i = 1, \dots, n, \underline{F}(x_i) \leq P((x_i]_{[\underline{F}, \overline{F}]}) \leq \overline{F}(x_i)\}.$$

It induces coherent upper and lower probabilities such that $\underline{F}(x_i) = \underline{P}((x_i]_{[\underline{F}, \overline{F}]})$ and $\overline{F}(x_i) = \overline{P}((x_i]_{[\underline{F}, \overline{F}]})$. Again, if we consider real numbers \mathbb{R} and the natural ordering on them, then $\forall r \in \mathbb{R}, (r]_{[\underline{F}, \overline{F}]} = (-\infty, r]$, and the above equation coincides with Equation (3.6).

Let us denote by A_i the sets $(x_i]_{[\underline{F}, \overline{F}]}$, for all $i = 1, \dots, n$. These sets are nested, since $\emptyset \subset A_1 \subseteq \dots \subseteq A_n = \mathcal{X}^2$. For all $i = 1, \dots, n$, let $\underline{F}(x_i) = \alpha_i$ and $\overline{F}(x_i) = \beta_i$. With these conventions, the credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$ can now be described by the following constraints bearing on probabilities of nested sets A_i :

$$i = 1, \dots, n \quad \alpha_i \leq P(A_i) \leq \beta_i \tag{3.20}$$

with $0 = \alpha_0 \leq \alpha_1 \leq \dots \leq \alpha_n = 1, 0 = \beta_0 < \beta_1 \leq \beta_2 \leq \dots \leq \beta_n = 1$ and $\alpha_i \leq \beta_i$.

As a consequence, a generalized p-box can be generated in two different ways:

- Either we start from two comonotone functions $\overline{F}, \underline{F}$ on the space \mathcal{X} , and the order on \mathcal{X} is then induced by the values of $\overline{F}, \underline{F}$,

²Since there is a complete pre-order on \mathcal{X} , we can have $x_i =_{[\underline{F}, \overline{F}]} x_{i+1}$ and $A_i = A_{i+1}$, which explains the non-strict inclusions. They would be strict if $<_{[\underline{F}, \overline{F}]}$ were a linear order.

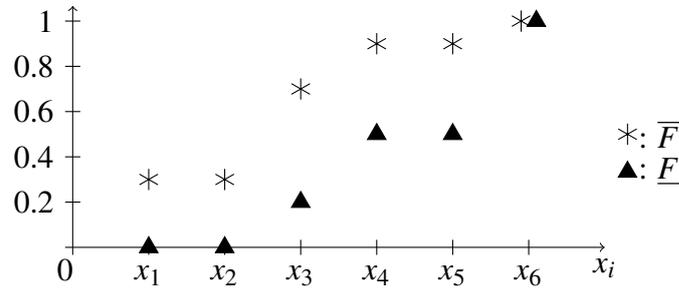


Figure 3.2: Generalized p-box $[F, \bar{F}]$ of Example 3.1

- or a generalized p-box is built by assigning upper and lower bounds on probabilities of nested sets, (i.e. sets A_i built or not from a complete ordering on \mathcal{X}), for instance intervals.

The second approach is likely to be more useful in practical assessments of generalized p-boxes.

Example 3.1. All along this section, we will use this example to illustrate results on generalized p-boxes. Consider a space \mathcal{X} made of six elements $\{x_1, \dots, x_6\}$. These elements could be, for instance, successive components on a production line. For various reasons (cost, production constraints, ...), when a component breaks down, the safety system only informs us whether the broken component is in the set $A_1 = \{x_1, x_2\}$, $A_2 = \{x_1, x_2, x_3\}$, $A_3 = \{x_1, \dots, x_5\}$, or the whole \mathcal{X} ($= A_4$). Asking an expert to evaluate the probability of breakdowns, he can only give us lower and upper probability bounds for each of these sets:

$$P(A_1) \in [0, 0.3] \quad P(A_2) \in [0.2, 0.7] \quad P(A_3) \in [0.5, 0.9]$$

Since these sets are nested, the uncertainty can be modeled by the generalized p-box pictured on Figure 3.2:

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|-----------|-------|-------|-------|-------|-------|-------|
| \bar{F} | 0.3 | 0.3 | 0.7 | 0.9 | 0.9 | 1 |
| F | 0 | 0 | 0.2 | 0.5 | 0.5 | 1. |

Outer approximation From Equations 3.20, we can easily derive a method that builds, from any credal set \mathcal{P} , a generalized p-box $[F, \bar{F}]$ inducing a credal set $\mathcal{P}_{[F, \bar{F}]}$ outer-approximating \mathcal{P} . This method is given by Algorithm 2. This algorithm shares many similarities with Algorithm 1: It depends on the chosen ranking of elements of \mathcal{X} , and provide the tightest gener-

alized p-boxes outer-approximating a given credal set, **given a specific ranking of elements**. Also note that, in the case of generalized p-boxes, there is only $|\mathcal{X}|!/2$ potential different results, since choosing a ranking and the reverse ranking result in the same Constraints (3.20), due to the duality between lower and upper probabilities.

Algorithm 2: Extracting from \mathcal{P} a generalized p-box $[\underline{F}, \overline{F}]$ such that $\mathcal{P} \subseteq \mathcal{P}_{[\underline{F}, \overline{F}]}$

Input: Credal Set \mathcal{P} on \mathcal{X} with $|\mathcal{X}| = n$

Output: Generalized p-box $[\underline{F}, \overline{F}]$ such that $\mathcal{P} \subseteq \mathcal{P}_{[\underline{F}, \overline{F}]}$

Take an (arbitrary) ranking $\{x_1, \dots, x_n\}$ of the n elements x of \mathcal{X}

for $i = 1, \dots, n$ **do**

\lfloor Build sets $A_i = \{x_1, \dots, x_i\}$ forming a nested collection (i.e., $A_1 \subset \dots \subset A_n = \mathcal{X}$)

for $i = 1, \dots, n$ **do**

\lfloor compute $\overline{P}(A_i)$ and $\underline{P}(A_i)$

for $i = 1, \dots, n$ **do**

\lfloor take $\overline{F}(x_i) = \overline{P}(A_i)$ and $\underline{F}(x_i) = \underline{P}(A_i)$

3.2.2 Connecting generalized p-boxes with possibility distributions

As mentioned earlier, Kozine and Utkin [130] discussed the problem of building p-boxes from partial information. They already noticed that, for usual p-boxes, sets A_i can be interpreted as nested confidence intervals with upper and lower bounds. It is therefore natural to search a connection with possibility theory, since possibility distributions can be interpreted as a collection of nested sets with lower probability bounds. Given a generalized p-box $[\underline{F}, \overline{F}]$, the following proposition holds:

Proposition 3.3. *Any generalized p-box $[\underline{F}, \overline{F}]$ on a finite set \mathcal{X} is representable (see Definition 3.4) by a pair of possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$, in the sense $\mathcal{P}_{[\underline{F}, \overline{F}]} = \mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$, where:*

$$\pi_{\overline{F}}(x_i) = \beta_i \quad \text{and} \quad \pi_{\underline{F}}(x_i) = 1 - \max \{ \alpha_j \mid j = 0, \dots, i, \alpha_j < \alpha_i \}$$

for $i = 1, \dots, n$, with $\alpha_0 = 0$.

Proof of Proposition 3.3. Consider the set of constraints given by Equation (3.20) and describing the convex set $\mathcal{P}_{[\underline{F}, \overline{F}]}$. These constraints can be split into two distinct sets: $(P(A_i) \leq \beta_i)_{i=1, n}$ and $(P(A_i^c) \leq 1 - \alpha_i)_{i=1, n}$. Now, rewrite constraints of Proposition 3.2, in the form $\forall \alpha \in (0, 1]: P \in \mathcal{P}_{\pi}$ if and only if $P(\{x \in \mathcal{X} \mid \pi(x) \leq \alpha\}) \leq \alpha$.

The first set of constraints $(P(A_i) \leq \beta_i)_{i=1,n}$ defines a credal set $\mathcal{P}_{\pi_{\bar{F}}}$ that is induced by the possibility distribution $\pi_{\bar{F}}$, while the second set of constraints $(P(A_i^c) \leq 1 - \alpha_i)_{i=1,n}$ defines a credal set $\mathcal{P}_{\pi_{\underline{F}}}$ that is induced by the possibility distribution $\pi_{\underline{F}}$, since $A_i^c = \{x_k, \dots, x_n\}$, where $k = \max\{j | \alpha_j < \alpha_i\}$. The credal set of the generalized p-box $[\underline{F}, \bar{F}]$, resulting from the merging of the two sets of constraints, namely $i = 1, \dots, n$, $\beta_i \leq P(A_i) \leq \alpha_i$, is thus $\mathcal{P}_{\pi_{\bar{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$. \square

If \underline{F} is injective, it induces a complete order $<_{[\underline{F}, \bar{F}]}$, and then $\pi_{\underline{F}}(x_i) = 1 - \alpha_{i-1}$.

Example 3.2. The possibility distributions $\pi_{\bar{F}}, \pi_{\underline{F}}$, for the generalized p-box defined in Example 3.1 are:

| | x_1 | x_2 | x_3 | x_4 | x_5 | x_6 |
|-----------------------|-------|-------|-------|-------|-------|-------|
| $\pi_{\bar{F}}$ | 0.3 | 0.3 | 0.7 | 0.9 | 0.9 | 1 |
| $\pi_{\underline{F}}$ | 1 | 1 | 1 | 0.8 | 0.8 | 0.5 |

So, generalized p-boxes allow to model uncertainty in terms of pairs of possibility distributions. In this case, contrary to the case of only one possibility distribution, the two bounds describing uncertainty on a particular event A can be tighter, i.e. no longer restricted to the form $[0, \alpha]$ or $[\beta, 1]$, since the corresponding interval containing $P(A)$ will be contained in the intersection of intervals of this form.

An interesting case is the one where, for all $i = 1, \dots, n-1$, $\underline{F}(x_i) = 0$ and $\bar{F}(x_n) = 1$. Then, $\pi_{\underline{F}} = 1$ and $\mathcal{P}_{\pi_{\bar{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}} = \mathcal{P}_{\pi_{\bar{F}}}$ and we retrieve the single distribution $\mathcal{P}_{\pi_{\bar{F}}}$. We recover $\mathcal{P}_{\pi_{\underline{F}}}$ if we take for all $i = 1, \dots, n$, $\bar{F}(x_i) = 1$. This means that any possibility distribution can be viewed as a generalized cumulative distribution function F (it can be understood either as an upper or a lower function of a generalized p-box) associated to the specific ordering the possibility degrees induce on \mathcal{X} , and therefore that possibility distributions are special cases of generalized p-boxes.

3.2.3 Connecting Generalized p-boxes and random sets

We mentioned before that p-boxes were special cases of random sets, in the sense that any p-boxes could be mapped into an equivalent random set. The next proposition shows that this result can be extended to generalized p-boxes:

Proposition 3.4. *Generalized p-boxes are special cases of random sets, in the sense that for any generalized p-box $[F, \bar{F}]$ defined on \mathcal{X} , there always exist a random set $(m, \mathcal{F})_{[F, \bar{F}]}$ such that $\mathcal{P}_{[F, \bar{F}]} = \mathcal{P}_{(m, \mathcal{F})_{[F, \bar{F}]}}$.*

Proof. See Section D.1 in Appendix D □

In order to prove Proposition 3.4, we show that the lower probabilities on events induced by a generalized p-box are the same as the belief function given by Algorithm 3. To do that, we first build the partition of the space \mathcal{X} induced by sets A_i , and we formulate lower probabilities on events by means of elements of this partition. We then calculate lower bounds of these lower probabilities on all events, and show that these bounds are reached. We then check that the lower probabilities on all events coincide with the belief function induced by the algorithm.

Algorithm 3 below provides an easy way to build the random set encoding a given generalized p-box. It is similar to existing algorithms [132, 171] for usual p-boxes, and extends them to more general spaces. The main idea of the algorithm is to use the fact that a generalized p-box can be seen as a random set whose focal sets are obtained by thresholding the cumulative distributions. Since the sets A_i are nested, they induce a partition of \mathcal{X} whose elements are of the form $G_i = A_i \setminus A_{i-1}$. The focal sets of the random set equivalent to the generalized p-box are made of unions of consecutive elements of this partition. Basically, the procedure comes down to considering a threshold $\theta \in [0, 1]$. When $\alpha_{i+1} > \theta \geq \alpha_i$ and $\beta_{j+1} > \theta \geq \beta_j$, then, the corresponding focal set is $A_{i+1} \setminus A_j$, with mass

$$m(A_{i+1} \setminus A_j) = \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j). \quad (3.21)$$

Example 3.3 illustrates the application of Algorithm 3.

Example 3.3. Consider again the generalized p-box given in Example 3.1 and let us build the associated random set by applying Algorithm 3. We have:

$$G_1 = \{x_1, x_2\} \quad G_2 = \{x_3\} \quad G_3 = \{x_4, x_5\} \quad G_4 = \{x_6\}$$

Algorithm 3: R-P-box \rightarrow random set transformation

Input: Generalized p-box $[E, \overline{F}]$ and corresponding nested sets $\emptyset = A_0, A_1, \dots, A_n = \mathcal{X}$, lower bounds α_i and upper bounds β_i

Output: Equivalent random set

for $i = 1, \dots, n$ **do**

\lfloor Build partition $G_i = A_i \setminus A_{i-1}$

Build set of values $\{\gamma_l | l = 1, \dots, 2n-1\} = \{\alpha_i | i = 1, \dots, n\} \cup \{\beta_i | i = 1, \dots, n\}$

With γ_l indexed such that $\gamma_1 \leq \dots \leq \gamma_l \leq \dots \leq \gamma_{2n-1} = \beta_n = \alpha_n = 1$

Set $\alpha_0 = \beta_0 = \gamma_0 = 0$

Set focal set $E_0 = \emptyset$

for $k = 1, \dots, 2n-1$ **do**

if $\gamma_{k-1} = \alpha_i$ **then**

$\lfloor E_k = E_{k-1} \cup G_{i+1}$

if $\gamma_{k-1} = \beta_i$ **then**

$\lfloor E_k = E_{k-1} \setminus G_i$

 Set $m(E_k) = \gamma_k - \gamma_{k-1}$

and

$$0 \leq 0 \leq 0.2 \leq 0.3 \leq 0.5 \leq 0.7 \leq 0.9 \leq 1$$

$$\alpha_0 \leq \alpha_1 \leq \alpha_2 \leq \beta_1 \leq \alpha_3 \leq \beta_2 \leq \beta_3 \leq \alpha_4$$

$$\gamma_0 \leq \gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4 \leq \gamma_5 \leq \gamma_6 \leq \gamma_7$$

which finally yields the following random set

$$m(E_1) = m(G_1) = 0 \quad m(E_2) = m(G_1 \cup G_2) = 0.2$$

$$m(E_3) = m(G_1 \cup G_2 \cup G_3) = 0.1 \quad m(E_4) = m(G_2 \cup G_3) = 0.2$$

$$m(E_5) = m(G_2 \cup G_3 \cup G_4) = 0.2 \quad m(E_6) = m(G_3 \cup G_4) = 0.2$$

$$m(E_7) = m(G_4) = 0.1$$

This random set can then be used as an alternative representation of the provided information. This representation lays bare the high imprecision of the information. This imprecision can only be alleviated by seeking more information.

Proposition 3.4 shows that generalized p-boxes are special cases of general random sets. Generalized p-boxes are thus more expressive than single possibility distributions and less

expressive than random sets, but, as emphasized in the introduction, less expressive (and, in this sense, simpler) models are often easier to handle in practice. As shown by the following remark, we can expect it to be the case for generalized p-boxes.

Remark 3.2. Let $[\underline{F}, \overline{F}]$ be a generalized p-box over \mathcal{X} , and G_i be the elements of the partition induced by nested subsets A_i , for $i = 1, \dots, n$. Let us call a subset E of \mathcal{X} *full* if it can be expressed as an union of consecutive elements G_k , i.e. $E = \bigcup_{k=i}^j G_k$, with $0 < i < j \leq n$. Then, we have an explicit expression for the induced lower probability of any full subset E :

$$\underline{P}(E) = \max(0, \alpha_j - \beta_{i-1}). \quad (3.22)$$

Now, for any event A , let $A_* = \bigcup_{E \subseteq A} E$ be the lower approximation of A by union of elements of the partition, with E all maximal *full* subsets included in A . We know that $\underline{P}(A) = \underline{P}(A_*)$. Then, the explicit expression for $\underline{P}(A)$ is

$$\underline{P}(A_*) = \sum_{E \subseteq A} \underline{P}(E),$$

which remains simple to compute and just becomes a sum of lower probabilities of those subsets formed of unions of consecutive G_k included in A .

This simple remark shows the potential advantages of using generalized p-boxes rather than general random sets, since the computation of lower probabilities is faster than checking which focal elements E_i are included in a given event A . Other computational aspects of generalized p-boxes related to other problems will be studied in subsequent chapters.

So far, results in this section mainly exploit the fact that a collection of nested subsets on a space \mathcal{X} induces a partition on this space, useful when computing lower probabilities of events. In the following we explain the links between this partition and the complete pre-ordering $\leq_{[\underline{F}, \overline{F}]}$ as well as the two possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$. First notice that Equation (3.22) can be restated in terms of the two possibility distributions $\pi_{\overline{F}}, \pi_{\underline{F}}$, rewriting $\underline{P}(A_*)$ as

$$\underline{P}(A_*) = \max(0, N_{\pi_{\underline{F}}}(\bigcup_{k=1}^j G_k) - \Pi_{\pi_{\overline{F}}}(\bigcup_{k=1}^{i-1} G_k)),$$

where $N_{\pi_i}(A), \Pi_{\pi_i}(A)$ are respectively the necessity and possibility degree of event A (given by Equations (3.13)) with respect to π_i . It makes $\underline{P}(A_*)$ even easier to compute.

We can also directly derive the random set equivalent to a given generalized p-box $[\underline{F}, \overline{F}]$: let us note $0 = \gamma_0 < \gamma_1 < \dots < \gamma_M = 1$ the distinct values taken by $\underline{F}, \overline{F}$ over elements x_i of \mathcal{X}

(note that M is finite and $M < 2n$). Then, for $j = 1, \dots, M$, the random set defined as:

$$\begin{cases} E_j = \{x_i \in \mathcal{X} \mid (\pi_{\bar{F}}(x_i) \geq \gamma_j) \wedge (1 - \pi_{\underline{F}}(x_i) < \gamma_j)\} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (3.23)$$

is the same as the one built by using Algorithm 3, but this formulation lays bare the link between Equation (3.17) and the possibility distributions $\pi_{\bar{F}}, \pi_{\underline{F}}$.

3.2.4 Generalized p-boxes and imprecise probability assignments

As in the case of random sets, there is no direct relationship between imprecise probability assignments and generalized p-boxes, in the sense that none of them generalizes the other. The two representations have comparable complexities, but do not involve the same kind of events (singletons for imprecise probability assignments, and nested collection of sets for generalized p-boxes). Nevertheless, given previous results, we can state how a set L of imprecise probability assignments can be approximated into a generalized p-box $[\underline{F}, \bar{F}]$, and vice-versa. We can also study more complex links between the two.

3.2.4.1 Approximations between the two representations

Let us first consider a set L of imprecise probability assignments on a space \mathcal{X} and some indexing of elements in \mathcal{X} . For all $i = 1, \dots, n$, let $l(x_i) = l_i$ and $u(x_i) = u_i$. A generalized p-box $[\underline{F}, \bar{F}]'$ outer-approximating the set L of imprecise probability assignments can be computed by means of Equations (3.10) of Section 3.1.2.3 in the following way:

$$\begin{aligned} \underline{F}'(x_i) = \underline{P}(A_i) = \alpha'_i &= \max\left(\sum_{x_i \in A_i} l_i, 1 - \sum_{x_i \notin A_i} u_i\right) \\ \bar{F}'(x_i) = \bar{P}(A_i) = \beta'_i &= \min\left(\sum_{x_i \in A_i} u_i, 1 - \sum_{x_i \notin A_i} l_i\right) \end{aligned} \quad (3.24)$$

where \underline{P}, \bar{P} are respectively the lower and upper probabilities of \mathcal{P}_L for events A_i , given by Equations (3.10). Each permutation of elements of \mathcal{X} would provide a different generalized p-box. There is no tightest outer approximation among them, although Equations (3.24) do give the tightest generalized p-box for a given permutation. Note that Equations (3.24) correspond to the application of Algorithm 2 for the specific case of imprecise probability assignments.

Now we consider a generalized p-box $[\underline{F}, \bar{F}]$ with nested sets $A_1 \subseteq \dots \subseteq A_n$. The set L' of

probability intervals on elements x_i outer-approximating $[\underline{F}, \overline{F}]$ is given by:

$$\begin{aligned} \underline{P}(x_i) &= l'_i = \max(0, \alpha_i - \beta_{i-1}) \\ \overline{P}(x_i) &= u'_i = \beta_i - \alpha_{i-1} \end{aligned} \quad (3.25)$$

where $\underline{P}, \overline{P}$ are the lower and upper probabilities of $\mathcal{P}_{[\underline{F}, \overline{F}]}$, given by Equation (3.22), with $\beta_0 = \alpha_0 = 0$. This is the tightest set of imprecise probability assignments induced by the generalized p-box and outer-approximating it.

Of course, transforming a set L of imprecise probability assignments into a generalized p-box $[\underline{F}, \overline{F}]$ and vice-versa generally induces a loss of information, as already noticed in Sections 3.1.2.3 and 3.2.1 for the general problem of finding an outer approximation in term of generalized p-boxes or of imprecise probability assignments. The two following propositions quantify this loss.

Proposition 3.5. *Given an initial set L of imprecise probability assignments over a space \mathcal{X} , and given the two consecutive transformations*

$$\text{Imp. prob. ass. } L \xrightarrow{(3.24)} \text{p-box } [\underline{F}, \overline{F}]' \xrightarrow{(3.25)} \text{Imp. prob. ass. } L'' \quad (3.26)$$

we have $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$, and the differences between bounds of intervals in the sets L'' and L are given, for $i = 1, \dots, n$, by

$$\begin{aligned} l_i - l''_i &= \min(l_i, 0 + \sum_{x_i \in A_{i-1}} (u_i - l_i), 0 + \sum_{x_i \in A_i^c} (u_i - l_i), (l_i + \sum_{\substack{x_j \neq x_i \\ x_j \in \mathcal{X}}} u_j) - 1, 1 - \sum_{x_i \in \mathcal{X}} l_i) \\ u''_i - u_i &= \min(0 + \sum_{x_i \in A_{i-1}} (u_i - l_i), 0 + \sum_{x_i \in A_i^c} (u_i - l_i), 1 - (u_i + \sum_{\substack{x_j \neq x_i \\ x_j \in \mathcal{X}}} l_j), \sum_{x_i \in \mathcal{X}} u_i - 1) \end{aligned} \quad (3.27)$$

with $A_0 = \emptyset$. Under the assumptions that set L is consistent and tight, these differences are positive.

Proof. See Section D.1 in Appendix D □

Proposition 3.6. *Given an initial generalized p-box $[\underline{F}, \overline{F}]$ over a space \mathcal{X} , and given the two consecutive transformations*

$$\text{p-box } [\underline{F}, \overline{F}] \xrightarrow{(3.25)} \text{Imp. prob. ass. } L' \xrightarrow{(3.24)} \text{p-box } [\underline{F}, \overline{F}]'' \quad (3.28)$$

we have that $\mathcal{P}_{[\underline{F}, \overline{F}]} \subseteq \mathcal{P}_{[\underline{F}, \overline{F}]''}$, and the differences between values of $[\underline{F}, \overline{F}]$ and $[\underline{F}, \overline{F}]''$ are,

for $i = 1, \dots, n$

$$\begin{aligned} \underline{F}(x_i) - \underline{F}''(x_i) &= \min\left(\sum_{j=1}^{i-1} (\alpha_j - \beta_j), \sum_{j=i+1}^{n-1} (\alpha_j - \beta_j)\right) \\ \overline{F}''(x_i) - \overline{F}(x_i) &= \min\left(\sum_{j=1}^{i-1} (\alpha_j - \beta_j), \sum_{j=i+1}^{n-1} (\alpha_j - \beta_j)\right) \end{aligned} \quad (3.29)$$

Proof. See Section D.1 in Appendix D □

Example 3.4 illustrates both the transformation procedure and the fact that this procedure implies an information loss.

Example 3.4. Let us take the same four imprecise probability assignments as in the example given by Masson and Denoeux [141], on the space $\mathcal{X} = \{w, x, y, z\}$, and summarized in the following table

| | w | x | y | z |
|-----|------|------|------|------|
| l | 0.10 | 0.34 | 0.25 | 0 |
| u | 0.28 | 0.56 | 0.46 | 0.08 |

we then consider the order R such that $w <_R x <_R y <_R z$. After application of Equations (3.24), we obtain the following generalized p-box

| | \underline{F}' | \overline{F}' |
|---------------------|------------------|-----------------|
| $A_1 = \{w\}$ | 0.10 | 0.28 |
| $A_2 = \{w, x\}$ | 0.46 | 0.75 |
| $A_3 = \{w, x, y\}$ | 0.92 | 1 |
| $A_4 = \mathcal{X}$ | 1 | 1 |

and if we generate imprecise probability assignments from this generalized p-box by applying Equations (3.25), we obtain the set L''

| | w | x | y | z |
|-------|------|------|------|------|
| l'' | 0.10 | 0.18 | 0.17 | 0 |
| u'' | 0.28 | 0.65 | 0.54 | 0.08 |

a result which is less informative than the first imprecise probability assignments.

When no natural order exists on \mathcal{X} , as many as $|\mathcal{X}|!/2$ different generalized p-boxes can be generated from an imprecise probability assignment L . As in the case of Algorithm 1, the problem of picking a ranking on elements x of \mathcal{X} such that the generalized p-box generated L is one of the most precise (in the sense of expected cardinality or of other information measures [128]) is not trivial, and remains an open problem.

3.2.4.2 Linking the two representations

We're now going to relate an imprecise probability assignment with the $|\mathcal{X}|!$ generalized p-boxes that can be generated from it. Let Σ_σ the set of all possible permutations σ of elements of \mathcal{X} , each defining a linear order. A generalized p-box according to permutation σ is denoted $[\underline{F}, \overline{F}]'_\sigma$ and called a σ -p-box. We then have the following proposition:

Proposition 3.7. *Let L be a set of imprecise probability assignments, and let $[\underline{F}, \overline{F}]'_\sigma$ be the σ -p-box obtained from L by applying Equations (3.24). Moreover, let L''_σ denote the set of imprecise probability assignments obtained from the σ -p-box $[\underline{F}, \overline{F}]'_\sigma$ by applying Equations (3.25). Then, the various credal sets thus defined satisfy the following property:*

$$\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{[\underline{F}, \overline{F}]'_\sigma} = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma} \quad (3.30)$$

Proof. See Section D.1 in Appendix D □

Concretely, Proposition 3.7 means that, given initial information modeled by a set L of imprecise probability assignments, this information can be entirely recovered by considering the set of all σ -p-boxes, varying all permutations. In other words, imprecise probability assignments are representable by generalized p-boxes. Since there are $|\mathcal{X}|!$ such permutations, representing a set of imprecise probability assignments L by a set of generalized p-boxes does not look very interesting at first glance. In practice, L can be exactly recovered if a reduced set \mathcal{S} of $|\mathcal{X}|/2$ permutations is used to generate the generalized p-boxes, provided that $\{x_{\sigma(1)}, \sigma \in \mathcal{S}\} \cup \{x_{\sigma(n)}, \sigma \in \mathcal{S}\} = \mathcal{X}$ in , that is the set of first and last elements of permutations in \mathcal{S} covers the whole domain \mathcal{X} . Since $\mathcal{P}_{[\underline{F}, \overline{F}]} = \mathcal{P}_{\pi_{\overline{F}}} \cap \mathcal{P}_{\pi_{\underline{F}}}$ (Proposition 3.3), the following corollary directly follows from Proposition 3.7:

Corollary 3.8. *Let L be a set of imprecise probability assignments, and let $[\underline{F}, \overline{F}]'_\sigma$ be the σ -p-box obtained from L by applying Equations (3.24), then we have:*

$$\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \left(\mathcal{P}_{\pi_{\underline{F}_\sigma}} \cap \mathcal{P}_{\pi_{\overline{F}_\sigma}} \right),$$

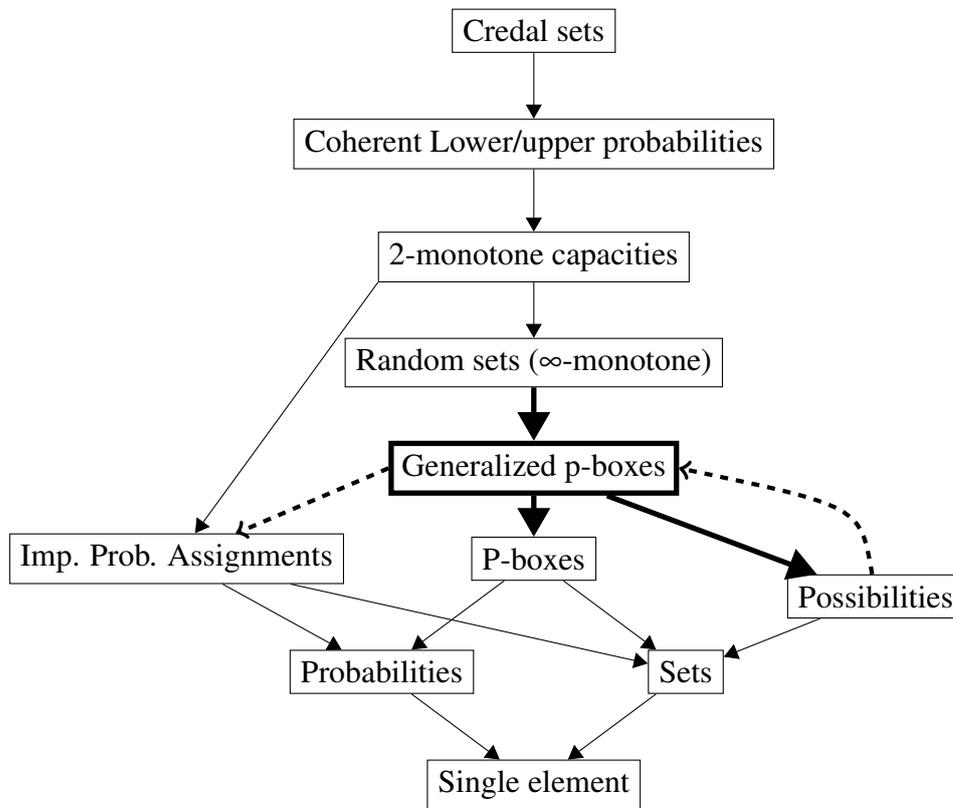


Figure 3.3: Representation relationships: summary with generalized p-boxes $A \longrightarrow B$: B is a special case of A . $A \dashrightarrow B$: B is representable by A .

where $\mathcal{P}_{\pi_{E_\sigma}}, \mathcal{P}_{\pi_{\bar{F}_\sigma}}$ are respectively the possibility distributions corresponding to \underline{F}_σ and \bar{F}_σ (by Proposition 3.3).

And this corollary allows us to model the information of an imprecise probability assignment by means of a set of possibility distributions. Also note that, since generalized p-boxes are special cases of random sets, Proposition 3.7 also provides a means to characterize imprecise probability assignments by a set of random sets (although perhaps not in an optimal way, in the sense that there could be other sets of random sets characterizing imprecise probability assignments, with a smaller cardinality).

Figure 3.3 summarizes the results obtained for generalized p-boxes by introducing them into the first summary given by Figure 3.1. New results related to the study of generalized p-boxes and to our findings are in bold.

3.2.5 Computing with generalized p-boxes: first results on propagation

In this section, we provide some first results investigating computational aspects of generalized p-boxes. In particular, we study the problem of propagating a p-box defined on \mathcal{X} through a model (here a function) $T : \mathcal{X} \rightarrow \mathcal{Y}$ with X an input variable and $Y = T(X)$ the output variable. First recall that a generalized p-box $[\underline{F}, \overline{F}]$ can be represented in three different ways:

- By a set of n constraints (3.20) such that

$$i = 1, \dots, n \quad \alpha_i \leq P(A_i) \leq \beta_i$$

with $0 = \alpha_0 \leq \alpha_1 \leq \dots \leq \alpha_n = 1$, $0 = \beta_0 < \beta_1 \leq \beta_2 \leq \dots \leq \beta_n = 1$ and $\alpha_i \leq \beta_i$.

- By an equivalent random set that can be built by considering a threshold $\theta \in [0, 1]$ and such that when $\alpha_{i+1} > \theta \geq \alpha_i$ and $\beta_{j+1} > \theta \geq \beta_j$ the corresponding focal set is $A_{i+1} \setminus A_j$, with weight

$$m(A_{i+1} \setminus A_j) = \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j).$$

- By two possibility distributions $\pi_{\underline{F}}, \pi_{\overline{F}}$ respectively modeled by random sets $(m, \mathcal{F})_{\pi_{\underline{F}}}, (m, \mathcal{F})_{\pi_{\overline{F}}}$ such that, for $i = 0, \dots, n-1$,

$$m_{\pi_{\underline{F}}}(A_{i+1}) = \alpha_{i+1} - \alpha_i \quad \text{and} \quad m_{\pi_{\overline{F}}}(A_{i+1}^c) = \beta_{i+1} - \beta_i$$

and each of these three different representations suggests a different propagation technique.

The first solution, propagating the nested sets and their confidence bounds consists of computing for each set A_i the propagated sets $T(A_i)$, and to consider the generalized p-box induced by the constraints:

$$\forall i = 1, \dots, n, \quad \alpha_i \leq P(T(A_i)) \leq \beta_i \quad (3.31)$$

where α_i, β_i are the confidence bounds originally related to the set A_i . Given this propagated generalized p-box (it is still a generalized p-box, since sets $T(A_i)$ are also nested), we can

build the counterpart of the random set given by equation (3.21), which is here:

$$\left. \begin{array}{l} \theta \in [0, 1] \\ \alpha_{i+1} > \theta \geq \alpha_i \\ \beta_{j+1} > \theta \geq \beta_j \end{array} \right\} \begin{array}{l} m(T(A_{i+1}) \setminus T(A_j)) = \\ \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j) \end{array}$$

and we note this random set $(m, \mathcal{F})_{T([E, \bar{F}])}$.

The second solution, directly propagating the focal elements of the random set given by equation (3.21), gives the following random set:

$$\left. \begin{array}{l} \theta \in [0, 1] \\ \alpha_{i+1} > \theta \geq \alpha_i \\ \beta_{j+1} > \theta \geq \beta_j \end{array} \right\} \begin{array}{l} m(T(A_{i+1} \setminus A_j)) = \\ \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j) \end{array}$$

that is potentially different from the one given by the first propagation. We note this second random set $(m, \mathcal{F})_{T((m, \mathcal{F}))}$.

The third solution consists of propagating both possibility distributions by the so-called extension principle [80]. This is equivalent to propagate the respective focal elements of each distribution through T , which gives us the random sets $(m, \mathcal{F})_{T(\pi_E)}$ and $(m, \mathcal{F})_{T(\pi_{\bar{F}})}$ respectively having, for $i = 0, \dots, n-1$, the following masses and focal elements

$$m_{T(\pi_E)}(T(A_i^c)) = \beta_{i+1} - \beta_i \text{ and } m_{T(\pi_{\bar{F}})}(T(A_{i+1})) = \alpha_{i+1} - \alpha_i$$

and, if we take from these two random sets the counterpart of the random set given by equation (3.21), we end up with the following random set:

$$\left. \begin{array}{l} \theta \in [0, 1] \\ \alpha_{i+1} > \theta \geq \alpha_i \\ \beta_{j+1} > \theta \geq \beta_j \end{array} \right\} \begin{array}{l} m(T(A_{i+1}) \setminus T(A_j^c)^c) = \\ \min(\alpha_{i+1}, \beta_{j+1}) - \max(\alpha_i, \beta_j) \end{array}$$

that we note $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$.

We can already note that the three random sets $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$, $(m, \mathcal{F})_{T((m, \mathcal{F}))}$, $(m, \mathcal{F})_{T([E, \bar{F}])}$ have the same bpa and that only focal elements differ. To compare the results of the three propagations, we thus have to compare the informative content of their respective focal elements.

In this perspective, the following proposition is helpful:

Proposition 3.9. *Let A and B be two subsets of a space \mathcal{X} such that $A \subset B$, and let f be a function from \mathcal{X} to another space \mathcal{Y} . Then, we have the following inclusion relations:*

$$f(B) \setminus f(A) \subseteq f(B \setminus A) \subseteq f(B) \setminus f(A^c)^c$$

and inclusion relationships become equalities if f is injective

Proof. We will first prove the first inclusion relationship, then the second one, each time showing that we have equality if f is injective.

Let us first prove that any element of $f(B) \setminus f(A)$ is in $f(B \setminus A)$. Let us consider an element y in $f(B) \setminus f(A)$. This implies:

$$\left. \begin{array}{l} y \in f(B) \\ y \notin f(A) \end{array} \right\} \Rightarrow \exists x \in \mathcal{X} \left\{ \begin{array}{l} f(x) \in f(B) \\ f(x) \notin f(A) \end{array} \right.$$

and this x is in B and not in A (i.e. in $B \setminus A$), which implies that $y = f(x)$ is in $f(B \setminus A)$. This means that $f(B) \setminus f(A) \subseteq f(B \setminus A)$, and we still have to show that this inclusion can be strict. To see it, consider the case where one of the element x in $B \setminus A$ is such that $f(x)$ takes the same value as $f(x')$, where x' is in A , thus this particular $f(x)$ is in $f(B \setminus A)$ and not in $f(B) \setminus f(A)$ (since by assumption it is in $f(A)$), showing that the inclusion can be strict. This case does not happen if f is injective (since if f is injective $f(x) = f(x')$ if and only if $x = x'$).

To prove the second inclusion relation, first note that $f(B \setminus A) = f(B \cap A^c)$ and that $(f(B) \setminus f(A^c)^c) = (f(B) \cap f(A^c))$. Known results immediately give $f(B \cap A^c) \subseteq f(B) \cap f(A^c)$. Strict inclusion happens in the case where we have an element x of X in B and in A , and another element x' not in A and not in B (i.e. x' is in A^c) for which $f(x) = f(x')$, thus we have that x and x' are not in $B \cap A^c$, but are respectively in B and A^c , and thus $f(x)$ is in $f(B) \cap f(A^c)$. Again, this case cannot happen when f is injective (since in this case, $x \neq x'$ implies $f(x) \neq f(x')$). \square

The above proposition tells us that, when f is not injective, we have in general

$$(m, \mathcal{F})_{T([E, \bar{F}])} \subseteq (m, \mathcal{F})_{T((m, \mathcal{F}))} \subseteq (m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$$

thus showing that $(m, \mathcal{F})_{T([E, \bar{F}])}$ is more optimistic than $(m, \mathcal{F})_{T((m, \mathcal{F}))}$, which is itself more optimistic than $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$. And in the case where T is injective, all these propagations are equivalent. However, restricting ourselves to injective functions can be very limiting. For

instance, if \mathcal{X} and \mathcal{Y} are subset of \mathbb{R} , requiring injectivity of T is equivalent to limiting ourselves to strictly monotone functions from \mathbb{R} to \mathbb{R} .

The question is then, if f is not injective, why should we choose one propagation rather than the other?

From a purely theoretical standpoint, computing $(m, \mathcal{F})_{T((m, \mathcal{F}))}$, the result of an exact propagation, is of course the best course of action. However, computing $(m, \mathcal{F})_{T((m, \mathcal{F}))}$ can be difficult, since a maximal number of $(n+1)n/2$ non-nested sets have to be propagated. $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$ appears more attractive from a computational standpoint, since it requires to propagate only $2n$ sets at most, whose nestedness can be used advantageously. Indeed, let $T : \mathbb{R}^N \rightarrow \mathbb{R}$ be a (non-linear) function from \mathbb{R}^N to the real line. Given the sets $A_0 \subset A_1 \subseteq \dots \subseteq A_n$, assume the global minimum and maximum of T are respectively in $A_i \setminus A_{i-1}$, and in $A_j \setminus A_{j-1}$, and their respective values are known. In the propagation, we no longer have to compute the lower bounds of all $T(A_k), T(A_l^c)$ such that $k > i > l$ nor the upper bounds of all $T(A_{k'}), T(A_{l'}^c)$ such that $k' > j > l'$. That is, by reusing function evaluations, we can avoid additional computations. However, $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$ only provides an outer approximation of $(m, \mathcal{F})_{T((m, \mathcal{F}))}$. $(m, \mathcal{F})_{T([E, \bar{F}])}$ provides an inner approximation of $(m, \mathcal{F})_{T((m, \mathcal{F}))}$ and is even easier to compute, since at most n nested sets have to be propagated to compute it. Nevertheless, $(m, \mathcal{F})_{T([E, \bar{F}])}$ can give a non-null weight to the empty set, and thus be incoherent.

Eventually, if faced with a practical problem, the best solution is to compute $(m, \mathcal{F})_{T((m, \mathcal{F}))}$ if possible. If not possible, computing $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$, yields $(m, \mathcal{F})_{T([E, \bar{F}])}$ for free (since for computing the former we need to propagate sets A_i). So, another solution is to bracket the information contained in $(m, \mathcal{F})_{T((m, \mathcal{F}))}$ using $(m, \mathcal{F})_{T(\pi_E, \pi_{\bar{F}})}$ and $(m, \mathcal{F})_{T([E, \bar{F}])}$. Computing $(m, \mathcal{F})_{T([E, \bar{F}])}$ only is not cautious.

The above results give us some first insights about how generalized p-boxes can be computationally handled. It also highlights the potential interests of the results relating generalized p-boxes with other uncertainty representations. As we shall see, those results related to p-boxes can also be used for particular instances of Neumaier's clouds [159].

3.3 Clouds

Clouds have been recently introduced and studied by Neumaier [159] as practical uncertainty models. He proposes clouds as a convenient tool to model and treat uncertainty in

high-dimensional problems where information is scarce and imprecise. In his original paper [159], Neumaier study very briefly the relationships with other models, simply mentioning that clouds seem to have, in general, poor relationships with other uncertainty representations (such as credal sets and random sets). We will show in this section that such a statement is debatable, since clouds do have strong links with previously studied representations. In particular, we will show that generalized p-boxes are equivalent to a specific sub-family of clouds.

We begin by recalling the definition of clouds, and undertake a study similar to the one achieved for generalized p-boxes.

3.3.1 Definition of clouds

Definition 3.6 (Cloud). A cloud is defined as a pair of mappings $\delta : \mathcal{X} \rightarrow [0, 1]$ and $\pi : \mathcal{X} \rightarrow [0, 1]$ from the space \mathcal{X} to $[0, 1]$, such that δ is point-wise lower than π (i.e. $\delta \leq \pi$), with $\pi(x) = 1$ for at least one element x in \mathcal{X} , and $\delta(y) = 0$ for at least one element y in \mathcal{X} . δ and π are respectively the lower and upper distributions of a cloud.

Mappings δ, π forming the cloud $[\pi, \delta]$ are mathematically equivalent to fuzzy membership functions. A cloud $[\pi, \delta]$ is thus mathematically equivalent to an interval-valued fuzzy set (IVF for short) with boundary conditions ($\pi(x) = 1$ and $\delta(y) = 0$)³. More precisely, it is mathematically equivalent to an interval-valued membership function whereby the membership value of each element x of X is $[\delta(x), \pi(x)]$. Since a cloud is equivalent to a pair of fuzzy membership functions, at most $2|\mathcal{X}| - 2$ values (notwithstanding boundary constraints on δ and π) are needed to fully determine a cloud on a finite set. Two subcases of clouds considered by Neumaier [159] are the thin and fuzzy clouds. A *thin cloud* is defined as a cloud for which $\delta = \pi$, while a so-called *fuzzy cloud* is a cloud for which $\delta = 0$.

Given a cloud $[\pi, \delta]$, Neumaier [159] defines the credal set $\mathcal{P}_{[\pi, \delta]}$ induced by this cloud on \mathcal{X} as:

$$\mathcal{P}_{[\pi, \delta]} = \{P \in \mathbb{P}_{\mathcal{X}} \mid P(\{x \in \mathcal{X} \mid \delta(x) \geq \alpha\}) \leq 1 - \alpha \leq P(\{x \in \mathcal{X} \mid \pi(x) > \alpha\})\} \quad (3.32)$$

And, interestingly enough, this definition gives a mean to interpret IVF sets in terms of credal sets, or in terms of imprecise probabilities, eventually ending up with a behavioral interpreta-

³In general, IVF do not have to have elements x, y such that $\pi(x) = 1$ and $\delta(y) = 0$. Neither does a cloud, but a cloud not satisfying them would result in an empty credal set

tion of IVF.

When \mathcal{X} is finite, let $0 = \gamma_0 < \gamma_1 < \dots < \gamma_M = 1$ be the ordered distinct values taken by both δ and π on elements of \mathcal{X} , then denote the strong and regular cuts as

$$B_{\overline{\gamma_i}} = \{x \in \mathcal{X} \mid \pi(x) > \gamma_i\} \text{ and } B_{\gamma_i} = \{x \in \mathcal{X} \mid \pi(x) \geq \gamma_i\} \quad (3.33)$$

for the upper distribution π and

$$C_{\overline{\gamma_i}} = \{x \in \mathcal{X} \mid \delta(x) > \gamma_i\} \text{ and } C_{\gamma_i} = \{x \in \mathcal{X} \mid \delta(x) \geq \gamma_i\} \quad (3.34)$$

for the lower distribution δ . Note that in the finite case, $B_{\overline{\gamma_i}} = B_{\gamma_{i+1}}$ and $C_{\overline{\gamma_i}} = C_{\gamma_{i+1}}$, with $\gamma_{M+1} = 1$, and also

$$\begin{aligned} \emptyset &= B_{\overline{\gamma_M}} \subset B_{\overline{\gamma_{M-1}}} \subseteq \dots \subseteq B_{\overline{\gamma_0}} = \mathcal{X}; \\ \emptyset &= C_{\gamma_M} \subseteq C_{\gamma_{M-1}} \subseteq \dots \subseteq C_{\gamma_0} = \mathcal{X} \end{aligned}$$

and since $\delta \leq \pi$, this implies that $C_{\gamma_i} \subseteq B_{\gamma_i}$, hence $C_{\gamma_i} \subseteq B_{\overline{\gamma_{i-1}}}$, for all $i = 1, \dots, M$. In such a finite case, a cloud is said to be discrete. In terms of constraints bearing on probabilities, the credal set $\mathcal{P}_{[\pi, \delta]}$ of a finite cloud is described by the finite set of M inequalities:

$$i = 0, \dots, M, \quad P(C_{\gamma_i}) \leq 1 - \gamma_i \leq P(B_{\overline{\gamma_i}}) \quad (3.35)$$

under the above inclusion constraints.

Note that some conditions must hold for $\mathcal{P}_{[\pi, \delta]}$ to be non-empty in the finite case. In particular, distribution δ must not be equal to π everywhere (i.e. $\delta \neq \pi$). Otherwise, consider the case where $C_{\gamma_i} = B_{\overline{\gamma_{i-1}}}$ ($= B_{\gamma_i}$), that is π and δ have the same γ_i -cut. There is no probability distribution satisfying the constraint $1 - \gamma_{i-1} \leq P(C_{\gamma_i}) \leq 1 - \gamma_i$ since $\gamma_{i-1} < \gamma_i$. So, finite clouds cannot be thin. Example 3.5 illustrates the notion of cloud and will be used in the next sections to illustrate various results.

Example 3.5. Let us consider a space $\mathcal{X} = \{u, v, w, x, y, z\}$ and the following cloud $[\pi, \delta]$, pictured in Figure 3.4, defined on this space:

| | u | v | w | x | y | z | |
|----------|------|-----|------|------|------|-----|--------|
| π | 0.75 | 1 | 1 | 0.75 | 0.75 | 0.5 | (3.36) |
| δ | 0.5 | 0.5 | 0.75 | 0.5 | 0 | 0 | |

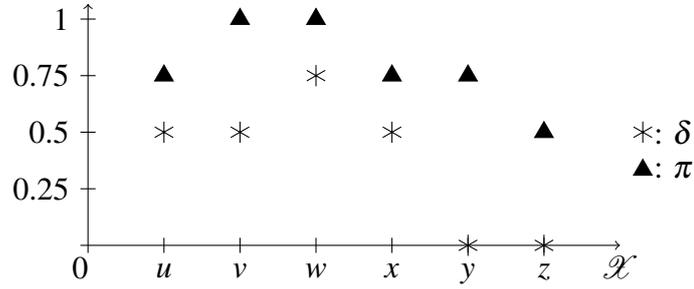


Figure 3.4: Cloud $[\pi, \delta]$ of Example 3.5

The values γ_i corresponding to this cloud are

$$0 \leq 0.5 \leq 0.75 \leq 1$$

$$\gamma_0 \leq \gamma_1 \leq \gamma_2 \leq \gamma_3$$

and the constraints associated to this cloud and corresponding to Equation (3.35) are

$$P(C_{\gamma_3} = \emptyset) \leq 1 - 1 \leq P(B_{\overline{\gamma_3}} = \emptyset)$$

$$P(C_{\gamma_2} = \{w\}) \leq 1 - 0.75 \leq P(B_{\overline{\gamma_2}} = \{v, w\})$$

$$P(C_{\gamma_1} = \{u, v, w, x\}) \leq 1 - 0.5 \leq P(B_{\overline{\gamma_1}} = \{u, v, w, x, y\})$$

$$P(C_{\gamma_0} = \mathcal{X}) \leq 1 - 0 \leq P(B_{\overline{\gamma_0}} = \mathcal{X})$$

3.3.2 Clouds in the setting of possibility theory

As for generalized p-boxes, we first relate clouds with possibility distributions. To do it, we first consider the case of *fuzzy clouds* $[\pi, \delta]$. In this case, $\delta = 0$ and, $C_{\gamma_i} = \emptyset$ for $i = 1, \dots, M$, which means that constraints given by Equations (3.35) reduce to

$$i = 0, \dots, M \quad 1 - \gamma_i \leq P(B_{\overline{\gamma_i}})$$

which induces a credal set equivalent to \mathcal{P}_π (direct from Proposition 3.2). This shows that *fuzzy clouds* are equivalent to possibility distributions.

3.3.2.1 General clouds and possibility distributions

The following proposition is a direct consequence of the preceding observation:

Proposition 3.10. *A cloud $[\pi, \delta]$ is representable by the pair of possibility distributions $1 - \delta$ and π , in the following sense:*

$$\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$$

Proof of Proposition 3.10. Consider a cloud $[\pi, \delta]$ and the constraints (3.35) inducing the credal set $\mathcal{P}_{[\pi, \delta]}$. As for generalized p-boxes, these constraints can be split into two sets of constraints, namely, for $i = 0, \dots, M$, $P(C_{\gamma_i}) \leq 1 - \gamma_i$ and $1 - \gamma_i \leq P(B_{\bar{\gamma}_i})$. Since $B_{\bar{\gamma}_i}$ are strong cuts of π , then by Proposition 3.2 we know that these constraints define a credal set equivalent to \mathcal{P}_π .

Note then that $P(C_{\gamma_i}) \leq 1 - \gamma_i$ is equivalent to $P(C_{\gamma_i}^c) \geq \gamma_i$ (where $C_{\gamma_i}^c = \{x \in \mathcal{X} \mid 1 - \delta(x) > 1 - \gamma_i\}$). By construction, $1 - \delta$ is a normalized possibility distribution. By interpreting these inequalities in the light of Proposition 3.2, we can see that they define the credal set $\mathcal{P}_{1-\delta}$. By merging the two sets of constraints, we get $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$. \square

This proposition shows that, as for generalized p-boxes, a cloud is representable by a pair of possibility distributions [96]. This similarity between clouds and generalized p-boxes is explored in Section 3.3.3. This result also confirms that a cloud $[\pi, \delta]$ is equivalent to its mirror cloud $[1 - \pi, 1 - \delta]$ (where $1 - \pi$ becomes the lower distribution, and $1 - \delta$ the upper one), as already mentioned by Neumaier [159]. Example 3.6 pursues Example 3.5 and shows the two possibility distributions induced from the cloud.

Example 3.6. We consider the same space \mathcal{X} and the same cloud as in Example 3.5. Then, possibility distributions $\pi, 1 - \delta$ are:

| | u | v | w | x | y | z |
|--------------|------|-----|------|------|------|-----|
| π | 0.75 | 1 | 1 | 0.75 | 0.75 | 0.5 |
| $1 - \delta$ | 0.5 | 0.5 | 0.25 | 0.5 | 1 | 1 |

3.3.2.2 Using possibility distributions to check non-emptiness of $\mathcal{P}_{[\pi, \delta]}$

Since not all clouds induce non-empty credal sets (e.g., a thin finite clouds), it is natural to search conditions under which a cloud $[\pi, \delta]$ induces a non-empty credal set $\mathcal{P}_{[\pi, \delta]}$. Such

conditions can be derived by using the links between clouds and possibility distributions.

Chateaufneuf [21] has found a characteristic condition under which the credal sets associated to two belief functions have a non-empty intersection. We can thus apply this result to a pair of possibility distributions and get the following necessary and sufficient condition for a cloud $[\pi, \delta]$ to have a non-empty credal set:

Proposition 3.11. *A cloud $[\pi, \delta]$ has a non-empty credal set if and only if*

$$\forall A \subseteq \mathcal{X}, \max_{x \in A} \pi(x) \geq \min_{y \notin A} \delta(y)$$

Proof of Proposition 3.11. Chateaufneuf's condition applied to possibility distributions π_1 and π_2 reads $\forall A \subseteq X, \Pi_1(A) + \Pi_2(A^c) \geq 1$. Choose $\pi_1 = \pi$ and $\pi_2 = 1 - \delta$. In particular $\Pi_2(A^c) = 1 - \min_{y \notin A} \delta(y)$. \square

Note that a naive test for non-emptiness based on this characterization would have exponential complexity. In the case of clouds, it can be simplified as follows: suppose the space $\mathcal{X} = \{x_1, \dots, x_n\}$ is indexed such that $\pi(x_1) \leq \pi(x_2) \leq \dots \leq \pi(x_n) = 1$ and consider an event A such that $\max_{x \in A} \pi(x) = \pi(x_i)$. The tightest constraint of the form $\max_{x \in A} \pi(x) = \pi(x_i) \geq \min_{y \notin A} \delta(y)$ is when choosing $A = \{x_1, \dots, x_i\}$. Hence, in the case of clouds, Chateaufneuf condition comes down to the following set of $n - 1$ inequalities to be checked:

$$j = 1, \dots, n - 1 \quad \pi(x_i) \geq \min_{j > i} \delta(x_j). \quad (3.37)$$

This gives us an efficient tool to check the non-emptiness of a given cloud, or to build a non-empty cloud from the knowledge of either δ or π . For instance, knowing δ , the cloud $[\pi, \delta]$ such that $\pi(x_i) = \min_{j > i} \delta(x_j), j = 1, \dots, n - 1$ is the most restrictive non-empty cloud that we can build, assuming the ordering $\pi(x_1) \leq \pi(x_2) \leq \dots \leq \pi(x_n) = 1$ (changing this assumption yields another non-empty cloud).

Now, consider the extreme case of a cloud for which $C_{\gamma_i} = B_{\bar{\gamma}_i}$ for all $i = 1, \dots, M$ in Equation (3.35). In this case, $P(B_{\bar{\gamma}_i}) = P(C_{\gamma_i}) = 1 - \gamma_i$ for all $i = 1, \dots, M$. Suppose distribution π takes distinct values on all elements of \mathcal{X} . Rank-ordering \mathcal{X} in increasing values of $\pi(x)$ ($\forall i, \pi(x_i) > \pi(x_{i-1})$) enforces $\delta(x_i) = \pi(x_{i-1})$, with $\delta(x_1) = 0$. Let δ_π be this lower distribution. The (almost thin) cloud $[\delta_\pi, \pi]$ satisfies Equations (3.37), and since $P(B_{\bar{\gamma}_i}) = 1 - \gamma_i$, the induced credal set $\mathcal{P}_{[\pi, \delta]}$ contains the single probability measure P with distribution

$p(x_i) = \pi(x_i) - \pi(x_{i-1})$ for all $x_i \in \mathcal{X}$, with $\pi(x_0) = 0$. So if a finite cloud $[\pi, \delta]$ is such that if $\delta > \delta_\pi$, it has empty credal set; and if $\delta \leq \delta_\pi$, then the credal set is not empty.

Conditions given by Equations (3.37) can be easily extended to the case of any two possibility distributions π_1, π_2 for which we want to check whether $\mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$ is ensured to be empty or not. Such an extension is meaningful only if the setting of clouds does not cover all the cases where $\mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2} \neq \emptyset$. To check that this is the case, we first recall that given any two possibility distributions π_1, π_2 , we do have $\mathcal{P}_{\min(\pi_1, \pi_2)} \subseteq \mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$, but in general not the converse inclusion [88]. From this remark, we can conclude that

- $\mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2} \neq \emptyset$ as soon as $\min(\pi_1, \pi_2)$ is a normalized possibility distribution.
- Not all pairs of possibility distributions such that $\mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2} \neq \emptyset$ derive from a cloud $[1 - \pi_2, \pi_1]$. Indeed the normalization of $\min(\pi_1, \pi_2)$ does not imply that $1 - \pi_2 \leq \pi_1$.

Another example is given by the two possibility distributions π_1, π_2 defined on $\mathcal{X} = \{w, x, y, z\}$ such that $\pi_1(w) = 0.5$, $\pi_1(x) = 1$, $\pi_1(y) = 0.5$, $\pi_1(z) = 0.3$ and $\pi_2(w) = 0.3$, $\pi_2(x) = 0.5$, $\pi_2(y) = 1$, $\pi_2(z) = 0.5$. $\mathcal{P}_{\pi_1} \cap \mathcal{P}_{\pi_2}$ is not empty (distribution $p(x) = 0.5$, $p(y) = 0.5$ is inside both credal sets), and neither $[1 - \pi_2, \pi_1]$ nor $[\pi_2, 1 - \pi_1]$ is a cloud.

Note that there may exist clouds $[\pi, \delta]$ with non-empty credal set while $\delta(x) = \pi(x)$ for some element x of \mathcal{X} . For instance, if for all $x \in \mathcal{X}$, $\delta(x) = \pi(x)$ if $\pi(x) < 1$ and $\delta(x) = 0$ if $\pi(x) = 1$, it defines a non-empty credal set since $\sup_{x \in \mathcal{X}} \min(\pi(x), 1 - \delta(x)) = 1$.

3.3.3 Generalized p-boxes as a special kind of clouds

Previous results show that, similarly to generalized p-boxes, clouds correspond to pairs of possibility distributions. Moreover, the constraints defining a finite cloud are similar to the ones defining a generalized p-box on a finite set, as per Equations (3.20). The lemma below lays bare the nature of the relationship between the two representations:

Proposition 3.12. *Let $[\pi, \delta]$ be a cloud defined on \mathcal{X} . Then, the three following statements are equivalent:*

- The cloud $[\pi, \delta]$ can be encoded as a p-box $[\underline{F}, \overline{F}]$ such that $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_{[\underline{F}, \overline{F}]}$
- δ and π are comonotonic ($\delta(x) < \delta(y) \Rightarrow \pi(x) \leq \pi(y)$)

(iii) Sets $\{B_{\bar{\gamma}_i}, C_{\gamma_j} | i, j = 0, \dots, M\}$ defined from Equations (3.33) and (3.34) form a nested sequence (i.e. these sets are completely (pre-)ordered with respect to inclusion).

Proof of Proposition 3.12. We use a cyclic proof to show that statements (i), (ii), (iii) are equivalent.

(i) \Rightarrow (ii) Since p-boxes and clouds are both representable by pairs of possibility distributions, then if (i) holds, we have $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_{1-\delta} \cap \mathcal{P}_\pi = \mathcal{P}_{\pi_E} \cap \mathcal{P}_{\pi_{\bar{F}}} = \mathcal{P}_{[\underline{F}, \bar{F}]}$ with $[\underline{F}, \bar{F}]$ the p-box equivalent to the cloud $[\pi, \delta]$. Using the Proposition 3.3 and the definition of a generalized p-box, $1 - \pi_{\underline{F}} = \delta$ and $\pi_{\bar{F}} = \pi$ must be comonotone, hence (i) \Rightarrow (ii).

(ii) \Rightarrow (iii) we will show that if (iii) does not hold, then (ii) does not hold either. Assume sets $\{B_{\bar{\gamma}_i}, C_{\gamma_j} | i, j = 0, \dots, M\}$ do not form a nested sequence, meaning that there exists two sets $C_{\gamma_j}, B_{\bar{\gamma}_i}$ with $j < i$ s.t. $C_{\gamma_j} \not\subseteq B_{\bar{\gamma}_i}$ and $B_{\bar{\gamma}_i} \not\subseteq C_{\gamma_j}$. This is equivalent to asserting $\exists x, y \in \mathcal{X}$ such that $\delta(x) \geq \gamma_j$, $\pi(x) \leq \gamma_i$, $\delta(y) < \gamma_j$ and $\pi(y) > \gamma_i$. This implies $\delta(y) < \delta(x)$ and $\pi(x) < \pi(y)$, and that δ, π are not comonotonic.

(iii) \Rightarrow (i) Assume the sets $B_{\bar{\gamma}_i}$ and C_{γ_j} form a globally nested sequence whose current element is A_k . Then the set of constraints defining a cloud can be rewritten in the form $\alpha_k \leq P(A_k) \leq \beta_k$, where $\alpha_k = 1 - \gamma_i$ and $\beta_k = \min\{1 - \gamma_j | B_{\bar{\gamma}_i} \subseteq C_{\gamma_j}\}$ if $A_k = B_{\bar{\gamma}_i}$; $\beta_k = 1 - \gamma_i$ and $\alpha_k = \max\{1 - \gamma_j | B_{\bar{\gamma}_i} \subseteq C_{\gamma_j}\}$ if $A_k = C_{\gamma_j}$. This ends the proof \square

Proposition 3.12 indicates that only those clouds for which δ and π are comonotonic can be encoded by generalized p-boxes, and from now on, we shall call such clouds *comonotonic*. To completely relate comonotonic clouds and generalized p-boxes, it remains to express a given comonotonic cloud $[\pi, \delta]$ as a generalized p-box $[\underline{F}, \bar{F}]$. As both clouds and generalized p-boxes correspond to pairs of possibility distribution, we can define $\pi = \pi_{\bar{F}}$ and $\delta = 1 - \pi_{\underline{F}}$, where δ, π are the distributions of the cloud and $\pi_{\bar{F}}, 1 - \pi_{\underline{F}}$ are the possibility distributions describing the generalized p-box equivalent to the cloud $[\pi, \delta]$. By using Proposition 3.3, \underline{F}, \bar{F} can then be computed for all x in \mathcal{X} :

$$\bar{F}(x) = \pi(x) \text{ and } \underline{F}(x) = \min\{\delta(y) | y \in \mathcal{X}, \delta(y) > \delta(x)\} \quad (3.38)$$

Conversely, note that any generalized p-box $[\underline{F}, \bar{F}]$ can be encoded by a comonotonic cloud, simply taking $\delta = 1 - \pi_{\underline{F}}$ and $\pi = \pi_{\bar{F}}$ (See Proposition 3.3), meaning that generalized p-boxes are special cases of clouds and are equivalent to comonotonic clouds. Also note that a comonotonic cloud $[\pi, \delta]$ and the corresponding generalized p-box $[\underline{F}, \bar{F}]$ induce the same complete pre-orders on elements of \mathcal{X} , that we will note $\leq_{[\underline{F}, \bar{F}]}$ to remain coherent with

previous notations. We will consider that elements x of \mathcal{X} are indexed accordingly, as already specified.

In practice, this relation between comonotonic clouds and generalized p-boxes means that all the results that hold for generalized p-boxes also hold for comonotonic clouds, and conversely. In particular, a comonotonic cloud $[\pi, \delta]$ can be encoded as an equivalent random set, and if we adapt Equations (3.21) to the case of the comonotonic cloud $[\pi, \delta]$, we get the random set (m, \mathcal{F}) with focal elements E_j such that for $j = 1, \dots, M$

$$\begin{cases} E_j = \{x \in \mathcal{X} \mid (\pi(x) \geq \gamma_j) \wedge (\delta(x) < \gamma_j)\} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (3.39)$$

Note that in the formalism of clouds this random set can be expressed in terms of the sets $\{B_{\bar{\gamma}_i}, C_{\gamma_i} \mid i = 0, \dots, M\}$. Namely, for $j = 1, \dots, M$:

$$\begin{cases} E_j = B_{\bar{\gamma}_{j-1}} \setminus C_{\gamma_j} = B_{\gamma_j} \setminus C_{\gamma_j} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (3.40)$$

Example 3.7 illustrates the above relations on the cloud $[\pi, \delta]$ used in Example 3.5, which is comonotonic.

Example 3.7. From the cloud of Example 3.5, $C_{\gamma_3} \subset C_{\gamma_2} \subset B_{\bar{\gamma}_2} \subset C_{\gamma_1} \subset B_{\bar{\gamma}_1} \subset B_{\bar{\gamma}_0}$, and the constraints defining $\mathcal{P}_{[\pi, \delta]}$ can be transformed into

$$\begin{aligned} 0 &\leq P(C_{\gamma_2} = \{w\}) \leq 0.25 \\ 0.25 &\leq P(B_{\bar{\gamma}_2} = \{v, w\}) \leq 0.5 \\ 0.25 &\leq P(C_{\gamma_1} = \{u, v, w, x\}) \leq 0.5 \\ 0.5 &\leq P(B_{\bar{\gamma}_1} = \{u, v, w, x, y\}) \leq 1. \end{aligned}$$

They are equivalent to the generalized p-box $[E, \bar{F}]$ pictured on Figure 3.5:

| | u | v | w | x | y | z | |
|-----------|------|------|-----|------|------|-----|--------|
| \bar{F} | 0.75 | 1 | 1 | 0.75 | 0.75 | 0.5 | (3.41) |
| F | 0.5 | 0.75 | 1 | 0.5 | 0.5 | 0 | |

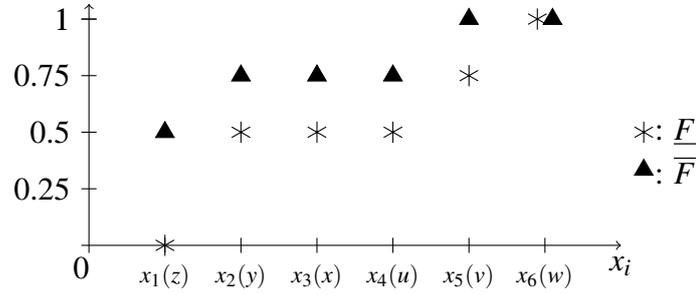


Figure 3.5: Generalized p-box $[\underline{F}, \overline{F}]$ corresponding to cloud of Example 3.5

The following ranking of elements of \mathcal{X} is compatible with the two distributions (see Figure 3.5):

$$z <_{[\underline{F}, \overline{F}]} y <_{[\underline{F}, \overline{F}]} x =_{[\underline{F}, \overline{F}]} u <_{[\underline{F}, \overline{F}]} v <_{[\underline{F}, \overline{F}]} w$$

The corresponding random set, given by Equations (3.40) or (3.39), is:

$$\begin{aligned} m(\{x_5, x_6\}) &= 0.25 \\ m(\{x_2, x_3, x_4, x_5\}) &= 0.25 \\ m(\{x_1, x_2\}) &= 0.5 \end{aligned}$$

These results provide insight in uncertainty representations based on pairs of comonotonic possibility distributions. They emphasize different views of the same tool. Comonotonic clouds being special cases of clouds, it is then natural to wonder if some of the results presented in this section extend to clouds that are not comonotonic (and consequently called *non-comonotonic*). In particular, can uncertainty modeled by a non-comonotonic cloud be exactly modeled by an equivalent random set?

3.3.4 The Nature of Non-comonotonic Clouds

We will now study the case of non-comonotonic clouds. For this kind of clouds, Proposition 3.10 linking clouds and possibility distributions still holds, but non-comonotonic clouds are no longer equivalent to generalized p-boxes, thus results valid for comonotonic clouds cannot be used anymore. As we shall see, non-comonotonic clouds appear to be less interesting,

at least from a practical point of view, than comonotonic ones.

3.3.4.1 Characterization

One way of characterizing an uncertainty model is to find the maximal natural number n such that the lower measure induced by this uncertainty model is always n -monotone. This is how we will proceed with non-comonotonic clouds. Let $[\pi, \delta]$ be a non-comonotonic cloud, and $\mathcal{P}_{[\pi, \delta]}$ the induced credal set. The question is: what is the (ensured) n -monotonicity of the associated lower probability \underline{P} of $\mathcal{P}_{[\pi, \delta]}$? To address this question, let us start with an example:

Example 3.8. Consider a space \mathcal{X} of five elements $\{v, w, x, y, z\}$ and the following non-comonotonic cloud $[\pi, \delta]$ pictured on Figure 3.6:

| | v | w | x | y | z | |
|----------|---|-----|------|-----|------|--------|
| π | 1 | 1 | 0.5 | 0.5 | 0.25 | (3.42) |
| δ | 0 | 0.5 | 0.25 | 0 | 0 | |

This cloud is non-comonotonic, since $\pi(v) > \pi(x)$ and $\delta(v) < \delta(x)$. The credal set $\mathcal{P}_{[\pi, \delta]}$ can also be defined by the following constraints:

$$\begin{aligned} P(C_{\gamma_2} = \{w\}) &\leq 1 - 0.5 \leq P(B_{\bar{\gamma}_2} = \{v, w\}) \\ P(C_{\gamma_1} = \{w, x\}) &\leq 1 - 0.25 \leq P(B_{\bar{\gamma}_1} = \{v, w, x, y\}) \end{aligned}$$

with $\gamma_2 = 0.5$ and $\gamma_1 = 0.25$. Now, consider the events $B_{\bar{\gamma}_2}, C_{\gamma_1}^c, B_{\bar{\gamma}_2} \cap C_{\gamma_1}^c, B_{\bar{\gamma}_2} \cup C_{\gamma_1}^c$. We can check that

$$\begin{aligned} \underline{P}(B_{\bar{\gamma}_2}) &= 0.5 & \underline{P}(C_{\gamma_1}^c) &= 0.25 \\ \underline{P}(B_{\bar{\gamma}_2} \cap C_{\gamma_1}^c = \{v\}) &= 0 & \underline{P}(B_{\bar{\gamma}_2} \cup C_{\gamma_1}^c = \{v, w, y, z\}) &= 0.5 \end{aligned}$$

since at most a 0.5 probability mass can be assigned to x . Then the inequality $\underline{P}(B_{\bar{\gamma}_2} \cap C_{\gamma_1}^c) + \underline{P}(B_{\bar{\gamma}_2} \cup C_{\gamma_1}^c) < \underline{P}(B_{\bar{\gamma}_2}) + \underline{P}(C_{\gamma_1}^c)$ holds, which shows that the lower probability induced by the cloud is not 2-monotone.

This example is sufficient to show that at least some non-comonotonic clouds induce lower probability measures that are not 2-monotone. The following proposition gives a general characterization of such non-comonotonic clouds:

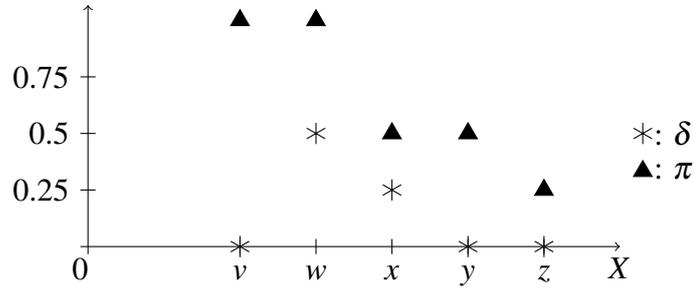


Figure 3.6: Cloud $[\pi, \delta]$ of Example 3.8

Proposition 3.13. *Let $[\pi, \delta]$ be a non-comonotonic cloud and assume there is a pair of events $B_{\gamma_i}, C_{\gamma_j}$ in the cloud s.t. $B_{\gamma_i} \cap C_{\gamma_j} \notin \{B_{\gamma_i}, C_{\gamma_j}, \emptyset\}$ and $B_{\gamma_i} \cup C_{\gamma_j} \neq \mathcal{X}$ (i.e., $B_{\gamma_i}, C_{\gamma_j}$ are just overlapping and do not cover the whole space \mathcal{X}). Then, the lower probability measure of the credal set $\mathcal{P}_{[\pi, \delta]}$ is not 2-monotone.*

Proof. See Section D.2 in Appendix D □

The proof comes down to showing that for any non-comonotonic cloud with a pair $B_{\gamma_i}, C_{\gamma_j}$ of events such that $B_{\gamma_i} \cap C_{\gamma_j} \notin \{B_{\gamma_i}, C_{\gamma_j}, \emptyset\}$ and $B_{\gamma_i} \cup C_{\gamma_j} \neq \mathcal{X}$ the situation exhibited in the above example always occurs, namely the existence of two subsets of the form B_{γ_i} and $C_{\gamma_j}^c$ for which 2-monotonicity fails. Proposition 3.13 also shows that non-comonotonic clouds satisfying this proposition cannot be viewed as random sets. Note that, although comonotonic clouds and clouds described by Proposition 3.13 cover a large number of possible discrete clouds, there remain "small" subfamilies of non-comonotonic clouds not covered by Proposition 3.13. That is, clouds such that, for any pair of events $B_{\gamma_i}, C_{\gamma_j}$ we have $B_{\gamma_i} \cap C_{\gamma_j} \in \{B_{\gamma_i}, C_{\gamma_j}, \emptyset\}$ or $B_{\gamma_i} \cup C_{\gamma_j} \neq \mathcal{X}$. Since such subfamilies are very peculiar and present poor interest, we will not explore them further here. Some considerations about the clouds satisfying $B_{\gamma_i} \cap C_{\gamma_j} \in \{B_{\gamma_i}, C_{\gamma_j}, \emptyset\}$, that we call nested-disjoint, are given in Appendix E

Since non-comonotonic clouds appear less tractable than comonotonic ones, it is desirable to provide tools that would make them more convenient to use. Such tools are given in the next two sections, in the form of guaranteed outer and inner approximations of the exact probability bounds induced by a non-comonotonic cloud. To this aim, we rely on previous results.

3.3.4.2 Outer approximation of a non-monotonic cloud

Given a cloud $[\pi, \delta]$, we have proven that $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_\pi \cap \mathcal{P}_{1-\delta}$, where π and $1 - \delta$ are possibility distributions. As a consequence, the upper and lower probabilities of $\mathcal{P}_{[\pi, \delta]}$ on any event can be bounded from above (resp. from below), using the possibility measures and the necessity measures induced by π and $\bar{\pi} = 1 - \delta$. The following bounds, originally considered by Neumaier [159], provide, for all event A of \mathcal{X} , an outer approximation of the range of $P(A)$:

$$\max(N_\pi(A), N_{1-\delta}(A)) \leq \underline{P}(A) \leq P(A) \leq \bar{P}(A) \leq \min(\Pi_\pi(A), \Pi_{1-\delta}(A)), \quad (3.43)$$

where $\underline{P}(A), \bar{P}(A)$ are the lower and upper probabilities induced by $\mathcal{P}_{[\pi, \delta]}$. Remember that probability bounds generated by possibility distributions alone are of the form $[0, \beta]$ or $[\alpha, 1]$. Using a cloud and applying Equation (3.43) lead to tighter bounds of the form $[\alpha, \beta] \subseteq [0, 1]$, and thus to more precise information, while remaining simple to compute. Nevertheless, these bounds are not, in general, the infimum and the supremum of $P(A)$ over $\mathcal{P}_{[\pi, \delta]}$ (i.e., $\max(N_\pi, N_{1-\delta})$ is not the lower envelope of $\mathcal{P}_{[\pi, \delta]}$). To see this, consider the following example:

Example 3.9. Let $[\pi, \delta]$ be a cloud defined on a space \mathcal{X} , such that distributions δ and π takes up to four different values on elements x of \mathcal{X} (including 0 and 1). These values are such that $0 = \gamma_0 < \gamma_1 < \gamma_2 < \gamma_3 = 1$, and the distributions δ, π are such that

$$\begin{aligned} \pi(x) &= 1 \text{ if } x \in B_{\bar{\gamma}_2}; \\ &= \gamma_2 \text{ if } x \in B_{\bar{\gamma}_1} \setminus B_{\bar{\gamma}_2}; \\ &= \gamma_1 \text{ if } x \notin B_{\bar{\gamma}_1}. \\ \delta(x) &= \gamma_2 \text{ if } x \in C_{\gamma_2}; \\ &= \gamma_1 \text{ if } x \in C_{\gamma_1} \setminus C_{\gamma_2}; \\ &= 0 \text{ if } x \notin C_{\gamma_1}. \end{aligned}$$

Since $P(B_{\bar{\gamma}_1}) \geq 1 - \gamma_1$ and $P(C_{\gamma_2}) \leq 1 - \gamma_2$, from Equations (3.35), we can check that $\underline{P}(B_{\bar{\gamma}_1} \setminus C_{\gamma_2}) = \underline{P}(B_{\bar{\gamma}_1} \cap C_{\gamma_2}^c) = \gamma_2 - \gamma_1$. Now, by definition of a necessity measure, $N_\pi(B_{\bar{\gamma}_1} \cap C_{\gamma_2}^c) = \min(N_\pi(B_{\bar{\gamma}_1}), N_\pi(C_{\gamma_2}^c)) = 0$ since $\Pi_\pi(C_{\gamma_2}) = 1$ because $C_{\gamma_2} \subseteq B_{\bar{\gamma}_1}$ and $\Pi_\pi(B_{\bar{\gamma}_1}) = 1$. Considering distribution δ , we can have $N_{1-\delta}(B_{\bar{\gamma}_1} \cap C_{\gamma_2}^c) = \min(N_{1-\delta}(B_{\bar{\gamma}_1}), N_{1-\delta}(C_{\gamma_2}^c)) = 0$ since $N_{1-\delta}(B_{\bar{\gamma}_1}) = \Delta_\delta(B_{\bar{\gamma}_1}^c) = 0$ and $C_{\gamma_1} \subseteq B_{\bar{\gamma}_1}$ (which means that the elements x of \mathcal{X} that are in $B_{\bar{\gamma}_1}^c$ are such that $\delta(x) = 0$). Equation (3.43) can thus result in a trivial lower bound (i.e. equal to 0), different from $\underline{P}(B_{\bar{\gamma}_1} \cap C_{\gamma_2}^c)$.

Bounds given by Equation (3.43), are the main motivation for clouds, after Neumaier [159]. Since these bounds are, in general, not the infimum and supremum of $P(A)$ on $\mathcal{P}_{[\pi, \delta]}$, Neumaier's claim that they're only vaguely related to Walley's previsions or to random sets is not surprising. If a cloud is comonotonic, Equation (3.43) remains valid but becomes less useful. Indeed, since comonotonic clouds are equivalent to generalized p-boxes, we can easily compute exact values of lower and upper probabilities of $\mathcal{P}_{[\pi, \delta]}$, e.g. via the random set representation (see Remark 3.2).

3.3.4.3 Inner approximation of a non-comonotonic cloud

The previous outer approximation is easy to compute and allows to make some of Neumaier's claims clearer. Nevertheless, it is still unclear how to practically use these outer bounds in subsequent treatments (e.g., propagation, fusion). The inner approximation of a cloud $[\pi, \delta]$ proposed now is a random set, which is easy to exploit in practice. This inner approximation is given by the following proposition:

Proposition 3.14. *Let $[\pi, \delta]$ be a non-comonotonic cloud defined on a space \mathcal{X} . Let us then define, for $j = 1, \dots, M$, the following random set (m, \mathcal{F}) :*

$$\begin{cases} E_j = \{x \in \mathcal{X} | (\pi(x) \geq \gamma_j) \wedge (\delta(x) < \gamma_j)\} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (3.44)$$

where $0 = \gamma_0 < \dots < \gamma_j < \dots < \gamma_M = 1$ are the distinct values taken by δ, π on elements of \mathcal{X} , E_j are the focal elements with masses $m(E_j)$ of the random set. This random set is an inner approximation of $[\pi, \delta]$, in the sense that the credal set $\mathcal{P}_{(m, \mathcal{F})}$ induced by this random set is such that $\mathcal{P}_{(m, \mathcal{F})} \subseteq \mathcal{P}_{[\pi, \delta]}$.

Proof. See Section D.2 in Appendix D □

In the case of non-comonotonic clouds satisfying Proposition 3.13, the inclusion is strict. This inner approximation appears to be a natural candidate, since on events of the type $\{B_{\bar{\gamma}_i}, C_{\gamma_i}, B_{\bar{\gamma}_i} \setminus C_{\gamma_i} | i = 0, \dots, M; j = 0, \dots, M; i \leq j\}$ it gives exact bounds, and is exact when the cloud $[\pi, \delta]$ is comonotonic. This finishes our study of non-monotonic clouds. It remains to study the links existing between clouds and imprecise probability assignments.

3.3.5 Clouds and imprecise probability assignments

Since in many cases a cloud is either equivalent to a random set or does not lead to 2-monotone capacities, there is no direct relationship between clouds and imprecise probability assignments. Nevertheless, we can study how to transform a set of imprecise probability assignments into a cloud. Such a transformation can be useful when one wishes to work with clouds even if information is obtained in terms of sets of imprecise probability assignments.

There are mainly two paths that can be followed to do this transformation:

- the first one is to use the fact that clouds are equivalent to pairs of possibility distributions, and to extend existing transformations that transform a set of imprecise probability assignments into a single possibility distribution.
- The second uses the correspondence between generalized p-boxes and comonotonic clouds, and simply apply the results obtained for generalized p-boxes.

Section 3.3.5.1 proposes a transformation following the first path, while Section 3.3.5.2 explores the second one and compares the two approaches.

3.3.5.1 Exploiting probability-possibility transformations

The problem of transforming a probability distribution into a quantitative possibility distribution has been addressed by many authors (see Dubois and Prade [94] for an extended discussion). A consistency principle between (precise) probabilities and possibility distributions was first informally stated by Zadeh [219]: what is probable should be possible. It was later translated by Dubois and Prade [81, 98] as a mathematical constraint. Given a possibility distribution π obtained by the transformation of a probability measure P , one should have, for all events A of \mathcal{X} :

$$P(A) \leq \Pi(A) \tag{3.45}$$

In this case, the possibility measure Π is said to dominate P , and the transformation from probability to possibility then consists of choosing a possibility distribution amongst the ones inducing a possibility measure dominating P . Dubois and Prade [78, 98] propose to add the following ordinal equivalence constraint, such that for two elements x, y in \mathcal{X}

$$p(x) \leq p(y) \iff \pi(x) \leq \pi(y) \tag{3.46}$$

and to choose the least specific possibility distribution (π' is more specific than π if $\pi' \leq \pi$) respecting these two constraints.

Dubois and Prade [81] showed that the solution exists and is unique. Let us consider probability masses such that the order on probability masses is $p_1 \leq \dots \leq p_n$ with $p_j = p(x_j)$. When all probability masses are different, Dubois and Prade probability-possibility transformation can be formulated as

$$\pi_i = \sum_{j=1}^i p_j \quad (3.47)$$

with $\pi_i = \pi(x_i)$. When some elements have equal probability, the above equation must be used on the ordered partition induced by the probability weights, using uniform probabilities inside each element of the partition. Note that the above transformation is equivalent to apply Algorithm 1 with elements ranked accordingly to the increasing order of their probability masses.

Reversing the ordering of the p_i 's in the above formula yields another possibility distribution $\bar{\pi}_i = \sum_{j=i}^n p_j$, with $\bar{\pi}_i = \bar{\pi}(x_i)$. Letting $\delta = 1 - \bar{\pi}$, distribution δ is of the form δ_π introduced in Section 3.3.2.2, that is, $[\pi, \delta]$ is a cloud such that $\delta_i = \pi_{i-1}$ for all $i > 1$, with $\delta_1 = 0$ and $\delta_i = \delta(x_i)$. It is the tightest cloud containing P , in the sense that $\mathcal{P}(\pi) \cap \mathcal{P}(\bar{\pi}) = \{P\}$. This shows that, at least when probability masses are precise, transformation into possibility distributions can be extended to get a second possibility distribution such that this pair of comonotonic distributions is equivalent to a cloud. Moreover, the fact that $\mathcal{P}(\pi) \cap \mathcal{P}(\bar{\pi}) = \{P\}$ shows that the cloud models exactly the same information as the (precise) probability distribution. δ corresponds to the application of Algorithm 1, this time with a ranking corresponding to the decreasing order of probability masses. The generalized p-box $[E, \bar{F}]$ associated to the cloud $[\delta_\pi, \pi]$ is the one built through Algorithm 2 with a ranking corresponding to the increasing order of their probability masses.

When working with imprecise probability assignments, i.e. with a set L of imprecise probability assignments, the order induced by probability weights on \mathcal{X} is a partial order \leq_L (actually, an interval order) defined by:

$$x \leq_L y \iff u(x) \leq l(y) \quad (3.48)$$

and two elements x, y are incomparable if intervals $[l(x), u(x)], [l(y), u(y)]$ intersect. The problem of transforming a set L of imprecise probability assignments into a possibility distribution π_L such that $\mathcal{P}_L \subseteq \mathcal{P}_{\pi_L}$ by extending Dubois and Prade transformation is studied in detail by

Masson and Denoeux [141]. We first recall their method, before proposing its extension to clouds.

Let \mathcal{C}_L be the set of linear extensions of the partial order \leq_L : a linear extension $<_l \in \mathcal{C}_L$ is a ranking of \mathcal{X} compatible with the partial order \leq_L . Let σ_l be the permutation such that $\sigma_l(x)$ is the rank of element x in the linear extension $<_l$. Given this partial order, Masson and Denoeux [141] propose the following procedure transforming the set of imprecise probability assignments into a possibility distribution:

1. For each linear order $<_l \in \mathcal{C}_L$ and each element x , solve

$$\pi^l(x) = \max_{\{P(y)|y \in \mathcal{X}\}} \sum_{\sigma_l(y) \leq \sigma_l(x)} P(y) \quad (3.49)$$

under the constraints

$$\left\{ \begin{array}{l} \sum_{x \in \mathcal{X}} P(x) = 1 \\ \forall x \in \mathcal{X}, l(x) \leq P(x) \leq u(x) \\ P(\sigma_l^{-1}(1)) \leq P(\sigma_l^{-1}(2)) \leq \dots \leq P(\sigma_l^{-1}(n)) \end{array} \right. \quad (3.50)$$

2. The most informative distribution π dominating all distributions π^l is:

$$\pi(x) = \max_{<_l \in \mathcal{C}} \pi^l(x). \quad (3.51)$$

This procedure ensures that the resulting possibility distribution π dominates every probability distribution contained in \mathcal{P}_L . In other words, the convex set \mathcal{P}_π is such that $\mathcal{P}_L \subseteq \mathcal{P}_\pi$.

To extend this transformation to a pair of possibility distributions equivalent to a cloud, we consider that the possibility distribution π given by Equation (3.51) is the upper distribution of a cloud $[\pi, \delta]$. To build the lower distribution δ of a cloud containing \mathcal{P}_L , we need to build a second possibility distribution π_δ such that $\mathcal{P}_L \subseteq \mathcal{P}_{\pi_\delta}$ and such that the pair $[1 - \pi_\delta, \pi]$ defines a cloud (with $1 - \pi_\delta = \delta$). To achieve this, we propose to use the same method as Masson and Denoeux [141], simply reversing the inequality under the summation sign in Equation (3.49). The procedure to build π_δ then becomes

1. For each order $\prec_l \in \mathcal{C}_L$ and each element x , solve

$$\pi_\delta^l(x) = \max_{\{p(y)|y \in \mathcal{X}\}} \sum_{\sigma_l(x) \leq \sigma_l(y)} p(y) \quad (3.52)$$

$$= 1 - \min_{\{p(y)|y \in \mathcal{X}\}} \sum_{\sigma_l(y) < \sigma_l(x)} p(y) = 1 - \delta^l(x) \quad (3.53)$$

with the same constraints as in the first transformation.

2. The most informative distribution dominating all distributions $\pi_\delta^l(x)$ is:

$$\pi_\delta(x) = 1 - \delta(x) = \max_{\prec_l \in \mathcal{C}} \pi_\delta^l(x) \quad (3.54)$$

Example 3.10 illustrates this procedure.

Example 3.10. Let us take the same four imprecise probability assignments as in the example given by Masson and Denoeux [141], on the space $\mathcal{X} = \{w, x, y, z\}$, and summarized in the following table

| | w | x | y | z |
|-----|------|------|------|------|
| l | 0.10 | 0.34 | 0.25 | 0 |
| u | 0.28 | 0.56 | 0.46 | 0.08 |

The partial order is given by $L_y < L_x; L_z < \{L_x, L_w, L_y\}$. There are three possible linear extensions $\prec_l \in \mathcal{C}_L$

$$\prec_l^1 = (L_z, L_w, L_y, L_x)$$

$$\prec_l^2 = (L_z, L_w, L_x, L_y)$$

$$\prec_l^3 = (L_z, L_y, L_w, L_x)$$

corresponding to the following π_δ 's:

| \prec_l^i | $\pi_\delta(w)$ | $\pi_\delta(x)$ | $\pi_\delta(y)$ | $\pi_\delta(z)$ |
|-------------|-----------------|-----------------|-----------------|-----------------|
| 1 | 1 | 0.16 | 0.63 | 1 |
| 2 | 1 | 0.9 | 0.46 | 1 |
| 3 | 0.75 | 0.5 | 1 | 1 |
| max | 1 | 0.9 | 1 | 1 |

and, finally, the obtained cloud is:

| | w | x | y | z |
|----------|------|-----|-----|------|
| π | 0.64 | 1 | 1 | 0.08 |
| δ | 0 | 0.1 | 0 | 0 |

where π is the possibility distribution obtained by Masson and Denoeux [141] for the same example by applying the first transformation. Note that the cloud is only a little more informative than the upper distribution taken alone (indeed, the only added constraint is that $p(x) \leq 0.9$).

We can verify the following property:

Proposition 3.15. *Given a set L of imprecise probability assignments, the cloud $[1 - \pi_\delta, \pi]$ built from the two possibility distributions π_δ, π obtained via the above procedures is such that the induced credal set $\mathcal{P}_{[1 - \pi_\delta, \pi]}$ contains \mathcal{P}_L . In the degenerate case of a precise probability distribution, this cloud contains this distribution only.*

Proof of Proposition 3.15. The two possibility distributions π, π_δ are such that $\mathcal{P}_L \subset \mathcal{P}_\pi$ and $\mathcal{P}_L \subset \mathcal{P}_{\pi_\delta}$ by construction, so $\mathcal{P}_L \subset (\mathcal{P}_\pi \cap \mathcal{P}_{\pi_\delta})$. The final result is thus more precise than a single possibility distribution dominating \mathcal{P}_L . When L reduces to precise masses $\{p\}$, the transformations give the following possibility distributions (elements of \mathcal{X} are ordered in accordance with the order of probability masses):

$$\pi(x_i) = \sum_{j \leq i} p_j \quad (3.55)$$

and

$$\pi_\delta(x_i) = \sum_{j \geq i} p_j = 1 - \sum_{j < i} p_j = 1 - \delta(x_i) = 1 - \pi(x_{i-1}). \quad (3.56)$$

Hence, the only probability distribution in the cloud $[\pi, \delta]$ is given by $p_i = \pi(x_i) - \pi(x_{i-1})$. \square

This method allows to get a cloud encompassing the information contained in any set of imprecise probability assignments. It directly extends known methods used in possibility theory, and can benefit from the efficient algorithm provided by Masson and Denoeux [141].

3.3.5.2 Using generalized p-boxes

We have previously shown that generalized p-boxes and comonotonic clouds were equivalent representations. Thus, another method consists of directly using transformations from imprecise probability assignments to generalized p-boxes and get an approximation as a comonotonic cloud, simply by applying Algorithm 2.

Consider the following example:

Example 3.11. Let us consider the same probability intervals as in example 3.10 and the following order relationship R on the elements: $z <_R w <_R y <_R x$. We can then build the generalized p-box $[\underline{F}, \overline{F}]$ associated to this order (using Equations (3.24)), and then take the comonotonic cloud $[\pi, \delta]$ associated to this p-box (by using transformations in Proposition 3.3). The following table summarizes the obtained distributions

| | w | x | y | z |
|----------------------|------|------|------|------|
| $\overline{F} = \pi$ | 0.36 | 1 | 0.66 | 0.08 |
| \underline{F} | 0.1 | 1 | 0.44 | 0 |
| δ | 0 | 0.44 | 0.1 | 0 |

And, by Proposition 3.5 and results related to generalized p-boxes, we know that the credal set $\mathcal{P}_{[\pi, \delta]}$ induced by this cloud is such that $\mathcal{P}_L \subseteq \mathcal{P}_{[\pi, \delta]}$ and that we can recover the information modeled by a set L of imprecise probability assignments by means of at least $|\mathcal{X}|/2$ clouds built by this method.

Both methods transform a set L of imprecise probability assignments into a cloud $[\pi, \delta]$ such that $\mathcal{P}_L \subseteq \mathcal{P}_{[\pi, \delta]}$, thus guaranteeing that no extra information is added in the transformation.

In general, since finite clouds can model precise probability distributions without any loss of information, the cloud resulting from any transformation of a discrete probability distribution should contain this probability distribution only. Both methods proposed here satisfy these requirements in the finite case.

However, if we compare the clouds resulting from Examples 3.10 and 3.11, it is clear that the cloud resulting from the second method (Example 3.11) is more precise than the one resulting from the first one (Example 3.10). Moreover, using the first method, it is in general impossible to recover the information provided by the original set L of imprecise probability assignments. This shows that the first method can be very conservative. This is mainly due to the fact that even if the method considers every possible ranking of elements, it is only based

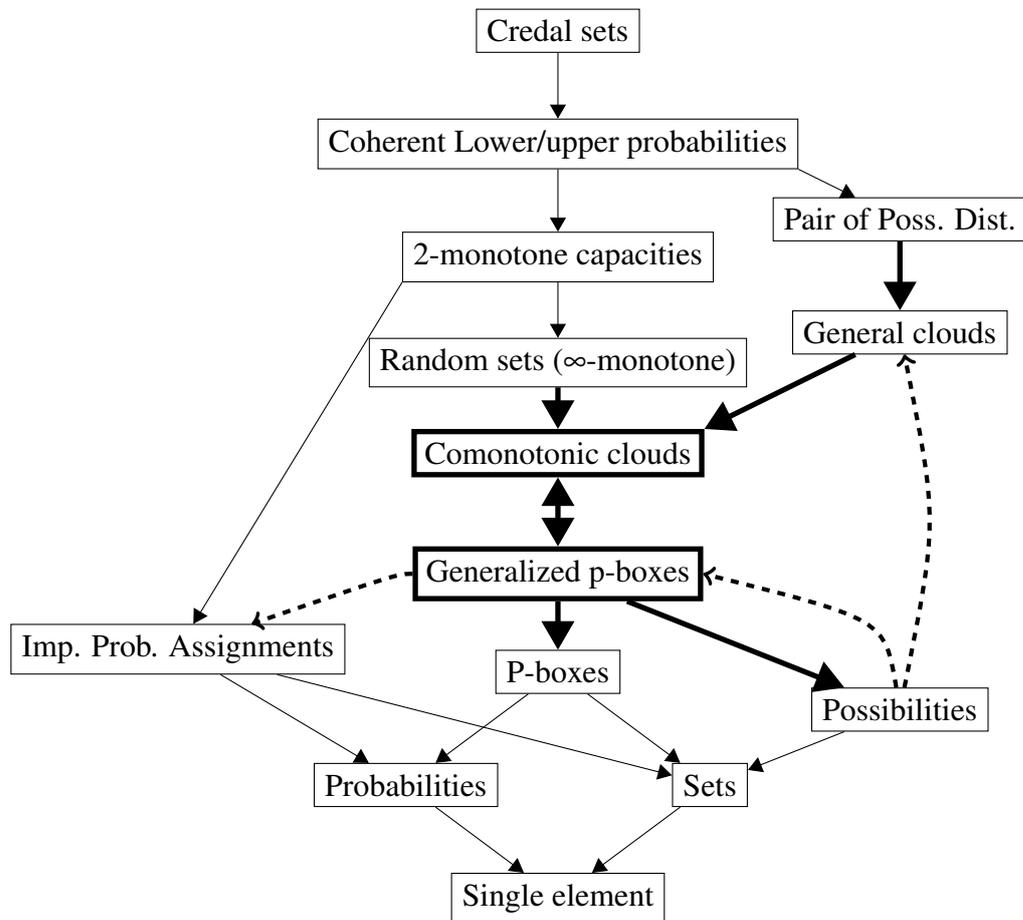


Figure 3.7: Representation relationships: summary with clouds $A \longrightarrow B$: B is a special case of A . $A \dashrightarrow B$: B is representable by A

on the partial order induced by imprecise probability assignments. Also, it is not based on the use of upper probabilities, as Algorithm 2 is.

If a ranking of elements is naturally present in a considered problem, then the second method seems to be the best solution. If no natural order is present, it is hard to justify the fact of considering one particular order rather than another one, and the first method should be applied. In this case, one has to be aware that a lot of information can be lost in the process. One may also use the ranking inducing one of the most precise comonotonic clouds, but we already mentioned that this question remains open. Also, it is not sure than the second method can always produce a more informative result than the first one. Figure 3.7 completes Figure 3.3 by inserting both general clouds and comonotonic clouds into it. As for the previous figures, new relationships resulting from the present work are in bold.

3.4 A word on continuous representations on the real line

In many applications, the available information concerning some parameters is defined on the (continuous) real line. Even if in practice the computational processing of such models will require sooner or later a discretization of the real line, thus coming back to the finite case, the continuous representation is in many cases mathematically convenient and often highlights interesting properties. Analytical formulations and closed forms can also be useful to build efficient numerical methods or algorithms.

Thus, although the major part of this work deals with discretized models or with models defined on finite spaces, we think that it is both interesting and important to devote some attention to models defined on the continuous real line. We restrict ourselves to the study of models defined on measurable spaces, because non-measurable sets are unlikely to occur in applications, and to avoid mathematical subtleties arising with the consideration of countably additive probabilities on infinite spaces (Countable additivity for generalized p-boxes is considered in Appendix F). Whether countable additivity should be considered or not by uncertainty theories remains a matter of debate, which is far beyond the scope of this work (see Miranda [144] for a short discussion).

3.4.1 Practical continuous representations on the real line

In this section, we briefly review how some of the models defined in Section 3.1 can be defined on the real line.

P-boxes The definition of p-boxes essentially remains the same, instead that the cumulative distributions $\underline{F}, \overline{F}$ are allowed to be continuous distributions defined over the real line. The credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$ associated to such a p-box is then the set of probability densities p whose cumulative distributions F are between $\underline{F}, \overline{F}$ (i.e. $\underline{F}(r) \leq F(r) \leq \overline{F}(r)$ for all $r \in \mathbb{R}$).

Imprecise probability assignments When defined on the continuous real line, imprecise probability assignment L become lower and upper density functions l and u , such that

$$\begin{aligned} \forall x \in \mathbb{R} \quad l(x) &\leq u(x) \\ \int_{\mathbb{R}} l(x) dx &\leq 1 \leq \int_{\mathbb{R}} u(x) dx \end{aligned}$$

and the credal set \mathcal{P}_L associated to this imprecise probability assignment is the set of all (normalized) density functions that lie between l and u . Other models defined in terms of lower/upper density functions and based on ratio of probability densities are considered by Walley [203, Ch.4.6.]

Random sets As in Section 3.1.2.4, random sets will be noted (m, \mathcal{F}) , the only difference being that m is now a probability density defined on $[0, 1]$ and that to each value $\alpha \in [0, 1]$ is associated a focal element E_α . Equations for belief and plausibility measures of an event $A \subseteq \mathbb{R}$ simply become

$$Pl(A) = \int_0^1 \mathbf{1}_{(A \cap E_\alpha \neq \emptyset)} m(\alpha) d\alpha$$

$$Bel(A) = \int_0^1 \mathbf{1}_{(A \subseteq E_\alpha)} m(\alpha) d\alpha$$

with $\mathbf{1}_{(A)}$ the indicator function which has value one if A is true, zero otherwise. If m is a uniform density over $[0, 1]$, above equations simply become:

$$Pl(A) = \int_0^1 \mathbf{1}_{(A \cap E_\alpha \neq \emptyset)} d\alpha$$

$$Bel(A) = \int_0^1 \mathbf{1}_{(A \subseteq E_\alpha)} d\alpha.$$

Smets [186] considers a more peculiar case, where focal elements E_α are restricted to closed intervals of \mathbb{R} . The interest of only considering closed intervals is the possibility to derive analytical formulas that extend various notions of finite random sets while remaining tractable. Moreover, as mentioned in Section 3.1.2.4, in practical applications random sets defined over the reals are often restricted to closed intervals. Continuous random sets still induce corresponding credal sets.

Possibility distributions We define a possibility distribution on the real line as a lower semi-continuous mapping $\pi : \mathbb{R} \rightarrow [0, 1]$ such that $\pi(x) = 1$ for at least one value $x \in \mathbb{R}$. In this case, relationships between possibility distributions, random sets and credal sets continue to hold. In particular, a possibility distribution π can be represented by a continuous random set with a uniform mass density m over $[0, 1]$ and the multi-valued mapping $\alpha \rightarrow \pi_\alpha$, with $\alpha \in [0, 1]$ and

π_α the α -cut of π . Similarly, the credal set \mathcal{P}_π induced by the possibility distribution π can still be described through its α -cuts. In particular, Proposition 3.2 still holds. Also, when π is such that every α -cut is a closed interval, it can be embedded straightforwardly in Smets [186] framework.

In the next section, we show that the main results of Sections 3.2 and 3.3 extend to models defined on the continuous real line.

3.4.2 Continuous clouds on the real line

We consider clouds defined on a bounded interval $[\underline{r}, \bar{r}]$ of the real line. First, let us recall that, as in the discrete case, a cloud $[\pi, \delta]$ defined on the real line is a pair of distributions such that, for any element $r \in \mathbb{R}$, $[\delta(r), \pi(r)]$ is an interval and there is an element r for which $\delta(r) = 0$, and another r' for which $\pi(r') = 1$. *Thin clouds* ($\pi = \delta$) and *Fuzzy clouds* ($\delta = 0$) have the same definition as in the case of finite spaces. The credal set $\mathcal{P}_{[\pi, \delta]}$ induced by a cloud on the real line is such that:

$$\mathcal{P}_{[\pi, \delta]} = \{P \in \mathbb{P}_{\mathbb{R}} \mid P(\{r \in \mathbb{R} \mid \delta(r) \geq \alpha\}) \leq 1 - \alpha \leq P(\{r \in \mathbb{R} \mid \pi(r) > \alpha\})\}, \quad (3.57)$$

where P is a σ -additive probability distribution.

As Proposition 3.2 has been proven for possibility distributions π defined on very general spaces [31], and since the proof of Proposition 3.12 extends directly to continuous models on the real line, results whose proof is based on these propositions also hold on the real line. In particular, the following statements still hold:

- if $[\pi, \delta]$ is a cloud, $1 - \delta, \pi$ are possibility distributions, and $\mathcal{P}_{[\pi, \delta]} = \mathcal{P}_\delta \cap \mathcal{P}_\pi$,
- if $[\underline{F}, \bar{F}]$ is a generalized p-box defined on the reals, then $\mathcal{P}_{[\underline{F}, \bar{F}]} = \mathcal{P}_{\pi_{\underline{F}}} \cap \mathcal{P}_{\pi_{\bar{F}}}$ with, for all $r \in \mathbb{R}$:

$$\pi_{\bar{F}}(r) = \bar{F}(r)$$

and

$$\pi_{\underline{F}}(r) = 1 - \sup \{\underline{F}(r') \mid r' \in \mathbb{R}; \underline{F}(r') < \underline{F}(r)\}$$

with $\pi_{\underline{F}}(\underline{r}) = 0$.

- *generalized* p-boxes and comonotonic clouds are equivalent representations.

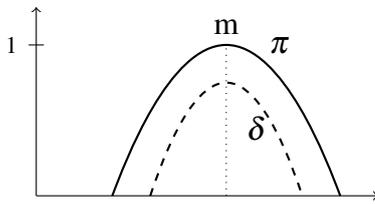


Fig. 3.8.A: Comonotonic cloud

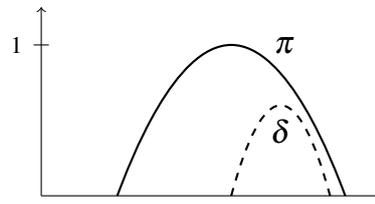


Fig. 3.8.B: Non-comonotonic cloud

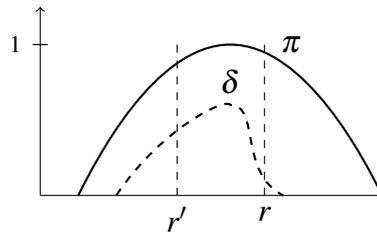


Fig. 3.8.C: Weakly comonotonic cloud

Figure 3.8: Illustration of (weakly) comonotonic and non-comonotonic clouds on the real line.

Note that, for clouds on the real line, we can define a weaker notion of comonotonicity: a (continuous) cloud $[\pi, \delta]$ is said to be *weakly* comonotonic if the sign of the derivative of distributions δ, π is the same in every point r of the real line \mathbb{R} . Being *weakly* comonotonic is not sufficient to be equivalent to a generalized p-box, since if π and δ are weakly comonotonic, then it is possible to find two values r and r' such that $\delta(r) < \delta(r')$ and $\pi(r) > \pi(r')$. In this case, the (pre-)ordering jointly induced by the two distributions is not complete, and the definition of comonotonicity given in section 3.2 is not satisfied. Figures 3.8.A, 3.8.B and 3.8.C respectively illustrate the notion of comonotonic, non-comonotonic and weakly comonotonic cloud on the reals. Figure 3.8.A illustrates a comonotonic cloud (and, consequently, a generalized p-box) for which elements are ordered according to their distance to the mode m (i.e., for this particular cloud, two values x, y in \mathbb{R} are such that $x <_{[\underline{F}, \overline{F}]} y$ if and only if $|m - x| > |m - y|$).

We can now extend the propositions linking clouds and generalized p-boxes with random sets. In particular, the following result extends Proposition 3.13 to the continuous case:

Proposition 3.16. *Let the distributions $[\pi, \delta]$ describe a continuous cloud on the reals and $\mathcal{P}_{[\pi, \delta]}$ be the induced credal set. Then, the random set defined by the Lebesgue measure on the unit interval $\alpha \in [0, 1]$ and the multimapping $\alpha \rightarrow E_\alpha$ such that*

$$E_\alpha = \{r \in \mathbb{R} | (\pi(r) \geq \alpha) \wedge (\delta(r) < \alpha)\} \quad (3.58)$$

describe a credal set $\mathcal{P}_{(m,\mathcal{F})_{[\pi,\delta]}}$ which is an inner approximation of $\mathcal{P}_{\pi,\delta}$ ($\mathcal{P}_{(m,\mathcal{F})_{[\pi,\delta]}} \subseteq \mathcal{P}_{\pi,\delta}$).

Proof. See Section D.2 in Appendix D □

The proof comes down to using sequences of discrete clouds outer- and inner-approximating $[\delta, \pi]$ and converging to it, and then to consider inner-approximations of those discrete clouds given by Proposition 3.14. This proposition has two corollaries:

Corollary 3.17. *Let $[\pi, \delta]$ be a comonotonic cloud with continuous distributions on the real line. Then the credal set $\mathcal{P}_{[\pi,\delta]}$ induced by $[\pi, \delta]$ is equivalent to the credal set induced by a continuous random set with uniform mass density, whose focal sets are of the form, for $\alpha \in [0, 1]$:*

$$E_\alpha = \{r \in \mathbb{R} | (\pi(r) \geq \alpha) \wedge (\delta(r) < \alpha)\} \quad (3.59)$$

To obtain the result, simply observe that the inner-approximation of Proposition 3.14 becomes exact for discrete comonotonic clouds, which are special cases of random sets. In particular, this is true for the sequences of discrete comonotonic clouds outer- and inner-approximating $[\pi, \delta]$ and converging to it. So, this sequence of random sets converge to a continuous random set at the limit. Another interesting particular case is the one of uniformly continuous p-boxes.

Corollary 3.18. *The credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$ described by two continuous cumulative distributions $\underline{F}, \overline{F}$ forming a regular p-box on the reals is equivalent to the credal set described by the continuous random set with uniform mass density, whose focal sets are sets of the form $[x(\alpha), y(\alpha)]$ where $x(\alpha) = \sup \{x \in \mathbb{R} | \overline{F}(x) < \alpha\}$ and $y(\alpha) = \inf \{x \in \mathbb{R} | \underline{F}(x) > \alpha\}$.*

This is because strictly increasing continuous p-boxes are special cases of comonotonic clouds (or, equivalently, of generalized p-boxes). To check that, in this case, $E_\alpha = [x(\alpha), y(\alpha)]$, it suffices to consider the possibility distributions $\pi_{\underline{F}}, \pi_{\overline{F}}$ and to see that $\inf_{r \in \mathbb{R}} (\pi_{\overline{F}}(r) \geq \alpha) = x(\alpha)$ and $\sup_{r \in \mathbb{R}} (1 - \pi_{\underline{F}}(r) < \alpha) = y(\alpha)$. The strict increasingness property can be relaxed to intervals where the cumulative functions are constant, provided one consider pseudo-inverses when building the continuous random set.

From a practical and computational perspective, these results are appealing. For example, they can facilitate the computation of lower and upper expectations over continuous generalized p-boxes (see, for examples, results of Section 6.2). Another interesting point is that all

the framework developed by Smets [186] concerning belief functions on reals can be applied to comonotonic clouds (generalized p-boxes). Results given in this section extend and give alternative proofs to some results given by Alvarez [5] concerning continuous p-boxes. A theoretical study of generalized p-boxes defined on a complete chain and of the associated set of finitely additive probabilities is given in Appendix F. Among other things, this study checks to which extent the results presented above extend to this more general setting. The main results of this study are that:

- such generalized p-boxes still induce lower probabilities (and previsions) that are ∞ -monotone, and can therefore be expressed through particular Choquet Integrals,
- they can be approximated by sequences of finite discrete degenerate p-boxes (somehow extending proposition 3.17),
- equivalence with random sets only hold for some practical subcases. This is mainly due to the fact that such extensions also consider confidence measures on non-measurable sets.

3.4.3 Thin continuous clouds

In the case of thin clouds, constraints defining the credal set, given by Equation (3.35), reduce to $P(\pi(x) \geq \alpha) = P(\pi(x) > \alpha) = 1 - \alpha$ for all $\alpha \in (0, 1)$. Recall that on a finite space \mathcal{X} these constraints are generally conflicting, because for some α , $P(\{x \in \mathcal{X} | \pi(x) \geq \alpha\}) > P(\{x \in \mathcal{X} | \pi(x) > \alpha\})$ will hold. When the *thin* cloud is a distribution defined on the real line, this is no longer a difficulty, and the following proposition holds:

Proposition 3.19. *If π is a continuous possibility distribution on the real line, then its credal set $\mathcal{P}_\pi \cap \mathcal{P}_{1-\pi}$ is not empty.*

Proof of Proposition 3.19. Let $F(x) = \Pi((-\infty, x])$, with $x \in \mathbb{R}$. F is the distribution function of a probability measure P_π such that for all $\alpha \in [0, 1]$, $P_\pi(\{x \in \mathbb{R} | \pi(x) > \alpha\}) = 1 - \alpha$, where the sets $\{x \in \mathbb{R} | \pi(x) > \alpha\}$ form a continuous nested sequence (see [78] p. 285). Such a probability lies in \mathcal{P}_π . Moreover,

$$P_\pi(\{x \in \mathbb{R} | \pi(x) > \alpha\}) = P_\pi(\{x \in \mathbb{R} | \pi(x) \geq \alpha\})$$

due to uniform continuity of π . We also have

$P_\pi(\{x \in \mathbb{R} | \pi(x) > \alpha\}) = 1 - \Pi(\{x \in \mathbb{R} | \pi(x) \geq \alpha\}^c) = 1 - \Delta(\{x \in \mathbb{R} | \pi(x) \geq \alpha\})$ again due to

uniform continuity. Since

$$1 - \Delta(\{x \in \mathbb{R} | \pi(x) \geq \alpha\}) = \sup_{\{x | \pi(x) \geq \alpha\}} 1 - \pi(x), \text{ this means } P_\pi \in \mathcal{P}(1 - \pi). \quad \square$$

A *thin* cloud is a particular case of comonotonic cloud. It induces a complete pre-ordering on the reals. If there are no ties, meaning that this pre-order is linear, it means that for any $\alpha \in [0, 1]$, there is only one value $r \in \mathbb{R}$ for which $\pi(r) = \alpha$, and that $\mathcal{P}_\pi \cap \mathcal{P}_{1-\pi}$ contains only one probability measure. In particular, if the order is the natural order of real numbers, this *thin* cloud reduces to a usual cumulative distribution.

When the pre-order has ties, it means that for some $\alpha \in [0, 1]$, there are several values in $r \in \mathbb{R}$ such that $\pi(r) = \alpha$. Using Corollary 3.17, we can model the credal set $P_\pi \in \mathcal{P}(1 - \pi)$ by the random set with uniform mass density, whose focal sets are of the form

$$E_\alpha = \{r \in \mathbb{R} | \pi(r) = \alpha\} \quad (3.60)$$

In this case, we can check that the belief measure of this random set is in accordance with Equation (3.35), since $Bel(\{r \in \mathbb{R} | \pi(r) \geq \alpha\}) = 1 - \alpha$ for any $\alpha \in [0, 1]$.

Finally, consider the specific case of a thin cloud modeled by an unimodal distribution π (formally, a fuzzy interval). In this case, each focal set associated to a value α is a doubleton $\{x(\alpha), y(\alpha)\}$ where $\{x | \pi(x) \geq \alpha\} = [x(\alpha), y(\alpha)]$. Noticeable probability distributions that are inside the credal set modeled by such a *thin* cloud are the cumulative distributions F_+ and F_- such that for all α in $[0, 1]$, $F_+^{-1}(\alpha) = x(\alpha)$ and $1 - F_-^{-1}(\alpha) = y(\alpha)$ (they respectively correspond to the case where the mass density of the random set is concentrated on values $x(\alpha)$ and $y(\alpha)$). All probability measures with cumulative functions of the form $\lambda \cdot F_+ + (1 - \lambda) \cdot F_-$ also belong to the credal set (for $\lambda = \frac{1}{2}$, this distribution corresponds to the case where mass density is evenly divided between elements $x(\alpha)$ and $y(\alpha)$). Other distributions inside this set are considered by Dubois *et al.* [78].

3.5 Combinations of uncertainty representations into higher order models

In the preceding sections, we have presented, studied and linked a number of uncertainty representations: sets, possibility distributions, probability distributions, generalized p-boxes, clouds, probability intervals, random sets, credal sets.

In applications, it is common to encounter combinations of uncertainty representations studied in this work. For example, a random set can be seen as a probability distribution bearing over sets (i.e., as a combination of probability distributions and sets). In this section, we give a brief review of the literature about such models. We then give more attention to fuzzy random variables and recall their basics.

3.5.1 A quick review of the literature

The combinations mentioned above are often referred as hierarchical models in the literature. Such hierarchical models arise when there is a 'correct' or 'ideal' uncertainty model on some variable X , but there is some uncertainty concerning what this 'ideal' first order model actually is. This uncertainty is itself modeled by a so-called 2^{nd} order model. It is not difficult to conceive how $3^{rd}, 4^{th}, \dots$ order models can be defined likewise. However, in practice, 2^{nd} order models are by far the most often used, and the higher the level of a model, the more difficult it is to give it a meaningful interpretation. In this section, we only do a short and non-exhaustive review of existing hierarchical models related to representations studied in previous sections:

- Bayesian 2^{nd} order models are for sure the most commonly encountered hierarchical models. In this case, both the 1^{st} and 2^{nd} order models are precise probabilities. Robustness studies [16] or combination of expert opinions [28] are two practical examples where Bayesian hierarchical models are often used. A common example of such a model consists in a random variable following a normal distribution whose uncertainties on the mean and variance are themselves modeled by normal distributions.
- Desirability functions, introduced by de Cooman [44] and based on Walley's [203] theory of imprecise probabilities, constitute a very general 2^{nd} order model where both the 1^{st} and 2^{nd} order models are allowed to be credal sets. This model encompasses, among other representations, fuzzy random variables, type-2 fuzzy sets, Bayesian models, imprecise probabilities. Note that this model is less general than the one defined by Walley [203, CH.5.], since desirability functions only allows credal sets induced by coherent lower probabilities at the 2^{nd} level.
- Fuzzy random variables [221, 168, 32, 45], which are formally equivalent to (normalized) fuzzy belief structures [214, 217, 83], also constitute popular models. Although they've been originally interpreted as a 1^{st} order model by Puri and Ralescu [168], they

can also be given a 2^{nd} order interpretation [32]. This 2^{nd} order model is, for example, used in so-called hybrid uncertainty propagation techniques, studied in Section 7.2.1 (see Baudrit *et al.* [12] and Cooper *et al.* [29]). They also come up as the result of an information fusion process studied in Section 4.2. They can also be instrumental in statistical inferences [20]

- Sets of random sets are considered by Augustin [7], who call them generalized basic probability assignment, and by Miranda *et al.* [146], who link this model to Walley's [203] lower previsions by giving them a behavioral interpretation. Instead of considering a precise basic probability assignment over sets, this model considers (closed convex) sets of precise basic probability assignments. Although Augustin [7] shows that, formally, it is mainly equivalent to 1^{st} order credal sets, this model is of practical importance, as assigning imprecise weights to subsets can be a convenient way to elicit and represent beliefs. Denoeux [58] studies, under the TBM interpretation (see Appendix A), interval-valued probability basic assignments over random sets, which can be seen as a special case of the model considered by Augustin [7].
- Type-2 fuzzy sets [218, 143] are other examples of 2^{nd} order models, where the 2^{nd} order models are fuzzy sets and the 1^{st} order models are interval-valued fuzzy sets. Note that, if we interpret the 2^{nd} order model as a possibility distribution and the 1^{st} order model as a cloud, then previous sections indicate us how Type-2 fuzzy sets could be interpreted in terms of imprecise probabilities. Evaluating the potential of Type-2 fuzzy sets as imprecise probabilistic models is an open question, and is out of the scope of this work.
- Denoeux [59] also studies, under the TBM interpretation, the case of fuzzy-valued imprecise fuzzy belief structures, where uncertainty on the basic probability assignments, bearing themselves on fuzzy sets, is described by means of fuzzy sets. To each levels $\alpha \in [0, 1]$ of the fuzzy sets describing the uncertainty on the basic probability assignments, he associates a set of random (fuzzy) sets described by interval-valued probability basic assignments.

In the next section, we give more details about the basics of a hierarchical model that will be used in the sequels of this work and, as noticed above, appears in practical treatments of uncertainty, namely so-called fuzzy random variables or fuzzy belief structures.

3.5.2 Fuzzy random variables

Let \mathcal{X} be the space⁴ on which variable X can take its values. A random fuzzy variable on \mathcal{X} is here defined as a (discrete) mapping $m : \wp(\widetilde{\mathcal{F}}_{\mathcal{X}}) \rightarrow [0, 1]$ where $\wp(\widetilde{\mathcal{F}}_{\mathcal{X}})$ is the set of all fuzzy sets \widetilde{F} that can be defined on \mathcal{X} and such that $\sum_{\widetilde{F} \in \wp(\widetilde{\mathcal{F}}_{\mathcal{X}})} m(\widetilde{F}) = 1$. We assume here the number of fuzzy focal elements to be finite. This definition is equivalent to a basic probability assignment assigned over fuzzy sets rather than on simple sets. For this reason, we denote $(m, \widetilde{\mathcal{F}})$ a random fuzzy variable, with $\widetilde{\mathcal{F}} = \{\widetilde{F} \in \wp(\widetilde{\mathcal{F}}_{\mathcal{X}}) | m(\widetilde{F}) > 0\}$ the set of fuzzy sets having a (strictly) positive mass, and call elements of this set *focal elements* of the random fuzzy variable.

3.5.2.1 Interpretation as a 1st order model

We consider a fuzzy random variable $(m, \widetilde{\mathcal{F}})$ of n focal elements defined on a finite space \mathcal{X} and an (arbitrary) indexing \widetilde{F}_i of focal elements ($i = 1, \dots, n$). Let $\{\gamma_0 = 0 < \gamma_1 < \dots < \gamma_M = 1\}$ be the range $\{\widetilde{F}_i(x) | x \in \mathcal{X}, i = 1, \dots, n\}$ of distinct values taken by the focal elements. The 1st order model corresponding to this fuzzy random variables is the random set which, for $i = 1, \dots, n$ and $j = 1, \dots, M$, have the following focal sets E_{i,γ_j}

$$\begin{cases} E_{i,\gamma_j} = \{x \in \mathcal{X} | \widetilde{F}_i(x) \geq \gamma_j\} \\ m(E_{i,\gamma_j}) = m(\widetilde{F}_i)(\gamma_j - \gamma_{j-1}) \end{cases} \quad (3.61)$$

Plausibility and belief measures on any event $A \subseteq \mathcal{X}$ are then given by:

$$Pl_{(m, \widetilde{\mathcal{F}})_1}(A) = \sum_{\substack{i=1, \dots, n \\ j=1, \dots, M}} m(E_{i,\gamma_j}) \mathbf{1}_{(E_{i,\gamma_j} \cap A \neq \emptyset)} = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \sum_{j=1, \dots, M} (\gamma_j - \gamma_{j-1}) \mathbf{1}_{(E_{i,\gamma_j} \cap A \neq \emptyset)}$$

$$Bel_{(m, \widetilde{\mathcal{F}})_1}(A) = \sum_{\substack{i=1, \dots, n \\ j=1, \dots, M}} m(E_{i,\gamma_j}) \mathbf{1}_{(E_{i,\gamma_j} \subseteq A)} = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \sum_{j=1, \dots, M} (\gamma_j - \gamma_{j-1}) \mathbf{1}_{(E_{i,\gamma_j} \subseteq A)}$$

or, equivalently, by

$$Pl_{(m, \widetilde{\mathcal{F}})_1}(A) = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \int_0^1 \mathbf{1}_{(E_{i,\gamma} \cap A \neq \emptyset)} d\gamma = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \sup_{x \in A} \widetilde{F}_i(x) \quad (3.62)$$

$$Bel_{(m, \widetilde{\mathcal{F}})_1}(A) = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \int_0^1 \mathbf{1}_{(E_{i,\gamma} \subseteq A)} d\gamma = \sum_{i=1, \dots, n} m(\widetilde{F}_i) \sup_{x \in A^c} \widetilde{F}_i(x) \quad (3.63)$$

⁴We assume this space is finite or measurable

with $\bar{P}_{\tilde{F}_i}$ and $\underline{P}_{\tilde{F}_i}$ the upper and lower probabilities of the credal set $\mathcal{P}_{\tilde{F}_i}$ induced by the focal element \tilde{F}_i .

We can also link a fuzzy random variable to a continuous random set. In order to do this, let us define, for all $i = 1, \dots, n$, the value α_i as $\alpha_i = \sum_{j=1}^i m_j$, with $\alpha_0 = 0$. Let us now consider a fuzzy random variable $(m, \tilde{\mathcal{F}})$ and the associated continuous random set $(m, \tilde{\mathcal{F}})_c$ defined by a uniform density m over $[0, 1]$ and the mapping

$$\forall \alpha \in (\alpha_i, \alpha_{i+1}], i = 1, \dots, n, \alpha \rightarrow E_\alpha = E_{i, \phi_i(\alpha)} \quad (3.64)$$

where $\phi_i(\alpha) = \frac{\alpha - \alpha_i}{m_i}$ is a linear function mapping $(\alpha_i, \alpha_{i+1}]$ to $(0, 1]$, and $E_{i, \phi_i(\alpha)}$ is the set given by Equation (3.59) of Corollary 3.17, for the focal element \tilde{F}_i of $(m, \tilde{\mathcal{F}})$ and for level $\phi_i(\alpha)$. We can then formulate the following proposition:

Proposition 3.20. *Let $(m, \tilde{\mathcal{F}})$ be a fuzzy random variable and $(m, \tilde{\mathcal{F}})_c$ the associated continuous random set $(m, \tilde{\mathcal{F}})_c$ on \mathcal{X} . Then $(m, \tilde{\mathcal{F}})_c$ is equivalent to the 1st order interpretation of $(m, \tilde{\mathcal{F}})$, and we have, For all events $A \subseteq \mathcal{X}$:*

$$Bel_{(m, \tilde{\mathcal{F}})_c}(A) = Bel_{(m, \tilde{\mathcal{F}})_1}(A) \quad \text{and} \quad Pl_{(m, \tilde{\mathcal{F}})_c}(A) = Pl_{(m, \tilde{\mathcal{F}})_1}(A)$$

where $Bel_{(m, \tilde{\mathcal{F}})_c}, Pl_{(m, \tilde{\mathcal{F}})_c}$ are the belief and plausibility measures of $(m, \tilde{\mathcal{F}})_c$, and $Bel_{(m, \tilde{\mathcal{F}})_1}, Pl_{(m, \tilde{\mathcal{F}})_1}$ are the belief and plausibility measures given by Equations (3.63) and (3.62).

Proof. Let us note $\beta_i = \phi_i(\alpha)$ and notice that $d\alpha = m_i d\beta_i$ for all $i = 1, \dots, n$. Then,

$$\begin{aligned} Pl_{(m, \tilde{\mathcal{F}})_c}(A) &= \int_0^1 \mathbf{1}_{(A \cap E_\alpha \neq \emptyset)} d\alpha = \sum_{i=1}^n \int_{\alpha_{i-1}}^{\alpha_i} \mathbf{1}_{(A \cap E_{i, \phi_i(\alpha)} \neq \emptyset)} d\alpha \\ &= \sum_{i=1}^n m_i \int_0^1 \mathbf{1}_{(A \cap E_{i, \beta_i} \neq \emptyset)} d\beta_i = \sum_{i=1, \dots, n} m(\tilde{F}_i) \underline{P}_{\tilde{F}_i}(A) \\ &= \sum_{i=1, \dots, n} m(\tilde{F}_i) \sup_{x \in A^c} \tilde{F}_i(x) \end{aligned}$$

which is equivalent to Equation (3.62). Since plausibility and belief measures are dual measures, the proof is finished. \square

The above discussion shows that, as a 1st order level, a fuzzy random variable can be viewed as:

- a regular random set, either finite (Equation (3.61)), or continuous (see Mapping (3.64))

- a convex combination of the focal elements \tilde{F}_i with their associated weights m_i (Equations (3.63) and (3.62)).

Conversely, Proposition 3.20 indicates that any random set can, in principle, be mapped into a corresponding fuzzy random variable (eventually degenerated into a classical random set when no focal elements are nested).

Equations (3.62) and (3.63) coincide with Yen's [217] definition of the plausibility and belief measures for fuzzy belief structures (Note that Yen's proposition is based on optimization criteria, and not on some imprecise probabilistic derivation). We can also notice that they're coherent with Smets definition of a fuzzy event [182] (they reduce to Smets definition when focal sets are crisp). The above generalizations of plausibility and belief measures to fuzzy random variables (or to fuzzy belief structures) have thus strong theoretical justifications and appear to us as the best choice. Nevertheless, other choices are possible: Denoeux [59], following Zadeh [220], prefers a generalization of plausibility and belief measures based on min – max operators, on the basis that this generalization is less sensible to small changes in the fuzzy focal elements⁵. Yager [212] also proposes to replace the min – max operators by any combination of t-norms and t-conorms.

This 1st order interpretation of fuzzy random variables corresponds to the interpretation of Puri and Ralescu [168]. We now briefly describe other interpretations of fuzzy random variables as 2nd order models, and explore the links between such interpretations and the 1st order interpretation of this section.

3.5.2.2 Interpretations as a 2nd order model

A fuzzy random variable $(m, \tilde{\mathcal{F}})$ defined on \mathcal{X} is equivalent to a basic probability assignment over a set of fuzzy sets. There are two ways of interpreting it as a 2nd order model: either we consider that the basic probability assignment is the 2nd order level model, and the 1st order level model is then a fuzzy set known with uncertainty, or we consider that our 2nd order level model is a fuzzy set, describing our uncertainty about what could be the 1st order representation.

First interpretation Let us consider an arbitrary indexing \tilde{F}_i , $i = 1, \dots, n$ of the focal elements in $\tilde{\mathcal{F}}$ and $m_i = m(\tilde{F}_i)$ the associated basic probability assignments. Then, the first

⁵This is not surprising, since the generalization proposed here is based on linear operators, which are more sensible than maxitive ones

interpretation of $(m, \widetilde{\mathcal{F}})$ as a 2^{nd} order model means that we consider that there is a probability m_i that our first order model describing uncertainty on a variable X assuming values in \mathcal{X} is \widetilde{F}_i (nothing more, nothing less), inducing the corresponding credal set $\mathcal{P}_{\widetilde{F}_i}$. In this case, equations (3.63) and (3.62) indicate that the lower and upper probabilities of the 1^{st} order model can be seen as the mean over m of the lower probabilities induced by the 1^{st} order models of this interpretation. In the case of random sets, the first order model reduce to a classical set.

Second interpretation The second interpretation means that the nested sets on which is defined a fuzzy set become credal sets rather than sets of values. If the first order model is defined over \mathcal{X} , the second order fuzzy set consists in lower (or, equivalently, upper) confidence bounds given to collection of nested credal sets. We then retrieve the usual notion of 2^{nd} order possibility distribution defined over credal sets. Given a fuzzy random variable $(m, \widetilde{\mathcal{F}})$, to a cut of level α of this 2^{nd} order possibility distribution then corresponds the credal set $\mathcal{P}_{(m, \widetilde{\mathcal{F}})_\alpha}$, to which is given a confidence degree $1 - \alpha$ and that is induced by the following random set (see also Equation (3.61)):

$$\left\{ \begin{array}{l} E_{i,\alpha} = \{x \in \mathcal{X} | \widetilde{F}_i(x) \geq \alpha\} \\ m(E_{i,\alpha}) = m(\widetilde{F}_i) \end{array} \right.$$

This interpretation of fuzzy random variables has recently received a lot of attention from various authors, and we refer to De cooman [45], Walley [205], De Cooman and Walley [53], Couso and Sanchez [34], Baudrit *et al.* [10] for theoretical discussions of this model, and Ferson and Ginzburg [102] and Baudrit *et al.* [12] for more practical considerations.

3.6 Conclusions and perspectives

We've seen above that there exists a whole range of uncertainty representations, going from very general models (e.g. credal sets) to more specific ones (e.g. possibility distributions). More general models (random sets, credal sets), by their higher expressiveness and generality, have a unifying power that is used here to relate the different models. Less general models have the advantages of being easier to handle than more general ones and are of great practical importance for elicitation tasks and to make uncertainty treatments more tractable (and thus applicable to more problems). At the very least, when resources are limited, they can provide

quick gross approximation that can, in some cases, be sufficient to provide satisfying answers. It also often happens that available information can be modeled by simpler representations, and in these cases there is no need to use more complex models. This is why we think that the study and evaluation of such simpler models is of great importance, especially in applications. With respect to this objective, the main results of this section can be summarized as follows:

- The introduction and study of a generalized form of p-box that bridges the gap between possibility distributions and p-boxes, while remaining a special case of random sets. This model looks promising for two reasons: first, its interpretation in terms of lower and upper confidence bounds on nested sets makes it attractive for elicitation procedures, and second, it can be a good compromise between possibility distributions and random sets, offering more expressiveness than the former and having a higher computational tractability than the latter.
- The study of clouds, that have been related to other uncertainty representations, a study that was only partially done by Neumaier [159]. In particular, we have shown that comonotonic clouds are equivalent to generalized p-boxes, and that non-comonotonic clouds are likely to be less interesting representations, at least from a practical point of view.

As practical perspectives to this section, we can cite:

- the potential usefulness of clouds and of generalized p-boxes still largely remains to be evaluated, however promising these models may look. In particular, computational aspects have to be explored, similarly to what De campos *et al.* [42] have done for imprecise probabilistic assignments. Section 3.2.5 provides first results of such an exploration, for the specific problem of uncertainty propagation. Investigating other computational aspects (i.e. information fusion) is partially done in subsequent chapters, and is part of our on-going research.
- The instrumentality of clouds and generalized p-boxes in elicitation procedures also has to be evaluated, possibly using some previous studies done with possibility distributions [170].
- Practical applications of clouds and generalized p-boxes. In this perspective, a first application is done by Fuchs and Neumaier [111], where they use multi-dimensional comonotonic clouds in a robustness study.

More theoretical perspectives include, but are not restricted to:

- The evaluation of type-2 fuzzy sets as (2^{nd} order) imprecise probabilistic models, by using results related to clouds, which are formally equivalent to interval-valued fuzzy sets.
- The study of clouds and generalized p-boxes in more general settings, for instance by interpreting them through Walley's [203] lower previsions and by considering finitely additive probabilities on infinite spaces. Appendix D, where generalized p-boxes defined on complete chains are studied, shows that results holding on finite spaces do not always extend to more general cases. The extension of such results is also the topic of on-going research work, done in collaboration with Enrique Miranda and Matthias Troffaes.
- The study of an "imprecise possibility theory", where sets of possibility distributions dominating an uncertainty measure are considered rather than single possibility distributions. Such a study seems natural to us, and in the spirit of Augustin's [7] generalized basic probability assignments (sets of random sets) and of Walley's [203] lower previsions (sets of probabilities). Moreover, some of the results of this chapter (for example, Propositions 3.10, 3.3 and Corollary 3.8) suggest that such a study could be useful.
- The general characterization of non-empty credal sets representable by pairs of possibility distributions. We have seen that clouds do not cover all such cases, and that some subcases of clouds remain unexplored. In particular, it would be useful to check under which conditions such credal sets induce ∞ -monotone lower probabilities (i.e. random sets), and when it is not the case, if the induced lower probability is always not 2-monotone (as it is in Proposition 3.13)

Material contained in this chapters can be found in papers [67, 74, 70, 71, 69, 68, 64]

Chapter 4

Treating multiple sources of information

“When you have eliminated the impossible, whatever remains, however improbable, must be the truth”

— Conan Doyle (1859–1930)

Contents

| | | |
|------------|---|------------|
| 4.1 | Basics of Information fusion in uncertainty | 121 |
| 4.1.1 | A classification of fusion operators | 121 |
| 4.1.2 | Mathematical properties of fusion operators | 122 |
| 4.1.3 | Basic fusion operators in uncertainty theories | 127 |
| 4.1.3.1 | Probability theory | 128 |
| 4.1.3.2 | Imprecise probability theory | 130 |
| 4.1.3.3 | Random set theory | 133 |
| 4.1.3.4 | Possibility theory | 136 |
| 4.1.3.5 | Links between basic rules | 138 |
| 4.1.3.6 | Fusion rules for clouds | 141 |
| 4.2 | Treating the conflict by adaptive rules using maximal coherent subsets . | 147 |
| 4.2.1 | Maximal coherent subsets (MCS) rule: basic methodology | 149 |
| 4.2.1.1 | MCS in imprecise probability theories. | 149 |
| 4.2.1.2 | Maximal coherent subsets (MCS) in practice. | 151 |
| 4.2.2 | Level-wise MCS on the real line with possibility distributions | 152 |
| 4.2.2.1 | Extracting MCS of intervals | 153 |
| 4.2.2.2 | Level-wise MCS on possibility distributions | 154 |
| 4.2.2.3 | Exploiting the fuzzy belief structure | 157 |

| | | |
|------------|---|------------|
| 4.2.2.4 | Taking additional information into account | 161 |
| 4.3 | Towards a cautious conjunctive rule in random set theory | 166 |
| 4.4 | Assessing sources quality: a general framework | 172 |
| 4.4.1 | Rational requirements and general methodology | 172 |
| 4.4.2 | Evaluating sources in probability | 173 |
| 4.4.3 | Evaluating sources in uncertainty theories | 176 |
| 4.4.3.1 | Informativeness | 178 |
| 4.4.3.2 | Calibration | 179 |
| 4.4.3.3 | Global scores | 180 |
| 4.5 | Conclusions and perspectives | 181 |

In practical applications, it often happens that multiple sources (experts, sensors, databases, . . .) provide information about the value that could assume a variable X (e.g. model parameter, physical quantity, . . .). In such a situation, two different but correlated problems are (i) the construction of a synthetic interpretable model or message from these multiple pieces of information, and (ii) the evaluation of the information quality. In this work, we restrict ourselves to cases where uncertainty is represented numerically, and where it bears on "objective" variables, that is variables whose value is, at least in principle, observable (i.e., we do not consider the problems of reaching a subjective consensus or of aggregating preferences)

The construction of a synthetic model is commonly referred to as information fusion. It consists in aggregating information coming from different sources into a single representation. This representation is then easier to use or analyze in subsequent treatments than scattered information. In general, this representation is required to be both informative and more reliable than information taken from a single source. If these two goals are relatively easy to achieve when information brought by the sources is consistent, this is not the case when inconsistencies between sources occur. Moreover, the fact that sources can be dependent (e.g. two experts having common backgrounds), more or less reliable, biased, . . . added to the fact that information can come in a great variety of shapes makes the problem even more difficult to solve. Two of the most important problems encountered in information fusion are to cope with conflicting information and to deal with potentially dependent sources. After a brief reminder of basics of information fusion in uncertainty theories (Section 4.1), we provide in this chapter some results regarding these two problems. To deal with inconsistencies in the information, we propose to use the logical notion of maximal coherent subsets (MCS) [173] (Section 4.2). This notion appears as a very natural way to conciliate the two objectives of gaining information and of remaining in agreement with all sources. Regarding the problem of dependent sources, we study, inside random set theory, the merging of sources whose dependencies are

badly known. Some preliminary results are given, that should eventually lead to a practical cautious merging rule (Section 4.3).

Assessing the quality of the information delivered by a source is another recurrent problem when receiving information from multiple sources. When multiple sources provide information, it is natural to wonder which are the more reliable among them. If subjective assessments of reliability are admissible in some situations, they are not in industrial risk analysis or safety studies. We thus propose in this chapter methods to assess sources reliability from previous situations where the same sources have given information about some other variables (Section 4.4). These methods are inspired from probabilistic and possibilistic evaluation methods respectively proposed and discussed by Cooke [28] and Sandri *et al.* [174]. Once the sources reliability is known, it can serve various purposes (be used in information fusion process, in code validation procedures [196], ...).

4.1 Basics of Information fusion in uncertainty

There are mainly two ways to characterize a fusion operator: by its behavior or by the properties it satisfies. In Section 4.1.1, we introduce a classification of fusion operators which draws from previous works of Bloch [18] and of Dubois and Prade [93, 86]. In Section 4.1.2, the main properties that a fusion operator can satisfy are recalled. Section 4.1.3 studies the main operators used in the different imprecise probability theories considered in this work (i.e., possibility theory, random sets, imprecise probabilities), and gives a brief account of their links.

All along this section, φ will denote a general fusion operator, and we assume sources provide their information about a variable X taking its values on a domain \mathcal{X} . We also note $\llbracket N \rrbracket$ the subset $\{1, \dots, N\}$ of natural numbers from 1 to N .

Since the formal definitions of fusion rules and properties differs between uncertainty theories, we start by informal formulations of such behaviors and properties, and then to provide formal definitions within each of the uncertainty theories considered in this work (i.e. possibility theory, random set theory, imprecise probability theory).

4.1.1 A classification of fusion operators

Fusion operators can follow three main kinds of behaviors:

- **conjunctive:** a *conjunctive* operator is the counterpart to a set intersection. The uncertainty in the information resulting from a *conjunction* is less than the uncertainty of each source alone. A conjunctive operator makes the assumption that **all** sources are reliable, and usually results in very precise information. If there is some conflict in the information (i.e. at least one source is not fully reliable), then the result of the conjunction can be poorly reliable, or even empty.
- **disjunctive:** a *disjunctive* operator is the counterpart to a set union. The uncertainty in the information resulting from a *disjunction* is higher than the uncertainty of all sources together. A disjunctive operator makes the assumption that **at least one** source is reliable. The result of a disjunctive operator can be considered as very reliable, but is also often (too) poorly informative.
- **Trade-off:** the result of a *trade-off* operator lies between conjunctive and disjunctive behaviors. Trade-off operators are typically used when sources are partially conflicting. As its name indicates, It tries to make a trade-off between disjunction and conjunction so as to achieve a good balance between informativeness and reliability. In the following, we will distinguish two compromise operators:
 - **adaptive:** an operator will be called *adaptive* if the fusion result depends on the context (Bloch [18] calls them context-dependent). The aim of such an operator is to go from a *conjunctive* behavior to a *disjunctive* behavior as the conflict (disagreement) among sources increases. Adaptive operators reduce to disjunction (conjunction) in case of total conflict (agreement) between sources. In-between, they act as trade-off operators. Methods using maximal coherent subsets studied in Section 4.2 are good representatives of adaptive operators.
 - **non-adaptive:** a compromise operator is *non-adaptive* when it always acts the same way, irrespectively to the context. Arithmetic weighted mean (or convex combination) is a popular and typical example of such operators, and is undoubtedly the most widely used compromise operator in practice.

4.1.2 Mathematical properties of fusion operators

It is common to require a fusion operator φ or the result of the fusion process to satisfies some mathematical properties, considered as desirable by the analyst. In this section, we assume and discuss the main properties encountered in the literature. The proposed list heavily

draws on the works of Walley [202] in the framework of imprecise probabilities and of Ousalah [161] in the framework of possibility theory. We will thus use their terminologies in the sequel. To fix ideas, let O_1, \dots, O_N be the N uncertainty models representing sources opinions, indexed in an arbitrary way and with O_i representing the information provided by source i . We denote $O_{(i:j)} := \varphi(O_i, \dots, O_j)$ the result of applying the fusion operator to representations i to j . Concepts of empty result; inclusion; convexity and convex hull; total and partial conflict; total and partial consistency that we use to define some properties will then be declined for each uncertainty theory.

- Property I) Total Consistency (Cons.):** this property is satisfied when the core (Equation 3.2) of the lower uncertainty measure induced by the result $O_{(1:N)}$ is non-empty. Total Consistency comes down to eliminate any conflict in the result, either by the fusion process itself or by removing it by post-processing. This property is commonly required when subsequent inferences or decisions have to be made. As total consistency comes down to eliminate the existing conflict in the information, this property is not desirable if we want to keep track of it (this is the case, for example, with the Transferable Belief Model [189]).
- Property II) Associativity / Aggregation of aggregates [202] (Ass.):** φ is associative if $\varphi(O_1, \dots, O_N) = \varphi(\varphi(O_1, \dots, O_{N-1}), O_N)$. Associativity allows information to be fused step-wise (first O_1 with O_2 , then the result with O_3 , etc.). This property is important, for instance, if sources provide information at different times and individual pieces of information cannot be stored while waiting for all sources to have delivered information. Otherwise, we do not regard associativity as important, not to say appealing, as it can drastically limits the range of possible fusion operators (see Walley [202]). Note that associativity is also convenient in computations.
- Property III) Commutativity / Symmetry [202] (Comm.):** φ is commutative if $\varphi(O_1, \dots, O_N) = \varphi(O_{\sigma(1)}, \dots, O_{\sigma(N)})$, with σ a permutation of $\llbracket N \rrbracket$. We regard commutativity as essential when there is no information allowing to distinguish the sources between them.
- Property IV) Idempotence (Idem.) :** φ is idempotent if $\varphi(O, O) = O$. The property of idempotence can be regarded as a cautious attitude with respect to the potential dependency between sources. If all sources provide the same information, then idempotence ensures that there will be no reinforcement effect, that is the likelihood of values that sources consider more likely to occur is not reinforced.

- Property V)** Weak zero preservation [161] / Unanimity [202] (**W-Z-P**): φ satisfies this property if an element is considered impossible by the fusion result when it is so by all sources. This property ensures that the fusion operator only accounts for sources information and do not add implicit information in the result, making some values suddenly possible without any source supporting this opinion. This property seems compelling to us, and should be required. This property is called *strong* zero preservation by Oussalah [161]¹, but we choose to call it weak, since it puts less constraints on the result than its (here) strong counterpart.
- Property VI)** Strong zero preservation [161] / Conjunction [202] (**S-Z-P**): φ satisfies this property if an element is considered impossible by the fusion result when it is so by at least one source. Requiring this property enforces φ to have a conjunctive behavior, and only makes sense when sources agree together. It can be considered as dictatorial, as one source can make a value impossible.
- Property VII)** Weak maximal plausibility [161] / Indeterminacy [202] (**W-M-P**): φ verifies this property if an element considered possible by all sources is also considered possible by the fusion result. This property is trivially satisfied if sources are totally conflicting (as there is no elements considered possible by all sources). In other cases, this property ensures a minimal consistency with the information given by the sources, and we thus consider it as desirable.
- Property VIII)** Strong maximal plausibility [161] / Total reconciliation [202] (**S-M-P**): φ satisfies this property if an element considered possible by at least one source must also be considered possible by the fusion result. Requiring strong maximal plausibility enforces φ to have a disjunctive behavior, and the fusion result to be fully consistent with the information delivered by every source. It is very strong, and not desirable if we want to have the possibility to reduce uncertainty via the fusion process.
- Property IX)** Information relevance [163] / Reconciliation and Strong Reconciliation [202] (**Recon.**): φ satisfies reconciliation property if the conjunction of $O_{(1:N)}$ with any representation O_i , $i \in \llbracket N \rrbracket$ is non-empty (i.e. if $O_{(1:N)}$ is partially consistent with every sources). It satisfies strong reconciliation property if, for any subset $I \subset \llbracket N \rrbracket$ such that the conjunction of representations O_i , $i \in I$ (whose result is denoted O_I) is non-empty, the conjunction of $O_{(1:N)}$ with O_I is not

¹Delmotte [55] also requires this property in the framework of possibility theory, and formulates it in terms of supports of possibility distributions

empty. Strong reconciliation requires the fusion result to be partially consistent with all partially consistent subsets of sources. Unless there are some reasons to discard some sources, strong reconciliation appears as a natural requirement.

Property X) Insensitivity to complete and relative ignorance [202] (**InsIgn.**): Consider $O_{(1:N)}$ and a $N + 1^{th}$ source supplying information. φ satisfies insensitivity to complete ignorance if $\varphi(O_{(1:N)}, O_{N+1}) = O_{(1:N)}$ when O_{N+1} represents complete ignorance. φ satisfies insensitivity to relative ignorance if O_{N+1} is less informative than all the other sources taken together (i.e. if it includes the disjunction of O_1, \dots, O_N).

Property XI) Convexity (**Conv.**): this property is satisfied if $O_{(1:N)}$ is convex when O_1, \dots, O_N are. In general, satisfying convexity increases computational tractability, but there are no obvious reasons to always require it. Note that if the final result of the fusion is not convex, it is always possible to take its convex hull (losing some information in the process, but gaining computational tractability).

Property XII) Robustness / Continuity [202] (**Robus.**) : φ satisfy robustness or continuity if a small change in the opinion of one source only causes a small change in $O_{(1:N)}$, a small change being usually defined with respect to a distance function d . Since a model of the available information can often be regarded as approximative, robustness is often considered as desirable, particularly when the fusion procedure is some part of an automatic process. Nevertheless, there is no obvious reasons to always require this property.

These properties are rather general and can be formulated for each uncertainty theory considered here, but other (more ad hoc) properties that generalize less easily can be considered with respect to a particular theory (e.g. sensitivity to the core in possibility theory [55], strong Pareto in imprecise probability theory [202], insensitivity to refinement in evidence theory [188]).

When building or choosing a fusion operator, a natural thing to do is to determine which properties it should follow. As the importance of properties can vary with the problem to solve, we can only provide some guidelines about which properties should or should not be satisfied. Our opinion is that satisfying Weak Zero Preservation (Property V, W-Z-P) and Weak Maximal Plausibility (Property VII, W-M-P) is a minimal requirement to ensure that the fusion operator do not add some extra information and that it does not discard values supported by all sources. If no information about the sources (e.g. reliability) allows to differentiate them, then

satisfying the properties of Commutativity (Property III, Comm.) and of Strong Reconciliation (Property IX, Recon.) should be required. We consider Insensibility to relative ignorance (Property X, InsIgn.) as strongly desirable, since if it is not verified, the fusion result is likely to quickly become too imprecise to be really useful. Strong zero preservation (Property VI, S-Z-P) is desirable only when all sources are partially consistent with each others. Conversely, Strong maximal plausibility (Property VIII, S-M-P) is desirable when all sources are totally conflicting, and that none can be discarded. Total consistency (Property I, Cons.) should be required when some inferences or decision must be drawn from information, but regarding it as a general minimal requirement is debatable, since it may be interesting to keep track of inconsistencies in the information. Idempotence (Property IV, Idem.) is desirable when no reinforcement effects are wanted (e.g., when sources cannot be considered independent). The main advantage of Associativity (Property II, Ass.) and of convexity (Property XI, Conv.) is to alleviate the computational burden, but we do not regard them as always desirable for a fusion operator. Although Robustness (Property XII, Robus.) is often required, we do not regard it as desirable, except in automatic procedures, where it is important to ensure a certain robustness.

The above argumentation is based on the rational basis that a fusion operator should aim at gaining information (Properties W-Z-P V, S-Z-P VI, InsIgn. X) while remaining coherent with the sources (Properties Recon. IX, W-M-P VII) and with the information we have (or do not have) about their relations (Properties Comm. III, Idem. IV). We consider that inconsistency is an information in itself, and thus should not be automatically removed (Property Cons. I), and that practical considerations only comes after rational ones (Properties Ass. II, Conv. XI, Robus. XII).

However, there is no real consensus among researchers about which properties should be regarded as essential, desirable, optional or not desirable, and in which context they should be so. For instance, Walley [202] regards total consistency (Property I) as fundamental, as it is central in his behavioral theory, while it is central in the Transferable Belief Model developed by Smets [189] to consider total consistency as optional as long as no decisions have to be taken. Denoeux [61] and Smets [188] often regard associativity (Property II) as strongly desirable, if not essential. In possibility theory, Oussalah *et al.* [163] and Delmotte [55] consider robustness (Property XII) as critical.

4.1.3 Basic fusion operators in uncertainty theories

In this section, we review the main basic fusion operators in the various imprecise probability theories considered in this work. We also give a word about their links and about basic fusion operators with generalized p-boxes, pursuing our study of this representation.

First, let us relate in a general way our classification of fusion operators given in Section 4.1.1 with properties of Section 4.1.2.

As said in the previous section, conjunctive operators are retrieved by requiring Strong Zero preservation VI. Such operators are sensible choices only when the result of conjunction is non-empty (i.e. all sources are partially consistent with each others). Conjunctive operators usually satisfy the properties of Convexity XI, Insensibility to Ignorance X, Weak Maximal Plausibility VII, Strong Reconciliation IX and obviously Weak Zero Preservation V. Total consistency I and Robustness XII are in general not satisfied (Total consistency when there is no conflict at all among sources). In any cases, they do not satisfy Strong Maximal Plausibility VIII. As said above, conjunctive operators should be chosen when sources are strongly consistent.

Disjunctive operators are retrieved by requiring Strong Maximal Plausibility VIII. They satisfy properties of Strong Reconciliation IX, Total Consistency I, Robustness XII and forcefully Weak Maximal Plausibility VII. They do not satisfy Properties of Insensibility to partial Ignorance X, Strong Zero Preservation VI and Weak Zero preservation V. In general, they also do not satisfy Convexity XI. Given the argumentation of Section 4.1.2 about the desirability of properties, disjunctive operators are, in general, to be discarded, except when all sources are totally conflicting with each others.

Adaptive trade-off operators generally satisfy properties of Total Consistency I, Insensibility to partial Ignorance X, Weak Zero Preservation V, Weak Maximal Plausibility VII and Strong Reconciliation IX. They do not, in general, satisfy Properties of Maximal Plausibility VIII, Strong Zero Preservation VI, Convexity XI and Robustness XII, and are very difficult to conciliate with Associativity II. They represent good solutions to the fusion problem, but can be limited by the potential computational burden associated to them.

Non-adaptive compromise operators generally satisfy properties of Total Consistency I, Weak Zero Preservation V, Weak Maximal Plausibility VII, Convexity XI, Robustness XII and Strong Reconciliation IX. They do not, in general, satisfy Properties of Strong Maximal Plausibility VIII, Strong Zero Preservation VI and Insensitivity to partial Ignorance X. The mean, a representative of non-adaptive compromise operators, is undoubtedly the most widely

and commonly used fusion operator. Given two numbers x, y and associated weights λ_x, λ_y , the general form of a mean² φ is given by the functional equation

$$\varphi(x, y, \lambda_x, \lambda_y) = f^{-1} \left(\frac{\lambda_x f(x) + \lambda_y f(y)}{\lambda_x + \lambda_y} \right) \quad (4.1)$$

where $f(x)$ is a continuous and strictly monotonic function, with f^{-1} its inverse. See Aczel [4, Sec.5.3.2] for a full characterization. Although mean operators can be convenient to aggregate subjective opinions or preferences of a group, we will argue in the sequel that they are not well fitted to the case when information concern a variable whose true value is not well known.

In the next sections, we review how are defined the main operators in probability theory, imprecise probability, random set theory and possibility theory, and how properties defined in Section 4.1.2 are particularized for each theory.

4.1.3.1 Probability theory

Let \mathcal{X} be the space of interest. In probability theory, N probability distributions p_1, \dots, p_N are built from the sources information, and the fusion result $P_{(1:N)} = \varphi(P_1, \dots, P_N)$ must be another probability distribution on \mathcal{X} . Two main approaches exist to solve the problem: the first concerns direct aggregation methods, using most of the time mean operators satisfying Equation (4.1), while the second uses a Bayesian approach to fuse probabilities. Cooke [28] provides a long and critical survey of the various methods. Shorter and critical reviews are provided by French [110], Genest and Zidek [113] and more recently by Clemen and Winkler [26]. Bayesian methods will not be considered here, since we regard them as more related to prior information revision and updating problems than to fusion problems. Examples of Bayesian methods are given by Winkler [211], who considers the use of natural-conjugate distributions³, Morris [155], Mosleh and Apostolakis [156], who mainly considers normal models, and Lindley [137].

For direct aggregation methods, Cooke [28, Ch.11] considers a family of operators that he calls r -norm, and which constitutes a subfamily of the general equation (4.1) where $f(x) = x^r$. Inside this family, only the arithmetic weighted mean ($r = 1$), often called linear pooling, and the geometric weighted mean ($r = 0$), often called logarithmic pooling, have appealing properties. If $\lambda_1, \dots, \lambda_N$ are non-negative weights summing to one, they respectively reads,

²In the sense that the result is between x and y

³Natural-conjugate distributions are families of likelihood and prior distributions such that the prior and posterior densities belongs to the same family, thus making calculations easier

for all $x \in \mathcal{X}$,

$$P_{(1:N)}(x) = \sum_{i=1}^N \lambda_i p_i(x) \quad (4.2)$$

and

$$P_{(1:N)}(x) = k \prod_{i=1}^N p_i(x)^{\lambda_i} \quad (4.3)$$

with k a normalization constant. The main arguments for using logarithmic pooling is that it is more coherent with Bayesian methods and it is the only operator having the so-called independence preservation property, an argument which has been criticized by many authors [28, 110] (See Genest *et al.* [112] for a study of logarithmic pooling). Linear opinion pooling is the only operator to possess the marginalization property, meaning that it does not matter whether we apply it to the distributions p_i or to the induced measures P_i (i.e., $P_{(1:N)}(A) = \sum_{x \in A} P_{(1:N)}(x) = \sum_{i=1}^n \lambda_i P_i(A)$). This means that the fused probability of an event depends only on the source probabilities for this event, and is not influenced by the probabilities of other events. See McConway [142] and Cooke [28] for more details. Walley [202] also shows that, inside a behavioral theory of (precise) probability, the arithmetic weighted mean is the only fusion procedure (including Bayesian ones) that satisfies the property of Weak Zero Preservation V. All these arguments show that, inside a probabilistic approach, the arithmetic weighted mean appears as the most appealing and sensible operator to fuse probabilities.

There are no disjunctive or conjunctive operators in probability theory, only trade-off operators, although some authors [191] consider the arithmetic mean as disjunctive and the normalized product as conjunctive. However, neither the arithmetic mean nor the normalized product do respectively satisfy Strong Maximal Plausibility VIII and Strong Zero Preservation VI. Moreover, if μ_1, μ_2 denote two additive probabilities over \mathcal{X} , we have $\min(\mu_1, \mu_2) \leq \varphi(\mu_1, \mu_2) \leq \max(\mu_1, \mu_2)$, whether φ is the arithmetic mean or the normalized product.

Potential shortcomings of using the arithmetic mean (which is by far the best option inside probability theory) is given by the following example: consider two experts giving their opinion about the potential value of a real-valued variable X . Suppose the first gives an opinion centered around small values, and the other around high values. Aggregating their opinion by the arithmetic mean would give a distribution whose mean would be between the two values, a value none of the expert considers as possible. There is no means with probabilistic modeling to convey the idea that either the opinion that X is small **or** that it is high is right, and that both cannot be right at the same time. This is not the case for other theories that allow for

imprecision in their representations. We will now give a review of the main operators used in each of them.

4.1.3.2 Imprecise probability theory

We consider that information delivered by the N sources is modeled by the credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$ on a domain \mathcal{X} . Again, we will often limit ourselves to credal sets $\mathcal{P}_{\underline{P}}$ defined by coherent lower probability measures \underline{P} , as they are more tractable in practice. Given a set of probabilities \mathcal{P} (convex or not), we will denote by $\mathcal{H}(\mathcal{P})$ its convex hull (i.e., the set of all convex combinations of its extreme points $\mathbf{ext}_{\mathcal{P}}$)

Conjunction Given credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$, their conjunction $\mathcal{P}_{\cap(1:N)}$ is defined as

$$\mathcal{P}_{\cap(1:N)} = \bigcap_{i=1}^N \mathcal{P}_i \quad (4.4)$$

and we say that the credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$ (and associated sources) are conflicting if $\mathcal{P}_{\cap(1:N)}$ is empty. Credal sets whose intersection is not empty are said to be *non-conflicting*. If credal sets $\mathcal{P}_{\underline{P}_1}, \dots, \mathcal{P}_{\underline{P}_N}$ are induced by coherent lower probabilities $\underline{P}_1, \dots, \underline{P}_N$ and are non-conflicting, then the credal set $\mathcal{P}_{\underline{P}_{\max(1:N)}}$ induced by $\underline{P}_{\max(1:N)} = \max(\underline{P}_1, \dots, \underline{P}_N)$ only gives an outer approximation of $\mathcal{P}_{\underline{P}_{\cap(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{\underline{P}_i}$, and we have $\mathcal{P}_{\underline{P}_{\max(1:N)}} \supseteq \mathcal{P}_{\underline{P}_{\cap(1:N)}}$. Although $\mathcal{P}_{\underline{P}_{\cap(1:N)}}$ is still induced by a coherent lower probability $\underline{P}_{\cap(1:N)}$ on events, exactly computing this lower probability usually requires the resolution of $\wp(\mathcal{X})$ linear programs [40, 42].

If the credal sets $\mathcal{P}_{L_1}, \dots, \mathcal{P}_{L_N}$ are given by imprecise probability assignments, then de Campos *et al.* [42] have shown that the conjunction $\mathcal{P}_{L_{\cap(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{L_i}$ again corresponds to a credal set induced by an imprecise probability assignment, and that this imprecise probability assignment is given, for every element $x \in \mathcal{X}$, by:

$$l_{L_{\cap(1:N)}}(x) = \left\{ \max_{i \in [N]} (l_i(x)), 1 - \sum_{y \in \mathcal{X} \setminus x} \min_{i \in [N]} (u_i(y)) \right\}$$

$$u_{L_{\cap(1:N)}}(x) = \left\{ \min_{i \in [N]} (u_i(x)), 1 - \sum_{y \in \mathcal{X} \setminus x} \max_{i \in [N]} (l_i(y)) \right\}$$

where $l_{L_{\cap(1:N)}}$ and $u_{L_{\cap(1:N)}}$ are the lower and upper bounds of the imprecise probability assignment inducing $\mathcal{P}_{L_{\cap(1:N)}}$.

Disjunction Given credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$, their disjunction $\mathcal{P}_{\cup(1:N)}$ is defined as

$$\mathcal{P}_{\cup(1:N)} = \bigcup_{i=1}^N \mathcal{P}_i \quad (4.5)$$

in practice, it is mathematically convenient (and equivalent from a behavioral point of view [202]) to consider the convex hull $\mathcal{H}(\mathcal{P}_{\cup(1:N)})$ of $\mathcal{P}_{\cup(1:N)}$. When the credal sets $\mathcal{P}_{P_1}, \dots, \mathcal{P}_{P_N}$ are induced by coherent lower probabilities, their disjunction $\mathcal{H}(\mathcal{P}_{\cup(1:N)}) = \mathcal{H}(\bigcup_{i=1}^N \mathcal{P}_{P_i})$ will not, in general, be induced by a coherent lower probability. Nevertheless, we can easily compute the tightest lower probability inducing a credal set outer approximating $\mathcal{H}(\mathcal{P}_{\cup(1:N)})$. It is given by $\underline{P}_{\min(1:N)} = \min(\underline{P}_1, \dots, \underline{P}_N)$ and the induced credal set $\mathcal{P}_{\underline{P}_{\min(1:N)}}$ is such that $\mathcal{P}_{\underline{P}_{\min(1:N)}} \supseteq \mathcal{H}(\mathcal{P}_{\cup(1:N)})$. Similarly, if the credal sets $\mathcal{P}_{L_1}, \dots, \mathcal{P}_{L_N}$ are induced by imprecise probability assignments, then $\mathcal{H}(\mathcal{P}_{L_{\cup(1:N)}}) = \mathcal{H}(\bigcup_{i=1}^N \mathcal{P}_{L_i})$ will not, in general, be induced by an imprecise probability assignment. However, the tightest imprecise probability assignment L_{out} inducing a credal set $\mathcal{P}_{L_{out}}$ outer approximating $\mathcal{H}(\mathcal{P}_{L_{\cup(1:N)}})$ is simple to compute, and is given, for every $x \in \mathcal{X}$, by

$$\begin{aligned} l_{out}(x) &= \min_{i \in \llbracket N \rrbracket} (l_i(x)) \\ u_{out}(x) &= \max_{i \in \llbracket N \rrbracket} (u_i(x)) \end{aligned}$$

where l_i, u_i are the bounds of the imprecise probability assignment L_i .

Convex combination Given the credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$, their respective sets of extreme points $\mathbf{ext}_1, \dots, \mathbf{ext}_N$, and non-negative weights $\lambda_1, \dots, \lambda_N$ summing up to one and associated to the credal sets, their convex combination is defined as

$$\mathcal{P}_{\Sigma(1:N)} = \sum_{i \in \llbracket N \rrbracket} \lambda_i \mathcal{P}_i \quad (4.6)$$

and $\mathcal{P}_{\Sigma(1:N)}$ is equivalent to the credal set associated to the set $\mathbf{ext}_{\Sigma(1:N)} = \{\sum_{i \in \llbracket N \rrbracket} \lambda_i p_i \mid p_i \in \mathbf{ext}_i\}$ of extreme points built from the convex combination of extreme points of $\mathcal{P}_1, \dots, \mathcal{P}_N$. If credal sets $\mathcal{P}_{P_1}, \dots, \mathcal{P}_{P_N}$ are induced by coherent lower probabilities, then their convex combination $\mathcal{P}_{\underline{P}_{\Sigma(1:N)}}$ is easy to compute and is induced by the coherent lower probability $\underline{P}_{\Sigma(1:N)} = \sum_{i \in \llbracket N \rrbracket} \lambda_i \underline{P}_i$, with \underline{P}_i the coherent lower probability inducing \mathcal{P}_{P_i} . Similarly, if the credal sets $\mathcal{P}_{L_1}, \dots, \mathcal{P}_{L_N}$ are induced by imprecise probability assignments, then their convex combination $\mathcal{P}_{L_{\Sigma(1:N)}}$ is still induced by an imprecise probability assignment $L_{\Sigma(1:N)}$ such that, for every

| Property | Formulation | Property | Formulation |
|-----------|---|------------|--|
| I Cons. | $\mathcal{P}_{\varphi(1:N)} \neq \emptyset$ | VII W-M-P | $\mathcal{P}_{\varphi(1:N)} \supseteq \mathcal{P}_{\cap(1:N)}$ |
| II Ass. | $\mathcal{P}_{\varphi(1:N)} = \varphi(\mathcal{P}_{\varphi(1:N-1)}, \mathcal{P}_N)$ | VIII S-M-P | $\mathcal{P}_{\varphi(1:N)} \supseteq \mathcal{H}(\mathcal{P}_{\cup(1:N)})$ |
| III Comm. | $\mathcal{P}_{\varphi(1:N)} = \varphi(\mathcal{P}_{\sigma(1)}, \dots, \mathcal{P}_{\sigma(N)})$ | IX Recon. | $\forall K, \left(\mathcal{P}_{\varphi(1:N)} \cap \bigcap_{i \in K} \mathcal{P}_i \right) \neq \emptyset$ |
| IV Idem. | $\varphi(\mathcal{P}, \mathcal{P}) = \mathcal{P}$ | X InsIgn. | $\varphi(\mathcal{P}_{\varphi(1:N)}, \mathcal{P}_{N+1}) = \mathcal{P}_{\varphi(1:N)}$ if $\mathcal{P}_{N+1} \supseteq \mathcal{P}_{\cap(1:N)}$ |
| V W-Z-P | $\mathcal{P}_{\varphi(1:N)} \subseteq \mathcal{H}(\mathcal{P}_{\cup(1:N)})$ | XI Conv. | $\mathcal{P}_{\varphi(1:N)} = \mathcal{H}(\mathcal{P}_{\varphi(1:N)})$ |
| VI S-Z-P | $\mathcal{P}_{\varphi(1:N)} \subseteq \mathcal{P}_{\cap(1:N)}$ | XII Robus. | $d(\mathcal{P}_{\varphi(1:N)}, \mathcal{P}'_{\varphi(1:N)}) \rightarrow 0$ as $d(\mathcal{P}_i, \mathcal{P}'_i) \rightarrow 0$ with $\mathcal{P}'_{\varphi(1:N)} = \varphi(\mathcal{P}'_1, \dots, \mathcal{P}'_N)$ |

Table 4.1: Properties of Section 4.1.2 for credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$ with φ the fusion operator, $\mathcal{P}_{\varphi(1:i)} = \varphi(\mathcal{P}_1, \dots, \mathcal{P}_i)$, $K \subseteq \llbracket N \rrbracket$ any maximal subset such that $(\bigcap_{i \in K} \mathcal{P}_i) \neq \emptyset$ and d the supremum norm between credal sets.

$x \in \mathcal{X}$

$$l_{\Sigma(1:N)}(x) = \sum_{i \in \llbracket N \rrbracket} \lambda_i(l_i(x))$$

$$u_{\Sigma(1:N)}(x) = \sum_{i \in \llbracket N \rrbracket} \lambda_i(u_i(x))$$

Table 4.1 summarizes how properties of Section 4.1 particularize to the case of credal sets (most formulations are equivalent to those given by Walley [202]). Recall that, given two credal sets $\mathcal{P}_1, \mathcal{P}_2$, the supremum norm reads

$$d(\mathcal{P}_1, \mathcal{P}_2) = \max \left\{ \sup_{P_1 \in \mathcal{P}_1} \inf_{P_2 \in \mathcal{P}_2} \|P_1 - P_2\|, \sup_{P_2 \in \mathcal{P}_2} \inf_{P_1 \in \mathcal{P}_1} \|P_1 - P_2\| \right\}$$

with $\|P_1 - P_2\| = \sup_{E \subseteq \mathcal{X}} |P_1(E) - P_2(E)|$. Conjunction is found back if we require both properties VII and VI, while requiring properties V and VIII enforces to take the convex hull of the disjunction.

4.1.3.3 Random set theory

We consider now that information is given by N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ defined on \mathcal{X} . In order to define usual fusion operators on random sets and to relate them to properties of Section 4.1.2, we must first choose a particular relation of inclusion between random sets, with respect to which will be defined the fusion operators. We choose here to concentrate on the s -ordering and on the associated notions of inclusion and of specialization, because this ordering has a unifying characteristic when compared to the others, and seems the more "natural" ordering within random set theory. See appendix C for more details about orderings and inclusion notions of random sets. If the notion of convex sets on \mathcal{X} makes sense, given a random set (m, \mathcal{F}) , we define its convex hull $\mathcal{H}((m, \mathcal{F}))$ as the random set which have the same bpa m , but where every focal element $E \in \mathcal{F}$ is replaced by its convex hull $\mathcal{H}(E)$.

Given N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, we define a joint basic probability assignment (bpa) $m_{(1:N)}$ defined on \mathcal{X}^N as a joint distribution having $m_i, i = 1, \dots, N$ for marginals, with m_i the bpa of the random set $(m, \mathcal{F})_i$. That is, for all $j \in 1, \dots, N$ and $i_j \in \{1, \dots, |\wp(\mathcal{X})|\}$, $m_{(1:N)}$ is such that:

$$m_{(1:N)}(\cdot \times E_{i_j}) := \sum_{\substack{i_1, \dots, i_{j-1} \in \{1, \dots, |\wp(\mathcal{X})|\} \\ i_{j+1}, \dots, i_N \in \{1, \dots, |\wp(\mathcal{X})|\}}} m_{(1:N)}(E_{i_1} \times \dots \times E_{i_j} \times \dots \times E_{i_N}) = m_j(E_{i_j}) \quad (4.7)$$

note that $m_{(1:N)}$ can receive positive mass only if E_{i_j} is in \mathcal{F}_j for $j = 1, \dots, N$.

Conjunction Given N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, a conjunction can be obtained by the following procedure:

1. Build a joint bpa $m_{(1:N)}$ satisfying Equation (4.7).
2. Allocate each joint mass $m_{(1:N)}(\times_{j=1}^N E_j)$ to the set $\cap_{j=1}^N E_j$, with $E_j \in \mathcal{F}_j$ for $j = 1, \dots, N$.

Note that, in the above procedure, the empty set can have a non-null mass (when it is the case, Property I of total consistency is not satisfied). The empty set is thus equivalent to a random set where the empty set has a mass equal to one.

We note $\mathcal{M}_{\cap(1:N)}$ the set of all possible bpa resulting from a conjunction of the marginal random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, and $(m, \mathcal{F})_{\cap(1:N)}$ a particular random set of $\mathcal{M}_{\cap(1:N)}$. $\mathcal{M}_{\cap(1:N)}$ can be characterized by four different situations:

1. $\mathcal{M}_{\cap(1:N)}$ contains only normalized random sets, that is, for every $(m, \mathcal{F})_{\cap(1:N)} \in \mathcal{M}_{\cap(1:N)}$, $m_{\cap(1:N)}(\emptyset) = 0$. It means that for all sets $\{E_j | j = 1, \dots, N, E_j \in \mathcal{F}_j\}$, we have $\cap_{j=1}^N E_j \neq \emptyset$. The random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ are then said to be *logically consistent* or *totally consistent*.
2. $\mathcal{M}_{\cap(1:N)}$ contains both subnormalized and normalized random sets. It means that there are sets $\{E_j | j = 1, \dots, N, E_j \in \mathcal{F}_j\}$ such that $\cap_{j=1}^N E_j = \emptyset$, but that there are also some random sets $(m, \mathcal{F})_{\cap(1:N)} \in \mathcal{M}_{\cap(1:N)}$ such that $m_{\cap(1:N)}(\emptyset) = 0$. In this case, Random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ are said to be *non-conflicting*.
3. $\mathcal{M}_{\cap(1:N)}$ contains only subnormalized random sets. It means that all random sets $(m, \mathcal{F})_{\cap(1:N)} \in \mathcal{M}_{\cap(1:N)}$ are such that $m_{\cap(1:N)}(\emptyset) > 0$. Random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ are then said to be *conflicting*.
4. All random sets $(m, \mathcal{F})_{\cap(1:N)}$ in $\mathcal{M}_{\cap(1:N)}$ are such that $m_{\cap(1:N)}(\emptyset) = 1$. Random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ are then said to be *totally conflicting*, and the associated set $\mathcal{M}_{\cap(1:N)}$ is said to be *empty*.

We also say of conflicting and non-conflicting random sets that they are *partially consistent*, in the sense that the result of a conjunctive combination of these random sets does not necessarily satisfies the property of total consistency, but the joint bpa mass can be assigned to other sets than the empty one.

If random sets are interpreted as hierarchical uncertainty representations where the 1st order level models are credal sets modeling ignorance on focal sets and the 2nd order models are precise probabilities, then a random set conjunction defined above comes down to build a precise joint probability as 2nd order model while conjunctively merging the corresponding 1st order models with Equation (4.4). In this case, the joint precise probability at the 2nd level is often interpreted as a convenient means to model the dependencies between sources.

If random sets are interpreted as 1st order imprecise probabilistic models, then the conjunction defined above appears less justified, and using Equation (4.4) seems to be a better choice (However, we'll see in Section 4.1.3.5 that we can link this equation to $\mathcal{M}_{\cap(1:N)}$).

One of the commonest combination rule used with random sets is obtained when bpa's m_1, \dots, m_N are judged to be independent. We note by $(m, \mathcal{F})_{\odot(1:N)}$ the specific conjunctive random set obtained with this assumption. We then have, for any collection of sets

$\{E_j | j = 1, \dots, N, E_j \in \mathcal{F}_j\}$

$$m_{\odot(1:N)}(\cap_{j=1}^N E_j) = \prod_{j=1}^N m_j(E_j). \quad (4.8)$$

Where $m_{\odot(1:N)}$ is the bpa of $(m, \mathcal{F})_{\odot(1:N)}$. The well known Dempster [57] combination rule can then be retrieved by uniformly redistributing the mass $m_{\odot(1:N)}(\emptyset)$ among focal sets $E \in \mathcal{F}_{\odot(1:N)}$ of $(m, \mathcal{F})_{\odot(1:N)}$, thus normalizing the conjunction result. We note $(m, \mathcal{F})_{\otimes(1:N)}$ the random set resulting from such a normalization. This makes the fusion result totally consistent, but if $m_{\odot(1:N)}(\emptyset) > 0$, the belief measure associated to $(m, \mathcal{F})_{\otimes(1:N)}$ no longer verifies $Bel_{(m, \mathcal{F})_{\otimes(1:N)}} \geq \max_{i=1, \dots, N} \{Bel_{(m, \mathcal{F})_i}\}$, similarly to what happens in probability theory for the product aggregation rule. Evenly distributing the mass $m_{\odot(1:N)}(\emptyset)$ among focal sets is not the only way of making the conjunction result totally consistent, and other authors have argued that other normalization operators appear more reasonable in some cases (see Smets [188] for a recent and critical survey). Most of the conjunctive combination rules proposed in the literature consist in choosing a particular random set in $\mathcal{M}_{\cap(1:N)}$, based on different rational arguments (for example, Cattaneo [19] proposes to minimize first the conflict, that is the mass allocated to the empty set, and then to maximize the imprecision, while Denoeux [61] uses a cautious approach based on Smets [184] canonical decomposition).

Disjunction Given N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, a disjunction can be obtained by the following procedure:

1. Build a joint bpa $m_{(1:N)}$ satisfying Equation 4.7.
2. Allocate each joint mass $m_{(1:N)}(\times_{j=1}^N E_j)$ to the set $\cup_{j=1}^N E_j$, with $E_j \in \mathcal{F}_j$ for $j = 1, \dots, N$.

We note $\mathcal{M}_{\cup(1:N)}$ the set of all possible bpa resulting from a disjunction of the marginal random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$, and $(m, \mathcal{F})_{\cup(1:N)}$ a particular random set of $\mathcal{M}_{\cup(1:N)}$.

Augustin [7] considers another possible disjunctive combination of random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$: he proposes to take the convex hull of the union of all marginal bpas m_1, \dots, m_N , and to consider the resulting set of bpas as the result of the disjunction. This is equivalent to apply Equation (4.5) to the credal sets induced by random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$.

Convex combination Given N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ and associated non-negative weights $\lambda_1, \dots, \lambda_N$ summing to one, the random set $(m, \mathcal{F})_{\Sigma(1:N)}$ resulting from their convex

| Property | Formulation | Property | Formulation |
|-----------|---|------------|--|
| I Cons. | $m_{\varphi_{(1:N)}}(\emptyset) = 0$ | VII W-M-P | $(m, \mathcal{F})_{\varphi_{\cap(1:N)}} \sqsubseteq_s (m, \mathcal{F})_{\varphi_{(1:N)}}$ |
| II Ass. | $(m, \mathcal{F})_{\varphi_{(1:N)}} = \varphi((m, \mathcal{F})_{\varphi_{(1:N-1)}}, (m, \mathcal{F})_N)$ | VIII S-M-P | $(m, \mathcal{F})_{\varphi_{\cup(1:N)}} \sqsubseteq_s (m, \mathcal{F})_{\varphi_{(1:N)}}$ |
| III Comm. | $(m, \mathcal{F})_{\varphi_{(1:N)}} = \varphi((m, \mathcal{F})_{\sigma(1)}, \dots, (m, \mathcal{F})_{\sigma(N)})$ | IX Recon. | $\{(m, \mathcal{F})_{\varphi_{(1:N)}}\} \cup \{(m, \mathcal{F})_i i \in K\}$ are not tot. conflict. $\forall K$ |
| IV Idem. | $\varphi((m, \mathcal{F}), (m, \mathcal{F})) = (m, \mathcal{F})$ | X Insign. | $\varphi((m, \mathcal{F})_{\varphi_{(1:N)}}, (m, \mathcal{F})_{N+1}) = (m, \mathcal{F})_{\varphi_{(1:N)}}$ if $(m, \mathcal{F})_{\varphi_{\cap(1:N)}} \sqsubseteq_s (m, \mathcal{F})_{N+1}$ |
| V W-Z-P | $(m, \mathcal{F})_{\varphi_{(1:N)}} \sqsubseteq_s (m, \mathcal{F})_{\varphi_{\cup(1:N)}}$ | XI Conv. | $(m, \mathcal{F})_{\varphi_{(1:N)}} = \mathcal{H}((m, \mathcal{F})_{\varphi_{(1:N)}})$ |
| VI S-Z-P | $(m, \mathcal{F})_{\varphi_{(1:N)}} \sqsubseteq_s (m, \mathcal{F})_{\varphi_{\cap(1:N)}}$ | XII Robus. | $d((m, \mathcal{F})_{\varphi_{(1:N)}}, (m, \mathcal{F})'_{\varphi_{(1:N)}}) \rightarrow 0$ as $d((m, \mathcal{F})_i, (m, \mathcal{F})'_i) \rightarrow 0$ with $(m, \mathcal{F})'_{\varphi_{(1:N)}} = \varphi((m, \mathcal{F})'_1, \dots, (m, \mathcal{F})'_N)$ |

Table 4.2: Properties of Section 4.1.2 for random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ with φ the fusion operator, $(m, \mathcal{F})_{\varphi_{(1:i)}} = \varphi((m, \mathcal{F})_1, \dots, (m, \mathcal{F})_i)$, $m_{\varphi_{(1:i)}}$ and $\mathcal{F}_{\varphi_{(1:i)}}$ the associated bpa and focal sets (φ_{\cap} and φ_{\cup} denote the conjunction and disjunction associated to $m_{\varphi_{(1:i)}}$). $K \subseteq \llbracket N \rrbracket$ denote any maximal subset such that $\mathcal{M}_{\cap K}$ is non-empty (i.e., random sets $\{(m, \mathcal{F})_i | i \in I\}$ are not totally conflicting), and d a distance measure between random sets [125].

combination has a bpa $m_{\Sigma_{(1:N)}}$ which, for every subset $E \subseteq \mathcal{X}$, has value

$$m_{\Sigma_{(1:N)}}(E) = \sum_{i=1}^N \lambda_i m_i(E). \quad (4.9)$$

Table 4.2 summarizes how properties of Section 4.2 particularize to the case of random sets and the notion of inclusion associated to s -ordering. φ consists here of building a joint bpa $m_{(1:N)}$ and then to distribute the mass $m_{(1:N)}(\times_{j=1}^N E_j)$ among combinations of conjunctions and disjunctions of sets E_j , with $E_j \in \mathcal{F}_j$ for $j = 1, \dots, N$. φ_{\cap} and φ_{\cup} denote the conjunction and disjunction associated to the joint bpa obtained by the φ operator.

4.1.3.4 Possibility theory

We consider that information delivered by the N sources is modeled by possibility distributions π_1, \dots, π_N on a domain \mathcal{X} . Recall that a distribution π_1 is said to be included in another

distribution π_2 ($\pi_1 \subseteq \pi_2$) if $\pi_2 \geq \pi_1$. Given a possibility distribution π and its α -cuts π_α , its convex hull $\mathcal{H}(\pi)$ is obtained by taking the convex hull of each α -cut, that is $\mathcal{H}(\pi)_\alpha = \mathcal{H}(\pi_\alpha)$ for any $\alpha \in [0, 1]$ (provided the convex hull of an α -cut is defined). The empty set is here equivalent to $\pi(x) = 0$ for all $x \in \mathcal{X}$. Given a set of possibility distributions, their respective cores (i.e. level 1 α -cut) and supports (i.e. level 0 α -cut), they are said to be: totally conflicting if the intersection of their supports is empty; partially conflicting or partially consistent if the intersection of their supports and cores are respectively non-empty and empty; totally consistent if the intersection of their cores is non-empty.

Conjunction Given N possibility distributions π_1, \dots, π_N , their usual conjunction is given, for all $x \in \mathcal{X}$, by

$$\pi_{\top(1:N)}(x) = \top_{i=1, \dots, N} \pi_i(x) \quad (4.10)$$

where \top is a triangular norm [127], or t-norm for short. A t-norm is a function $\top : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that is associative, commutative, non-decreasing in each variable and has 1 as identity element (i.e., $\top(x, 1) = x$). The most commonly used t-norms are the minimum and the product. The minimum is the largest point-wise t-norm and the only one to possess the idempotence property, making it the most conservative (and, therefore, cautious) conjunctive operator in possibility theory. The product is often associated to an assumption of independence between sources. Note that distributions π_1, \dots, π_N are totally conflict, partially consistent and totally consistent when and only when $\pi_{\min(1:N)}$ is respectively such that: $\pi_{\min(1:N)}(x) = 0$ for all $x \in \mathcal{X}$; $\pi_{\min(1:N)} < 1$ and is positive for at least one element $x \in \mathcal{X}$; $\pi_{\min(1:N)}(x) = 1$ for at least one element $x \in \mathcal{X}$.

Disjunction Given N possibility distributions π_1, \dots, π_N , their usual disjunction is given, for all $x \in \mathcal{X}$, by

$$\pi_{\perp(1:N)}(x) = \perp_{i=1, \dots, N} \pi_i(x) \quad (4.11)$$

where \perp is a triangular conorm, or t-conorm for short. A t-conorm is a function $\perp : [0, 1] \times [0, 1] \rightarrow [0, 1]$ that is associative, commutative, non-decreasing in each variable and has 0 as identity element (i.e., $\perp(x, 0) = x$). T-conorms are dual to t-norms, in the sense that, to any t-norm \top can be associated its complementary t-conorm \perp such that

$$\perp(x, y) = 1 - \top(1 - x, 1 - y) \quad (4.12)$$

and conversely, to any t-conorm is associated its dual t-norm. The most commonly used t-conorm is the maximum t-conorm, which is the smallest point-wise t-conorm, the dual of the minimum t-norm and the only t-conorm to possess the idempotence property.

Convex Combination Given N possibility distributions π_1, \dots, π_N and associated non-negative weights $\lambda_1, \dots, \lambda_N$ summing up to one, the possibility distribution $\pi_{\Sigma(1:N)}$ resulting from their convex combination is given, for every value $x \subseteq \mathcal{X}$ by

$$\pi_{\Sigma(1:N)}(x) = \sum_{i=1}^N \lambda_i \pi_i(x). \quad (4.13)$$

Table 4.3 summarizes how properties of Section 4.1.2 particularize to the case of possibility distributions (most formulations are equivalent to those given by Oussalah [161]). For the robustness property, we consider the distance between two possibility distributions π_1, π_2 given by

$$1 - \frac{\sum_{x \in \mathcal{X}} \min(\pi_1(x), \pi_2(x))}{\sum_{x \in \mathcal{X}} \max(\pi_1(x), \pi_2(x))} \quad (4.14)$$

and, when \mathcal{X} is continuous, sums simply become integrals. Note that this distance is based on a similarity measure considered by Pappis and Karacapilidis [164] for fuzzy sets, and also reduce to the Jacquard index when possibility distributions model classical sets.

4.1.3.5 Links between basic rules

In this section, we recall some of the links existing between the basic fusion rules of each uncertainty theories.

Possibility distributions and random set theory First, recall that, given a random set (m, \mathcal{F}) defined over a space \mathcal{X} , its contour function is given by the values $Pl(\{x\})$ for all elements $x \in \mathcal{X}$. This contour function is formally equivalent to a fuzzy set, which is normalized if and only if there is an element $x \in \mathcal{X}$ that is in every focal set of (m, \mathcal{F}) .

Now, if π_1, π_2 are two possibility distributions, and $(m, \mathcal{F})_{\pi_1}, (m, \mathcal{F})_{\pi_2}$ the corresponding random sets with nested focal sets, then Dubois and Yager [101] show that the random sets $(m, \mathcal{F})_{\pi_{\min}}$ and $(m, \mathcal{F})_{\pi_{\max}}$ with nested focal sets corresponding to the possibility distributions

| Property | Formulation | Property | Formulation |
|-----------|---|------------|---|
| I Cons. | $\exists x \in \mathcal{X}$ s.t. $\pi_{\varphi(1:N)} = 1$ | VII W-M-P | $\pi_{\varphi(1:N)} \geq \pi_{\min(1:N)}$ |
| II Ass. | $\pi_{\varphi(1:N)} = \varphi(\pi_{\varphi(1:N-1)}, \pi_N)$ | VIII S-M-P | $\pi_{\varphi(1:N)} \geq \pi_{\max(1:N)}$ |
| III Comm. | $\pi_{\varphi(1:N)} = \varphi(\pi_{\sigma(1)}, \dots, \pi_{\sigma(N)})$ | IX Recon. | $\forall K, \min(\pi_{\varphi(1:N)}, \min_{i \in K} \pi_i) \neq \emptyset$ |
| IV Idem. | $\varphi(\pi, \pi) = \pi$ | X InsIgn. | $\varphi(\pi_{\varphi(1:N)}, \pi_{N+1}) = \pi_{\varphi(1:N)}$ if $\pi_{N+1} \geq \pi_{\varphi(1:N)}$ |
| V W-Z-P | $\pi_{\varphi(1:N)} \leq \pi_{\max(1:N)}$ | XI Conv. | $\pi_{\varphi(1:N)} = \mathcal{H}(\pi_{\varphi(1:N)})$ |
| VI S-Z-P | $\pi_{\varphi(1:N)} \leq \pi_{\min(1:N)}$ | XII Robus. | $d(\pi_{\varphi(1:N)}, \pi'_{\varphi(1:N)}) \rightarrow 0$ as $d(\pi_i, \pi'_i) \rightarrow 0$ with $\pi'_{\varphi(1:N)} = \varphi(\pi'_1, \dots, \pi'_N)$ |

Table 4.3: Properties of Section 4.1.2 for possibility distributions π_1, \dots, π_N with φ the fusion operator, $\pi_{\varphi(1:i)} = \varphi(\pi_1, \dots, \pi_i)$. $K \subseteq \llbracket N \rrbracket$ denote any maximal subset such that $\min_{i \in K} \pi_i \neq \emptyset$, and d the distance measure between possibility distributions given by Equation (4.14).

$\pi_{\min(1:2)}$ and $\pi_{\max(1:2)}$ can be obtained in the following way:

- Let $\{0 = \gamma_0 < \gamma_1 < \dots < \gamma_M\}$ be the set of all distinct values taken by π_1, π_2 over \mathcal{X} .
- Build a joint bpa $m_{\pi(1:2)}$ such that $m(E_{\bar{\gamma}_i, \pi_1} \times E_{\bar{\gamma}_i, \pi_2}) = \gamma_{i+1} - \gamma_i$ for all $i = 0, \dots, M-1$, with $E_{\bar{\gamma}_i, \pi_j}$ the strong γ_i -cut of distribution π_j .
- To get $\pi_{\min(1:2)}, \pi_{\max(1:2)}$, consider respectively the conjunctive and disjunctive allocation of $m_{\pi(1:2)}$.

In other words, the minimum t-norm and maximum t-conorm in possibility theory are equivalent to consider random set conjunctions and disjunctions with a complete (positive) correlation between α -cuts. The extension to N possibility distributions follows from the associativity and commutativity of t-norms and t-conorms. Note that it is generally possible to find a conjunctively fused random set $(m, \mathcal{F})_{\cap(1:2)}$ built from $(m, \mathcal{F})_{\pi_1}, (m, \mathcal{F})_{\pi_2}$ that is different from $(m, \mathcal{F})_{\pi_{\min}}$ and still has $\pi_{\min(1:2)}$ as its contour function. In this latter case, $(m, \mathcal{F})_{\cap(1:2)}$ is not equivalent to the nested random set induced by π_{\min} , and we have $(m, \mathcal{F})_{\pi_{\min}} \sqsubseteq_{Pl} (m, \mathcal{F})_{\cap(1:2)}$

Dubois and Yager [101] also relate different specific joint bpa whose conjunctive or disjunctive allocations allow to retrieve a contour function having the same values as the application of other t-norms or t-conorms. In this latter case, some information is lost when restricting to the contour function. For instance, the contour function of the random set $(m, \mathcal{F})_{\odot(1:N)}$ cor-

responding to the product of marginal bpas is the same as the possibility distribution $\pi_{\Pi(1:N)}$ obtained with the product t-norm, but we have $(m, \mathcal{F})_{\pi_{\Pi(1:N)}} \sqsubseteq_{Pl} (m, \mathcal{F})_{\odot(1:N)}$, with $(m, \mathcal{F})_{\pi_{\Pi(1:N)}}$ the random set with nested focal sets induced by $\pi_{\Pi(1:N)}$. See Dubois and Yager [101] for an ampler discussion. A similar result holds for the convex combination: given N random sets $(m, \mathcal{F})_{\pi_1}, \dots, (m, \mathcal{F})_{\pi_N}$ induced by N possibility distributions π_1, \dots, π_N and non-negative weights $\lambda_1, \dots, \lambda_N$, the contour function of the random set $(m, \mathcal{F})_{\Sigma(1:N)}$ is equal to the possibility distribution $\pi_{\Sigma(1:N)}$ (this is simply due to the fact that $Pl_{\Sigma(1:N)}(\{x\}) = \sum_{i=1}^N \lambda_i Pl_i(\{x\})$ for every $x \in \mathcal{X}$), and we still have $(m, \mathcal{F})_{\pi_{\Sigma(1:N)}} \sqsubseteq_{Pl} (m, \mathcal{F})_{\Sigma(1:N)}$, with $(m, \mathcal{F})_{\pi_{\Sigma(1:N)}}$ the random set with nested focal sets induced by $\pi_{\Sigma(1:N)}$.

Possibility distributions and credal sets Recall that, given two possibility distributions π_1, π_2 defined on a space \mathcal{X} , the induced credal sets $\mathcal{P}_{\pi_1}, \mathcal{P}_{\pi_2}$ are such that $\mathcal{P}_{\pi_1} \subseteq \mathcal{P}_{\pi_2}$ if and only if $\pi_1 \leq \pi_2$.

Some elements of comparison between possibilistic conjunction and credal sets conjunction are already given in Section 3.3.2.2. Given N possibility distributions π_1, \dots, π_N and the respective induced credal sets $\mathcal{P}_{\pi_1}, \dots, \mathcal{P}_{\pi_N}$, we have $\mathcal{P}_{\pi_{\min(1:N)}} \subseteq \bigcap_{i=1}^N \mathcal{P}_{\pi_i}$ [88]. This is also true if min is replaced by any other t-norms, since min is the largest possible t-norm. A sufficient condition to have $\mathcal{P}_{\pi_{\min(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{\pi_i}$ will be given in Section 4.1.3.6.

Regarding disjunction, the credal set $\mathcal{P}_{\pi_{\max(1:N)}}$ induced by the possibility distribution $\pi_{\max(1:N)}$ is the tightest credal set induced by a lower coherent probability such that $\mathcal{H}(\bigcup_{i=1}^N \mathcal{P}_{\pi_i}) \subseteq \mathcal{P}_{\pi_{\max(1:N)}}$. This is due to the maxitivity property of possibility measures. As for random sets, one could relate a credal set resulting from disjunction, conjunction and convex combination to the contour function induced by this credal set, which would be given by the values $\bar{P}(\{x\})$ for all $x \in \mathcal{X}$. Studying those relations remains an open problem.

Random sets and credal sets We consider N random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ defined over the space \mathcal{X} and the induced credal sets $\mathcal{P}_{(m, \mathcal{F})_1}, \dots, \mathcal{P}_{(m, \mathcal{F})_N}$. It has been showed by Chateauneuf [21] that, in general, the conjunction of credal sets $\mathcal{P}_{\cap(m, \mathcal{F})_{(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{(m, \mathcal{F})_i}$ is no longer representable by a random set, and does not coincide with a conjunctively fused random set built from $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$. Chateauneuf [21] also shows that the credal set $\mathcal{P}_{\cap(m, \mathcal{F})_{(1:N)}}$ is equivalent to the set of all normalized random sets $(m(\emptyset) = 0)$ that are in $\mathcal{M}_{\cap(1:N)}$.

Relating random set disjunction with the credal set $\mathcal{P}_{\cup(m, \mathcal{F})_{(1:N)}} = \mathcal{H}(\bigcup_{i=1}^N \mathcal{P}_{(m, \mathcal{F})_i})$ corresponding to the convex hull of the disjunction of credal sets $\mathcal{P}_{(m, \mathcal{F})_1}, \dots, \mathcal{P}_{(m, \mathcal{F})_N}$ is more

difficult. On the one hand, it is known that the coherent lower probability of $\mathcal{P}_{\cup(m,\mathcal{F})_{(1:N)}}$ is not in general an ∞ -monotone capacity, and thus will not be induced by an equivalent random set. On the other hand, the set $\mathcal{M}_{\cup(1:N)}$ of all disjunctively fused random sets does not seem to be related to $\mathcal{P}_{\cup(m,\mathcal{F})_{(1:N)}}$. To see this, simply consider the case where $\mathcal{X} = \{x_1, x_2\}$ and where two sources provide the uniform probability as their opinion: $\mathcal{P}_{\cup(m,\mathcal{F})_{(1:2)}}$ will be the uniform probability, while the random set such that $m(\mathcal{X}) = 1$ will be in $\mathcal{M}_{\cup(1:2)}$ (i.e. ignorance on \mathcal{X}). Credal set disjunction can nevertheless be related to Augustin [7] proposition recalled in Section 4.1.3.3, since considering the convex hull $\mathcal{H}(\cup_{i=1}^N m_i)$ of all bpas of $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ is equivalent to $\mathcal{P}_{\cup(m,\mathcal{F})_{(1:N)}}$.

The relation between the convex combinations of random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ and of the induced credal sets $\mathcal{P}_{(m,\mathcal{F})_1}, \dots, \mathcal{P}_{(m,\mathcal{F})_N}$ is more direct. If $\lambda_1, \dots, \lambda_N$ are the non-negative weights summing to one associated to the sources, then the random set $(m, \mathcal{F})_{\Sigma(1:N)}$ resulting from the convex combination is the random set inducing the credal set $\mathcal{P}_{\Sigma(1:N)}$.

These relations and the previous sections show that, although there exist some congruencies between basic fusion rules of each uncertainty theories, results will in general not be directly related.

4.1.3.6 Fusion rules for clouds

Let $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ be N clouds defined over a space \mathcal{X} and modeling sources uncertainty. Due to their relationships with possibility distributions (Property 3.10), it seems natural to define their basic fusion rules as follows:

- **Conjunction:** we define the conjunction $[\pi, \delta]_{\cap}$ of the clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ as

$$[\pi, \delta]_{\cap(1:N)} = [\pi_{\cap(1:N)}, \delta_{\cap(1:N)}] = [\min_{i=1}^N(\pi_i), \max_{i=1}^N(\delta_i)]. \quad (4.15)$$

Note that this conjunction is different from the one usually considered for Interval-Valued Fuzzy Sets. Nevertheless, it is coherent with the fact that conjunctive fusion operators should reduce uncertainty (in the sense that $\mathcal{P}_{[\pi,\delta]_{\cap(1:N)}} \subseteq \mathcal{P}_{[\pi,\delta]_i}$ for any $i \in \llbracket N \rrbracket$), and also with the notion of conjunction used with bipolar information [15].

- **Disjunction:** we define the disjunction $[\pi, \delta]_{\cup}$ of the clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ as

$$[\pi, \delta]_{\cup(1:N)} = [\pi_{\cup(1:N)}, \delta_{\cup(1:N)}] = [\max_{i=1}^N(\pi_i), \min_{i=1}^N(\delta_i)] \quad (4.16)$$

- **Convex combination:** given non-negative weights $\lambda_1, \dots, \lambda_N$ summing up to one and associated to sources, we define the convex combination of the clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ as

$$[\pi, \delta]_{\Sigma(1:N)} = [\pi_{\Sigma(1:N)}, \delta_{\Sigma(1:N)}] = \left[\sum_{i=1}^N \lambda_i \pi_i, \sum_{i=1}^N \lambda_i \delta_i \right] \quad (4.17)$$

These operations still result in clouds, even if conjunction and convex combination may result in clouds that do not satisfy the boundary conditions usually associated to clouds (i.e., the existence of elements x and y in \mathcal{X} such that $\delta(x) = 0$ and $\pi(y) = 1$, see Section 3.3). Satisfying these boundary conditions corresponds to satisfying coherence (Property I of Section 4.1.2). The conjunction is said to be empty if the resulting cloud $[\pi, \delta]_{\cap(1:N)}$ is such that, for at least one element $x \in \mathcal{X}$, we have $\pi_{\cap(1:N)}(x) < \delta_{\cap(1:N)}(x)$. Note that, even if all clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ are comonotonic (see Section 3.3.3), then there is no guarantee that the clouds resulting from one of the above operations will still be comonotonic. The next proposition indicates a sufficient condition for the resulting clouds to be comonotonic.

Proposition 4.1. *Consider N clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ whose mappings $\{\delta_i, \pi_i | i = 1, \dots, N\}$ are all comonotonic. Let $\leq_{[\underline{E}, \overline{F}](1:N)}$ be the coarsest pre-order on \mathcal{X} refining all the pre-orders $\leq_{[\underline{E}, \overline{F}]_1}, \dots, \leq_{[\underline{E}, \overline{F}]_N}$ respectively induced by clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$. Then, the following clouds:*

$$[\pi, \delta]_{\cap(1:N)} = [\pi_{\cap(1:N)}, \delta_{\cap(1:N)}] = \left[\min_{i=1}^N (\pi_i), \max_{i=1}^N (\delta_i) \right] \quad (4.18)$$

$$[\pi, \delta]_{\cup(1:N)} = [\pi_{\cup(1:N)}, \delta_{\cup(1:N)}] = \left[\max_{i=1}^N (\pi_i), \min_{i=1}^N (\delta_i) \right] \quad (4.19)$$

$$[\pi, \delta]_{\Sigma(1:N)} = [\pi_{\Sigma(1:N)}, \delta_{\Sigma(1:N)}] = \left[\sum_{i=1}^N \lambda_i \pi_i, \sum_{i=1}^N \lambda_i \delta_i \right] \quad (4.20)$$

are comonotonic and induce the pre-order $\leq_{[\underline{E}, \overline{F}](1:N)}$ on \mathcal{X} .

Proof. The fact that all mappings $\{\delta_i, \pi_i | i = 1, \dots, N\}$ are comonotonic ensures that there exist a coarsest pre-order $\leq_{[\underline{E}, \overline{F}](1:N)}$ on \mathcal{X} refining all the pre-orders $\leq_{[\underline{E}, \overline{F}]_1}, \dots, \leq_{[\underline{E}, \overline{F}]_N}$.

Now, let us consider the pre-order $\leq_{[\underline{E}, \overline{F}](1:N)}$ and any pair of elements x and y in \mathcal{X} such that $x \leq_{[\underline{E}, \overline{F}](1:N)} y$. This means that $\delta_i(x) \leq \delta_i(y)$ and $\pi_i(x) \leq \pi_i(y)$ for all $i = 1, \dots, N$.

Due to the monotonic property of t-norms and t-conorms, and in particular of minimum

and maximum, we have the following inequalities:

$$\begin{aligned}\min_{i=1}^N(\pi_i(x)) &\leq \min_{i=1}^N(\pi_i(y)) \\ \max_{i=1}^N(\delta_i(x)) &\leq \max_{i=1}^N(\delta_i(y)) \\ \max_{i=1}^N(\pi_i(x)) &\leq \max_{i=1}^N(\pi_i(y)) \\ \min_{i=1}^N(\delta_i(x)) &\leq \min_{i=1}^N(\delta_i(y))\end{aligned}$$

therefore, the respective pre-orders $\leq_{[\underline{E}, \overline{F}]_{\cap(1:N)}}$ and $\leq_{[\underline{E}, \overline{F}]_{\cup(1:N)}}$ induced by clouds $[\pi, \delta]_{\cap(1:N)}$ and $[\pi, \delta]_{\cup(1:N)}$ are also such that $x \leq_{[\underline{E}, \overline{F}]_{\cap(1:N)}} y$ and $x \leq_{[\underline{E}, \overline{F}]_{\cup(1:N)}} y$. Since this is true for every possible pair of elements $x, y \in \mathcal{X}$, it follows that $\leq_{[\underline{E}, \overline{F}]_{(1:N)}}$, $\leq_{[\underline{E}, \overline{F}]_{\cap(1:N)}}$ and $\leq_{[\underline{E}, \overline{F}]_{\cup(1:N)}}$ are the same pre-orders. The same holds for the cloud $[\pi, \delta]_{\Sigma(1:N)}$, since the arithmetic weighted mean is also a monotone operation. \square

We now study relationships between basic fusion rules for clouds and the other uncertainty theories.

Relations with possibility theory. Clouds conjunctions, disjunctions and convex combination are respectively equivalent to applying the minimum t-norm, maximum t-conorm and arithmetic weighted mean separately to possibility distributions π_1, \dots, π_N and $1 - \delta_1, \dots, 1 - \delta_N$ and then to consider the associated cloud.

In this spirit, an obvious generalization of the conjunction and disjunction of clouds would be to consider a t-norm \top and its dual t-conorm \perp and then to respectively define the associated conjunctively and disjunctively fused clouds $[\pi, \delta]_{\top(1:N)}, [\pi, \delta]_{\perp(1:N)}$ as:

$$[\pi, \delta]_{\top(1:N)} = [\pi_{\top(1:N)}, \delta_{\top(1:N)}] = [\top_{i=1}^N(\pi_i), \perp_{i=1}^N(\delta_i)] \quad (4.21)$$

$$[\pi, \delta]_{\perp(1:N)} = [\pi_{\perp(1:N)}, \delta_{\perp(1:N)}] = [\perp_{i=1}^N(\pi_i), \top_{i=1}^N(\delta_i)] \quad (4.22)$$

and, by Equation (4.12), above conjunction and disjunction are respectively equivalent to apply \top and \perp separately to possibility distributions π_1, \dots, π_N and $1 - \delta_1, \dots, 1 - \delta_N$, and then to consider the associated cloud. Note that such operations are totally consistent with possibility theory, since when clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ are fuzzy ($\delta_i = 0, i = 1, \dots, N$), then we retrieve the usual conjunctions, disjunctions and convex combinations of possibility distri-

butions π_1, \dots, π_N . The above equations are still coherent with fusion operators defined in bipolar possibility theory [15]. Also, since t-norms and t-conorms are monotonic operators, Property 4.1 still holds for any $[\pi, \delta]_{\top(1:N)}, [\pi, \delta]_{\perp(1:N)}$.

Relations with random set theory. Recall that if all clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ are comonotone but induce non-compatible orderings $\leq_{[E, \bar{F}]_1}, \dots, \leq_{[E, \bar{F}]_N}$ (i.e., each pair of mappings $[\pi, \delta]_i, i = 1, \dots, N$ is comonotone, but two mappings $\pi_i, \pi_j, j \neq i$ can be non-comonotone), the clouds $[\pi, \delta]_{\cap(1:N)}, [\pi, \delta]_{\cup(1:N)}, [\pi, \delta]_{\Sigma(1:N)}$ resulting from conjunction, disjunction and convex combination will not forcefully be comonotone. Since most non-comonotonic clouds cannot be represented by an equivalent random set (Proposition 3.13), there is no direct relationships between clouds fusion rules and random set fusion rules that always hold. Nevertheless, when comonotonic clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ satisfy Proposition 4.1, we have the following relation

Proposition 4.2. Consider N clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ satisfying the conditions of Proposition 4.1, and $\leq_{[E, \bar{F}]_{(1:N)}}$ the pre-order induced from $[\pi, \delta]_1, \dots, [\pi, \delta]_N$. Let $0 = \gamma_0 < \gamma_1 < \dots < \gamma_M = 1$ be the distinct values taken by $[\pi, \delta]_1, \dots, [\pi, \delta]_N$, and $E_{i,j}$ the set given by Equation (3.39) for cloud $[\pi, \delta]_i$. We note $(m, \mathcal{F})_{[\pi, \delta]_i}, (m, \mathcal{F})_{[\pi, \delta]_{\cap(1:N)}}$ and $(m, \mathcal{F})_{[\pi, \delta]_{\cup(1:N)}}$ the random sets respectively induced by the clouds $[\pi, \delta]_i, [\pi, \delta]_{\cap(1:N)}$ and $[\pi, \delta]_{\cup(1:N)}$. Then, clouds $[\pi, \delta]_{\cap(1:N)}$ and $[\pi, \delta]_{\cup(1:N)}$ can be built in the following way:

- Build a joint bpa $m_{[\pi, \delta]_{(1:N)}}$ such that $m(E_{1,j} \times \dots \times E_{N,j}) = \gamma_j - \gamma_{j-1}$ for all $j = 1, \dots, M$
- The random sets $(m, \mathcal{F})_{[\pi, \delta]_{\cap(1:N)}}$, $(m, \mathcal{F})_{[\pi, \delta]_{\cup(1:N)}}$ induced by $[\pi, \delta]_{\cap(1:N)}, [\pi, \delta]_{\cup(1:N)}$ can respectively be retrieved by taking the conjunctive and disjunctive allocation of $m_{[\pi, \delta]_{(1:N)}}$.

Proof. Given the cloud $[\pi, \delta]_i$, we note $\bar{\pi}_i$ the distribution $1 - \delta_i$. We also note $(m, \mathcal{F})_{\pi_i}$ and $(m, \mathcal{F})_{\bar{\pi}_i}$ the following random sets:

$$\left\{ \begin{array}{l} E_{\pi_i, j} = \{x \in \mathcal{X} \mid \pi_i(x) \geq \gamma_j\} \\ m(E_{\pi_i, j}) = \gamma_j - \gamma_{j-1} \end{array} \right\} \quad \left\{ \begin{array}{l} E_{\bar{\pi}_i, j} = \{x \in \mathcal{X} \mid \bar{\pi}_i(x) \geq 1 - \gamma_{j-1}\} \\ m(E_{\bar{\pi}_i, j}) = \gamma_j - \gamma_{j-1} \end{array} \right\}$$

Note that $(m, \mathcal{F})_{\pi_i}, (m, \mathcal{F})_{\bar{\pi}_i}$ are the random sets induced by the possibility distributions $\pi_i, \bar{\pi}_i$ and are given by Equation (3.17).

We will only provide the proof for the conjunction, the proof for the disjunction being similar. First, let us consider the cloud $[\pi, \delta]_{\cap(1:N)}$ resulting from the conjunction. The random

set $(m, \mathcal{F})_{[\pi, \delta]_{\cap(1:N)}}$ induced by this cloud (which is comonotonic, by Proposition 4.1) reads, for $j = 1, \dots, M$:

$$\begin{cases} E_{[\pi, \delta]_{\cap, j}} = \{x \in \mathcal{X} \mid (\min_{i=1, \dots, N} \pi_i(x)) \geq \gamma_j \wedge (\max_{i=1, \dots, N} \delta_i(x)) < \gamma_j\} \\ m(E_{[\pi, \delta]_{\cap, j}}) = \gamma_j - \gamma_{j-1} \end{cases}$$

with $E_{[\pi, \delta]_{\cap, j}}$ the focal sets and $m(E_{[\pi, \delta]_{\cap, j}})$ its mass. Given an index j , we have the following equalities:

$$\begin{aligned} E_{[\pi, \delta]_{\cap, j}} &= \{x \in \mathcal{X} \mid \left(\min_{i=1, \dots, N} (\pi_i(x)) \right) \geq \gamma_j \wedge \left(\max_{i=1, \dots, N} (\delta_i(x)) \right) < \gamma_j\} \\ &= \{x \in \mathcal{X} \mid \left(\min_{i=1, \dots, N} (\pi_i(x)) \right) \geq \gamma_j\} \cap \{x \in \mathcal{X} \mid \left(\max_{i=1, \dots, N} (\delta_i(x)) \right) \leq \gamma_{j-1}\} \\ &= \{x \in \mathcal{X} \mid \left(\min_{i=1, \dots, N} (\pi_i(x)) \right) \geq \gamma_j\} \cap \{x \in \mathcal{X} \mid \left(\min_{i=1, \dots, N} (\bar{\pi}_i(x)) \right) \geq 1 - \gamma_{j-1}\} \end{aligned}$$

and by using the relation between the minimum in possibility theory and random sets (see Section 4.1.3.5),

$$\begin{aligned} &= \left(\bigcap_{i=1, \dots, N} \{x \in \mathcal{X} \mid \pi_i(x) \geq \gamma_j\} \right) \cap \left(\bigcap_{i=1, \dots, N} \{x \in \mathcal{X} \mid \bar{\pi}_i(x) \geq 1 - \gamma_{j-1}\} \right) \\ &= \bigcap_{i=1, \dots, N} (\{x \in \mathcal{X} \mid \pi_i(x) \geq \gamma_j\} \cap \{x \in \mathcal{X} \mid 1 - \delta_i(x) > 1 - \gamma_{j-1}\}) \\ &= \bigcap_{i=1, \dots, N} \{x \in \mathcal{X} \mid \pi_i(x) \geq \gamma_j \wedge \delta_i(x) < \gamma_j\} \end{aligned}$$

and, since $E_{i, j} = \{x \in \mathcal{X} \mid \pi_i(x) \geq \gamma_j \wedge \delta_i(x) < \gamma_j\}$, this finishes the proof. \square

Proposition 4.2 is similar to the link between possibilistic conjunction and random set conjunction. It also comes down to consider random set conjunctive and disjunctive allocations, while assuming a complete correlation between α -cuts, and in the case of fuzzy clouds, we retrieve the result of Dubois and Yager [101]. Note that, if clouds do not satisfy Proposition 4.1, then the procedure described in Proposition 4.2 give the random sets inner approximating $[\pi, \delta]_{\cap(1:N)}$, $[\pi, \delta]_{\cup(1:N)}$ (see Proposition 3.14).

Relations with credal sets. As for possibility theory, clouds resulting from the conjunction and disjunction generally induce respectively smaller and larger credal sets than the credal set conjunction and disjunction. This is formalized by the next proposition

Proposition 4.3. Let $\mathcal{P}_{[\pi, \delta]_1}, \dots, \mathcal{P}_{[\pi, \delta]_N}$ be the credal sets induced by the clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$, we then have the following relationships with clouds $[\pi, \delta]_{\cap(1:N)}, [\pi, \delta]_{\cup(1:N)}$:

$$\begin{aligned} \mathcal{P}_{[\pi, \delta]_{\cap(1:N)}} &\subseteq \bigcap_{i=1}^N \mathcal{P}_{[\pi, \delta]_i} \\ \mathcal{P}_{[\pi, \delta]_{\cup(1:N)}} &\supseteq \mathcal{H} \left(\bigcup_{i=1}^N \mathcal{P}_{[\pi, \delta]_i} \right) \end{aligned}$$

and the first inclusion is turned into an equality if clouds $[\pi, \delta]_1, \dots, [\pi, \delta]_N$ satisfy Proposition 4.1

Proof. By Proposition 3.10, we have $\mathcal{P}_{[\pi, \delta]_{\cap(1:N)}} = \mathcal{P}_{\pi_{\min(1:N)}} \cap \mathcal{P}_{\bar{\pi}_{\min(1:N)}}$, with $\bar{\pi}_i = 1 - \delta_i$. Since $\mathcal{P}_{\pi_{\min(1:N)}} \subseteq \bigcap_{i=1}^N \mathcal{P}_{\pi_i}$, and likewise for possibility distributions $\bar{\pi}_i$, this gives

$$\left(\mathcal{P}_{[\pi, \delta]_{\cap(1:N)}} \right) = \left(\mathcal{P}_{\pi_{\min(1:N)}} \cap \mathcal{P}_{\bar{\pi}_{\min(1:N)}} \right) \subseteq \left(\bigcap_{i=1}^N \mathcal{P}_{\pi_i} \cap \bigcap_{i=1}^N \mathcal{P}_{\bar{\pi}_i} \right) = \left(\bigcap_{i=1}^N \mathcal{P}_{[\pi, \delta]_i} \right)$$

and we have the inclusion relationship between conjunction. The inclusion between disjunctions can be proved likewise, since $\mathcal{P}_{\pi_{\max(1:N)}} \supseteq \mathcal{H}(\bigcup_{i=1}^N \mathcal{P}_{\pi_i})$.

Now let's turn to the case where all mappings $\{\delta_i, \pi_i | i = 1, \dots, N\}$ are comonotonic (clouds satisfying the conditions of Proposition 4.1). In this case, this means that every cloud $[\pi, \delta]_i$ can be mapped into an equivalent generalized p-box inducing the pre-order $\leq_{[E, \bar{F}](1:N)}$ and defined on a same collection $\emptyset \subset A_1 \subseteq \dots \subseteq A_{|\mathcal{X}|} = \mathcal{X}$ of nested sets, with $|\mathcal{X}|$ the cardinality of \mathcal{X} . These N generalized p-boxes correspond to N collection of constraints corresponding to Equations (3.20):

$$i = 1, \dots, |\mathcal{X}| \quad j = 1, \dots, N \quad \alpha_{i,j} \leq P(A_i) \leq \beta_{i,j}$$

with $\alpha_{i,j}, \beta_{i,j}$ respectively the lower and upper bounds on the probability of A_i induced by the cloud $[\pi, \delta]_j$. Since all the constraints are defined on the same collection of nested subsets, their conjunction is given by the $|\mathcal{X}|$ constraints:

$$i = 1, \dots, |\mathcal{X}| \quad \max_{j=1, \dots, N} (\alpha_{i,j}) \leq P(A_i) \leq \min_{j=1, \dots, N} (\beta_{i,j})$$

which are equivalent to the cloud $[\pi, \delta]_{\cap(1:N)}$, and also to $\bigcap_{i=1}^N \mathcal{P}_{[\pi, \delta]_i}$. This shows the equality.

□

And from this proposition, we can derive the following corollaries, of practical importance:

Corollary 4.4. *Let π_1, \dots, π_N be N possibility distributions defined on \mathcal{X} such that π_1, \dots, π_N are comonotone. Then, we have*

$$\mathcal{P}_{\pi_{\min(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{\pi_i}$$

Corollary 4.5. *Let $[\underline{E}, \overline{F}]_1, \dots, [\underline{E}, \overline{F}]_N$ be N usual p -boxes defined on the real line \mathbb{R} . Let $[\underline{E}, \overline{F}]_{\cap(1:N)}$ denote the p -box $[\max_{i=1, \dots, N}(\underline{E}_i), \min_{i=1, \dots, N}(\overline{F}_i)]$ Then, we have*

$$\mathcal{P}_{[\underline{E}, \overline{F}]_{\cap(1:N)}} = \bigcap_{i=1}^N \mathcal{P}_{[\underline{E}, \overline{F}]_i}$$

The above corollaries and propositions confirm that the comonotonicity property is very appealing, both for theoretical and practical reasons. It reinforces the idea that comonotonic clouds are likely to be more useful than their non-comonotonic counterparts. Above properties also suggest that multiple clouds should be elicited by considering a common ordering on the space \mathcal{X} , and that when extracting clouds from credal sets (for instance, by Algorithm 2), one should preferably always consider the same ordering of elements.

4.2 Treating the conflict by adaptive rules using maximal coherent subsets

In this section, we address the problem of dealing with partially conflicting information. In such cases, neither conjunctive nor disjunctive fusion operators will provide satisfactory results, the former resulting in poorly reliable representations, and the latter in very imprecise results.

As said in Section 4.1.1, trade-off operators can be used to provide a result between the conjunction and the disjunction, with the aim to balance gain and reliability of information. In most cases, the classical convex combination is used. However, it is in general not easy to determine meaningful weights $\lambda_1, \dots, \lambda_N$ (we will suggest some methods to do so in Section 4.4). Moreover, when sources provide information concerning an observable parameter

or a variable whose exact value is unknown, convex combinations can be criticized on the basis that they are voting-like procedures, and that their result could promote values that none sources judged plausible at first, an undesirable feature when modelling uncertainty about physical variables. Thus, convex combination seems more fitted to the cases where sources express preferences or utilities, and when a consensus has to be reached (in which case previous criticisms are no longer relevant).

There are other non-adaptive trade-off operators representing alternatives to convex combination: they include (among others) mean operators satisfying Equation (4.1)) and ordered weighted averaging operators introduced by Yager [213, 215]. Yet, such operators and convex combination often suffer from the same defects. Also, it is not always clear when to choose one operator rather than another one (when an important amount of data is available, it is possible to choose the operator by some training process, so that it best-fits the situation). Also, these operators always behave in the same way, irrespectively of the amount of conflict between information.

Adaptive rules are other trade-off operators, whose behavior depend on the amount of conflict in the information. They range from a conjunctive behavior (total consistency) to a disjunctive behavior (total conflict). In-between, they act as trade-off operators. They often consist in combination of disjunctive and conjunctive operators. Adaptive rules are particularly well suited when no specific knowledge is available about the sources (e.g., reliability) and when choosing a particular non-adaptive operator appears difficult. A typical example is when multiple experts whose reliability is unknown provide opinions about the value of a variable. Although conceptually attractive, such rules often require more computational effort than the other operators. Also, since their behavior depends on the global amount of inconsistency in the information, they usually do not satisfy Associativity (Property II), which means that all the sources have to be considered at once, and that no step-by-step computations are possible.

So far, trade-off operators other than classical convex combination have mainly been studied in the framework of possibility theory [162, 56, 215, 93]. Only few works study such operators in the framework of imprecise probability theory [154, 202], in which propositions to cope with partially conflicting information often considers second-order models [157, 194, 129]. Similarly, using trade-off operators outside convex combination to deal with partially conflicting information is seldom considered in random set theory, where most propositions to deal with partially consistent information consist in normalization procedures redistributing the mass associated to the empty set among other sets after a conjunctive fusion (see Smets [188] for a thorough and critical review). Note that although the Dubois and Prade [84]

rule is often seen as such a redistribution, it can also be interpreted as an adaptive trade-off operator, where pairs of focal sets are conjunctively allocated if their intersection is non-empty, and disjunctively allocated otherwise. In short, there exist a lot of rules that tries to cope with conflict, and it is often not clear which one should be chosen in a particular application. In this work, we propose to use fusion rules based on the notion of maximal coherent subsets, originating from logic [173]. As we shall see, this notion is conceptually simple, attractive and has many interesting properties, but can be limited by the computational burden associated to it.

4.2.1 Maximal coherent subsets (MCS) rule: basic methodology

Using MCS to fuse information is an attractive concept, since using MCS precisely aims at gaining a maximal amount of information while remaining coherent with all the information sources. It is a natural way to cope with the (somewhat "opposed") main objectives pursued by the general problem of information fusion. Given N sources, fusion rule based on MCS consist in applying conjunctive operator(s) inside subsets of sources that are consistent and then disjunctive operator(s) across these subgroups. It is thus an adaptive rule, since conjunctive and disjunctive behaviors are retrieved respectively when the N sources are all consistent or totally conflicting with each others.

4.2.1.1 MCS in imprecise probability theories.

We first study generic fusion rules applying the notion of maximal coherent subset in the different theories of uncertainty.

MCS with credal sets Let $\mathcal{P}_1, \dots, \mathcal{P}_N$ be N credal sets modeling the source information. $K \subset \llbracket N \rrbracket$ is a MCS if it is such that $\bigcap_{i \in K} \mathcal{P}_i \neq \emptyset$ and if it is maximal with this property. Let K_1, \dots, K_k be the MCS of the N credal sets. The MCS fusion rule resulting in the credal set $\mathcal{P}_{MCS(1:N)}$ is defined as:

$$\mathcal{P}_{MCS(1:N)} = \mathcal{H} \left(\bigcup_{j=1}^k \bigcap_{i \in K_j} \mathcal{P}_i \right) \quad (4.23)$$

the credal set $\mathcal{P}_{MCS(1:N)}$ generally fails Properties of associativity II, strong zero preservation VI, strong maximal plausibility VIII and robustness XII and satisfies all the other prop-

erties of Section 4.1.2 (see Table 4.1). The MCS rule giving $\mathcal{P}_{MCS_{(1:N)}}$, as well as some variations have been studied by Walley [202]. It satisfies all the properties that he regards as desirable, but he points out the fact that, in some situations, the failing of robustness (Prop. XII) for this particular rule can be problematic.

The notion of maximal coherent subsets is also used in other works done in the framework of imprecise probabilities, although it is not mentioned explicitly: the result of the aggregation procedure proposed by Troffaes [194] can be seen as a convex combination of maximal coherent subsets of credal sets, and Moral and Sagrado [154] use maximal coherent subsets as a step in their fusion procedure.

MCS with random sets Let $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ be N random sets modeling the source information. A random set $(m, \mathcal{F})_{MCS_{(1:N)}}$ resulting from the application of the MCS rule is given by the following steps:

1. Build a joint bpa $m_{(1:N)}$ satisfying Equation (4.7).
2. For each joint mass $m_{(1:N)}(\times_{i=1}^N E_i)$, $K \subset \llbracket N \rrbracket$ is a MCS if it such that $\cap_{i \in K} E_i \neq \emptyset$ and if it is maximal with this property. Let K_1, \dots, K_k be the MCS for this joint mass.
3. Allocate the joint mass $m_{(1:N)}(\times_{i=1}^N E_i)$ to the set $\cup_{j=1}^k \cap_{i \in K_j} E_i$, with $E_i \in \mathcal{F}_i$ for $i = 1, \dots, N$.

Again, it can be checked that $(m, \mathcal{F})_{MCS_{(1:N)}}$ generally fails Properties of associativity II, strong zero preservation VI, strong maximal plausibility VIII, robustness XII and convexity XI (which can be satisfied by taking the convex hull of $(m, \mathcal{F})_{MCS_{(1:N)}}$). That it satisfies Property of idempotence IV depends on how the joint bpa is built, and it satisfies all the other properties of Section 4.1.2 (see Table 4.2). If $m_{(1:N)}(\times_{i=1}^N E_i) = \prod_{i=1}^N m_i(E_i)$ and $N = 2$, then we retrieve Dubois and Prade rule of combination [84].

As an application of the notion of maximal coherent subset in random set theory, we can mention the work of Ayoun and Smets [8], who considers MCS not on focal elements, but on sources (i.e., they divide sources in consistent subgroups, since in the application sources potentially consider different objects).

MCS with possibility distributions Let π_1, \dots, π_N be the N possibility distributions modeling the source information. $K \subset \llbracket N \rrbracket$ is a MCS if it is such that $\min_{i \in K} \pi_i \neq \emptyset$ and if it is

maximal with this property. The possibility distribution $\pi_{MCS(1:N)}$ resulting from the application of the MCS fusion rule is:

$$\pi_{MCS(1:N)} = \max_{j=1}^k \min_{i \in K_j} \pi_i \quad (4.24)$$

With K_j the different MCS. Again, it can be checked that $\pi_{MCS(1:N)}$ generally fails Properties of associativity II, strong zero preservation VI, strong maximal plausibility VIII, robustness XII and convexity XI (which can be satisfied by taking the convex hull of $\pi_{MCS(1:N)}$). It satisfies all the other properties of Section 4.1.2 (see Table 4.3). In the above rule, operators min and max can be replaced respectively by any t-norm and its dual t-conorm, but Property of idempotence IV would not be satisfied anymore.

4.2.1.2 Maximal coherent subsets (MCS) in practice.

The above rules have the advantages that they try to gain a maximal amount of information while taking account of all sources. They also need a minimal amount of information about these sources: in possibility theory and imprecise probability theory, both Equations (4.24) and (4.23) are parameters free, and can be applied with only the source information.

Within random set theory, applying the MCS fusion rule requires to build a joint bpa, which implies making some assumptions on the interactions existing between sources. In absence of such information, a cautious approach following the least-commitment principle (LCP) can be adopted.

As conceptually attractive and elegant as it may be, applying a fusion rule based on MCS usually requires an important computational effort, which is an important drawback in practical applications where calculation time is a critical issue or where information is provided by a lot of sources. Similarly to logic, extracting coherent subsets from arbitrary spaces is of exponential complexity (actually, the problem is NP-complete, see for example Malouf [140]), and this complexity adds up to the complexity of applying conjunctive and disjunctive fusion rules in the considered theory.

To reduce this complexity and increase the tractability of MCS based methods, one can either build algorithms providing approximate solutions that converge to the true solution (e.g. by using MCMC like methods [210]), or use MCS methods in a restricted framework in which computations are easier to achieve. In the next section, we adopt the second solution, by studying a MCS method that applies to information modeled by convex possibility distributions on

the real line. Although restrictive, this framework is likely to be useful in many real-life applications, where variables and parameters take real values and where possibility distributions model a limited amount of information.

4.2.2 Level-wise MCS on the real line with possibility distributions

In this section, we consider the case where the information provided by the N sources concerns a variable X taking its value on the real line \mathbb{R} and can be modeled by N convex possibility distributions π_i , $i = 1, \dots, N$. It can be, for instance, N experts providing nested intervals with their confidence levels.

We propose to summarize or fusion the information by applying a level-wise MCS method to the N distributions. That is, for each level $\alpha \in [0, 1]$, we apply an MCS method to the N α -cuts, from which we retrieve a fuzzy random variable, or fuzzy belief structure $(m, \widetilde{\mathcal{F}})$. Such a method can be seen as an extension of Equation (4.24) or as a particular case of the MCS method in random set theory, where a complete correlation is assumed between α -cuts. This makes the method consistent with the relation (recalled in Section 4.1.3.5) existing between conjunctive and disjunctive rules of possibility theory and of random set theory. Moreover, this particular setting allows for fast and easy computations. In the sequel, we use the following example to illustrate the proposed MCS fusion rule:

Example 4.1. Four sources (experts, computer code, sensor, ...) provide information in term of a best-estimate and a conservative interval, and the possibility distributions are supposed to have trapezoidal shapes. The information, represented in Figure 4.1, is summarized in Table 4.4.

Table 4.4: Information of Example 4.1 sources

| Source | Conservative interval | Best estimate |
|--------|-----------------------|---------------|
| 1 | [1,5] | [2,4] |
| 2 | [1,13] | [3,6] |
| 3 | [3,11] | [7] |
| 4 | [5,13] | [10,12] |

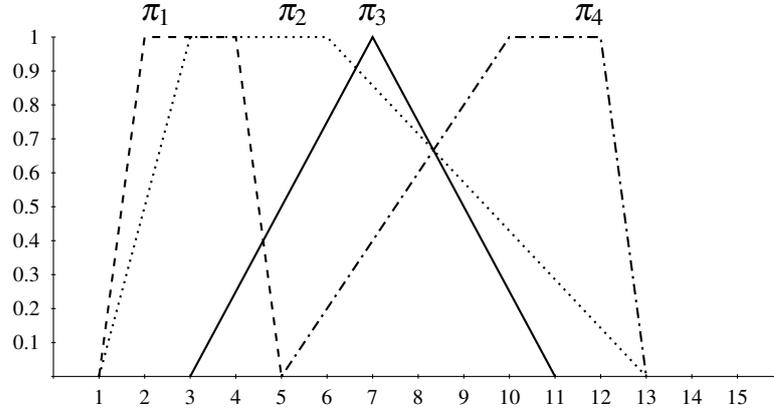


Figure 4.1: Example 4.1 distributions

4.2.2.1 Extracting MCS of intervals

We first recall results found by Dubois *et al.* [76, 77] and concerning the case where sources supply N intervals $I_i = [a_i, b_i]$, $i = 1, \dots, N$. In this case, the MCS are the subsets $K_j \subset \llbracket N \rrbracket$ of sources such that $\bigcap_{i \in K_j} I_i \neq \emptyset$, and applying the MCS method to such intervals is equivalent to find $I_{MCS} = \bigcup_j \bigcap_{i \in K_j} I_i$, which is usually a union of disjoint intervals. Dubois *et al.* [76, 77] show that Algorithm 4, described below, allows to extract subsets K_j of coherent sources from the N intervals I_i .

Algorithm 4: Maximal coherent subsets of intervals

Input: N intervals

Output: List of m maximal coherent subsets K_j

List = \emptyset ;

j=1 ;

$K = \emptyset$;

Order in an increasing order $\{a_i | i = 1, \dots, N\} \cup \{b_i | i = 1, \dots, N\}$;

Rename them $\{c_i | i = 1, \dots, 2N\}$ with $type(i) = a$ if $c_i = a_k$ and $type(i) = b$ if $c_i = b_k$;

for $i = 1, \dots, 2N - 1$ **do**

if $type(i) = a$ **then**

 Add Source k to K s.t. $c_i = a_k$;

if $type(i+1) = b$ **then**

 Add K to List ($K_j = K$) ;

 j = j + 1 ;

else

 Remove Source k from K s.t. $c_i = b_k$;

Algorithm 4 is based on increasingly sorting the interval end-points into a sequence $(c_i)_{i=1, \dots, 2N}$

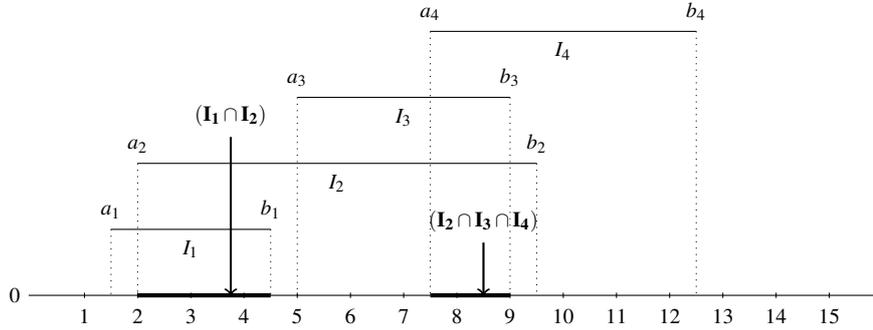


Figure 4.2: Maximal coherent subsets on Intervals (0.5-cuts of Example 4.1)

that is scanned in this order. Each time (and only then) it meets an element c_i of type a , (i.e. the lower bound of an interval) followed by an element c_{i+1} of type b (i.e. the upper bound of another interval), a maximally coherent set of intervals is obtained. Once end-points of intervals $I_i, i = 1, \dots, N$ have been sorted, Algorithm 4 complexity is linear in the number N of sources, whereas extracting maximal coherent subsets is generally of exponential complexity. This greater efficiency is due to the facts that there is a natural ordering between real numbers and that we consider intervals. Algorithm 4 could thus be easily adapted to any similar situation (i.e. ordered space \mathcal{X} where we consider sets I_i containing all elements respectively higher and lower than their lowest and greatest elements).

Figure 4.2 illustrates the situation for α -cuts of level 0.5 of Example 4.1. Using Algorithm 4, we find two maximal coherent subsets : $K_1 = \{I_1, I_2\}$ and $K_2 = \{I_2, I_3, I_4\}$. After applying the maximal coherent subset method, the result is $(I_1 \cap I_2) \cup (I_2 \cap I_3 \cap I_4) = [2, 4.5] \cup [7.5, 9]$, as pictured in bold lines on the figure. They can be thought of as the most likely intervals where the unknown value may lie.

4.2.2.2 Level-wise MCS on possibility distributions

We now consider that sources provide N possibility distributions $\pi_i, i = 1, \dots, N$ whose α -cuts are intervals (i.e. possibility distributions π_i are formally equivalent to fuzzy intervals). This means that, for each level $\alpha \in [0, 1]$, their α -cuts form a set of N intervals $E_{i,\alpha}$, with $E_{i,\alpha}$ the α -cut of π_i . It is then possible to apply Algorithm 4 to them : Let $K_{j,\alpha}$ be the maximal subsets of coherent intervals such that $\bigcap_{i \in K_{j,\alpha}} E_{i,\alpha} \neq \emptyset$. Define $E_{MCS,\alpha}$ as the union of the

partial results associated to $K_{j,\alpha}$, as suggested by Dubois and Prade [93] :

$$E_{MCS,\alpha} = \bigcup_{j=1,\dots,f(\alpha)} \bigcap_{i \in K_{j,\alpha}} E_{i,\alpha} \quad (4.25)$$

where $f(\alpha)$ is the number of distinct maximal subsets $K_{j,\alpha}$ of coherent intervals at level α . In general, $E_{MCS,\alpha}$ is a union of disjoint intervals, and it does not hold that $E_{MCS,\alpha} \supset E_{MCS,\beta}$ for any two values $\beta, \alpha \in [0, 1]$ such that $\beta > \alpha$. So, the result is not a possibility distribution, since the sets $E_{MCS,\alpha}$ are not nested.

In practice, for a finite collection of distributions π_i , there is a finite set of $p + 1$ values $0 = \beta_1 \leq \dots \leq \beta_p \leq \beta_{p+1} = 1$ such that the sets $E_{MCS,\alpha}$ will be nested for values $\alpha \in (\beta_k, \beta_{k+1}]$, $k = 1, \dots, p$. Algorithm 5 gives a simple method to compute these threshold values β_k . It simply computes the height of $\min(\pi_i, \pi_j)$ for every pair of possibility distributions π_i, π_j . This value is the threshold above which π_i and π_j do not belong to the same coherent subset anymore.

Algorithm 5: Values β_k of fuzzy belief structure

Input: N possibility distributions π_i

Output: List of values β_k

List = \emptyset ;

i=1 ;

for $k = 1, \dots, N$ **do**

for $l = k + 1, \dots, N$ **do**

$\beta_i = \max(\min(\pi_k, \pi_l))$;

 i=i+1 ;

if β_i not in List **then**

 Add β_i to List ;

Order List by increasing order ;

By applying Equation (4.25) for all levels $\alpha \in (\beta_k, \beta_{k+1}]$, we retrieve a non-normalized fuzzy set F_k with membership range $(\beta_k, \beta_{k+1}]$, since sets $E_{MCS,\alpha}$ are nested in that range. We note \tilde{F}_k the normalized fuzzy set obtained by changing, for all $x \in \mathbb{R}$, the membership function $v_{F_k}(x)$ of F_k into

$$v_{\tilde{F}_k}(x) = \frac{\max(v_{F_k}(x) - \beta_k, 0)}{\beta_{k+1} - \beta_k}$$

that is, we expand the range $(\beta_k, \beta_{k+1}]$ to $[0, 1]$. If we now assign a weight $m_k = \beta_{k+1} - \beta_k$ to \tilde{F}_k , and do this for every $k = 1, \dots, p$, the result is a random fuzzy variable that we note $(m, \mathcal{F})_{MCS}$, whose focal elements are normalized fuzzy sets \tilde{F}_k with weights m_k . Weight m_k can be interpreted as the confidence given to adopting \tilde{F}_k as the information provided by all

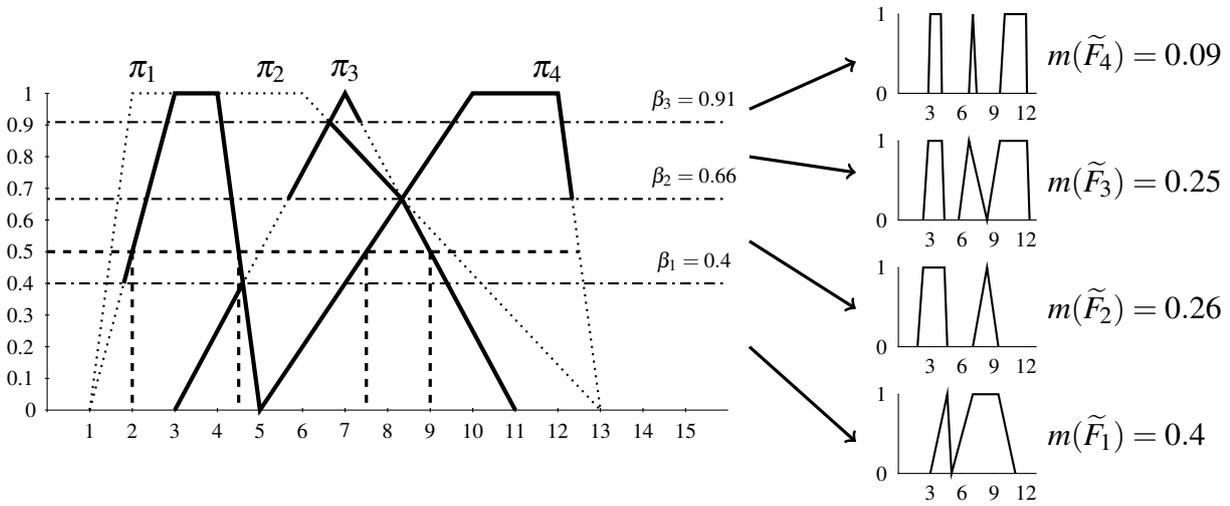


Figure 4.3: Result of MCS method on Example 4.1 (—) and 0.5-cut (---)

the sources.

Figure 4.3 illustrates the result of applying the method to Example 4.1. The 0.5-cut is exactly the result of Figure 4.2. The result of Equation (4.25) for each level $\alpha \in (0, 1]$ is in bold.

Note that in Equation (4.25), if we consider only the maximal coherent subsets $K_{j,0}$ of level 0 (i.e. distributions whose support intersects) and build for every level $\alpha \in [0, 1]$ the set $E_{MCS_0,\alpha}$ defined as

$$E_{MCS_0,\alpha} = \bigcup_{j=1,\dots,f(0)} \bigcap_{i \in K_{j,0}} E_{i,\alpha}$$

then we retrieve a random set whose focal elements are nested and are the α -cuts of the possibility distribution resulting from Equation (4.24). This shows that the proposed method is an extension of the possibilistic MCS fusion rule, where we allow maximal coherent subsets to evolve with level α . Similarly, the method is a particular instance of MCS fusion rule in random set theory, where a positive total correlation between α -cuts is assumed. The link with credal sets MCS fusion rule is less clear, except if we consider random sets as 2^{nd} order imprecise probabilistic models, where 2^{nd} order models are precise probabilities (i.e. the bpas) and the 1^{st} order model is an interval modeling our knowledge (i.e. the focal sets).

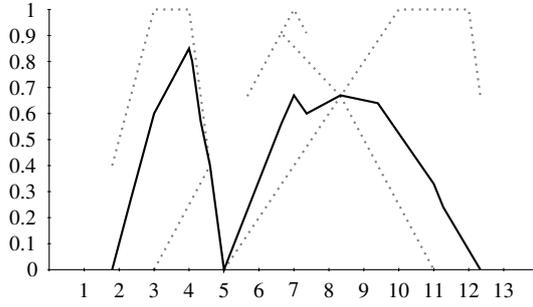
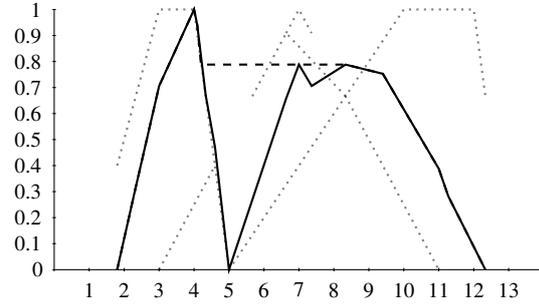

 Fig.4.4.A Distribution π_c


Fig.4.4.B Normalized (—) and convexified (---)

Figure 4.4: Contour function π_c extracted from Example 4.1, with fuzzy focal elements (gray lines)

4.2.2.3 Exploiting the fuzzy belief structure

From Figure 4.3, we can see that the fuzzy belief structure resulting from the MCS method is likely to be hard to use in practice, or to be interpreted by non-experts. It is therefore desirable to have some tools that allow to extract summarized and useful information from the whole structure. Such tools are proposed here.

Building the contour function. Summarizing the information in a more synthetic model such as a unique possibility distribution allow to provide an analyst or decision maker with a simplified and more interpretable model. We propose to take the contour function π_c of the fuzzy belief structure $(m, \tilde{\mathcal{F}})_{MCS}$, that is

$$\forall x \in \mathcal{X}, \quad \pi_c(x) = Pl(\{x\}) = \sum_{i=1}^p m_i v_{\tilde{F}_i}(x), \quad (4.26)$$

which boils down to computing the weighted arithmetic mean of the membership functions of the fuzzy focal sets \tilde{F}_i . If needed, one can then normalize this distribution π_c (by computing $\pi'_c(x) = \pi_c(x)/h(\pi_c)$ where $h(\pi_c)$ is the height of π_c) and/or take its convex hull.

Figures 4.4.A and 4.4.B respectively show the contour function and its normalized and convexified versions that can be extracted from Example 4.1, together with the fuzzy focal sets in the background. As we can see, the result is a bimodal distribution with one mode centered around value 8 and the other with a value of 4, this last value being the most plausible. This is so because these areas are the only ones supported by three sources whose information are highly (even if not totally) consistent. We can expect that the true value lies in one of these

two areas, but it is hard to tell which one. Indeed, in this case, one should either take the normalized convex hull of π_c as the final representation of the parameter X , or find out the reason for the conflict (if feasible).

If we consider distribution π_c as the result of a fusion operator φ applied to distributions π_1, \dots, π_N , then it satisfies Properties of commutativity (III), idempotence (IV), insensibility to relative ignorance (X), strong reconciliation (IX), weak maximal plausibility (VII) and weak zero preservation (V) (see table 4.3). It is also less sensible to small changes than other rules and satisfy all the requirements advocated by Oussalah *et al.* [163] and Delmotte [55]. Also note that the same properties, this time for table 4.2, are also satisfied if we consider that the random set equivalent to the fuzzy belief structure $(m, \widetilde{\mathcal{F}})_{MCS}$ (see Equation (3.61)) is the result of fusing the random sets $(m, \mathcal{F})_{\pi_1}, \dots, (m, \mathcal{F})_{\pi_N}$ corresponding to π_1, \dots, π_N . Moreover, in this case, $(m, \widetilde{\mathcal{F}})_{MCS}$ also satisfy Property of total consistency (I).

Extracting subgroups of coherent sources For each threshold in $(\beta_k, \beta_{k+1}]$, $k = 1, \dots, p$, Algorithm 4 exploits the same MCS $K_j^{(\beta_k, \beta_{k+1}]}$ of sources. Changing the value of this threshold yields a finite collection of coverings of the set of sources. Increasing the threshold from 0 to 1, we go from the largest sets of agreeing sources (i.e. those sources for which the supports of distributions π_i intersect), to the smallest sets of agreeing sources (i.e. those for which cores intersect). Subsets $K_j^{(\beta_k, \beta_{k+1}]}$ can be interpreted as clusters of sources that agree up to a confidence level β_{k+1} .

Analyzing these clusters can give some information as to which groups of sources are consistent, i.e. agree together with a high confidence level (possibly using some common evidence to supply information) and which ones are strongly conflicting with each other (and which items of information are plausibly based on different pieces of evidence). The finite collection of groups from Example 4.1 are summarized in the following table

| Subsets | Clusters | Max. Conf. level |
|-------------------|--------------------|------------------|
| $K^{(0,0.4]}$ | [1, 2, 3][2, 3, 4] | 0.4 |
| $K^{(0.4,0.66]}$ | [1, 2][2, 3, 4] | 0.66 |
| $K^{(0.66,0.91]}$ | [1, 2][2, 3][4] | 0.91 |
| $K^{(0.91,1]}$ | [1, 2][3][4] | 1.0 |

In this example, not much can be concluded from these clusters of coherent sources. Nevertheless, presenting the information in this form seems natural, and can trigger further investi-

gation as to why some sources seems to be more conflicting with the others (such as source 4 in our example).

Measuring the gain of information An interesting piece of information to have is how much precision has been gained by the fusion process. We consider that the overall imprecision of the information provided by all the sources is equal to

$$IP = |\pi_{\max(1:N)}| = \int_X \pi_{\max(1:N)}(x) dx$$

with $|\pi_{\max(1:N)}|$ the fuzzy cardinality of $\pi_{\max(1:N)}$. If we now consider the fuzzy belief structure $(m, \widetilde{\mathcal{F}})_{MCS}$ resulting from the fusion process, its imprecision can be measured as

$$IP' = \sum_{k=1}^p m_k |\widetilde{F}_k|$$

The difference $GP = IP - IP'$ quantifies the precision gained due to the fusion process. This index is 0 in case of total conflict and when the sources supply the same possibility distribution. Indeed, the MCS method increases the precision when sources are consistent with one another but supply distinct pieces of information. In Example 4.1, we have $IP = 11.195$, $IP' = 5.412$ and the normalized index $(IP - IP')/IP$ is 0.52, which indicates a reasonable gain of precision after fusion.

Since the fusion process follows a level-wise application of the MCS method, it is natural to investigate the level-wise counterpart of both IP and IP' . That is, we can compute, for each threshold $\alpha \in (0, 1]$

$$IP(\alpha) = |E_{\max(1:N),\alpha}| \quad IP'(\alpha) = |E_{MCS,\alpha}|$$

with $E_{\max(1:N),\alpha}$ the α -cut of $\pi_{\max(1:N)}$. These evaluations depending on α , they can be seen as gradual numbers [95, 109]. Recall that a gradual number is formally a mapping from $(0, 1]$ to the real line \mathbb{R} , such as $IP(\alpha)$ and $IP'(\alpha)$. $IP(\alpha)$ measures the imprecision of the continuous belief structure $(m, \mathcal{F})_{\pi_{\max(1:N)}}$ with uniform distribution on $[0, 1]$ and which assigns to each $\alpha \in [0, 1]$ the set $E_{\max(1:N),\alpha} = \bigcup_{i=1,\dots,N} E_{i,\alpha}$, that is $IP(\alpha)$ gradually evaluate the imprecision of the belief structure resulting from the level-wise disjunction of α -cuts. It is a monotonic gradual number, since the disjunctions of α -cuts are nested. The gradual number $IP'(\alpha)$ measures the imprecision of $(m, \widetilde{\mathcal{F}})_{MCS}$ likewise. However it is generally neither continuous nor monotone (cuts are not nested). The gradual number $GP(\alpha) = IP(\alpha) - IP'(\alpha)$ is thus a

level-wise measure of the precision gained by applying the maximal coherent subset method. The following equality formalize their link with their scalar counterparts IP, IP' and GP :

$$IP = \int_0^1 IP(\alpha) d\alpha,$$

and likewise for IP' and GP . Since $m_k|\tilde{F}_k| = \int_{\beta_{k-1}}^{\beta_k} |E_{MCS,\alpha}| d\alpha$, we effectively have $IP' = \int_0^1 IP'(\alpha) d\alpha$. The validity of the other equality $IP = \int_0^1 IP(\alpha) d\alpha$ follows from the definition of fuzzy cardinality.

Measuring the confidence in an event, in a source Once $(m, \tilde{\mathcal{F}})_{MCS}$ has been accepted as a good representative of the information provided by the sources, plausibility and belief measures of an event A , given by Equations (3.62) and (3.63), provide natural upper and lower confidence levels given by the group of sources to this event. In particular, if $A = \pi_i$, plausibility and belief can be used to evaluate the resulting upper and lower “trust” in the information given by source i in view of all the sources.

In Example 4.1, values $[Bel(\pi_i), Pl(\pi_i)]$ for sources 2 and 4 are, respectively, $[0.38, 1]$ and $[0, 0.93]$ (using, for example, Equations (3.62)-(3.63)). We see that information provided by source 2 is judged totally plausible by the group, and also strongly supported (source 2 is undoubtedly the less conflicting of the four). Because one source completely disagrees with source 4, its belief value drops to zero, but the information delivered by it is nevertheless judged fairly plausible (since source 4 is not very conflicting with sources 2 and 3).

Although belief and plausibility functions are natural candidates to measure the overall confidence in a source, there are cases where their informativeness will be poor. For instance, if a distribution π_i is in total conflict with the others, the resulting fuzzy belief structure $(m, \tilde{\mathcal{F}})_{MCS}$ will give the following measures for π_i : $[Bel(\pi_i), Pl(\pi_i)] = [0, 1]$ (total ignorance). It means that in the presence of strong conflict, the MCS method grants no confidence in individual sources, even though no source can be individually discarded.

An alternative to reduce this imprecision is to use a fuzzy equivalent [167] of the so-called pignistic probability [187] (see Appendix C), namely

$$BetP_{MCS}(A) = \sum_{k=1}^p m(\tilde{F}_k) \frac{|\min(\tilde{F}_k, v_A)|}{|\tilde{F}_k|} \quad (4.27)$$

with v_A the membership functions of (fuzzy) event A , and $|\min(\tilde{F}_k, v_A)|/|\tilde{F}_k|$ the degree of

subsethood, or relative cardinality, of \tilde{F}_k in A . $BetP_{MCS}(A)$ is zero if A is strongly conflicting with every focal set \tilde{F}_k and one if every \tilde{F}_k is included in A (here, \tilde{F}_k is included in A iff $v_{\tilde{F}_k}(x) < v_A(x) \forall x \in \mathbb{R}$). In Example 4.1, Equation (4.27) applied to sources 2 and 4 ($A = \pi_2$ and $A = \pi_4$) respectively gives confidence 0.80 and 0.49, confirming that source 2 is more trusted by the group than source 4.

Note that other formulas instead of $|\min(\tilde{F}_k, v_A)|/|\tilde{F}_k|$ could have been chosen to measure the subsethood of \tilde{F}_k in A [54, 97]. One could also choose to consider the continuous random set associating set $E_{MCS,\alpha}$ to each level $\alpha \in [0, 1]$ and to use the continuous extension of the pignistic probability proposed by Smets [186], which would give yet another result. Further research is needed to know the properties of each of these measures and the relations existing between them, and it is presently not clear when to choose one measure rather than the others. From our standpoint, the important criteria satisfied by these measures is that they are consistent ways to measure the coherence of A with respect to the fuzzy belief structure coming out from the MCS method (e.g., in our example, source 2 would be judged more reliable than source 4, irrespectively of the chosen formula for the pignistic probability, and only the scalar evaluations would change).

All these tools (building the contour function, extracting coherent subgroups, measuring the precision gain and the resulting confidence in events) provide users and analyst with synthetic and interpretable messages. That such tools should be made available is important for future practical applications. Another important issue, not considered so far, is the ability of the method to take additional information about the sources into account. Methodologies to do so are proposed in the next section.

4.2.2.4 Taking additional information into account

We consider here three kind of additional information or assumptions concerning the sources and/or the space \mathcal{X} , that are often encountered in practice: the number of reliable sources, numbers quantifying the reliability of sources (possibly given by methods studied in Section 4.4), and the existence of a metric on space \mathcal{X} .

Number of reliable sources Suppose we have information on the number r of sources that can be expected to be reliable, or at least that some assumptions can be made about this

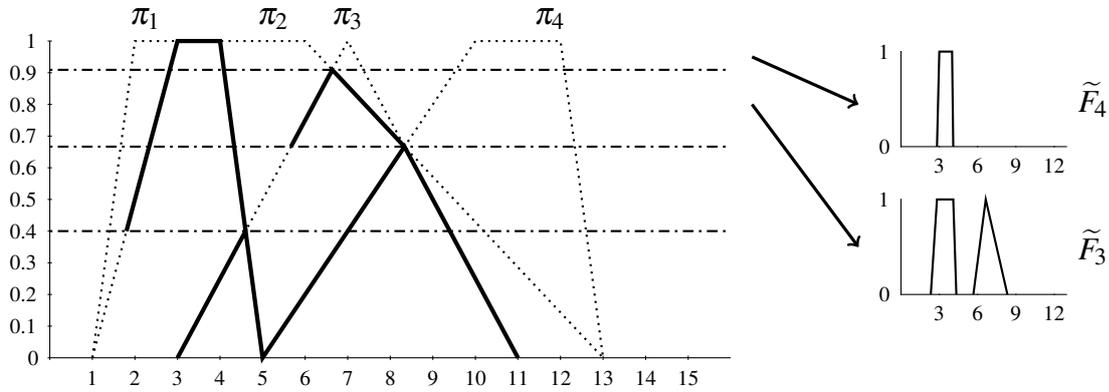


Figure 4.5: Result on Example 4.1 of MCS method with number of reliable sources $r = 2$

number. Given this number, we propose to adapt Equation (4.25) as follows

$$E_{MCS,r,\alpha} = \bigcup_{\substack{j=1,\dots,f(\alpha) \\ |K_{j,\alpha}| \geq r}} \bigcap_{i \in K_{j,\alpha}} E_{i,\alpha} \tag{4.28}$$

with $|K_{j,\alpha}|$ the number of sources in $K_{j,\alpha}$. For each level $\alpha \in [0, 1]$, only those subsets counting more than r sources are considered. This lessens the contribution of isolated sources or of small subgroups of consistent sources, and ensures that the result will be at least as informative as the result provided by the original method (Equation (4.25)). Figure 4.5 illustrates the fuzzy belief structure resulting from Example 4.1 when $r = 2$.

Accounting for the reliability of sources Suppose that some numerical evaluation of the reliability of each source is available. Denote λ_i the reliability of source i , and suppose, without loss of generality, that $\lambda_i \in [0, 1]$, value 1 meaning that the source is fully reliable, 0 representing a useless source. There are at least two ways of taking this reliability indices into account, the first one increasing the result imprecision by modifying (i.e. discounting) the possibility distributions, the second one decreasing the imprecision by discarding poorly reliable subgroups of sources:

- *Discounting*: discounting consists of transforming all $\pi_i, i = 1, \dots, N$ into distributions π'_i whose imprecision increases all the more as λ_i is low. In other words, the lower λ_i is,

the more irrelevant π_i becomes. A common discounting operation is the following:

$$\forall x \in \mathbb{R}, \quad \pi'_i(x) = \max(1 - \lambda_i, \pi_i(x))$$

Once discounted, sources are assumed to be reliable. The effect of the discounting operation on MCS method possesses a nice interpretation. Indeed, applying the MCS method to discounted sources means that the information modeled by π_i will only be considered for levels higher than $1 - \lambda_i$, since below that level, source i is present in every MCS K_j , as no information coming from it will be considered. A drawback of this method is that if values λ_i are low for each source, the result will be highly imprecise.

- *Discarding unreliable sources*: we propose to compute the overall reliability λ_K of a subgroup K as

$$\lambda_K = \perp_{i \in K}(\lambda_i) \quad (4.29)$$

where \perp is a t-conorm (here considered as an aggregation operator [13]). Choosing a particular t-conorm to aggregate reliability scores then depends on the assumed dependence between sources. For example, the maximum t-conorm $\perp(x, y) = \max(x, y)$ corresponds to the cautious assumption that agreeing sources are dependent (i.e. use the same information), thus the highest reliability score is not reinforced by the fact that multiple sources agree. On the contrary, the t-conorm $\perp(x, y) = x + y - xy$ (the dual of the product t-norm) can be associated to the hypothesis that sources are independent (reliability score increases as more sources agree together). A limit value $\underline{\lambda}$ can then be fixed, such that only subsets of sources having a reliability score over this limit are kept. Equation (4.25) then becomes

$$E_{MCS_r, \alpha} = \bigcup_{\substack{j=1, \dots, f(\alpha) \\ \lambda_{K_j, \alpha} \geq \underline{\lambda}}} \bigcap_{i \in K_j, \alpha} E_{i, \alpha}. \quad (4.30)$$

Remark that this method does not modify the pieces of information π_i .

Figures 4.6 and 4.7 show the result of applying the above methods (respectively discounting and discarding) to Example 4.1 when reliability indices are $\lambda_1 = 0.2$, $\lambda_2 = 0.6$, $\lambda_3 = 0.8$, $\lambda_4 = 0.2$. For the discounting method, we consider $X \in [1, 14]$, and for the discarding method, we consider independent sources and $\underline{\lambda} = 0.5$. Figure 4.6 well illustrates the higher imprecision that can result from applying a discounting method.

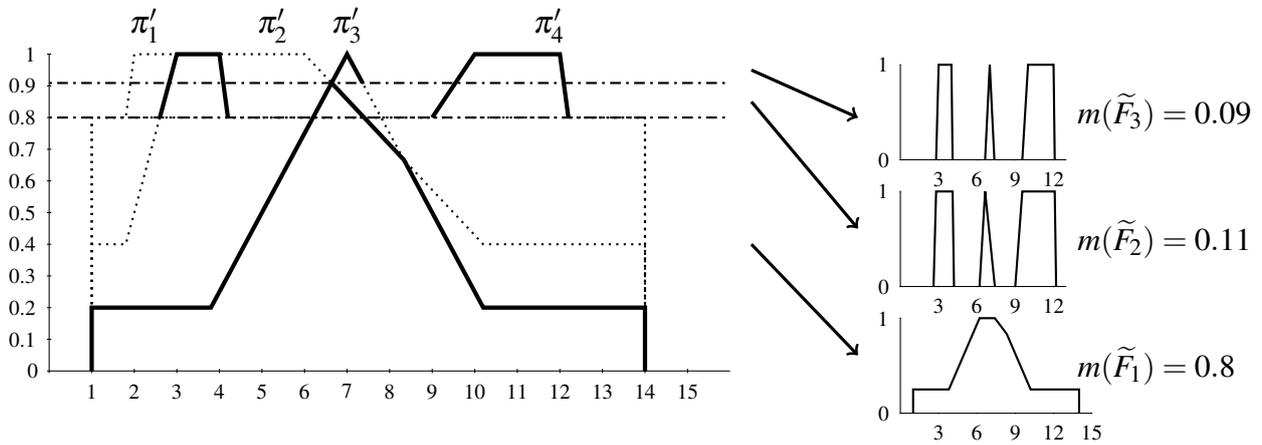


Figure 4.6: Result of MCS method on Example 4.1 with reliability scores $\lambda = (0.2, 0.6, 0.8, 0.2)$ and discounting method

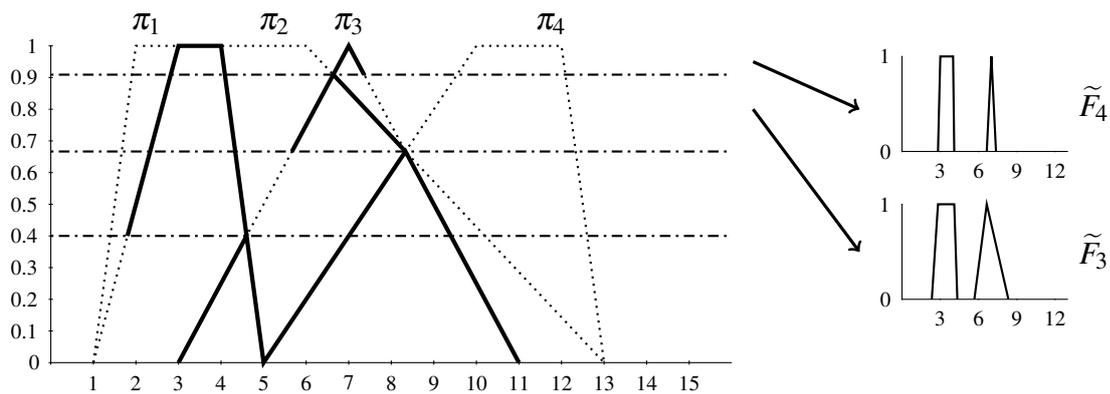


Figure 4.7: Result of MCS method on Example 4.1 with reliability scores $\lambda = (0.2, 0.6, 0.8, 0.2)$ and discarding method ($\perp(x, y) = x + y - xy$)

Accounting for the metric In the original MCS method described here, if an isolated source is totally conflicting with the others, then it will constitute a maximal coherent subset of its own. If the notion of distance on \mathcal{X} makes sense, this will be true whatever the distance of the isolated source distribution from the others is. However, in some applications [163], it is sometimes desirable to take the distance between distributions into account, with the aim of neglecting the information lying outside a certain zone.

Let $k_\alpha = \max_{j=1,\dots,f(\alpha)} |K_{j,\alpha}|$ be the maximal number of consistent sources at level α . Denote $E_{K_j,\alpha} = \bigcap_{i \in K_{j,\alpha}} E_{i,\alpha}$. At each level α , following Oussalah *et al.* [163], a so-called consensus zone can be defined as the interval:

$$E_{K,\alpha} = \mathcal{H} \left(\bigcup_{j, |K_{j,\alpha}|=k_\alpha} (E_{K_j,\alpha}) \right) = [\underline{k}_\alpha, \bar{k}_\alpha],$$

with \mathcal{H} the convex hull. Now, let $A = [\underline{a}, \bar{a}]$, $B = [\underline{b}, \bar{b}]$ be two intervals. We define the closeness $C(A, B)$ between A and B as

$$C(A, B) = \inf_{a \in A, b \in B} (d(a, b))$$

where $d(a, b)$ is the distance between two points a and b of the space \mathcal{X} . Note that $C(A, B)$ is not a distance (it does not satisfy triangle inequality), but is a measure of consistency between sets A and B accounting for the metric. Indeed, it will be 0 as soon as $A \cap B \neq \emptyset$. Since the proposed method emphasizes the concept of consistency, this choice appears sensible⁴. Moreover, between two thresholds β_k, β_{k+1} , the closeness $C(E_{K_j,\alpha}, E_{K_i,\alpha})$ between any two sets $E_{K_j,\alpha}, E_{K_i,\alpha}$ $i \neq j$ is an increasing function of α , due to the nestedness of these sets⁵.

Given the consensus zone $E_{K,\alpha}$, we can now fix a distance threshold d_0 and adapt Equation (4.25) in the following way, so that it takes account of the metric of \mathcal{X}

$$E_{MCS_d,\alpha} = \bigcup_{\substack{j=1,\dots,f(\alpha) \\ C(E_{K_j,\alpha}, E_{K,\alpha}) \leq d_0}} \bigcap_{i \in K_{j,\alpha}} E_{i,\alpha}. \quad (4.31)$$

applying Equation (4.31) means that information too far away from consensus zones are regarded as outliers and deleted. Figure 4.8 illustrates the method on Example 4.1 when $d_0 = 1$.

Overall, the proposed modifications allowing to take additional information into account only slightly modify the original method (Equations (4.28), (4.30) and (4.31) remain quite

⁴Genuine distances between sets like the Hausdorff distance are less meaningful in our context.

⁵this would not be true for the Hausdorff distance.

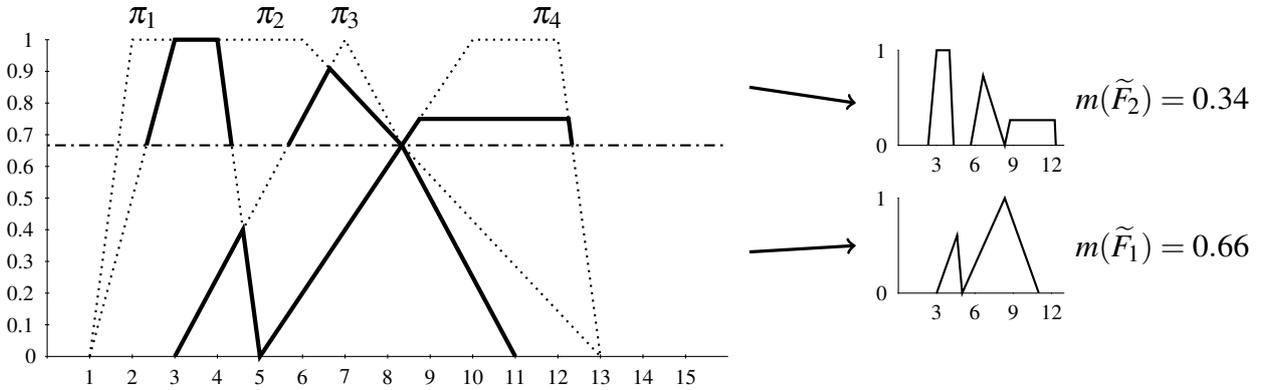


Figure 4.8: Result of MCS method on Example 4.1 taking metric into account with $d_0 = 1$

close to Equation (4.25)). It thus brings more flexibility to the method, while preserving its computational attractiveness.

4.3 Towards a cautious conjunctive rule in random set theory

We now turn to another common problem of information fusion, that is coping with dependencies between sources. We here look at the specific situation where uncertainty is represented by random sets and the dependencies between sources are badly known. Random set theory is an attractive theory to treat such a problem, since the construction of a joint bpa allows to model specific dependencies, while the Least-Commitment Principle (LCP, see Appendix C for details) provides guidance allowing to adopt a cautious attitude when some information are missing (here, information concerning the dependence between sources).

First recall that, given a random set (m, \mathcal{F}) , its expected cardinality $|C|_{(m, \mathcal{F})}$ is given by

$$|C|_{(m, \mathcal{F})} = \sum_{E \in \mathcal{F}} m(E) |E| \tag{4.32}$$

and is the simplest measure of the imprecision of (m, \mathcal{F}) . Also recall that this cardinality is equal to the one of the possibility distribution equivalent to the contour function (i.e., $|C|_{(m, \mathcal{F})} = \sum_{x \in \mathcal{X}} Pl(\{x\})$).

We consider two sources 1, 2, providing their information about X in terms of two random

sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$. Combining them by the means of Dempster's combination rule to obtain $(m, \mathcal{F})_{\otimes(1:2)}$ (the normalized version of $(m, \mathcal{F})_{\odot(1:2)}$) is justified only when sources can be considered as independent and reliable, that is when taking the stochastic product of m_1, m_2 to build the joint bpa $m_{(1:2)}$ and its conjunctive allocation appear reasonable.

When independence cannot be assumed and the dependence structure existing between sources is badly known, a solution to merge random sets conjunctively is to apply the "least commitment principle" to the merging of random sets. Such a cautious rule, denoted here \wedge (and $(m, \mathcal{F})_{\wedge(1:2)}$ the resulting random set), should at least satisfy the property of idempotence (Property IV of Table 4.2)

Following the LCP, $(m, \mathcal{F})_{\wedge(1:2)}$ should be chosen in the set $\mathcal{M}_{\cap(1:2)}$, so that it is one of the least x -committed element, with x one of the inclusion ordering in $\{s, pl, q, w, v, d, dd\}$. Again, we choose the s -ordering, for the same reasons as before: for its unifying characteristic, and because it is from a theoretical standpoint the most appealing notion of inclusion in random set theory. However, since the s -ordering is a partial ordering, the s -least committed element in $(m, \mathcal{F})_{\wedge(1:2)}$ is in general not unique.

In order to avoid such incomparabilities, we propose here to define the least-committed element $(m, \mathcal{F})_{\wedge(1:2)}$ as one with maximal expected cardinality. The reasons for choosing expected cardinality is that (i) it is the most simple way to measure imprecision of random sets and (ii) it is coherent with the s -ordering, that is a joint random set having maximum expected cardinality will be among the s -least committed ones. From this requirement follows the following proposition:

Proposition 4.6. *Let $(m, \mathcal{F})_1$ be a specialization of $(m, \mathcal{F})_2$, then the result of the least committed rule \wedge is $(m, \mathcal{F})_1 \wedge (m, \mathcal{F})_2 = (m, \mathcal{F})_{\wedge(1:2)} = (m, \mathcal{F})_1$.*

Proof. The result $(m, \mathcal{F})_{\wedge(1:2)}$ must be a specialization of both $(m, \mathcal{F})_1$ and $(m, \mathcal{F})_2$, by definition. The fact that $(m, \mathcal{F})_1$ is a specialization of $(m, \mathcal{F})_2$ implies that the set $\mathcal{M}_{\cap(1:2)}$ of possible solutions reduces to the specializations of $(m, \mathcal{F})_1$ (since every specialization of $(m, \mathcal{F})_1$ is also a specialization of $(m, \mathcal{F})_2$). And the specialization of $(m, \mathcal{F})_1$ that has maximal expected cardinality is $(m, \mathcal{F})_1$ itself. □

We tend to think that Proposition 4.6 should be satisfied by any rule tagged as cautious, and at the very least by those based on the s -ordering. Nevertheless, Proposition 4.6 concerns very peculiar cases, and does not provide (practical) guidelines as to how general random sets should be cautiously merged. In order to propose such practical guidelines, we will first recall

and then use the concept of commensurate bpas, first introduced by Dubois and Yager [101] to relate fuzzy connectives with random set conjunctive fusion rules. In the sequel of this section, we will allow a bpa to bear on non-distinct focal sets, that is we consider it as a relation between $\wp(\mathcal{X})$ and $[0, 1]$ (i.e., several masses can be assigned to the same subset). We will name a bpa whose focal elements are all distinct *regular*.

Definition 4.1. Let m be a bpa with focal sets A_1, \dots, A_n and associated weights m^1, \dots, m^n . A *split* of m is a bpa m' with focal sets A'_1, \dots, A'_n and associated weights m'^1, \dots, m'^n such that

$$\sum_{A'_i=A_i} m'^i = m^i$$

In other words, a split is a new bpa where the original weight given to a focal set is separated in smaller weights given to the same focal set, with the sum of weights given to a specific focal element being constant.

Definition 4.2. Two bpas m_1, m_2 , belonging to some random sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$ are said to be equivalent if, for any subset $E \subseteq \mathcal{X}$ $Pl_1(E) = Pl_2(E)$ and $Bel_1(E) = Bel_2(E)$.

And one can show [101] that two bpas are equivalent if and only if they are splits of a common regular bpa.

Definition 4.3. Let m_1, m_2 be two bpas with respective focal sets $\{A_1, \dots, A_n\}, \{B_1, \dots, B_k\}$ and associated weights $\{m_1^1, \dots, m_1^n\}, \{m_2^1, \dots, m_2^k\}$. Then, m_1 and m_2 are said to be commensurate if $k = n$ and there is a permutation σ of $\llbracket n \rrbracket$ such that $m_1^j = m_2^{\sigma(i)} \forall i = 1, \dots, n$.

Two bpas are commensurate if their distribution of weights over focal sets can be described by the same vector of numbers.

Algorithm 6 has been proposed by Dubois and Yager [101] to make any two (regular) bpa commensurate, given a prescribed ranking of focal elements and by successive splitting. Once this commensuration done, they propose a conjunctive merging rule, denoted here \oplus , resulting in a random set $(m, \mathcal{F})_{\oplus(1:2)} \in \mathcal{M}_{\cap(1:2)}$

The random set $(m, \mathcal{F})_{\oplus(1:2)}$ resulting from Dubois and Yager's rule then have, for $k = 1, \dots, m$, masses $m_{\oplus(1:2)}^k = m_{R_1}^k = m_{R_2}^k$ and focal elements $R_{\oplus(1:2)}^k = R_1^k \cap R_2^k$. The whole procedure is illustrated by the following example.

Example 4.2. Two random sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$, their commensuration and the result of Dubois and Yager's rule are summarized in the following table:

Algorithm 6: Commensuration Algorithm

Input: Random sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$ on \mathcal{X}

Output: Two commensurate random sets $(m, \mathcal{F})_{R_1}, (m, \mathcal{F})_{R_2}$ respectively equivalent to $(m, \mathcal{F})_1, (m, \mathcal{F})_2$

Choose an indexing of focal elements $\{A_1, \dots, A_n\}, \{B_1, \dots, B_k\}, A_i \in \mathcal{F}_1, B_i \in \mathcal{F}_2$

for $i = 1, \dots, n$ **do**

 | Compute $\alpha_i = \sum_{j=i}^n m_1(A_j)$

for $i = 1, \dots, k$ **do**

 | Compute $\beta_i = \sum_{j=i}^k m_2(B_j)$

Compute $\{\gamma_1, \dots, \gamma_m\} = \{\alpha_1, \dots, \alpha_n\} \cup \{\beta_1, \dots, \beta_k\}$ with $1 = \gamma_1 > \dots > \gamma_m > \gamma_{m+1} = 0$

for $k = 1, \dots, m$ **do**

 | Define $m_{R_1}^k = m_{R_2}^k = \gamma_k - \gamma_{k-1}$

 | Define $R_1^k = A_i$ such that $\alpha_i \geq \gamma_k > \alpha_{i+1}$

 | Define $R_2^k = B_j$ such that $\beta_j \geq \gamma_k > \beta_{j+1}$

| m_1 | | m_2 | |
|-------|----|-------|----|
| A_1 | .5 | B_1 | .6 |
| A_2 | .3 | B_2 | .2 |
| A_3 | .2 | B_3 | .1 |
| | | B_4 | .1 |

Alg. 6 \rightarrow

| l | $m_{R_i}^l$ | R_1^l | R_2^l | $R_{1 \oplus 2}^l$ |
|-----|-------------|---------|---------|--------------------|
| 1 | .5 | A_1 | B_1 | $A_1 \cap B_1$ |
| 2 | .1 | A_2 | B_1 | $A_2 \cap B_1$ |
| 3 | .2 | A_2 | B_2 | $A_2 \cap B_2$ |
| 4 | .1 | A_3 | B_3 | $A_3 \cap B_3$ |
| 5 | .1 | A_3 | B_4 | $A_3 \cap B_4$ |

This example shows that the resulting merged random set $(m, \mathcal{F})_{\oplus(1:2)}$ heavily depends on the chosen ranking of focal sets \mathcal{F}_1 and \mathcal{F}_2 . Actually, it can be shown that any conjunctively merged bba can be produced by following this procedure.

Definition 4.4. Two commensurate generalised bpas are said to be *equi-commensurate* if each of their focal sets has the same weight.

Equicommensurate bpas can be obtained in the same way as commensurate bpas are obtained with Algorithm 6, by successively splitting each original bpa until all weights are equal. Note that, provided the ordering on focal sets is the same, the result of applying Dubois and Yager’s rule to equicommensurate bpas remains $(m, \mathcal{F})_{\oplus(1:2)}$, and is still in $\mathcal{M}_{\cap(1:2)}$.

Proposition 4.7. Any merged bpa in $\mathcal{M}_{\cap(1:2)}$ can be reached by means of Dubois and Yager

rule using appropriate commensurate bpas equivalent to m_1 and m_2 and the two appropriate rankings of focal sets.

Proof. We assume masses (of marginal and merged bpas) are rational numbers. Let $(m, \mathcal{F}) \in \mathcal{M}_{\cap(1;2)}$ be the conjunctively merged bpa we want to reach by using Dubois and Yager's rule. Let $m(A_i, B_j)$ denote the mass allocated to $A_i \cap B_j$ in (m, \mathcal{F}) . It is of the form $k_{12}(A_i, B_j) \times 10^{-n}$ where k_{12}, n are integers. By successive splitting followed by a reordering of elements R_1^j , we can always reach m . For instance, let k_R be equal to the greatest common divisor of all values $k_{12}(A_i, B_j)$. Then, $k_{12}(A_i, B_j) = q_{ij} \times k_R$, for an integer q_{ij} . Then, it suffices to re-order elements R_1^k by a re-ordering σ such that for q_{ij} of them, $R_1^k = A_i$ and $R_2^{\sigma(k)} = B_j$. Then, by applying Dubois and Yager's rule, we obtain the result m . From a practical standpoint, restricting ourselves to rational numbers has no importance: rational numbers being dense in reals, this means that we can always get as close as we want to any merged bpa. \square

Given the above definitions and results, it seems natural to derive practical guidelines of a cautious merging rule \wedge by looking for appropriate rankings of focal sets so that the merged bpa obtained via commensuration has maximal expected cardinality. The answer is : rankings should be extensions of the partial ordering induced by inclusion (i.e. $A_i < A_j$ if $A_i \subset A_j$). This is due to the following result:

Lemma 4.1. *Let A, B, C, D be four sets such that $A \subseteq B$ and $C \subseteq D$. Then, we have the following inequality*

$$|A \cap D| + |B \cap C| \leq |A \cap C| + |B \cap D| \quad (4.33)$$

Proof. From the assumption, the inequality $|(B \setminus A) \cap C| \leq |(B \setminus A) \cap D|$ holds. Then consider the following equivalent inequalities:

$$\begin{aligned} |(B \setminus A) \cap C| + |A \cap C| &\leq |A \cap C| + |(B \setminus A) \cap D| \\ |B \cap C| &\leq |A \cap C| + |(B \setminus A) \cap D| \\ |A \cap D| + |B \cap C| &\leq |A \cap C| + |A \cap D| + |(B \setminus A) \cap D| \\ |A \cap D| + |B \cap C| &\leq |A \cap C| + |B \cap D| \end{aligned}$$

hence the inequality (4.33) is true. \square

When using equi-commensurate bpas, masses in the formula of expected cardinality can be factorized, and expected cardinality then becomes $|C|_{(m, \mathcal{F})_{\oplus(1;2)}} = m_{R_{1 \oplus 2}} \sum_{i=1}^l |R_{1 \oplus 2}^i| =$

$m_{R_1 \oplus 2} \sum_{i=1}^l |R_1^i \cap R_2^i|$, where $m_{R_1 \oplus 2}$ is the smallest mass enabling equi-commensuration. We are now ready to prove the following proposition

Proposition 4.8. *There exists a conjunctive merging rule \wedge constructing a random set $(m, \mathcal{F})_{\wedge(1:2)}$ by the commensuration method, where focal elements are ranked according to the partial order of inclusion, with $(m, \mathcal{F})_{\wedge(1:2)} \in \mathcal{M}_{\cap(1:2)}$ minimally committed for expected cardinality.*

Proof. Assume some $(m, \mathcal{F})_{\widehat{(1:2)}} \in \mathcal{M}_{\cap(1:2)}$ is minimally committed for expected cardinality. By Proposition 4.7, it can be obtained by commensuration. Let m_{R_1}, m_{R_2} be the two equi-commensurate bpas with n elements each derived from the two original bpas m_1, m_2 . Suppose that the rankings used display four focal sets $R_1^i, R_1^j, R_2^i, R_2^j$, $i < j$, such that $R_1^i \supset R_1^j$ and $R_2^i \subseteq R_2^j$. By Lemma 4.1, $|R_1^j \cap R_2^j| + |R_1^i \cap R_2^i| \leq |R_1^j \cap R_2^i| + |R_1^i \cap R_2^j|$. Hence, if we permute focal sets R_1^i, R_1^j before applying Dubois and Yager's merging rule, we end up with a merged bpa $m_{R'_1 \oplus 2}$ such that $|C|(m, \mathcal{F})_{R_1 \oplus 2} \leq |C|(m, \mathcal{F})_{R'_1 \oplus 2}$. Since any merged bpa can be reached by sufficient splittings of m_1, m_2 and by inducing the proper ranking of focal sets of the resulting bpas m_{R_1}, m_{R_2} , there is a merged bpa $(m, \mathcal{F})_{\widehat{(1:2)}} \in \mathcal{M}_{\cap(1:2)}$ maximizing expected cardinality can be reached by Dubois and Yager's rule, using rankings of focal sets in accordance with the inclusion ordering. \square

Nevertheless, ranking focal sets in accordance with the partial order induced by inclusion is neither a sufficient nor a necessary condition to obtain a merged random set with maximized expected cardinality (examples are provided by Destercke *et al.* [66]). Still, these first results about how to cautiously merge random sets by maximizing expected cardinality are promising because:

- They provide first practical guidelines to cautiously and conjunctively merge random sets with respect to the s -ordering.
- They are coherent with cautious conjunctive merging in possibility theory, that is when $(m, \mathcal{F})_1, (m, \mathcal{F})_2$ are consonant, $(m, \mathcal{F})_{\wedge(1:2)}$ corresponds to the minimum merging rule of possibility theory.
- They are coherent with q -least committed approach in the case of consonant bpas, and consequently coherent with other cautious approaches [99, 100]. Coherence with non-consonant cases remains to be explored.

A potential disadvantage of the proposed approach is that it does not seem fully coherent with a pl -least committed approach [66] (and thus, with imprecise probability theory). Con-

solidating these results and making them more attractive from a practical standpoint would require:

- to find more constraints to add to the ordering of focal elements, so that $(m, \mathcal{F})_{\wedge(1:2)}$ would be uniquely defined (possibly using results related to the contour function of random sets [87]),
- to compare the proposed rule to other existing approaches to merge non-independent sources in the context of random set theory [19, 61] (we have already made some first comparisons [66], but a more thorough and axiomatic comparison is needed)
- to verify which properties does the proposed rule satisfy, as, for example, associativity.

4.4 Assessing sources quality: a general framework

Disposing of evaluations of sources reliability can be useful in many ways: to select subgroup of sources, weight them in fusion processes or simply to send back the information to sources. When sources give information about observable parameters or variables, subjective assessments of reliability are usually not acceptable. In this section, we propose methods based on rational requirements allowing to evaluate sources reliability from past assessments of variables or parameters, whose true value has been subsequently known.

The rational requirements on which are based the methods presented in this section have first been considered by Cooke [28], who apply them to expert opinions in a probabilistic setting. Later, they have been considered by Sandri *et al.*[174], also to treat expert opinions, but in a possibilistic setting.

4.4.1 Rational requirements and general methodology

A means to objectively evaluate information delivered by sources is to center the evaluation on past performances of the same sources. This can be done by considering so-called *seed variables*. A seed variable is a variable whose exact value is not known by the source when it provides information about it, but is either known by the analyst or will subsequently be determined by physical experiments or other means. Here, we note a seed variable T and its (discrete) domain \mathcal{T} (if T is a physical variable, \mathcal{T} will often be the discretized real line). If

known or precisely observed, the exact value of T is noted t^* , but we allow here the value of a seed variable to be imperfectly known (i.e. represented by an uncertainty model).

A source is then evaluated by two quantitative criteria, called here *informativeness* and *calibration*, based on the information provided on seed variable T :

- **Informativeness** is a score measuring the precision of the information given by the source on T . The more precise the source, the higher informativeness.
- **Calibration** is a score measuring the coherence between the information given by the source about T and the knowledge the analyst has about T .

A good source is then a source that receives an high score, i.e., a source that is both informative and coherent with available knowledge. Following Cooke [28, Ch.9], a good evaluation method gives scores that:

1. reward sources that are both informative and well calibrated,
2. are relevant, that is are influenced only by observations or by knowledge on seed variables,
3. are meaningful, that is are comparable, irrespectively of the nature and number of seed variables.

In the following, we regard these rules as basic requirements that should follow evaluation procedures. They are sensible, and do not constraint too much the evaluation procedure.

4.4.2 Evaluating sources in probability

The probabilistic method to evaluate sources is based on the use of so-called scoring rules [138], which were originally introduced and used in subjective probability elicitation procedures. However, Cooke [28] argues quite convincingly that, inside probability theory, they are also well fitted to the evaluation of sources (and in particular, of experts).

Let T be the seed variable, taking its values on finite domain \mathcal{T} . First recall that, given two probability distributions $P_1 = \{p_{1,1}, \dots, p_{1,|\mathcal{T}|}\}$ and $P_2 = \{p_{2,1}, \dots, p_{2,|\mathcal{T}|}\}$, the Kullback-

Leibler (KL) divergence [133] (also called relative entropy⁶) of P_2 from P_1 reads

$$KL(P_1, P_2) = \sum_{i=1}^{|\mathcal{T}|} p_{1,i} \log \left(\frac{p_{1,i}}{p_{2,i}} \right) \quad (4.34)$$

In the probabilistic setting, a source S provide information modeled in terms of a probability distribution $P_S = \{p_{S,1}, \dots, p_{S,|\mathcal{T}|}\}$, where $p_{S,i}$ is the probability mass given to element t_i of \mathcal{T} .

Informativeness Let P_U be the uniform probability on \mathcal{T} , that is $p_{U,i} = 1/|\mathcal{T}|$ for $i = 1, \dots, |\mathcal{T}|$. The informativeness Inf_S of source S is then computed as

$$Inf_S = KL(P_S, P_U).$$

If multiple seed variables are used, then the resulting informativeness is simply the arithmetic mean over all informativeness scores.

Calibration If r observations of the seed variable T are available, then let r_i be the number of observations corresponding to element t_i . To these observations corresponds an empirical probability $P_R = \{p_{R,1}, \dots, p_{R,|\mathcal{T}|}\}$, with $p_{R,i} = r_i/r$. The KL-divergence $KL(P_R, P_S)$ then provides a measure of the closeness of P_S , the source of information, from the empirically built distribution P_R . This divergence has value 0 if and only if $P_R = P_S$. The final calibration score Cal_S of source S is then derived by the following statistical hypothesis test

$$Cal_S = 1 - \chi_{|\mathcal{T}|-1}^2(2 * r * KL(P_R, P_S))$$

with $\chi_{|\mathcal{T}|-1}^2$ a chi-square distribution with $|\mathcal{T}| - 1$ degrees of freedom. Note that this measure is based on a convergence property of the KL divergence.

When considering multiple seed variables T_1, \dots, T_M taking only one value on a real-bounded domain, Cooke [28] proposes to model sources information on each seed variable by the same set of inter-percentiles $P_q = \{p_{q,1}, \dots, p_{q,B}\}$ extracted from a set of percentiles q_1, \dots, q_{B+1} (typically, the 0,5,50,95 and 100 % percentiles), with $p_{q,i} = q_{i+1} - q_i$. This set of percentiles is induced by the source information for each seed variable, and even if the values of the real line to which corresponds the percentiles q_1, \dots, q_{B+1} can be different for each seed

⁶Also sometimes quoted as KL-distance, although it does not satisfy the property of symmetry

variables, distribution P_q will remain the same. We note $I_{S,i,j} = [q_{S,i,j}, q_{S,i+1,j}]$ the interval corresponding to the i^{th} inter-percentile extracted from source S information concerning the j^{th} seed variable T_j and $Q_{S,j} = \{q_{S,1,j}, \dots, q_{S,B+1,j}\}$ the set of boundary values. Once the M values t^*_1, \dots, t^*_M assumed by the seed variables T_1, \dots, T_M have been observed, the empirical distribution $P_R = \{p_{R,1}, \dots, p_{R,B}\}$ is build in the following way: the value $p_{R,i}$ is r_i/M , that is the number of observations that are in the interval $I_{S,i,j}$, $j = 1, \dots, M$. Thus, for every source, we have $P_S = P_q$, and a different P_R for each source. Note that, to compute informativeness in such cases, the discretized uniform distribution P_U is computed with the same procedure as P_R .

The following example gives a short illustration of the method:

Example 4.3. Consider two economists S_1, S_2 (the sources) which are asked their opinion about the values of some portfolios in the two upcoming days (the seed variables). Denote these (unknown) values T_1, T_2 . Assume that the values evolve between $[4500, 5000]$. The analyst chooses to model information given by the means of percentiles $q_1 = 0\%, q_2 = 5\%, q_3 = 50\%, q_4 = 95\%, q_5 = 100\%$, forming the probability $P_q = \{0.05, 0.45, 0.45, 0.05\}$ ($B = 4$).

Information provided by the two economists are summarized as follows:

| | S_1 | S_2 |
|-------|--|--|
| T_1 | $Q_{S_1,1} = \{4500, 4700, 4800, 4950, 5000\}$ | $Q_{S_2,1} = \{4500, 4600, 4750, 4900, 5000\}$ |
| T_2 | $Q_{S_1,2} = \{4500, 4800, 4900, 4950, 5000\}$ | $Q_{S_2,2} = \{4500, 4650, 4750, 4850, 5000\}$ |

Assume now that the observations are $t^*_1 = 4680$ and $t^*_2 = 4752$, respectively for the first and second days. Then, the empirical distributions built from these observations are $P_R = \{1.0, 0, 0, 0\}$ for the first economist, and $P_R = \{0, 0.5, 0.5, 0\}$ for the second.

Global score The global score Sc_S of a source S is then computed as the product of informativeness and calibration scores

$$Sc_S = Inf_S Cal_S.$$

Cooke [28] also proposes to add a parameter acting as a minimal threshold of calibration, so that sources receiving too low calibration scores receive a global score of zero, thus avoiding the case of very precise but badly calibrated sources receiving an high global score. Since computed scores are then used to combine the distributions coming from the different sources

by arithmetic weighted mean, Cooke [28] also proposes some way to tune this threshold, so that the combined final probability distribution is optimized with respect to calibration criterium.

Within probability theory, this evaluation method is sound and well justified. However, it has a certain number of drawbacks, among which:

- The need of at least 10 seed variables or observations for the statistical test used in calibration to be robust and discriminative
- The propensity to confuse imprecision and variability in a single representation, as emphasized by Sandri *et al.* [174]
- The fact that, when different seed variables are considered (as in Example 4.3), no notions of individual calibration exists, which can produce results where good sources are well-calibrated only when they are uninformative (very imprecise), and badly calibrated in the cases where they are informative. This is also illustrated by Sandri *et al.* [174]

These shortcomings are not imputable to the method and mainly come from the necessity of staying within probability theory, which does not allow to use set-based representation or calculus and, as already argued, tends to mix up imprecision and variability. The problem of extending the method to cases where observations are themselves pervaded with uncertainty is considered by Kraan [131].

4.4.3 Evaluating sources in uncertainty theories

We define the *imprecision index* IG as follows:

Definition 4.5. Let $\underline{\mu}$ be a function defined on the power set $\wp(\mathcal{X})$ of the finite space \mathcal{X} and such that:

- $\underline{\mu}(\mathcal{X}) = 1$ (boundary conditions)
- $A \subseteq B \Rightarrow \underline{\mu}(A) \leq \underline{\mu}(B)$ (monotonicity)
- $\forall A, B \subseteq \mathcal{X}, A \cap B = \emptyset, \underline{\mu}(A \cup B) \geq \underline{\mu}(A) + \underline{\mu}(B)$ (super-additivity)

and let $m_{\underline{\mu}}$ be its Möbius inverse (see Remark 3.1). Then, the *imprecision index* $IG_{\underline{\mu}}$ of $\underline{\mu}$ is

defined as the value

$$IG_{\underline{\mu}} = \sum_{E \subseteq \mathcal{X}} m_{\underline{\mu}}(E) |E| \quad (4.35)$$

with $|E|$ the cardinality of E (0 when $E = \emptyset$).

Note that in the above definition, $\underline{\mu}$ corresponds to a super-additive capacity for which the boundary condition on the empty-set has been dropped (in possibility and random set theory, this corresponds to unnormalized possibility distributions and to random sets with non-null weight given to the empty set).

Given an uncertainty representation and its associated lower confidence measure, Equation (4.35) can be considered as a measure of its imprecision. When the representation is a possibility distribution or a random set, the *imprecision index* IG respectively comes down to compute the cardinality of the possibility distribution and the expected cardinality of the random set.

The case where $\underline{\mu}$ is a lower capacity modeling a general coherent lower probability is studied by Abellan and Moral [3] (they consider $\ln |E|$ instead of $|E|$ in Equation (4.35), but most of their results remain valid with $|E|$). Let \mathcal{P}_S and \mathcal{P}'_S be any two credal sets defined on a domain \mathcal{X} such that $\mathcal{P}_S \subseteq \mathcal{P}'_S$, and $\underline{P}_S, \underline{P}'_S$ the induced coherent lower probabilities. The imprecision index satisfies the following properties:

1. Monotonicity: $\mathcal{P}_S \subseteq \mathcal{P}'_S \Rightarrow IG_{\underline{P}_S} \leq IG_{\underline{P}'_S}$
2. Positivity: $IG_{\underline{P}_S} > 0$
3. Bounded: $IG_{\underline{P}_S} \in [1, |X|]$, respectively when \mathcal{P}_S reduce to a probability on \mathcal{X} and ignorance on the set \mathcal{X} .

And the same properties hold for possibility distributions and random sets, except that the lower bound of the imprecision index is 0, which happens in case of complete conflict (i.e., the whole mass is given to the empty set, and $\underline{\mu}(\emptyset) = 1$).

In the following, we consider a seed variable T taking values on \mathcal{T} (which can be, again, the discrete real line), for which a source has provided information. Let S be a source whose information given on T induces a lower capacity (see Definition 3.1) $\underline{\mu}_S$ on \mathcal{T} . We note \mathcal{P}_S the non-empty core (or, equivalently, the credal set) induced by this capacity. We also note $m_{\underline{\mu}_S}$ the mass function on \mathcal{T} given by the Möbius inverse (see Definition 3.3).

4.4.3.1 Informativeness

Let $\underline{\mu}_{Ign}$ be the lower capacity representing ignorance on \mathcal{T} , that is the capacity such that $\underline{\mu}_{Ign}(E) = 0$ for all events $E \subset \mathcal{T}$, and $\underline{\mu}_{Ign}(\mathcal{T}) = 1$. The Möbius inverse of $\underline{\mu}_{Ign}$ is noted $m_{\underline{\mu}_{Ign}}$, and reduces to $m_{\underline{\mu}_{Ign}}(\mathcal{T}) = 1$ and $m_{\underline{\mu}_{Ign}} = 0$ for other events. The value $IG_{\underline{\mu}_{Ign}} = |\mathcal{T}|$ then follows.

Since $IG_{\underline{\mu}_S}$ is a measure of the imprecision of the information, the value $IG_{\underline{\mu}_{Ign}} - IG_{\underline{\mu}_S} = |\mathcal{T}| - IG_{\underline{\mu}_S}$ measures the precision of the information given by source S . We then define the informativeness of a source S for the seed variable T as the normalized index

$$Inf_S = \frac{|\mathcal{T}| - IG_{\underline{\mu}_S}}{|\mathcal{T}|}$$

which has value 0 if the source gives no information, and is maximal ($Inf_S = 1$) if the source provides a single probability distribution (i.e., provides information with maximal precision).

Example 4.4. Consider a space $\mathcal{T} = \{t_1, t_2, t_3\}$, and two sources S_1, S_2 , the first providing an opinion in terms of confidence bounds over nested sets (i.e., possibility distributions), the second providing probability bounds on each elements (i.e., imprecise probability assignments). The respective information are summarized in the following table:

| | S_1 | | S_2 | |
|---------------|-----------------|--|-----------|---------------------------|
| Set | \underline{P} | | Set | \underline{P} \bar{P} |
| $\{t_2\}$ | 0.75 | | $\{t_1\}$ | 0.2 0.5 |
| \mathcal{T} | 1 | | $\{t_2\}$ | 0.3 0.6 |
| | | | $\{t_3\}$ | 0 0.3 |

The lower capacities $\underline{\mu}_{S_1}, \underline{\mu}_{S_2}$ induced by S_1, S_2 information are respectively computed by Equation (3.13) and Equations (3.10). The next table shows the result of applying the Möbius inverse to $\underline{\mu}_{S_1}, \underline{\mu}_{S_2}$

| | | S_1 | | S_2 | | | |
|----------------|------|----------------|------|----------------|-----|----------------|------|
| Set | m | Set | m | Set | m | Set | m |
| $\{t_1\}$ | 0 | $\{t_1, t_3\}$ | 0 | $\{t_1\}$ | 0.2 | $\{t_1, t_3\}$ | 0.2 |
| $\{t_2\}$ | 0.75 | $\{t_2, t_3\}$ | 0 | $\{t_2\}$ | 0.3 | $\{t_2, t_3\}$ | 0.2 |
| $\{t_3\}$ | 0 | \mathcal{T} | 0.25 | $\{t_3\}$ | 0 | \mathcal{T} | -0.1 |
| $\{t_1, t_2\}$ | 0 | | | $\{t_1, t_2\}$ | 0.2 | | |

and we get, respectively, $IG_{\underline{\mu}_{S_1}} = 1.5, IG_{\underline{\mu}_{S_2}} = 1.4$, which once normalized give $Inf_{\underline{\mu}_{S_1}} = 0.5, Inf_{\underline{\mu}_{S_2}} = 0.533\dots$, from which we can conclude that S_2 is slightly more informative than S_1 .

4.4.3.2 Calibration

We differentiate here the case where the value of a seed variable is precisely known with the case where it is only known with some uncertainty. Let t^* be the value of the seed variable when it is precisely known, and $\underline{\mu}_T$ be the lower capacity induced by the information we have on the true value of T (coming from an imprecise observation or measurement).

Precisely known value In this case, we propose to simply measure the calibration of the source by the upper confidence level given by the source to the value, that is

$$Cal_S = \bar{\mu}_S(t^*)$$

with $\bar{\mu}_S$ the dual upper confidence measure of $\underline{\mu}_S$. Indeed, $\bar{\mu}_S(t^*)$ measures to which degree source S judges value t^* plausible, hence it is a good measure of the coherence between the source information and the observed value. It is maximal if and only if source S judges t^* totally plausible, and only depend on this last value.

Imprecisely known value In this case, observations on T are modeled by an uncertainty model inducing a lower capacity $\underline{\mu}_{t^*}$. To measure the coherence of source S information with observations, we propose to use the equivalent of an inclusion index, that is, to measure in some way the proportion of $\underline{\mu}_{t^*}$ in $\underline{\mu}_S$. Let φ be some conjunctive fusion operator, as described in Section 4.1.3, and $\underline{\mu}_{t^* \cap S} = \varphi(\underline{\mu}_{t^*}, \underline{\mu}_S)$ be the (not forcefully coherent) lower confidence

measure resulting from the conjunction. Then, we propose to measure Cal_S , the calibration of source S , as

$$Cal_S = \frac{IG_{\underline{\mu}_{t^*} \cap S}}{IG_{\underline{\mu}_{t^*}}}$$

If $\underline{\mu}_{t^*}, \underline{\mu}_S$ are induced by two corresponding credal sets $\mathcal{P}_S, \mathcal{P}_{t^*}$ and if φ is Equation (4.4), then we retrieve the inclusion index proposed by Abellan and Gomez [1]. Similarly, if $\underline{\mu}_{t^*}, \underline{\mu}_S$ are induced by possibility distributions π_{t^*}, π_S , and if φ corresponds to Equation (4.10) with the minimum t-norm, then we retrieve an usual inclusion index between fuzzy sets (see, for instance, Smithson [190]). Finally, if $\underline{\mu}_{t^*}, \underline{\mu}_S$ reduce to intervals I_{t^*}, I_S and φ is the intersection, then we retrieve the usual inclusion index $|I_{t^*} \cap I_S|/|I_{t^*}|$ of I_{t^*} in I_S .

Note that the chosen fusion operator should satisfy at least some properties of Section 4.1.2. It seems natural to require it to be both commutative (Property III) and idempotent (Property IV), because an index of inclusion should not depend on the order of combination of $\underline{\mu}_{t^*}, \underline{\mu}_S$ and, if $\underline{\mu}_{t^*} = \underline{\mu}_S$, then the inclusion index (and, therefore, the calibration) should be maximal.

One could also use similarity indices (e.g., counterparts of Jacquard index for intervals), but the inconvenient with such indices is that they would mix up informativeness and calibration, since a source giving very imprecise but well calibrated information could obtain a very bad score.

4.4.3.3 Global scores

Let T_1, \dots, T_M be M seed variables taking their values on $\mathcal{T}_1, \dots, \mathcal{T}_M$, and for which sources S_1, \dots, S_N have provided information. Let $Inf_{S_i, T_j}, Cal_{S_i, T_j}$ respectively be the informativeness and calibration of the source i in regard with seed variable j .

For each source, we then propose to compute the global informativeness and calibration scores, Inf_{S_i} and Cal_{S_i} , as the simple arithmetic means

$$Inf_{S_i} = \frac{1}{M} \sum_{j=1}^M Inf_{S_i, T_j} \quad Cal_{S_i} = \frac{1}{M} \sum_{j=1}^M Cal_{S_i, T_j}$$

and, from these aggregated scores, to compute the global score of source i as

$$Sc_{S_i} = Inf_{S_i} Cal_{S_i}.$$

Note that other aggregation functions than the product can be chosen, but since methods are designed to minimize interaction between informativeness and calibration, the use of the product appears justified. But nothing prevents the use of other conjunctive aggregation functions, or even of sets of such functions (in this case, global scores would become imprecise, and the order between sources partial). The use of conjunctive operators ensures that no sources will have an high score by having only high informativeness scores and poor calibration scores (or the converse).

Note that the above methods can be easily extended to cases where:

- spaces $\mathcal{T}_1, \dots, \mathcal{T}_M$ are the real line. In this case, either one can consider discretized real lines or the continuous counterparts of above methods (provided these last ones are easy enough to use).
- Each source delivers information on a (different) subset of available seed variables. In this case, global scores are only computed on the basis of those seed variables for which a source has given information.

Above methods are general and simple enough to provide practical and useful tools in the assessment of sources quality. The result of such methods can then be used to weight sources in the fusion process (e.g. in a weighted mean), to modify sources information so as to take account of their reliability (i.e., discounting operations), to select subgroups of "best" sources, or to simply analyze the performances of sources and send a synthetic feedback to experts.

In Chapter 7, we apply the above methodologies to the result of uncertainty studies performed with various computer codes simulating accidental conditions of a nuclear reactor, for which data coming from an experimental small-scale facility were available.

4.5 Conclusions and perspectives

In this chapter, we have considered two problems: fusing information concerning a common variable and coming from multiple sources, as well as the problem as assessing the quality of those sources in a formal and, as much as possible, objective way.

For the first problem, we have proposed and studied basic fusion operators for clouds, linking them to basic fusion operators proposed in other uncertainty theories (which were shortly reviewed). This has allowed us to emphasize even more the interest of comonotonic clouds.

To deal with partially inconsistent information, we have considered the application of MCS-based methods. They are theoretically sound and conceptually attractive, but are in general computationally greedy. This is why we have studied a frame of application where such methods remain tractable (almost linear in the number of sources), namely the one where information is modeled by possibility distributions and MCS are applied level-wise to those distributions.

We have also briefly addressed the problem of merging information coming from potentially dependent sources, and have given first promising results that would eventually lead to a practical cautious merging rule of random sets. This approach appears interesting, since for the special case of possibility distributions, we retrieve the minimum conjunctive rule.

With respect to the above problems, perspectives include

- the implementation of the proposed MCS fusion rule, and its application to real-world problems, in order to validate its practical usefulness and meaningfulness with users. We're currently planing to integrate it to the uncertainty treatment software (SUNSET) developed at IRSN and to apply it to results of the OCDE research programme BEMUSE (see chapter 7)
- the comparison of MCS fusion rule in real-world problems with other adaptive fusion rules of possibility theory, since we already know from Section 4.2.2.3 that it competes well with these other rules from an axiomatic and theoretical standpoint.
- further research concerning the cautious conjunctive merging of random sets, including: thorough comparison of the proposed rule with other similar rules [61, 19], theoretical study to check which properties of table 4.2 are satisfied by the rule (in particular associativity), additional constraints so that the ordering of focal elements is uniquely defined.

With respect to source quality evaluation, we have proposed a general methodology inspired by previous works [28, 174] and allowing to assess source quality from previous performances, in a way as objective as possible. This is done by following some rules of common sense, initially proposed by Cooke [28].

Perspectives concerning this method are mainly its implementation and use in real-world problems. We have already partially⁷ implemented this method in the SUNSET software, and have applied it to the results of the BEMUSE programme (see chapter 7), in order to evaluate

⁷only the probabilistic and possibilistic methods have been integrated so far

the quality and responses of different computer codes on a particular accidental scenario. Overall, people to whose we have presented the result of our study recognized the need and the interest of such methods in every-day problems.

The material contained in this chapter can be found in papers [65, 72, 73]

Chapter 5

Independence and uncertainty

“With four parameters I can fit an elephant and with five I can make him wiggle his trunk”
— John von Neumann (1903-1957)

Contents

| | | |
|------------|---|------------|
| 5.1 | Finding our way in the jungle of independence: towards a taxonomy . . . | 187 |
| 5.1.1 | Judgment of (ir)relevance: a classification | 187 |
| 5.1.1.1 | Irrelevance judgments in probability theory | 189 |
| 5.1.1.2 | Irrelevance judgments in set theory | 190 |
| 5.1.2 | (Ir)relevance in imprecise probability theories: first steps towards a taxonomy | 191 |
| 5.1.2.1 | Unknown interaction | 192 |
| 5.1.2.2 | Strong and repetition independence | 193 |
| 5.1.2.3 | Epistemic irrelevance and independence | 196 |
| 5.1.2.4 | Random set independence | 198 |
| 5.1.2.5 | Possibilistic non-interaction | 199 |
| 5.1.2.6 | Irrelevance and independence: a general picture | 203 |
| 5.2 | Relating irrelevance notions to event-trees: first results | 206 |
| 5.2.1 | Event-trees | 206 |
| 5.2.2 | Probability trees | 207 |
| 5.2.3 | Imprecise probability trees | 209 |
| 5.2.4 | Forward irrelevance in event trees | 210 |
| 5.2.5 | Usefulness and meaningfulness of the result | 213 |

| | | |
|-----|--|-----|
| 5.3 | A consonant approximation of consonant and independent random sets | 215 |
| 5.4 | Conclusions and perspectives | 220 |

When assessing and representing uncertainty bearing on multiple variables, a current practice is to build and/or assess marginal representations on each single variable, and then to combine these representations into a joint uncertainty representations defined on all variables, taking account of potential relations between variables. This strategy have the advantage that marginal models are often more accessible or easier to build than a direct joint model, but it also implies that potential relationships between variables have to be taken into account.

Within classical probability theory, this is usually achieved by combining marginal probability distributions into a joint probability distribution, assuming precise dependency structures.

Notions of (in)dependence thus plays a crucial role in the construction of joint models that will subsequently be used to manipulate uncertainty in order to make decisions and make particular inferences. As in previous Chapters, using more expressive frameworks allow for more flexibility, and shifting to such frameworks when also often means that we have to take a new look at some questions, because some solutions that previously met general consensus can be extended in more than one way. There is subsequently a need to make sense of (at least some of) these generalizations. This is the case for the classical (formal) notion of stochastic independence, which is often used within classical probability theory to combine marginal distributions into a joint one.

This is why we review (Section 5.1) the main existing notions of independence in uncertainty theories, and propose a tentative taxonomy aiming at unifying and interpreting those different notions, while leaving some questions open. Again, the more general language of credal sets will be useful to compare these various notions. We also address the question of relating existing notions of independence to the event-tree independence notion recently proposed by Shafer [179, ch.5,8], since event-trees constitute an attractive framework to define and interpret notions of independence. Some first results are provided and discussed (Section 5.2).

As for uncertainty representations, not all notions of independence are directly comparable, in terms of generated joint structures (i.e. inclusion of one in the other) or of interpretations. However, similarly to uncertainty representations, one can use a notion of independence to approximate another one, possibly losing some information or expressiveness, but gaining in ease of use and in computational efficiency. This is illustrated in Section 5.3, where a notion

of independence proper to possibility theory is used to approximate another notion classically used in random set theory. Potential interests of such an approximation when propagating uncertainty are then investigated.

All along this chapter, we will note X_1, \dots, X_N the input variables, $\mathcal{X}_1, \dots, \mathcal{X}_N$ the respective spaces on which they take their values, and x_i an element of \mathcal{X}_i . For $1 \leq k \leq \ell \leq N$, we denote by $\mathcal{X}_{(k:\ell)} := \times_{i=k}^{\ell} \mathcal{X}_i$ the Cartesian product of the $k - \ell + 1$ sets $\mathcal{X}_k, \dots, \mathcal{X}_\ell$, and by $X_{(k:\ell)} := (X_k, \dots, X_\ell)$ a variable that assumes values in $\mathcal{X}_{(k:\ell)}$. Similarly, $x_{(k:\ell)} := (x_k, \dots, x_\ell) \in \mathcal{X}_{(k:\ell)}$ denotes an element of $\mathcal{X}_{(k:\ell)}$. Index $k : k$ is assimilated to k (e.g., $\mathcal{X}_{(k:k)} = \mathcal{X}_k$).

5.1 Finding our way in the jungle of independence: towards a taxonomy

The notion of independence between variables is often associated to a (qualitative) structural judgment asserting that two or more variables are not related in some way. For this judgment to be useful in uncertainty treatment, it has to be interpreted and formally translated. Here, we will adopt the name irrelevance rather than independence as a general term, as it is better fitted to describe all judgments of absence of (some) relations between variables. We will reserve the term independence to symmetric notions and to cases where no confusion is possible. As for information fusion in Chapter 4, we first introduce some general notions before seeing how they apply in the different uncertainty theories.

5.1.1 Judgment of (ir)relevance: a classification

Similarly to Walley [203], we associate (ir)relevance statements to structural judgments concerning the properties of variables. A judgment of (ir)relevance is thus always *epistemic*, in the sense that it is based on our current knowledge or observations, and could evolve with the arrival of new knowledge. To differentiate (ir)relevance statements, we classify them in different categories, built from various works concerning independence in uncertainty theories [203, Ch.9], [144], [216], [6]. Namely, an (ir)relevance statement can be:

- *Non-Informative (nInf.)* or *Informative (Inf.)*: by non-informative, we mean irrelevance statement expressing the absence of knowledge about the relations linking variables, leading to maximal imprecision in the joint representation (since this representation

must encompass any potential relation). By informative, we mean irrelevance statements expressing the knowledge that there is an absence of relations, leading to tighter joint representations.

- *Subjective (Sub.)* or *Objective (Obj.)*: an (ir)relevance statement is said to be *Subjective* if it affects beliefs about the values that variables may jointly assume. It is said *Objective* if it describes an intrinsic property of the relations between the variables or of the process generating their values.
- *Symmetric (Sym)* or *Asymmetric (Asym.)*: Let X_1 and X_2 be two variables. A statement is *Symmetric* when stating that X_1 is irrelevant to X_2 automatically implies that X_2 is irrelevant to X_1 in the same sense. Such statements often allows to build easily joint uncertainty models from (local) marginal representations and to limit considerations to such local models. They also easily extend to any number of variables. These features are often referred to as *factorization* or *decompositional* properties in the literature.

However, such statements do not allow to express asymmetric and directional notions of irrelevance, which are handled by asymmetrical irrelevance statements. A statement is asymmetrical when stating that X_1 is irrelevant to X_2 do not implies that X_2 is irrelevant to X_1 . The symmetrical counterpart is obtained when the statement is **explicitly** made in both ways (and this two-way judgment having to be justified, while it is implicitly accepted in the case of symmetric notions). Extending these statements to any number of variables usually requires more justifications and is less straightforward, and making this kind of statements does not automatically allow to only consider marginal local models. Asymmetrical statements have two main origins, that we call *causal* and *evidential*:

- causal type of asymmetric irrelevance is used when one want to express that two variables are not causally related (for example, in Bayesian networks). It can be used, for instance, to express that a particular habit is not (one of) the cause of a particular disease.
- evidential asymmetric irrelevance roughly express the idea that learning the value of one variable will not change beliefs about the values that the other variable can assume. However, such a statement does not exclude that learning the value of the latter can change our beliefs about the former.

We first review how this classification translates in the most classical uncertainty theories: probability theory and set theory.

5.1.1.1 Irrelevance judgments in probability theory

Consider two variables X_1, X_2 that respectively assume values in $\mathcal{X}_1, \mathcal{X}_2$. Let P_{X_1}, P_{X_2} be the two unique probability distributions describing the uncertainty on these variables, and $P_{X_{(1:2)}}$ the joint probability distribution. Let $w : \mathcal{X}_1 \rightarrow \mathbb{R}, z : \mathcal{X}_2 \rightarrow \mathbb{R}$ be two mappings, respectively from spaces \mathcal{X}_1 and \mathcal{X}_2 to the real line \mathbb{R} , and E_P denote the expected value with respect to P . In probability theory, all the following definitions of independence lead to the same mathematical form, that is the one corresponding to classical (symmetric) stochastic independence:

1. $\forall x_{(1:2)} \in \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(x_{(1:2)}) = P_{X_1}(x_1)P_{X_2}(x_2)$
2. $\forall A \times B \in \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(X_{(1:2)} \in A \times B) = P_{X_1}(X_1 \in A)P_{X_2}(X_2 \in B)$
3. $\forall A \times B \in \mathcal{X}_{(1:2)}, P_{X_2}(X_2 \in B) > 0, P_{X_{(1:2)}}(X_1 \in A | X_2 \in B) = P_{X_1}(X_1 \in A)$
4. $\forall A \times B \in \mathcal{X}_{(1:2)}, P_{X_1}(X_1 \in A) > 0, P_{X_{(1:2)}}(X_2 \in B | X_1 \in A) = P_{X_2}(X_2 \in B)$
5. $\forall w, z, E_{P_{X_{(1:2)}}}(w(X_1)z(X_2)) = E_{P_{X_1}}(w(X_1))E_{P_{X_2}}(z(X_2))$

with $P_{X_{(1:2)}}(X_1 \in A | X_2 \in B)$ the conditional probability of $X_1 \in A$, knowing (only) that $X_2 \in B$. Nevertheless, each of the above definitions would be classified differently with respect to the above proposed classification, as summarized in the following table:

| Independence Type | name | Inf./nInf. | Sub./Obj. | Sym/Asym. |
|-------------------|-------------------------|------------|-----------|----------------------|
| 1, 2 | Stochastic independence | Inf. | Obj. | Sym |
| 5 | Non-correlation | Inf. | Sub. | Sym |
| 3 | Epistemic irrelevance | Inf. | Sub. | Asym. _{2→1} |
| 4 | Epistemic irrelevance | Inf. | Sub. | Asym. _{1→2} |

Table 5.1: Classification of probabilistic independence types

Where $\text{Asym.}_{i \rightarrow j}$ denotes epistemic irrelevance of variable X_i to variable X_j . Although all these definitions reduce to the same formal symmetrical notion, it is useful to already make some distinctions in their interpretations, as each of these interpretations will have a different counterpart in imprecise probability settings.

That these concepts collapse into one formal definition in classical probabilities and differs in imprecise probabilistic settings is due to many reasons: classical probabilistic conditioning having more than one counterpart [90], the equivalence between expectations and probabilities

on events no longer holding [203, Ch.2], some symmetries breaking apart into asymmetric notions.

5.1.1.2 Irrelevance judgments in set theory

In set theory, the basic notion of irrelevance is the one of *non-interactivity* or of *logical independence*, which comes down to the following definition

Definition 5.1 (Logical independence). Consider two sets $\mathcal{X}_1, \mathcal{X}_2$, domains of variables X_1, X_2 . Then, X_1, X_2 are said to be *logically independent* if the possible joint values they can assume (i.e. that could be observed) on $\mathcal{X}_{(1:2)}$ are given by

$$\mathcal{X}_{(1:2)} = \mathcal{X}_1 \times \mathcal{X}_2 = \{x_{(1:2)} | x_{(1:2)} \in \mathcal{X}_{(1:2)}\}$$

It is equivalent to state that no combination of values in those two domains are forbidden. Within the proposed classification, a judgment of logical independence between X_1, X_2 is objective, non-informative and symmetric. Conversely, two variables will be logically dependent if there exist some values $x_{(1:2)}$ of the Cartesian product $\mathcal{X}_{(1:2)}$ that X_1, X_2 cannot jointly assume.

Similarly, we can define the notions of relational irrelevance and of functional irrelevance as follows:

Definition 5.2 (Relational irrelevance). Consider two sets $\mathcal{X}_1, \mathcal{X}_2$, on which variables X_1, X_2 assume their value. Then, X_1 is *relationally irrelevant* to X_2 if there exist no relation $R : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ from \mathcal{X}_1 to \mathcal{X}_2 other than the relation such that $R(x_1) = \mathcal{X}_2$ for any $x_1 \in \mathcal{X}_1$.

X_1, X_2 are said to be relationally independent when both X_1 and X_2 are relationally irrelevant of each other. Notions of relational relevance and dependence follows: for example, if X_1, X_2 are the ages of two different persons given in years, then the information $X_1 \leq X_2$ induces a relational dependency between the two variables.

Definition 5.3 (Functional Irrelevance). Consider two sets $\mathcal{X}_1, \mathcal{X}_2$, on which variables X_1, X_2 assume their value. Then, X_1 is *functionally irrelevant* to X_2 if there exist no function $f : \mathcal{X}_1 \rightarrow \mathcal{X}_2$ from \mathcal{X}_1 to \mathcal{X}_2 linking X_1, X_2 .

and X_1, X_2 are said to be functionally independent when both X_1 AND X_2 are functionally irrelevant to each other. X_1 is said to be functionally relevant to X_2 if there is a function

$f: \mathcal{X}_1 \rightarrow \mathcal{X}_2$ linking the two, and X_1, X_2 are said to be functionally dependent if they are linked by a bijective function. Contrary to logical (in)dependence, both the notions of relational (ir)relevance and functional (ir)relevance are asymmetric. We also have the following relations between these notions:

$$\text{Logical independence} \Leftrightarrow \text{Relational independence} \Rightarrow \text{Functional independence}$$

All these notions are objective and non-informative. Such kinds of dependencies are discussed by Ferson and Kreinovich [105] in the case where sets are intervals of real values.

Also note that, in the above definitions, nothing prevents sets $\mathcal{X}_1, \mathcal{X}_2$ to be sets of probabilities, and the above definitions of irrelevance (or independence) can well be applied to credal sets (it is suggested by Couso *et al.* [33]). In particular, it could be interesting to study how such dependencies could be taken into account when combining sets of probabilities (Section 4.1.3.2).

5.1.2 (Ir)relevance in imprecise probability theories: first steps towards a taxonomy

Since the question of interpreting and modeling irrelevance between variables is central in uncertainty reasoning, it has been studied by many authors, often within the bounds of a particular theory, both for possibility theory [219, 6, 41, 43], random set theory [57, 216, 107, 30] and imprecise probability theory [203, Ch.9], [33, 144]. In this work, we restrict our classification to the main notions of irrelevance existing in imprecise probability theories, and only to unconditional notions of irrelevance. Our study is also mainly formal, and only a minimal amount of information is provided about interpretations. Indeed, considering conditional irrelevance as well as potential interpretations would require a study of its own, and is out of the scope of the present work.

We thus consider two variables, X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$, and $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ the credal sets induced by uncertainty representations on each of these variables. If credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ reduce to all probabilities in $\mathbb{P}_{\mathcal{X}_1}, \mathbb{P}_{\mathcal{X}_2}$ (i.e. ignorance on $\mathcal{X}_1, \mathcal{X}_2$), then we note these credal sets $\mathcal{I}_{X_1}, \mathcal{I}_{X_2}$.

We will also relate irrelevance notions in cases where both credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ can be represented by random sets (i.e., if their induced lower probability is ∞ -montone). In this case, we will note $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ the induced random sets, and for any focal set $E \in \mathcal{F}_{X_j}$,

we will often interpret $m(E)$ as the probability that our knowledge on X_j is modeled by the credal set $\mathfrak{I}_E = \{P \in \mathbb{P}_{\mathcal{X}} | P(E) = 1\}$, that is the credal set representing ignorance on E . This corresponds to the 2nd order interpretation mentioned in Section 3.5.2, where the 2nd order level model is a precise probability, and the 1st order level model is reduced to a set. We can now define the notions of *unknown interaction*, *strong independence*, *repetition independence*, *epistemic irrelevance* and *independence*, *random set independence*, and *possibilistic non-interaction*. All along this section, we will use a common example to illustrate them all (the same as the one considered by Abellan and Klir [2], and similar to the one used by Couso *et al.* [33]). This example, given below, has the advantage of being simple enough to illustrate most notions in one sweep, while not being too simple, so that differences between notions can be emphasized.

Example 5.1. Consider two variables X, Y taking their values on (binary) spaces $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2\}$. The only information we have about these variables are that $p(x_1) \leq p(x_2)$ and $p(y_1) \leq p(y_2)$, generating two credal sets $\mathcal{P}_X, \mathcal{P}_Y$, which can be described by the following sets of extreme points:

$$\mathbf{ext}_{\mathcal{P}_X} = \{(0.5, 0.5); (0.5, 1)\} \quad \text{and} \quad \mathbf{ext}_{\mathcal{P}_Y} = \{(0.5, 0.5); (0.5, 1)\}$$

and the equivalent random sets $(m, \mathcal{F})_X, (m, \mathcal{F})_Y$ are given by

$$m_X(\{x_2\}) = 0.5, m_X(\{\mathcal{X}\}) = 0.5 \quad \text{and} \quad m_Y(\{y_2\}) = 0.5, m_Y(\{\mathcal{Y}\}) = 0.5$$

that are also possibility distributions, since focal sets are nested.

For convenience, in the subsequent examples, we will note $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, Z the associated variable and $z_{ij} = x_i \times y_j$, $i, j = 1, 2$ a generic element of \mathcal{Z} . Similarly, we denote $p_{ij} = p(x_i \times y_j)$ and $P_{\mathcal{Z}} = (p_{11}, p_{12}, p_{21}, p_{22})$ joint probabilities on \mathcal{Z}

5.1.2.1 Unknown interaction

Definition 5.4 (UI). Let $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ be two marginal credal sets describing uncertainty on two variables, X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$. Then, a judgment of *unknown interaction* between X_1, X_2 is equivalent to building the joint credal set $\mathcal{P}_{UI, X_{(1,2)}}$ such that

$$\mathcal{P}_{UI, X_{(1,2)}} = \{P_{X_{(1,2)}} \in \mathbb{P}_{\mathcal{X}_{(1,2)}} | P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2} \in \mathcal{P}_{X_2}\}$$

where P_{X_1}, P_{X_2} are respectively the marginal probabilities of $P_{X_{(1,2)}}$ on $\mathcal{X}_1, \mathcal{X}_2$

Unknown interaction [33] is equivalent to stating that we do not know the relationship between X_1 and X_2 . In other word, the resulting joint structure include all possible combination of marginal uncertainties, hence all type of possible (in)dependence between X_1, X_2 . With respect to our classification, it is a non-informative, subjective, symmetric type of independence. Note that unknown interaction has no counterpart in classical probability theory, since even if both marginal are precise, the joint representation resulting from an assumption of unknown interaction is a credal set.

When credal sets can be modeled by random sets $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$, Fetz [107] shows that *unknown interaction* comes down to considering the set $\mathcal{M}_{X(1:2)}$ of random sets such that $(m, \mathcal{F})_{X(1:2)} \in \mathcal{M}_{X(1:2)}$ if

- $A \times B \in \mathcal{F}_{X(1:2)}$ if and only if $A \in \mathcal{F}_{X_1}, B \in \mathcal{F}_{X_2}$
- $\forall A \in \mathcal{F}_{X_1}, \sum_{B \subseteq \mathcal{F}_{X_2}} m_{X(1:2)}(A \times B) = m_{X_1}(A)$
- $\forall B \in \mathcal{F}_{X_2}, \sum_{A \subseteq \mathcal{F}_{X_1}} m_{X(1:2)}(A \times B) = m_{X_2}(B)$

in this last case, *unknown interaction* is equivalent to assuming unknown interaction between 2^{nd} order model (marginal bpa), and to assume that the 1^{st} order joint model are the Cartesian products of focal sets.

Example 5.2. The joint credal set resulting from unknown interaction between X, Y and the two marginal credal sets of Example 5.1 gives the credal set $\mathcal{P}_{UI,Z}$ on \mathcal{L} that have the following extreme points:

$$\text{ext}_{\mathcal{P}_{UI,Z}} = \{(0, 0, 0, 1); (0, 0.5, 0.5, 0); (0.5, 0, 0, 0.5); (0, 0.5, 0, 0.5); (0, 0, 0.5, 0.5)\}$$

Note that the uniform distribution is inside this credal set (i.e. it corresponds to the arithmetic mean of the second and third joint probabilities)

5.1.2.2 Strong and repetition independence

Definition 5.5 (SI). Let $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ be two marginal credal sets describing uncertainty on two variables, X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$. Then, a judgment of *strong independence* between X_1, X_2 is equivalent to building the joint credal set $\mathcal{P}_{SI, X(1:2)}$ such that

$$\mathcal{P}_{SI, X(1:2)} = \{P_{X(1:2)} \in \mathbb{P}_{\mathcal{X}(1:2)} | P_{X(1:2)} = P_{X_1} \otimes P_{X_2}, P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2} \in \mathcal{P}_{X_2}\}$$

with \otimes the classical stochastic product.

Strong independence (also called type-1 independence [203, Ch.9]) is equivalent to assuming that both X_1 and X_2 are stochastically independent and are governed by some imprecisely known random processes $P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2} \in \mathcal{P}_{X_2}$. With respect to our classification, strong independence is informative, objective and symmetric.

Fetz [107] shows that, when credal sets can be modeled by random sets $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$, the joint credal set $\mathcal{P}_{SI, X(1:2)}$ induced by a judgment of strong independence can also be obtained by building the credal set

$$\mathcal{P}_{SI, X(1:2)} = \{P \in \mathbb{P}_{\mathcal{X}(1:2)} \mid P = \sum_{A \in \mathcal{F}_{X_1}} \sum_{B \in \mathcal{F}_{X_2}} m(A)m(B)P_{X_1, A} \otimes P_{X_2, B}\}$$

such that

- \otimes is the classical stochastic product.
- $\forall A \times B \subset \mathcal{X}(1:2), P_{X_1, A} \in \mathfrak{I}_A, P_{X_2, B} \in \mathfrak{I}_B$
- for a fixed $A \in \mathcal{F}_{X_1}$, choose the same $P_{X_1, A}$ for all $B \in \mathcal{F}_{X_2}$
- for a fixed $B \in \mathcal{F}_{X_2}$, choose the same $P_{X_2, B}$ for all $A \in \mathcal{F}_{X_1}$

which is equivalent, in a 2^{nd} order interpretation of random sets, to assume that X_1, X_2 are governed by two imprecisely known and stochastically independent random processes, whose supports are themselves independently known with uncertainty. This means that we assume one of the five possible probabilistic independence between the 2^{nd} order (precise) uncertainty models (see table 5.1), and that, at the 1^{st} order level, we assume the existence of precise probabilities, for which the only available information is the extent of the support (i.e., the focal sets). This is why marginal probabilities on each focal set are forced to remain the same. Within classical probability theory, strong independence corresponds to extensions of cases 1 and 2 of table 5.1.

The credal set $\mathcal{P}_{SI, X(1:2)}$ also has a simple characterization in terms of its extreme points, and we have:

$$\mathbf{ext}_{\mathcal{P}_{SI, X(1:2)}} = \{P_{X_1} \otimes P_{X_2} \mid P_{X_1} \in \mathbf{ext}_{\mathcal{P}_{X_1}}, P_{X_2} \in \mathbf{ext}_{\mathcal{P}_{X_2}}\}$$

where $\mathbf{ext}_{\mathcal{P}_{X_i}}$ is the set of extreme points of the credal set \mathcal{P}_{X_i} , and \otimes the stochastic product. This computationally attractive property, which allows to easily build $\mathcal{P}_{SI, X(1:2)}$ by fo-

cusing only on extreme points, and the fact that strong independence satisfy so-called *d-separation* [166, 35] explains that this notion has been studied by many authors and extensively used in imprecise probabilistic extensions of popular graphical models such as naive Bayesian classifiers [222]. See Cozman [37] for a recent and good review on graphical models and the use of independence in such models.

Example 5.3. The joint credal set resulting from strong independence between X, Y and the two marginal credal sets of Example 5.1 gives the credal set $\mathcal{P}_{SI,Z}$ on \mathcal{Z} that have the following extreme points:

$$\mathbf{ext}_{\mathcal{P}_{SI,Z}} = \{(0, 0, 0, 1); (0.25, 0.25, 0.25, 0.25); (0, 0, 0.5, 0.5); (0, 0.5, 0, 0.5)\}$$

and it can be checked that all these extreme points can be found back by linear combination of extreme points of $\mathbf{ext}_{\mathcal{P}_{UI,Z}}$, thus we have $\mathcal{P}_{SI,Z} \subset \mathcal{P}_{UI,Z}$

Definition 5.6 (RI). Let $\mathcal{P}_{X_1} = \mathcal{P}_{X_2} = \mathcal{P}_X$ be two identical marginal credal sets describing uncertainty on two different variables, X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$, with $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$. Then, a judgment of *repetition independence* between X_1, X_2 is equivalent to building the joint credal set $\mathcal{P}_{RI, X_{(1:2)}}$ such that

$$\mathcal{P}_{RI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid P_{X_{(1:2)}} = P_X \otimes P_X, P_X \in \mathcal{P}_X\}$$

with \otimes the classical stochastic independence.

Repetition independence (also called type-2 independence [203, Ch.9]) corresponds to the case where X_1, X_2 have the same nature and can be assumed to follow an identical but imprecisely known random process. With respect to our classification, repetition independence is informative, objective and symmetric. Note that this kind of independence is very popular in classical statistics, where it corresponds to the assumption of "independently and identically" distributed variables.

Example 5.4. The joint credal set resulting from repetition independence between X, Y and the two marginal credal sets of Example 5.1 gives the credal set $\mathcal{P}_{RI,Z}$ on \mathcal{Z} that have the following extreme points:

$$\mathbf{ext}_{\mathcal{P}_{RI,Z}} = \{(0, 0, 0, 1); (0.25, 0.25, 0.25, 0.25)\}$$

and we have $\mathcal{P}_{RI,Z} \subset \mathcal{P}_{SI,Z}$, since $\mathbf{ext}_{\mathcal{P}_{RI,Z}} \subset \mathbf{ext}_{\mathcal{P}_{SI,Z}}$

5.1.2.3 Epistemic irrelevance and independence

Definition 5.7 (EIrr). Let $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ be two marginal credal sets describing uncertainty on two variables, X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$. Then, judging that X_1 is *epistemically irrelevant* to X_2 is equivalent to building the joint credal set $\mathcal{P}_{EIrr_{1 \rightarrow 2}, X_{(1:2)}}$ such that

$$\mathcal{P}_{EIrr_{1 \rightarrow 2}, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid \forall x_{(1:2)} \in \mathcal{X}_{(1:2)}, p_{X_{(1:2)}}(x_{(1:2)}) = p_{X_1}(x_1)p_{X_2}(x_2|x_1), \\ P_{X_1} \in \mathcal{P}_{X_1}, P_{X_2}(\cdot|x_1) \in \mathcal{P}_{X_2}\}$$

with $P_{X_2}(\cdot|x_1)$ the conditional probability on \mathcal{X}_2 given x_1 .

Enforcing $P_{X_2}(\cdot|x_1) \in \mathcal{P}_{X_2}$ means that, upon learning $X_1 = x_1$, our uncertainty about X_2 remain the same (is still described by \mathcal{P}_{X_2}). Nevertheless, the particular probability assignment $p_{X_2}(x_2|x_1)$ is allowed to be different for different values $x_1 \in \mathcal{X}_1$, i.e. it is not assumed that X_2 is governed by some unique random process independent of the value assumed by X_1 . This last point is essential to the notion of epistemic irrelevance, since if $P_{X_2}(\cdot|x_1)$ is assumed to be the same for every value $x_1 \in \mathcal{X}_1$, then strong independence is retrieved (as shown by DeCooman and Miranda [50]).

With respect to our classification, epistemic irrelevance is informative and subjective. It is also asymmetric, since assessing that X_1 is *epistemically irrelevant* to X_2 will lead to a joint credal set different from the one obtained by assessing that X_2 is *epistemically irrelevant* to X_1 . Assessing that X_1 is *epistemically irrelevant* to X_2 do not imply any kind of knowledge about how the value X_2 could influence uncertainty on X_1 . The symmetric notion of *epistemic independence* is defined as follow

Definition 5.8 (EInd). Two variables, X_1, X_2 are *epistemically independent* if X_1 is judged *epistemically irrelevant* to X_2 , and X_2 *epistemically irrelevant* to X_1 . We denote $\mathcal{P}_{EInd, X_{(1:2)}}$ the corresponding joint credal set.

Within classical probability theory, the notion of epistemic irrelevance and independence are extensions of cases 3 and 4 of table 5.1. Notions of epistemic irrelevance and independence are the most natural within Walley's behavioral theory of imprecise probabilities [203], while strong independence and repetition independence are more related to a Bayesian sensitivity interpretation of credal sets (in which credal sets model a unique but imprecisely known probability). This explains that epistemic irrelevance and independence have received a lot of attention from researchers in the field of imprecise probabilities [200, 153, 38, 147].

Example 5.5. The joint credal set resulting from assessing that Y is epistemically irrelevant to X and built from the two marginal credal sets of Example 5.1 gives the credal set $\mathcal{P}_{EIrrY \rightarrow X, Z}$ on \mathcal{L} that have the following extreme points:

$$\mathbf{ext}_{\mathcal{P}_{EIrrY \rightarrow X, Z}} = \{(0, 0, 0, 1); (0.25, 0.25, 0.25, 0.25); (0, 0, 0.5, 0.5) \\ (0, 0.5, 0, 0.5); (0.25, 0.25, 0, 0.5); (0, 0.5, 0.25, 0.25)\}$$

and extreme points that are not in $\mathbf{ext}_{\mathcal{P}_{SI, Z}}$ are obtained when $p_Y(y_1|x_1) \neq p_Y(y_1|x_2)$, but are still in $\mathbf{ext}_{\mathcal{P}_Y}$. For instance, the last extreme point in $\mathbf{ext}_{\mathcal{P}_{EIrrY \rightarrow X, Z}}$ is obtained by choosing $p_X(x_1) = 0.5, p_Y(y_1|x_1) = 0.5$ and $p_Y(y_1|x_2) = 0$, and cannot be expressed as a stochastic product of extreme points of $\mathbf{ext}_{\mathcal{P}_X}, \mathbf{ext}_{\mathcal{P}_Y}$. Assessing that X is epistemically irrelevant of Y result in the joint credal set $\mathcal{P}_{EIrrX \rightarrow Y, Z}$ on \mathcal{L} that have the following extreme points:

$$\mathbf{ext}_{\mathcal{P}_{EIrrX \rightarrow Y, Z}} = \{(0, 0, 0, 1); (0.25, 0.25, 0.25, 0.25); (0, 0, 0.5, 0.5) \\ (0, 0.5, 0, 0.5); (0.25, 0, 0.25, 0.5); (0, 0.25, 0.5, 0.25)\}$$

and $\mathcal{P}_{EIrrX \rightarrow Y, Z} \neq \mathcal{P}_{EIrrY \rightarrow X, Z}$, but both includes $\mathcal{P}_{SI, Z}$.

We also have that the joint credal set $\mathcal{P}_{EIInd, Z}$ resulting from epistemic independence between X, Y is such that $\mathcal{P}_{EIInd, Z} = \mathcal{P}_{EIrrY \rightarrow X, Z} \cap \mathcal{P}_{EIrrX \rightarrow Y, Z}$. In this specific example, we have $\mathcal{P}_{SI, Z} = \mathcal{P}_{EIInd, Z}$, but in general we only have $\mathcal{P}_{SI, Z} \subseteq \mathcal{P}_{EIInd, Z}$ (see Couso *et al.* [33] for examples)

Epistemic irrelevance has a counterpart in random set theory, in which Dempster's rule of conditioning is used. Shafer [178, Ch.7.5] calls it *cognitive independence* (CI), and the notion is also briefly studied by Ben Yaghlane *et al.* [216] under the name *irrelevance*. Here, we just mention that a deeper study of this notion would involve the study of the maximal set of joint random sets and of the associated joint credal set satisfying the constraint imposed by the notion (i.e., factorization of upper probabilities). Such a study appears necessary if we wish to make sense of this notion (if possible), both in random set theory and imprecise probability theory. Also note that, contrary to the notion of epistemic irrelevance (Definition 5.7), cognitive independence in random set theory is symmetric by definition, even if it intends to express an asymmetric and evidential notion of irrelevance.

When credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ can be modeled by random sets, another work that has to be done is to define the constraints that have to be imposed on the joint bpas and on the combination of focal sets so that the resulting structure is equivalent to epistemic irrelevance of one variable to the other or of epistemic independence, as has be done by Fetz [107] for

unknown interaction and strong independence. It would not be surprising for these constraints to be close to the one used to retrieve strong independence with random sets, as epistemic irrelevance and strong independence are closely related [50].

5.1.2.4 Random set independence

Definition 5.9 (RSI). Let $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ be two marginal random sets describing uncertainty on two variables X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$, and $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ the induced credal sets. Then, *random set independence* between X_1, X_2 is equivalent to build the joint credal set

$$\mathcal{P}_{RSI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid \forall A \subseteq \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(A) \leq \sum_{\substack{(E_{X_1} \times E_{X_2}) \cap A \neq \emptyset \\ E_{X_i} \in \mathcal{F}_{X_i}}} m_{X_1}(E_{X_1})m_{X_2}(E_{X_2})\}$$

with $m_{X_i}(E_{X_i})$ the mass given to focal set E_{X_i} in $(m, \mathcal{F})_{X_i}$

Random set independence [57, 207] (called evidential independence by Shafer [178, Ch.7.4]) can be seen as the counterpart to cases 1 and 2 of table 5.1 within random set theory. It is thus totally coherent within the bounds of this theory, where it is the natural extension of stochastic independence, since it comes down to building the joint random set $(m, \mathcal{F})_{X_{(1:2)}}$ allocating products of focal set masses to the Cartesian products of focal sets. It is a subjective, informative and symmetric concept of independence.

If random sets $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ are considered as 1st order imprecise probabilistic models, it appears difficult to make sense of the joint credal set $\mathcal{P}_{RSI, X_{(1:2)}}$. In this case, random set independence can nevertheless be used as a convenient mathematical tool, since it can be proved (see Fetz [107] or Couso [30]) that the credal set $\mathcal{P}_{RSI, X_{(1:2)}}$ includes the joint credal sets obtained with assessments of epistemic irrelevance, epistemic independence and strong independence. $\mathcal{P}_{RSI, X_{(1:2)}}$ can thus be used as an instrumental guaranteed outer approximation of $\mathcal{P}_{EI, X_{(1:2)}}, \mathcal{P}_{SI, X_{(1:2)}}$.

Ben Yaghlane *et al.* [216] have also shown that the joint credal set $\mathcal{P}_{RSI, X_{(1:2)}}$ is the unique solution to the following set of constraints:

- $\forall A \times B \in \mathcal{X}_{(1:2)}, \underline{P}_{X_{(1:2)}}(A \times B) = \underline{P}_{X_1}(A)\underline{P}_{X_2}(B)$
- $\forall A \times B \in \mathcal{X}_{(1:2)}, \bar{P}_{X_{(1:2)}}(A \times B) = \bar{P}_{X_1}(A)\bar{P}_{X_2}(B)$
- $\underline{P}_{X_{(1:2)}}$ is an ∞ -monotone capacity

with $\underline{P}_{X_1}, \underline{P}_{X_2}, \underline{P}_{X(1:2)}, \overline{P}_{X_1}, \overline{P}_{X_2}, \overline{P}_{X(1:2)}$ respectively the lower and upper probabilities respectively induced by $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ and by the joint credal set $\mathcal{P}_{X(1:2)}$ satisfying the above constraints. This indicates that, if one wants to express a notion of independence and restrict its expressiveness within the bounds of random set theory (for reasons such as computational convenience), then random set independence has to be used.

Example 5.6. The joint credal set resulting from random set independence between X, Y and the two marginal credal sets of Example 5.1 gives the following joint random set $(m, \mathcal{F})_Z$:

$$\begin{aligned} m_Z(z_{22}) &= 0.25 & m_Z(z_{21} \times z_{22}) &= 0.25 \\ m_Z(\mathcal{L}) &= 0.25 & m_Z(z_{12} \times z_{22}) &= 0.25 \end{aligned}$$

and the induced credal set $\mathcal{P}_{RSI,Z}$ on \mathcal{L} have the following extreme points:

$$\text{ext}_{\mathcal{P}_{RSI,Z}} = \{ (0, 0, 0, 1); (0, 0, 0.5, 0.5); (0, 0.5, 0, 0.5); (0.25, 0, 0, 0.75); (0.25, 0.25, 0.25, 0.25) \\ (0.25, 0.25, 0, 0.5); (0.25, 0, 0.25, 0.5); (0, 0.5, 0.25, 0.25); (0, 0.25, 0.5, 0.25) \}$$

and we have that $\mathcal{P}_{RSI,Z}$ includes in joint credal sets resulting from epistemic irrelevance.

Note that even if $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ are not induced by random sets, Definition 5.9 can be generalized by a proper use of Möbius inverse (Definition 3.3) and by considering products of negative masses and positive masses. Although the interpretation of such a joint structure has still to be clarified, it could be advantageously used in practical application as guaranteed outer approximation, provided the inclusion relation with strong independence, epistemic irrelevance and independence still holds. Abellan and Klir [2] briefly study under which assumptions such an extension would coincide with strong independence, but do not elaborate further on the relationship.

Should random sets $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ be seen as 2^{nd} order imprecise probabilistic models, then interpreting random set independence is almost straightforward, as it corresponds to assume one of the five possible probabilistic independence between the 2^{nd} order (precise) uncertainty models (see table 5.1), and to take the Cartesian product as our joint uncertainty model at the 1^{st} order level.

5.1.2.5 Possibilistic non-interaction

Definition 5.10 (PI). Let π_{X_1}, π_{X_2} be two marginal possibility distributions describing uncertainty on two variables X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$. Then, *possibilistic non-interaction*

between X_1, X_2 is equivalent to build the joint possibility distribution $\pi_{PI, X_{(1:2)}}$ such that, for all $x_{(1:2)}$ in $\mathcal{X}_{(1:2)}$

$$\pi_{PI, X_{(1:2)}}(x_{(1:2)}) = \min(\pi_{X_1}(x_1), \pi_{X_2}(x_2))$$

We can then associate to $\pi_{PI, X_{(1:2)}}$ the joint credal set $\mathcal{P}_{PI, X_{(1:2)}}$ such that:

$$\mathcal{P}_{PI, X_{(1:2)}} = \{P_{X_{(1:2)}} \in \mathbb{P}_{\mathcal{X}_{(1:2)}} \mid \forall A \subseteq \mathcal{X}_{(1:2)}, P_{X_{(1:2)}}(A) \geq N_{PI, X_{(1:2)}}(A)\}$$

with $N_{PI, X_{(1:2)}}(A)$ the necessity measure of event A induced by $\pi_{PI, X_{(1:2)}}$. With respect to our classification, possibilistic non-interaction is informative, symmetric and subjective.

Possibilistic non-interaction has been first introduced in the framework of possibility theory by Zadeh [218], and the term non-interaction was used on the basis that, inside the Cartesian product of each α -cut, variables X_1, X_2 are judged non-interactive (i.e. logically independent). Fetz [107] calls this notion fuzzy set independence.

Let $(m, \mathcal{F})_{X_1}, (m, \mathcal{F})_{X_2}$ be the two random sets induced by π_{X_1}, π_{X_2} . Then, the joint random set $(m, \mathcal{F})_{PI, X_{(1:2)}}$ induced by $\pi_{PI, X_{(1:2)}}$ can be built in the following way:

- Let $\{0 = \gamma_0 < \gamma_1 < \dots < \gamma_M\}$ be the set of all distinct values taken by π_{X_1}, π_{X_2} , respectively on $\mathcal{X}_1, \mathcal{X}_2$.
- Build a joint bpa $m_{PI, X_{(1:2)}}$ such that $m(E_{\bar{\gamma}_i, \pi_{X_1}} \times E_{\bar{\gamma}_i, \pi_{X_2}}) = \gamma_{i+1} - \gamma_i$ for all $i = 0, \dots, M - 1$, with $E_{\bar{\gamma}_i, \pi_{X_j}}$ the strong γ_i -cut of distribution π_{X_j} .

which well shows that $\pi_{PI, X_{(1:2)}}$ is equivalent to assuming a complete dependence between α -cuts, and consequently between levels of confidence. Inclusion relationships between possibilistic non-interaction and random set independence notions have been studied by Tonon and Chen [192].

If we now compare the credal set $\mathcal{P}_{PI, X_{(1:2)}}$ with joint credal sets generated by other notions of irrelevance, it appears that it has poor relationships with them (it generally neither includes nor is included in any of them). This is not so surprising, since possibilistic non-interaction was first motivated within possibility theory, which does not generalize classical probabilities and is at odd with it (see Figure 3.1), thus there is no obvious reasons for possibilistic irrelevance to generalize in some way one of the probabilistic independences summarized in table 5.1. Similarly to random set independence, if π_{X_1}, π_{X_2} are interpreted as 1st imprecise probabilistic level, making sense of possibilistic irrelevance appears difficult. Nevertheless, given marginal credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ induced by π_{X_1}, π_{X_2} , the joint credal set $\mathcal{P}_{PI, X_{(1:2)}}$ is the unique solution to the following constraints:

- For any $\gamma_i \in \{\gamma_0, \dots, \gamma_{M-1}\}$, $\underline{P}_{X(1:2)}(E_{\gamma_i, \pi_{X_1}} \times E_{\gamma_i, \pi_{X_2}}) = 1 - \gamma_i$
- $\underline{P}_{X(1:2)}$ is a necessity measure.

with $\underline{P}_{X(1:2)}$ the lower probability of the joint credal set satisfying the above constraints.

If possibility distributions π_{X_1}, π_{X_2} are interpreted as 2^{nd} order imprecise probabilistic model, then the joint structure $\pi_{PI, X(1:2)}$ is equivalent to considering a complete correlation between 2^{nd} order (precise) models and to considering the Cartesian product at the 1^{st} order level.

Example 5.7. The joint credal set resulting from possibilistic non-interaction between X, Y and the two marginal credal sets of Example 5.1 gives the following joint possibility distribution π_Z :

$$\pi_Z(z_{11}) = \pi_Z(z_{12}) = \pi_Z(z_{21}) = 0.5 \quad \pi_Z(z_{22}) = 1$$

the induced random set $(m, \mathcal{F})_{\pi_Z}$ is such that

$$m_{\pi_Z}(\{z_{22}\}) = 0.5 \quad m_{\pi_Z}(\{\mathcal{Z}\}) = 0.5$$

and the induced credal set $\mathcal{P}_{PI, Z}$ on \mathcal{Z} have the following extreme points:

$$\mathbf{ext}_{\mathcal{P}_{PI, Z}} = \{(0, 0, 0, 1); (0, 0, 0.5, 0.5); (0, 0.5, 0, 0.5); (0.5, 0, 0, 0.5)\}$$

and it can be checked that $\mathcal{P}_{PI, Z}$ is neither included in nor includes other credal sets considered up to now, except for $\mathcal{P}_{UI, Z}$ (simply note that a probability assignment such that $p_Z(z_{11}) = 0.5$ cannot be reached by convex combination of extreme points of other credal sets, except for $\mathcal{P}_{UI, Z}$ and $\mathcal{P}_{PI, Z}$, and similarly, that the uniform distribution on \mathcal{Z} , which is in all other credal sets, cannot be expressed through a convex combination of elements in $\mathbf{ext}_{\mathcal{P}_{PI, Z}}$).

The following proposition shows that the notion of possibilistic non-interaction can be interpreted as a non-informative irrelevance notion, where all the information that is kept is the information on elementary elements of the Cartesian product:

Proposition 5.1. *Let π_1, π_2 be two possibility distributions, $\mathcal{P}_{\pi_1}, \mathcal{P}_{\pi_2}$ the induced credal sets and $\mathcal{P}_{UI, \pi(1:2)}$ the joint credal set induced from marginal credal sets $\mathcal{P}_{\pi_1}, \mathcal{P}_{\pi_2}$ and a judgment of Unknown Interaction. Then, we have, for any $x_{(1:2)} \in \mathcal{X}_{(1:2)}$:*

$$\bar{P}_{UI, \pi(1:2)}(\{x_{(1:2)}\}) = \bar{P}_{UI, \pi(1:2)}(x_1 \times x_2) = \min(\bar{P}_{\pi_1}(\{x_1\}), \bar{P}_{\pi_2}(\{x_2\})) = \min(\pi_1(x_1), \pi_2(x_2))$$

with $\bar{P}_{UI,\pi_{(1:2)}}, \bar{P}_{\pi_1}, \bar{P}_{\pi_2}$ the upper probabilities of $\mathcal{P}_{UI,\pi_{(1:2)}}, \mathcal{P}_{\pi_1}, \mathcal{P}_{\pi_2}$ respectively.

Proof. Consider a generic element $x_{(1:2)} \in \mathcal{X}_{(1:2)}$, and let $p'_1 \in \mathcal{P}_{\pi_1}, p'_2 \in \mathcal{P}_{\pi_2}$ be two probability distributions such that $p'_1(x_1) = \pi_1(x_1)$ and $p'_2(x_2) = \pi_2(x_2)$. Now, let $p'_{(1:2)}$ be a joint probability distributions on $\mathcal{X}_{(1:2)}$ which have p'_1, p'_2 as marginals. By using the Frechet bounds, we know that $p'_{(1:2)}(x_1 \times x_2)$ is bounded as follows

$$\max(p'_1(x_1) + p'_2(x_2) - 1, 0) \leq p'_{(1:2)}(x_1 \times x_2) \leq \min(p'_1(x_1), p'_2(x_2))$$

and that these bounds can be reached by some joint distribution $p'_{(1:2)}$. Since $p'_1(x_1), p'_2(x_2)$ are upper bounds, and since a judgment of Unknown Interaction consider every possible joint distribution built from marginals that are in $\mathcal{P}_{\pi_1}, \mathcal{P}_{\pi_2}$, we have $\bar{P}_{UI,\pi_{(1:2)}}(x_1 \times x_2) = \min(\bar{P}_{\pi_1}(\{x_1\}), \bar{P}_{\pi_2}(\{x_2\}))$, and this finishes the proof. \square

This proposition shows that taking the joint possibility distribution induced by possibilistic non-interaction is equivalent to making an unknown interaction judgment between credal sets and then keeping only the information related to the singletons of $\mathcal{X}_{(1:2)}$

To conclude with possibilistic non-interaction and possibilistic joint models in general, it appears that in a quantitative framework, one cannot fully model complex notions of (ir)relevance or independence only with the language of possibility distributions. It is also the conclusion reached by Miranda and de Cooman [147] in their study of epistemic dependence in quantitative possibility theory. Nevertheless, as they notice, possibility theory and distributions are very useful in qualitative framework, in which some important previous work studying independence in possibility theory took place [6, 75]. It is also important to recall that, by being the simplest model of imprecise probabilities, possibility distributions are also the most convenient from a computational standpoint. Hence, it is always useful to be able to provide joint possibility distributions that are guaranteed outer-approximation of a given joint uncertainty representation, these distributions then playing the role of first "quick and clean" approximations (see Section 5.3).

Our short walk in the jungle of independence notions in quantitative imprecise probability theories would not be complete without mentioning *Kuznetsov's* [?] condition of *independence* (KI), studied by Cozman [36]. This condition, based on lower and upper expectations reached by random variables, mainly impose that intervals of expectations follows the rules of interval arithmetic [152]. It is the most natural extension of case 5 in Table 5.1 inside imprecise probabilist theories, and we consider it as more related to non-correlation than to independence. Cozman [36] shows that, given marginal credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$, the most

conservative joint credal set constructible from Kuznetsov’s independence condition, that we note $\mathcal{P}_{KI, \mathcal{X}_{(1:2)}}$, includes the joint credal set built from strong independence assessment and is included in the joint credal set built from epistemic independence. Within our classification, Kuznetsov’s independence is informative, subjective and symmetric.

5.1.2.6 Irrelevance and independence: a general picture

We just have recalled the main notions of irrelevance and independence existing within imprecise probability theory. From this review, it is clear that there are many questions to be solved. Nevertheless, we can sketch some general conclusions.

First, note that epistemic notions of irrelevance (all notions of previous section) requires to first assume logical independence between spaces. This pre-requisite assumption of logical independence between spaces also applies between sets of probability distributions.

Second, independence notions are useful to demonstrate the poor expressive power of precise probabilities and classical sets. Let $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ be two credal sets on $\mathcal{X}_1, \mathcal{X}_2$, then, if they respectively reduce to precise probabilities P_{X_1}, P_{X_2} , the following notions collapse into the classical stochastic independence (as already suggested by table 5.1):

$$SI \equiv EIrr \equiv EInd \equiv RSI \equiv KI \equiv CI \text{ (Cogn. Ind.)} \equiv RI$$

with repetition independence applying only when $P_{X_1} = P_{X_2}$. Similarly, if the two credal sets reduce to $\mathfrak{I}_{X_1}, \mathfrak{I}_{X_2}$, that is credal sets equivalent to sets $\mathcal{X}_1, \mathcal{X}_2$, the following notions collapse into the Cartesian product $\mathcal{X}_{(1:2)}$:

$$SI \equiv EIrr \equiv EInd \equiv RSI \equiv KI \equiv CI \text{ (Cogn. Ind.)} \equiv UI \equiv PI.$$

Given two marginal credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$, Figure 5.1 summarizes the inclusion relationships between the different irrelevance assessments reviewed in the previous section, with notions RI, RSI, CI, PI applying when possible. This figure allows to compare how much each irrelevance notion reduces the imprecision of the resulting joint uncertainty. For instance, we can see that an assessment of epistemic independence following an assessment of strong independence would not reduce the uncertainty. However, this does not mean that an assessment of epistemic independence following one of strong independence would be useless, let alone that the latter implies the former. A simple illustration is the case of a coin whose characteristics are unknown: while it is reasonable to judge the results of each successive flips to

$$\begin{array}{ccccccc}
\mathcal{P}_{SI, X_{(1:2)}} \subseteq \mathcal{P}_{KI, X_{(1:2)}} \subseteq \mathcal{P}_{EInd, X_{(1:2)}} \subseteq \left\{ \begin{array}{l} \mathcal{P}_{EIrr, X_{1 \rightarrow 2}} \\ \mathcal{P}_{EIrr, X_{2 \rightarrow 1}} \end{array} \right\} \subseteq \mathcal{P}_{RSI, X_{(1:2)}} \subseteq \mathcal{P}_{CI, X_{(1:2)}} \subseteq \mathcal{P}_{UI, X_{(1:2)}} \\
\cup \qquad \cup \\
\mathcal{P}_{RI, X_{(1:2)}} \qquad \mathcal{P}_{PI, X_{(1:2)}}
\end{array}$$

Figure 5.1: Inclusion relationships of joint models, with marginal credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$

be stochastically independent (strong independence), it would be unreasonable to also judge them epistemically independent, since it would not allow one to learn from the successive flipping results (i.e., to know if the coin is loaded or fair). Note that there is an important subcase on which the joint uncertainty representations resulting from Strong independence, Kusnetsov independence, Epistemic irrelevance/independence and random set independence all agree: whenever $A \subseteq \mathcal{X}_1, B \subseteq \mathcal{X}_2$, these five notions produce the same lower and upper probabilities on $A \times B$, which are

$$\underline{P}_{X_{(1:2)}}(A \times B) = \underline{P}_{X_1}(A)\underline{P}_{X_2}(B) \quad \text{and} \quad \bar{P}_{X_{(1:2)}}(A \times B) = \bar{P}_{X_1}(A)\bar{P}_{X_2}(B)$$

Table 5.2 summarizes how each notions of irrelevance/independence can be classified, whether they are symmetric or not and if they are (always) expressible within a particular uncertainty theory (i.e., imprecise probability theory, random set theory, possibility theory), given our current knowledge. Also note that, if decompositional types of irrelevance can easily be extended to any Cartesian product $\mathcal{X}_{(1:N)}$ of N spaces, it is not the case for evidential and non-symmetric types of irrelevance.

Besides this formal study stands the question of interpreting and using irrelevance statements in practice. It appears that a given irrelevance notion can have multiple interpretations, can fit different theories and serve different purposes. Eventually, the choice of a particular notion should be guided by:

- Available evidence/knowledge about the variables, their relations and about the considered problem.
- The framework of application.
- The computational convenience of a particular notion.

| Irrelevance notion | Inf./nInf. | Obj./Sub. | Sym/Asym. | Expressible in | | |
|------------------------|------------|-----------|-----------|----------------|----|---|
| | | | | IP | RS | P |
| Logical Ind. | nInf. | Obj. | Sym | ✓ | ✓ | ✓ |
| Relational Irr. | nInf. | Obj. | Asym. | ✓ | ✓ | ✓ |
| Functional Irr. | nInf. | Obj. | Asym. | ✓ | ✓ | ✓ |
| Unknown Int. | nInf. | Sub. | Sym | ✓ | ✓ | × |
| Possibilistic non-Int. | Inf./nInf. | Sub. | Sym | ✓ | ✓ | ✓ |
| Cognitive ind. | Inf. | Sub. | Asym. | ? | ✓ | × |
| Random set Ind. | Inf. | Sub. | Sym | ✓ | ✓ | × |
| Epistemic Irr. | Inf. | Sub. | Asym. | ✓ | ? | × |
| Kusnetsov Ind. | Inf. | Sub. | Sym | ✓ | ? | × |
| Strong Ind. | Inf. | Obj. | Sym | ✓ | ✓ | × |
| Repetition Ind. | Inf. | Obj. | Sym | ✓ | ? | × |

Table 5.2: Irrelevance notions in uncertainty: a summary (? : matter of further research)

For example, even if the notion of random set independence can sometimes appear conservative or somewhat ad hoc, its computational convenience and the fact that it allows a simple use of sampling techniques [117, 116] are instrumental when the computational cost of using strong independence or epistemic independence is too high. The same arguments hold for unknown interaction and possibilistic non-interaction, which are equipped with a full-fledged counterpart of interval arithmetic (see, for example, Williamson and Downs [209] for probabilistic arithmetic and Kaufmann and Gupta [126] for possibilistic arithmetic), making them very efficient tools providing fast approximations.

To conclude, it is important, when making an (ir)relevance statement, to motivate this statement, e.g., by using our knowledge of how things work, by considering available observations (frequencies) or evidences, by making experiments. . . Such a motivation requires careful thinking, since as emphasized recently by Shafer [179, Ch.5.] (and, in a shorter way, by Couso [30]), witnessing formal independence in the observations (i.e., stochastic product of frequencies) does not forcefully imply independence or irrelevance in the processes generating these frequencies, similarly to the known fact that a non-null statistical correlation does not imply a true (causal) relationship between two variables.

5.2 Relating irrelevance notions to event-trees: first results

A quite interesting idea to motivate and interpret (ir)relevance statements is to consider them inside event-trees. In his book, Shafer [179] uses event-trees (first instances of which probably dates back to Huygens [120]) in order to develop a complete theory of probability and of causal conjecture. In this work, he elaborates on the idea that probabilistic thinking should be associated to a "story" or a protocol, associated to a particular observer, and describing how the world works. In particular, he links (in)dependence with Reichenbach's [172] seminal idea of common cause.

This notion of event tree is central in the later approach developed by Shafer and Vovk [180], where they develop a theory of (imprecise) probability based on a game-theoretic framework and involving lower and upper expectations. Recently, de Cooman and Hermans [47, 48] have shown that this theory can be related to Walley's [203] behavioural approach to imprecise probabilities, and they have introduced imprecise probability trees as a bridge between the two. By showing that many results can be imported from one theory to the other, they make significant progress towards the unification of the two theories. Given this relation between the two theories, it makes sense to wonder if and how (ir)relevance statements of one theory fits into the other theory, and if it can help in interpreting and understanding them. Partial answers are given in this section, where we give results showing how the recent notion of forward irrelevance [50], consisting in iterated epistemic irrelevance statements, fits into imprecise probability trees and relates to event-tree independence. Discussions about preliminary results for other irrelevance statements can be found in Appendix G.

5.2.1 Event-trees

An event tree is composed of situations linked together, and it represents what relevant events may possibly happen in what particular order in the world, according to a particular subject, i.e., an event tree shows the probability "story" viewed by this subject, and its uncertainty about what will happens. It is formally equivalent to a rooted tree in graph theory. Here, we restrict ourselves to trees with finite depth and width. The notions that we now introduce are illustrated in Figure 5.2. A *situation* is a node in the tree. The *initial situation* is the root of the tree. A *terminal situation* is a leaf of the tree; all other situations, including the initial one, are called *non-terminal*. A *path* in the tree is a sequence of situations from the initial to a terminal situation. A path *goes through* a situation s if s belongs to it. The set \mathcal{X} of all possible paths, or equivalently, of all terminal situations, is called the *sample space*, and is equivalent

to spaces we considered in earlier sections and chapters. An *event* is modeled by any set of terminal situations. Situations immediately following a non-terminal situation s are called *daughters* of s , and the set of such daughters is denoted by $D(s)$. The link between a situation s and one of its daughters t is called a *move* from s to t . If a situation s is before a situation t in the tree, we say that s *strictly precedes* t , and denote this as $s < t$; and if a situation s is before or equal to a situation t , we say that s *precedes* t , and denote this as $s \leq t$. Two situations are called *disjoint* if there is no path they both belong to. A *cut* is a set of disjoint situations, such that every path goes through exactly one situation in the cut. If each situation in a cut V (strictly) precedes some situation in another cut U , then V is said to (*strictly*) *precede* U , and we denote this as $V \leq U$ ($V < U$).

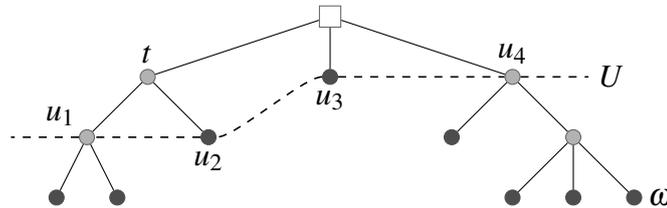
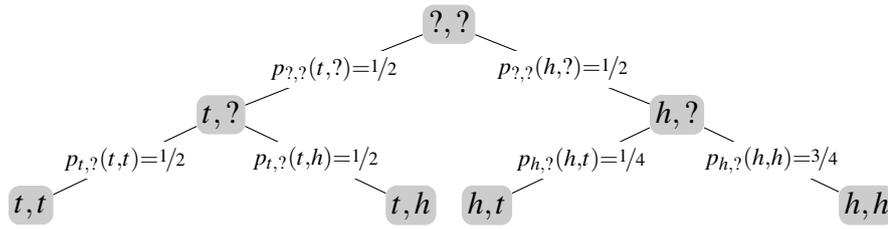


Figure 5.2: An event tree, with initial situation \square , non-terminal situations (such as t) in grey, and terminal situations (such as ω) in black. Also depicted is a cut $U = \{u_1, \dots, u_4\}$. Observe that $t < u_1$ and that $D(t) = \{u_1, u_2\}$. Also, u_4 and t are disjoint, but not u_4 and ω .

5.2.2 Probability trees

Branching probabilities p_s for a non-terminal situation s are non-negative numbers summing up to one, and each of them is attached to a different move originating in s : we denote by $p_s(t)$ the probability to go from s to its daughter t ; p_s is a probability mass assignment on the set $D(s)$. A (precise) *probability tree* is an event tree for which every non-terminal situation has such branching probabilities. $p_s(t) = p(t|s)$ is interpreted as the probability to reach t , conditional on the fact that we are in situation s , and P_s is thus a local predictive probabilistic model for what will happen right after s . Such a tree defines a joint probabilistic model on the sample space \mathcal{X} such that for any $x \in \mathcal{X}$, $p(x)$ is the product of all probabilities on the branches of the path reaching x .

Example 5.8. We illustrate the concept of probability tree with an event-tree describing two successive flipping of coins:



The labels for the situations are explicit, e.g., $h, ?$ means that the first coin has landed ‘heads’, and the second still has to be flipped. As indicated on the edges of the tree, the first flip is made with a coin judged fair. Same coin is kept if it lands ‘heads’, otherwise a biased coin is used if it lands ‘tails’. We can already note that this modeling requires to model our uncertainty with quite strong statements, that is with precise probabilities for each coin. The joint probabilistic model defined by the above tree is $p(t, t) = 1/4, p(h, t) = 1/8, p(h, h) = 3/8, p(t, h) = 1/4$

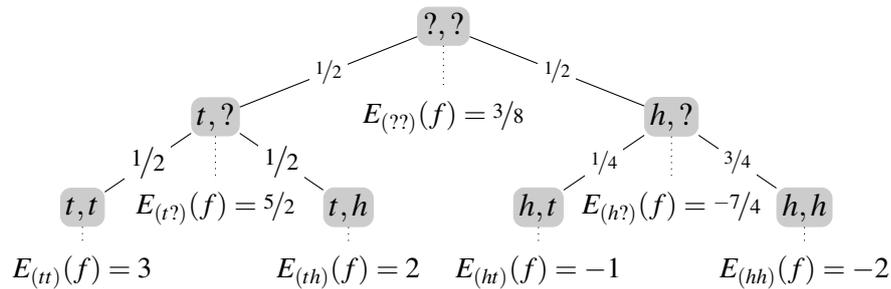
Let us now consider a non-terminal situation s and a function $f_s : D(s) \rightarrow \mathbb{R}$ affecting a value to each daughter of s . We denote by $E_s(f_s) = E(f_s|s) = \sum_{t \in D(s)} p_s(t) f_s(t)$ the expected value of f_s given that we are in situation s . Let us denote by $\mathcal{L}(D(s))$ the set of all real-valued functions on $D(s)$, then, the local probabilistic model P_s can be equivalently described by expected values $E_s(f_s)^1$, with $f_s \in \mathcal{L}(D(s))$ (simply note that $p_s(t)$ is retrieved when $f_s(t) = 1$ and $f_s = 0$ elsewhere).

If we now consider a function $f : \mathcal{X} \rightarrow \mathbb{R}$ on the sample space \mathcal{X} , then the expectation $E_s(f)$ of f in any situation s can be calculated from local models P_s by using a rule of *iterated expectation* [179, Ch.3.], also referred as *concatenation formula* by de Cooman and Hermans [47, 48]: for any situation t , we have $E_t(f) = E_t(E(f|D(t)))$, with $E(f|D(t))$ the function that assumes the value $E_s(f)$ for each $s \in D(t)$. If $x \in \mathcal{X}$ is a terminal situation, then we have $E_x(f) = f(x)$. The (conditional) probabilities of any event $A \subseteq \mathcal{X}$ can be retrieved by considering the indicator function $\mathbf{1}_{(A)}$ such that $\mathbf{1}_{(A)}(x) = 1$ if $x \in A$, zero otherwise².

Example 5.9. In the probability tree of Example 5.8, let us consider the function f such that $p(t, t) = 3, p(h, t) = -1, p(h, h) = -2, p(t, h) = 2$. Expected values obtained for the different situations are summarized below.

¹Called prevision in de Finetti language [108], who used P rather than E , E being used for so-called coherent extensions in his work

²Within probability trees, we could have limited ourselves to probabilities of events, but considering expectations is necessary with credal sets and imprecise probability trees, see Appendix A



5.2.3 Imprecise probability trees

Now, an *imprecise probability tree*³ is simply an event tree in which to each non-terminal situation s is associated a closed convex set \mathcal{P}_s of branching probabilities p_s , describing a subject’s uncertainty about which move is going to be observed just after s (i.e., immediate predictions [48]).

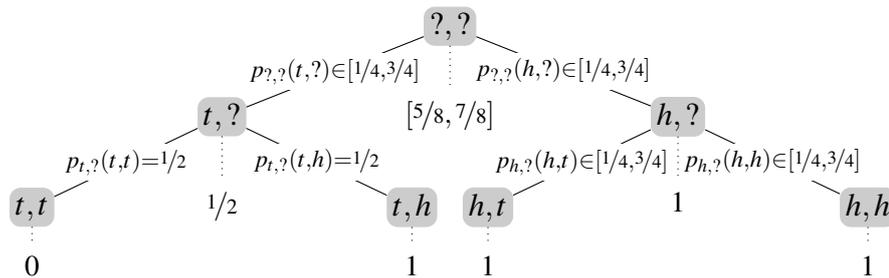
To an imprecise probability tree, we can associate coherent lower expectations. First of all, for any non-terminal situation s , and for any real-valued function h on $D(s)$, we can consider the lower expectation $\underline{E}_s(h) = \min \{E_{P_s}(h) | P_s \in \mathcal{P}_s\}$ with $E_{P_s}(h)$ the expectation of h in s given the local probabilistic model P_s . \underline{E}_s ⁴ and \mathcal{P}_s are equivalent *local* predictive models for what is going to be observed immediately after s , these models being now allowed to be imprecise (whereas in probability tree they are necessarily precise). We can also consider *global* predictive models for imprecise probability trees: Let f be a function on the sample space \mathcal{X} . For every situation t , we consider the lower expectation $\underline{E}(f|t)$ conditional on t : lower expectation of f , given that the actual path goes through t .

The global models $\underline{E}(\cdot|t)$ can be calculated from the local models \underline{E}_s by backwards recursion, using the *Concatenation Formula* [47, 48]: for any given situation t , $\underline{E}(f|t) = \underline{E}_t(\underline{E}(f|D(t)))$, where $\underline{E}(f|D(t))$ is the function on $D(t)$ that assumes the value $\underline{E}(f|s)$ in each $s \in D(t)$; and for a terminal situation $x \in \mathcal{X}$, we have $\underline{E}(f|x) = f(x)$. Lower and upper probabilities of events are retrieved as in the precise case, by considering lower and upper expectations of indicator functions on events.

Example 5.10. Let us illustrate this with the same event tree as in Example 5.8, but this time with imprecise local probabilistic models for some situations.

³Shafer [179, Ch. 12] uses the term ‘martingale tree’.

⁴In Walley’s work [203], similarly to de Finetti [108], lower expectations are called lower previsions and denoted $\underline{P}, \underline{E}$ standing for the so-called natural extension



As indicated on the edges of the tree, the knowledge about the first coin is modelled by the imprecise probability assignments $p(h) \in [1/4, 3/4]$ and $p(t) \in [1/4, 3/4]$. If it lands ‘heads’, we keep the same coin, otherwise the second flip is made with a fair coin ($p(h) = p(t) = 1/2$). Also indicated are the different steps in the calculation of the lower and upper probability of getting ‘heads’ at least once, using the Concatenation Formula.

5.2.4 Forward irrelevance in event trees

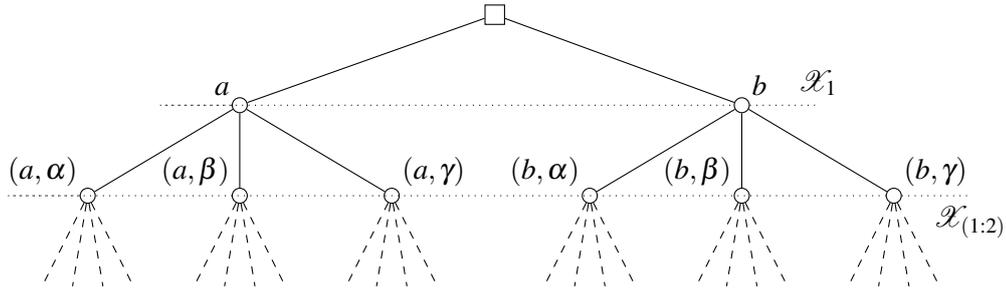
Let us briefly recall the notion of forward irrelevance, discussed in detail by De Cooman and Miranda [50], before relating it to independence in event trees. First recall that the notion of epistemic irrelevance for credal set is an asymmetric notion (see Section 5.1.2.3).

Now, assume that uncertainty bears on (random) variables X_1, \dots, X_N , respectively assuming values in $\mathcal{X}_1, \dots, \mathcal{X}_N$. We assume logical independence (Section 5.1.1.2) between all these variables, since it is a pre-requisite to express forward irrelevance. A function f defined on $\mathcal{X}_{(1:N)}$ is called $\mathcal{X}_{(k:\ell)}$ -measurable if $f(x_{(1:N)}) = f(y_{(1:N)})$ for all $x_{(1:N)}$ and $y_{(1:N)}$ in $\mathcal{X}_{(1:N)}$ such that $x_{(k:\ell)} = y_{(k:\ell)}$, that is an $\mathcal{X}_{(k:\ell)}$ -measurable function is totally determined by the values it takes on $\mathcal{X}_{(k:\ell)}$. We denote by $\mathcal{L}(\mathcal{X}_{(k:\ell)})$ the set of all $\mathcal{X}_{(k:\ell)}$ -measurable functions, and by $f_{(k:\ell)}$ a generic function in this set.

We now consider the specific example where *the X_k constitute a stochastic process with "time" variable k* , implying in particular that it is known in advance that the value of random variable X_ℓ will be revealed before that of $X_{\ell+1}$, where $\ell = 1, 2, \dots, N - 1$. Such a specific situation can be modeled by a special event tree (also called a *standard tree* [179, Ch. 2]) where the situations (nodes) s have the general form $x_{(1:k)} \in \mathcal{X}_{(1:k)}$, $k = 0, \dots, N$. For $k = 0$ there is some abuse of notation, as we let $\mathcal{X}_{(1:0)} := \{\square\}$ and $x_{(1:0)} := \square$. The sets $\mathcal{X}_{(1:k)}$ constitute special cuts of the tree, where the value of X_k is revealed (known). We have $\mathcal{X}_{(1:1)} < \mathcal{X}_{(1:2)} < \dots < \mathcal{X}_{(1:N)}$, and this sequence of cuts is also called a *standard filter* [179, Ch. 2]. It is clear that $D(x_{(1:k)}) = \{x_{(1:k)}\} \times \mathcal{X}_{k+1}$ for $k = 0, 1, \dots, N - 1$, that is the daughters of a situation in cut $\mathcal{X}_{(1:k)}$ are the values that X_{k+1} can assume. The sample space of such a tree is $\mathcal{X}_{(1:N)}$,

and with the variable X_k there corresponds a set $\mathcal{L}(\mathcal{X}_k)$ of \mathcal{X}_k -measurable functions on this sample space. For instance, in the standard tree of Example 5.9, functions characterising the second coin flip are such that $f(t, h) = f(h, h)$ and $f(t, t) = f(h, t)$. Below is another example of a standard tree.

Example 5.11. We see here the first two cuts of a standard tree, with $\mathcal{X}_1 = \{a, b\}$ and $\mathcal{X}_2 = \{\alpha, \beta, \gamma\}$.



As in the previous section, we consider that our uncertainty in each situation about what will be observed next is described by a local credal set, or equivalently by lower expectations of real-valued functions defined on the set of daughters of the situation. To any non-terminal node $x_{(1:k)}$ ($k = 0, 1, \dots, N - 1$) then corresponds a (coherent) local predictive lower expectation $\underline{E}_{x_{(1:k)}}$ defined on $\mathcal{L}(D(x_{(1:k)}))$ (in other words on $\mathcal{L}(\mathcal{X}_{k+1})$). Recall that $\underline{E}_{x_{(1:k)}}$ is equivalent to a local credal set $\mathcal{P}_{x_{(1:k)}}$. This local predictive model represents beliefs or knowledge about the value of X_{k+1} , given that the k previous variables $X_{(1:k)}$ assume the values $x_{(1:k)}$. This means that to each node is attached a credal set of conditional probabilities. For instance, in Example 5.11, to situation a would corresponds a local predictive model \underline{E}_a describing our uncertainty about which values X_2 would assume on $D(a) = \{\alpha, \beta, \gamma\}$ given that $X_1 = a$.

For *standard imprecise probability trees*, the Concatenation Formula given above for deriving the global lower previsions $\underline{E}(\cdot|x_{(1:\ell)})$ on $\mathcal{L}(\mathcal{X}_{(1:N)})$ from the local models $\underline{E}_{x_{(1:k)}}$ completely coincides with the formulae for *Marginal Extension*, derived by Miranda and De Cooman [148]. Recall that this formula allows to build, from assessments of (local) credal sets of conditional probabilities, joint uncertainty models. However, using *Marginal Extension* in general requires to assess as many credal sets as there are nodes in the tree, and this number can increase exponentially with the number of variables.

A way to reduce the number of needed assessments is to use an assessment of *forward irrelevance*, meaning that for $1 \leq k \leq N - 1$, uncertainty about the value of the ‘future’ random variable X_{k+1} won’t be changed by learning new information about the values of the ‘past’ random variables $X_{(1:k)}$: the past random variables X_1, \dots, X_k are *epistemically irrelevant* to the

future random variable X_{k+1} , for $1 \leq k \leq N - 1$. This is expressed by the following condition involving the local models: for all $0 \leq k \leq N - 1$, any function f_{k+1} in $\mathcal{L}(\mathcal{X}_{k+1})$, and all $x_{(1:k)}$ in $\mathcal{X}_{(1:k)}$:

$$\underline{E}_{x_{(1:k)}}(f_{k+1}) = \underline{E}(f_{k+1}|x_{(1:k)}) = \underline{E}_{k+1}(f_{k+1}), \quad (5.1)$$

where \underline{E}_{k+1} is the so-called *marginal* lower prevision on $\mathcal{L}(\mathcal{X}_{k+1})$, and is equivalent to specifying a marginal credal set \mathcal{P}_{k+1} , which expresses the uncertainty about the value of X_{k+1} , irrespective of the values assumed by the other random variables. Equation (5.1) indicates that our belief about the value assumed by X_{k+1} do not depend on the values $x_{(1:k)}$ observed for variables X_1, \dots, X_k . Invoking the Concatenation Formula now leads to a very specific way of combining the marginal lower expectations $\underline{E}_1, \dots, \underline{E}_N$ into a joint lower expectation, reflecting the assessment of forward irrelevance. This joint lower prevision, called the *forward irrelevant product*, is studied in detail by De Cooman and Miranda [50], who also use it to prove very general laws of large numbers [49].

We now proceed to show that *forward irrelevance* is exactly the same thing as Shafer's notion of event-tree independence, when applied to standard imprecise probability trees. In Shafer's [179] terminology, *a situation s influences a variable X* if there is at least one situation $t \in D(s)$ such that uncertainty about the value of X modifies when moving from s to t ; when we adapt this definition to imprecise probability trees this means that $\underline{E}(f|s) \neq \underline{E}(f|t)$, where f is some function whose value depends on (and only on) the outcome of X . Two variables X and Y are called *event-tree independent* if *there is no situation that influences both of them* [179, Ch.8].

In a standard imprecise probability tree, a situation $x_{(1:k)}$ influences a variable X_m if there is at least one situation $x_{(1:k+1)}$ in $D(x_{(1:k)})$ and one function $f_m \in \mathcal{L}(\mathcal{X}_m)$ such that $\underline{E}(f_m|x_{(1:k)}) \neq \underline{E}(f_m|x_{(1:k+1)})$. Note that in a standard imprecise probability tree, the only situations $x_{(1:k)}$ that can influence X_m are such that $k < m$, since in all other situations, the value of X_m has already been revealed 'for some time'. In addition, it is easy to check that X_m is always influenced by any situation $x_{(1:m-1)}$ in the cut $\mathcal{X}_{(1:m-1)}$ right before the value of X_m is revealed (i.e. the value of X_m is no longer uncertain).

Proposition 5.2. *Let X_1, \dots, X_N be N random variables. Then there is forward irrelevance, or in other words, the random variables $X_{(1:k)}$ are epistemically irrelevant to X_{k+1} for $1 \leq k \leq N - 1$ if and only if the random variables X_1, \dots, X_N are event-tree independent in the corresponding standard imprecise probability tree.*

Proof. We deal with the ‘only if’ part first. Suppose the random variables $X_{(1:N)}$ are forward irrelevant. Consider any X_k and function $f_k \in \mathcal{L}(\mathcal{X}_k)$, where $1 \leq k \leq N$. Then it follows from the forward irrelevance condition (5.1) and the Concatenation Formula that $\underline{E}_k(f_k) = \underline{E}_{x_{(1:k-1)}}(f_k) = \underline{E}(f_k|x_{(1:k-1)})$ for all $x_{(1:k-1)}$ in $\mathcal{X}_{(1:k-1)}$. Applying the Concatenation Formula again leads to $\underline{E}(f_k|x_{(1:k-2)}) = \underline{E}_{x_{(1:k-2)}}(\underline{E}(f_k|x_{(1:k-2)}, \cdot)) = \underline{E}_{x_{(1:k-2)}}(\underline{E}_k(f_k)) = \underline{E}_k(f_k)$, and if we continue the backwards recursion, we see that

$$\underline{E}_k(f_k) = \underline{E}(f_k|x_{(1:k-1)}) = \underline{E}(f_k|x_{(1:k-2)}) = \cdots = \underline{E}(f_k|x_{(1:2)}) = \underline{E}(f_k|x_1) = \underline{E}(f_k|\square).$$

This implies that the only situations that (may) influence X_k are the ones in the cut $\mathcal{X}_{(1:k-1)}$ immediately before X_k is revealed. Therefore, no situation can influence more than one variable, and there is event-tree independence.

Next, we turn to the ‘if’ part. Assume that all variables are event-tree independent in the standard tree. This implies that no variable X_k can be influenced by a situation $x_{(1:\ell)}$ corresponding to a time $\ell < k - 1$ (If X_k were influenced by such a situation, then we know that this situation also always influences $X_{\ell+1}$, and $\ell + 1 < k$, thus we end up with a contradiction). So for all $x_{(1:k-1)} \in \mathcal{X}_{(1:k-1)}$ and all $f_k \in \mathcal{L}(\mathcal{X}_k)$:

$$\underline{E}(f_k|x_{(1:k-1)}) = \underline{E}(f_k|x_{(1:k-2)}) = \cdots = \underline{E}(f_k|x_{(1:2)}) = \underline{E}(f_k|x_1) = \underline{E}(f_k|\square).$$

Now of course $\underline{E}(f_k|\square) = \underline{E}(f_k) = \underline{E}_k(f_k)$, where \underline{E}_k is the marginal lower expectation for X_k , and it follows from the Concatenation Formula that $\underline{E}(f_k|x_{(1:k-1)}) = \underline{E}_{x_{(1:k-1)}}(f_k)$. This shows that (5.1) is satisfied, so indeed there is forward irrelevance. \square

5.2.5 Usefulness and meaningfulness of the result

The above result is a first step towards a unification of independence notions used in Walley’s [203] behavioral theory with independence notion in event-trees, these last ones being central in the recent theory developed by Shafer and Vovk [180]. It is of course desirable to extend it to other structural judgments [203, Ch.9] about variables and to more general situations. Some preliminary ideas concerning other independence notions (such as epistemic independence) can be found in Appendix G. Also, there is quite a number of interesting things (both practically and theoretically) to say about this result, despite its somewhat preliminary nature.

First, assume we want to consider a theory of uncertain (random) processes, where prob-

abilities are no longer necessarily precise. Since the concept of independence in classical random processes has many counterparts when allowing imprecision, it is natural to wonder what is the most useful and meaningful of them? There are a number of reasons to prefer the asymmetric notion of epistemic irrelevance, and its generalization to multiple variables (forward irrelevance), rather than the symmetric notion of epistemic independence. Some of these reasons, that we find compelling, are the following:

- When a notion that is (more or less, as shows table 5.1) automatically symmetrical, breaks apart into two asymmetrical counterparts when using a more powerful language, symmetry becomes something that has to be justified: it can't be imposed without giving it another thought.
- An assessment of epistemic independence is stronger, and leads to higher values of lower expectations and to a smaller joint uncertainty. This means that epistemic independence leads to make stronger commitments about what could happen, and these may be unwarranted when it is only epistemic irrelevance that one want to model.
- Joint credal sets, lower probabilities and expectations based on an epistemic irrelevance assessment are generally speaking straightforward to calculate, as the discussion of the Concatenation Formula in Section 5.2.4 testifies (See also other related works [47, 49]). But calculating joint lower previsions from marginals based on an epistemic independence assessment is quite often a very complicated matter [203, Ch. 9.3.2].
- Finally, in a random process, it is known that the value of X_k *will be available before* the value of X_{k+1} . Stating that X_k is epistemically independent of X_{k+1} amounts to judging that (i) getting to know the value of X_k won't change his beliefs about X_{k+1} [forward irrelevance]; and (ii) getting to know the value of X_{k+1} won't change his beliefs about X_k [backward irrelevance]. Since we always know X_k before X_{k+1} (ii) is either counterfactual (since there is no longer any kind of uncertainty concerning the value of X_k when learning the value of X_{k+1}) or useless. In this case, we think it is much more natural in such situations context to let go of (ii) and therefore to resort to epistemic (forward) irrelevance.

Note that similar arguments hold for strong independence and repetition independence, which are also symmetric and induce lower expectations higher than epistemic independence.

Second, this result, together with the concatenation formula, make the use of epistemic independence in practical applications easier, by allowing for local computations. It could even

be a step towards the use of such independence notions in credal networks, since, as argued by Shafer [179, Ch.16], Bayes nets can be seen as economical representations of probability trees.

Third, results given by de Cooman and Miranda [50] relating forward irrelevance to strong and repetition independence allow to also express these two notions in standard imprecise probability trees, by adding suitable constraints on the local credal sets in each situation. It has to be noted that computing lower and upper expectations would then become more difficult, since local computations could no longer be used (see Appendix G).

5.3 A consonant approximation of consonant and independent random sets

Section 5.1 and Figure 5.1 indicate that it can be difficult to compare some notions of irrelevance and to interpret all of them in a single framework. However, in some cases, it can be useful to be able to approximate one notion by the other (e.g., for computational convenience, or because one wants to work within a particular framework or theory). In particular, it is always useful to be able to approximate a given joint uncertainty model by a joint possibility distribution, for the reason that possibility distributions are the simplest imprecise probabilistic models, and are thus easier to manipulate. Here, we consider the case where marginal uncertainty is described by possibility distributions, and where we assume random set independence between them.

Let π_i be a possibility distribution describing uncertainty on variable X_i , $i = 1, \dots, N$, and denote the equivalent consonant random set by $(m, \mathcal{F})_{\pi_i}$. Let $\alpha_1 = 1 > \alpha_2 > \dots > \alpha_M > \alpha_{M+1} = 0$ be the collection of distinct values taken by distributions π_i , $i = 1, \dots, N$ (or, in the case of continuous distributions on the real line, the collection of chosen discretization levels). Then, $(m, \mathcal{F})_{\pi_i}$ is given, for $j = 1, \dots, M$, by

$$\left\{ \begin{array}{l} E_{i,j} = A_{i,\alpha_j} \\ m(E_{i,j}) = \alpha_j - \alpha_{j+1} = m_{i,j} \end{array} \right.$$

with $E_{i,j}$ the α_j -cut of π_i . Each marginal random set thus have M focal elements. The joint random set $(m, \mathcal{F})_{RSI, X_{(1:N)}}$ resulting from an assessment of random set independence then has M^N focal sets, an exponentially growing number that can quickly become intractable in prac-

tice. As shown by the following proposition, it is possible to outer approximate $(m, \mathcal{F})_{RSI, X_{(1:N)}}$ by a joint possibility distribution $\pi'_{X_{(1:N)}}$ which only have M focal sets, that is a number independent of the input space dimensions.

Proposition 5.3. *The most specific non-interactive possibility distribution $\pi'_{X_{(1:N)}}$ inducing a random set $(m, \mathcal{F})_{\pi'_{X_{(1:N)}}$ outer approximating $(m, \mathcal{F})_{RSI, X_{(1:N)}}$ (in the sense of s -inclusion) and whose focal sets are in $\{\times_{i=1}^N E_{i,j} | j = 1, \dots, M\}$ is such that, for any $x_{(1:N)} \in \mathcal{X}_{(1:N)}$,*

$$\pi'_{X_{(1:N)}}(x_{(1:N)}) = \min_{i=1, \dots, N} \{(-1)^{N+1}(\pi_i(x_i) - 1)^N + 1\} \quad (5.2)$$

Proof. See appendix D □

This proposition extends to the general case a result given by Dubois and Prade [89] for the 2-dimensional case. It shows that if one transforms each distribution π_i into $\pi'_i = (-1)^{N+1}(\pi_i - 1)^N + 1$ and then builds a joint model with an assumption of possibilistic non-interaction, then the result is a guaranteed outer approximation of $(m, \mathcal{F})_{RSI, X_{(1:N)}}$. This has the practically important advantage to go from exponential to constant complexity in the number of input dimensions. Of course, such a drastic reduction is not without cost, and for a particular distribution π_i , Equation (5.2) will converge to 1 if $\pi_i(x_i) > 0$ as N increases, and is 0 if $\pi(x_i)_i = 0$. This means that, as N increases, the outer-approximation converges towards the Cartesian product of supports of distributions π_i . It is thus legitimate to wonder about (i) the speed of convergence of non-null values to 1 and (ii) the usefulness of the proposed outer-approximation when compared to other quick and cheap methods providing outer approximations, such as the use of probabilistic arithmetic [209] when propagating uncertainty?

Figures 5.3 and 5.4 give ideas about the rate of convergence, by drawing the evolution of possibility degree values versus the number of dimensions (Figure 5.3), and by sketching the evolution of a triangular possibility distribution on the real line, with center 0 and support $[-1, 1]$ (Figure 5.4). We can see that, if the loss of information is important (and thus the approximation likely to be gross), part of this information remain, even for high dimensions.

Let us now investigate if this outer approximation can be useful in some ways. In particular, we compare the propagation of the proposed outer approximation $\pi'_{X_{(1:N)}}$ with the propagation of tightest classical p-boxes $[F, \bar{F}]_i$ outer-approximating π_i using probabilistic arithmetic. Recall that probabilistic arithmetic [209] allows to apply the four basic operations $\{+, -, \times, \div\}$ with an assumption of unknown interaction to p-boxes in a very efficient way. Given two real-valued variables X, Y and some (classical) p-boxes $[F, \bar{F}]_X, [F, \bar{F}]_Y$ describing

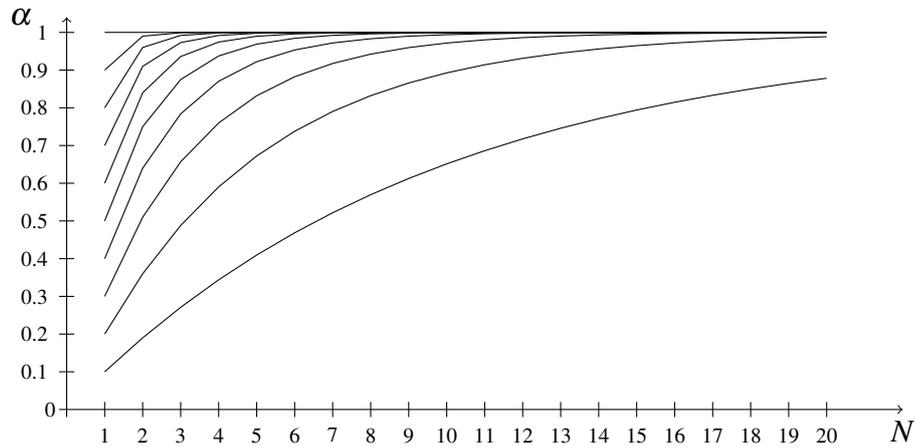


Figure 5.3: Evolution of distributions degree (α) versus input space dimension (N)

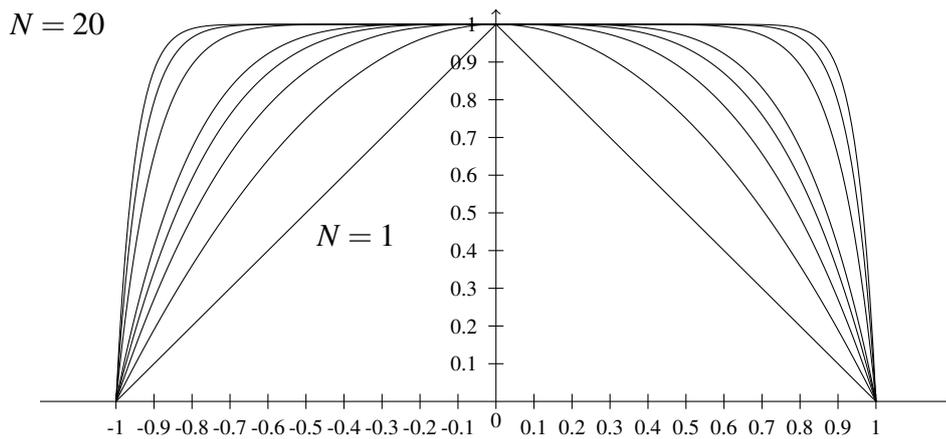


Figure 5.4: Evolution of a triangular possibility distribution for different input space dimensions (1,2,3,4,5,10,15,20)

our uncertainty on them, result of applying each arithmetic operations read, for any $z \in \mathbb{R}$:

$$\underline{E}_{X+Y}(z) = \sup_{\substack{x,y \in \mathbb{R} \\ x+y=z}} \{\max(\underline{E}_X(x) + \underline{E}_Y(y) - 1, 0)\}$$

$$\overline{F}_{X+Y}(z) = \inf_{\substack{x,y \in \mathbb{R} \\ x+y=z}} \{\min(\overline{F}_X(x) + \overline{F}_Y(y), 1)\}$$

$$\underline{E}_{X-Y}(z) = \sup_{\substack{x,y \in \mathbb{R} \\ x+y=z}} \{\max(\underline{E}_X(x) + \overline{F}_Y(-y), 0)\}$$

$$\overline{F}_{X-Y}(z) = \inf_{\substack{x,y \in \mathbb{R} \\ x+y=z}} \{\min(\overline{F}_X(x) + 1 - \underline{E}_Y(-y), 1)\}$$

$$\underline{E}_{X \times Y}(z) = \sup_{\substack{x,y \in \mathbb{R} \\ x \times y=z}} \{\max(\underline{E}_X(x) + \underline{E}_Y(y) - 1, 0)\}$$

$$\overline{F}_{X \times Y}(z) = \inf_{\substack{x,y \in \mathbb{R} \\ x \times y=z}} \{\min(\overline{F}_X(x) + \overline{F}_Y(y), 1)\}$$

$$\underline{E}_{X \div Y}(z) = \sup_{\substack{x,y \in \mathbb{R} \\ x \times y=z}} \{\max(\underline{E}_X(x) + \overline{F}_Y(1/y), 0)\}$$

$$\overline{F}_{X \div Y}(z) = \inf_{\substack{x,y \in \mathbb{R} \\ x \times y=z}} \{\min(\overline{F}_X(x) + 1 - \underline{E}_Y(1/y), 1)\}$$

and, provided the model T through which uncertainty has to be propagated is expressible by a combination of arithmetic operations, above equations can be applied for each such operations.

A comparison of both approaches is given by Example 5.12. It can be seen on this example that when the model T contains no repeated variable (i.e. can be reduced to an analytical form where each variable appears once), then both methods provide comparable results. However, it is not the case when there are repeated variables, and/or when extrema are not forcefully on boundaries of Cartesian product (non-isotone models), and when propagation of the focal sets of the outer approximation of Proposition 5.3 is done exactly (and not by using fuzzy arithmetic). Also note that, when T is not expressible analytically, propagating $\pi'_{X(1:N)}$ can be used and still requires to propagate exactly M focal sets, while in this latter case the counterpart of probabilistic arithmetic requires to solve linear systems, and present an increased computational complexity.

Example 5.12. In this example, we consider two simple models, with variables that are assumed to be random set independent and whose uncertainty is described by possibility distributions. For the two of them, we compare the result of applying extension principle (i.e. exact

propagation of the focal sets) to the outer approximation of Proposition 5.3 with the result that would give the application of probabilistic arithmetic.

Let us first consider the simple model $Y = A + B - C$, with variables A, B, C positive real-valued variables represented by the same possibility distribution, summarized in the following table (together with the result of Transformation (5.2)):

| π_A, π_B, π_C | $\Rightarrow_{(5.2)}$ | π'_A, π'_B, π'_C |
|-----------------------|-----------------------|--------------------------|
| Masses (m) | Focal Sets | Transf. masses (m') |
| 0.1 | [1, 2] | 0.01 |
| 0.7 | [0.5, 3] | 0.511 |
| 0.2 | [0.1, 5] | 0.488 |

and Figure 5.5 show the two p-boxes resulting from probabilistic arithmetic applied to p-boxes derived from original distributions π and from extension principle applied to transformed distributions π' . Although neither of them is contained in the other, there is no great differences between the two, and since both outer-approximate the exact propagation and requires comparable computational effort, they can be used conjointly, as their conjunction provide a tighter bounds of the exact result. Note that, with such simple models, exact propagation is often feasible.

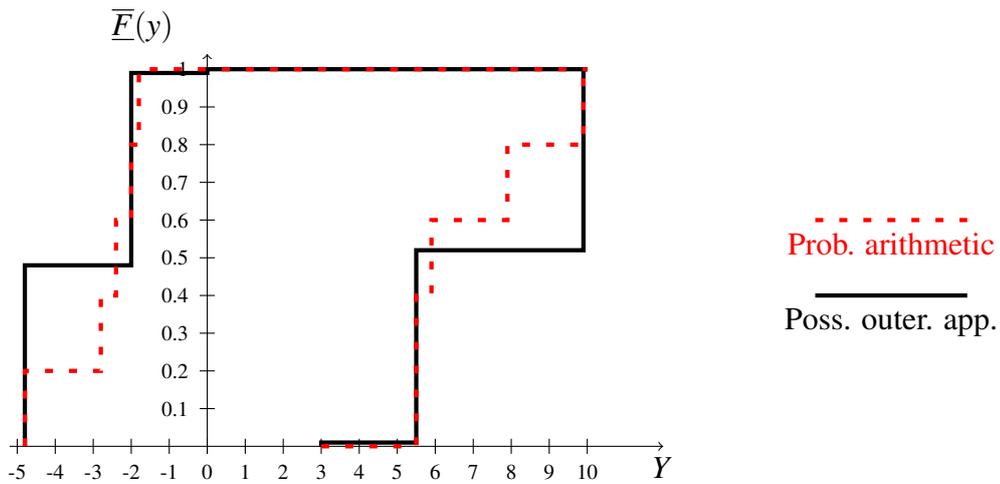


Figure 5.5: Comparison of probabilistic arithmetic and outer approximation of Proposition 5.3

Cases where the proposed outer approximation is likely to be more useful is those where (i) repeated variables occur and/or (ii) extrema are not reached on bounds of Cartesian products, but remains relatively easy to locate. To exemplify this, let us consider the following model: $Y = T(X_{(1:2)}) = (X_1^2 + X_2^2) / (2X_1 + 1)(X_2^3 - 1.9)$ with X_1, X_2 assuming values on the real line \mathbb{R} . The table below summarizes both marginal Models on X_1, X_2 , the joint possibility distribution resulting from the Transformation (5.2) and the result after (exact) propagation of this joint distribution through T .

| π_{X_1} | | π_{X_2} | | $\pi'_{X_{(1:2)}}$ | | π'_Y |
|-------------|---------------------|-------------|------------------------------|-----------------------------|----------------------------|----------------------------------|
| m_{X_1} | \mathcal{F}_{X_1} | m_{X_2} | \mathcal{F}_{X_2} | $m'_{X_{(1:2)}}$ | $\mathcal{F}'_{X_{(1:2)}}$ | \mathcal{F}'_Y |
| | | | | $[1, 2] \times [2, 3]$ | 0.01 | $[0.1036, 0.2732]$ |
| 0.1 | $[1, 2]$ | 0.5 | $[2, 3]$ | $[0.5, 3] \times [2, 3]$ | 0.24 | T $[0.1013, 0.3484]$ |
| 0.7 | $[0.5, 3]$ | 0.4 | $[2, 5] \Rightarrow_{(5.2)}$ | $[0.5, 3] \times [2, 5]$ | 0.39 | \Rightarrow $[0.0395, 0.3484]$ |
| 0.2 | $[0.1, 5.1]$ | 0.1 | $[2, 10]$ | $[0.1, 5.1] \times [2, 5]$ | 0.17 | $[0.0368, 0.5478]$ |
| | | | | $[0.1, 5.1] \times [2, 10]$ | 0.19 | $[0.0113, 0.5478]$ |

We see that the result is a distribution whose support is the interval $[0.0113, 0.5478]$, while an application of probabilistic arithmetic would provide the interval $[0.007, 2.7868]$ as the core of the resulting p-box⁵, and $[0.0003, 17.08]$ as its support. It shows that in situations where probabilistic arithmetic performs poorly, the proposed outer approximation can provide a much better result, and can therefore be of real usefulness.

5.4 Conclusions and perspectives

In this chapter, we have studied irrelevance notions in uncertainty theories. The notion of irrelevance or independence is important in many aspects of uncertainty treatment, both theoretically and practically. As we have seen in our review, it is also a very complex notion, since even for unconditional irrelevance, there are many possible extensions and interpretations of the classical notions of logical independence for sets and stochastic independence for

⁵Here, the core is the interval corresponding to the set of all possible dirac measures inside the p-box, that is, given a p-box $[F, \bar{F}]$, it is the interval $[x, y]$ such that $x = \inf_{r \in \mathbb{R}} \{\bar{F}(r) | \bar{F}(r) = 1\}$ and $y = \sup_{r \in \mathbb{R}} \{F(r) | F(r) = 0\}$

probability distributions. And, of course, there are even more when considering conditional independence. Making a whole and unified picture of all these notions and interpreting all of them inside a common framework appears difficult. Nevertheless, links and relations do exist, and many of them remain to be clarified.

Often, in probability theory, different interpretations resulted in the same formal joint representations, thus making the question of interpretation appearing as less essential (at least in practical applications). It is no longer the case here, and the question of interpretation becomes essential, even for practical applications where two different interpretations will give two different results. Moreover, it can be difficult to decide what notion of independence is the most fitted to a particular situation, and we are only able to give some guidelines about the choice of a particular notion.

A very interesting theoretical frame to study independence notions is the one of event-trees. Our first results demonstrate that there are close links between the notion of independence regarded as the most meaningful by Walley [203] and the notion of event-tree independence developed by Shafer [179].

We have also shown how some notions of irrelevance can be approximated by other ones, by concentrating on the specific question of outer-approximating random set independence by possibilistic non-interaction. This allows to lower the complexity of joint uncertainty structures and to facilitate their subsequent manipulations. The proposed outer approximation appears particularly interesting when few variables are uncertain and when the model is complex enough (i.e. non-monotonic, presence of repeated variable in the analytical formula). In any cases, it provides a way to compute a quick and rough guaranteed outer approximation.

Our perspectives regarding the study of independence are mainly theoretical, eventually resulting in results of practical importance. They include:

- Answering question marks in Table 5.2, either positively by providing a suitable interpretation, as well as a formal way to express it, or negatively by showing that a given notion makes poor sense in a given theory.
- Pursuing our study relating structural judgments of imprecise probability theory to event trees. In particular, the notions of permutability and of epistemic independence appears interesting to study (see Appendix G for some first ideas).

The material contained in this chapter can be found in paper [63]

Chapter 6

Decision Making

“Real stupidity beats artificial intelligence, every time”

— Terry Pratchett (1948–?)

Uncertainty treatment and, more generally, plausible reasoning are rather impersonal and theoretical processes. Although they can involve some choices (of a particular theory, of a specific fusion rule, . . .), these choices should be guided primarily by rationality requirements, and not by personal preferences.

In this chapter, we are concerned with a far less impersonal matter: decision making. Let a be an action that we can apply to a situation (e.g., going out for a picnic, buying a car, taking one particular direction, . . .), freely chosen from a set A of feasible actions. The problem of decision making consists in choosing, within A , an *optimal* or best course of action with respect to some criterion, given our current knowledge of the situation.

Although the problem of decision making is not related, *per se*, to the problem of uncertainty treatment, the two problematics are closely related: sooner or later, available information is used to make a decision and select a course of action. Roughly speaking, we see decision making as the step where we stop to manipulate information in order to make a decision (In this sense, we’re close in spirit to the TBM model proposed by Smets [189], which differentiates credal and pignistic levels).

So, even if this work is not devoted to decision making (itself a wide and vivid area of research), it is useful to study practical problems related to it, because of the close link between decision making, uncertainty treatment and risk analysis. Here, we will work within a

restricted frame, since we assume:

- i that the result of an action a depends of the value assumed by a variable X on \mathcal{X} , which is only known with uncertainty.
- ii that to any action a can be associated a real-valued and precise gain (or utility) $u_a : \mathcal{X} \rightarrow \mathbb{R}$ that is a mapping from \mathcal{X} to \mathbb{R} , and $u_a(x)$ reflects the interest of choosing a when X assume value $x \in \mathcal{X}$.
- iii that only crisp actions can be chosen, that is we do not consider randomized actions (i.e., convex mixtures of actions)
- iv that we are in a static environment, that is we do not consider dynamical problems involving the choice of sequences of actions.
- v that the choice of an action a does not modify uncertainty on X , i.e., we assume so-called *act-state independence*.

If there is no uncertainty about the (single) value assumed by X , then the set of optimal actions is simply given by

$$opt(A) := \arg \max_{a \in A} (u_a)$$

However, the value of X is often only known with some uncertainty, and in this case, choosing an optimal action, even in our restricted framework, is more difficult.

6.1 Decision making in uncertainty theories

Roughly speaking, defining optimal actions is equivalent to inducing some preferences between these actions, that is, an action a_1 is in the set of optimal actions if there is no other action a_2 preferred to a_1 . This is equivalent to define a partial pre-order relation \geq between the actions, and to say that a_1 is preferred to a_2 if and only if $a_1 \geq a_2$. However, there are many ways to define this partial pre-order in uncertainty. Here, we restrict ourselves with a short review, and refer to given references (and to references therein) for ampler discussions. It is sensible to first remove those actions whose gains are point-wise dominated by other actions, since whatever the value assumed by X , those actions will give less utility than the one(s) they're dominated by. Following Troffaes [195], we denote $opt_{\geq} A$ the set of actions such that

$$opt_{\geq}(A) := \{a \in A \mid \nexists c \in A, u_c \geq u_a\} \quad (6.1)$$

6.1.1 Classical expected utility

Provided uncertainty can be modeled by precise probabilities and the use of (linear) utility scale is accepted, optimal decisions are often chosen accordingly to principle of *expected utility*. While first suggestions to use expected utility in games can be traced back to Huygens, its every-day use in statistics and decision theory is mainly due to the works of Von Neumann and Morgenstern [201] and of Savage [175], who justified the use and uniqueness of expected utility as a means of selecting optimal decisions with very different assumptions and sets of axioms. Let P_X be the probability distribution modeling uncertainty of X , and $E_{P_X}(u_a)$ the expectation of u_a with respect to P_X . Then, the set of optimal decisions in A is defined as

$$opt_{P_X}(A) := \arg \max_{a \in opt_{\geq}(A)} E_{P_X}(u_a) \quad (6.2)$$

However, when uncertainty on X cannot be properly modeled by a unique probability distribution, the use of expected utility to choose an optimal action is usually not satisfactory, as shows the following example

Example 6.1. We use the example given by Troffaes [195]: let X be the outcome of a coin toss ($\mathcal{X} = \{h, t\}$). All that can be said about the probability of getting heads is that it lays between 0.28 and 0.7. Let us consider the following set of actions $A = \{a_1, \dots, a_6\}$ and the associated utilities summarized below

| <i>Utility</i> | <i>Heads</i> | <i>Tails</i> |
|----------------|--------------|--------------|
| u_{a_1} | 4 | 0 |
| u_{a_2} | 0 | 4 |
| u_{a_3} | 3 | 2 |
| u_{a_4} | 1/2 | 3 |
| u_{a_5} | 47/20 | 47/20 |
| u_{a_6} | 41/10 | −3/10 |

and as none of these actions is dominated by another one, $opt_{\geq}(A) = A$, and we have

$$opt_{P_X} = \begin{cases} \{a_2\} & \text{if } p_X(h) < 2/5 \\ \{a_2, a_3\} & \text{if } p_X(h) = 2/5 \\ \{a_3\} & \text{if } 2/5 < p_X(h) < 2/3 \\ \{a_1 a_3\} & \text{if } p_X(h) = 2/3 \\ \{a_1\} & \text{if } p_X(h) > 2/3 \end{cases}$$

which shows that using maximized expected utility (by choosing a precise probability fitting available information) is not very robust and can lead to different conclusions. Indeed, there is no obvious reason to restrict the potential choice to one optimal action, until we are forced to act. In other words, we do not have to enforce the (pre)-order on possible actions to be complete.

Note that there is no problem with (6.2) when information is sufficient to describe uncertainty on X by a single probability P_X . This is why extensions of (6.2) to uncertainty theories generalizing probability theory (at least formally) should satisfy (6.2) when reduced to precise probabilities. All extensions considered in the next sections satisfy this condition.

6.1.2 Decision making in imprecise probability theory

Recently, Troffaes [195] has provided a nice theoretical and short review of (most of) the existing extensions of (6.2) in imprecise probability theory. Another short and good review, more computationally oriented, is provided by Utkin and Augustin [198].

We now consider that uncertainty on X is modeled by a credal set \mathcal{P}_X instead of a single probability. We denote $\underline{E}_{\mathcal{P}_X}(u_a) = \min_{P \in \mathcal{P}_X}(E_P(u_a))$ the lower expectation of u_a given \mathcal{P}_X , and $\overline{E}_{\mathcal{P}_X}(u_a)$ the upper expectation defined likewise (replacing min by max). Equation (6.2) can be extended in two main ways: either by relaxing the completeness of the order between actions, meaning that we end up with a set of optimal and incomparable actions, or by adjusting (6.2), so that the (pre)-order is still complete but now depends on both the lower and upper expectations. We first recall two solutions belonging to the second trend, before shifting to those belonging to the first one (i.e. relaxing completeness). In the latter case, most derived criteria consist in pair-wise comparisons of actions. It is therefore useful to recall that (6.2) can be seen as the result of pair-wise comparison of acts, that is a is preferred to c , or $a \geq_{P_X} c$ whenever $E_{P_X}(u_a) \geq E_{P_X}(u_c)$, or equivalently when $E_{P_X}(u_a - u_c) \geq 0$.

Γ -maximin [114] and Γ -maximax A straightforward way to extend (6.2) is simply to replace the usual expectation by the lower or the upper expectation of utilities. In the case of the lower expectation \underline{E} , this gives the so-called Γ -maximin criterion

$$opt_{\underline{E}_X}(A) := \max_{a \in opt_{\geq}(A)} \underline{E}_{\mathcal{P}_X}(u_a) \quad (6.3)$$

which corresponds to a worst-case analysis, that is we assume the worst case, and pick the least worst among them. Use of Γ -maximin can be justified by a principle of cautiousness in risk analysis, or in games where you know that the opponent is assumed to choose the probability in \mathcal{P}_X so that the reward is minimal. In Example 6.1, $opt_{\underline{E}_X} = a_5$.

Conversely, replacing the expectation with the upper expectation \bar{E} gives the Γ -maximax criterion

$$opt_{\bar{E}_X}(A) := \arg \max_{a \in opt_{\geq}(A)} \bar{E}_{\mathcal{P}_X}(u_a) \quad (6.4)$$

corresponding this time to a best-case analysis, or to an optimistic view, in which we hope to get the maximal reward. Using Γ -maximax in Example 6.1 results in $opt_{\bar{E}_X} = a_2$.

Hurwicz's criterion [122] This criterion was originally considered for cases of complete ignorance, and consisted in selecting optimal actions by considering a weighted average of the worst and best rewards, with the worst reward receiving a weight α , and the best a weight $1 - \alpha$, with α considered as a pessimism index. Given a weight α , the extension of Hurwicz criterion when uncertainty is modeled by \mathcal{P}_X reads

$$opt_{H_\alpha}(A) := \arg \max_{a \in opt_{\geq}(A)} \alpha \underline{E}_{\mathcal{P}_X}(u_a) + (1 - \alpha) \bar{E}_{\mathcal{P}_X}(u_a) \quad (6.5)$$

and Γ -maximin, Γ -maximax criteria are respectively retrieved by taking $\alpha = 1$ and $\alpha = 0$. In Example 6.1, only actions a_2, a_3, a_5 can be found optimal with Hurwicz's criteria for different values α .

Maximality We now drop the assumption that the (pre)-order on actions must be complete, i.e., we allow optimal actions to be a set of incomparable actions. The first extension, initially considered by Walley [203, ch.3.], consists in considering that $a >_{\underline{E}} c$ whenever $\underline{E}_{\mathcal{P}_X}(a - c) > 0$, that is, in Walley's term, we are ready to pay a (strictly) positive price to exchange action a

with c . This induces a partial order $>_{\underline{E}}$ on the actions, and the *Maximality* criterion reads

$$opt_{>_{\underline{E}}}(A) := \{a \in opt_{\geq}(A) \mid \nexists c \in opt_{\geq}(A) \underline{E}_{\mathcal{P}_X}(u_c - u_a) > 0\} \quad (6.6)$$

and this criterion, apart from its behavioral interpretation, can also be seen as a robust version of (6.2) on \mathcal{P}_X . In Example 6.1, the solution with *Maximality* criterion is $opt_{>_{\underline{E}}} = \{a_1, a_2, a_3, a_5\}$.

Interval Dominance Another robust version of (6.2) that gives another partial order $>_{I_{\underline{E}}}$, such that $a >_{I_{\underline{E}}} c$ whenever $\underline{E}_{\mathcal{P}_X}(u_a) \geq \bar{E}_{\mathcal{P}_X}(u_c)$, in other words, interval $[\underline{E}_{\mathcal{P}_X}(u_a), \bar{E}_{\mathcal{P}_X}(u_a)]$ is on the right hand-side of $[\underline{E}_{\mathcal{P}_X}(u_c), \bar{E}_{\mathcal{P}_X}(u_c)]$. *Interval dominance* then follows as

$$opt_{>_{I_{\underline{E}}}}(A) := \{a \in opt_{\geq}(A) \mid \nexists c \in opt_{\geq}(A) \bar{E}_{\mathcal{P}_X}(u_c) \geq \underline{E}_{\mathcal{P}_X}(u_a)\} \quad (6.7)$$

Since the partial order $>_{\underline{E}}$ refines $>_{I_{\underline{E}}}$, *Interval dominance* criterion usually results in larger sets than *maximality*, and in Example 6.1 gives $opt_{>_{I_{\underline{E}}}} = \{a_1, a_2, a_3, a_5, a_6\}$

E-admissibility this criterion corresponds to the most straightforward robustification of (6.2), and is given by

$$opt_{\mathcal{P}_X}(A) := \bigcup_{P \in \mathcal{P}_X} opt_P(A) \quad (6.8)$$

and in *E-admissibility* applied to Example 6.1 yields $opt_{\mathcal{P}_X} = \{a_1, a_2, a_3\}$

Currently, there is no consensus among which criterion is the "best" choice. In our opinion, that such an absolute best choice exists is dubious, and a criterion should be chosen with respect to the properties we want it to satisfy. For further discussions about properties verified by the criteria given above, we refer to Troffaes [195], Utkin and Augustin [198] and Jaffray and Jeleva [122].

6.1.3 Decision making in random set theory

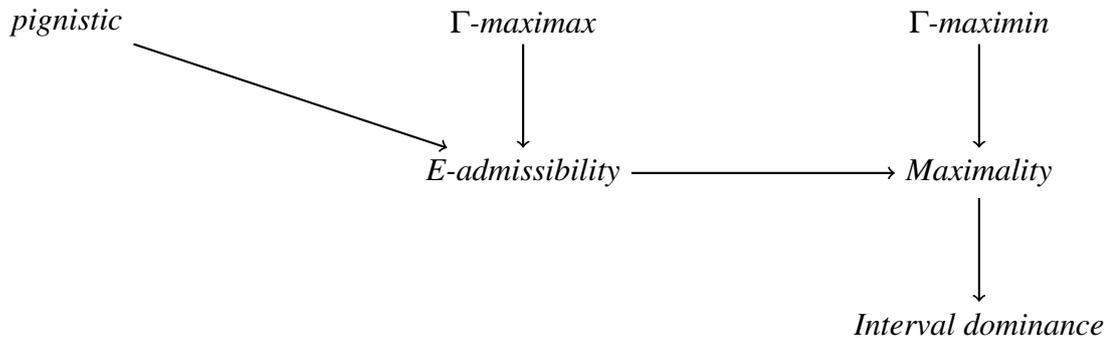
Let \mathcal{P}_X be now a credal set induced by a random set $(m, \mathcal{F})_X$ modeling uncertainty on X , and denote $BetP_X$ the pignistic probability derived from $(m, \mathcal{F})_X$ (see Appendix C). Inside the TBM, Smets [187, 185] justifies axiomatically the use of the pignistic probability $BetP_X$

as a means to determine an optimal action. We define the *Pignistic* criterion as

$$opt_{BetP_X}(A) := \arg \max_{a \in opt_{\geq}(A)} E_{BetP_X}(u_a) \tag{6.9}$$

then follows. In Example 6.1, the pignistic probability is $BetP_X(h) = 0.49, BetP_X(t) = 0.51$, and $opt_{BetP_X} = a_3$.

Note that, since $BetP_X$ comes down to taking the gravity center of \mathcal{P}_X , it can be applied (in principle) to any credal set \mathcal{P}_X , and not only to those whose lower probability is an ∞ -monotone capacity. Also, since $BetP_X \in \mathcal{P}_X$, we have that the optimal action chosen through pignistic criterion is also *E-admissible*, and the following implications between criterion hold:



with $A \rightarrow B$ means that if an act a is in the set of optimal acts in the sense of A , then it is also in the set of optimal acts in the sense of B .

6.2 Practical computations of lower/upper expectations: the case of p-boxes

As seen above, determining a set of optimal actions in imprecise probability theory often necessitates the computations of lower/upper expectations of various utilities. When \mathcal{X} is finite, determining optimal actions with respect to above criteria usually involves solving a finite collection of linear programs, and thus remain feasible, even if the number of linear programs to solve can be pretty high (Utkin and Augustin [198] provide efficient algorithm for various criteria).

Things get more complex when \mathcal{X} is not finite (e.g., the real line), since one would then have to solve an infinite collection of linear programs, which is not feasible in practice. It is

then useful to consider particular cases, for which simplified solutions can be found. Eventually, such solutions could suggest some ways to derive more efficient solutions for more general cases. In this work, we consider the special case where the model is a usual p-box defined over the real-line \mathbb{R} , and the utility u_a for which we want to compute lower/upper expectation is a continuous real function¹.

6.2.1 General problem statement and proposed solutions

Let $[\underline{F}, \overline{F}]$ be a p-box on \mathbb{R} , describing our uncertainty about a variable X (here a closed interval on \mathbb{R} , and let u_a be the utility associated to action a . Computing the lower and upper expectations of u_a , with respect to $[\underline{F}, \overline{F}]$, amounts to solving

$$\underline{E}_{[\underline{F}, \overline{F}]}(u_a) = \inf_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} u_a(x) dF(x) \quad \overline{E}_{[\underline{F}, \overline{F}]}(u_a) = \sup_{\underline{F} \leq F \leq \overline{F}} \int_{\mathbb{R}} u_a(x) dF(x) \quad (6.10)$$

that is, to find, inside $\mathcal{P}_{[\underline{F}, \overline{F}]}$, "optimal" distributions F reaching the infimum and supremum of $\int_{\mathbb{R}} u_a(x) dF(x)$, respectively for the lower and upper expectations. There are at least two general ways to find solutions to (6.10), that we explore here: the use of linear programming and random sets.

Numerically solving (6.10) by linear programming can be done by approximating the (searched) cumulative distribution function F by a set of N points $F(x_i)$, $i = 1, \dots, N$, by translating (6.10) into the corresponding linear programming problem with N optimization variables and with constraints equivalent to those constraining F (i.e. $\underline{F} \leq F \leq \overline{F}$). Those linear programming problems are of the form

$$\underline{E}_{[\underline{F}, \overline{F}]}^*(u_a) = \inf \sum_{k=1}^N u_a(x_k) z_k \quad \text{or} \quad \overline{E}_{[\underline{F}, \overline{F}]}^*(u_a) = \sup \sum_{k=1}^N u_a(x_k) z_k$$

subject to

$$\begin{aligned} z_i &\geq 0, \quad i = 1, \dots, N, \quad \sum_{k=1}^N z_k = 1, \\ \sum_{k=1}^i z_k &\leq \overline{F}(x_i), \quad \sum_{k=1}^i z_k \geq \underline{F}(x_i), \quad i = 1, \dots, N. \end{aligned}$$

where the z_k are the optimization variables, and objective functions $\underline{E}^*(u_a)$ ($\overline{E}^*(u_a)$) are re-

¹With respect to usual Euclidean topology

spectively approximations of the lower (upper) expectation. Note that, when N is large, solving this linear program can be computationally greedy, and not very efficient. Indeed, the optimization problems have N variables and $3N + 1$ constraints. On the other hand, by taking a small value of N , we run the risk of obtaining bad approximations of the exact solution.

Using random sets, we know from Section 3.4 and Appendix F that, provided u_a is continuous, we can safely consider the continuous random set $(m, \mathcal{F})_{[F, \bar{F}]}$ whose bpa m is a uniform law on the unit interval $[0, 1]$ and whose focal elements corresponds to the mapping that associates to each value $\alpha \in [0, 1]$ the interval $[x_\alpha, y_\alpha] = [\bar{F}^{-1}(\alpha), \underline{F}^{-1}(\alpha)]$ where $\bar{F}^{-1}(\alpha) = \sup\{x \in \mathbb{R} | \bar{F}(x) < \alpha\}$ and $\underline{F}^{-1}(\alpha) = \inf\{x \in \mathbb{R} | \underline{F}(x) > \alpha\}$. For easiness of notation, we denote by $\Gamma_{[F, \bar{F}], \alpha}$ this interval. Given this correspondence between p-boxes and random sets, we can rewrite (6.10) into

$$\underline{E}_{[F, \bar{F}]}(u_a) = \int_0^1 \inf_{x \in \Gamma_{[F, \bar{F}], \alpha}} u_a(x) d\alpha, \quad (6.11)$$

$$\bar{E}_{[F, \bar{F}]}(u_a) = \int_0^1 \sup_{x \in \Gamma_{[F, \bar{F}], \alpha}} u_a(x) d\alpha. \quad (6.12)$$

Again, finding an analytical solution to this integral is, in general, not feasible, but the solution can easily be (outer or inner) approximated by considering a finite number $0 \leq \alpha_1 < \dots < \alpha_M \leq 1$ of levels α_i and solving the discretized version of Equations (6.11), (6.12). In the latter case, the main difficulty is to find infimum and supremum of u_a in intervals $\Gamma_{[F, \bar{F}], \alpha_i}$. As in linear programming, computational effort increase with the number of discretization levels, but taking too few of them could lead to high approximation errors, and so would the choice of poor heuristics to detect extrema in the case of complex function u_a .

Note that the cumulative distribution F reaching infimum or supremum depends of the form of u_a , meaning that, if u_a is known to follow some behavior, it is possible to find the analytical form of the searched cumulative distribution F , eventually leading to more efficient numerical methods to approximate solutions of (6.10). The simplest examples (for which solutions are well known) of such typical cases are monotone functions.

Let u_a be such a monotone function non-decreasing (non-increasing) in \mathbb{R} , then the well

known result [204]:

$$\begin{aligned} \underline{E}_{[E,\bar{F}]}(u_a) &= \int_{\mathbb{R}} u_a(x) d\bar{F}(x) \left(\underline{E}_{[E,\bar{F}]}(u_a) = \int_{\mathbb{R}} u_a(x) dF(x) \right), \\ \bar{E}_{[E,\bar{F}]}(u_a) &= \int_{\mathbb{R}} u_a(x) dF(x) \left(\bar{E}_{[E,\bar{F}]}(u_a) = \int_{\mathbb{R}} u_a(x) d\bar{F}(x) \right), \end{aligned}$$

follows. Using equations (6.11),(6.12), we get

$$\begin{aligned} \underline{E}_{[E,\bar{F}]}(u_a) &= \int_0^1 u_a(x_\alpha) d\alpha \left(\underline{E}_{[E,\bar{F}]}(u_a) = \int_0^1 u_a(y_\alpha) d\alpha \right) \\ \bar{E}_{[E,\bar{F}]}(u_a) &= \int_0^1 u_a(y_\alpha) d\alpha \left(\bar{E}_{[E,\bar{F}]}(u_a) = \int_0^1 u_a(x_\alpha) d\alpha \right) \end{aligned}$$

and (lower/upper) expectations are totally determined by extreme values of the mappings.

6.2.2 Unimodal u_a

We now consider a slightly more complex case, where u_a has one maximum on \mathbb{R} in point $a \in \mathbb{R}$. Although still simple, this case can happen in practice (see example given by Utkin [197]), and will be instrumental to show (some of) the interests of considering jointly linear programming and random set solutions. Situation where u_a has one minimum is similar, due to the duality between upper/lower expectations (i.e., $\underline{E}(u_a) = -\bar{E}(-u_a)$).

Proposition 6.1. *If the function u_a has a single maximum at point $a \in \mathbb{R}$, then the upper and lower expectations of $u_a(X)$ on $[E, \bar{F}]$ are*

$$\bar{E}_{[E,\bar{F}]}(u_a) = \int_{-\infty}^a u_a(x) dF + u_a(a) [\bar{F}(a) - E(a)] + \int_a^{\infty} u_a(x) d\bar{F}, \tag{6.13}$$

$$\underline{E}_{[E,\bar{F}]}(u_a) = \int_{-\infty}^{\bar{F}^{-1}(\alpha)} u_a(x) d\bar{F} + \int_{\underline{F}^{-1}(\alpha)}^{\infty} u_a(x) dF, \tag{6.14}$$

or, equivalently

$$\bar{E}_{[E,\bar{F}]}(u_a) = \int_0^{E(a)} u_a(y_\alpha) d\alpha + [\bar{F}(a) - E(a)]u_a(a) + \int_{\bar{F}(a)}^1 u_a(x_\alpha) d\alpha \tag{6.15}$$

$$\underline{E}_{[\underline{F}, \overline{F}]}(u_a) = \int_0^\alpha u_a(x_\alpha) d\alpha + \int_\alpha^1 u_a(y_\alpha) d\alpha, \quad (6.16)$$

where α is one of the solutions of the equation

$$u_a(\overline{F}^{-1}(\alpha)) = u_a(\underline{F}^{-1}(\alpha)). \quad (6.17)$$

Proof using linear programming (sketch). We assume that functions $u_a, \underline{F}, \overline{F}$ are differentiable in \mathbb{R} . Then the following primal and dual optimization problems can be written for computing the lower expectation of the function u_a :

Primal problem:

$$\text{Minimize } v = \int_{-\infty}^{\infty} u_a(x) \rho(x) dx$$

subject to

$$\rho(x) \geq 0, \int_{-\infty}^{\infty} \rho(x) dx = 1,$$

$$-\int_{-\infty}^x \rho(x) dx \geq -\overline{F}(x),$$

$$\int_{-\infty}^x \rho(x) dx \geq \underline{F}(x).$$

Dual problem:

$$\text{Max. } w = c_0 + \int_{-\infty}^{\infty} (-c(t)\overline{F}(t) + d(t)\underline{F}(t)) dt$$

subject to

$$c_0 + \int_x^{\infty} (-c(t) + d(t)) dt \leq u_a(x), c_0 \in \mathbb{R},$$

$$c(x) \geq 0, d(x) \geq 0.$$

The proof of equations (6.13)-(6.14) and (6.17) then follows in three main steps:

1. We propose a feasible solution of the primal problem.
2. We then consider the feasible solution of the dual problem corresponding to the one proposed for the primal problem.
3. We show that the two solutions coincide and, therefore, according to the basic duality theorem of linear programming, these solutions are optimal ones.

And proving the third point mainly comes down to showing that there are two points a', a'' such that $a' \leq a \leq a''$, that is from each side of the maximum, satisfying the following two equations:

$$\overline{F}(a') = \underline{F}(a''). \quad (6.18)$$

and

$$u_a(a'') = u_a(a'). \quad (6.19)$$

The first corresponding to the feasible solution of primal problem, and the second to the dual. These two conditions in turns lead to exhibit the existence and the role of the level α in (6.17).

The full proof is provided in Appendix D □

Proof using random sets. Let us now consider equations (6.12)-(6.11). Looking first at equation (6.12), we see that for values $\alpha \in [0, 1]$ lower than $\alpha_* = \underline{F}(a)$, the supremum of u_a on $\Gamma_{[\underline{F}, \bar{F}], \alpha}$ is $u_a(y_\alpha)$, since u_a is increasing between $[\infty, a]$. For values α between $\alpha_* = \underline{F}(a)$ and $\alpha^* = \bar{F}(a)$, the supremum of u_a on $\Gamma_{[\underline{F}, \bar{F}], \alpha}$ is $u_a(a)$. And for values grater than $\alpha^* = \bar{F}(a)$, we can make the same reasoning as for the increasing part of u_a (except that it is now decreasing). Finally, this gives us the following formula:

$$\bar{E}_{[\underline{F}, \bar{F}]}(u_a) = \int_0^{\underline{F}(a)} u_a(y_\alpha) d\alpha + \int_{\underline{F}(a)}^{\bar{F}(a)} u_a(a) d\alpha + \int_{\bar{F}(a)}^1 u_a(x_\alpha) d\alpha \quad (6.20)$$

which is equivalent to (6.15). Let us now turn to the lower expectation. For values of α before α_* and after α^* , finding the infimum is again not a problem (it is respectively $u_a(x_\alpha)$ and $u_a(y_\alpha)$). Between α_* and α^* , since we know that u_a is increasing before $x = a$ and decreasing after, infimum is either $h(x_\alpha)$ or $h(y_\alpha)$. This gives us equation

$$\underline{E}_{[\underline{F}, \bar{F}]}(u_a) = \int_0^{\underline{F}(a)} u_a(x_\alpha) d\alpha + \int_{\underline{F}(a)}^{\bar{F}(a)} \min(u_a(x_\alpha), u_a(y_\alpha)) d\gamma + \int_{\bar{F}(a)}^1 u_a(y_\alpha) d\gamma \quad (6.21)$$

and by using results from the first equation or the fact that both x_α, y_α are non-decreasing functions of α , we know that there is a level α such that $u_a(\bar{F}^{-1}(\alpha)) = u_a(\underline{F}^{-1}(\alpha))$, and for which the above equation simplify in equation (6.17). □

Of course, both proofs lead to similar formulas and, in applications, would lead to the same exact lower and upper expectations. Nevertheless, it is interesting to note that each view suggests a different way to approximate the exact solution. Namely, the proof involving linear programming suggested to us a more analytical and explicit solution, where the main difficulty is to find the level α satisfying Equation (6.18). If an analytical solution is not available, then the solution is generally approximated by scanning a larger or smaller range of possible values for α (see Utkin [197] for an example).

On the other hand, the proof is shorter in the case of random set, but analytical results are more difficult to derive. Compared to the linear programming view, equations (6.15),(6.16),(6.21)

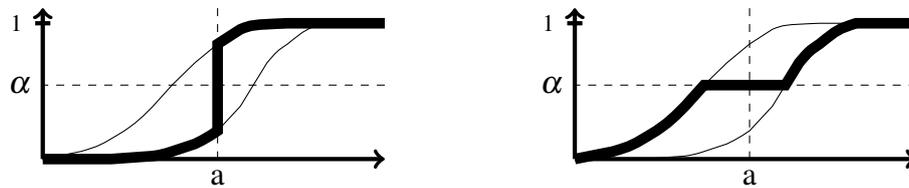


Figure 6.1: u_a with one maximum in a , illustration of cumulative distributions F reaching upper expected value $\bar{E}_{[F,\bar{F}]}(u_a)$ (left) and lower expected value $\underline{E}_{[F,\bar{F}]}(u_a)$ (right)

suggest numerical methods based on a discretization of the unit interval $[0, 1]$ rather than a heuristic search of the level α satisfying equation (6.18). Note that in the worst case, two evaluations are needed at each of the discretized levels (using equation (6.21)).

Figure 6.1 provides an illustration of the shape of distributions functions on which lower and upper expectations are reached. It shows that the lower expectation $\underline{E}_{[F,\bar{F}]}(u_a)$ is obtained with a distribution having an horizontal jump avoiding the higher values of u_a , while the upper expectation $\bar{E}_{[F,\bar{F}]}(u_a)$ is reached by concentrating probability mass on the maximum a

6.2.3 Many extrema

We now consider another univariate case, far more general this time, where u_a has alternate local maximum at points a_i and minimum at points b_i , $i = 1, 2, \dots$, such that

$$b_0 < a_1 < b_1 < a_2 < b_2 < \dots \tag{6.22}$$

Proposition 6.2. *If local maxima (a_i) and minima (b_i) of the function u_a satisfy condition (6.22), then the optimal distribution F for computing the lower unconditional expectation $\underline{E}_{[F,\bar{F}]}(u_a)$ is vertical (has "jumps") at points b_i , $i = 1, \dots$ of the size*

$$\min(\bar{F}(b_i), \alpha_{i+1}) - \max(F(b_i), \alpha_i).$$

Between jumps indexed $i - 1$ and i , the optimal probability distribution function F is of the

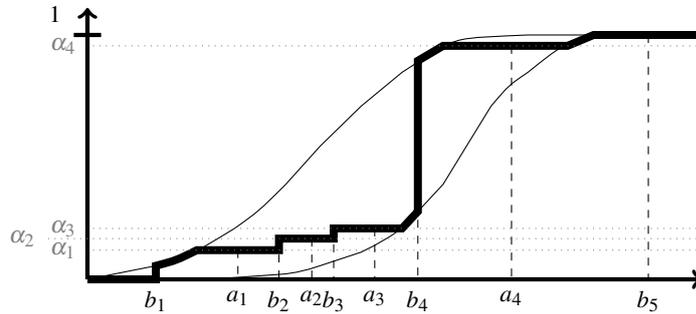


Figure 6.2: u_a with alternate extrema, illustration of cumulative distributions F reaching lower expected value $\underline{E}_{[\underline{F}, \bar{F}]}(u_a)$

form:

$$F(x) = \begin{cases} \bar{F}(x), & x < a' \\ \alpha, & a' \leq x \leq a'' \\ \underline{F}(x), & a'' < x \end{cases}$$

where α is the root of the equation

$$u_a \left(\max \left(\bar{F}^{-1}(\alpha), b_{i-1} \right) \right) = u_a \left(\min \left(\underline{F}^{-1}(\alpha), b_i \right) \right)$$

in interval $[\underline{F}(a_i), \bar{F}(a_i)]$,

$$a' = \max \left(\underline{F}^{-1}(\alpha), b_{i-1} \right), \quad a'' = \min \left(\underline{F}^{-1}(\alpha), b_i \right).$$

The upper expectation $\bar{E}_{[\underline{F}, \bar{F}]}(u_a)$ can be found from the dual relation $\bar{E}_{[\underline{F}, \bar{F}]}(u_a) = -\underline{E}_{[\underline{F}, \bar{F}]}(-u_a)$.

Proofs are a bit more complex than for Proposition 6.1, but follows similar reasonings, and are omitted here. Sketches are provided by Utkin and Destercke [199].

Figure 6.2 illustrates Proposition 6.2. The solution again consists in concentrating probability masses on lower values of u_a , while avoiding the higher ones. Other situations also considered by Utkin and Destercke [199] (multivariate case with different assumptions of independence, conditional expectations) tends to suggest that this result can be generalized in most cases and could be used in very general situations.

Further works on this topic include the design and implementation of algorithms derived

from the studied situations, the exploration of other and more general cases, the extensions of presented results to other practical representations such as possibility distributions, clouds, imprecise probability assignments,

6.3 Decision in industrial risk analysis

Decision making in industrial risk analysis and in safety studies is usually a bit different from the problem stated above. First, the aim of such studies is not to take optimal, but rather safe actions. In such problems, variable X will in general assume values on the real line and the decision will generally depends on the probability of exceeding a threshold, that is on the (possibly imprecise) evaluation of $P((-\infty, x^*])$, with x^* the threshold. With this respect, two main kind of studies and associated decision can be taken:

- *Prospective studies*: we call prospective those studies where nothing has been done so far, and the decision maker wants to know for which value x^* the probability $P([x^*, \infty))$ of exceeding this threshold will be below a certain level $1 - \alpha$ (typical values are 0.1, 0.001, . . .), that is for which value x^* do we have at least $\underline{P}([-\infty, x^*]) = \alpha$. Systems are then dimensioned with respect to that x^* , so that the associated risk is judged acceptable. For example, X can characterize the water level of a river, and x^* is then used to dimension a dam ensuring that no flood will occur. In terms of behavior, it means that the decision maker wants to (or has to) buy a gamble at a fixed price α , and the decision consists in determining for which value x^* this price appears acceptable.
- *Retrospective studies*: we call retrospective those studies where the situation is fixed, and the decision maker wants to assess the probability that X do not exceed some critical levels x^* , in order to know if some action has to be undertaken to lower this probability. In this case, x^* is fixed and the study consists in checking that $\underline{P}((-\infty, x^*])$ is above some level α . If it is not, then some action should be done to increase $\underline{P}((-\infty, x^*])$. For example, x^* can be a critical temperature in a nuclear reactor core, and the corresponding action could be to activate some coolant system or not. In terms of behavior, it means that the decision maker is somewhat forced to buy the gamble $\mathbf{1}_{((-\infty, x^*])}$ for a price α , and wants to keep this price acceptable.

Decision making for industrial risk analysis and safety studies is somewhat reversed compared to the classical definition of decision making, since the problem is not to determine the action that would give us the highest "reward" but rather to check that the minimal reward

will be at least α , or to act in order to reach this minimal level. Note that here, "reward" is usually inversely proportional to spent money, since the higher the decision maker safety requirements² (the higher α), the more money will be spent.

The material contained in this chapter can be found in paper [199].

²Or, equivalently, the more risk-adverse he is

Chapter 7

Illustrative applications

“As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality”

— Albert Einstein (1879–1955)

In this chapter, we detail two illustrative applications using some of the methods studied previously. The first one (Section 7.1) applies information evaluation methodologies and basic fusion operators to the results of uncertainty studies performed with nuclear compute codes, and the second (Section 7.2) proposes and illustrates on a case study a numerical method of hybrid propagation allowing to propagate uncertainty in an efficient way while coping with numerical accuracy.

7.1 Information evaluation and fusion applied to nuclear computer codes

In this section, we apply methods developed in Section 4.4 to evaluate and fuse information coming from multiple sources. They are applied to results of uncertainty studies achieved with nuclear computer codes. All computations have been done with the SUNSET software developed at the IRSN, in which the methods have been implemented.

Only probabilistic and possibilistic approaches will be used here, since this is sufficient to illustrate the benefits of allowing some imprecision in uncertainty representations.

7.1.1 Introduction to the problem

Evaluating nuclear power plant performance during transient conditions is a very important issue in thermal-hydraulic research since nuclear energy was used to produce electricity. Across the years, a huge amount of experimental data has been produced from very simple loops and from Integral Test Facilities. A lot of computer codes have also been developed and made available to the nuclear community in order to simulate variables of interest during transient conditions. It is important to evaluate the predicting reliability of such codes by comparing their results to experimental data obtained from small scale facilities. Past years have witnessed an increasing interest in the combination of such codes with uncertainty analysis, allowing for a more realistic modeling of the parameter knowledge, and thus helpful to make better previsions. Nevertheless, practitioners often find it difficult to compare and to analyze the final results of such uncertainty analyses, as well as to assess the agreement level of such results with experimental data. This is why we have applied methods from section 4.4 to the results of the BEMUSE (Best Estimate Methods - Uncertainty and Sensitivity Evaluation) programme [160] performed by the NEA (Nuclear Energy Agency).

Our study focuses here on the results of the first step of the programme, in which ten participants from nine organisations were brought together in order to compare their respective uncertainty analysis with experimental data coming from the experiment L2-5 performed on the loss-of-fluid test (LOFT) facility. Although most participants (9 out of 10) used similar methodologies to complete their uncertainty evaluations, their results were quite different, due to the fact that different codes were used and that the number, models and physical nature of inputs were different for each participant.

Since a nuclear reactor generates internal heat, this heat has to be removed by a coolant system. A loss-of-coolant accident happens when the flow of coolant is reduced (in our case, by the simulation of a guillotine rupture of an inlet pipe). When such an accident happens, emergency systems are designed to stop the fission process. Nevertheless, even after such a stop, significant amount of heat may still be generated due to radioactive decay, and this heat can cause important damage to the facility, resulting in catastrophic consequences for the reactor, its facility and vicinity. Safety systems thus have to ensure that parameters such as pressure and temperature remain below critical levels.

The ten participants of the BEMUSE programme, as well as the code they used and their organization are summarized in table 7.1. In the first step of BEMUSE programme, the L2-5 experiment has been chosen to apply uncertainty methodologies on a large break loss-of-coolant accident (LB-LOCA transient) performed on an integral test facility. The L2-5 ex-

| Used code | Participants |
|-----------|------------------------|
| ATHLET | GRS, NRI2 |
| CATHARE | CEA, IRSN |
| MARS | KAERI |
| RELAP5 | KINS, NRI1, UNIPI, UPC |
| TRACE | PSI |

Table 7.1: Participants of BEMUSE programme and used codes

periment has been completed on 16 June 1982 in the LOFT facility at INEL (Idaho National Engineering Laboratory). This facility simulated the major components and the system responses of a commercial PWR (Pressurized Water Reactor) during a loss-of-coolant accident (LOCA). The core was a semi-scale one with an active height of 1.70m. The experimental assembly included five major subsystems which were instrumented with measurement devices.

As an output of their uncertainty analysis, each participant had to provide lower bounds, reference values and upper bounds for four scalar output parameters as well as the time trends of two output parameters (maximum cladding temperature and upper plenum pressure). For each of these output parameters, experimental values are available (thus, they can be taken as so-called seed variables to assess sources predictive quality). Here, we have only considered the four scalar output parameters. These four scalar output parameters are:

1. The first Peak Cladding Temperature (1PCT) during the blowdown phase
2. The second Peak Cladding Temperature (2PCT) during the reflood phase
3. The Time of accumulator injection (T_{inj})
4. The Time of complete quenching (T_q)

Table 7.2 summarizes the values given by the participants for the lower bounds, reference calculation and upper bounds for each output. Obtained experimental values are also recalled.

| | 1PCT (Kfl) | | | 2PCT (Kfl) | | | T_{inj} (s) | | | T_q (s) | | |
|-----------|------------|------|------|------------|------|------|---------------|------|------|-----------|------|-------|
| | Low | Ref | Up | Low | Ref | Up | Low | Ref | Up | Low | Ref | Up |
| CEA | 919 | 1107 | 1255 | 674 | 993 | 1176 | 14.8 | 16.2 | 16.8 | 30 | 69.7 | 98 |
| GRS | 969 | 1058 | 1107 | 955 | 1143 | 1171 | 14 | 15.6 | 17.6 | 62.9 | 80.5 | 103.3 |
| IRSN | 872 | 1069 | 1233 | 805 | 1014 | 1152 | 15.8 | 16.8 | 17.3 | 41.9 | 50 | 120 |
| KAERI | 759 | 1040 | 1217 | 598 | 1024 | 1197 | 12.7 | 13.5 | 16.6 | 60.9 | 73.2 | 100 |
| KINS | 626 | 1063 | 1097 | 608 | 1068 | 1108 | 13.1 | 13.8 | 13.8 | 47.7 | 66.9 | 100 |
| NRI1 | 913 | 1058 | 1208 | 845 | 1012 | 1167 | 13.7 | 14.7 | 17.7 | 51.5 | 66.9 | 87.5 |
| NRI2 | 903 | 1041 | 1165 | 628 | 970 | 1177 | 12.8 | 15.3 | 17.8 | 47.4 | 62.7 | 82.6 |
| PSI | 961 | 1026 | 1100 | 887 | 972 | 1014 | 15.2 | 15.6 | 16.2 | 55.1 | 78.5 | 88.4 |
| UNUPI | 992 | 1099 | 1197 | 708 | 944 | 1118 | 8.0 | 16.0 | 23.5 | 41.4 | 62.0 | 81.5 |
| UPC | 1103 | 1177 | 1249 | 989 | 1157 | 1222 | 12 | 13.5 | 16.5 | 56.5 | 63.5 | 66.5 |
| Exp. Val. | 1062 | | | 1077 | | | 16.8 | | | 64.9 | | |

Table 7.2: Scalar output values by participants (Exp. Val. : Experimental value)

7.1.2 Modeling the information

The lower and upper given by the participants were respectively the lowest and highest values obtained for 156 runs of their computer codes. According to order statistics [27], we considered, as a first approximation, these values as the 1% and 99% percentiles. So, the *Low* and *Up* provide, for each participant and each variable, the two percentiles $q_2 = 1\%$ and $q_4 = 99\%$.

Given a particular output, let us call q_{\min} and q_{\max} the minimal and maximal values of the lower and upper bounds of this output, taken over all participants. Then, for each output, we take $[q_l, q_u]$ as the interval $[q_{\min}, q_{\max}]$ increased by 2% (e.g. for 1PCT, $q_{\min} = 626$ (KINS), $q_{\max} = 1255$ (CEA) and $[q_l, q_u] = [620, 1261]$). Note that, for a given output, the interval $[q_l, q_u]$ is common to all participants, to make sure that their informativeness scores will be comparable. According to this information, we take the following models:

Probabilistic model: Since the reference values *Ref* are often close to the middle of interval $[Low, Up]$, and as nominal values are often associated to the median of the distribution, we have chosen to take, for each participant and output, the following distribution : $(q_1 = 0\%, q_2 = 1\%, q_3 = 50\%, q_4 = 99\%, q_5 = 100\%) = (q_l, Low, Ref, Up, q_u)$. For example, the distribution corresponding to the information given by NRI1 for the 2PCT is $(q_1 = 592, q_2 = 845, q_3 = 1012, q_4 = 1167, q_5 = 1228)$. The only exception to this rule is the distribution of KINS for T_{inj} , since concentrating 50% of the probability mass on a single value would make

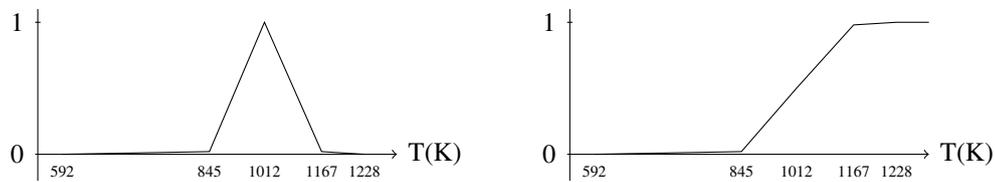


Figure 7.1: Probability (right) and possibility (left) dist. of NRI1 for the 2PCT

no sense. Thus, the distribution of KINS for T_{inj} is ($q_1 = 0\% = 7.8, q_2 = 1\% = 13.1, q_2 = 99\% = 13.8, q_4 = 100\% = 23.7$).

Possibilistic model: The interval $[q_l, q_u]$ common to each source is considered as containing with certainty the true unknown value. The interval $[Low, Up]$ provides for each source a 98% confidence interval, while it is natural to consider the nominal value Ref as the most plausible one. For each source, the possibility distribution that fits this information is s.t. $\pi(q_l) = 0, \pi(Low) = 0.02, \pi(Ref) = 1, \pi(Up) = 0.02, \pi(q_u) = 0$ (with linear interpolation between each points). When taken as an imprecise probabilistic model, this possibility distribution dominates the chosen probabilistic model (see [11]). Figure 7.1 illustrates both models built from the information of NRI2 concerning the second PCT.

7.1.3 Evaluating the sources

For the evaluation steps, the four scalar parameters were considered as seed variables, as their (precise) experimental values are known. Evaluation is performed according to the methodologies described in Section 4.4, with the uncertainty models given above. Table 7.3 summarizes the obtained informativeness, calibration and global scores for both approaches.

The results shown in Table 7.3 confirms that the two methodologies, being based on the same rational requirements, gives comparable results, the few noticeable differences (e.g., informativeness scores of KINS, Calibration scores of GRS and NRI2) being explainable by the formal differences existing between the two methodologies (see Section 4.4, Sandri *et al.* [174] and Destercke and Chojnacki [62] for ampler discussions)

Also, one of the reasons why at least 10 seed variables should be used within probabilistic methodology is illustrated by our results, where only 4 seed variables were used. Indeed, the probabilistic approach results in six different calibration scores, and have a reduced discriminative power when compared to the possibilistic approach, for which each source have

| | Prob. approach | | | Poss. approach | | |
|-------|----------------|------------|------------|----------------|-----------|-----------|
| | Inf. | Cal. | Global | Inf. | Cal. | Global |
| CEA | 8 (0.77) | 5 (0.16) | 6 (0.12) | 8 (0.71) | 6 (0.55) | 7 (0.40) |
| GRS | 4 (1.23) | 1 (0.98) | 1 (1.21) | 3 (0.84) | 7 (0.52) | 6 (0.44) |
| IRSN | 5 (0.98) | 2 (0.75) | 2 (0.73) | 6 (0.73) | 1 (0.83) | 1 (0.60) |
| KAERI | 9 (0.68) | 5 (0.16) | 7 (0.11) | 9 (0.70) | 8 (0.48) | 8 (0.34) |
| KINS | 3 (1.29) | 5 (0.16) | 5 (0.21) | 7 (0.72) | 3 (0.67) | 3 (0.49) |
| NRI1 | 7 (0.79) | 2 (0.75) | 3 (0.59) | 5 (0.75) | 5 (0.63) | 4 (0.47) |
| NRI2 | 6 (0.79) | 8 (0.13) | 8 (0.10) | 4 (0.78) | 2 (0.72) | 2 (0.56) |
| PSI | 1 (1.6) | 10 (0.004) | 10 (0.008) | 1 (0.88) | 10 (0.25) | 10 (0.22) |
| UNIPI | 10 (0.53) | 2 (0.75) | 4 (0.4) | 10 (0.69) | 4 (0.67) | 5 (0.46) |
| UPC | 2 (1.44) | 9 (0.02) | 9 (0.025) | 2 (0.87) | 9 (0.28) | 9 (0.24) |

Table 7.3: Results of sources evaluation (Inf.: informativeness ; Cal.: Calibration) by ranks (values)

received a different calibration score (note that this remains true for all imprecise probability theories).

Such comments are useful to highlight formal advantages or deficiencies of the methods, but are of little use to the analyst, to decision makers or to participants. On the contrary, the following observations concerning the results were found interesting by various researchers in the field of nuclear safety:

- **Ranking with respect to the used code:** the ranking of the participants is poorly correlated with the particular code used to achieve the computations. This indicates and confirms the importance of user-influence on the final results, irrespectively of the used code.
- **Coherence with informal observations:** in BEMUSE reports [160], it was observed that only UPC and PSI bounds did not envelop the PCT experimental values (respectively for the first and second PCT), one of the reason given to explain this was the very narrow uncertainty band considered by both UPC and PSI. Results give formal justification to such informal observations, since UPC and PSI both obtain the worst and best rankings respectively for calibration and informativeness.
- **Code evaluation/validation:** an important issue and a recurrent problem when modeling physical phenomena with complex computer codes is the validation of the results

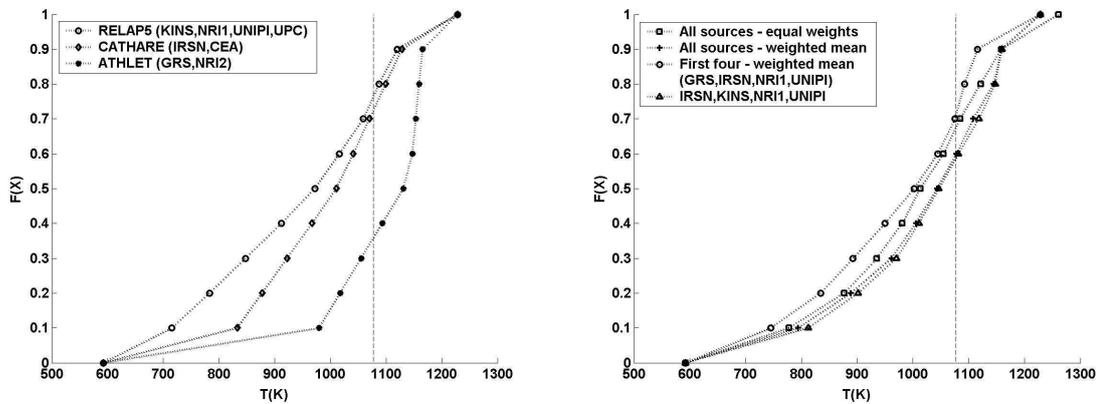


Figure 7.2: Application of probabilistic aggregation

provided by those codes [196]. Proposed methods can be used to achieve such a validation. Note that recent propositions done by Ferson *et al.* [106] can also be considered.

7.1.4 Merging the information supplied by the sources

We now apply the fusion operators introduced in Section 4.1.3 to the second PCT. Interests and defects of each operator are illustrated, as well as how they can help to analyze the information and the relations between sources (here, the participants of the BEMUSE programme).

Probabilistic aggregation Figure 7.2 shows the result of aggregating the probability distributions of some participants. Each arithmetic mean is used with the associated weights, except when specified so on the figure (i.e. all sources with equal weights).

As we see, grouping participants by used codes (left figure) gives poorly calibrated results. CATHARE and RELAP5 users tend to underestimate the experimental value, while ATHLET users tend to overestimate it. Few can be said about the agreement between code users.

The right figure shows how the scores given to each participant can be used to improve the aggregated distribution, both in term of precision and of quality. Interestingly enough, the best distributions are the one in which all sources are taken into account with their associated scores, and the one considering the four common participants being in the five best scored sources of each approach. Both these two distributions are slightly narrower and more centered around the experimental value than the two others. This shows that using the scores in the aggregation is useful and that the two approaches can help each other in the selection of the

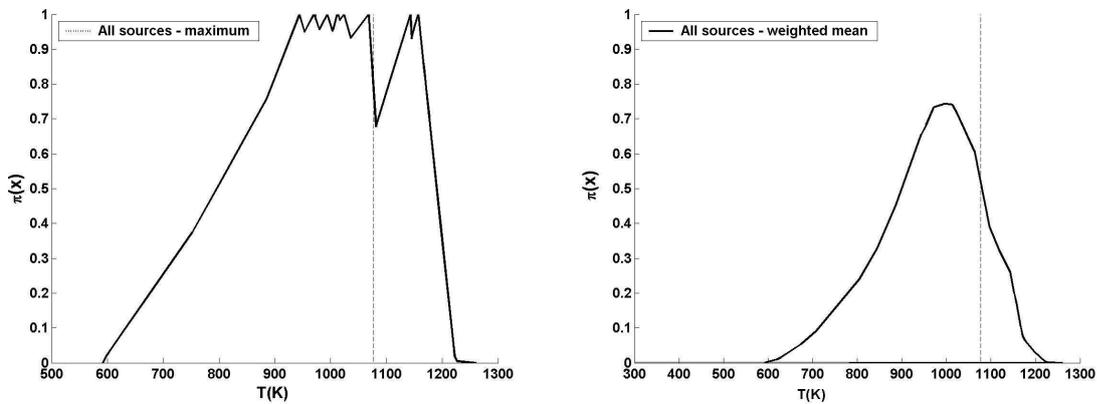


Figure 7.3: Application of possibilistic aggregation : disjunction (left) and weighted mean (right)

best sources. Here again, an eventual conflict between sources is hardly visible. The fact that the arithmetic mean tends to average the resulting distribution is shown in the right figure, in which resulting distributions, although different, remain close to each others. Indeed, we can see here that taking the average is not very discriminative, especially in our case where information given by sources are similar.

Possibilistic aggregation Figure 7.3 shows the result of applying the disjunctive operator (i.e. maximum) and the usual compromise operator (i.e. weighted arithmetic mean) to the set of all sources (taking smaller sets of sources do not bring any really useful extra information in these two cases). As expected, the result of the disjunction is quite imprecise and the arithmetic mean averages the contribution of all participants, resulting in a smooth distribution which has a peak around 1000 K.

Some interesting (and surprising) facts can be said about these distributions. The fact that, in the distribution resulting from the disjunction, more peaks are below rather than above the experimental value indicates that most sources tends to underestimate it. This is somewhat confirmed by the distribution resulting from the arithmetic mean, whose peak is slightly below the experimental value.

A more surprising characteristic is the relatively low possibility degree around the experimental value that exhibits the distribution resulting from the disjunction. Indeed, the possibility degree of the experimental value is around 0.8, which is low if we compare it to possibility degrees of values surrounding the experimental value. This drop comes from the fact that the

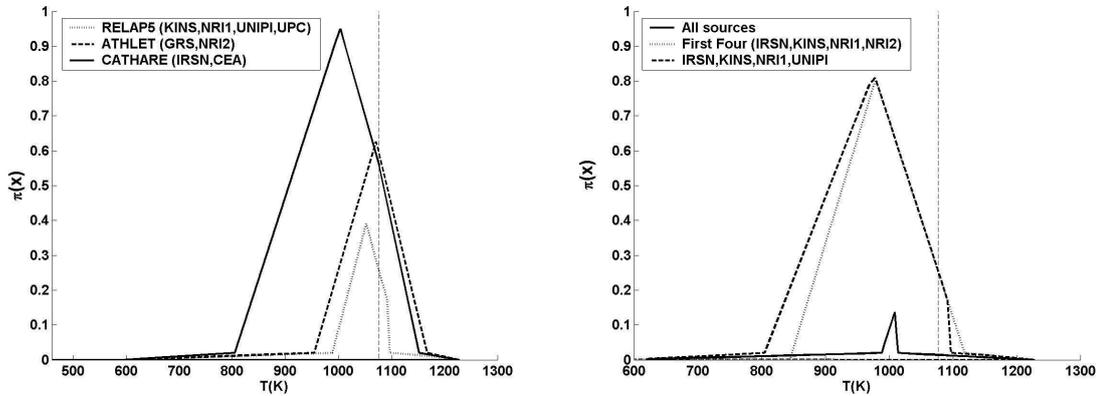


Figure 7.4: Application of possibilistic aggregation : conjunction (minimum)

reference value of most participant is not very close from experimental data (this is not the case for the first PCT) and that KINS, whose reference value is the closest to the experimental value, also gives a very low upper bound (in fact, the lowest outside of PSI, the only participant having an upper bound lower than the experimental value).

Figure 7.4 shows the result of applying the conjunctive operator (i.e. minimum) to various subgroups of participants. The eventual conflict among each subgroup is here directly visible. For instance, we see that, concerning the second PCT, the information given by both users of CATHARE code are coherent, while the information given by ATHLET users are more conflicting. The higher conflict shown by RELAP5 users is not surprising, since they are more numerous.

The right figure shows that the information given by all sources concerning the second PCT is highly conflicting (conflict ~ 0.9), and thus that the resulting conjunction, although very precise, is judged to be highly unreliable. Inversely, limiting ourselves to the most highly scored participants (either only by possibilistic approach or by both approaches) results in distributions that are reliable (conflict only ~ 0.2). We see that using conjunction with only the most reliable sources results in a distribution well balanced between precision and reliability.

Note that the MCS method (see Section 4.2.2) was not applied to the above data and possibility distributions for two main reasons, a bad and a good one:

- the bad one is that we have not found the time yet to implement the MCS method inside the SUNSET software and
- the good one is that, since information provided by the 10 sources display an high con-

sistency (removing only PSI from the group of all sources in Figure 7.4 would remove most of the conflict), we believe that applying the MCS to such data would not bring much more interesting information.

7.2 Hybrid propagation with numerical accuracy

In practice, when propagating uncertainty representations through a deterministic model, an exact propagation can rarely be achieved, except in problems involving only simple models and/or uncertainty representations. This difficulty has many sources: a single run of complex computer codes can take hours to complete (if not days or weeks); interval analysis or set propagation with non-linear models often involve the use of complex algorithms [123]; complexity of uncertainty representations often increase exponentially with the number of input dimensions (a problem often known as the curse of dimensionality, a term first coined by Bellman [14]).

It is thus important, when possible, to propose numerical methods allowing to reduce and optimize the number of required computations, and when doing so, to control the generated numerical error. Here, we propose such a numerical improvement for a popular uncertainty propagation method: the so-called hybrid propagation [9, 12]. The proposed numerical method, called RaFu (for Random/Fuzzy) and implemented in the SUNSET software, relies on the fact that the final desired result is often known before the propagation happens, and that computing only this final result often allows some simplifications.

7.2.1 RaFu method: efficiency in numerical hybrid propagation

We consider that uncertainty bearing on variables $X_{(1:N)}$, each assuming values on the real line \mathbb{R} , (we use the same notation as in the beginning of Chapter 5¹) has to be propagated through a deterministic and functional model $T : \mathcal{X}_{(1:N)} \rightarrow \mathcal{Y}$ from $\mathcal{X}_{(1:N)}$, the *input space*, to \mathcal{Y} , the *output space*.

Given variables $X_{(1:N)}$, Hybrid propagation [9, 12] proposes to differentiate variables tainted with aleatory uncertainty (stemming from natural variability) from variables tainted with epistemic uncertainty (stemming from a lack of knowledge).

¹ X_1, \dots, X_N are the input variables, $\mathcal{X}_1, \dots, \mathcal{X}_N$ the input spaces and x_i an element of \mathcal{X}_i . For $1 \leq k \leq \ell \leq N$, note $\mathcal{X}_{(k:\ell)} := \times_{i=k}^{\ell} \mathcal{X}_i$, $X_{(k:\ell)} := (X_k, \dots, X_\ell)$ a variable assuming values in $\mathcal{X}_{(k:\ell)}$, and $x_{(k:\ell)} := (x_k, \dots, x_\ell) \in \mathcal{X}_{(k:\ell)}$ an element of $\mathcal{X}_{(k:\ell)}$

7.2.1.1 Usual Hybrid method

Without loss of generality, let variables $X_{(1:k)}$ be tainted with aleatory uncertainty, and variables $X_{(k+1:N)}$ with epistemic uncertainty. Aleatory uncertainty on the first is modeled by means of probability distributions P_1, \dots, P_k , while epistemic uncertainty on the second is modeled by means of convex possibility distributions π_{k+1}, \dots, π_N .

A joint uncertainty model is then built as follows:

- (in)dependencies between variables $X_{(1:k)}$ are supposed fully known, and modeled through a joint probability distribution $P_{(1:k)}$.
- Possibilistic non-interaction is assumed between variables $X_{(k+1:N)}$, leading to a joint possibility distribution $\pi_{PI,(k+1:N)}$.
- Joint uncertainty model is the fuzzy random variable $(m, \widetilde{\mathcal{F}})_{H,(1:N)}$ (see Section 3.5.2), with $m_{H,(1:N)}(\pi_{PI,x_{(1:k)},{(k+1:N)}}) = P_{(1:k)}(x_{(1:k)})$ and $\pi_{PI,x_{(1:k)},{(k+1:N)}} \in \widetilde{\mathcal{F}}_{H,(1:N)}$ the possibility distribution such that, for $i = 1, \dots, k$, $\pi_i(x_i) = 1$ and zero elsewhere on \mathcal{X}_i , and possibilistic non-interaction is assumed between π_1, \dots, π_N

Hybrid propagation then consists in propagating $(m, \widetilde{\mathcal{F}})_{H,(1:N)}$ through the model T , to obtain the fuzzy random variable $(m, \widetilde{\mathcal{F}})_{T(H,(1:N))}$. In practice, making an analytical and exact propagation of $(m, \widetilde{\mathcal{F}})_{H,(1:N)}$ will be impossible in most cases, and in practice, $(m, \widetilde{\mathcal{F}})_{T(H,(1:N))}$ is often numerically approximated by a methodology similar to the following one [12]:

1. Generate M_P samples $x_{(1:k)_1}, \dots, x_{(1:k)_{M_P}}$ of $P_{(1:k)}$ by a sampling procedure (e.g. Monte-Carlo, LHS, ...)
2. For each value $(x_{(1:k)})_i, i = 1, \dots, M_P$, consider the possibility distribution $\pi_{T,i}$, given by the propagation of $\pi_{PI,(x_{(1:k)})_i,(k+1:N)}$ through T
3. since we have

$$\pi_{T,i\alpha} = T(\pi_{PI,(x_{(1:k)})_i,(k+1:N)}\alpha) \quad (7.1)$$

with $\pi_{PI,(x_{(1:k)})_i,(k+1:N)}\alpha$ the α -cut of $\pi_{PI,(x_{(1:k)})_i,(k+1:N)}$, approximate $\pi_{T,i}$ by computing (7.1) for a finite collection $\alpha_1 < \dots < \alpha_{M_\pi}$ of M_π α -cuts, building $\widehat{\pi}_{T,i}$

4. give a mass of $1/M_P$ to each distribution $\widehat{\pi}_{T,i}, i = 1, \dots, M_P$

the result is a fuzzy random variable $(\widehat{m}, \widehat{\mathcal{F}})_{T(H,(1:N))}$ approximating $(m, \widetilde{\mathcal{F}})_{T(H,(1:N))}$, with a uniform bpa distributed over distributions $\widehat{\pi}_{T,i}$, $i = 1, \dots, M_P$. The sampling procedure is summarized by Figure 7.5.

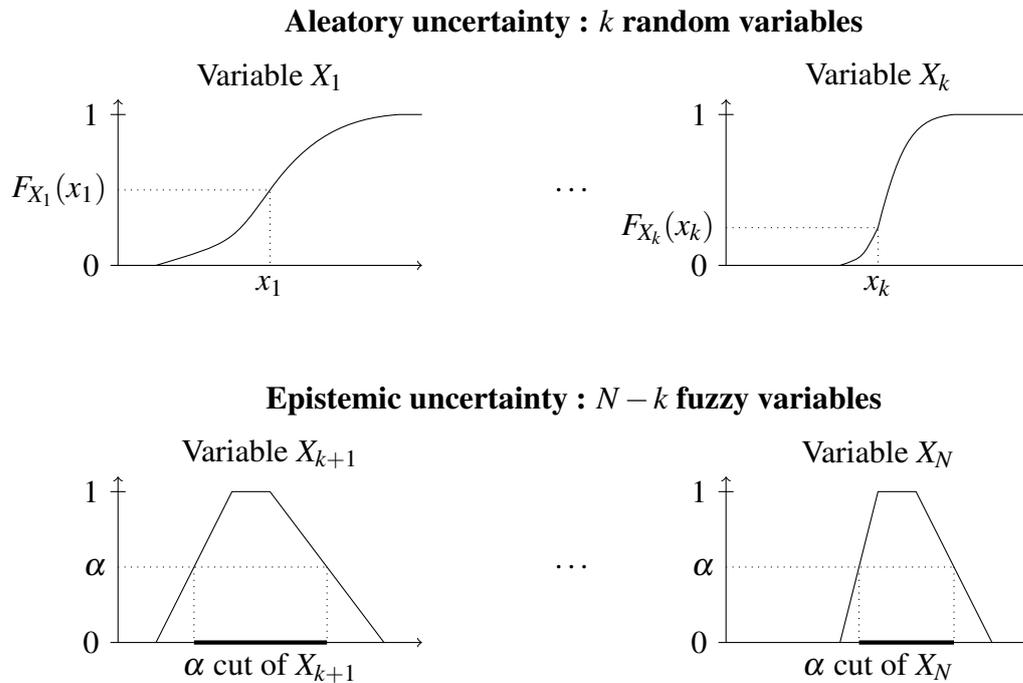


Figure 7.5: Sampling of variables $X_{(1:k)}$ and $X_{(k+1:N)}$ in hybrid numerical propagation.

As for the fuzzy random variable resulting from the fusion process of Section 4.2, information conveyed by $(\widehat{m}, \widehat{\mathcal{F}})_{T(H,(1:N))}$ and the associated representation are hard to handle in practice by a decision maker (particularly if he is not familiar with uncertainty representations). Thus, in order to improve understanding, $(\widehat{m}, \widehat{\mathcal{F}})_{T(H,(1:N))}$ has to be post-processed in some ways.

In this work, we will restrict ourselves to post-processings concerning cumulative functions evaluating the uncertainty of trespassing thresholds, since they are the most useful information in risk analysis studies. First recall that a fuzzy random variable (see Section 3.5.2) can be interpreted as a collection of nested credal sets \mathcal{P}_α , with $\alpha \in [0, 1]$, and $\mathcal{P}_\alpha \subseteq \mathcal{P}_\beta$ for any pair of values such that $\alpha \geq \beta$.

Within the hybrid method, the fuzzy random variable has M_P trapezoidal (discrete) fuzzy focal sets with equal weights, and for each value α , \mathcal{P}_α is induced by the random set having the M_P corresponding α -cuts as focal elements, each having equal weight, as pictured in

Figure 7.6. To simplify notations, we denote by \mathcal{P}_{T_α} the credal set corresponding to level α of $(m, \widetilde{\mathcal{F}})_{T(H, (1:N))}$. For each level $\alpha \in [0, 1]$, the credal set \mathcal{P}_{T_α} induces a p-box, denoted $[\underline{F}, \overline{F}]_{T_\alpha}$. Note that for any pair of values such that $\alpha \geq \beta$, $\overline{F}_{T_\alpha} < \overline{F}_{T_\beta}$ and $\underline{F}_{T_\alpha} > \underline{F}_{T_\beta}$.

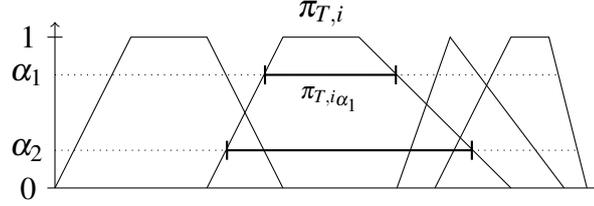


Figure 7.6: Random fuzzy variable.

Baudrit *et al.* [12] and Ferson and Ginzburg [102] have proposed two different post-processings based on p-boxes $[\underline{F}, \overline{F}]_{T_\alpha}$.

Baudrit *et al.* [12] post-processing, called homogeneous post-processing, consists in summarizing the information contained in $(m, \widetilde{\mathcal{F}})_{T(H, (1:N))}$ by a single pair $[\underline{F}, \overline{F}]_{T_E}$ of averaged cumulative distributions such that

$$\underline{F}_{T_E} = \int_0^1 \underline{F}_{T_\alpha} d\alpha ; \overline{F}_{T_E} = \int_0^1 \overline{F}_{T_\alpha} d\alpha$$

note that $[\underline{F}, \overline{F}]_{T_E}$ is equivalent to the p-box induced by the fuzzy random variable when interpreted as a 1st order uncertainty representation (see Section 3.5.2).

Ferson and Ginzburg [102] propose to keep only some relevant values of confidence levels α and to retain the p-boxes corresponding to those levels. For example, considering $[\underline{F}, \overline{F}]_{T_0}$ and $[\underline{F}, \overline{F}]_{T_1}$ comes down to only look at the most pessimistic and most optimistic p-boxes.

7.2.1.2 The RaFu method: more numerical efficiency through pre-processing

Two potential defects of the above methodology are the following:

1. Required number of computation: building $\widehat{(m, \widetilde{\mathcal{F}})}_{T(H, (1:N))}$ can be computationally too expensive. Consider the common choices $M_P = 100$ and $M_\pi = 21$ ($\alpha = \{0, 0.05, \dots, 1\}$) for M_P, M_π . Building $\widehat{(m, \widetilde{\mathcal{F}})}_{T(H, (1:N))}$ then requires 2100 computations. If it is reasonable when T can be quickly evaluated, 2100 computations is often unaffordable when dealing with complex computer codes.

$$\begin{matrix} 1 \\ \vdots \\ i \\ \vdots \\ M \end{matrix} \begin{pmatrix} F_{X_1}^{-1}(\alpha_{1,1}) & \cdots & F_{X_k}^{-1}(\alpha_{k,1}) & [\underline{x}, \bar{x}]_{\alpha_{k+1,1}} & \cdots & [\underline{x}, \bar{x}]_{\alpha_{N,1}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{X_1}^{-1}(\alpha_{1,i}) & \cdots & F_{X_k}^{-1}(\alpha_{k,i}) & [\underline{x}, \bar{x}]_{\alpha_{k+1,i}} & \cdots & [\underline{x}, \bar{x}]_{\alpha_{N,i}} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ F_{X_1}^{-1}(\alpha_{1,M}) & \cdots & F_{X_k}^{-1}(\alpha_{k,M}) & [\underline{x}, \bar{x}]_{\alpha_{k+1,M}} & \cdots & [\underline{x}, \bar{x}]_{\alpha_{N,M}} \end{pmatrix} \begin{matrix} \\ \\ \Rightarrow \\ \\ \end{matrix} T \begin{pmatrix} [\underline{y}, \bar{y}]_1 \\ \vdots \\ [\underline{y}, \bar{y}]_i \\ \vdots \\ [\underline{y}, \bar{y}]_M \end{pmatrix}$$

$$[\underline{y}, \bar{y}]_i = T(F_{X_1}^{-1}(\alpha_{1,i}), \dots, F_{X_k}^{-1}(\alpha_{k,i}), [\underline{x}, \bar{x}]_{\alpha_{k+1,i}}, \dots, [\underline{x}, \bar{x}]_{\alpha_{N,i}})$$

Figure 7.7: Illustration of sample matrix

- Controlling the error/rate of convergence: as numerical approximation means numerical error, it is desirable to have some means to control the error. Up to now, poor attention has been given to such questions when using the hybrid method.

It is therefore desirable to design methods allowing both to reduce the number of required computations to apply the hybrid method and to control or evaluate the numerical error resulting from the numerical propagation. The RaFu method intends to improve these two aspects, mainly by replacing the classical post-processing step by a pre-processing. Since, in practice, a decision maker (DM) will only be interested in some features of the structure $\widehat{(m, \mathcal{F})}_{T(H, (1:N))}$, the RaFu method consists in asking to this decision maker (DM), under the form of a triplet $(\gamma_S, \gamma_E, \gamma_A)$ of parameters, what specific feature of $\widehat{(m, \mathcal{F})}_{T(H, (1:N))}$ he is interested in, and which degree of numerical accuracy does he want to reach. An optimized sampling strategy of distributions $P_{(1:k)}$ and of α -cuts of π_{k+1}, \dots, π_N is then derived to satisfy the DM's choice with a minimal amount of computations. This sampling strategy will take the shape of a sample matrix counting M samples, as summarized in Figure 7.7.

Statistical parameter γ_S : this parameter encompasses two kind of information, contained in two sub-parameters γ_S^i, γ_S^o respectively related to information on input and output variables:

- γ_S^i : concerns information related to the joint distribution $P_{(1:k)}$; It can be, for example, assessment of copulas [158] linking some of the variables or rank correlations [121] between some variables, or simply stochastic independence between all variables
- γ_S^o : concerns information about the statistical value of interest for the DM. For ex-

ample the probability of exceeding a certain threshold value y^* , the entire cumulative distributions, or other statistical quantities such as the mean, the variance, ...

In other words, γ_S specify how samples will be drawn from distributions P_1, \dots, P_k (γ_S^i), and what is the awaited form of the final result (γ_S^o).

Epistemic parameter γ_E : this parameter is related to epistemic uncertainty and to the DM behavior with respect to this uncertainty. It determines how α -cuts or values from π_{k+1}, \dots, π_N have to be sampled. Typical choices for γ_E are:

- Choice 1 .** Fixed $\alpha \in [0, 1]$: for every sample $x_{(1:k)_i}, i = 1, \dots, M$ of P_1, \dots, P_k , take $\pi_{PI,(k+1:N)_\alpha}$ as the sample on π_{k+1}, \dots, π_N , with α a fixed value. This corresponds to fix a confidence level $1 - \alpha$ concerning epistemic uncertainty.
- Choice 2 .** Vector $\alpha = (\alpha_1, \dots, \alpha_J)$ of values: duplicate every sample $x_{(1:k)_i}, i = 1, \dots, M/J$ of P_1, \dots, P_k J times and associate the sampled cut $\pi_{PI,(k+1:N)_{\alpha_j}}$ to the j^{th} duplication. Numerical approximation method described in Section 7.2.1.1 corresponds to this choice (with J the number of discretized levels).
- Choice 3 .** Partially randomized value α : for every sample $x_{(1:k)_i}, i = 1, \dots, M$ of P_1, \dots, P_k , sample a random value α_i from a uniform law on $[0, 1]$ and take $\pi_{PI,(k+1:N)_{\alpha_i}}$ as sample on π_{k+1}, \dots, π_N . As we will see, this is equivalent to averaging over all α -cuts.
- Choice 4 .** Totally randomized value α : for every sample $x_{(1:k)_i}, i = 1, \dots, M$ of P_1, \dots, P_k , sample $N - k$ random values $\alpha_{i,k+1}, \dots, \alpha_{i,N}$ from independent uniform laws on $[0, 1]$, and samples $\pi_{k+1,\alpha_{i,k+1}}, \dots, \pi_{N,\alpha_{i,N}}$ from possibility distributions. This is equivalent to assume random set independence between π_{k+1}, \dots, π_N .

Thus, these two parameters γ_E, γ_S settle which kind of information will be sampled from distributions $P_1, \dots, P_k, \pi_{k+1}, \dots, \pi_N$, as well as the (in)dependency assumptions between them.

7.2.1.3 Integration of numerical error

Numerical Accuracy parameter γ_A : One of the interest of using Monte-Carlo sampling technique or one of its variant (e.g. importance sampling, MCMC, Latin HyperCube Sampling) is that they often comes with convergence theorems that are handy to control or bound approximation errors. Parameter γ_A is related to this numerical error, and has a direct effect on

the number M of samples (or, equivalently, computations) required by a specific application. It can take two different forms: either the DM specifies a numerical accuracy to be reached, and the number of required samples is deduced from it, or the DM provides a maximal number of samples (often limited by available resources), and the numerical accuracy reachable with this number of samples is then deduced.

We only recall here convergence results related to the use of order statistics [27] when evaluating percentiles of cumulated distribution. Let us note X_q the q percentile of a random variable X . From a sample of size M , the use of order statistics consists in considering the ordered values $x_{(1)} \leq \dots \leq x_{(M)}$ drawn from the random variable X . If the M values are drawn randomly and independently, the following equation

$$P(X_{(K)} < X_q) = \sum_{i=K}^M \binom{i}{M} q^i (1-q)^{M-i} \quad (7.2)$$

holds. This is equivalent to saying that the random variable $F_X(X_{(K)})$ follows a beta law of parameters K and $M - K + 1$. The main interest of this result is that $F_X(X_{(K)})$ does not depend of X distribution, therefore it allows to know the number M of samples required to derive a confidence interval for X bounding a given percentile (q) with a given numerical accuracy without knowing neither the values $X_{(i)}$ nor the distribution of X . For example, if a DM wants a conservative upper bound of the 95% percentile that covers it with a confidence of at least 95%, then, by using equation (7.2), it is straightforward to determine that at least 59 computations will be required, since if we draw 58 samples, $P(X_{(58)} < X_{95}) = (0.95)^{58} = 5.1\%$ (i.e. a confidence of 94.9%), while if 59 samples are drawn, $P(X_{(59)} < X_{95}) = (0.95)^{59} = 4.8\%$. This is particularly interesting in risk analysis involving costly computer codes (see, for example, the BEMUSE programme [160]).

Note that results from order statistics to pre-determine the number of required samples cannot be used in the cases where the DM specifies a confidence interval with a minimal width, or if the statistical quantity of interest is not a percentile but another value (e.g. the mean or variance). In this case, a first propagation has to be done, with a prescribed number of M samples from which will be estimated a first confidence interval. Number of propagated samples can then be increased, accordingly to the DM (dis)satisfaction. Note that integrating numerical accuracy add yet another kind of imprecision, deriving from the use of numerical approximation methods. Figure 7.8 illustrates the whole procedure by a flowchart.

The RaFu method is based on the same theoretical assumptions as the original hybrid method, and in this respect does not bring much novelty. However, it tries to solve some of its

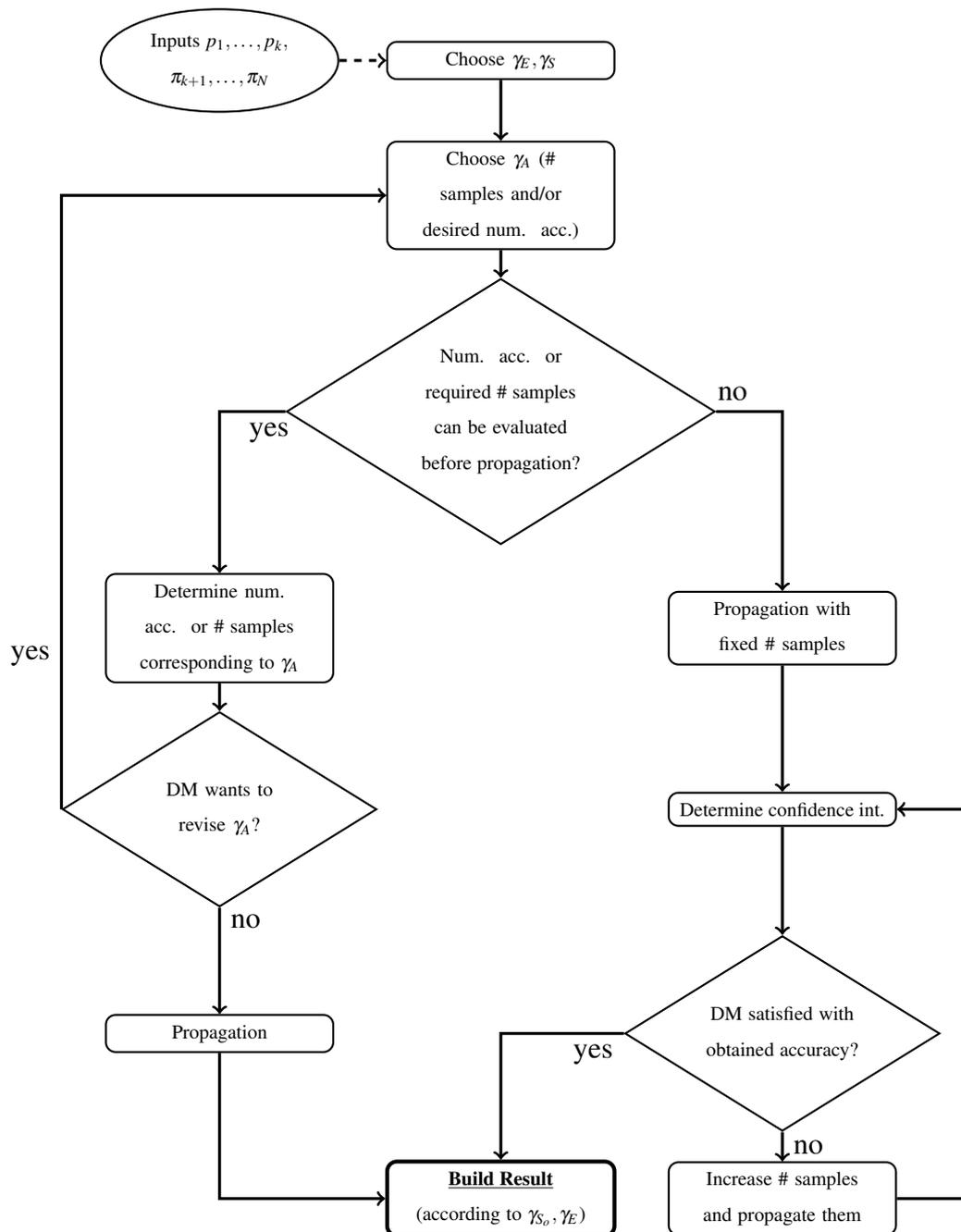


Figure 7.8: RaFu method : flowchart (# samples: number of samples).

practical shortcomings, i.e., reducing its computational cost, making it more "user-friendly" (an important feature in applications) and integrating the notion of numerical approximation error, an important aspect of safety studies involving complex computer codes.

7.2.1.4 Links with existing post-processing

As indicate the two following propositions, it is possible to express both Baudrit *et al.* [12] and Ferson and Ginzburg [102] post-processing methods by suitable choice of the triplet $(\gamma_S, \gamma_E, \gamma_A)$.

Proposition 7.1. *The result of the post-treatment giving $[\underline{F}, \overline{F}]_{T_E}$ can be interpreted as the following choices over γ_S, γ_E :*

- $\gamma_S = F(x) ; \forall x \in \mathbb{R}$ (entire cumulative distribution)
- $\gamma_E = \text{randomized } \alpha \text{ for each sample.}$

Proof. Let us consider, for a value $y \in \mathbb{R}$, the lower probability $\underline{P}_{T_E}([-\infty, y]) = \underline{F}_{T_E}(y)$ associated to Baudrit *et al.*'s post-treatment. Since this lower probability corresponds to the lower expectation of the indicator function $\mathbf{1}_{([-\infty, y])}$ of the event $[-\infty, y]$, it is given by the following formula:

$$\underline{P}_{T_E}([-\infty, y]) = \int_{\kappa=0}^1 \int_{\alpha_1=0}^1 \cdots \int_{\alpha_k=0}^1 \mathbf{1}_{(T(F_{X_1}^{-1}(\alpha_1), \dots, F_{X_k}^{-1}(\alpha_k), \pi_{PI, (k+1:N)_\kappa}) \subset [-\infty, y])} d\kappa d\alpha_1 \dots d\alpha_k \quad (7.3)$$

where distributions P_1, \dots, P_k are assumed to be independent, without loss of generality. This holds for every $y \in \mathbb{R}$, and since making a Monte-Carlo sampling with parameters of Proposition 7.1 is equivalent to a numerical evaluation of Integral (7.3), this finishes the proof for the lower distribution \underline{F}_{T_E} . The proof for the upper one is similar (inclusion in indicator function become a non-empty intersection). \square

Proposition 7.2. *Given a fixed κ , the result of the post-treatment giving $[\underline{F}, \overline{F}]_\kappa$ can be interpreted as the following choices over γ_S, γ_E :*

- $\gamma_S = F(x) ; \forall x$ (whole cumulative distribution)
- $\gamma_E = \kappa$

Proof. We follow a reasoning similar to the one used in the previous proof, except that the integral to evaluate becomes

$$\underline{P}_\kappa([-\infty, y]) = \int_{\alpha_1=0}^1 \dots \int_{\alpha_k=0}^1 \mathbf{1}_{(T(F_{X_1}^{-1}(\alpha_1), \dots, F_{X_k}^{-1}(\alpha_k), \pi_{PI, (k+1:N)_\kappa}) \subset [-\infty, y])} d\alpha_1 \dots d\alpha_k$$

□

Similarly, it can be checked that random set independence assumption is retrieved when every sample is taken randomly in the RaFu method. Pre-processing DM choices rather than post-processing it allows to gain a factor proportional to M_π (the number of discretized α -cuts) in the number of required computations when only some lower and upper cumulative distributions have to be estimated, while keeping the same numerical quality in the final approximation. Of course, some information is lost in the process, but if this information is not relevant for the DM, there are no obvious reasons to keep it, particularly when the number and costs of computations become important issue.

7.2.2 Case-study

We apply the RaFu method on a simplified model used by EDF (French integrated energy operator) to compute the overflowing height for a river dike [139]. Although this model is quite simple, it provides a realistic industrial application in which we can distinguish between aleatory and epistemic uncertainty. This model approximates the overflowing height H of a river and depends on six parameters which are summarized in table 7.4. It reads

$$H = \left(\frac{Q}{K_s \sqrt{\frac{Z_u - Z_d}{L} B}} \right)^{\frac{3}{5}} \quad (7.4)$$

7.2.3 Modeling uncertainty sources

We assume the river width (B) is constant on all the length of the river (L). Both this width and length are assumed to be well known (i.e. no uncertainty on these parameters).

The value of the river flow rate (Q) depends on a huge number of physical phenomena

| Symbol | Name |
|--------|-----------------------|
| Q | River flow rate |
| B | River width |
| K_s | Strickler coefficient |
| Z_u | Upriver water level |
| Z_d | Downriver water level |
| L | River length |

Table 7.4: Summary of parameters used in equation (7.4)

(e.g. climatic and meteorologic conditions, period of the year, ...) that are highly variable over time and/or space. The flow rate value can therefore be interpreted as an aleatory value due to the natural variability of various physical phenomena. As a lot of measurements are usually available for river flow rates, it is possible to fit the data to a probability law modeling this variability. Experience has shown that this variability can be well represented by classical lognormal or Gumbel laws.

Water levels Z_u and Z_d depend on sedimentary conditions that are peculiar to the considered river bed. Due to various reasons, these sedimentary conditions are usually not well known, but are not the consequence of some physical variability or of some random event (since we consider a specific river). The uncertainty of the water levels being due to a lack of information, it is therefore of epistemic nature, and should be modeled by a fuzzy variable.

Similarly, the Strickler coefficient K_s is a model parameter used instead of a physical model to describe the dependance between the flow velocity and the slope of the river. It is also specific to the considered river bed, and the complexity of the river nature makes it difficult to estimate with precision. In our context, the uncertainty linked to such a non-measurable parameter should be modeled by a fuzzy variable as well. Table 7.5 gives the chosen models for our application (considered values and uncertainties are typical values). As an example, figure 7.9 illustrates the distribution modeling the epistemic uncertainty on K_s .

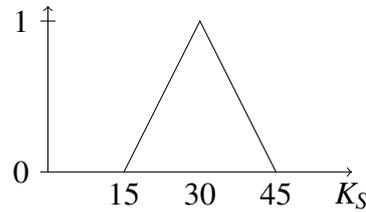


Figure 7.9: Triangular fuzzy number modeling K_s

| Variable | Unit | Model |
|----------|-------------|---|
| Q | m^3s^{-1} | Lognormal law ($m = 7.04$ and $\sigma = 0.6$) |
| Z_u | m | Triangular fuzzy number (54,55,56) |
| Z_d | m | Triangular fuzzy number (49,50,51) |
| K_s | | Triangular fuzzy number (15,30,45) |

Table 7.5: Uncertainty models

7.2.4 RaFu method application

Figure 7.10 shows the results of three applications of the RaFu method, each with 1000 samples. In these applications, the parameter γ_S^2 was the whole cumulative distribution(s) (i.e. $\gamma_S = \{F(x) \forall x\}$), while the various γ_E corresponded to Ferson’s post-treatment for $\alpha = \{0, 1\}$ and to Baudrit et al.’s post-treatment (i.e. $\gamma_E = \text{random } \alpha$ in each sample). For sake of clarity, numerical accuracy is not considered in this figure.

Let us note that, because fuzzy variables (epistemic uncertainty) are modeled by means of triangular fuzzy numbers, taking $\gamma_E = \{\alpha = 1\}$ comes down to suppress this epistemic uncertainty, thus the result is a classical unique cumulative distribution (we consider that both Z_u, Z_d, K_s are precisely known).

Had we built the whole random fuzzy variable to get these five curves, $p \times 1000$ interval computations would have been necessary to reach the same numerical accuracy, where p would have been the chosen number of discretized α -cuts.

Figure 7.11 illustrates how numerical accuracy can influence the result. This figure shows the 95 % confidence interval (i.e. this interval covers the true value with a 95 % confidence)

²Here, only one variable is of aleatory nature, therefore there is no need to specify dependencies

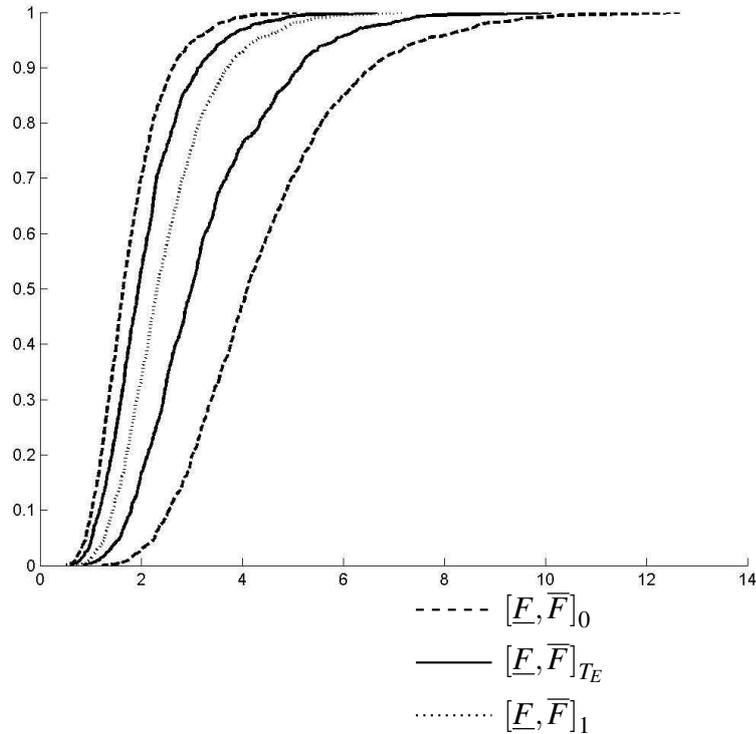


Figure 7.10: Result of Rafu Method with 1000 samples

for the 95 % percentile evaluation and for the three chosen values of γ_E . Since for $\gamma_E = \{\alpha = 0\}$, intervals reduce to single values, we have five series of 1000 values (corresponding to lower/upper bounds of $\gamma_E := \alpha = \{1, 0, T_E\}$). Using order statistics and equation (7.2), we have that the lower and upper bounds of the 95 % numerical confidence interval respectively correspond to the 936th and 964th sorted values. Best estimates are given by the 950th sorted value. Among other things, this figure shows us that the numerical approximation effect is not negligible, even for a relatively high number of computations (here, 1000), and thus should be taken into account in the analysis.

The material contained in this chapter can be found in papers [23, 24, 62].

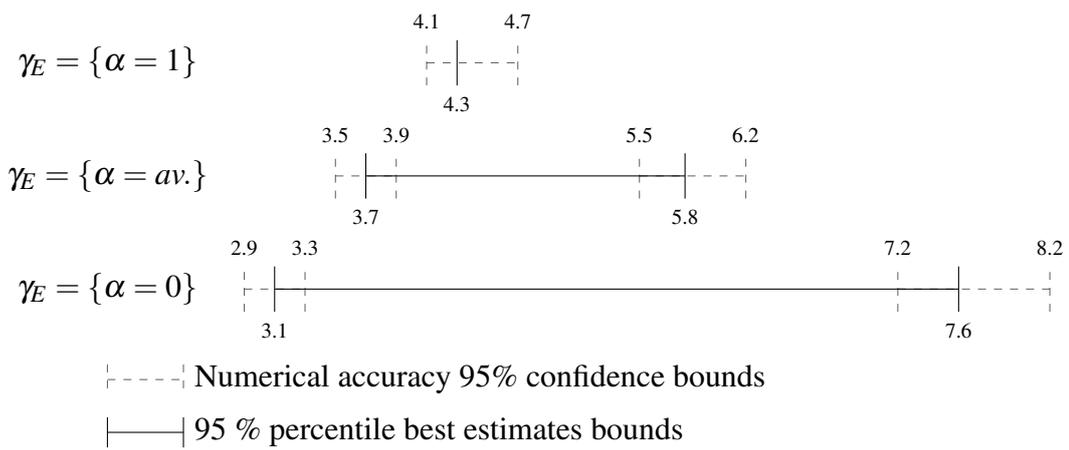


Figure 7.11: Evaluation of the 95% percentile

Chapter 8

Conclusions, perspectives and open problems

“Knowledge would be fatal, it is the uncertainty that charms one. A mist makes things beautiful.”

— Oscar Wilde (1854–1900)

In this work, we have studied various aspects of uncertainty treatment, focusing on the double objective of progressing towards a more unified handling of uncertainty and of proposing tractable solutions for practical problems, by advantageously using the specific features of the different theories considered here.

The main conclusion from Chapter 3, in which we studied practical uncertainty representations, is that generalized p-boxes are interesting uncertainty representations constituting the missing links between these other popular and practical uncertainty models that are possibility distributions, p-boxes and clouds. The fact that they can be interpreted as pairs of lower/upper confidence bounds over collection of nested sets make them attractive for elicitation processes (this still has to be confirmed by experiments), and the fact that they constitute special cases of random sets, while generalizing possibility distributions, let us think that they can be a good compromise between the two, being more tractable than the former and more expressive than the latter. Within this perspective, works to do concerns the practical handling of generalized p-boxes. We have started such a work by studying the propagation and fusion of generalized p-boxes, respectively in Chapter 3 and Chapters 4. On-going works concerns the problem of conditioning on generalized p-boxes, marginalization of generalized p-boxes,

and computation of expectation with generalized p-boxes (in this last case, results from Chapter 6 can probably be extended). Another work that remains to be done is to study to which extend results concerning p-boxes can be extended to generalized p-boxes (e.g., the use of copulas [158], of probabilistic arithmetic [209]).

From Chapter 4, we can conclude some guidelines about the use of fusion operators: conjunctive and disjunctive operators should respectively be used when information is consistent and totally inconsistent, while non-adaptive compromise operators should be used with caution. Adaptive fusion operators, and particularly the use of maximal coherent subsets (MCS), appears to us as the best solution to deal with partial inconsistency in the information and to reconcile sources, at least theoretically. Nevertheless, this approach can quickly lead to computational difficulties, and there is a great need for efficient algorithms to apply it, especially when there are numerous sources. We have studied in details a framework where such efficient algorithms are available, where information is modeled by quasi-concave possibility distributions on the real line, and where MCS methodology is applied level-wise. The question of taking account of eventual dependencies between sources¹ when merging their information is still open, and although propositions exist here and there, they still have to be better axiomatized and studied thoroughly. We have given some first results related to the study of a cautious fusion rule in random set theory maximizing expected cardinality, and which is coherent with cautious possibilistic fusion rules. Further (and on-going) research is needed to characterize this rule, its drawbacks and advantages.

When past assessments of sources are available, we have also proposed a general method to evaluate the quality of the information provided by sources. This method has been applied, within the framework of probability and possibility theory, to the results of a benchmark of nuclear computer codes in Chapter 7.

Conclusions from Chapter 5 is that allowing imprecision in uncertainty representations makes the issue of interpretation essential when using independence notions, both from a theoretical and practical standpoint. Concerning this issue, many problems remains to be solved in both directions, since nowadays, the use of a particular independence notion is mainly dictated by practical convenience (i.e., strong independence in credal nets, random set independence in Monte-Carlo sampling). As indicates our first results, Shafer's recent theory [179] based event-trees is an attractive framework to study and motivate notions of independence, both theoretically and practically (since the use of backward propagation makes computations easier).

¹Note that some solutions have been proposed in the probabilistic setting [124], but they appear again too precise, since dependencies between sources are unlikely to be known with such precision

By studying how the notion random set independence could be outer-approximated by the notion of possibilistic non-interaction, we have explored the usefulness of approximating one notion of independence by another one. Similar studies for other notions of independence could be useful in the practical handling of uncertainty.

Concerning the problem of decision making, briefly studied in Chapter 6, we have proposed first results eventually leading to practical algorithms allowing to compute lower and upper expectations of continuous (utility) functions when the uncertainty on a variable are described by p-boxes. On-going works include the formalization of such algorithms, and the consideration of mixed strategies (i.e., randomized actions).

Some examples of applications have been given in Chapter 7, one of them concerning the treatment of the outputs of multiple computer codes, and the other one the practical propagation of uncertainty by the so-called hybrid method. As this method can be computationally greedy, we have proposed a particular sampling method, called the RaFu method and developed in the IRSN software SUNSET, allowing the number of required computations. This computational reduction is achieved by pre-processing rather than by post-processing some of the decision maker choices.

However, to make propagation of imprecise uncertainty models through complex models affordable and attractive to industrial users, next research efforts should focus on the use of surface responses in combination with imprecise uncertainty models, or in the extension of efficient simulation technics like MCMC techniques to such models.

Appendix A

Uncertainty theories: a short introduction

We give here a short introduction to uncertainty theories used in this work. Further details and discussions can be found in the references (and in the references therein).

A.1 Probability theory: a short introduction

Probability theory is surely the oldest theory allowing to model uncertainty about which values a variable X assume in \mathcal{X} . The theory of probability dates back to Jacob Bernoulli and its *ars coniectandi* [17], and we refer to the first chapters of de Finetti's [108], Shafer's [179] or Walley's [203] for recent reviews of the history of probability and its interpretations.

A *probability mass* p on a finite space \mathcal{X} is a non-negative mapping $p : \mathcal{X} \rightarrow [0, 1]$ such that $\sum_{x \in \mathcal{X}} p(x) = 1$. In the sequel, we will note $\mathbb{P}_{\mathcal{X}}$ the set of all probability masses on \mathcal{X} .

A subset $A \subseteq \mathcal{X}$ of space \mathcal{X} is called an *event*, and given p , the *probability measure* P of the event A is $P(A) = \sum_{x \in A} p(x)$. This measure evaluates the likelihood that event A will happen.

Given a real-valued function $f : \mathcal{X} \rightarrow \mathbb{R}$ and the probability mass p on \mathcal{X} , the expected value $E(f)$ of f is $E(f) = \sum_{x \in \mathcal{X}} p(x)f(x)$. We will note $\mathcal{L}(\mathcal{X})$ the set of all such functions. Remark that the probability $P(A)$ of an event A corresponds to the expectation of the indicator function of A , denoted $\mathbf{1}_{(A)}$, which is such that it takes values one on elements $x \in A$, and zero on elements $x \in A^c$, with A^c the complement of A .

Probability masses can be characterized both in terms of probability measures on event, or

of expected values of functions in $\mathcal{L}(\mathcal{X})$. In terms of events, a probability measure verify the two following axioms:

$$\forall A, B \subset \mathcal{X}, P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{Additivity} \quad (\text{A.1})$$

$$\forall A \subset \mathcal{X}, P(A) = 1 - P(A^c) \quad \text{Duality} \quad (\text{A.2})$$

which, in terms of expected values, read:

$$\forall f, g \in \mathcal{L}(\mathcal{X}), E(f + g) = E(f) + E(g) \quad \text{Additivity} \quad (\text{A.3})$$

$$\forall f \in \mathcal{L}(\mathcal{X}), E(f) = -E(-f) \quad \text{Duality} \quad (\text{A.4})$$

and, in probability theory, the two languages have the same expressiveness, i.e. it does not matter whether we speak in terms of probabilities of events or in terms of expected values of functions in $\mathcal{L}(\mathcal{X})$.

Similarly, a probability density p defined on the real line \mathbb{R} is totally defined by its cumulative distribution $F : \mathbb{R} \rightarrow [0, 1]$, defined as

$$\forall x \in \mathbb{R}, F(x) = \int_{-\infty}^x p(x) dx$$

provided F is right-continuous, or that we restrict ourselves to so-called σ -additive distributions (see Miranda *et al.* [149] for a exhaustive discussion of this topic).

In his work on subjective probability [108], de Finetti privileges the language of expected values, and use the terms *prevision* for $E(f)$. In his theory, the value $E(f)$ is associated to the *fair price* at which a given subject would buy or sell the game f to someone, such that $f(x)$ ¹ is returned if the value of variable X turns out to be x . In this theory, provided the subject can be forced at any moment to buy or sell the game at the given price, then De Finetti shows that for the subject to be coherent (i.e., not engage himself in a sure loss, or in a so-called Dutch-book), $E(f)$ must obey the laws of probability.

Although of major importance, many authors in the recent years have pinpointed arguments converging to the conclusion that probability masses cannot adequately account for all kind of uncertainty, and are likely to be too precise in a number of cases, eventually leading to make too strong commitments when making previsions and taking decisions.

There are many examples where uncertainty cannot be faithfully modeled by a precise

¹ $f(x)$ can be a gain or a loss

probability:

- confronted to a game $f(x)$, it seems natural that a subject could be allowed to give different prices at which he is ready to buy or sell it, allowing some imprecision in his expected gain.
- when eliciting information from an expert, this information rarely corresponds to a single possible probability distribution, and except for the will to model uncertainty by a single probability, there are no reasons to force one to choose one distribution rather than another.
- a similar argument holds when only few samples or observations are available.
- when building a joint uncertainty model from marginal ones, it is often unlikely that this joint model will be faithfully modeled by a joint probability mass, even if marginal models are themselves probabilities, simply because dependencies structures are usually not so well known.
- it is counterintuitive to take the same model for two (quite) different states of knowledge, namely ignorance and the fact that we know that everything is equi-probable.

for other arguments, see, for example, Walley [203]. Of course, there are numerous cases where usual probability theory will do the job. But there are also numerous cases in which it will not, and these are the cases that interest us.

A.2 Imprecise probability theory

Instead of considering single probabilities as model of uncertainty, imprecise probability theory takes closed convex sets of probabilities as their basic uncertainty models. Such sets are often called *credal sets* [136], and that is the name we will adopt here. Thus, uncertainty on X is characterized by a credal set \mathcal{P}_X which is a closed convex set of probability distributions on \mathcal{X}

This approach is very close to robust statistics [119] and share many similarities with it, but a great difference between the two approaches is that, in robust statistics, the existence of a precise but imprecisely known probability P_T is assumed, while in imprecise probability theory, credal sets are the basic uncertainty representation, and the existence of an underlying precise probability is not forcefully assumed.

Where, for an event $A \subseteq \mathcal{X}$, a precise probability induced a precise probability measure, a credal set \mathcal{P}_X induces lower $\underline{P}(A)$ and upper $\overline{P}(A)$ probability measures defined as:

$$\underline{P}(A) = \inf_{p \in \mathcal{P}_X} P(A) \text{ and } \overline{P}(A) = \sup_{p \in \mathcal{P}_X} P(A)$$

and the lower probability measure satisfy the following properties:

$$\forall A, B \subset \mathcal{X}, \underline{P}(A \cup B) \leq \underline{P}(A) + \underline{P}(B) \quad \text{Sub-Additivity} \quad (\text{A.5})$$

$$\forall A \subset \mathcal{X}, \underline{P}(A) = 1 - \overline{P}(A^c) \quad \text{Duality} \quad (\text{A.6})$$

and the difference $\overline{P}(A) - \underline{P}(A)$ reflects the imprecision of the available information.

Now, given a function f in $\mathcal{L}(\mathcal{X})$, a credal set \mathcal{P}_X also induces lower and upper expectations defined as

$$\underline{E}(f) = \inf_{p \in \mathcal{P}_X} E_p(f) \text{ and } \overline{E}(f) = \sup_{p \in \mathcal{P}_X} E_p(f)$$

with E_p the expectation of f given the probability mass p . And we have

$$\forall f, g \in \mathcal{L}(\mathcal{X}), E(f + g) \leq E(f) + E(g) \quad \text{Additivity} \quad (\text{A.7})$$

$$\forall f \in \mathcal{L}(\mathcal{X}), \underline{E}(f) = -\overline{E}(-f) \quad \text{Duality} \quad (\text{A.8})$$

Note that, thanks to the duality property, one can focus entirely on lower probabilities or expectations, since once they are defined respectively for all events or all functions in $\mathcal{L}(\mathcal{X})$, upper probabilities and upper expectations immediately follow. Similarly to the case of precise probabilities, lower and upper probabilities of an event A correspond to lower and upper expectations of the indicator function $\mathbf{1}_{(A)}$

But, contrary to precise probabilities, the two languages are not equivalent: any credal set can be characterized by its lower expectations of functions in $\mathcal{L}(\mathcal{X})$, but it is not, in general, possible to characterize a credal set only by its lower probability on events.

Inversely, we can start with functions g that are in a subset \mathcal{H} of $\mathcal{L}(\mathcal{X})$, and some lower bounds $L(g)$ of their corresponding expectation $E(g)$. We can then consider the set \mathcal{P}_L of probability masses $p \in \mathbb{P}_{\mathcal{X}}$ compatible with these lower bounds, that is

$$\mathcal{P}_L = \{p \in \mathbb{P}_{\mathcal{X}} \mid \forall g \in \mathcal{H}, L(g) \leq E_p(g)\}$$

lower bounds are called *consistent* if $\mathcal{P}_L \neq \emptyset$ and *coherent* or *tight* if we have $\underline{E}_{\mathcal{P}_L}(g) = L(g)$ for any $g \in \mathcal{H}$ (usually, we have only $\underline{E}_{\mathcal{P}_L}(g) \geq L(g)$).

In walley's [203] behavioral theory, consistency is called avoiding sure loss, tightness is also called coherence, and lower/upper expectations are called lower/upper previsions. Note that in this theory, and as recalled above, speaking in terms of expectations rather than probabilities is not merely a matter of choice, since credal sets generally cannot be characterized by their lower/upper probabilities alone. In the same line of thought as De Finetti [108], Walley considers that the lower prevision $\underline{E}(f)$ represents the maximal price at which a subject is ready to buy the game f , while the upper prevision $\overline{E}(f)$ represents the minimal price at which he is ready to sell it. When both prices coincide for every game on $\mathcal{L}(\mathcal{X})$, usual probabilities are retrieved.

A.3 Random set theory

Random set theory allows to model uncertainty of X by the means of a formal model that is a mapping $m : \wp(\mathcal{X}) \rightarrow [0, 1]$ from the power set of \mathcal{X} to the unit interval, such that

$$\forall A \subseteq \mathcal{X}, 0 \leq m(A) \tag{A.9}$$

$$\sum_{A \in \wp(\mathcal{X})} m(A) = 1 \tag{A.10}$$

i.e., the mapping m is normalized and non-negative, and we call it a *basic probability assignment* (bpa). Another name often found in the literature is the Dempster-Shafer model. It is common to call subsets receiving positive mass *focal elements*. In general, we will denote \mathcal{F} the set of focal elements, and (m, \mathcal{F}) a whole random set.

As a model of uncertainty coping with imprecision, random sets were first introduced by Dempster [57], who linked them to lower and upper probabilities generated by imprecise observations. The same formal model was then considered by Shafer [178] in his theory of evidence, which does not refer to an underlying precise probability distribution but deals with degree of beliefs, and in which random sets are called *belief functions*. This model was then extensively taken over by Smets [189] in its Transferable Belief Model (TBM). See Molchanov [151] for a recent theoretical account concerning random sets.

A.3.1 Shafer's belief functions and Smet's TBM

In Shafer's theory of evidence [178], the mass $m(E)$ attributed to E quantifies our belief that the actual value of X lies in the set E , and nowhere else. Now, given an event A , belief $Bel(A)$,

plausibility $Pl(A)$ and commonality $Q(A)$ measures are defined as:

$$Bel(A) = \sum_{E, E \subseteq A} m(E) \quad (\text{Belief}).$$

$$Pl(A) = 1 - Bel(A^c) = \sum_{E, E \cap A \neq \emptyset} m(E) \quad (\text{Plausibility}).$$

$$Q(A) = \sum_{E, E \supseteq A} m(E) \quad (\text{Commonality}).$$

Belief and plausibility measures respectively provide upper and lower confidence degree in the fact that event A will happen (or is true), since the belief measure quantifies the mass of belief that **surely** supports A , while plausibility measure quantifies the mass of belief that **could** supports A . Commonality measure quantifies the mass that could freely support any part of A . It can be argued that commonality measure gives an image of the imprecision of m , since the higher the masses given to broad subsets, the higher the commonality measure.

And it can be checked that the belief measure induced by a random set has the following property:

$$\forall n \in \mathbb{N} > 0, \forall \text{Collection } \mathcal{A} = \{A_1, \dots, A_n\}, A_i \subseteq \mathcal{X}, Bel\left(\bigcup_{A_i \in \mathcal{A}} A_i\right) \geq \sum_{I \subseteq \mathcal{A}} (-1)^{|I|+1} Bel\left(\bigcap_{A_i \in I} A_i\right)$$

Following Shafer's idea that m quantifies beliefs, Smets introduce the so-called Transferable Belief Model. The two main features of this model are that:

- It differentiate the *credal state*, where beliefs are held and possibly changed, from the *pignistic² state*, in which the so-called pignistic probability (see Appendix C) is used to make a decision based on beliefs
- It allows for an *open world* assumption, that it is a non-null mass $m(\emptyset) > 0$ can be given to the empty set, assuming that the true state of the world is possibly somewhere outside the considered universe \mathcal{X} .

A.3.2 Dempster's random sets

Let P_Y be a probability distribution on a space \mathcal{Y} . Then, Dempster interpret a random set as a so-called multi-valued mapping $\Gamma : \mathcal{Y} \rightarrow \wp(\mathcal{X})$ from the space \mathcal{Y} to another space \mathcal{X} . In

²From Latin Pignus, to decide

Dempster's view, Γ represents the imprecise observation on \mathcal{X} of some instance of a (random) variable Y assuming values in \mathcal{Y} and having P as probability measure.

For all event $A \subseteq \mathcal{X}$, we can define $A^* = \{y \in \mathcal{Y} | \Gamma(y) \cap A \neq \emptyset\}$ and $A_* = \{y \in \mathcal{Y} | \Gamma(y) \subseteq A\}$. For any such event A , we can then define lower P_* and upper P^* probabilities such that

$$P_*(A) = P_Y(A_*) \leq P(A) \leq P_Y^*(A) = P(A^*)$$

In Dempster's view, random sets are explicitly referring to the imprecise observation of some variable Y .

Dempster's random sets and Shafer's basic probability assignments are related in the following way: for any subset $E \subseteq \mathcal{X}$ let us define the mass $m(E)$ as

$$m_E = P_Y(\{y \in \mathcal{Y} | \Gamma(y) = E\})$$

then, the belief and plausibility measures derived from this bpa are equal to the lower and upper probabilities defined above. Nevertheless, an important potential difference is that, in Dempster's view, the mass $P_Y(y)$ is not forcefully allowed to distribute "freely" among elements of the set $\Gamma(y)$, and is usually constrained to be allocated to one and only one element of $\Gamma(y)$.

A.3.3 Random sets as credal sets

A bpa m and the associated belief measure Bel can also be related to a credal set \mathcal{P}_{Bel} , since the belief measure Bel can be viewed as a coherent lower probability, and we have

$$\mathcal{P}_{Bel} = \{P \in \mathbb{P}_{\mathcal{X}} | \forall A \subseteq \mathcal{X}, Bel(A) \leq P(A) \leq Pl(A)\}$$

And it can be proved that \mathcal{P}_{Bel} is the convex hull of all probability distributions compatible with a multi-valued mapping Γ inducing a lower probability $\underline{P} = Bel$. Nevertheless, if $P_Y(y)$ is constrained to be allocated to one and only one element of $\Gamma(y)$, there can be slight differences between the two models [146, 34] (but these differences are not relevant in our work). Also note that, in the case where the mass given to the empty set is non-null, then the resulting credal-set is empty (i.e. induced probabilistic bounds are not consistent)

Above interpretations and use of random set formalism are the most widely use in uncertainty treatment, and we refer to Smets [183] for a review of other interpretations.

A.4 Possibility theory

The first proposal to replace probability theory with a theory formally equivalent to possibility theory in order to deal with uncertainty probably dates back to Shackle [177] with the introduction of potential surprise distributions (equivalent to $1 - \pi$) in economy. It was later considered by Zadeh [219] and developed by Dubois and Prade [85].

The basic tool of possibility theory are possibility distributions. Given a variable X assuming values in \mathcal{X} , a possibility distribution is a mapping $\pi : \mathcal{X} \rightarrow [0, 1]$ from the space \mathcal{X} to the unit interval, quantifying the uncertainty about X .

Several set-functions can be defined from a possibility distribution π [79], namely the *possibility, necessity and sufficiency measures*:

$$\Pi(A) = \sup_{x \in A} \pi(x) \quad (\text{Possibility measures}).$$

$$N(A) = 1 - \Pi(A^c) \quad (\text{Necessity measures}).$$

$$\Delta(A) = \inf_{x \in A} \pi(x) \quad (\text{Sufficiency measures}).$$

The possibility degree of an event A evaluates the extent to which this event is plausible, i.e., consistent with the available. Necessity degrees express the certainty of events, by duality. In this context, distribution π is potential (in the spirit of Shackle's), i.e. $\pi(x) = 1$ does not guarantee the existence of x . Their characteristic property are: $N(A \cap B) = \min(N(A), N(B))$ and $\Pi(A \cup B) = \max(\Pi(A), \Pi(B))$ for any pair of events A, B of X .

On the contrary $\Delta(A)$ measures the extent to which all states of the world where A occurs are plausible. Sufficiency (or guaranteed possibility) distributions [79] generally denoted by δ , are understood as degree of empirical support and obey an opposite convention: $\delta(x) = 1$ guarantees (= is sufficient for) the existence of x .

It can be shown (already in [178]) that a necessity measure N is induced by a random set whose focal elements are nested, that is form a complete ordering with respect to inclusion. If we let $0 = \alpha_0 < \alpha_1 < \dots < \alpha_M$ be the distinct values of π on \mathcal{X} , then π is equivalent to the random set having, for $i = 1, \dots, M$, the following focal sets E_i with masses $m(E_i)$:

$$\left\{ \begin{array}{l} E_i = \{x \in \mathcal{X} \mid \pi(x) \geq \alpha_i\} = \\ m(E_i) = \alpha_i - \alpha_{i-1} \end{array} \right.$$

and the same amount of information is then contained in this random set and in the distribution $\pi(x) = Pl(\{x\})$. The *open-world* assumption then comes down to allow $\pi < 1$, and $m(\emptyset) = 1 - \sup_{x \in \mathcal{X}} \pi(x)$.

It is then easy to see that a possibility distribution π induce a particular credal set \mathcal{P}_π [92, 46], which is non-empty if and only if $\pi(x) = 1$ for at least one $x \in \mathcal{X}$. The credal set \mathcal{P}_π is defined as:

$$\mathcal{P}_\pi = \{P \in \mathbb{P}_{\mathcal{X}} \mid \forall A \subseteq \mathcal{X}, N(A) \leq P(A) \leq \Pi(A)\}$$

Appendix B

Some notions of order theory

In this appendix, we introduce few notions of order theory that are needed in this work. See Davey and Priestley [39] for an extended introduction to the subject.

Consider some space \mathcal{X} and a (binary) relation \leq on this space. Let us first define some properties that the binary relation can satisfy:

- The relation \leq is *reflexive* if, for any element $x \in \mathcal{X}$, we have

$$x \leq x \quad (\text{Reflexivity}) \quad (\text{B.1})$$

- The relation \leq is *antisymmetric* if, for any pair of elements $x, y \in \mathcal{X}$, we have

$$(x \leq y \text{ and } y \leq x) \Rightarrow x = y \quad (\text{Antisymmetry}) \quad (\text{B.2})$$

- The relation \leq is *transitive* if, for any triplet of elements $x, y, z \in \mathcal{X}$, we have

$$(x \leq y \text{ and } y \leq z) \Rightarrow x \leq z \quad (\text{Transitivity}) \quad (\text{B.3})$$

- The relation \leq is *complete* if, for any pair of elements $x, y \in \mathcal{X}$, we have

$$x \leq y \text{ or } y \leq x \quad (\text{Completeness}) \quad (\text{B.4})$$

Order relations are then characterized by the properties they satisfy. First, all of them satisfy the properties of reflexivity and of transitivity.

A relation that satisfy only reflexivity and transitivity is called a *partial pre-order*, also called *partial quasiorder*. Inside a *partial pre-order*, two elements are said to be:

- *incomparable* if $\neg(x \leq y \text{ or } y \leq x)$, with \neg the logical negation
- *equivalent* if $(x \leq y \text{ and } y \leq x)$ but $y \neq x$

A *complete pre-order*, also called *pre-order* or *quasiorder* is a relation satisfying the properties of reflexivity, transitivity and completeness. This means that every pairs of elements can be compared, but that there remain some elements that are judged *equivalent*.

A *partial order* is a relation satisfying the properties of reflexivity, transitivity and asymmetry. This means that there are non equivalent and distinct elements, but that some elements are *incomparable*.

Finally, an *linear order*, also called *total order*, *simple order*, or simply *chain* or *order*, is a relation satisfying the four properties of reflexivity, transitivity, completeness and asymmetry. The best known examples of this notion are probably the set of real numbers or of natural numbers equipped with the natural order on numbers (any number is comparable to another one, and if a number is both lower and upper than another one, then they are equal).

Given a *partial (pre-)order* \leq on \mathcal{X} , a *linear extension* \leq_L of \leq on \mathcal{X} is a total order such that, whenever $x \leq y$ for two elements $x, y \in \mathcal{X}$, it also holds that $x \leq_L y$. In other words, a linear extension is a linear order that is consistent with the original *partial (pre-)order* \leq , making incomparable elements comparable and arbitrating equivalences.

Appendix C

Random sets: inclusion, least commitment principle, pignistic probability

We recall here some results related to the random set formalism and to the TBM interpretation of this formalism. In particular, we recall the various notions of inclusion existing within this theory, and what is behind the so-called Least-commitment principle (LCP) and pignistic probability (*BetP*).

C.1 Inclusion relations between random sets

There exist many notions of inclusions between random sets, based on different measures and/or notions. here, we recall the principal ones, and we refer to Denoeux for additional notions. Each inclusion notion gives rise to a corresponding partial order between random sets. The pl -, q - and s -orderings were introduced by Dubois and Prade [84], while Denoeux [61] recently introduced other orderings (w - and v -orderings) based on Smets [184] canonical decomposition.

Let us first recall the notions of specialization, as well as Smets canonical decomposition of belief functions.

Consider an arbitrary indexing of subsets E_i of $\wp(\mathcal{X})$, $i = 1, \dots, |\wp(\mathcal{X})|$. Given two random sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$ defined on \mathcal{X} , let $\mathbf{m}_1, \mathbf{m}_2$ be $|\wp(\mathcal{X})| \times 1$ vectors of weights, where the i^{th} element of \mathbf{m}_j is the mass $m(E_i)$ of subset E_i in $(m, \mathcal{F})_j$. Then, $(m, \mathcal{F})_2$ is a specialization of $(m, \mathcal{F})_1$ if, given the $\mathbf{m}_1, \mathbf{m}_2$, there exist a $\wp(\mathcal{X}) \times \wp(\mathcal{X})$ stochastic matrix

S such that

$$\mathbf{m}_2 = S \cdot \mathbf{m}_1$$

with S_{ij} the element in the i^{th} line and j^{th} column of S , and $S_{ij} > 0$ if and only if $E_i \subset E_j$. In short, $(m, \mathcal{F})_2$ is a specialization of $(m, \mathcal{F})_1$ if the masses of $(m, \mathcal{F})_1$ "flow downs" to subsets of $(m, \mathcal{F})_2$ (i.e., $m_1(E)$ is reallocated among subsets of E in m_2).

Let (m, \mathcal{F}) be a random set such that $m(\mathcal{X}) > 0$ (so-called non-dogmatic bpa). Then, Smets canonical decomposition [184] consists in affecting to every subset $A \subseteq \mathcal{X}$ a weight $w(A)$ with $w \in [0, \infty)$ and such that w can be obtained from commonality measure in the following way:

$$w(A) = \prod_{B \supseteq A} Q(B)^{(-1)^{|B \setminus A|+1}}$$

We can now define the following partial orders based on different extensions of set inclusions:

- *pl-ordering*: $(m, \mathcal{F})_2 \sqsubseteq_{pl} (m, \mathcal{F})_1$ if and only if, for all subset $E \subseteq \mathcal{X}$, $Pl_{(m, \mathcal{F})_2}(E) \leq Pl_{(m, \mathcal{F})_1}(E)$
- *q-ordering*: $(m, \mathcal{F})_2 \sqsubseteq_Q (m, \mathcal{F})_1$ if and only if, for all subset $E \subseteq \mathcal{X}$, $Q_{(m, \mathcal{F})_2}(E) \leq Q_{(m, \mathcal{F})_1}(E)$
- *s-ordering*: $(m, \mathcal{F})_2 \sqsubseteq_s (m, \mathcal{F})_1$ if and only if $(m, \mathcal{F})_2$ is a specialization of $(m, \mathcal{F})_1$
- *w-ordering*: $(m, \mathcal{F})_2 \sqsubseteq_w (m, \mathcal{F})_1$ if and only if, for all subset $E \subseteq \mathcal{X}$, $Q_{(m, \mathcal{F})_2}(E) \leq Q_{(m, \mathcal{F})_1}(E)$

and each relation \sqsubseteq_x with $x \in \{w, s, pl, q\}$ induce a partial order on random sets. Also note that $(m, \mathcal{F})_2 \sqsubseteq_{pl} (m, \mathcal{F})_1$ is equivalent to $\mathcal{P}_{(m, \mathcal{F})_2} \subseteq \mathcal{P}_{(m, \mathcal{F})_1}$, with $\mathcal{P}_{(m, \mathcal{F})_i}$ the credal set induced by $(m, \mathcal{F})_i$. Note that some inclusion notions are stronger than others, and, given two random sets $(m, \mathcal{F})_1, (m, \mathcal{F})_2$, we have the following relations

$$(m, \mathcal{F})_2 \sqsubseteq_w (m, \mathcal{F})_1 \Rightarrow (m, \mathcal{F})_2 \sqsubseteq_s (m, \mathcal{F})_1 \Rightarrow \begin{cases} (m, \mathcal{F})_2 \sqsubseteq_{pl} (m, \mathcal{F})_1 \\ (m, \mathcal{F})_2 \sqsubseteq_Q (m, \mathcal{F})_1 \end{cases}$$

Denoeux [61] (to which we refer for ampler discussion) considers additional orderings tagged by the letters v, d, dd and they complete the above picture in the following way:

$$\left. \begin{array}{l} (m, \mathcal{F})_2 \sqsubseteq_v (m, \mathcal{F})_1 \Rightarrow (m, \mathcal{F})_2 \sqsubseteq_{dd} (m, \mathcal{F})_1 \\ (m, \mathcal{F})_2 \sqsubseteq_w (m, \mathcal{F})_1 \Rightarrow (m, \mathcal{F})_2 \sqsubseteq_d (m, \mathcal{F})_1 \end{array} \right\} \Rightarrow (m, \mathcal{F})_2 \sqsubseteq_s (m, \mathcal{F})_1 \Rightarrow \begin{cases} (m, \mathcal{F})_2 \sqsubseteq_{pl} (m, \mathcal{F})_1 \\ (m, \mathcal{F})_2 \sqsubseteq_Q (m, \mathcal{F})_1 \end{cases}$$

C.2 Least-commitment principle (LCP)

The least-commitment principle can be informally stated as the motto "one should never presuppose more belief than justified by evidence".

In terms of random sets, it is translated by the fact that, when a set of constraints do not allow to identify a unique random set, but rather a set \mathcal{M} of compatible random sets, then one should select the least-committed random set with respect to one of the ordering $\sqsubseteq_x, x \in \{w, s, pl, q\}$.

In general, there are multiple least-committed random sets, since the above orderings are partial. Note that, the stronger the ordering notion, the more incompatibilities it generates, and the larger the set of potential least-committed random sets.

C.3 Pignistic probability $BetP$

As briefly recalled in Appendix A, the Transferable Belief Model (TBM) has two levels: a credal one, in which beliefs are entertained, and a pignistic one, in which a decision is taken based on beliefs. Based on a set of rational requirements [187], Smets justify the use of the so-called pignistic probability to determine this decision. The pignistic probability $BetP$ is defined as follow:

Definition C.1. Let (m, \mathcal{F}) be a random set defined on \mathcal{X} . The pignistic probability of an element $x \in \mathcal{X}$ is then defined as:

$$BetP(x) = \sum_{E \in \mathcal{F}, x \in E} \frac{1}{|E|} \frac{m(E)}{1 - m(\emptyset)}$$

And it can be checked that $BetP(x)$ is a probability mass on \mathcal{X} . It comes down to distribute, for each focal element $E, m(E)$ uniformly among elements of E and to normalize the

obtained distribution. Actually, it is the probability mass equivalent to the gravity center of the credal set $\mathcal{P}_{(m, \mathcal{F})}$ induced by the corresponding normalized random set, and it is also equivalent to the Shapley value [181] in game theory.

Appendix D

Proofs

This appendix contains longer proofs not essentials to the understanding of te whole manuscript.

D.1 Proofs of Section 3.2

Proof of Proposition 3.4. From the nested sets $A_1 \subseteq A_2 \subseteq \dots \subseteq A_n = X$ we can build a partition s.t. $G_1 = A_1, G_2 = A_2 \setminus A_1, \dots, G_n = A_n \setminus A_{n-1}$. Once we have a finite partition, every possible set $B \subseteq X$ can be approximated from above and from below by pairs of sets $B_* \subseteq B^*$ [165]:

$$B^* = \bigcup \{G_i, G_i \cap B \neq \emptyset\}$$
$$B_* = \bigcup \{G_i, G_i \subseteq B\}$$

made of a finite union of the partition elements intersecting or contained in this set B . Then $\underline{P}(B) = \underline{P}(B_*), \bar{P}(B) = \bar{P}(B^*)$, so we only have to care about unions of elements G_i in the sequel. Especially, for each event $B \subset G_i$ for some i , it is clear that $\underline{P}(B) = 0 = Bel(B)$ and $\bar{P}(B) = \bar{P}(G_i) = Pl(B)$. So, to prove Proposition 3.4, we have to show that lower probabilities given by a generalized p-box $[\underline{F}, \bar{F}]$ and by the corresponding random set built through algorithm 3 coincide on every possible union of elements G_i . We will first concentrate on unions of conscutive elements G_i , and then to any union of such elements.

Union of consecutive elements G_i Let us first consider union of consecutive elements $\bigcup_{k=i}^j G_k$ (when $k = 1$, we retrieve the sets A_j). Finding $\underline{P}(\bigcup_{k=i}^j G_k)$ is equivalent to computing

the minimum of $\sum_{k=i}^j P(G_k)$ under the constraints

$$i = 1, \dots, n \quad \alpha_i \leq P(A_i) = \sum_{k=1}^i P(G_k) \leq \beta_i \quad (\text{D.1})$$

which reads

$$\alpha_j \leq P(A_{i-1}) + \sum_{k=i}^j P(G_k) \leq \beta_j \quad (\text{D.2})$$

so $\sum_{k=i}^j P(G_k) \geq \max(0, \alpha_j - \beta_{i-1})$. This lower bound is optimal, since it is always reachable:

- if $\alpha_j > \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}$, $P(\bigcup_{k=i}^j G_k) = \alpha_j - \beta_{i-1}$, $P(\bigcup_{k=j+1}^n G_k) = 1 - \alpha_j$.
- If $\alpha_j \leq \beta_{i-1}$, take P s.t. $P(A_{i-1}) = \beta_{i-1}$, $P(\bigcup_{k=i}^j G_k) = 0$, $P(\bigcup_{k=j+1}^n G_k) = 1 - \beta_{i-1}$.

And we can see, by looking at Algorithm 3, that $Bel(\bigcup_{k=i}^j G_k) = \max(0, \alpha_j - \beta_{i-1})$: focal elements of Bel are subsets of $\bigcup_{k=i}^j G_k$ if $\beta_{i-1} < \alpha_j$ only.

Union of non-consecutive elements Now, let us consider a union A of non-consecutive elements s.t. $A = (\bigcup_{k=i}^{i+l} G_k \cup \bigcup_{k=i+l+m}^j G_k)$ with $m > 1$. As in the previous case, we must compute $\min\left(\sum_{k=i}^{i+l} P(G_k) + \sum_{k=i+l+m}^j P(G_k)\right)$ to find the lower probability on $\underline{P}(A)$. An obvious lower bound is given by

$$\min\left(\sum_{k=i}^{i+l} P(G_k) + \sum_{k=i+l+m}^j P(G_k)\right) \geq \min\left(\sum_{k=i}^{i+l} P(G_k)\right) + \min\left(\sum_{k=i+l+m}^j P(G_k)\right) \quad (\text{D.3})$$

and, by the result obtained for consecutive elements, this lower bound is equal to

$$\max(0, \alpha_{i+l} - \beta_{i-1}) + \max(0, \alpha_j - \beta_{i+l+m-1}) = Bel(A) \quad (\text{D.4})$$

Consider the two following cases and the probability assignments showing that bounds are attained:

- $\alpha_{i+l} < \beta_{i-1}$, $\alpha_j < \beta_{i+l+m-1}$ and probability masses $P(A_{i-1}) = \beta_{i-1}$,
 $P(\bigcup_{k=i}^{i+l} G_k) = \alpha_{i+l} - \beta_{i-1}$, $P(\bigcup_{k=i+l+m}^j G_k) = \beta_{i+l+m-1} - \alpha_{i+l}$,
 $P(\bigcup_{k=i+l+m}^j G_k) = \alpha_j - \beta_{i+l+m-1}$ and $P(\bigcup_{k=j+1}^n G_k) = 1 - \alpha_j$
- $\alpha_{i+l} > \beta_{i-1}$, $\alpha_j > \beta_{i+l+m-1}$ and probability masses $P(A_{i-1}) = \beta_{i-1}$, $P(\bigcup_{k=i}^{i+l} G_k) = 0$,
 $P(\bigcup_{k=i+l+m}^j G_k) = \alpha_j - \beta_{i-1}$, $P(\bigcup_{k=i+l+m}^j E_k) = 0$ and $P(\bigcup_{k=j+1}^n G_k) = 1 - \alpha_j$

A same line of thought can be followed for the two remaining cases. As in the consecutive case, the lower bound is reachable without violating any of the restrictions associated to the generalized p-box. We have $\underline{P}(A) = Bel(A)$ and the extension of this result to any number n of "discontinuities" in the sequence of G_k is straightforward.

The proof is complete, since for every possible union A of elements G_k , we have $\underline{P}(A) = Bel(A)$ □

Proof of Proposition 3.5. Let \mathcal{X} be a finite set and define a ranking of their elements $x_i < x_j$ if and only if $i < j$. Given this ranking, and to prove Proposition 3.5, we start from a set L with, for $i = 1, \dots, n$, initial bounds u_i, l_i . We then apply successively Equations (3.27) and (3.25), with the aim of expressing bounds u'_i, l'_i of the set L'' in terms of initial bounds u_i, l_i . Expressions for $(l_i - l'_i)$ and $(u'_i - u_i)$ then follows. The positiveness of these two differences is sufficient to prove inclusion between credal sets \mathcal{P}_L and $\mathcal{P}_{L''}$. To shorten the proof, we focus on lower bounds (proof for upper bounds is similar).

Let us consider the p-box $[\underline{F}, \overline{F}]'$ built from a given reachable non-empty set L of probability intervals, given, for $i = 1, \dots, n$, by equations

$$\begin{aligned} \underline{P}(A_i) &= \alpha'_i = \max\left(\sum_{x_j \in A_i} l_j, 1 - \sum_{x_j \notin A_i} u_j\right) \\ \overline{P}(A_i) &= \beta'_i = \min\left(\sum_{x_j \in A_i} u_j, 1 - \sum_{x_j \notin A_i} l_j\right) \end{aligned}$$

with $\underline{P}, \overline{P}$ the lower and upper probabilities \mathcal{P}_L . Now, given these bounds, we can compute the set L'' of probability intervals s.t.

$$l''_i = \underline{P}'(x_i) = \max(0, \alpha'_i - \beta'_{i-1}) \tag{D.5}$$

with \underline{P}' the lower probability of $\mathcal{P}_{[\underline{F}, \overline{F}]}'$. When expressed in term of values l_i, u_i of the original set L , l''_i is given by

$$\begin{aligned} l''_i &= \max\left(0, \sum_{x_j \in A_i} l_j - \sum_{x_j \in A_{i-1}} u_j, \sum_{x_j \in A_i} l_j + \sum_{x_j \in A_{i-1}^c} l_j - 1, \right. \\ &\quad \left. 1 - \sum_{x_j \in A_i^c} u_j - \sum_{x_j \in A_{i-1}} u_j, \sum_{x_j \in A_{i-1}^c} l_j - \sum_{x_j \in A_i^c} u_j\right) \end{aligned}$$

and, given that the set L is consistent (Equation 3.8) and tight (Equations 3.9), we have that

$l_i'' \leq l_i$. To get Equation (3.27) giving $(l_i - l_i'')$, simply note that:

$$\begin{aligned} l_i - \sum_{x_j \in A_i} l_j &= - \sum_{x_j \in A_{i-1}} l_j \\ A_i^c \cup A_{i-1} &= \mathcal{X} \setminus x_i \\ l_i - \sum_{x_j \in A_{i-1}^c} l_j &= - \sum_{x_j \in A_i^c} l_j \end{aligned}$$

The same procedure can be followed for the bounds u_i'' , and we have $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$. The set L'' is tight (since $\mathcal{P}_L \subseteq \mathcal{P}_{L''}$) and consistent (by construction, the new bounds $[l_i'', u_i'']$ are reached by one distribution in the p-box $[\underline{F}, \overline{F}]$, and this distribution is also in $\mathcal{P}_{L''}$, thus set L'' is tight, or reachable) \square

Proof of Proposition 3.6. Proof of proposition 3.6 follows the same line of thought as the proof of Proposition 3.5.

Let us consider an original generalized p-box $[\underline{F}, \overline{F}]$ with bounds α_i, β_i on sets A_i . The set L' of probability intervals corresponding to this generalized p-box is given by equations

$$\begin{aligned} \underline{P}(x_i) &= l_i' = \max(0, \alpha_i - \beta_{i-1}) \\ \overline{P}(x_i) &= u_i' = \beta_i - \alpha_{i-1}, \end{aligned}$$

where $\underline{P}, \overline{P}$ are the lower and upper probabilities of $\mathcal{P}_{[\underline{F}, \overline{F}]}$. From the set L' , we can get the lower bound \underline{F}'' of $[\underline{F}, \overline{F}]''$ by using equations

$$\underline{P}'(A_i) = \alpha_i'' = \max\left(\sum_{x_i \in A_i} l_i', 1 - \sum_{x_i \notin A_i} u_i'\right)$$

with \underline{P}' the lower probability of $\mathcal{P}_{L'}$. In terms of the original p-box bounds α_i, β_i , this gives us

$$\begin{aligned} \alpha_i'' &= \max\left(\sum_{j=1}^i \alpha_j - \sum_{j=1}^{i-1} \beta_j, 1 + \sum_{j=i}^{n-1} \alpha_j - \sum_{j=i+1}^n \beta_j\right) \\ \alpha_i'' &= \max\left(\sum_{j=1}^i \alpha_j - \sum_{j=1}^{i-1} \beta_j, \alpha_i + \sum_{j=i+1}^{n-1} \alpha_j - \sum_{j=i+1}^{n-1} \beta_j\right) \end{aligned}$$

Given that $\forall j, \alpha_j \leq \beta_j$ by definition of a generalized p-box, we have $\alpha_i'' \leq \alpha_i$ and Equation (3.29) follows. The same procedure can again be done for the upper bound to check that $\beta_i'' \geq \beta_i$, and we get $\mathcal{P}_{[E,\bar{F}]} \subseteq \mathcal{P}_{[E,\bar{F}]''}$. \square

Proof of Proposition 3.7. To prove this proposition, we must first recall a result given by De Campos *et al.* [42]: given two sets of probability intervals L and L' defined on a space \mathcal{X} and the induced credal sets \mathcal{P}_L and $\mathcal{P}_{L'}$, the conjunction $\mathcal{P}_{L \cap L'} = \mathcal{P}_L \cap \mathcal{P}_{L'}$ of these two sets can be modeled by the set $(L \cap L')$ of probability intervals that is such that for every element x of \mathcal{X} ,

$$l_{(L \cap L')}(x) = \max(l_L(x), l_{L'}(x)) \text{ and } u_{(L \cap L')}(x) = \min(u_L(x), u_{L'}(x))$$

and these formulas extend directly to the conjunction of any number of set of probability intervals on \mathcal{X} (due to the associativity and commutativity of operators max and min).

To prove Proposition 3.7, we will show, by using the above conjunction, that $\mathcal{P}_L = \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$. Since, by Proposition 3.5 and for any $\sigma \in \Sigma_\sigma$, $\mathcal{P}_L \subseteq \mathcal{P}_{[E,\bar{F}]'_\sigma} \subseteq \mathcal{P}_{L''_\sigma}$, showing this equality is sufficient to prove the whole proposition.

Let us note that the above inclusion relationships alone ensure us that $\mathcal{P}_L \subseteq \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{[E,\bar{F}]'_\sigma} \subseteq \bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$. So, all we have to show is that the inclusion relationship is in fact an equality.

Since we know that both \mathcal{P}_L and $\bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$ can be modeled by set of probability intervals, we will show that the lower bounds l on every element x in these two sets coincide (and the proof for upper bounds is similar).

For all x in \mathcal{X} , $l_{L''_\Sigma}(x) = \max_{\sigma \in \Sigma_\sigma} \{l_{L''_\sigma}(x)\}$, with L''_Σ the set of probability intervals corresponding to $\bigcap_{\sigma \in \Sigma_\sigma} \mathcal{P}_{L''_\sigma}$ and L''_σ the set of probability intervals corresponding to a particular permutation σ . We must now show that, for all x in \mathcal{X} , $l_{L''_\Sigma}(x) = l_L(x)$.

From Proposition 3.7, we already know that, for any permutation σ and for all x in \mathcal{X} , we have $l_{L''_\sigma}(x) \leq l_L(x)$. So we must now show that, for a given x in \mathcal{X} , there is one permutation σ such that $l_{L''_\sigma}(x) = l_L(x)$. Let us consider a permutation placing the given element at the front. If x is the first element $x_{\sigma(1)}$, then Equation (3.27) has value 0 for this element, and we thus have $l_{L''_\sigma}(x) = l_L(x)$. Since if we consider every possible ranking, every element x of \mathcal{X} will be first in at least one of these rankings, this completes the proof. \square

D.2 Proofs of Section 3.3

To proof Proposition 3.13, we first state a short Lemma allowing us to emphasize the mechanism behind the proof of the latter proposition.

Lemma D.1. *Let $(F_1, F_2), (G_1, G_2)$ be two pairs of sets such that $F_1 \subset F_2$, $G_1 \subset G_2$, $G_1 \not\subset F_2$ and $G_1 \cap F_1 \neq \emptyset$. Let also π_F, π_G be two possibility distributions such that the corresponding random sets are defined by mass assignments $m_F(F_1) = m_G(G_2) = \lambda$, $m_F(F_2) = m_G(G_1) = 1 - \lambda$. Then, the lower probability of the non-empty credal set $\mathcal{P} = \mathcal{P}_{\pi_F} \cap \mathcal{P}_{\pi_G}$ is not 2-monotone.*

Note that in the above lemma, $[1 - \pi_G, \pi_F]$ is not a cloud, since the inequality $\pi_G + \pi_F \geq 1$ does not hold, even if by construction, $\mathcal{P} = \mathcal{P}_{\pi_F} \cap \mathcal{P}_{\pi_G}$ is not empty. Non-emptiness of $\mathcal{P} = \mathcal{P}_{\pi_F} \cap \mathcal{P}_{\pi_G}$ comes from $\pi_F(x) = \pi_G(x) = 1$ for an element $x \in G_1 \cap F_1$, thus $\min(\pi_G, \pi_F)$ is normalized (see Section 3.3.2.2). Example 3.8 and Proposition 3.13 shows that this situation described in Lemma D.1 also occurs in non-comonotonic clouds.

Proof of Lemma D.1. To prove Lemma D.1, we first recall a useful result by Chateauneuf [21] concerning the intersection of credal sets induced by random sets. This result is then applied to the possibility distributions defined in Lemma D.1 to prove that the associated lower probability is not 2-monotone. The main idea is to exhibit two events such that 2-monotonicity is not satisfied for them. Consider the set \mathcal{M} of matrices M of the form

$$\begin{array}{|c|c|c|}
 \hline
 & G_1 & G_2 \\
 \hline
 F_1 & m_{11} & m_{12} \\
 \hline
 F_2 & m_{21} & m_{22} \\
 \hline
 \end{array} \tag{D.6}$$

where

$$\begin{aligned}
 m_{11} + m_{12} &= m_{22} + m_{12} = \lambda \\
 m_{21} + m_{22} &= m_{21} + m_{11} = 1 - \lambda \\
 \sum m_{ij} &= 1
 \end{aligned}$$

Each such matrix is a normalized (i.e. such that $m(\emptyset) = 0$) joint mass distribution for the random sets induced from possibility distributions π_F, π_G , viewed as marginal belief functions.

Following Chateauneuf [21], the lower probability \underline{P} induced by the credal set $\mathcal{P} = \mathcal{P}_{\pi_F} \cap \mathcal{P}_{\pi_G}$ has, for any event $E \subseteq \mathcal{X}$, value

$$\underline{P}(E) = \min_{M \in \mathcal{M}} \sum_{(F_i \cap G_j) \subset E} m_{ij} \quad (\text{D.7})$$

Now consider the four events $F_1, G_1, F_1 \cap G_1, F_1 \cup G_1$. Studying the relations between sets and the constraints on the values m_{ij} , we can see that

$$\begin{aligned} \underline{P}(F_1) &= \lambda \\ \underline{P}(G_1) &= 1 - \lambda \\ \underline{P}(F_1 \cap G_1) &= 0. \end{aligned}$$

For $F_1 \cap G_1$, just consider the matrix $m_{12} = \lambda, m_{21} = 1 - \lambda$. To show that the lower probability is not even 2-monotone, it is enough to show that $\underline{P}(F_1 \cup G_1) < 1$. To achieve this, consider the following mass distribution

$$\begin{aligned} m_{11} &= \min(\lambda, 1 - \lambda) \\ m_{12} &= \lambda - m_{11} \\ m_{21} &= 1 - \lambda - m_{11} \\ m_{22} &= \min(\lambda, 1 - \lambda) \end{aligned}$$

it can be checked that the matrix corresponding to this distribution is in the set \mathcal{M} , and yields

$$\begin{aligned} P(F_1 \cup G_1) &= m_{12} + m_{11} + m_{21} \\ &= m_{11} + \lambda - m_{11} + 1 - \lambda - m_{11} \\ &= 1 - m_{11} = 1 - \min(\lambda, 1 - \lambda) \\ &= \max(1 - \lambda, \lambda) < 1 \end{aligned}$$

since $(F_2 \cap G_2) \not\subseteq (F_1 \cup G_1)$ (due to the fact that $G_1 \not\subseteq F_2$). Then the inequality

$$\underline{P}(F_1 \cup G_1) + \underline{P}(F_1 \cap G_1) < \underline{P}(F_1) + \underline{P}(G_1) \quad (\text{D.8})$$

holds, which ends the proof. □

Proof of Proposition 3.13. To prove Proposition 3.13, we again use the result by Chateauneuf [21] as in the proof of Lemma D.1, that is we exhibit a pair of events for which 2-monotonicity fails. Chateauneuf results are clearly applicable to clouds, since possibility distributions are equivalent to nested random sets. Consider a finite cloud described by the general Equation (3.35) and the following matrix Q of weights q_{ij}

| | $C_{\gamma_1}^c$ | \cdots | $C_{\gamma_j}^c$ | \cdot | $C_{\gamma_{i+1}}^c$ | \cdots | $C_{\gamma_M}^c$ |
|-------------------------------|------------------|----------|-----------------------|---------|---------------------------|----------|------------------|
| $B_{\overline{\gamma_0}}$ | q_{11} | \cdots | q_{1j} | \cdot | $q_{1(i+1)}$ | \cdots | q_{1M} |
| \vdots | \vdots | \ddots | | | \vdots | | \vdots |
| $B_{\overline{\gamma_{j-1}}}$ | q_{j1} | \cdots | \mathbf{q}_{jj} | \cdot | $\mathbf{q}_{j(i+1)}$ | \cdots | q_{jM} |
| \vdots | \vdots | \vdots | \vdots | | \vdots | \ddots | \vdots |
| $B_{\overline{\gamma_i}}$ | $q_{(i+1)1}$ | \cdots | $\mathbf{q}_{(i+1)j}$ | \cdot | $\mathbf{q}_{(i+1)(i+1)}$ | \cdots | $q_{(i+1)M}$ |
| \vdots | \vdots | \vdots | \vdots | | \vdots | \ddots | \vdots |
| $B_{\overline{\gamma_{M-1}}}$ | q_{M1} | \cdots | q_{Mj} | \cdot | $q_{M(i+1)}$ | \cdots | q_{MM} |

Respectively call Bel_1 and Bel_2 the belief functions equivalent to the possibility distributions respectively generated by the collections of sets $\{B_{\overline{\gamma_i}} | i = 0, \dots, M-1\}$ and $\{C_{\gamma_j}^c | j = 1, \dots, M\}$. From Equation (3.17), $m_1(B_{\overline{\gamma_i}}) = \gamma_{i+1} - \gamma_i$ for $i = 0, \dots, m-1$, and $m_2(C_{\gamma_j}^c) = \gamma_j - \gamma_{j-1}$ for $j = 1, \dots, M$. As in the proof of Lemma D.1, we consider every possible joint random set such that $m(\emptyset) = 0$ built from the two marginal belief functions Bel_1, Bel_2 .

Following Chateauneuf, let \mathcal{Q} be the set of matrices Q s.t.

$$q_{i\cdot} = \sum_{j=1}^M q_{ij} = \gamma_i - \gamma_{i-1}$$

$$q_{\cdot j} = \sum_{i=1}^M q_{ij} = \gamma_j - \gamma_{j-1}$$

$$\text{If } i, j \text{ s.t. } B_{\overline{\gamma_i}} \cap C_{\gamma_j}^c = \emptyset \text{ then } q_{ij} = 0$$

and the lower probability of the credal set $\mathcal{P}_{[\pi, \delta]}$ on event E is such that

$$\underline{P}(E) = \min_{Q \in \mathcal{Q}} \sum_{(B_{\overline{\gamma_i}} \cap C_{\gamma_j}^c) \subset E} q_{ij}. \quad (\text{D.9})$$

Now, by hypothesis, there are at least two overlapping sets $B_{\bar{\gamma}_i}, C_{\gamma_j}$ $i > j$ that are not included in each other (i.e. $B_{\bar{\gamma}_i} \cap C_{\gamma_j} \notin \{B_{\bar{\gamma}_i}, C_{\gamma_j}, \emptyset\}$). Let us consider the four events $B_{\bar{\gamma}_i}, C_{\gamma_j}^c, B_{\bar{\gamma}_i} \cap C_{\gamma_j}^c, B_{\bar{\gamma}_i} \cup C_{\gamma_j}^c$. Considering Equation (D.9), the matrix Q and the relations between sets, inclusions $B_{\bar{\gamma}_M} \subset \dots \subset B_{\bar{\gamma}_0}, C_{\gamma_0}^c \subset \dots \subset C_{\gamma_M}^c$ and, for $i = 0, \dots, M, C_{\gamma_i} \subset B_{\bar{\gamma}_i}$ imply:

$$\begin{aligned} \underline{P}(B_{\bar{\gamma}_i}) &= 1 - \gamma_i \\ \underline{P}(C_{\gamma_j}^c) &= \gamma_j \\ \underline{P}(B_{\bar{\gamma}_i} \cap C_{\gamma_j}^c) &= 0 \end{aligned}$$

for the last result, just consider the mass distribution $q_{kk} = \gamma_{k-1} - \gamma_k$ for $k = 1, \dots, M$.

Next, consider event $B_{\bar{\gamma}_i} \cup C_{\gamma_j}^c$ (which is different from X by hypothesis). Suppose all masses are such that $q_{kk} = \gamma_{k-1} - \gamma_k$, except for masses (in boldface in the matrix) $q_{jj}, q_{(i+1)(i+1)}$. Then, $C_{\gamma_j}^c \subset C_{\gamma_{i+1}}^c, B_{\bar{\gamma}_i} \subset B_{\bar{\gamma}_{j-1}}, C_{\gamma_j}^c \not\subset B_{\bar{\gamma}_{j-1}}$ by definition of a cloud and $B_{\bar{\gamma}_i} \cap C_{\gamma_j}^c \neq \emptyset$ by hypothesis. Finally, using Lemma D.1, consider the mass distribution

$$\begin{aligned} q_{(i+1)j} &= \min(\gamma_{i+1} - \gamma_i, \gamma_j - \gamma_{j-1}) \\ q_{(i+1)(i+1)} &= \gamma_{i+1} - \gamma_i - q_{(i+1)j} \\ q_{jj} &= \gamma_j - \gamma_{j-1} - q_{(i+1)j} \\ q_{j(i+1)} &= \min(\gamma_{i+1} - \gamma_i, \gamma_j - \gamma_{j-1}) \end{aligned}$$

It always gives a matrix in the set \mathcal{Q} . By considering every subset of $B_{\bar{\gamma}_i} \cup C_{\gamma_j}^c$, we thus get the following inequality

$$\underline{P}(B_{\bar{\gamma}_i} \cup C_{\gamma_j}^c) \leq \gamma_{j-1} + 1 - \gamma_{i+1} + \max(\gamma_{i+1} - \gamma_i, \gamma_j - \gamma_{j-1}). \quad (\text{D.10})$$

And, similarly to what was found in Lemma D.1, we get

$$\underline{P}(B_{\bar{\gamma}_i} \cup C_{\gamma_j}^c) + \underline{P}(B_{\bar{\gamma}_i} \cap C_{\gamma_j}^c) < \underline{P}(B_{\bar{\gamma}_i}) + \underline{P}(C_{\gamma_j}^c), \quad (\text{D.11})$$

which shows that the lower probability is not 2-monotone. □

Proof of Proposition 3.14. First, we know that the random set given in Proposition 3.14 is

equivalent to

$$\begin{cases} E_j = B_{\overline{\gamma_{j-1}}} \setminus C_{\gamma_j} = B_{\gamma_j} \setminus C_{\gamma_j} \\ m(E_j) = \gamma_j - \gamma_{j-1} \end{cases} \quad (\text{D.12})$$

Now, if we consider the matrix given in the proof of Proposition 3.13, this random set comes down, for $i = 1, \dots, M$ to assign masses $q_{ii} = \gamma_i - \gamma_{i-1}$. Since this is a legal assignment, we are sure that for all events $E \subseteq \mathcal{X}$, the belief function of this random set is such that $Bel(E) \geq \underline{P}(E)$, where \underline{P} is the lower probability induced by the cloud. The proof of Proposition 3.13 shows that this inclusion is strict for clouds satisfying the latter proposition (since the lower probability induced by such clouds is not 2-monotone). \square

Proof of Proposition 3.16. We build a sequence of outer and inner approximations of the continuous random set that converge to the belief measure of the continuous random set, while the corresponding clouds of which they are inner approximations themselves converge to the uniformly continuous cloud.

First, consider a finite collection of equidistant levels α_i s.t. $0 = \alpha_0 < \alpha_1 < \dots < \alpha_n = 1$ ($\alpha_{i-1} - \alpha_i = 1/n \forall i = 1, \dots, n$). Then, consider the following discrete non-comonotonic clouds $[\underline{\delta}_n, \underline{\pi}_n]$, $[\overline{\delta}_n, \overline{\pi}_n]$ that are respectively outer and inner approximations of the cloud $[\pi, \delta]$: for every value r in \mathbb{R} , do the following transformation

$$\begin{aligned} \pi(r) = \alpha \text{ with } \alpha \in [\alpha_{i-1}, \alpha_i] \text{ then } \underline{\pi}_n(r) = \alpha_i \text{ and } \overline{\pi}_n(r) = \alpha_{i-1} \\ \delta(r) = \alpha' \text{ with } \alpha' \in [\alpha_{j-1}, \alpha_j] \text{ then } \underline{\delta}_n(r) = \alpha_{j-1} \text{ and } \overline{\delta}_n(r) = \alpha_j \end{aligned}$$

This construction is illustrated in Figure D.1 for the particular case when both π and δ are unimodal. In this particular case, for $i = 1, \dots, n$

$$\begin{aligned} \{x \in \mathbb{R} | \underline{\pi}(x) \geq \alpha\} &= [x(\alpha_{i-1}), y(\alpha_{i-1})] \text{ with } \alpha \in [\alpha_{i-1}, \alpha_i] \\ \{x \in \mathbb{R} | \underline{\delta}(x) > \alpha\} &= [u(\alpha_i), v(\alpha_i)] \text{ with } \alpha \in [\alpha_{i-1}, \alpha_i] \end{aligned}$$

$$\begin{aligned} \{x \in \mathbb{R} | \overline{\pi}(x) \geq \alpha\} &= [x(\alpha_i), y(\alpha_i)] \alpha \in [\alpha_{i-1}, \alpha_i] \\ \{x \in \mathbb{R} | \overline{\delta}(x) > \alpha\} &= [u(\alpha_{i-1}), v(\alpha_{i-1})] \alpha \in [\alpha_{i-1}, \alpha_i] \end{aligned}$$

Given the above transformations, $\mathcal{P}(\underline{\pi}_n) \subset \mathcal{P}(\pi) \subset \mathcal{P}(\overline{\pi}_n)$, and $\lim_{n \rightarrow \infty} \mathcal{P}(\underline{\pi}_n) = \mathcal{P}(\pi)$ and also $\lim_{n \rightarrow \infty} \mathcal{P}(\overline{\pi}_n) = \mathcal{P}(\pi)$. Similarly, $\mathcal{P}(1 - \underline{\delta}_n) \subset \mathcal{P}(1 - \delta) \subset \mathcal{P}(1 - \overline{\delta}_n)$, $\lim_{n \rightarrow \infty} \mathcal{P}(1 -$

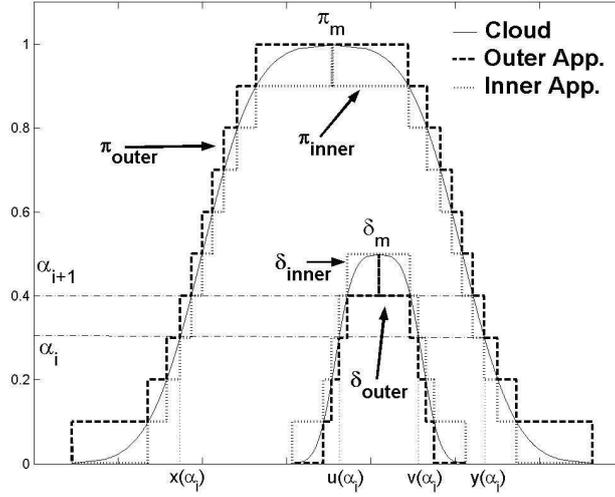


Figure D.1: Inner and outer approximations of a non-comonotonic clouds

$\bar{\delta}_n = \mathcal{P}(1 - \delta)$ and $\lim_{n \rightarrow \infty} \mathcal{P}(1 - \underline{\delta}_n) = \mathcal{P}(1 - \delta)$. Since the set of probabilities induced by the cloud $[\pi, \delta]$ is $\mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta)$, it is clear that the two credal sets $\mathcal{P}(\underline{\pi}_n) \cap \mathcal{P}(1 - \underline{\delta}_n)$ and $\mathcal{P}(\bar{\pi}_n) \cap \mathcal{P}(1 - \bar{\delta}_n)$, are respectively inner and outer approximations of $\mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta)$. Moreover:

$$\lim_{n \rightarrow \infty} \mathcal{P}(\underline{\pi}_n) \cap \mathcal{P}(1 - \underline{\delta}_n) = \mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta)$$

and

$$\lim_{n \rightarrow \infty} \mathcal{P}(\bar{\pi}_n) \cap \mathcal{P}(1 - \bar{\delta}_n) = \mathcal{P}(\pi) \cap \mathcal{P}(1 - \delta).$$

The random sets that are inner approximations (by proposition 3.14) of the finite clouds $[\underline{\delta}_n, \underline{\pi}_n]$ and $[\bar{\delta}_n, \bar{\pi}_n]$ converge to the continuous random set defined by the Lebesgue measure on the unit interval $\alpha \in [0, 1]$ and the multimapping $\alpha \rightarrow E_\alpha$ such that

$$E_\alpha = \{r \in \mathbb{R} | (\pi(r) \geq \alpha) \wedge (\delta(r) < \alpha)\}. \tag{D.13}$$

In the limit, it follows that this continuous random set is an inner approximation of the continuous cloud. \square

D.3 Proofs of Section 5.3

Proof of Proposition 5.3. Note that the bpa of each $(m, \mathcal{F})_{\pi_i}$ form the same vector of masses $(m_{i,1}, \dots, m_{i,M})$, and to simplify notations, we will refer to masses only by their index, and $m_j := m_{i,j}$ for some i . To prove Proposition 5.3, we're first going to express the value that should assume, on elements $x_{(1:N)}$ of $\mathcal{X}_{(1:N)}$, a possibility distribution outer-approximating $(m, \mathcal{F})_{RSI, X_{(1:N)}}$. We're going to express it in terms of masses m_j , $j = 1, \dots, M$, and then we will show that this expression is equivalent to the distribution $\pi'_{X_{(1:N)}}$ given by Equation (5.2).

Let us express the value of the outer approximation in terms of masses $m_{i,j}$. First, note that focal sets of $(m, \mathcal{F})_{RSI, X_{(1:N)}}$ have the general form $\times_{i=1}^N E_{i,j_i}$, with mass $\prod_{i=1}^N m_{j_i}$.

For a given value $j \in \llbracket M \rrbracket$, the focal sets of $(m, \mathcal{F})_{RSI, X_{(1:N)}}$ that are included in $\times_{i=1}^N E_{i,j}$ but not in $\times_{i=1}^N E_{i,j-1}$ are, for $k = 1, \dots, N$,

$$\left\{ \bigotimes_{\substack{I \subseteq \llbracket N \rrbracket \\ |I|=k}} E_{i,j} \times \bigotimes_{\substack{I \subseteq \llbracket N \rrbracket \\ |I|=n-k}} E_{i,j_i} \mid j_i < j \right\}$$

with \otimes standing for cartesian product, and $|I|$ for the cardinality of I . Note that for a fixed value k , there are $\binom{N}{k}$ different subset of $\llbracket N \rrbracket$ having cardinality k . Following Dubois and Prade [89], we can define a mass function defined on focal sets that are cartesian products of the type $\times_{i=1}^N E_{i,j}$ (i.e., α_j -cuts of distributions π_i) by

$$m^*(\times_{i=1}^N E_{i,j}) = \sum_{k=1}^N \binom{N}{k} m_{k,j} \sum_{j_1, \dots, j_{n-k} < j} m_{j_1} \dots m_{j_{n-k}}$$

and, as all the vectors of weights are the same, we can reduce the polynomial $\sum_{j_1, \dots, j_{n-k} < j} m_{j_1} \dots m_{j_{n-k}}$ and get

$$m^*(\times_{i=1}^N E_{i,j}) = \sum_{k=1}^N \binom{N}{k} m_{k,j} \left(\sum_{l < j} m_l \right)^{N-k}$$

this mass function sums up to one, corresponds to a possibility distribution, and outer-approximates $(m, \mathcal{F})_{RSI, X_{(1:N)}}$. Now, let us consider (as done by Dubois and Prade [89]) an element $x_{(1:N)} \in (\times_{i=1}^N E_{i,j}) \setminus (\times_{i=1}^N E_{i,j-1})$ (recall that $E_{i,j} \subseteq E_{i,j-1}$ for any $i \in \llbracket N \rrbracket$ and $j \in \{2, \dots, M\}$), that is an element $x_{(1:N)}$ that is in the cartesian product of α_j -cuts, but not α_{j-1} -cuts. Note that only these elements have to be considered, since the outer-approximation is consonant with focal sets of the type $\times_{i=1}^N E_{i,j}$. Given the outer-approximating mass m^* given above on sets

$\times_{i=1}^N E_{i,j}$, we have

$$\begin{aligned} \pi'_{X_{(1:N)}}(x_{(1:N)}) &= \sum_{i \geq j} m^* (\times_{k=1}^N E_{k,i}) \\ &= \sum_{i \geq j} \left(\sum_{k=1}^N \binom{N}{k} m_i \sum_{j_1, \dots, j_{n-k} < i} m_{j_1} \dots m_{j_{n-k}} \right) \\ &= \sum_{i \geq j} \left(\sum_{k=1}^N \binom{N}{k} m_i \left(\sum_{l < i} m_l \right)^{N-k} \right) = \alpha'_j \end{aligned}$$

the equality from the second to the third line being again obtained by reducing the polynomial $\sum_{j_1, \dots, j_{n-k} < i} m_{j_1} \dots m_{j_{n-k}}$.

Given our choice of $x_{(1:N)}$, we also have that $\min_{i=1, \dots, N} (\pi_i(x_i)) = \alpha_j$. What we want to check is that, if by applying Equation (5.2), we have $\min_{i=1, \dots, N} (\pi'_i(x_i)) = \alpha'_j$? To answer this, first remark that $\pi_i(x_i) = \alpha_j = \sum_{i \geq j} m_i$, and that Equation (5.2) can be rewritten $(-1)^{N+1} (\sum_{i \geq j} m_i - 1)^N + 1$. Checking that $\min_{i=1, \dots, N} (\pi'_i(x_i)) = \alpha'_j$ then amounts to prove the following equality:

$$\sum_{i \geq j} \left(\sum_{k=1}^N \binom{N}{k} m_i \left(\sum_{l < i} m_l \right)^{N-k} \right) = (-1)^{N+1} (\sum_{i \geq j} m_i - 1)^N + 1 \tag{D.14}$$

checking $N = 1$ is trivial, and $N = 2$ has been originally checked by Dubois and Prade [89]. We now prove its validity for any N . First, note that

$$\left(\sum_{l \leq i} m_l \right)^N = \left(m_j + \sum_{l < i} m_l \right)^N = \sum_{k=0}^N \binom{N}{k} (m_i)^k \left(\sum_{l < i} m_l \right)^{N-k} = \left(\sum_{l < i} m_l \right)^N + \sum_{k=1}^N \binom{N}{k} (m_i)^k \left(\sum_{l < i} m_l \right)^{N-k}$$

and

$$\sum_{k=1}^N \binom{N}{k} (m_i)^k \left(\sum_{l < i} m_l \right)^{N-k} = \left(\sum_{l \leq i} m_l \right)^N - \left(\sum_{l < i} m_l \right)^N$$

if we now denote $C_i := \sum_{l \leq i} m_l$ the cumulation of masses m_1, \dots, m_i , the first member of equation (D.14) can be rewritten

$$\sum_{i \geq j} \left(\left(\sum_{l \leq i} m_l \right)^N - \left(\sum_{l < i} m_l \right)^N \right) = \sum_{i \geq j} ((C_i)^N - (C_{i-1})^N) = (C_M)^N - (C_{j-1})^N$$

and, likewise, the second member can be rewritten

$$\begin{aligned} (-1)^{N+1} \left(\left(\sum_{i \geq j} m_i \right) - 1 \right)^N + 1 &= (-1)^{N+1} \left(- \sum_{i < j} m_i \right)^N + 1 = (-1)^{N+1} (-1)^N (C_{j-1})^N + 1 \\ &= (-1)^{2N+1} (C_{j-1})^N + 1 = (C_M)^N - (C_{j-1})^N \end{aligned}$$

since $C_M = 1$ by definition. This finishes our proof. □

D.4 Proofs of Section 6.2

Full Proof of Proposition 6.1 using linear programming. We assume that functions $u_a, \underline{F}, \overline{F}$ are differentiable in \mathbb{R} . Then the following primal and dual optimization problems can be written for computing the lower expectation of the function u_a :

| | |
|---|--|
| <p>Primal problem:</p> <p>Minimize $v = \int_{-\infty}^{\infty} u_a(x) \rho(x) dx$</p> <p>subject to</p> <p>$\rho(x) \geq 0, \int_{-\infty}^{\infty} \rho(x) dx = 1,$</p> <p>$-\int_{-\infty}^x \rho(x) dx \geq -\overline{F}(x),$</p> <p>$\int_{-\infty}^x \rho(x) dx \geq \underline{F}(x).$</p> | <p>Dual problem:</p> <p>Max. $w = c_0 + \int_{-\infty}^{\infty} (-c(t)\overline{F}(t) + d(t)\underline{F}(t)) dt$</p> <p>subject to</p> <p>$c_0 + \int_x^{\infty} (-c(t) + d(t)) dt \leq u_a(x), c_0 \in \mathbb{R},$</p> <p>$c(x) \geq 0, d(x) \geq 0.$</p> |
|---|--|

The proof of equations (6.13)-(6.14) and (6.17) then follows in three main steps:

1. We propose a feasible solution of the primal problem.
2. We then consider the feasible solution of the dual problem corresponding to the one proposed for the primal problem.
3. We show that the two solutions coincide and, therefore, according to the basic duality theorem of linear programming, these solutions are optimal ones.

First, we consider the primal problem. Let a' and a'' be real values. We can check that the function

$$\rho(x) = \begin{cases} d\bar{F}(x)/dx, & x < a' \\ 0, & a' \leq x \leq a'' \\ d\underline{F}(x)/dx, & a'' < x \end{cases}$$

is a feasible solution to the primal problem if the following conditions are respected:

$$\int_{-\infty}^{\infty} \rho(x) dx = 1,$$

which, given the above solution, can be rewritten

$$\int_{-\infty}^{a'} d\bar{F} + \int_{a''}^{\infty} d\underline{F} = 1,$$

which is equivalent to the equality

$$\bar{F}(a') = \underline{F}(a''). \tag{D.15}$$

We now interest ourselves in the dual problem. Let us first consider the sole constraint

$$c_0 + \int_x^{\infty} (-c(t) + d(t)) dt \leq u_a(x), \tag{D.16}$$

which is the equivalent of the primal constraint $\rho(x) \geq 0$. We then consider the following feasible solution to the dual problem as $c_0 = u_a(\infty)$,

$$c(x) = \begin{cases} h'(x), & x < a' \\ 0, & x \geq a' \end{cases} \quad d(x) = \begin{cases} 0, & x < a'' \\ -u_a'(x), & x \geq a'' \end{cases}.$$

The inequalities $c(x) \geq 0$ and $d(x) \geq 0$ are valid provided we have the inequalities $a' \leq a \leq a''$ (i.e. interval $[a', a'']$ encompasses maximum of u_a). By integrating $c(x)$ and $d(x)$, we get the increasing function

$$C(x) = - \int_x^{\infty} c(t) dt = \begin{cases} u_a(x) - u_a(a'), & x < a' \\ 0, & x \geq a' \end{cases}$$

and the decreasing function

$$D(x) = \int_x^\infty d(t) dt = \begin{cases} u_a(a'') - u_a(\infty), & x < a'' \\ u_a(x) - u_a(\infty), & x \geq a'' \end{cases}.$$

Let us rewrite condition (D.16) as follows:

$$c_0 + C(x) + D(x) \leq u_a(x). \quad (\text{D.17})$$

If $x < a'$, equation (D.17) reads

$$c_0 + u_a(x) - u_a(a') + u_a(a'') - u_a(\infty) = u_a(x).$$

Hence

$$u_a(a'') = u_a(a'). \quad (\text{D.18})$$

If $a' < x < a''$, we have $c_0 + u_a(a'') - u_a(\infty) \leq h(x)$ which means that for all $x \in (a', a'')$ we have $h(a'') (= u_a(a')) \leq h(x)$ (i.e. $u_a(a'')$ and a' are the minimal values of the function $u_a(x)$ in interval $x \in (a', a'')$.) If $x \geq a''$, then we get the trivial equality $c_0 + u_a(x) - u_a(\infty) = u_a(x)$. The two proposed solutions are valid if and only if equation (D.15) is valid for the primal problem and equation (D.18) is valid for the dual problem. In order to do it, let us consider the function

$$\varphi(\alpha) = u_a(\overline{F}^{-1}(\alpha)) - u_a(\underline{F}^{-1}(\alpha)),$$

which, being a subtraction of two continuous functions (by supposition), is continuous. Since the function u_a has its maximum at point $x = a$, then, by taking $\alpha = \overline{F}(a)$, we get the inequality

$$\varphi(\gamma) = u_a(\overline{F}^{-1}(\underline{F}(a))) - u_a(a) \leq 0$$

and, by taking $\gamma = \overline{F}(a)$, we get the inequality

$$\varphi(\gamma) = u_a(a) - u_a(\underline{F}^{-1}(\overline{F}(a))) \geq 0.$$

Consequently, there exists γ in the interval $(\underline{F}(a), \overline{F}(a))$ such that $\varphi(\gamma) = 0$ (since φ is continuous). Therefore, there exist $a' = \overline{F}^{-1}(\gamma)$ and $a'' = \underline{F}^{-1}(\gamma)$ (hence, equality (D.15) holds) such that equality $u_a(a') = u_a(a'')$ in (D.18) is valid. We find the values of the objective

functions

$$v_{\min} = \int_0^{a'} u_a(x) d\bar{F} + \int_{a''}^{\infty} u_a(x) d\underline{F},$$

$$w_{\max} = c_0 + \int_0^{\infty} (-c(t)\bar{F}(t) + d(t)\underline{F}(t)) dt.$$

and, by using integration by parts together with equations (D.15)-(D.18), we can show that equality $w_{\max} = v_{\min}$ holds, with γ the particular solution of equation (6.17) for which optimum is reached, as was to be proved. □

Appendix E

Remarks on Nested-Disjoint clouds

Some of the non-monotonic clouds violating conditions of Proposition 3.13 correspond to the case where, for any pair of events $\{B_{\bar{\gamma}_i}, C_{\gamma_j} | i, j = 1, \dots, n\}$, we have $B_{\bar{\gamma}_i} \cap C_{\gamma_j} \in \{B_{\bar{\gamma}_i}, C_{\gamma_j}, \emptyset\}$, with at least one pair $B_{\bar{\gamma}_i}, C_{\gamma_j}$ of non-empty sets such that $B_{\bar{\gamma}_i} \cap C_{\gamma_j} = \emptyset$. In other words, $\forall i, j, B_{\bar{\gamma}_i}$ and C_{γ_j} are either nested or disjoint. They can be called *nested-disjoint* clouds and they are fully described by the existence of three indices $j > k \geq l$ s.t.

$$\text{the sets } \{B_{\bar{\gamma}_i} | i < j\} \cup \{C_{\gamma_i} | i \leq k\} \text{ form a nested sequence} \quad (\text{E.1})$$

$$C_{\gamma_i} \cap B_{\bar{\gamma}_j} = \emptyset \quad \forall i \quad k \geq i \geq l \quad (\text{E.2})$$

$$C_{\gamma_i} = \emptyset \quad \forall i > k \text{ and } C_{\gamma_i} \neq \emptyset \quad \forall i \leq k \quad (\text{E.3})$$

$$B_{\bar{\gamma}_i} \cap C_{\gamma_f} = B_{\bar{\gamma}_i} \text{ or } C_{\gamma_f} \quad i > k, f < l. \quad (\text{E.4})$$

and these four statements induce the fact that $(C_{\gamma_i} \cup B_{\bar{\gamma}_j}) \subset B_{\bar{\gamma}_{j-1}}$, since we know that $B_{\bar{\gamma}_j} \subset B_{\bar{\gamma}_{j-1}}, C_{\gamma_i} \cap B_{\bar{\gamma}_j} = \emptyset$ (Statement (E.2)) and $C_{\gamma_i} \cap B_{\bar{\gamma}_{j-1}} \in \{C_{\gamma_i}, B_{\bar{\gamma}_{j-1}}\}$ (Statement (E.1)). Given these facts, $C_{\gamma_i} \cap B_{\bar{\gamma}_{j-1}} = C_{\gamma_i}$, otherwise we end up with a contradiction. The structure of this particular case is summarized by Figure E.1 (where only the most important sets are represented).

We have strong reasons to think that these particular clouds, although not generalized p-boxes, can still be represented by random sets. A first reason is that we can associate to the sets $B_{\bar{\gamma}_1}, \dots, B_{\bar{\gamma}_j}$ a possibility distribution (i.e. they are nested and are associated to lower probability bounds), while the sets $\{B_{\bar{\gamma}_i} | i < j\} \cup \{C_{\gamma_i} | i \leq k\}$ form a nested sequence and can thus be associated to a generalized p-box. The nested-disjoint clouds could then be seen as a convex mixture of two random sets, thus giving again a random set. Secondly, this conjecture is reinforced by the following simple example: Let us consider a cloud whose cuts are such

that $B_{\overline{\gamma_2}}, B_{\overline{\gamma_1}}, C_{\gamma_2}$ with $B_{\overline{\gamma_2}} \subset B_{\overline{\gamma_1}}, C_{\gamma_2} \subset B_{\overline{\gamma_1}}$ and $C_{\gamma_1} \cap B_{\overline{\gamma_2}} = \emptyset$, together with the two weights $\gamma_2 > \gamma_1$. This cloud is nested-disjoint, and the belief function such that $m(B_{\overline{\gamma_2}}) = 1 - \gamma_2$, $m(B_{\overline{\gamma_1}}) = \gamma_2 - \gamma_1$ and $m(C_{\gamma_1}^c) = \gamma_1$ models the same credal set as this cloud (since the presence of C_{γ_1} only induces a bound over its upper probability).

Finally, it should be noted that, in the case of continuous clouds, this subclass does not exist, since when distribution δ is such that $\sup_{x \in \mathcal{X}} \delta(x) > 0$ and δ, π are not comonotonic, we can always exhibit two cuts that are overlapping.

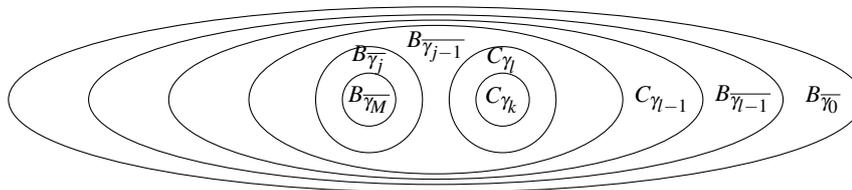


Figure E.1: Structure of a nested-disjoint cloud

Appendix F

Generalized p-boxes on complete chain

In this appendix, we study the notion of generalized p-boxes defined on a totally ordered space \mathcal{X} (i.e., a complete chain), and where \mathcal{X} is not necessarily finite. Such spaces are quite general, since they encompass in one sweep all finite spaces on which a total order is defined, as well as product spaces \mathbb{R}^d , provided these last ones are equipped with a total order between elements. Note that, in order to be as general as possible, and to be coherent with Walley's [203] approach, we also consider *finitely additive* probabilities, which was not the case previously.

The discussed results are part of an on-going work with Enrique Miranda and Matthias Troffaes [150]. As we shall see, many of the results concerning generalized p-boxes on pre-ordered finite spaces (see Section 3.2) extends more or less straightforwardly to the case of (not necessarily finite) totally ordered spaces, but some of them do not extend so easily. Most proofs are omitted, and we limit ourselves to main notions and results.

As before, we denote by $\mathcal{L}(\mathcal{X})$ the set of all bounded and real-valued functions f on \mathcal{X} . Let \underline{E} be the lower expectation of some credal set. Then, a lower expectation \underline{E} defined on a lattice of gambles $\mathcal{K} \subseteq \mathcal{L}(\mathcal{X})$ is called n -monotone when if for all $p \in \llbracket N \rrbracket$, $p \leq n$, and all f, f_1, \dots, f_p in \mathcal{K} :

$$\sum_{I \subseteq \{1, \dots, p\}} (-1)^{|I|} \underline{E} \left(f \wedge \bigwedge_{i \in I} f_i \right) \geq 0. \quad (\text{F.1})$$

With \wedge the point-wise minimum. A lower expectation which is n -monotone for all $n \in \llbracket N \rrbracket$ is called *completely monotone*, and is the counterpart of ∞ -monotone capacities.

F.1 Characterization of generalized p-boxes

As said before, \mathcal{X}, \leq is here an order complete chain, and we take $x < y$ as a brief notation for $x \leq y$ and $x \not\geq y$. So \leq is transitive, reflexive, and anti-symmetric, and for any two elements $x, y \in \mathcal{X}$ we have either $x < y$, $x = y$, or $x > y$. For simplicity, we assume that \mathcal{X} has a smallest element $0_{\mathcal{X}}$ and a largest element $1_{\mathcal{X}}$.

We will use the same notation for an event A and its indicator function $\mathbf{1}_{(A)}$. This means that, in the following, $\underline{E}(A)$ will denote the lower expectation of $\mathbf{1}_{(A)}$.

F.1.1 Definition

With the term *distribution function*, we shall refer to any function $F : \mathcal{X} \rightarrow [0, 1]$ which is non-decreasing (with respect to \leq) and satisfies moreover $F(1_{\mathcal{X}}) = 1$. $F(x)$ provides information about the cumulative probability on the interval $[0_{\mathcal{X}}, x]$. Note that we need not to impose $F(0_{\mathcal{X}}) = 0$. Also note that distribution functions are not assumed to be right-continuous. Given a distribution F on \mathcal{X} and a value $x \in \mathcal{X}$, we note $F(x^+)$ the right-limit $F(x^+) = \inf\{F(y), y > x\}$ and $F(x^-)$ the left-limit $F(x^-) = \sup\{F(y), y < x\}$ ¹.

Definition F.1. A *generalized probability box*, or *generalized p-box*, is a pair $[\underline{F}, \overline{F}]$ of distribution functions from \mathcal{X} to $[0, 1]$, satisfying $\underline{F}(1_{\mathcal{X}}) = \overline{F}(1_{\mathcal{X}}) = 1$ and $\underline{F} \leq \overline{F}$. If \mathcal{X} is a closed interval on \mathbb{R} , then we call the pair $[\underline{F}, \overline{F}]$ a *p-box*.

Similarly to its definition on other spaces, a generalized p-box is interpreted as lower and upper probabilities $\underline{P}_{[\underline{F}, \overline{F}]}$ on the set of events

$$\mathcal{H} = \{[0_{\mathcal{X}}, x] | x \in \mathcal{X}\} \cup \{(y, 1_{\mathcal{X}}] | y \in \mathcal{X}\}$$

by

$$\underline{P}_{[\underline{F}, \overline{F}]}([0_{\mathcal{X}}, x]) := \underline{F}(x) \text{ and } \underline{P}_{[\underline{F}, \overline{F}]}((y, 1_{\mathcal{X}}]) = 1 - \overline{F}(y). \quad (\text{F.2})$$

for classical p-boxes, it was already mentioned by Walley [203, Section 4.6.6] and proved by Troffaes [193, p. 93, Thm. 3.59] that $\underline{P}_{[\underline{F}, \overline{F}]}$ is a coherent lower probability. The proof extends easily to the present case, and thus $\underline{P}_{[\underline{F}, \overline{F}]}$ is a coherent lower probability on \mathcal{H} , and

¹Or, equivalently, $F(x^+) = \lim_{y \rightarrow x} (F(y))$ and $F(x^-) = \lim_{y \rightarrow x} (F(y))$

corresponds to the lower envelope of a credal set $\mathcal{P}_{[\underline{F}, \overline{F}]}$, which is here the set of all finitely additive probabilities satisfying constraints (F.2).

Given a generalized p-box, we can consider the set of distribution functions that lie between \underline{F} and \overline{F} ,

$$\Phi(\underline{F}, \overline{F}) = \{F \mid \underline{F} \leq F \leq \overline{F}\}.$$

We can easily express the lower expectation $\underline{E}_{[\underline{F}, \overline{F}]}$ in terms of $\Phi(\underline{F}, \overline{F})$: $\underline{E}_{[\underline{F}, \overline{F}]}$ is the lower envelope of the lower expectations of the cdfs F that lie between \underline{F} and \overline{F} :

$$\underline{E}_{[\underline{F}, \overline{F}]}(f) = \inf_{F \in \Phi(\underline{F}, \overline{F})} \underline{E}_F(f) \tag{F.3}$$

for all functions f on \mathcal{X} . A similar result for p-boxes in the unit interval can be found in [203, Section 4.6.6].

We now study and characterize this lower expectation $\underline{E}_{[\underline{F}, \overline{F}]}$. In order to do so, we will first study it on a restricted and useful set \mathcal{H} of indicator functions, that is the field of events generated by \mathcal{K} , and will then extend the results to any kind of event and then to any function in $\mathcal{L}\mathcal{X}$.

F.1.2 The Field \mathcal{H}

The field of events \mathcal{H} generated by the domain \mathcal{K} of p-boxes corresponds to events of the type

$$[0_{\mathcal{X}}, x_1] \cup (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$$

for $x_1 < x_2 < x_3 < \dots < x_{2n+1}$ in \mathcal{X} (if n is 0 we simply take this expression to be $[0_{\mathcal{X}}, x_1]$) and

$$(x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$$

for $x_2 < x_3 < \dots < x_{2n+1}$ in \mathcal{X} . Clearly, these events form a field: the union and intersection of any two events in \mathcal{H} is again in \mathcal{H} , and the complement of any event in \mathcal{H} also is again in \mathcal{H} .

In the case of a precise cdf F^2 , P_F has a unique extension to a probability charge on \mathcal{H} .

Proposition F.1. \underline{E}_F restricted to \mathcal{H} is a probability charge. Moreover, for any $A \in \mathcal{H}$ which contains $0_{\mathcal{X}}$, that is $A = [0_{\mathcal{X}}, x_1] \cup (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$, it holds that

$$\underline{E}_F(A) = F(x_1) + \sum_{k=1}^n (F(x_{2k+1}) - F(x_{2k})) \quad (\text{F.4})$$

and for any $A \in \mathcal{H}$ which does not contain $0_{\mathcal{X}}$, that is $A = (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$,

$$\underline{E}_F(A) = \sum_{k=1}^n (F(x_{2k+1}) - F(x_{2k}))$$

This proposition can be extended to p-boxes in the following way

Proposition F.2. For any $A \in \mathcal{H}$ which contains $0_{\mathcal{X}}$, that is $A = [0_{\mathcal{X}}, x_1] \cup (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$, it holds that

$$\underline{E}_{[\underline{F}, \overline{F}]}(A) = \underline{F}(x_1) + \sum_{k=1}^n \max\{0, \underline{F}(x_{2k+1}) - \overline{F}(x_{2k})\} \quad (\text{F.5})$$

and for any $A \in \mathcal{H}$ which does not contain $0_{\mathcal{X}}$, that is $A = (x_2, x_3] \cup \dots \cup (x_{2n}, x_{2n+1}]$,

$$\underline{E}_{[\underline{F}, \overline{F}]}(A) = \sum_{k=1}^n \max\{0, \underline{F}(x_{2k+1}) - \overline{F}(x_{2k})\} \quad (\text{F.6})$$

Among other things, this proposition shows that Remark 3.2 can be extended to the case where \mathcal{X} is an infinite totally ordered space. Next results show that p-boxes on totally ordered spaces induce a completely monotone lower expectation (see Equation (F.1)), thus somewhat extending Proposition 3.4 (since on finite spaces, there is a duality between ∞ -monotonicity and the fact of being a random set).

F.2 The lower expectation of p-boxes: A Choquet Integral Representation

As shown in [149], the lower expectation \underline{E}_F of a (precise) distribution function F on $[0, 1]$ is completely monotone. Generalizing this result to distribution functions on a totally ordered

²A precise cdf is such that $\underline{F} = \overline{F} =: F$

space \mathcal{X} is straightforward. However, given this, and Eq. (F.3), we cannot establish the complete monotonicity of $\underline{E}_{[F, \bar{F}]}$, because the lower envelope of a set of completely monotone lower expectation is not necessarily completely monotone. Next results indicate that such an envelope is indeed completely monotone in the case of generalized p-boxes.

Let $\underline{P}_{[F, \bar{F}]}^{\mathcal{H}}$ denote the restriction of $\underline{E}_{[F, \bar{F}]}$ to \mathcal{H} (That is, to the lower probabilities induced by $[F, \bar{F}]$ of events in \mathcal{H}). Note that we have already derived closed form expressions for $\underline{P}_{[F, \bar{F}]}^{\mathcal{H}}$ in Proposition F.2.

Theorem F.3. $\underline{P}_{[F, \bar{F}]}^{\mathcal{H}}$ is a completely monotone coherent lower probability.

And the inner measure $\underline{P}_{[F, \bar{F}]}^{\mathcal{H}*}$ of $\underline{P}_{[F, \bar{F}]}^{\mathcal{H}}$, defined as

$$\underline{P}_{[F, \bar{F}]}^{\mathcal{H}*}(A) = \sup_{C \in \mathcal{H}, C \subseteq A} \underline{P}_{[F, \bar{F}]}^{\mathcal{H}}(C), \tag{F.7}$$

coincides with $\underline{E}_{[F, \bar{F}]}$ on all events [203, Cor. 3.1.9, p. 127]. It allows us moreover to characterise the lower expectation on all function on $\mathcal{L} \mathcal{X}$, as is shown in the following theorem:

Theorem F.4. The lower expectation $\underline{E}_{[F, \bar{F}]}$ of $\underline{P}_{[F, \bar{F}]}$ to all functions in $\mathcal{L} \mathcal{X}$ is given by the Choquet integral $(C) \int \cdot d\underline{P}_{[F, \bar{F}]}^{\mathcal{H}*}$ and is completely monotone as well.

This theorem confirm that generalized p-boxes on totally ordered spaces continue to be completely monotone, as they are for finite totally (pre-)ordered spaces. As the above expression is not very manageable, it is desirable to dispose of more manageable ones. Nevertheless, the use of an equivalent random set is not possible anymore, since the duality between random set representation and complete monotonicity do not longer forcefully hold on infinite spaces. In order to give more manageable expressions, we will consider the *upper limit topology* on \mathcal{X} which is the topology generated by the base

$$\tau := \{(x, y] : x, y \in \mathcal{X}, x < y\} \cup \{[0_x, x] : x \in \mathcal{X}\}.$$

As we have indicated (Proposition F.3), $\underline{P}_{[F, \bar{F}]}$ is a completely monotone (and therefore coherent) lower probability on the field \mathcal{H} generated by τ . For any function f on \mathcal{X} , let us define its *lower oscillation* as the gamble

$$\underline{osc}(f)(d) := \sup_{C \in \tau: d \in C} \inf_{x \in C} f(x). \tag{F.8}$$

which comes down to consider the supremum of all lower semi-continuous functions dominated by f (recall that a function g dominates another function f if $f \leq g$).

Given a subset A of \mathcal{X} , it follows from Eq. (F.8) that the lower oscillation of $\mathbf{1}_{(A)}$ is the indicator function of the set

$$B := \{d \in A : \exists C \in \tau \text{ s.t. } d \in C \subseteq A\} = \bigcup_{C \in \tau: C \subseteq A} C = \text{int}(A) = \underline{\text{osc}}(\mathbf{1}_{(A)}); \quad (\text{F.9})$$

note that B is the union of the elements of the base τ that are included in A , and is therefore the topological interior of A in the upper limit topology. The following Lemma then follows:

Lemma F.1. *For any subset A of \mathcal{X} , $\underline{E}(A) = \underline{E}(B)$, where $B = \text{int}(A)$ is given by Eq. (F.9).*

What this result tells us is that the natural extension of a generalized p-box is characterised by its restriction to lower semi-continuous gambles (and, in the case of events, to the restriction to open sets). It allows us to rewrite the Choquet integral of Theorem F.4 as

$$\underline{E}(f) = \inf \underline{\text{osc}}(f) + \int_{\inf \underline{\text{osc}}(f)}^{\sup \underline{\text{osc}}(f)} \underline{E}(\{\underline{\text{osc}}(f) \geq x\}) dx = \underline{E}(\underline{\text{osc}}(f)), \quad (\text{F.10})$$

an expression which is indeed more manageable.

Above results also allow us to determine the expression of the lower expectation \underline{E} on the subsets of \mathcal{X} which are open in the upper limit topology. Before that, we need to introduce some preliminary notions. Let B be an open subset of \mathcal{X} , and let us show that B is a union of pairwise disjoint open intervals of \mathcal{X} . Recall that by open we are referring here to the upper limit topology, so the subinterval $(a, b]$ is also open for any a, b in \mathcal{X} such that $a \leq b$.

Definition F.2. [176] A set S is called *full* if $[a, b] \subseteq S$ for any $a \leq b$ in S . Given a set B and an element x of B , the *full component* $C(x, B)$ of x in B is the largest full set S which satisfies $x \in S \subseteq B$.

The full components $\{C(x, A) : x \in A\}$ of a set $A \subseteq \mathcal{X}$ form a partition of A [176, 4.4(a)]. This partition has a countable number of elements in the particular case where $\mathcal{X} = [0, 1]$, but not necessarily in general (since, contrary to the case where \mathcal{X} is the real line, we can consider the set $(a, a]$). The following Lemma shows that if A is open, each of these full components is an open interval.

Lemma F.2. *Let B be a subset of \mathcal{X} . If B is open, so are its full components.*

This allow us to state the following lemma, showing that the lower expectation \underline{E} is additive on full components.

Lemma F.3. *Let B be an arbitrary subset of \mathcal{X} , and let $(B_\lambda)_{\lambda \in \Lambda}$ be the full components of B . Then $\underline{E}(\cup_{\lambda \in \Lambda} B_\lambda) = \sum_{\lambda \in \Lambda} \underline{E}(B_\lambda)$.*

Note that this lemma extends to totally ordered spaces the part of Remark 3.2 stating that the lower probability of disjoint unions of consecutive elements was the sum of the lower probabilities of every such union.

All these results allow us to deduce that the lower expectation \underline{E} is characterised by the value it takes on the full components of the open sets. From Lemma F.2, these full components are open intervals of \mathcal{X} , and are therefore of the form $[0_{\mathcal{X}}, x], (x, y], [0_{\mathcal{X}}, x)$ or (x, y) , for $x \leq y$ in \mathcal{X} . It follows from Proposition F.2 that $\underline{E}([0_{\mathcal{X}}, x]) = \underline{F}(x)$ and $\underline{E}((x, y]) = \max\{0, \underline{F}(y) - \overline{F}(x)\}$ for any $y \leq x$ in \mathcal{X} . On the other hand, we deduce from Theorem F.4 and Equation (F.7) that

$$\underline{E}([0_{\mathcal{X}}, x)) = \underline{F}(x-) \text{ and } \underline{E}(x, y) = \max\{0, \underline{F}(y-) - \overline{F}(x)\} \tag{F.11}$$

for any $x \leq y$ in \mathcal{X} .

F.3 Approximating lower expectation by limits of p-boxes

Now that we have studied and characterized the lower expectation induced by a p-box on totally ordered spaces, we give an expression allowing to approximate the lower expectation by limits of discrete p-boxes. This approach is quite close to the one considered in chapter 6 using linear programming for computing lower expectations of p-boxes defined on the real line. Nevertheless, there are at least three important differences between the two approaches: the space here is more general, we consider finitely additive probabilities, and no continuity assumptions are made about functions f . Consider a p-box $[\underline{F}, \overline{F}]$ on \mathcal{X} . Let $(\underline{F}_n)_n, (\overline{F}_n)_n$ be increasing and decreasing sequences of cdfs converging point-wise to \underline{F} and \overline{F} , respectively.

For ease of notation, denote by \underline{P}_n the lower probability associated with $[\underline{F}_n, \overline{F}_n]$, that is, $\underline{P}_n = \underline{P}_{[\underline{F}_n, \overline{F}_n]}^{\mathcal{H}}$, and \mathcal{P}_n the induced credal set, and let \underline{E}_n be the lower expectation corresponding to \mathcal{P}_n . Since $\underline{F}_n \leq \underline{F}$ and $\overline{F}_n \geq \overline{F}$, it follows that $\Phi(\underline{F}, \overline{F}) \subseteq \Phi(\underline{F}_n, \overline{F}_n)$ (or, equivalently, $\mathcal{P}_{[\underline{F}, \overline{F}]} \subseteq \mathcal{P}_n$, and Eq. (F.3) implies that $\underline{E}_n \leq \underline{E}$). Moreover, the same argument implies that $\underline{E}_n \leq \underline{E}_{n+1}$ for any $n \in \llbracket N \rrbracket$, so $\lim_n \underline{E}_n = \sup_n \underline{E}_n \leq \underline{E}$. The following proposition shows that the converse holds too:

Proposition F.5. $\underline{E}(f) = \lim_n \underline{E}_n(f)$ for any gamble f .

Next, we use this Proposition to establish an expression for the lower expectation of a generalised p-box in terms of discrete generalized p-boxes. For any natural number $n \geq 1$, and $i \in \{2, \dots, n\}$, define the sets $A_1^n := \bar{F}^{-1}([0, \frac{1}{n}])$, $A_i^n := \bar{F}^{-1}((\frac{i-1}{n}, \frac{i}{n}])$, $B_1^n := \underline{F}^{-1}([0, \frac{1}{n}])$ and $B_i^n := \underline{F}^{-1}((\frac{i-1}{n}, \frac{i}{n}))$. Clearly, both $\{A_1^n, \dots, A_n^n\}$ and $\{B_1^n, \dots, B_n^n\}$ are partitions of \mathcal{X} . Define \underline{F}_n and \bar{F}_n by

$$\bar{F}_n(x) = \frac{i}{n} \text{ if } x \in A_i^n, \quad \underline{F}_n(x) = \begin{cases} \frac{i-1}{n} & \text{if } x \in B_i^n \text{ and } x \neq 1_{\mathcal{X}}, \\ 1 & \text{if } x = 1_{\mathcal{X}}. \end{cases} \quad (\text{F.12})$$

Lemma F.4. *The following statements hold for all $x \in \mathcal{X}$:*

- (i) *For any $n \in \llbracket N \rrbracket$, \underline{F}_n and \bar{F}_n are cdfs, $\underline{F}_n(x) \leq \underline{F}(x)$, and $\bar{F}(x) \leq \bar{F}_n(x)$.*
- (ii) *$\lim_n \underline{F}_n(x) = \underline{F}(x)$ and $\lim_n \bar{F}_n(x) = \bar{F}(x)$.*
- (iii) *$(\underline{F}_{2^n})_n, (\bar{F}_{2^n})_n$ are increasing and decreasing sequences of cdfs such that $\underline{F}(x) = \lim_n \underline{F}_{2^n}(x)$ and $\bar{F}(x) = \lim_n \bar{F}_{2^n}(x)$.*

If we can find a simple expression for the lower expectations induced by \underline{P}_n for our particular choice of \underline{F}_n and \bar{F}_n , then we also have a simple expression for $\underline{E}_{[\underline{F}, \bar{F}]}$ via Proposition F.5. To derive such a simple expression, we consider $\underline{G}_1, \dots, \underline{G}_n$ and $\bar{G}_1, \dots, \bar{G}_n$ defined by

$$\underline{G}_i(x) = \begin{cases} 1 & \text{if } \underline{F}_n(x) \geq \frac{i}{n} \\ 0 & \text{otherwise} \end{cases} \quad \bar{G}_i(x) = \begin{cases} 1 & \text{if } \bar{F}_n(x) \geq \frac{i}{n} \\ 0 & \text{otherwise} \end{cases}$$

and the following proposition holds

Proposition F.6. *For each $n \in \llbracket N \rrbracket$, $\underline{E}_n = \frac{1}{n} \sum_{i=1}^n \underline{E}_{\underline{G}_i, \bar{G}_i}$.*

Hence, all we need to characterise the lower expectation induced by $[\underline{F}, \bar{F}]_n$ is to determine the lower expectation induced by a *degenerate p-box*, i.e. one where the lower and upper cdfs only assume the values 0 and 1. Note that a degenerate p-box (\underline{G}, \bar{G}) is uniquely determined by

$$I_{(\underline{G}, \bar{G})} = \{x \in \mathcal{X} : \underline{G}(x) < \bar{G}(x)\} = \{x \in \mathcal{X} : \underline{G}(x) = 0 \text{ and } \bar{G}(x) = 1\}.$$

as $\underline{G}(x) = \bar{G}(x) = 0$ on the left of this interval, and $\underline{G}(x) = \bar{G}(x) = 1$ on the right of it. The following proposition characterize the lower expectation $\underline{E}_{\underline{G}, \bar{G}}$ induced by such degenerate p-boxes:

Proposition F.7. *Let $(\underline{G}, \overline{G})$ be a degenerate p-box, and $f \in \mathcal{L}(\mathcal{X})$. If $0_{\mathcal{X}} \notin I_{(\underline{G}, \overline{G})}$,*

$$(i) \text{ If } I_{(\underline{G}, \overline{G})} = (a, b) \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \inf_{z \in (a, b]} f(z).$$

$$(ii) \text{ If } I_{(\underline{G}, \overline{G})} = (a, b] \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \lim_{y \rightarrow b} \inf_{z \in (a, y]} f(z).$$

$$(iii) \text{ If } I_{(\underline{G}, \overline{G})} = [a, b) \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \lim_{x \rightarrow a} \inf_{z \in (x, b]} f(z).$$

$$(iv) \text{ If } I_{(\underline{G}, \overline{G})} = [a, b] \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \lim_{x \rightarrow a} \lim_{y \rightarrow b} \inf_{z \in (x, y]} f(z).$$

On the other hand, if $0_{\mathcal{X}} \in I_{(\underline{G}, \overline{G})}$, then

$$(a) \text{ If } I_{(\underline{G}, \overline{G})} = [0_{\mathcal{X}}, b) \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \inf_{z \in [0_{\mathcal{X}}, b]} f(z).$$

$$(b) \text{ If } I_{(\underline{G}, \overline{G})} = [0_{\mathcal{X}}, b] \text{ then } \underline{E}_{\underline{G}, \overline{G}}(f) = \lim_{y \rightarrow b} \inf_{z \in [0_{\mathcal{X}}, y]} f(z).$$

Concluding, if we consider now the lower expectation \underline{E}'_n of $(\underline{F}_{2^n}, \overline{F}_{2^n})$ as defined in Eq. (F.12), it follows from Proposition F.5 and Lemma F.4 that $(\underline{E}'_n)_n$ is an increasing sequence of functionals that converges point-wise to the lower expectation \underline{E} of $[\underline{F}, \overline{F}]$. By Proposition F.6, \underline{E}'_n can be calculated as a convex combination of natural extensions of degenerate p-boxes, whose expressions follow from Proposition F.7.

We thus end up with expressions allowing to compute the lower expectation of any generalized p-box defined on totally ordered spaces as a combination of lower expectations of degenerate generalized p-boxes, this last lower expectations being easier to compute. Among other things, these results also show that extending results found on finite spaces is far from being always trivial, not to say impossible, as show the next section.

F.4 Relating random sets with p-boxes on the unit interval

In the case of generalized p-boxes defined on finite (pre-)ordered spaces, Proposition 3.4 indicates that they are equivalent to special kinds of random sets. This extends relations already hinted by Ferson *et al.* [104] for p-boxes defined on the discretized real line.

Here, we study the extensions of such results to p-boxes defined on the (continuous) unit interval $[0, 1]$ on the real. Note that, up to a transformation, this is equivalent to study p-boxes defined on any closed interval of the real line. Relating random sets to such p-boxes is easier than relating random sets to generalized p-boxes defined on a totally ordered space \mathcal{X} , due

to some properties: for instance, the number of full components of any open set B is always countable (which is not the case on \mathcal{X} , see Definition F.2); and, using the Borel σ -field on $[0, 1]$, we can establish a link between these p-boxes and random sets, while considering fields of measurable sets on \mathcal{X} would require many mathematical subtleties not forcefully desirable for a first study.

Coherently with previous notations, we denote by $\mathcal{L}([0, 1])$ the set of bounded real-valued functions on $[0, 1]$. Given \underline{F} and \overline{F} , let us define their *pseudo-inverses* by

$$L(\alpha) := \sup \{x \in [0, 1] \mid \overline{F}(x) < \alpha\}$$

$$U(\alpha) := \inf \{x \in [0, 1] \mid \underline{F}(x) > \alpha\}$$

for all $\alpha \in [0, 1]$. These are non-decreasing maps from $[0, 1]$ to $[0, 1]$. Because $\underline{F} \leq \overline{F}$, it follows that $L \leq U$. Let $\Gamma : [0, 1] \rightarrow \wp([0, 1])$ be the multi-valued mapping given by $\Gamma(\alpha) = [L(\alpha), U(\alpha)]$, with $[L(\alpha), U(\alpha)]$ a closed interval whose lowest and highest elements are respectively $L(\alpha)$ and $U(\alpha)$. Define, for each $\alpha \in [0, 1]$, the lower prevision \underline{Q}_α on $\mathcal{L}([0, 1])$ as follows:

$$\underline{Q}_\alpha(f) = \inf \{f(x) \mid x \in \Gamma(\alpha)\};$$

i.e., \underline{Q}_α is the lower expectation relative to \mathfrak{J} , the credal sets modeling ignorance on the set $\Gamma(\alpha)$. Any such lower expectation is coherent and completely monotone, as was shown in [51, Theorem 10]. For any function f , we define the lower expectation $\underline{Q}_{[\underline{F}, \overline{F}]}$ on f by

$$\underline{Q}_{[\underline{F}, \overline{F}]}(f) = \int_0^1 \underline{Q}_\alpha(f) d\alpha_*$$

which is a Lebesgue inner integral. $\underline{Q}_{[\underline{F}, \overline{F}]}$ is a coherent lower expectation, induced by the random set defined by Γ . Now, the question we want to investigate in this section is to which extend the random set defined by Γ is related to the p-box $[\underline{F}, \overline{F}]$, that is what are the relationships between the lower expectations $\underline{Q}_{[\underline{F}, \overline{F}]}$ and \underline{E} .

Our results show that, in general, lower expectations $\underline{Q}_{[\underline{F}, \overline{F}]}$ and \underline{E} do not coincide, consequently Proposition 3.4 do not in general fully extend to more general cases, and there is no longer one-to-one correspondence between p-boxes and specific random sets. However, we will show that the relation between p-boxes and random sets continue to hold for particular cases of practical interest.

F.4.1 Lower expectation on general functions

Let us first introduce some notations. Given a distribution function F , let \tilde{F} be the mapping given by $\tilde{F}(x) = F(x^+) = \inf\{F(y) : y > x\}$ for any $x \in [0, 1]$, $\tilde{F}(1) = 1$. In other words, $\tilde{F}(x)$ is the right-continuous approximation of F . Note that the functionals L, U defined above do not change if we replace $\underline{F}, \overline{F}$ by their right-continuous approximations $\underline{\tilde{F}}, \overline{\tilde{F}}$, and as a consequence $\underline{Q}_{\underline{\tilde{F}}, \overline{\tilde{F}}}(f) = \underline{Q}_{\underline{F}, \overline{F}}(f)$ for any function f in $\mathcal{L}([0, 1])$.

For any precise p-box described by a distribution function F , we will denote \underline{Q}_F and, for any $\alpha \in [0, 1]$, $\underline{Q}_{\alpha, F}$ the particular lower expectations $\underline{Q}_{[\underline{F}, \overline{F}]}$ and \underline{Q}_α induced by F . Similarly, we will denote L_F and U_F the mappings L and U induced by F , and Γ_F the associated multi-valued mapping.

Our next proposition shows that the lower expectations $\underline{Q}_{[\underline{F}, \overline{F}]}(f)$ and $\underline{P}_{[\underline{F}, \overline{F}]}(f)$ do not coincide on all functions:

Proposition F.8. *The lower previsions $\underline{Q}_{[\underline{F}, \overline{F}]}$ and $\underline{P}_{[\underline{F}, \overline{F}]}$ coincide on the class \mathcal{K} of events if and only if \underline{F} and \overline{F} are right-continuous.*

Proof. First consider a precise p-box F . Given the definition of lower expectations \underline{Q}_F and $\underline{Q}_{\alpha, F}$,

$$\underline{Q}_{\alpha, F}([0, x]) = 1 \Leftrightarrow U_F(\alpha) \leq x \Leftrightarrow F(y) > \alpha \quad \forall y > x;$$

from this, we can deduce that $[0, F(x^+)) \subseteq \{\alpha : \underline{Q}_{\alpha, F}([0, x]) = 1\} \subseteq [0, F(x^+)]$, whence $\underline{Q}_F([0, x]) = F(x^+)$ (whereas $\underline{P}_F([0, x]) = F(x)$). Similarly,

$$\underline{Q}_{\alpha, F}((x, 1]) = 1 \Leftrightarrow L_F(\alpha) > x \Leftrightarrow \exists y > x \text{ s.t. } F(y) < \alpha,$$

whence $(F(x^+), 1] \subseteq \{\alpha : \underline{Q}_{\alpha, F}((x, 1]) = 1\} \subseteq [F(x^+), 1]$ and consequently $\underline{Q}_F((x, 1]) = 1 - F(x^+)$ (whereas $\underline{P}_F((x, 1]) = 1 - F(x)$). We deduce from this that the coherent lower expectation \underline{Q}_F coincides with \underline{P}_F on \mathcal{K} if and only if F is right-continuous. An equivalent reasoning can be separately used on \underline{F} and \overline{F} when the p-box is not reduced to a precise one. \square

This show that $\underline{Q}_{[\underline{F}, \overline{F}]}$ and $\underline{P}_{[\underline{F}, \overline{F}]}$ do not coincide in general, and thus that the two models are not equivalent. The next example show that this is still true even if we consider only right-continuous distribution functions F .

Example F.1. Let us consider the distribution function F on $[0, 1]$ given by $F(x) = x$ for all x , and let A be the set of the irrational numbers on $[0, 1]$. It follows from the definition of

L_F and U_F that $L_F(\alpha) = U_F(\alpha) = \alpha$ for all α in $[0, 1]$, whence $\Gamma(\alpha) = \{\alpha\}$ for all α and consequently $\underline{Q}_\alpha(A) = \mathbb{I}_A(\alpha)$. Hence, $\underline{Q}_{[E, \bar{F}]}(A) = \int_0^1 A d\alpha_* = 1$.

On the other hand, it follows from Theorem F.4 that $\underline{E}_F(A) = \sup\{P_F(C) : C \subseteq A, C \in \mathcal{H}\}$. But since the only element of \mathcal{H} which is included in A is the empty set, we deduce that $\underline{E}_F(A) = 0$. ♦

F.4.2 Lower expectation on continuous functions

We now proceed to demonstrate that the lower expectation given by our random set expression coincide with \underline{E} when the function f is continuous. For this, we shall first state that any (precise) distribution function has a unique expectation when considering continuous functions. A similar result in the case of distribution functions on the unit interval was established in [149, Section 3.3] (where it is also shown that, when considering non-continuous functions, the lower and upper expectations induced by a precise distribution function do not forcefully coincide).

Let f be a continuous function on $[0, 1]$, i.e., a function such that $f(d^-) = f(d) = f(d^+)$ for any $d \in [0, 1]$. Let on the other hand F be a cdf, and let \bar{E} be the dual upper expectation of \underline{E} , given by $\bar{E}(f) = -\underline{E}(-f)$ for all functions f . Then \bar{E} is the upper envelope of the set of expectations given F to all functions, and $\underline{E}(f) = \bar{E}(f)$ if and only if expectation of F to f is unique. We have the following proposition:

Proposition F.9. *Let F be a precise distribution function and let f be a continuous gamble. Then $\underline{E}(f) = \bar{E}(f)$.*

Using this Proposition, we are going to prove that $\underline{Q}_{[E, \bar{F}]}$ coincides with \underline{E} on continuous gambles. We will first relate the functional $\underline{Q}_{[E, \bar{F}]}$ to the functional that we can define for each of the distribution functions F that belong to $\Phi(\underline{E}, \bar{F})$. Let F be such a distribution function, we then have the two following lemmas:

Lemma F.5. $\underline{Q}_\alpha(f) = \inf_{F \in \Phi(\underline{E}, \bar{F})} \underline{Q}_{\alpha, F}(f)$ for any $\alpha \in [0, 1]$ and any $f \in \mathcal{L}([0, 1])$.

Lemma F.6. *Let \tilde{F} be the right-continuous approximation of F . Their expectations to continuous functions f in $\mathcal{L}[0, 1]$ coincide, that is $\underline{E}_F(f) = \underline{E}_{\tilde{F}}(f)$ for any continuous functions f in $\mathcal{L}[0, 1]$.*

These two lemmas together allow us to state the following theorem:

Theorem F.10. For any continuous function f on $[0, 1]$, $\underline{E}(f) = \underline{Q}_{\underline{E}, \overline{F}}(f)$.

Proof. For any continuous function f ,

$$\underline{E}(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{E}_F(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{E}_{\tilde{F}}(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{Q}_{\tilde{F}}(f),$$

where the second equality follows from Lemma F.6 and the third from the fact that $\underline{Q}_{\tilde{F}}$ is the expectation induced by $\underline{P}_{\tilde{F}}$ (from Proposition F.8), which is unique for continuous functions (Proposition F.9). Since $\underline{Q}_{\underline{E}, \tilde{F}}(f) = \underline{Q}_{\underline{E}, \overline{F}}(f)$ for any function f on $[0, 1]$,

$$\begin{aligned} \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{Q}_{\tilde{F}}(f) &= \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{Q}_F(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \int_0^1 \underline{Q}_{\alpha, F}(f) d\alpha \\ &\geq \int_0^1 \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{Q}_{\alpha, F}(f) d\alpha = \int_0^1 \underline{Q}_{\alpha}(f) d\alpha = \underline{Q}_{\underline{E}, \overline{F}}(f), \end{aligned}$$

where the one but last equality follows from Lemma F.5. Hence, $\underline{E}(f) \geq \underline{Q}_{\underline{E}, \overline{F}}(f)$ for any continuous function f .

By Proposition F.8, $\underline{Q}_{\underline{E}, \tilde{F}}$ is an extension of $\underline{P}_{\underline{E}, \tilde{F}}$ to events in \mathcal{K} , and therefore dominates $\underline{E}_{\underline{E}, \tilde{F}}$ on all functions. But Lemma F.6 implies that

$$\underline{E}_{\underline{E}, \overline{F}}(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{E}_F(f) = \inf_{F \in \phi(\underline{E}, \overline{F})} \underline{E}_{\tilde{F}}(f) = \inf_{F \in \phi(\underline{E}, \tilde{F})} \underline{E}_F(f) = \underline{E}_{\underline{E}, \tilde{F}}(f)$$

for any continuous function f . Here, the third equality follows from the fact that given $F \in \phi(\underline{E}, \tilde{F})$, there exists some $F' \in \phi(\underline{E}, \overline{F})$ such that $\tilde{F} = \tilde{F}'$, which implies that $\underline{E}_F = \underline{E}_{F'} = \underline{E}_{\tilde{F}'}$ on continuous functions. We deduce that

$$\underline{E}_{\underline{E}, \overline{F}}(f) = \underline{E}_{\underline{E}, \tilde{F}}(f) \leq \underline{Q}_{\underline{E}, \tilde{F}}(f) = \underline{Q}_{\underline{E}, \overline{F}}(f)$$

for any continuous function f , and consequently we have the equality. □

Consequently, we can safely use the random set induced by Γ to compute lower expectations induced by $[\underline{E}, \overline{F}]$ on continuous functions. Recall that Example F.1 indicates that this equality between $\underline{Q}_{[\underline{E}, \overline{F}]}$ and \underline{E} does not extend in general to all gambles. Nevertheless, the case where expectations have to be computed for continuous functions is general enough to be of practical interest.

F.5 Conclusions

In this appendix, we have mainly explored an extension of generalized p-boxes presented in Section 3.2, that is the case where generalized p-boxes are defined on totally ordered spaces that are no longer necessarily finite. This setting encompasses in one sweep both generalized p-boxes on such spaces, p-boxes defined on the real line and on product spaces of the real line (provided elements are totally ordered).

We have shown that many of the results from Section 3.2, but not all, could be extended to this more general case, however not without introducing many mathematical subtleties.

In particular, generalized p-boxes on totally ordered spaces remain completely monotone, and this allows to give a closed and manageable form of the lower expectation induced by such a p-box in term of a Choquet integral. We have also shown that the correspondence with random sets do not hold anymore in general, thus demonstrating that one has to be cautious when extending results to more general cases. However, the correspondence still holds when computing lower expectations of continuous functions.

Other interesting results are those showing that a generalized p-box on totally ordered spaces is totally characterized by the values it takes on the open sets of the upper-limit topology, and that lower expectations induced by a generalized p-box can be approximated by limits of degenerate p-boxes.

There are still a few open problems and future lines of research steaming from this study; one would be the study, for generalized p-boxes defined on totally ordered spaces, of the properties we have established in Section F.4. A number of complications arise in that case because of the topological structure within \mathcal{X} . A more general open problem would be the connection of generalized p-boxes with other uncertainty models, like clouds. In particular, they could be useful model when linguistic assessments are both positive and negative assessments (see [45]). A first step, which perhaps would not be too difficult to do, would be to extend our results to completely (pre-)ordered spaces, that is to drop the property of asymmetry on the relation on \mathcal{X} .

Appendix G

(Ir)relevance statements, structural judgments and event trees

In Section 5.2.4, we studied how the notion of forward irrelevance could be related to the notion of independence in event-trees. We saw that, for particular event-trees (i.e., standard ones) the two notions were equivalent.

As discussed in Section 5.2.4, forward irrelevance statements are likely to be the most useful and sensible type of independence to use in a number of situations, particularly those involving uncertain processes. However, other statements of independence, or even of structural properties of the uncertainty about variables, are likely to be more useful in some other situations. This is why we now briefly discuss two related matters:

- How some results relating forward irrelevance with other notions of independence, namely strong and repetition independence, translate in event trees.
- How the symmetric notion of epistemic independence, discussed by Walley [203, Ch.9] could be set into event trees as well.

G.1 Forward irrelevance, strong independence and repetition independence

Recall that we consider variables X_1, \dots, X_N assuming values on $\mathcal{X}_1, \dots, \mathcal{X}_N$, and that variables X_k are indexed following a "time" index k (i.e., they form a process), that is, the value of

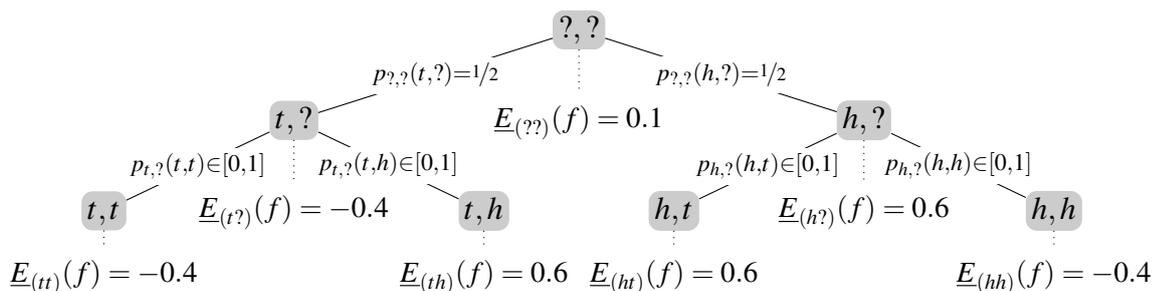
X_k is always known before.

By using the marginal extension [148], it is possible to build a joint model $\mathcal{P}_{(1:N)}$ by combining local credal sets $\mathcal{P}_{x_{(1:k)}} = \{P(\cdot|x_{(1:k)})\}$ defined on $\mathcal{X}_{\parallel+\infty}$ and for all $k = 0, \dots, N - 1$, $x_{(1:k)} \in \mathcal{X}_{(1:k)}$, with some abuse of notations for $x_{(1:0)}$, meaning that nothing has been observed yet. These credal sets are equivalent to local uncertainty models concerning the value of X_{k+1} , knowing that $X_{(1:k)} = x_{(1:k)}$. A statement of *forward irrelevance* allow to reduce the number of local credal sets to assess, since it comes down to consider that $\mathcal{P}_{x_{(1:k)}} = \mathcal{P}_{k+1}$, for any $x_{(1:k)} \in \mathcal{X}_{(1:k)}$ and for all $k = 0, \dots, N - 1$. In other words, our local predictive model about the value of X_{k+1} do not depend of values of variables $X_{(1:k)}$. In the corresponding standard event tree, this means that our local models $\mathcal{P}_{x_{(1:k)}}$ attached to situation $x_{(1:k)}$ do not depend on the situation we have reached in the tree.

Since we have equivalence between independence in standard event trees and forward irrelevance of variables X_1, \dots, X_N , we can use the results relating forward irrelevance to strong and repetition independence [50] to discuss these two notions inside standard event trees.

First, strong independence between marginal credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$ can be obtained by choosing, for each $k = 1, \dots, N - 1$ the *same* probability $P(\cdot|x_{(1:k)})$ in \mathcal{P}_k for all $x_{(1:k)}$ in $\mathcal{X}_{(1:k)}$, that is by assuming, in addition to forward irrelevance, a functional dependence between the sets $\mathcal{P}_{x_{(1:k)}}$. The following example shows that this added constraints on credal sets \mathcal{P}_k indeed lead to tighter results.

Example G.1. Again, we illustrate the concept of strong independence with an event-tree describing two successive flipping of coins. We consider two successive flips of different coins. The first coin is known to be fair, and \mathcal{P}_1 reduces to probability $P(h) = 1/2, P(t) = 1/2$, while nothing is known about the second coin, which could have two identical sides, and \mathcal{P}_2 is such that $P(h) \in [0, 1]$. Now, consider the function such that $f(t, h) = f(h, t) = 0.6$ and $f(t, t) = f(h, h) = -0.4$. On the tree is also indicated the lower expectations of this function obtained by assuming strong independence between the two flips.



The fact that strong independence lead to a tighter uncertainty is clearly visible in the

example, since $\underline{E}_{(??)}(f) = 0.1$, while simple forward irrelevance would have lead to $\underline{E}_{(??)}(f) = -0.4$. This is due to the fact that strong independence enforce $p_{h,?}(h,t) = p_{t,?}(t,t)$.

This also shows that assuming strong independence generally imply that backward recursion and local computations cannot be used any longer to computer lower expectations. This means that computing with an assumption of strong independence becomes more complex than computing with an assumption of epistemic independence, since one has to consider lower expectations generated by every possible combinations of extreme points in credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$.

Second, when $\mathcal{X}_1 = \dots = \mathcal{X}_N = \mathcal{X}$ and $\mathcal{P}_1 = \dots = \mathcal{P}_N = \mathcal{P}$, repetition independence is obtained by choosing the same probability $P(\cdot)$ and $P(\cdot|x_{(1:k)})$ in \mathcal{P} for all $x_{(1:k)}$ in $\mathcal{X}_{(1:k)}$ **and** all $1 \leq k \leq N - 1$. As for strong independence, local computations cannot be used to computing lower expectations with an assumption of repetition independence, nevertheless an assumption of repetition independence requires less computational effort than one of strong independence, since one only has to consider one computation per extreme points of \mathcal{P} .

G.2 Towards a characterization of epistemic independence in event-trees

We now examine, on the simple example of two successive coin flipping, how notions of epistemic independence, the symmetric counterpart of epistemic irrelevance, could be related to specific event-trees. Note that materials presented in this section are still very preliminary.

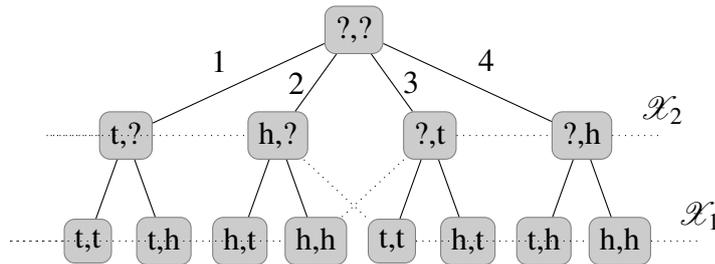
Let us consider two variables X_1, X_2 assuming values in $\mathcal{X}_1, \mathcal{X}_2$. Given a joint uncertainty model on these two variables, an assessment of epistemic independence can be translated by the fact that $\underline{E}(f_1|x_2) = \underline{E}(f_1)$ and $\underline{E}(f_2|x_1) = \underline{E}(f_2)$ for any $x_1 \in \mathcal{X}_1$ and $x_2 \in \mathcal{X}_2$, with f_i an \mathcal{X}_i -measurable function, and \underline{E} the lower expectations associated to situations in an event-tree.

Now, to make sense of the notion of epistemic independence in event trees, we need to build trees such that the values of X_1, X_2 can be observed in any possible ordering. In our small example, we will also need to use the notion of *weak independence* [179, Chapter 8]: two variables in a tree are *weakly independent* if there is no move that influences them both¹, i.e., if for any situation s outside the initial one, we have either $\underline{E}_s(f_1) = \underline{E}_{m(s)}(f_1)$ or $\underline{E}_s(f_2) = \underline{E}_{m(s)}(f_2)$ for any functions f_1, f_2 , and with $m(s)$ the *mother* of s , that is the situation immediately pre-

¹But different moves originating from one situation can influence different variables

ceding s .

Example G.2. Let us consider the following event tree, describing two successive flips of two coins, and allowing for the two flips to be observed in any possible ordering. Labels are again explicit enough. Moves for which both variables X_1, X_2 could be influenced have been numbered (other moves can only influence one variable, since the value of the other is already known). Also pictured are the cuts where the value of a particular variable is revealed.



The following table summarizes under which conditions variables X_1 (first flip), X_2 (second flip) are influenced by each move (Moves 1, 2, 3, 4 being the only ones for which both variables can be influenced):

| Infl. | 1 | 2 | 3 | 4 |
|-------|--|--|--|--|
| X_1 | Always | Always | $\underline{E}_{?,t}(f_1) \neq \underline{E}(f_1)$ | $\underline{E}_{?,h}(f_1) \neq \underline{E}(f_1)$ |
| X_2 | $\underline{E}_{t,?}(f_2) \neq \underline{E}(f_2)$ | $\underline{E}_{h,?}(f_2) \neq \underline{E}(f_2)$ | Always | Always |

Note that X_1, X_2 are weakly independent in the above tree if and only if inequalities of the above table turn into equalities. It can be seen that this is equivalent to epistemic independence between X_1, X_2

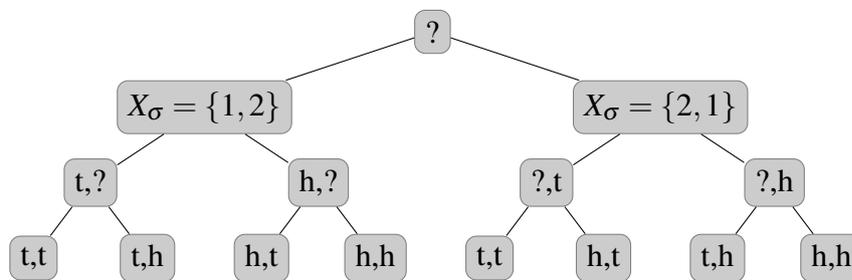
The above example suggest that, when considering variables X_1, \dots, X_N assuming values in $\mathcal{X}_1, \dots, \mathcal{X}_N$, epistemic independence could be related to weak independence in an event tree where $d(\square) = \mathcal{X}_{(1:N)}$, and where, at each step of the tree the value of a variable X_i is known, and the daughter of a situation consist of the cartesian products of all spaces of variables whose value is not yet known.

However, using such trees to characterize epistemic independence do not look very appealing at first sight, for various reasons:

- It is not obvious which form the immediate predictive model \mathcal{P}_1 , bearing on $\mathcal{X}_{(1:N)}$, should have in order to ensure epistemic independence of X_1, \dots, X_N .

- As emphasized by Shafer [179], the notion of weak independence is rather unstable, compared to the stronger notion of independence.
- Built trees do not appear very "intuitive" at first sight, and poorly related to the standard trees used to characterize forward irrelevance.

However, Shafer [179] also mentions that it is possible to transform a tree so that weak independence between two variables become classical independence (i.e., not influenced by the same situation). It seems possible to do something likewise in our case, and this would lead to the following tree:



That is, a tree where we have introduced an additional variable X_σ , that we will call the *sorting variable*, and which would determine in which order are revealed the variables X_1, \dots, X_N . Such a variable do not increase the final dimension of the final space, but would allow to relate more easily epistemic independence to local predictive models. Note that each subtree after the sorting variable would be equivalent to a standard tree where the order of observation is determined by the value of X_σ . This suggests that epistemic independence would be equivalent to forward irrelevance in each of these sub-trees.

Bibliography

- [1] J. Abellan and M. Gomez. Measures of divergence on credal sets. *Fuzzy Sets and System*, 157(11), 2006.
- [2] J. Abellan and G. Klir. Additivity of uncertainty measures on credal sets. *Int. J. of General Systems*, 34:691–713, 2005.
- [3] J. Abellan and S. Moral. A non-specificity measure for convex sets of probability distributions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 8:357–367, 2000.
- [4] J. Aczel. *Lectures on functional equations and their applications*. Academic Press, NY, 1966.
- [5] D. A. Alvarez. On the calculation of the bounds of probability of events using infinite random sets. *I. J. of Approximate Reasoning*, 43:241–267, 2006.
- [6] N. B. Amor, K. Mellouli, S. Benferhat, D. Dubois, and H. Prade. A theoretical framework for possibilistic independence in a weakly ordered setting. *I. J. of uncertainty, fuzziness and Knowledge-Based Systems*, 10:117–155, 2002.
- [7] T. Augustin. Generalized basic probability assignments. *I. J. of General Systems*, 34(4):451–463, 2005.
- [8] A. Ayoun and P. Smets. Data association in multi-target detection using transferable belief model. *I. J. of intelligent systems*, 16(10), October 2001.
- [9] G. Bardossy and J. Fodor. *Evaluation of uncertainties and risks in geology: new mathematical approaches for their handling*. Springer, Berlin, 2004.
- [10] C. Baudrit, I. Couso, and D. Dubois. Joint propagation of probability and possibility in risk analysis: towards a formal framework. *Int. J. of Approximate Reasoning*, 45:82–105, 2007.

- [11] C. Baudrit and D. Dubois. Practical representations of incomplete probabilistic knowledge. *Computational Statistics and Data Analysis*, 51(1):86–108, 2006.
- [12] C. Baudrit, D. Guyonnet, and D. Dubois. Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. *IEEE Trans. Fuzzy Systems*, 14:593–608, 2006.
- [13] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation Functions: A Guide for Practitioners*. Springer, Berlin, 2008.
- [14] R. Bellman. *Dynamic Programming*. 1957.
- [15] S. Benferhat, D. Dubois, S. Kaci, and H. Prade. Bipolar possibility theory in preference modelling: representation, fusion and optimal solutions. *Information Fusion*, 7:135–150, 2006.
- [16] J. O. Berger. An overview of robust Bayesian analysis. *Test*, 3:5–124, 1994. With discussion.
- [17] J. Bernoulli. *Ars Conjectandi*. Thurnisius, Basel, 1713.
- [18] I. Bloch. Information combination operators for data fusion : A comparative review with classification. *IEEE Trans. on Syst., Man, and Cybern. A*, 26(1):52–67, January 1996.
- [19] M. Cattaneo. Combining belief functions issued from dependent sources. In *Proc. Third International Symposium on Imprecise Probabilities and Their Application (ISIPTA'03)*, pages 133–147, Lugano, Switzerland, 2003.
- [20] M. Cattaneo. Likelihood-based statistical decisions. In *Proc. 4th International Symposium on Imprecise Probabilities and Their Applications*, pages 107–116, 2005.
- [21] A. Chateauneuf. Combination of compatible belief functions and relation of specificity. In *Advances in the Dempster-Shafer theory of evidence*, pages 97–114. John Wiley & Sons, Inc, New York, NY, USA, 1994.
- [22] A. Chateauneuf and J.-Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Mobius inversion. *Mathematical Social Sciences*, 17(3):263–283, 1989.

- [23] E. Chojnacki, J. Baccou, and S. Destercke. Numerical sensitivity and efficiency in the treatment of epistemic and aleatory uncertainty. In *Proc. Fifth Int. Conf. on Sensitivity Analysis of Model Output*, 2007.
- [24] E. Chojnacki, J. Baccou, and S. Destercke. Numerical sensitivity and efficiency in the treatment of epistemic and aleatory uncertainty. *Submitted to Int. J. of Intelligent Systems*, 2008.
- [25] G. Choquet. Theory of capacities. *Annales de l'institut Fourier*, 5:131–295, 1954.
- [26] R. Clemen and R. Winkler. Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19(2):187–203, 1999.
- [27] W. Conover. *Practical non-parametric statistic*. Wiley, New York, 3rd edition, 1999.
- [28] R. Cooke. *Experts in uncertainty*. Oxford University Press, Oxford, UK, 1991.
- [29] J. Cooper, S. Ferson, and L. Ginzburg. Hybrid processing of stochastic and subjective uncertainty. *Risk Analysis*, 16:785–791, 1996.
- [30] I. Couso. Independence concepts in evidence theory. In *Proc. of the 5th Int. Symp. on Imprecise Probability: Theories and Applications*, 2007.
- [31] I. Couso, S. Montes, and P. Gil. The necessity of the strong alpha-cuts of a fuzzy set. *Int. J. on Uncertainty, Fuzziness and Knowledge-Based Systems*, 9:249–262, 2001.
- [32] I. Couso, S. Montes, and P. Gil. *Statistical analysis, modeling and management of fuzzy data*, chapter Second order possibility measure induced by a fuzzy random variable, pages 127–144. Springer, 2002.
- [33] I. Couso, S. Moral, and P. Walley. A survey of concepts of independence for imprecise probabilities. *Risk Decision and Policy*, 5:165–181, 2000.
- [34] I. Couso and L. Sanchez. Higher order models for fuzzy random variables. *Fuzzy Sets and Systems*, 159:237–258, 2008.
- [35] F. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.
- [36] F. Cozman. Constructing sets of probability measures through kuznetsov's independence condition. In *Proc. 2nd Int. Symp. on Imprecise Probabilities and Their Applications*, 2001.

- [37] F. Cozman. Graphical models for imprecise probabilities. *I. J. of Approximate Reasoning*, 39:167–184, 2005.
- [38] F. Cozman and P. Walley. Graphoid properties of epistemic irrelevance and independence. *Annals of Mathematics and Artificial Intelligence*, 45:173–195, 2005.
- [39] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.
- [40] J. de Campos, M. Lamata, and S. Moral. Logical connectives for combining fuzzy measures. In Z. Ras and L. Saitta, editors, *Methodologies for Intelligent Systems*, volume 3, pages 11–18. North-Holland, Amsterdam, 1988.
- [41] L. de Campos and J. Huete. Independence concepts in possibility theory: Part i. *Fuzzy Sets and Systems*, 103:127–152, 1999.
- [42] L. de Campos, J. Huete, and S. Moral. Probability intervals: a tool for uncertain reasoning. *I. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2:167–196, 1994.
- [43] G. de Cooman. Possibility theory III: possibilistic independence. *International Journal of General Systems*, 25:353–371, 1997.
- [44] G. de Cooman. Precision-imprecision equivalence in a broad class of imprecise hierarchical uncertainty models. *J. of Statistical Planning and Inference*, 105:175–198, 2002.
- [45] G. de Cooman. A behavioural model for vague probability assessments. *Fuzzy Sets and Systems*, 154:305–358, 2005.
- [46] G. de Cooman and D. Aeyels. Supremum-preserving upper probabilities. *Information Sciences*, 118:173–212, 1999.
- [47] G. de Cooman and F. Hermans. Imprecise probability trees: bridging two theories of imprecise probability. *Submitted*, page 23 pages, 2007.
- [48] G. de Cooman and F. Hermans. On coherent immediate prediction: Connecting two theories of imprecise probabilities. In G. de Cooman, J. Vernarova, and M. Zaffalon, editors, *ISIPTA'07 - Proceedings of the fifth International Symposium on Imprecise Probability: Theories and Applications*, 2007.

- [49] G. de Cooman and E. Miranda. Weak and strong laws of large numbers for coherent lower previsions. *Journal of Statistical Planning and Inference*, 2006. Submitted for publication.
- [50] G. de Cooman and E. Miranda. Forward irrelevance. *Submitted*, page 31 pages, 2007.
- [51] G. de Cooman, M. Troffaes, and E. Miranda. n -monotone lower previsions and lower integrals. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. 4th International Symposium on Imprecise Probabilities and Their Applications*, 2005.
- [52] G. de Cooman, M. C. M. Troffaes, and E. Miranda. n -Monotone exact functionals. 2006. Submitted for publication.
- [53] G. de Cooman and P. Walley. A possibilistic hierarchical model for behaviour under uncertainty. *Theory and Decision*, 52:327–374, 2002.
- [54] M. Delgado, D. Sanchez, M. Martin-Bautista, and M. Vila. A probabilistic definition of a nonconvex fuzzy cardinality. *Fuzzy Sets and Systems*, 126:177–190, 2002.
- [55] F. Delmotte. Detection of defective sources in the setting of possibility theory. *Fuzzy Sets and Systems*, 158:555–571, 2007.
- [56] F. Delmotte and P. Borne. Modeling of reliability with possibility theory. *IEEE Trans. on Syst., Man, and Cybern. A*, 28(1):78–88, 1998.
- [57] A. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics*, 38:325–339, 1967.
- [58] T. Denoeux. Reasoning with imprecise belief structures. *Int. J. of Approximate Reasoning*, 20:79–111, 1999.
- [59] T. Denoeux. Modeling vague beliefs using fuzzy-valued belief structures. *Fuzzy Sets and Systems*, 116:167–199, 2000.
- [60] T. Denoeux. Constructing belief functions from sample data using multinomial confidence regions. *I. J. of Approximate Reasoning*, 42:228–252, 2006.
- [61] T. Denoeux. Conjunctive and disjunctive combination of belief functions induced by non-distinct bodies of evidence. *Artificial Intelligence*, 172:234–264, 2008.
- [62] S. Destercke and E. Chojnacki. Methods for the evaluation and synthesis of multiple sources of information applied to nuclear computer codes. *Nuclear Engineering and Design*, In Press, 2008.

- [63] S. Destercke and G. de Cooman. Relating epistemic irrelevance to event trees. In *Int. Conf. on Soft Methods in Probability and Statistics (SMPS)*, 2008.
- [64] S. Destercke and D. Dubois. A unified view of some representations of imprecise probabilities. In J. Lawry, E. Miranda, A. Bugarin, and S. Li, editors, *Int. Conf. on Soft Methods in Probability and Statistics (SMPS)*, Advances in Soft Computing, pages 249–257, Bristol, 2006. Springer.
- [65] S. Destercke, D. Dubois, and E. Chojnacki. Fusion d’opinions d’experts et theories de l’incertain. In *Proc. Rencontres Francophones sur la logique floue et ses applications*, 2006.
- [66] S. Destercke, D. Dubois, and E. Chojnacki. Cautious conjunctive merging of belief functions. In *Proc. Eur. Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, pages 332–343, 2007.
- [67] S. Destercke, D. Dubois, and E. Chojnacki. On the relationships between random sets, possibility distributions, p-boxes and clouds. In *Proc. 28th Linz Seminar on Fuzzy Set Theory*, 2007.
- [68] S. Destercke, D. Dubois, and E. Chojnacki. Relating practical representations of imprecise probabilities. In *Proc. 5th Int. Symp. on Imprecise Probabilities: Theories and Applications*, 2007.
- [69] S. Destercke, D. Dubois, and E. Chojnacki. Transforming probability intervals into other uncertainty models. In *Proc. European Society for Fuzzy Logic and Technology conference*, 2007.
- [70] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: II clouds. *Int. J. of Approximate Reasoning (in press)*, 2007.
- [71] S. Destercke, D. Dubois, and E. Chojnacki. Computing with generalized p-boxes: preliminary results. In *Proc. Information Processing and Management of Uncertainty in Knowledge-based systems (IPMU)*, 2008.
- [72] S. Destercke, D. Dubois, and E. Chojnacki. Possibilistic information fusion using maximal coherent subsets. *IEEE Trans. on Fuzzy Systems (in press)*, 2008.
- [73] S. Destercke, D. Dubois, and E. Chojnacki. Possibilistic information fusion using maximal coherent subsets. In *Proc. IEEE Int. Conf. On Fuzzy Systems (FUZZ’IEEE)*, 2008.

- [74] S. Destercke, D. Dubois, and E. Chojnacki. Unifying practical uncertainty representations: I generalized p-boxes. *Int. J. of Approximate Reasoning (In press)*, 2008.
- [75] D. Dubois, L. F. del Cerro, A. Herzig, and H. Prade. A roadmap of qualitative independence. In D. Dubois, H. Prade, and E. Klement, editors, *Fuzzy sets, logics, and reasoning about knowledge*. Springer, 1999.
- [76] D. Dubois, H. Fargier, and H. Prade. Multi-source information fusion: a way to cope with incoherences. In Cepadues, editor, *Proc. of French Days on Fuzzy Logic and Applications (LFA)*, pages 123–130, La rochelle, 2000. Cepadues.
- [77] D. Dubois, H. Fargier, and H. Prade. Multiple-sources information fusion - a practical inconsistency-tolerant approach. In *Proc. of 8th Information Processing and Management of Uncertainty in Knowledge-Based Systems Conference (IPMU)*, pages 123–130. Springer, 2000.
- [78] D. Dubois, L. Foulloy, G. Mauris, and H. Prade. Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. *Reliable Computing*, 10:273–297, 2004.
- [79] D. Dubois, P. Hajek, and H. Prade. Knowledge-driven versus data-driven logics. *J. of Logic, Language and Information*, 9:65–89, 2000.
- [80] D. Dubois, E. Kerre, R. Mesiar, and H. Prade. *Fundamentals of fuzzy sets*, chapter Fuzzy interval analysis, pages 483–581. Kluwer, Boston, 2000.
- [81] D. Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. New York, 1980.
- [82] D. Dubois and H. Prade. On several representations of an uncertain body of evidence. In M. Gupta and E. Sanchez, editors, *Fuzzy Information and Decision Processes*, pages 167–181. North-Holland, 1982.
- [83] D. Dubois and H. Prade. Evidence measures based on fuzzy information. *Automatica*, 21(5):547–562, 1985.
- [84] D. Dubois and H. Prade. A set-theoretic view on belief functions: logical operations and approximations by fuzzy sets. *I. J. of General Systems*, 12:193–226, 1986.
- [85] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York, 1988.

- [86] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.
- [87] D. Dubois and H. Prade. Fuzzy sets, probability and measurement. *European Journal of Operational Research*, 40:135–154, 1989.
- [88] D. Dubois and H. Prade. Aggregation of possibility measures. In J. Kacprzyk and M. Fedrizzi, editors, *Multiperson Decision Making using Fuzzy Sets and Possibility Theory*, pages 55–63. Kluwer, Dordrecht, the Netherlands, 1990.
- [89] D. Dubois and H. Prade. Consonant approximations of belief functions. *I.J. of Approximate reasoning*, 4:419–449, 1990.
- [90] D. Dubois and H. prade. Evidence, knowledge, and belief functions. *Int. J. of Approximate Reasoning*, 6:295–319, 1992.
- [91] D. Dubois and H. Prade. Random sets and fuzzy interval analysis. *Fuzzy Sets and Systems*, (42):87–101, 1992.
- [92] D. Dubois and H. Prade. When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49:65–74, 1992.
- [93] D. Dubois and H. Prade. Possibility theory in information fusion. In G. D. Riccia, H. Lenz, and R. Kruse, editors, *Data fusion and Perception*, volume CISM Courses and Lectures N 431, pages 53–76. Springer Verlag, Berlin, 2001.
- [94] D. Dubois and H. Prade. Quantitative possibility theory and its probabilistic connections. In P. Grzegorzewski, O. Hryniewicz, and M. Angeles-Gil, editors, *Soft Methods in Probability, Statistics and Data Analysis*, Advances in Soft Computing, pages 3–26. Physica Verlag, Heidelberg - Germany, 2002.
- [95] D. Dubois and H. Prade. Fuzzy elements in a fuzzy set. In *Proc. 10th Int. Fuzzy Syst. Assoc. (IFSA) Congress*, pages 55–60, 2005.
- [96] D. Dubois and H. Prade. Interval-valued fuzzy sets, possibility theory and imprecise probability. In *Proceedings of International Conference in Fuzzy Logic and Technology (EUSFLAT'05)*, Barcelona, September 2005.
- [97] D. Dubois and H. Prade. Gradual elements in a fuzzy set. *Soft Computing*, 12:165–175, 2008.

- [98] D. Dubois, H. Prade, and S. Sandri. On possibility/probability transformations. In *Proc. of the Fourth International Fuzzy Systems Association World Congress (IFSA'91)*, pages 50–53, Brussels, Belgium, 1991.
- [99] D. Dubois, H. Prade, and P. Smets. New semantics for quantitative possibility theory. In G. de Cooman, T. Fine, and T. Seidenfeld, editors, *ISIPTA'01, Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications*. Shaker Publishing, 2001.
- [100] D. Dubois, H. Prade, and P. Smets. A definition of subjective possibility. *Journal of Approximate Reasoning*, In Press, 2008.
- [101] D. Dubois and R. Yager. Fuzzy set connectives as combination of belief structures. *Information Sciences*, 66:245–275, 1992.
- [102] S. Ferson and L. Ginzburg. Hybrid arithmetic. In *Proc. of ISUMA/NAFIPS'95*, 1995.
- [103] S. Ferson, L. Ginzburg, and R. Akcakaya. Whereof one cannot speak: when input distributions are unknown. *Risk Analysis*, To appear.
- [104] S. Ferson, L. Ginzburg, V. Kreinovich, D. Myers, and K. Sentz. Constructing probability boxes and dempster-shafer structures. Technical report, Sandia National Laboratories, 2003.
- [105] S. Ferson and V. Kreinovich. Modeling correlation and dependence among intervals. Technical Report UTEP-CS-06-04, Univ. of Texas, El Paso Computer Science Department, 2006.
- [106] S. Ferson, W. Oberkampf, and L. Ginzburg. Validation of imprecise probability models. In *3rd International Workshop on Reliable Engineering Computing*, 2008.
- [107] T. Fetz. Sets of joint probability measures generated by weighted marginal focal sets. In F. Cozman, R. Nau, and T. Seidenfeld, editors, *Proc. 2st International Symposium on Imprecise Probabilities and Their Applications*, 2001.
- [108] B. Finetti. *Theory of probability*, volume 1-2. Wiley, NY, 1974. Translation of 1970 book.
- [109] J. Fortin, D. Dubois, and H. Fargier. Gradual numbers and their application to fuzzy interval analysis. *IEEE Transactions on Fuzzy Systems*, Accepted for publication, 2006.

- [110] S. French. Group consensus probability distributions : a critical survey. *Bayesian Statistics*, 2:183–202, 1985.
- [111] M. Fuchs and A. Neumaier. Potential based clouds in robust design optimization. *Journal of statistical theory and practice*, To appear, 2008.
- [112] C. Genest, S. Weerahandi, and J. Zidek. Aggregating opinions through logarithmic pooling. *Theory and Decision*, 17:61–70, 1984.
- [113] C. Genest and J. Zidek. Combining probability distributions: A critique and an annotated bibliography. *Statistical Science*, 1(1):114–135, February 1986.
- [114] I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153, 1989.
- [115] M. Grabisch, J. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics on operations research*, 25(2):157–178, 2000.
- [116] J. Hall. Uncertainty-based sensitivity indices for imprecise probability distributions. *Reliability Engineering and System Safety*, 91:1443–1451, 2006.
- [117] J. Helton, J. Johnson, W. Oberkampf, and C. Sallaberry. Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty. *Reliability Engineering and System Safety*, 91:1414–1434, 2006.
- [118] J. Helton and W. Oberkampf, editors. *Alternative Representations of Uncertainty, Special issue of Reliability Engineering and Systems Safety*, volume 85. Elsevier, 2004.
- [119] P. Huber. *Robust statistics*. Wiley, New York, 1981.
- [120] C. Huygens. *Œuvres complètes de Christiaan Huygens*. Martinus Nijhoff, Den Haag, 1888-1950. Twenty-two volumes.
- [121] R. Iman and W. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics*, 11(3):311–334, 1982.
- [122] J.-Y. Jaffray and M. Jeleva. Information processing under imprecise risk with the Hurwicz criterion. In *Proc. of the fifth Int. Symp. on Imprecise Probabilities and Their Applications*, 2007.
- [123] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter. *Applied Interval Analysis*. London, 2001.

- [124] M. Jouini and R. Clemen. Copula models for aggregating expert opinions. *Operations Research*, 44(3):444–457, 1996.
- [125] A. Jousselme, D. Grenier, and E. Bosse. A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001.
- [126] A. Kaufmann and M. Gupta. *Introduction to Fuzzy Arithmetic: Theory and Applications*. 1985.
- [127] E. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publisher, Dordrecht, 2000.
- [128] G. Klir. Uncertainty and information measures for imprecise probabilities : An overview. In *Proc. 1st International Symposium on Imprecise Probabilities and Their Applications*, Ghent, Belgium, 1999.
- [129] I. Kozine and L. Utkin. Processing unreliable judgments with an imprecise hierarchical model. *Risk Decision and Policy*, 7:325–339, 2002.
- [130] I. Kozine and L. Utkin. Constructing imprecise probability distributions. *I. J. of General Systems*, 34:401–408, 2005.
- [131] B. C. P. Kraan. *Probabilistic Inversion in Uncertainty Analysis and Related Topics*. PhD thesis, Delft Institute of Applied Mathematics, 2002.
- [132] E. Kriegler and H. Held. Utilizing random sets for the estimation of future climate change. *I. J. of Approximate Reasoning*, 39:185–209, 2005.
- [133] S. Kullback. *Information Theory and Statistics*. 1968.
- [134] V. P. Kuznetsov. *Interval Statistical Methods*. Radio i Svyaz Publ., 1991. (in Russian).
- [135] J. Lemmer and H. Kyburg. Conditions for the existence of belief functions corresponding to intervals of belief. In *Proc. 9th National Conference on A.I.*, pages 488–493, Anaheim, 1991.
- [136] I. Levi. *The Enterprise of Knowledge*. MIT Press, London, 1980.
- [137] D. Lindley. Reconciliation of probability distributions. *Operations Search*, 31(5):866–880, 1983.
- [138] D. V. Lindley. Scoring rules and the inevitability of probability. *International Statistical Review*, 50:1–26, 1982. With discussion.

- [139] L. Magne and D. Vasseur, editors. *Risques industriels. Complexite, incertitude et decision : une approche interdisciplinaire*, chapter Maitriser les incertitudes pour mieux gerer les risques, pages 219–260. Lavoisier, 2006.
- [140] R. Malouf. Maximal consistent subsets. *Computational Linguistics*, 33:153–160, 2007.
- [141] M. Masson and T. Denoeux. Inferring a possibility distribution from empirical data. *Fuzzy Sets and Systems*, 157(3):319–340, february 2006.
- [142] K. McConway. Marginalization and linear opinion pools. *Journal of the American Statistical Association*, 76(374):410–414, June 1981.
- [143] J. Mendel. *Uncertain Rules-Based Fuzzy Logic Systems: Introduction and New Directions*. Prentice Hall, 2001.
- [144] E. Miranda. A survey of the theory of coherent lower previsions. *Int. J. of Approximate Reasoning*, In press, 2008.
- [145] E. Miranda, I. Couso, and P. Gil. Extreme points of credal sets generated by 2-alternating capacities. *I. J. of Approximate Reasoning*, 33:95–115, 2003.
- [146] E. Miranda, I. Couso, and P. Gil. Random sets as imprecise random variables. *Journal of Mathematical Analysis and Applications*, 307, 2005. 32–47.
- [147] E. Miranda and G. de Cooman. Epistemic independence in numerical possibility distribution. *Int. J. of Approximate Reasoning*, 32:23–42, 2003.
- [148] E. Miranda and G. de Cooman. Marginal extension in the theory of coherent lower previsions. *I.J. of Approximate Reasoning*, 46:188–225, 2007.
- [149] E. Miranda, G. de Cooman, and E. Quaeghebeur. Finitely additive extensions of distribution functions and moment sequences: the coherent lower prevision approach. *International Journal of Approximate Reasoning*, 2007. In press.
- [150] E. Miranda, M. Troffaes, and S. Destercke. Generalised p-boxes on totally ordered spaces. In *Proc. of the fourth international conference on soft methods in probabilities and statistics (SMPS)*, 2008.
- [151] I. Molchanov. *Theory of Random Sets*. Springer, London, 2005.
- [152] R. Moore. *Methods and applications of Interval Analysis*. SIAM Studies in Applied Mathematics. SIAM, Philadelphia, 1979.

- [153] S. Moral. Epistemic irrelevance on sets of desirable gambles. *Ann. Math. Artif. Intell.*, 45:197–214, 2005.
- [154] S. Moral and J. Sagrado. Aggregation of imprecise probabilities. In B. BouchonMeunier, editor, *Aggregation and Fusion of Imperfect Information*, pages 162–188. Physica-Verlag, Heidelberg, 1997.
- [155] P. Morris. Combining expert judgments : a bayesian approach. *Management Science*, 23:679–693, 1977.
- [156] A. Mosleh and G. Apostolakis. The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk Analysis*, 6(4):447–461, 1986.
- [157] R. Nau. The aggregation of imprecise probabilities. *Journal of Statistical Planning and Inference*, 105:265–282, 2002.
- [158] R. Nelsen. Copulas and quasi-copulas: An introduction to their properties and applications. In E. Klement and R. Mesiar, editors, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*, chapter 14. Elsevier, 2005.
- [159] A. Neumaier. Clouds, fuzzy sets and probability intervals. *Reliable Computing*, 10:249–272, 2004.
- [160] OCDE. Bemuse phase iii report: Uncertainty and sensitivity analysis of the loft 12-5 test. Technical Report NEA/NCIS/R(2007)4, NEA, May 2007.
- [161] M. Oussalah. Study of some algebraic properties of adaptative combination rules. *Fuzzy sets and systems*, 114:391–409, 2000.
- [162] M. Oussalah. On the use of hamacher’s t-norms family for information aggregation. *Information Science*, 153:107–154, 2003.
- [163] M. Oussalah, H. Maaref, and C. Barret. From adaptative to progressive combination of possibility distributions. *Fuzzy sets and systems*, 139:559–582, 2003.
- [164] C. Pappis and N. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56:171–174, 1993.
- [165] Z. Pawlak. *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic, Dordrecht, 1991.

- [166] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.
- [167] S. Petit-Renaud and T. Denoeux. Nonparametric regression analysis of uncertain and imprecise data using belief functions. *I. J. of Approximate Reasoning*, 35:1–28, 2004.
- [168] M. Puri and D. Ralescu. Fuzzy random variables. *J. Math. Anal. Appl.*, 114:409–422, 1986.
- [169] E. Quaeghebeur and G. D. Cooman. Extreme lower probabilities. *Fuzzy Sets and Systems*, In press, 2008.
- [170] E. Raufaste, R. D. S. Neves, and C. Mariné. Testing the descriptive validity of possibility theory in human judgments of uncertainty. *Artificial Intelligence*, 148:197–218, 2003.
- [171] H. Regan, S. Ferson, and D. Berleant. Equivalence of methods for uncertainty propagation of real-valued random variables. *I. J. of Approximate Reasoning*, 36:1–30, 2004.
- [172] H. Reichenbach. *The Direction of Time*. 1956.
- [173] N. Rescher and R. Manor. On inference from inconsistent premises. *Theory and Decision*, 1:179–219, 1970.
- [174] S. Sandri, D. Dubois, and H. Kalfsbeek. Elicitation, assessment and pooling of expert judgments using possibility theory. *IEEE Trans. on Fuzzy Systems*, 3(3):313–335, August 1995.
- [175] L. Savage. *Foundations of statistics*. Wiley, NY, 1954.
- [176] E. Schechter. *Handbook of Analysis and Its Foundations*. Academic Press, San Diego, CA, 1997.
- [177] G. L. S. Shackle. *Decision, Order and Time in Human Affairs*. Cambridge University Press, Cambridge, 1961.
- [178] G. Shafer. *A mathematical Theory of Evidence*. Princeton University Press, New Jersey, 1976.
- [179] G. Shafer. *The art of causal conjecture*. MIT Press, Cambridge, Massachusetts, 1996.
- [180] G. Shafer and V. Vovk. *Probability and finance: it's only a game!* Wiley, New York, 2001.

- [181] L. Shapley. A value for n-person games. In *Contributions to the theory of games*, pages 307–317. Princeton University Press, 1953.
- [182] P. Smets. The degree of belief in a fuzzy event. *Information Science*, 25:1–19, 1981.
- [183] P. Smets. The transferable belief model and other interpretations of dempster-shafer’s model. In *Proc. of the Sixth Annual Confernce on Uncertainty in Artifical Intelligence*, pages 375–384, 1990.
- [184] P. Smets. The canonical decomposition of a weighted belief. In *Proc. Int. Joint. Conf. on Artificial Intelligence*, pages 1896–1901, Montreal, 1995.
- [185] P. Smets. Decision making in a context where uncertainty is represented by belief functions. In R. Srivastava and T. Mock, editors, *Belief Functions in Business Decisions*, pages 17–61. Physica-Verlag, Heidelberg, 2002.
- [186] P. Smets. Belief functions on real numbers. *I. J. of Approximate Reasoning*, 40:181–223, 2005.
- [187] P. Smets. Decision making in the tbm: the necessity of the pignistic transformation. *I.J. of Approximate Reasoning*, 38:133–147, 2005.
- [188] P. Smets. Analyzing the combination of conflicting belief functions. *Information Fusion*, 8:387–412, 2006.
- [189] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66:191–234, 1994.
- [190] M. Smithson. Fuzzy set inclusion: linking fuzzy set methods with mainstreams techniques. *Sociological Methods and Research*, 33:431–461, 2005.
- [191] A. Tarantola. *Inverse Problem Theory and methods for model parameters estimation*. SIAM, Philadelphia, 2005.
- [192] F. Tonon and S. Chen. Inclusion properties for random relations under the hypotheses of stochastic independence and non-interactivity. *Int. J. of General Systems*, 34:615–624, 2005.
- [193] M. Troffaes. *Optimality, Uncertainty, and Dynamic Programming with Lower Previsions*. PhD thesis, Ghent University, Ghent, Belgium, 2005.
- [194] M. Troffaes. Generalising the conjunction rule for aggregating conflicting expert opinions. *I. J. of Intelligent Systems*, 21(3):361–380, March 2006.

- [195] M. Troffaes. Decision making under uncertainty using imprecise probabilities. *Int. J. of Approximate Reasoning*, 45:17–29, 2007.
- [196] T. Trucano, L. Swiler, T. Igusa, W. Oberkampf, and M. Pilch. Calibration, validation, and sensitivity analysis: What’s what. *Reliability Engineering and System Safety*, 91:1331–1357, 2006.
- [197] L. Utkin. Risk analysis under partial prior information and non-monotone utility functions. *I. J. of Information Technology and Decision Making*, To appear.
- [198] L. Utkin and T. Augustin. Powerful algorithms for decision making under partial prior information and general ambiguity attitudes. In *Proc. of the fourth Int. Symp. on Imprecise Probabilities and Their Applications*, 2005.
- [199] L. Utkin and S. Destercke. Computing expectations with p-boxes: two views of the same problem. In *Proc. of the fifth Int. Symp. on Imprecise Probabilities and Their Applications*, 2007.
- [200] P. Vicig. Epistemic independence for imprecise probabilities. *Int. J. of approximate reasoning*, 24:235–250, 2000.
- [201] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.
- [202] P. Walley. The elicitation and aggregation of beliefs. Technical report, University of Warwick, 1982.
- [203] P. Walley. *Statistical reasoning with imprecise Probabilities*. Chapman and Hall, New York, 1991.
- [204] P. Walley. Measures of uncertainty in expert systems. *Artificial Intelligence*, 83:1–58, 1996.
- [205] P. Walley. Statistical inference based on a second-order possibility distribution. *Int. J. of General Systems*, 26:337–383, 1997.
- [206] P. Walley. Towards a unified theory of imprecise probability. In *Proc. of the first Int. Symp. on Imprecise Probabilities and Their Applications*, 1999.
- [207] P. Walley and T. Fine. Towards a frequentist theory of upper and lower probability. *Annals of statistics*, 10:741–761, 1982.

- [208] A. Wallner. Extreme points of coherent probabilities in finite spaces. *Int. J. of Approximate Reasoning*, 44:339–357, 2007.
- [209] R. Williamson and T. Downs. Probabilistic arithmetic i : Numerical methods for calculating convolutions and dependency bounds. *I. J. of Approximate Reasoning*, 4:8–158, 1990.
- [210] N. Wilson. *Handbook of Defeasible Reasoning and Uncertainty Management. Vol. 5: Algorithms*, chapter Algorithms for Dempster-Shafer Theory, pages 421–475. Kluwer Academic, 2000.
- [211] R. Winkler. The consensus of subjective probability distributions. *Management Science*, 15(2), October 1968.
- [212] R. Yager. Generalized probabilities of fuzzy events from fuzzy belief structures. *Information sciences*, 28:45–62, 1982.
- [213] R. Yager. On ordered wieghted averaging aggregation operators in multicriteria decision making. *IEEE Trans. on Syst., Man, and Cybern.*, 18:183–190, 1988.
- [214] R. Yager and D. Filev. Including probabilistic uncertainty in fuzzy logic controller modeling using dempster-shafer theory. *IEEE Trans. on Systems, Man and Cybernet.*, 25(8):1221–1230, 1995.
- [215] R. R. Yager. New modes of owa information fusion. *I.J. of intelligent systems*, 13:183–190, 1998.
- [216] B. B. Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. the marginal case. *I. J. of Approximate Reasoning*, 29(1):47–70, 2002.
- [217] J. Yen. Generalizing the dempster-shafer theory to fuzzy sets. *IEEE Trans. on Systems, Man and Cybernet.*, 20(3):559–569, 1990.
- [218] L. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-i. *Information Sciences*, 8:199–249, 1975.
- [219] L. Zadeh. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1:3–28, 1978.
- [220] L. Zadeh. Fuzzy sets and information granularity. In R. Ragade, M. Gupta, and R. Yager, editors, *Advances in Fuzzy Sets Theory and Applications*, pages 3–18. North Holland, Amsterdam, 1979.

- [221] L. Zadeh. Fuzzy probabilities. *Inf. Processing and Management*, 20:363–372, 1984.
- [222] M. Zaffalon. The naive credal classifier. *J. Probabilistic Planning and Inference*, 105:105–122, 2002.

List of Figures

| | | |
|-----|---|-----|
| 1.1 | Relations entre représentations pratiques: résumé $A \longrightarrow B$: B est un cas particulier de A . $A \dashrightarrow B$: B est représentable par A | 17 |
| 1.2 | Sous-ensembles maximaux cohérents: illustration | 26 |
| 1.3 | Relations d'inclusion des modèles joints à partir de modèles marginaux $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ 33 | |
| 2.1 | Uncertainty treatment: general frame | 43 |
| 3.1 | Representation relationships: summary $A \rightarrow B$: B is a special case of A | 66 |
| 3.2 | Generalized p-box $[\underline{F}, \overline{F}]$ of Example 3.1 | 68 |
| 3.3 | Representation relationships: summary with generalized p-boxes $A \longrightarrow B$: B is a special case of A . $A \dashrightarrow B$: B is representable by A | 78 |
| 3.4 | Cloud $[\pi, \delta]$ of Example 3.5 | 85 |
| 3.5 | Generalized p-box $[\underline{F}, \overline{F}]$ corresponding to cloud of Example 3.5 | 91 |
| 3.6 | Cloud $[\pi, \delta]$ of Example 3.8 | 93 |
| 3.7 | Representation relationships: summary with clouds $A \longrightarrow B$: B is a special case of A . $A \dashrightarrow B$: B is representable by A | 102 |
| 3.8 | Illustration of (weakly) comonotonic and non-comonotonic clouds on the real line. | 106 |
| 4.1 | Example 4.1 distributions | 153 |
| 4.2 | Maximal coherent subsets on Intervals (0.5-cuts of Example 4.1) | 154 |

| | | |
|-----|---|-----|
| 4.3 | Result of MCS method on Example 4.1 (—) and 0.5-cut (---) | 156 |
| 4.4 | Contour function π_c extracted from Example 4.1, with fuzzy focal elements (gray lines) | 157 |
| 4.5 | Result on Example 4.1 of MCS method with number of reliable sources $r = 2$ | 162 |
| 4.6 | Result of MCS method on Example 4.1 with reliability scores $\lambda = (0.2, 0.6, 0.8, 0.2)$ and discounting method | 164 |
| 4.7 | Result of MCS method on Example 4.1 with reliability scores $\lambda = (0.2, 0.6, 0.8, 0.2)$ and discarding method ($\perp(x, y) = x + y - xy$) | 164 |
| 4.8 | Result of MCS method on Example 4.1 taking metric into account with $d_0 = 1$ | 166 |
| 5.1 | Inclusion relationships of joint models, with marginal credal sets $\mathcal{P}_{X_1}, \mathcal{P}_{X_2}$ | 204 |
| 5.2 | An event tree, with initial situation \square , non-terminal situations (such as t) in grey, and terminal situations (such as ω) in black. Also depicted is a cut $U = \{u_1, \dots, u_4\}$. Observe that $t < u_1$ and that $D(t) = \{u_1, u_2\}$. Also, u_4 and t are disjoint, but not u_4 and ω | 207 |
| 5.3 | Evolution of distributions degree (α) versus input space dimension (N) | 217 |
| 5.4 | Evolution of a triangular possibility distribution for different input space dimensions (1,2,3,4,5,10,15,20) | 217 |
| 5.5 | Comparison of probabilistic arithmetic and outer approximation of Proposition 5.3 | 219 |
| 6.1 | u_a with one maximum in a , illustration of cumulative distributions F reaching upper expected value $\bar{E}[\underline{F}, \bar{F}](u_a)$ (left) and lower expected value $\underline{E}[\underline{F}, \bar{F}](u_a)$ (right) | 235 |
| 6.2 | u_a with alternate extrema, illustration of cumulative distributions F reaching lower expected value $\underline{E}[\underline{F}, \bar{F}](u_a)$ | 236 |
| 7.1 | Probability (right) and possibility (left) dist. of NRI1 for the 2PCT | 243 |
| 7.2 | Application of probabilistic aggregation | 245 |

| | | |
|------|---|-----|
| 7.3 | Application of possibilistic aggregation : disjunction (left) and weighted mean (right) | 246 |
| 7.4 | Application of possibilistic aggregation : conjunction (minimum) | 247 |
| 7.5 | Sampling of variables $X_{(1:k)}$ and $X_{(k+1:N)}$ in hybrid numerical propagation. . . | 250 |
| 7.6 | Random fuzzy variable. | 251 |
| 7.7 | Illustration of sample matrix | 252 |
| 7.8 | RaFu method : flowchart (# samples: number of samples). | 255 |
| 7.9 | Triangular fuzzy number modeling K_s | 259 |
| 7.10 | Result of Rafu Method with 1000 samples | 260 |
| 7.11 | Evaluation of the 95% percentile | 261 |
| D.1 | Inner and outer approximations of a non-comonotonic clouds | 293 |
| E.1 | Structure of a nested-disjoint cloud | 302 |

List of Tables

| | | |
|-----|---|-----|
| 1.1 | Notions d'indépendance et de non-pertinence dans l'incertain: résumé (? : question à résoudre) | 32 |
| 4.1 | Properties of Section 4.1.2 for credal sets $\mathcal{P}_1, \dots, \mathcal{P}_N$ with φ the fusion operator, $\mathcal{P}_{\varphi(1:i)} = \varphi(\mathcal{P}_1, \dots, \mathcal{P}_i)$, $K \subseteq \llbracket N \rrbracket$ any maximal subset such that $(\bigcap_{i \in K} \mathcal{P}_i) \neq \emptyset$ and d the supremum norm between credal sets. | 132 |
| 4.2 | Properties of Section 4.1.2 for random sets $(m, \mathcal{F})_1, \dots, (m, \mathcal{F})_N$ with φ the fusion operator, $(m, \mathcal{F})_{\varphi(1:i)} = \varphi((m, \mathcal{F})_1, \dots, (m, \mathcal{F})_i)$, $m_{\varphi(1:i)}$ and $\mathcal{F}_{\varphi(1:i)}$ the associated bpa and focal sets (φ_{\cap} and φ_{\cup} denote the conjunction and disjunction associated to $m_{\varphi(1:i)}$). $K \subseteq \llbracket N \rrbracket$ denote any maximal subset such that $\mathcal{M}_{\cap K}$ is non-empty (i.e., random sets $\{(m, \mathcal{F})_i i \in I\}$ are not totally conflicting), and d a distance measure between random sets [125]. | 136 |
| 4.3 | Properties of Section 4.1.2 for possibility distributions π_1, \dots, π_N with φ the fusion operator, $\pi_{\varphi(1:i)} = \varphi(\pi_1, \dots, \pi_i)$. $K \subseteq \llbracket N \rrbracket$ denote any maximal subset such that $\min_{i \in K} \pi_i \neq \emptyset$, and d the distance measure between possibility distributions given by Equation (4.14). | 139 |
| 4.4 | Information of Example 4.1 sources | 152 |
| 5.1 | Classification of probabilistic independence types | 189 |
| 5.2 | Irrelevance notions in uncertainty: a summary (? : matter of further research) | 205 |
| 7.1 | Participants of BEMUSE programme and used codes | 241 |
| 7.2 | Scalar output values by participants (Exp. Val. : Experimental value) | 242 |

| | | |
|-----|---|-----|
| 7.3 | Results of sources evaluation (Inf.: informativeness ; Cal.: Calibration) by ranks (values) | 244 |
| 7.4 | Summary of parameters used in equation (7.4) | 258 |
| 7.5 | Uncertainty models | 259 |

ABSTRACT

It often happens that the value of some parameters or variables of a system are imperfectly known, either because of the variability of the modelled phenomena, or because the available information is imprecise or incomplete. Classical probability theory is usually used to treat these uncertainties. However, recent years have witnessed the appearance of arguments pointing to the conclusion that classical probabilities are inadequate to handle imprecise or incomplete information. Other frameworks have thus been proposed to address this problem: the three main are probability sets, random sets and possibility theory. There are many open questions concerning uncertainty treatment within these frameworks. More precisely, it is necessary to build bridges between these three frameworks to advance toward a unified handling of uncertainty. Also, there is a need of practical methods to treat information, as using these frameworks can be computationally costly. In this work, we propose some answers to these two needs for a set of commonly encountered problems. In particular, we focus on the problems of:

- Uncertainty representation
- Fusion and evaluation of multiple source information
- Independence modelling

The aim being to give tools (both of theoretical and practical nature) to treat uncertainty. Some tools are then applied to some problems related to nuclear safety issues.

KEYWORDS : Imprecise probabilities, belief functions, possibility theory, representation, information fusion, nuclear safety, independence

AUTEUR : Sébastien Destercke

TITRE : Représentation et combinaison d'informations incertaines: nouveaux résultats avec applications à la sûreté nucléaire

DIRECTEUR DE THESE : Didier Dubois

LIEU et DATE de SOUTENANCE : Institut de Recherche en Informatique de Toulouse, le 29/10/2008

RESUME (français)

Souvent, les valeurs de certains paramètres ou variables d'un système ne sont connues que de façon imparfaite, soit du fait de la variabilité des phénomènes physiques que l'on cherche à représenter, soit parce que l'information dont on dispose est imprécise, incomplète ou pas complètement fiable. Usuellement, cette incertitude est traitée par la théorie classique des probabilités. Cependant, ces dernières années ont vu apparaître des arguments indiquant que les probabilités classiques sont inadéquates lorsqu'il faut représenter l'imprécision présente dans l'information.

Des cadres complémentaires aux probabilités classiques ont donc été proposés pour remédier à ce problème : il s'agit, principalement, des ensembles de probabilités, des ensembles aléatoires et des possibilités. Beaucoup de questions concernant le traitement des incertitudes dans ces trois cadres restent ouvertes. En particulier, il est nécessaire d'unifier ces approches et de comprendre les liens existants entre elles, et de proposer des méthodes de traitement permettant d'utiliser ces approches parfois cher en temps de calcul. Dans ce travail, nous nous proposons d'apporter des réponses à ces deux besoins pour une série de problème de traitement de l'incertain rencontré en analyse de sûreté. En particulier, nous nous concentrons sur les problèmes suivants :

- Représentation des incertitudes
- Fusion/évaluation de données venant de sources multiples
- Modélisation de l'indépendance

L'objectif étant de fournir des outils, à la fois théoriques et pratiques, de traitement d'incertitude. Certains de ces outils sont ensuite appliqués à des problèmes rencontrés en sûreté nucléaire.

MOTS clés : probabilités imprécises, fonctions de croyance, théorie des possibilités, représentation, fusion d'information, sûreté nucléaire, indépendance

DISCIPLINE ADMINISTRATIVE : Informatique

LABORATOIRE D'ACCUEIL : Institut de Recherche en Informatique de Toulouse (IRIT), 118 Rte de Narbonne, 31062 Toulouse, France