Artificial Intelligence: Foundations, Theory, and Algorithms

Series editors
Barry O'Sullivan, Cork, Ireland
Michael Wooldridge, Oxford, United Kingdom

More information about this series at http://www.springer.com/series/13900

Fabio Cuzzolin

The Geometry of Uncertainty

The Geometry of Imprecise Probabilities

Springer

Fabio Cuzzolin
Department of Computing & Communication
Oxford Brookes University
Oxford, UK

ISSN 2365-3051 ISSN 2365-306X (electronic)
Artificial Intelligence: Foundations, Theory, and Algorithms
ISBN 978-3-030-63152-9 ISBN 978-3-030-63153-6 (eBook)
https://doi.org/10.1007/978-3-030-63153-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To my dearest parents Elsa and Albino, whom I owe so much to, my beloved wife Natalia and my beautiful son Leonardo

Preface

Uncertainty

Uncertainty is of paramount importance in artificial intelligence, applied science, and many other areas of human endeavour. Whilst each and every one of us possesses some intuitive grasp of what uncertainty is, providing a formal definition can prove elusive. Uncertainty can be understood as a lack of information about an issue of interest for a certain agent (e.g., a human decision maker or a machine), a condition of limited knowledge in which it is impossible to exactly describe the state of the world or its future evolution.

According to Dennis Lindley [1175]:

" *There are some things that you know to be true, and others that you know to be false; yet, despite this extensive knowledge that you have, there remain many things whose truth or falsity is not known to you. We say that you are uncertain about them. You are uncertain, to varying degrees, about everything in the future; much of the past is hidden from you; and there is a lot of the present about which you do not have full information. Uncertainty is everywhere and you cannot escape from it* ".

What is sometimes less clear to scientists themselves is the existence of a hiatus between two fundamentally distinct forms of uncertainty. The first level consists of somewhat '*predictable*' variations, which are typically encoded as probability distributions. For instance, if a person plays a fair roulette wheel they will not, by any means, know the outcome in advance, but they will nevertheless be able to predict the frequency with which each outcome manifests itself (1/36), at least in the long run. The second level is about '*unpredictable*' variations, which reflect a more fundamental uncertainty about the laws themselves which govern the outcome. Continuing with our example, suppose that the player is presented with ten different doors, which lead to rooms each containing a roulette wheel modelled by a different probability distribution. They will then be uncertain about the very game they are supposed to play. How will this affect their betting behaviour, for instance?

Lack of knowledge of the second kind is often called *Knightian* uncertainty [1007, 831], from the Chicago economist Frank Knight. He would famously distinguish 'risk' from 'uncertainty':

"*Uncertainty must be taken in a sense radically distinct from the familiar notion of risk, from which it has never been properly separated . . . The essential fact is that 'risk' means in some cases a quantity susceptible of measurement, while at other times it is something distinctly not of this character; and there are far-reaching and crucial differences in the bearings of the phenomena depending on which of the two*

is really present and operating . . . It will appear that a measurable uncertainty, or 'risk' proper, as we shall use the term, is so far different from an unmeasurable one that it is not in effect an uncertainty at all."
In Knight's terms, 'risk' is what people normally call *probability* or *chance*, while the term 'uncertainty' is reserved for second-order uncertainty. The latter has a measurable consequence on human behaviour: people are demonstrably averse to unpredictable variations (as highlighted by *Ellsberg's paradox* [569]).

This difference between predictable and unpredictable variation is one of the fundamental issues in the philosophy of probability, and is sometimes referred to as the distinction between *common cause* and *special cause* [1739]. Different interpretations of probability treat these two aspects of uncertainty in different ways, as debated by economists such as John Maynard Keynes [961] and G. L. S. Shackle.

Probability

Measure-theoretical probability, due to the Russian mathematician Andrey Kolmogorov [1030], is the mainstream mathematical theory of (first-order) uncertainty. In Kolmogorov's mathematical approach probability is simply an application of measure theory, and uncertainty is modelled using additive measures.

A number of authors, however, have argued that measure-theoretical probability theory is not quite up to the task when it comes to encoding second-order uncertainty. In particular, as we discuss in the Introduction, additive probability measures cannot properly model missing data or data that comes in the form of *sets*. Probability theory's frequentist interpretation is utterly incapable of modelling 'pure' data (without 'designing' the experiment which generates it). In a way, it cannot even properly model continuous data (owing to the fact that, under measure-theoretical probability, every point of a continuous domain has zero probability), and has to resort to the 'tail event' contraption to assess its own hypotheses. Scarce data can only be effectively modelled asymptotically.

Bayesian reasoning is also plagued by many serious limitations: (i) it just cannot model ignorance (absence of data); (ii) it cannot model pure data (without artificially introducing a prior, even when there is no justification for doing so); (iii) it cannot model 'uncertain' data, i.e., information not in the form of propositions of the kind 'A is true'; and (iv) again, it is able to model scarce data only asymptotically, thanks to the Bernstein–von Mises theorem [1841].

Beyond probability

Similar considerations have led a number of scientists to recognise the need for a coherent mathematical theory of uncertainty able to properly tackle all these issues. Both alternatives to and extensions of classical probability theory have been proposed, starting from de Finetti's pioneering work on subjective probability [403]. Formalisms include possibility-fuzzy set theory [2084, 533], probability intervals

[784], credal sets [1141, 1086], monotone capacities [1911], random sets [1344] and imprecise probability theory [1874]. New original foundations of subjective probability in behavioural terms [1877] or by means of game theory [1615] have been put forward. The following table presents a sketchy timeline of the various existing approaches to the mathematics of uncertainty.

Imprecise-probabilistic theories: a timeline

Approach	Proposer(s)	Seminal paper	Year
Interval probabilities	John Maynard Keynes	A treatise on probability	1921
Subjective probability	Bruno de Finetti	Sul significato soggettivo della probabilità	1931
Theory of previsions	Bruno de Finetti	La prévision: ses lois logiques, ses sources subjectives	1937
Theory of capacities	Gustave Choquet	Theory of capacities	1953
Fuzzy theory	Lotfi Zadeh, Dieter Klaua	Fuzzy sets	1965
Theory of evidence	Arthur Dempster, Glenn Shafer	Upper and lower probabilities induced by a multivalued mapping; A mathematical theory of evidence	1967, 1976
Fuzzy measures	Michio Sugeno	Theory of fuzzy integrals and its applications	1974
Credal sets	Isaac Levi	The enterprise of knowledge	1980
Possibility theory	Didier Dubois, Henri Prade	Théorie des possibilités	1985
Imprecise probability	Peter Walley	Statistical reasoning with imprecise probabilities	1991
Game-theoretical probability	Glenn Shafer, Vladimir Vovk	Probability and finance: It's only a game!	2001

Sometimes collectively referred to as *imprecise probabilities* (as most of them comprise classical probabilities as a special case), these theories in fact form, as we will see in more detail in Chapter 6, an entire hierarchy of encapsulated formalisms.

Belief functions

One of the most popular formalisms for a mathematics of uncertainty, the *theory of evidence* [1583] was introduced in the 1970s by Glenn Shafer as a way of representing epistemic knowledge, starting from a sequence of seminal papers [415, 417, 418] by Arthur Dempster [416]. In this formalism, the best representation of chance is a *belief function* rather than a traditional probability distribution. Belief functions assign probability values to *sets* of outcomes, rather than single events. In this sense, belief functions are closely related to *random sets* [1268, 1857, 826]. Important work on the mathematics of random set theory has been conducted in recent years by Ilya Molchanov [1302, 1304].

In its original formulation by Dempster and Shafer, the formalism provides a simple method for merging the evidence carried by a number of distinct sources (called *Dempster's rule of combination* [773]), with no need for any prior distributions [1949]. The existence of different levels of granularity in knowledge representation is formalised via the concept of a 'family of compatible frames'.

The reason for the wide interest the theory of belief functions has attracted over the years is that it addresses most of the issues probability theory has with the handling of second-order uncertainty. It starts from the assumption that observations are indeed set-valued and that evidence is, in general, in support of propositions rather than single outcomes. It can model ignorance by simply assigning mass to the whole sample space or 'frame of discernment'. It copes with missing data in the most natural of ways, and can coherently represent evidence on different but compatible sample spaces. It does not 'need' priors but can make good use of them whenever there is actual prior knowledge to exploit. As a direct generalisation of classical probability, the theory's rationale is relatively easier to grasp. Last but not least, the formalism does not require us to entirely abandon the notion of an event, as is the case for Walley's imprecise probability theory [1874]. In addition, the theory of evidence exhibits links to most other theories of uncertainty, as it includes fuzzy and possibility theory as a special case and it relates to the theory of credal sets and imprecise probability theory (as belief functions can be seen as a special case of convex sets of probability measures). Belief functions are infinitely-monotone capacities, and have natural interpretations in the framework of probabilistic logic, and modal logic in particular.

Since its inception, the formalism has expanded to address issues such as inference (how to map data to a belief function), conditioning, and the generalisation of the notion of entropy and of classical results from probability theory to the more complex case of belief functions. The question of what combination rule is most appropriate under what circumstances has been hotly debated, together with that of mitigating the computational complexity of working with sets of hypotheses. Graphical models, machine learning approaches and decision making frameworks based on belief theory have also been developed.

A number of questions still remain open, for instance on what is the correct epistemic interpretation of belief functions, whether we should actually manipulate intervals of belief functions rather than single quantities, and how to formulate an effective general theory for the case of continuous sample spaces.

Aim(s) of the book

The principal aim of this book is to introduce to the widest possible audience an original view of belief calculus and uncertainty theory which I first developed during my doctoral term in Padua. In this *geometric approach to uncertainty*, uncertainty measures can be seen as points of a suitably complex geometric space, and there manipulated (e.g. combined, conditioned and so on).

The idea first sprang to my mind just after I had been introduced to non-additive probabilities. Where did such objects live, I wondered, when compared to classical, additive probabilities defined on the same sample space? How is their greater complexity reflected in the geometry of their space? Is the latter an expression of the greater degree of freedom these more complex objects can provide?

For the reasons mentioned above, my attention was first drawn to belief functions and their combination rule which, from the point of view of an engineer, appeared to provide a possible principled solution to the sensor fusion problems one encounters in computer vision when making predictions or decisions based on multiple measurements or 'features'. Using the intuition gathered in the simplest case study of a binary domain, I then proceeded to describe the geometry of belief functions and their combination in fully general terms and to extend, in part, this geometric analysis to other classes of uncertainty measures.

This programme of work is still far from reaching its conclusion – nevertheless, I thought that the idea of consolidating my twenty-year work on the geometry of uncertainty in a monograph had some merit, especially in order to disseminate the notion and encourage a new generation of scientists to develop it further. This is the purpose of the core of the book, Parts II, III and IV.

In the years that it took for this project to materialise, I realised that the manuscript could serve the wider purpose of illustrating the rationale for moving away from probability theory to non-experts and interested practitioners, of which there are many. Incidentally, this forced me to reconsider from the very foundations the reasons for modelling uncertainty in a non-standard way. These reasons, as understood by myself, can be found in the Introduction, which is an extended version of the tutorial I gave on the topic at IJCAI 2016, the International Joint Conference on Artificial Intelligence, and the talk I was invited to give at Harvard University in the same year.

The apparent lack of a comprehensive treatise on belief calculus in its current, modern form (and, from a wider perspective, of uncertainty theory) motivated me to make use of this book to provide what turned out to be probably the most complete summary (to the best of my knowledge) of the theory of belief functions. The entire first part of the book is devoted to this purpose. Part I is not quite a 'manual' on belief calculus, with easy recipes the interested practitioner can just follow, but does strive to make a serious effort in that direction. Furthermore, the first part of the book concludes with what I believe to be the most complete compendium of the various approaches to uncertainty theory, with a specific focus on how do they relate to the theory of evidence. All major formalisms are described in quite some detail, but an effort was really made to cover, albeit briefly, all published approaches to a mathematics of uncertainty and variations thereof.

Finally, the last chapter of the book advances a tentative research agenda for the future of the field, inspired by my own reflections and ideas on this. As will become clearer in the remainder of this work, my intuition brings me to favour a random-set view of uncertainty theory, driven by an analysis of the actual issues with data that expose the limitations of probability theory. As a result, the research problems I propose tend to point in this direction. Importantly, I strongly believe that, to break the near-monopoly of probability theory in science, uncertainty theory needs to measure itself with the really challenging issues of our time (climate change, robust artificial intelligence), compete with mainstream approaches and demonstrate its superior expressive power on their own grounds.

Last but not least, the book provides, again to the best of my knowledge, the largest existing bibliography on belief and uncertainty theory.

Structure and topics

Accordingly, as explained, this book is divided into five Parts.

Part I, 'Theories of uncertainty', is a rather extensive recapitulation of the current state of the art in the mathematics of uncertainty, with a focus on belief theory. The Introduction provided in Chapter 1 motivates in more detail the need to go beyond classical probability in order to model realistic, second-order uncertainty, introduces the most significant approaches to the mathematics of uncertainty and presents the main principles of the theory of belief functions. Chapter 2 provides a succinct summary of the basic notions of the theory of belief functions as formulated by Shafer. Chapter 3 digs deeper by recalling the multiple semantics of belief functions, discussing the genesis of the approach and the subsequent debate, and illustrating the various original frameworks proposed by a number of authors which use belief theory as a basis, while developing it further in original ways. Chapter 4 can be thought of as a manual for the working scientist keen on applying belief theory. It illustrates in detail all the elements of the evidential reasoning chain, delving into all its aspects, including inference, conditioning and combination, efficient computation, decision making and continuous formulations. Notable advances in the mathematics of belief functions are also briefly described. Chapter 5 surveys the existing array of classification, clustering, regression and estimation tools based on belief function theory. Finally, Chapter 6 is designed to provide the reader with a bigger picture of the whole field of uncertainty theory, by reviewing all major formalisms (the most significant of which are arguably Walley's imprecise probability, the theory of capacities and fuzzy/possibility theory), with special attention paid to their relationship with belief and random set theory.

Part II, 'The geometry of uncertainty', is the core of this book, as it introduces the author's own geometric approach to uncertainty theory, starting with the geometry of belief functions. First, Chapter 7 studies the geometry of the space of belief functions, or *belief space*, both in terms of a simplex (a higher-dimensional triangle) and in terms of its recursive bundle structure. Chapter 8 extends the analysis to Dempster's rule of combination, introducing the notion of a conditional subspace and outlining a simple geometric construction for Dempster's sum. Chapter 9 delves into the combinatorial properties of plausibility and commonality functions, as equivalent representations of the evidence carried by a belief function. It shows that the corresponding spaces also behave like simplices, which are congruent to the belief space. The remaining Chapter 10 starts extending the applicability of the geometric approach to other uncertainty measures, focusing in particular on possibility measures (consonant belief functions) and the related notion of a consistent belief function.

Part III, 'Geometric interplays', is concerned with the interplay of uncertainty measures of different kinds, and the geometry of their relationship. Chapters 11 and 12 study the problem of transforming a belief function into a classical probability measure. In particular, Chapter 11 introduces the *affine family* of probability transformations, those which commute with affine combination in the belief space. Chapter 12 focuses instead on the *epistemic* family of transforms, namely 'relative belief' and 'relative plausibility', studies their dual properties with respect to Dempster's sum, and describes their geometry on both the probability simplex and the belief space. Chapter 13 extends the analysis to the consonant approximation problem, the problem of finding the possibility measure which best approximates a given belief function. In particular, approximations induced by classical Minkowski norms are derived, and compared with classical outer consonant approximations. Chapter 14 concludes Part III by describing Minkowski consistent approximations of belief functions in both the mass and the belief space representations.

Part IV, 'Geometric reasoning', examines the application of the geometric approach to the various elements of the reasoning chain illustrated in Chapter 4. Chapter 15 tackles the conditioning problem from a geometric point of view. Conditional belief functions are defined as those which minimise an appropriate distance between the original belief function and the 'conditioning simplex' associated with the conditioning event. Analytical expressions are derived for both the belief and the mass space representations, in the case of classical Minkowski distances. Chapter 16 provides a semantics for the main probability transforms in terms of credal sets, i.e., convex sets of probabilities. Based on this interpretation, decision-making apparatuses similar to Smets's transferable belief model are outlined.

Part V, 'The future of uncertainty', consisting of Chapter 17, concludes the book by outlining an agenda for the future of the discipline. A comprehensive statistical theory of random sets is proposed as the natural destination of this evolving field. A number of open challenges in the current formulation of belief theory are highlighted, and a research programme for the completion of our geometric approach to uncertainty is proposed. Finally, very high-impact applications in fields such as climate change, rare event estimation, machine learning and statistical learning theory are singled out as potential triggers of a much larger diffusion of these techniques and of uncertainty theory in general.

Acknowledgements

This book would not have come to existence had I not, during my doctoral term at the University of Padua, been advised by my then supervisor Ruggero Frezza to attend a seminar by Claudio Sossai at Consiglio Nazionale delle Ricerche (CNR). At the time, as every PhD student in their first year, I was struggling to identify a topic of research to focus on, and felt quite overwhelmed by how much I did not know about almost everything. After Claudio's seminar I was completely taken by the idea of non-additive probability, and ended up dedicating my entire doctorate to the study of the mathematics of these beautiful objects. Incidentally, I need to thank Nicola Zingirian and Riccardo Bernardini for enduring my ramblings during our coffee breaks back then.

I would like to acknowledge the encouragement I received from Glenn Shafer over the years, which has greatly helped me in my determination to contribute to belief theory in terms of both understanding and impact. I would also like to thank Art Dempster for concretely supporting me in my career and for the insights he shared with me in our epistolary exchanges (not to mention the amazing hospitality displayed during my visit to Harvard in 2016).

Running the risk of forgetting somebody important, I would like to thank Teddy Seidenfeld, Gert de Cooman, Matthias Troffaes, Robin Gong, Xiao-Li Meng, Frank Coolen, Thierry Denoeux, Paul Snow, Jonathan Lawry, Arnaud Martin, Johan Schubert, Anne-Laure Jousselme, Sebastien Destercke, Milan Daniel, Jim Hall, Alberto Bernardini, Thomas Burger and Alessandro Antonucci for the fascinating conversations that all deeply influenced my work. I am also thankful to Ilya Molchanov for his help with better understanding the links between belief function theory and random set theory.

Finally, I am grateful to my Springer editor Ronan Nugent for his patience and support throughout all those years that passed from when we started this project to the moment we could actually cheer its successful outcome.

Oxford, United Kingdom *Fabio Cuzzolin*

September 2020

Table of Contents

1 Introduction . . . 1
1.1 Mathematical probability . . . 1
1.2 Interpretations of probability . . . 3
1.2.1 Does probability exist at all? . . . 3
1.2.2 Competing interpretations . . . 3
1.2.3 Frequentist probability . . . 4
1.2.4 Propensity . . . 6
1.2.5 Subjective and Bayesian probability . . . 7
1.2.6 Bayesian versus frequentist inference . . . 9
1.3 Beyond probability . . . 9
1.3.1 Something is wrong with probability . . . 9
1.3.2 Pure data: Beware of the prior . . . 10
1.3.3 Pure data: Designing the universe? . . . 10
1.3.4 No data: Modelling ignorance . . . 11
1.3.5 Set-valued observations: The cloaked die . . . 12
1.3.6 Propositional data . . . 13
1.3.7 Scarce data: Beware the size of the sample . . . 14
1.3.8 Unusual data: Rare events . . . 17
1.3.9 Uncertain data . . . 18
1.3.10 Knightian uncertainty . . . 20
1.4 Mathematics (plural) of uncertainty . . . 23
1.4.1 Debate on uncertainty theory . . . 23
1.4.2 Belief, evidence and probability . . . 26

Part I Theories of uncertainty

2 Belief functions . . . 31
2.1 Arthur Dempster's original setting . . . 32
2.2 Belief functions as set functions . . . 34
2.2.1 Basic definitions . . . 34
2.2.2 Plausibility and commonality functions . . . 36
2.2.3 Bayesian belief functions . . . 37
2.3 Dempster's rule of combination . . . 38
2.3.1 Definition . . . 38

2.3.2 Weight of conflict . . . 40
2.3.3 Conditioning belief functions . . . 40
2.4 Simple and separable support functions . . . 41
2.4.1 Heterogeneous and conflicting evidence . . . 42
2.4.2 Separable support functions . . . 42
2.4.3 Internal conflict . . . 43
2.4.4 Inverting Dempster's rule: The canonical decomposition . . . 44
2.5 Families of compatible frames of discernment . . . 45
2.5.1 Refinings . . . 45
2.5.2 Families of frames . . . 46
2.5.3 Consistent and marginal belief functions . . . 48
2.5.4 Independent frames . . . 49
2.5.5 Vacuous extension . . . 50
2.6 Support functions . . . 51
2.6.1 Families of compatible support functions in the evidential language . . . 51
2.6.2 Discerning the relevant interaction of bodies of evidence . . . 52
2.7 Quasi-support functions . . . 53
2.7.1 Limits of separable support functions . . . 53
2.7.2 Bayesian belief functions as quasi-support functions . . . 54
2.7.3 Bayesian case: Bayes' theorem . . . 54
2.7.4 Bayesian case: Incompatible priors . . . 55
2.8 Consonant belief functions . . . 56

3 Understanding belief functions . . . 59
3.1 The multiple semantics of belief functions . . . 61
3.1.1 Dempster's multivalued mappings, compatibility relations . . 62
3.1.2 Belief functions as generalised (non-additive) probabilities . 64
3.1.3 Belief functions as inner measures . . . 65
3.1.4 Belief functions as credal sets . . . 66
3.1.5 Belief functions as random sets . . . 68
3.1.6 Behavioural interpretations . . . 69
3.1.7 Common misconceptions . . . 71
3.2 Genesis and debate . . . 73
3.2.1 Early support . . . 73
3.2.2 Constructive probability and Shafer's canonical examples . . 74
3.2.3 Bayesian versus belief reasoning . . . 75
3.2.4 Pearl's criticism . . . 76
3.2.5 Issues with multiple interpretations . . . 77
3.2.6 Rebuttals and justifications . . . 78
3.3 Frameworks . . . 79
3.3.1 Frameworks based on multivalued mappings . . . 80
3.3.2 Smets's transferable belief model . . . 90
3.3.3 Dezert–Smarandache theory (DSmT) . . . 93
3.3.4 Gaussian (linear) belief functions . . . 95

3.3.5 Belief functions on generalised domains 97
3.3.6 Qualitative models 98
3.3.7 Intervals and sets of belief measures 101
3.3.8 Other frameworks 103

4 Reasoning with belief functions 109
4.1 Inference 111
4.1.1 From statistical data 112
4.1.2 From qualitative data 121
4.1.3 From partial knowledge 123
4.1.4 A coin toss example 124
4.2 Measuring uncertainty 127
4.2.1 Order relations 127
4.2.2 Measures of entropy 128
4.2.3 Principles of uncertainty 132
4.3 Combination 133
4.3.1 Dempster's rule under fire 134
4.3.2 Alternative combination rules 139
4.3.3 Families of combination rules 147
4.3.4 Combination of dependent evidence 150
4.3.5 Combination of conflicting evidence 151
4.3.6 Combination of (un)reliable sources of evidence: Discounting 154
4.4 Belief versus Bayesian reasoning: A data fusion example 156
4.4.1 Two fusion pipelines 156
4.4.2 Inference under partially reliable data 157
4.5 Conditioning 159
4.5.1 Dempster conditioning 159
4.5.2 Lower and upper conditional envelopes 160
4.5.3 Suppes and Zanotti's geometric conditioning 161
4.5.4 Smets's conjunctive rule of conditioning 161
4.5.5 Disjunctive rule of conditioning 162
4.5.6 Conditional events as equivalence classes: Spies's definition 163
4.5.7 Other work 164
4.5.8 Conditioning: A summary 164
4.6 Manipulating (conditional) belief functions 165
4.6.1 The generalised Bayes theorem 165
4.6.2 Generalising total probability 167
4.6.3 Multivariate belief functions 169
4.6.4 Graphical models 171
4.7 Computing 176
4.7.1 Efficient algorithms 177
4.7.2 Transformation approaches 181
4.7.3 Monte Carlo approaches 188
4.7.4 Local propagation 190

4.8 Making decisions . . . 196
4.8.1 Frameworks based on utilities . . . 197
4.8.2 Frameworks not based on utilities . . . 206
4.8.3 Multicriteria decision making . . . 208
4.9 Continuous formulations . . . 213
4.9.1 Shafer's allocations of probabilities . . . 213
4.9.2 Belief functions on random Borel intervals . . . 216
4.9.3 Random sets . . . 219
4.9.4 Kramosil's belief functions on infinite spaces . . . 222
4.9.5 MV algebras . . . 225
4.9.6 Other approaches . . . 227
4.10 The mathematics of belief functions . . . 227
4.10.1 Distances and dissimilarities . . . 227
4.10.2 Algebra . . . 229
4.10.3 Integration . . . 233
4.10.4 Category theory . . . 233
4.10.5 Other mathematical analyses . . . 234

5 A toolbox for the working scientist . . . 237
5.1 Clustering . . . 238
5.1.1 Fuzzy, evidential and belief C-means . . . 239
5.1.2 EVCLUS and later developments . . . 240
5.1.3 Clustering belief functions . . . 243
5.2 Classification . . . 244
5.2.1 Generalised Bayesian classifier . . . 244
5.2.2 Evidential k-NN . . . 245
5.2.3 TBM model-based classifier . . . 246
5.2.4 SVM classification . . . 247
5.2.5 Classification with partial training data . . . 248
5.2.6 Decision trees . . . 249
5.2.7 Neural networks . . . 251
5.2.8 Other classification approaches . . . 252
5.3 Ensemble classification . . . 254
5.3.1 Distance-based classification fusion . . . 255
5.3.2 Empirical comparison of fusion schemes . . . 258
5.3.3 Other classifier fusion schemes . . . 258
5.4 Ranking aggregation . . . 260
5.5 Regression . . . 261
5.5.1 Fuzzy-belief non-parametric regression . . . 262
5.5.2 Belief-modelling regression . . . 262
5.6 Estimation, prediction and identification . . . 264
5.6.1 State estimation . . . 264
5.6.2 Time series analysis . . . 265
5.6.3 Particle filtering . . . 266
5.6.4 System identification . . . 266

5.7 Optimisation . . . 267

6 The bigger picture . . . 269
6.1 Imprecise probability . . . 271
6.1.1 Lower and upper probabilities . . . 272
6.1.2 Gambles and behavioural interpretation . . . 273
6.1.3 Lower and upper previsions . . . 273
6.1.4 Events as indicator gambles . . . 274
6.1.5 Rules of rational behaviour . . . 275
6.1.6 Natural and marginal extension . . . 275
6.1.7 Belief functions and imprecise probabilities . . . 276
6.2 Capacities (a.k.a. fuzzy measures) . . . 277
6.2.1 Special types of capacities . . . 278
6.3 Probability intervals (2-monotone capacities) . . . 280
6.3.1 Probability intervals and belief measures . . . 281
6.4 Higher-order probabilities . . . 282
6.4.1 Second-order probabilities and belief functions . . . 282
6.4.2 Gaifman's higher-order probability spaces . . . 282
6.4.3 Kyburg's analysis . . . 283
6.4.4 Fung and Chong's metaprobability . . . 283
6.5 Fuzzy theory . . . 283
6.5.1 Possibility theory . . . 284
6.5.2 Belief functions on fuzzy sets . . . 285
6.5.3 Vague sets . . . 287
6.5.4 Other fuzzy extensions of the theory of evidence . . . 288
6.6 Logic . . . 289
6.6.1 Saffiotti's belief function logic . . . 290
6.6.2 Josang's subjective logic . . . 291
6.6.3 Fagin and Halpern's axiomatisation . . . 291
6.6.4 Probabilistic argumentation systems . . . 292
6.6.5 Default logic . . . 292
6.6.6 Ruspini's logical foundations . . . 293
6.6.7 Modal logic interpretation . . . 294
6.6.8 Probability of provability . . . 296
6.6.9 Other logical frameworks . . . 296
6.7 Rough sets . . . 299
6.7.1 Pawlak's algebras of rough sets . . . 299
6.7.2 Belief functions and rough sets . . . 300
6.8 Probability boxes . . . 301
6.8.1 Probability boxes and belief functions . . . 302
6.8.2 Approximate computations for random sets . . . 302
6.8.3 Generalised probability boxes . . . 304
6.9 Spohn's theory of epistemic beliefs . . . 304
6.9.1 Epistemic states . . . 305
6.9.2 Disbelief functions and Spohnian belief functions . . . 306

6.9.3 α-conditionalisation . . . 306
6.10 Zadeh's generalised theory of uncertainty . . . 307
6.11 Baoding Liu's uncertainty theory . . . 308
6.12 Info-gap decision theory . . . 310
6.12.1 Info-gap models . . . 310
6.12.2 Robustness of design . . . 311
6.13 Vovk and Shafer's game-theoretical framework . . . 311
6.13.1 Game-theoretic probability . . . 312
6.13.2 Ville/Vovk game-theoretic testing . . . 313
6.13.3 Upper price and upper probability . . . 315
6.14 Other formalisms . . . 315
6.14.1 Endorsements . . . 315
6.14.2 Fril-fuzzy theory . . . 315
6.14.3 Granular computing . . . 316
6.14.4 Laskey's assumptions . . . 316
6.14.5 Harper's Popperian approach to rational belief change . . . 316
6.14.6 Shastri's evidential reasoning in semantic networks . . . 317
6.14.7 Evidential confirmation theory . . . 317
6.14.8 Groen's extension of Bayesian theory . . . 318
6.14.9 Padovitz's unifying model . . . 319
6.14.10 Similarity-based reasoning . . . 319
6.14.11 Neighbourhoods systems . . . 320
6.14.12 Comparative belief structures . . . 320

Part II The geometry of uncertainty

7 The geometry of belief functions . . . 323
7.1 The space of belief functions . . . 326
7.1.1 The simplex of dominating probabilities . . . 326
7.1.2 Dominating probabilities and L_1 norm . . . 328
7.1.3 Exploiting the Möbius inversion lemma . . . 329
7.1.4 Convexity of the belief space . . . 330
7.2 Simplicial form of the belief space . . . 331
7.2.1 Faces of $\mathcal{B}$ as classes of belief functions . . . 334
7.3 The differential geometry of belief functions . . . 335
7.3.1 A case study: The ternary case . . . 335
7.3.2 Definition of smooth fibre bundles . . . 337
7.3.3 Normalised sum functions . . . 339
7.4 Recursive bundle structure . . . 339
7.4.1 Recursive bundle structure of the space of sum functions . . . 339
7.4.2 Recursive bundle structure of the belief space . . . 340
7.4.3 Bases and fibres as simplices . . . 341
7.5 Research questions . . . 344

8 Geometry of Dempster's rule . . . 351
8.1 Dempster combination of pseudo-belief functions . . . 352
8.2 Dempster sum of affine combinations . . . 353
8.3 Convex formulation of Dempster's rule . . . 356
8.4 Commutativity . . . 357
8.4.1 Affine region of missing points . . . 358
8.4.2 Non-combinable points and missing points: A duality . . . 358
8.4.3 The case of unnormalised belief functions . . . 359
8.5 Conditional subspaces . . . 360
8.5.1 Definition . . . 360
8.5.2 The case of unnormalised belief functions . . . 362
8.5.3 Vertices of conditional subspaces . . . 363
8.6 Constant-mass loci . . . 364
8.6.1 Geometry of Dempster's rule in $\mathcal{B}_2$ . . . 364
8.6.2 Affine form of constant-mass loci . . . 369
8.6.3 Action of Dempster's rule on constant-mass loci . . . 370
8.7 Geometric orthogonal sum . . . 372
8.7.1 Foci of conditional subspaces . . . 372
8.7.2 Algorithm . . . 375
8.8 Research questions . . . 376

9 Three equivalent models . . . 389
9.1 Basic plausibility assignment . . . 390
9.1.1 Relation between basic probability and plausibility assignments . . . 391
9.2 Basic commonality assignment . . . 392
9.2.1 Properties of basic commonality assignments . . . 393
9.3 The geometry of plausibility functions . . . 393
9.3.1 Plausibility assignment and simplicial coordinates . . . 394
9.3.2 Plausibility space . . . 395
9.4 The geometry of commonality functions . . . 396
9.5 Equivalence and congruence . . . 399
9.5.1 Congruence of belief and plausibility spaces . . . 399
9.5.2 Congruence of plausibility and commonality spaces . . . 400
9.6 Pointwise rigid transformation . . . 402
9.6.1 Belief and plausibility spaces . . . 402
9.6.2 Commonality and plausibility spaces . . . 403

10 The geometry of possibility . . . 409
10.1 Consonant belief functions as necessity measures . . . 410
10.2 The consonant subspace . . . 411
10.2.1 Chains of subsets as consonant belief functions . . . 412
10.2.2 The consonant subspace as a simplicial complex . . . 414
10.3 Properties of the consonant subspace . . . 416
10.3.1 Congruence of the convex components of $\mathcal{CO}$ . . . 416

10.3.2 Decomposition of maximal simplices into right triangles . . . 418
10.4 Consistent belief functions . . . 418
10.4.1 Consistent knowledge bases in classical logic . . . 419
10.4.2 Belief functions as uncertain knowledge bases . . . 419
10.4.3 Consistency in belief logic . . . 420
10.5 The geometry of consistent belief functions . . . 422
10.5.1 The region of consistent belief functions . . . 422
10.5.2 The consistent complex . . . 423
10.5.3 Natural consistent components . . . 424
10.6 Research questions . . . 426

Part III Geometric interplays

11 Probability transforms: The affine family . . . 431
11.1 Affine transforms in the binary case . . . 433
11.2 Geometry of the dual line . . . 435
11.2.1 Orthogonality of the dual line . . . 436
11.2.2 Intersection with the region of Bayesian normalised sum functions . . . 436
11.3 The intersection probability . . . 438
11.3.1 Interpretations of the intersection probability . . . 439
11.3.2 Intersection probability and affine combination . . . 444
11.3.3 Intersection probability and convex closure . . . 447
11.4 Orthogonal projection . . . 447
11.4.1 Orthogonality condition . . . 448
11.4.2 Orthogonality flag . . . 449
11.4.3 Two mass redistribution processes . . . 449
11.4.4 Orthogonal projection and affine combination . . . 451
11.4.5 Orthogonal projection and pignistic function . . . 452
11.5 The case of unnormalised belief functions . . . 453
11.6 Comparisons within the affine family . . . 455

12 Probability transforms: The epistemic family . . . 469
12.1 Rationale for epistemic transforms . . . 471
12.1.1 Semantics within the probability-bound interpretation . . . 472
12.1.2 Semantics within Shafer's interpretation . . . 473
12.2 Dual properties of epistemic transforms . . . 474
12.2.1 Relative plausibility, Dempster's rule and pseudo-belief functions . . . 475
12.2.2 A (broken) symmetry . . . 475
12.2.3 Dual properties of the relative belief operator . . . 476
12.2.4 Representation theorem for relative beliefs . . . 479
12.2.5 Two families of Bayesian approximations . . . 481
12.3 Plausibility transform and convex closure . . . 481

12.4 Generalisations of the relative belief transform 482
12.4.1 Zero mass assigned to singletons as a singular case 482
12.4.2 The family of relative mass probability transformations 484
12.4.3 Approximating the pignistic probability and relative plausibility 485
12.5 Geometry in the space of pseudo-belief functions 489
12.5.1 Plausibility of singletons and relative plausibility 489
12.5.2 Belief of singletons and relative belief 490
12.5.3 A three-plane geometry 491
12.5.4 A geometry of three angles 493
12.5.5 Singular case 496
12.6 Geometry in the probability simplex 497
12.7 Equality conditions for both families of approximations 504
12.7.1 Equal plausibility distribution in the affine family 504
12.7.2 Equal plausibility distribution as a general condition 505

13 Consonant approximation 517
13.1 Geometry of outer consonant approximations in the consonant simplex 522
13.1.1 Geometry in the binary case 522
13.1.2 Convexity 523
13.1.3 Weak inclusion and mass reassignment 523
13.1.4 The polytopes of outer approximations 524
13.1.5 Maximal outer approximations 525
13.1.6 Maximal outer approximations as lower chain measures 526
13.2 Geometric consonant approximation 528
13.2.1 Mass space representation 529
13.2.2 Approximation in the consonant complex 531
13.2.3 Möbius inversion and preservation of norms, induced orderings 533
13.3 Consonant approximation in the mass space 534
13.3.1 Results of Minkowski consonant approximation in the mass space 535
13.3.2 Semantics of partial consonant approximations in the mass space 537
13.3.3 Computability and admissibility of global solutions 540
13.3.4 Relation to other approximations 541
13.4 Consonant approximation in the belief space 545
13.4.1 L_1 approximation 546
13.4.2 (Partial) L_2 approximation 546
13.4.3 L_∞ approximation 547
13.4.4 Approximations in the belief space as generalised maximal outer approximations 549
13.5 Belief versus mass space approximations 553
13.5.1 On the links between approximations in $\mathcal{M}$ and $\mathcal{B}$ 554

13.5.2 Three families of consonant approximations 555

14 Consistent approximation 575
14.1 The Minkowski consistent approximation problem 577
14.2 Consistent approximation in the mass space 578
14.2.1 L_1 approximation 578
14.2.2 L_∞ approximation 579
14.2.3 L_2 approximation 580
14.3 Consistent approximation in the belief space 581
14.3.1 L_1/L_2 approximations 581
14.3.2 L_∞ consistent approximation 585
14.4 Approximations in the belief versus the mass space 586

Part IV Geometric reasoning

15 Geometric conditioning 595
15.1 Geometric conditional belief functions 597
15.2 Geometric conditional belief functions in $\mathcal{M}$ 598
15.2.1 Conditioning by L_1 norm 598
15.2.2 Conditioning by L_2 norm 599
15.2.3 Conditioning by L_∞ norm 600
15.2.4 Features of geometric conditional belief functions in $\mathcal{M}$ 602
15.2.5 Interpretation as general imaging for belief functions 603
15.3 Geometric conditioning in the belief space 604
15.3.1 L_2 conditioning in $\mathcal{B}$ 604
15.3.2 L_1 conditioning in $\mathcal{B}$ 606
15.3.3 L_∞ conditioning in $\mathcal{B}$ 607
15.4 Mass space versus belief space conditioning 610
15.4.1 Geometric conditioning: A summary 610
15.5 An outline of future research 613

16 Decision making with epistemic transforms 627
16.1 The credal set of probability intervals 630
16.1.1 Lower and upper simplices 631
16.1.2 Simplicial form 631
16.1.3 Lower and upper simplices and probability intervals 632
16.2 Intersection probability and probability intervals 632
16.3 Credal interpretation of Bayesian transforms: Ternary case 634
16.4 Credal geometry of probability transformations 636
16.4.1 Focus of a pair of simplices 636
16.4.2 Probability transformations as foci 638
16.4.3 Semantics of foci and a rationality principle 639
16.4.4 Mapping associated with a probability transformation 639
16.4.5 Upper and lower simplices as consistent probabilities 641

16.5 Decision making with epistemic transforms 642
16.5.1 Generalisations of the TBM 642
16.5.2 A game/utility theory interpretation 643

Part V The future of uncertainty

17 An agenda for the future 655
17.1 A statistical random set theory 659
17.1.1 Lower and upper likelihoods 659
17.1.2 Generalised logistic regression 669
17.1.3 The total probability theorem for random sets 671
17.1.4 Limit theorems for random sets 684
17.1.5 Frequentist inference with random sets 689
17.1.6 Random-set random variables 690
17.2 Developing the geometric approach 692
17.2.1 Geometry of general combination 693
17.2.2 Geometry of general conditioning 700
17.2.3 A true geometry of uncertainty 701
17.2.4 Fancier geometries 706
17.2.5 Relation to integral and stochastic geometry 707
17.3 High-impact developments 710
17.3.1 Climate change 710
17.3.2 Machine learning 714
17.3.3 Generalising statistical learning theory 721

References 729