Universität Stuttgart
Institut für Technische und Numerische Mechanik

Dominik Hose

Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering

SHAKER VERLAG

Band 74 (2022)

Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering

Von der Fakultät Konstruktions-, Produktions- und Fahrzeugtechnik
der Universität Stuttgart
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von
Christian Dominik Hose
aus Offenbach am Main

Hauptberichter:	apl. Prof. Dr.-Ing. Michael Hanss
Mitberichter:	Prof. Scott Ferson, PhD
	Prof. Ryan Martin, PhD

Tag der mündlichen Prüfung: 20. Mai 2022

Institut für Technische und Numerische Mechanik
der Universität Stuttgart

2022

Schriften aus dem Institut für Technische und Numerische Mechanik der Universität Stuttgart

Herausgeber: Prof. Dr.-Ing. Prof. E.h. Peter Eberhard

Band 74/2022

Dominik Hose

Possibilistic Reasoning with Imprecise Probabilities: Statistical Inference and Dynamic Filtering

D 93 (Diss. Universität Stuttgart)

Shaker Verlag
Düren 2022

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Zugl.: Stuttgart, Univ., Diss., 2022

Printed in Germany.

ISBN 978-3-8440-8721-5
ISSN 1861-1651

Shaker Verlag GmbH • Am Langen Graben 15a • 52353 Düren
Phone: 0049/2421/99011-0 • Telefax: 0049/2421/99011-9
Internet: www.shaker.de • e-mail: info@shaker.de

Vorwort

Die vorliegende Dissertation entstand während meiner Zeit als wissenschaftlicher Mitarbeiter am Institut für Technische und Numerische Mechanik (ITM) der Universität Stuttgart. Ich werde die fünf Jahre, in denen ich dort arbeiten und forschen durfte, als wundervolle Zeit in Erinnerung behalten und blicke gerne darauf zurück.

Auch wenn sich das Verfassen meiner Doktorarbeit für mich oft wie ein großer Kraftakt angefühlt hat, so konnte ich mir doch immer der Unterstützung vieler Menschen sicher sein, die in dieser Zeit ein wichtiger Bestandteil meines Lebens waren und es hoffentlich auch in Zukunft sein werden. Ihnen gilt es an dieser Stelle zu danken, denn ohne sie wäre diese Arbeit wohl gar nicht – oder wenigstens unter deutlich unangenehmeren Umständen – zustande gekommen.

Mein Doktorvater, Michael Hanss, hatte hierbei sicherlich den unmittelbarsten Anteil. Ich habe seine lockere Art der Betreuung seiner Doktoranden als sehr angenehm empfunden. Diese äußerte sich insbesondere durch einen großen Vorschuss an Vertrauen und viele Freiheiten, die er mir jederzeit gewährte und die es mir ermöglichten, unbeirrt an meinen Ideen zu arbeiten, nachdem ich meinen Weg gefunden hatte – eine Promotion muss am Anfang ja bekanntlich wehtun. Auch wenn er sich bestimmt manchmal gewünscht hätte, diese Ideen wären etwas weniger 'verquast', ließ er mich meinen eigenen Weg gehen und zeigte trotzdem immer großes Interesse und Begeisterung für meine Forschung. Auch meine Mitberichter haben ihre Rollen mehr als erfüllt und mich weit über das übliche Maß hinaus in meiner Arbeit unterstützt. So hat mir Scott Ferson in unzähligen virtuellen Runden zusammen mit seinen Mitarbeitern – an dieser Stelle sei auch Ander Gray, Alexander Wimbush und Marco de Angelis gedankt – viele kritische, und dadurch wertvolle, Fragen gestellt und Anmerkungen geliefert, um meine Ergebnisse zu durchleuchten. Vor allem aber hat er mich immer wieder dazu herausgefordert, diese Ideen aufs Neue zu formulieren bis sich schlussendlich die hier dargelegte Form herauskristallisiert hatte. Auch wenn ich ihn vermutlich bei der Erfüllung seines Wunsches, ich möge eine allgemeinverständliche Theorie formulieren, um sie einem breiteren Publikum zugänglich zu machen, im Stich gelassen habe – er hat seinen Unmut darüber regelmäßig zum Ausdruck gebracht (und das zu Recht!) –, so glaube ich doch, dass das Ergebnis insbesondere durch sein Mitwirken an Qualität gewonnen hat. Ihm und mir, aber sicherlich auch allen anderen, die an ähnlichen Theorien arbeiten, steht eine Theorie der Unsicherheiten vor Augen, die belastbar und allgemeingültig, aber eben auch elegant und allgemeinverständlich ist. Allein der Umfang dieses Dokuments spricht jedoch offensichtlich gegen die Einfachheit der Materie und in diesem Sinne bin ich wohl an seinen Ansprüchen gescheitert, eine so geartete Theorie darzulegen. Mir war das bereits vor ihm klar, aber ihm muss zugutegehalten werden,

dass er bis zuletzt an meine Fähigkeiten geglaubt hat. Auch von Ryan Martin habe ich jederzeit großes, und vor allem ehrliches, Interesse an meiner Arbeit erfahren und viel Zuspruch und Ermunterung bekommen. Sein größtes Kompliment an mich ist in jedem Fall, dass er Teile meiner Arbeit bereits kurz nach der Lektüre meiner Dissertation, in seiner eigenen Forschung weiterentwickelt hat. Dadurch hat sich, so hoffe ich, mein Wunsch erfüllt, einen kleinen, aber bleibenden, wissenschaftlichen Beitrag geleistet zu haben. Ich vertraue darauf, dass Ryan in der Lage ist, sowohl meine als auch seine Ergebnisse und die von anderen in der von Scott geforderten Form zusammenzuführen.

Gewidmet habe ich diese Dissertation den beiden Frauen, die mich dorthin gebracht haben, wo ich heute bin: meiner Mutter und Nana, meiner Großmutter. Schon während meiner Grundschulzeit war Nana klar, dass 'der verkopfte Junge' einmal promovieren würde, wenn sie auch noch nicht wusste, worüber – sie hat nie an mir und meinen Fähigkeiten gezweifelt und mich immer unterstützt. Meine Mutter hat meine Laufbahn überhaupt erst ermöglicht, indem sie das Wohl ihrer beiden Söhne zu jeder Zeit über ihr eigenes gestellt hat. Eine glückliche Kindheit und Jugend sind ein wertvoller Schutz gegen die Unwägbarkeiten des Lebens und in meiner Familie habe ich immer den nötigen Rückhalt gefunden. Dazu gehören auch mein Bruder, dessen Anteil in Form hochwillkommener Ablenkung nicht zu unterschätzen ist, und mein Vater, der mein Interesse an den Natur- und Technikwissenschaften immer gefördert hat. Auch meine Gasteltern, Dorothea und George van Ravenswaay, mitsamt ihren Eltern, Kindern und Geschwistern sollen hier nicht unerwähnt bleiben. Ohne zu übertreiben kann ich sagen, dass ich mit ihnen eine zweite Familie in Kanada gefunden habe, und ohne sie wäre die vorliegende Schrift wohl in deutscher Sprache verfasst worden.

Das ITM ist sicherlich kein Ort, an dem man fünf Jahre tätig sein kann ohne in anfänglichen Kollegen spätere Freunde zu finden. Diese bildeten für mich alleine schon durch ihre vergleichbare Situation eine emotionale Stütze, ich konnte mit ihnen aber auch wunderbar fachlich zusammenarbeiten. Dazu gehören u.a. Alexander Brauchler, Benjamin Fröhlich, Dominik Hamann, Steffen Maier, Patrick Schmid, Georg Schneider und Elizaveta Shishova. Zuvorderst hat mich Markus Mäck über den Großteil meiner Promotion hinweg begleitet, zunächst als Betreuer meiner Masterarbeit und dann als enger Forschungspartner, der meinen gelegentlich überschwänglichen Enthusiasmus ertragen musste. Ich werde die gemeinsamen Konferenzen, Fach- und Privatdiskussionen mit ihm in bester Erinnerung behalten. Die feste Konstante während meiner universitären Laufbahn war Henrik Ebel. Wir haben über mehr als zehn Jahre hinweg zusammen studiert, gearbeitet und promoviert. Sein Beitrag zum Gelingen meiner Promotion ist in vielerlei Hinsicht unschätzbar und ich bin dankbar, ihn so lange als Mitstreiter an meiner Seite gehabt zu haben.

An dieser Stelle sollte klar geworden sein, dass das ITM für mich weit mehr als eine bloße Forschungs- und Arbeitsstelle war. Überhaupt hat mich erst Jörg Fehr dorthin geholt, der meinen Weg auch später mit viel Enthusiasmus und Unterstützung begleitet

hat. In vielerlei Hinsicht ist die familiäre Atmosphäre am ITM aber auch seinem Leiter, Peter Eberhard, zu verdanken, der seinen Mitarbeitern jederzeit den Rücken freihält und ihnen die notwendigen Rahmenbedingungen schafft, damit sie sich und ihre Forschung frei entfalten können. Er hat mir eine einmalige Erfahrung ermöglicht, um die mich viele andere Doktoranden sicherlich beneiden. Ebenfalls bedanke ich mich bei Pascal Ziegler für viele interessante Diskussionen, in denen teils sehr unterschiedliche Sichtweisen aufeinander getroffen sind. Auch die von mir betreuten studentischen Arbeiten von Andreas Baumann, Sebastian Feige, Lennart Frie, Julius Herb, Tom Könecke, Kim Kröner, Nils Poslovski, Nathanael Römheld, Henrik Schühle und Sophie Schütt haben mir erheblich beim Verfassen meiner Dissertation geholfen.

Das ITM und die Universität Stuttgart werden für mich nicht zuletzt immer auch deswegen eine besondere Bedeutung haben, weil ich dort meine Partnerin, Luzia Hahn, kennenlernen durfte. Sie hat zum Abschluss der Promotion nicht nur für den Erhalt meiner Vitalfunktionen gesorgt, sondern mir jederzeit viel Verständnis entgegengebracht und Rückhalt gegeben. Ihren Beitrag werde ich wahrscheinlich erst vollständig nachvollziehen können und zu schätzen wissen, wenn wir einmal die umgekehrten Rollen haben. In jedem Fall freue ich mich auf die nächsten Kapitel mit ihr, die wir dann gemeinsam in einem anderen Buch schreiben werden.

Natürlich müsste ich hier noch weitaus mehr Personen danken, aber einen Anspruch auf Vollständigkeit könnte wohl keine Liste erheben. Nichtsdestotrotz möchte ich auch folgende Personen erwähnt wissen, die mir in der einen oder anderen Weise auf meinem Weg zur Seite gestanden haben: Gudrun und Hellmuth Hose, Therese Hose, Ruth Jähnig, Oliver Pantle, Bettina und Michael Schütterle, Louisa Bergmann, Emily Höhning, Benedikt Konersmann, Marcel Konstantinov, Conny Le, Catharina Zipper, Nikolai Arnaudov, Tobias Kattman, Jakob Kirch, Raphael Leiteritz, Jakob Rempel, Frank Schneider, Katrin Stollenmeier, Tobias Willmann, Marion Seitz, Michael Blasek und Hildegard Bauer.

Vielen Dank euch allen!

– Für Mama und Nana

Contents

Zusammenfassung

Numerische Simulationen sind ein integraler Bestandteil technischer Entwurfsprozesse für verschiedenste Arten von Systemen. Die dazu verwendeten Modelle zeichnen sich häufig dadurch aus, dass sie eine Vielzahl von Effekten berücksichtigen, die eine möglichst große Realitätsnähe zum Ziel haben. Aufgrund ihrer stetigen Weiterentwicklung und der damit einhergehenden zunehmenden Berücksichtigung auch weniger einflussreicher Effekte und Dynamiken wächst die Komplexität dieser Simulationsmodelle meist kontinuierlich.

Beim Abgleich solcher Modelle mit Messdaten aus den abbgebildeten physikalischen Systemen sind zwei Effekte häufig zu beobachten. Zum einen ist die präzise Bestimmung bzw. die Spezifikation aller darin enthaltenen Parameter aus verschiedenen Gründen oft nicht möglich – teilweise drücken diese auch überhaupt keine physikalischen Größen aus, die sich z.B. messen ließen. Zum anderen beinhalten viele physikalische Prozesse ein natürliches Maß an natürlicher Variabilität, die sich auch mit hochgenauen deterministischen Modellen nicht beschreiben lässt. Sollen nun die realen Messungen zur Kalibrierung der unbekannten Parameter verwendet werden, ist offensichtlich, dass verrauschte Daten niemals eine exakte Bestimmung erlauben. Dadurch bedingt können wiederum keine beliebig genauen Aussagen, insbesondere Vorhersagen, über das Systemverhalten getroffen werden. Solche Beobachtungen bieten einen direkten Zugang zur Unsicherheitsanalyse, die zum Ziel hat, ebendiese Unsicherheiten zu beschreiben, zu quantifizieren und damit zu rechnen.

Diese Arbeit beschäftigt sich mit possibilistischen, d.h. möglichkeitsbasierten, Methoden zur statistischen Inferenz mit unscharfen Wahrscheinlichkeiten und ist einem jungen Teilgebiet der Statistik und Unsicherheitsforschung zuzuordnen, welche ihren Ursprung in der Evidenztheorie nach Dempster und Shafer, aber auch in der Theorie der unscharfen Mengen von Zadeh hat. Dazu wird ein ganzheitlicher Ansatz verfolgt, der eine solide mathematische Basis mit einer praktischen Umsetzung verknüpft. In diesem Sinne liefert diese Arbeit einen Beitrag zur angewandten Theorie der unscharfen Wahrscheinlichkeiten, indem sie theoretische und numerische Erkenntnisse mit Anwendungen aus den Ingenieurwissenschaften verknüpft und aufzeigt, wie possibilistische Methoden gewinnbringend dort eingesetzt werden können.

Im ersten Kapitel der Arbeit wird die Möglichkeitstheorie historisch und philosophisch in die Theorie der Unsicherheitsanalyse bzw. in die zugrundeliegende Statistik eingeordnet. Die verschiedenen Arten von Unsicherheiten, insbesondere epistemische und aleatorische, werden erörtert, und es werden verschiedene mathematische Beschreibungsformen – Intervalle, scharfe und unscharfe Wahrscheinlichkeiten, und sogenannte Möglichkeiten – erläutert.

Der grundlegenden Behauptung des ersten Kapitels, nämlich dass die Möglichkeitstheorie sich zur Beschreibung von unscharfen Wahrscheinlichkeiten, also Mengen von Wahrscheinlichkeitsmaßen, eignet, wird sodann im zweiten Kapitel nachgegangen. Beginnend mit einem axiomatischen Aufbau der Möglichkeits- oder Possibilitätstheorie wird diese auf

vielfältige Weise in Bezug sowohl zur klassischen Wahrscheinlichkeitstheorie als auch zu anderen Theorien zur Beschreibung unscharfer Wahrscheinlichkeiten gesetzt. Ein erstes Hauptresultat ist hierbei die (Imprecise-Probability-to-Possibility-)Transformationsvorschrift, die sich mithilfe weniger fundamentaler Prinzipien motivieren lässt und es erlaubt, ebendiese unscharfen Wahrscheinlichkeiten possibilistisch zu beschreiben.

Diese Transformationsvorschrift stellt auch im Weiteren, insbesondere im dritten Kapitel, ein essentielles Werkzeug im Umgang mit unscharfen Wahrscheinlichkeiten dar. Dort wird aufgezeigt, wie sich verschiedenste Arten von (unscharf-)probabilistischen Informationen, z.B. stochastische Dominanz, unscharfe Erwartungswerte, aber auch präzise Wahrscheinlichkeitsverteilungen, oder eben die Abwesenheit solcher Informationen modellieren lassen. Weiterhin wird aufgezeigt, wie sich auf Basis des neu entwickelten, impliziten Erweiterungsprinzips mit Möglichkeitsverteilungen rechnen lässt, wie sich diese kombinieren lassen und wie sich verschiedene Arten von Abhängigkeiten beschreiben lassen.

Diese Vorüberlegungen bereiten die Grundlage für eine neue Theorie possibilistischer Inferenzmodelle im vierten Kapitel. Diese bilden den Übergang zur Statistik und erklären, wie sich Informationen aus Daten gewinnbringend in possibilistischen Beschreibungen verwenden lassen. Der Zusammenhang zwischen jenen Modellen und diesen Beschreibungen bildet der sogenannte Pivot-Schritt, welcher klassische Konzepte der (allgemeinen) Inferenzmodelle nach Martin und Liu und der Möglichkeitstheorie zusammenführt und dadurch einen neuartigen Zugang zur Statistik – insbesondere der frequentistischen – schafft.

Ein Schwerpunkt dieser Arbeit ist die numerische Umsetzung der possibilistischen Unsicherheitsanalyse, die auf mehrere Arten erfolgen kann, z.B. mittels sampling-basierter oder intervall-basierter Verfahren. Beide werden im fünften Kapitel mit Bezug auf ihre Vor- und Nachteile im Detail besprochen und es werden Methoden zur Implementierung des Erweiterungsprinzips und der Imprecise-Probability-to-Possibility-Transformation bzw. ihrer Varianten vorgestellt. Zusammengefasst zeichnen sich Samplingverfahren sowohl durch ein hohes Maß an Flexibilität als auch durch eine unkomplizierte Implementierung aus; der Vorteil intervallbasierter Methoden liegt in konservativen Berechnunen und Abschätzungen, die sich insbesondere zur Anwendung auf sicherheitskritische Systeme eignen.

Im sechsten Kapitel werden die theoretischen und numerischen Resultate auf ein praktisches Beispiel aus der Regelungstechnik angewandt, nämlich das Filterproblem, in dem aus Beobachtungen eines dynamischen Systems und seiner mathematischen Beschreibungen der derzeitige Systemzustand abgeleitet werden soll. Dazu wird ein possibilistischer Partikelfilter hergeleitet, wobei wiederum die vorteilhaften Eigenschaften der sampling-basierten Umsetzung des Erweiterungsprinzips ausgenutzt werden. Anhand eines Lokalisierungsbeispiels aus der Robotik wird die Funktionsweise dieses Filterprinzips demonstriert.

Das letzte Kapitel beginnt mit einigen Schlussbemerkungen und endet mit einer Erörterung von möglichen zukünftigen Forschungsthemen im Bereich der Möglichkeitstheorie.

Abstract

Numerical simulations are an integral part of the design process for various types of systems. The models used for this purpose are often characterized by many effects they take into account intended to achieve high-fidelity, i.e. being as close to reality as possible. Due to continuous refinements and the ensuing increased consideration of less influential effects and dynamics, the complexity of these simulation models usually grows over time.

Two effects can often be observed when checking such models against measured data from the described physical systems. On the one hand, a precise specification of all parameters in these models is often impossible for various reasons—sometimes, they do not even describe any physical quantities that could be measured. Moreover, many physical processes contain a natural degree of variability which cannot be described by deterministic models, not even by highly accurate ones. If actual measurements are to be used to calibrate the model, i.e. its unknown parameters, it is evident that noisy data will never allow an exact inference. This inexactness, in turn, implies that arbitrarily accurate statements, especially predictions, cannot be made about the behavior of the physical system. Such observations provide a direct motivation for uncertainty quantification, which aims to describe, quantify, and compute these very uncertainties.

This thesis considers possibilistic methods for statistical inference with imprecise probabilities. In this sense, it belongs to a young subfield of statistics and uncertainty quantification which has its origins in the Dempster-Shafer theory of evidence and the theory of fuzzy sets by Zadeh. More precisely, it makes a contribution to the applied theory of imprecise probabilities by linking theoretical and numerical insights with practical applications in engineering and showing how possibilistic methods can be used profitably.

In the first chapter, possibility theory is contextualized historically and philosophically among the various competing theories of uncertainty quantification and statistics. Different types of uncertainty, primarily epistemic and aleatory ones, are discussed, and various mathematical frameworks for their description—intervals, precise and imprecise probabilities, and so-called possibilities—are motivated and introduced.

The fundamental claim of the first chapter, namely that possibility theory is suitable for describing imprecise probabilities, i.e. sets of probability measures, is then investigated in the second chapter. Beginning with an axiomatic approach to possibility theory, it is linked to classical probability theory and other theories of imprecise probabilities next. The first main result is the Imprecise-Probability-to-Possibility Transformation, which can be motivated by a few fundamental principles and allows to describe imprecise probabilities possibilistically.

This transformation rule is also an essential tool for dealing with imprecise probabilities in the following, especially in the third chapter. Therein, it is shown how different kinds

of (imprecise-)probabilistic information, e.g. stochastic dominance, imprecise expected values, and precise probabilities, or the absence of such information, can be modeled. Moreover, it is shown how to compute with possibilities based on the newly derived implicit extension principle, how they can be combined, and how different kinds of dependencies between imprecise or random variables may be described.

These preliminary considerations prepare the ground for a novel theory of possibilistic inferential models in the fourth chapter that constitute the connection to statistics and explain how data can refine possibilistic descriptions. This connection is established via the Pivotal Step, which brings together concepts of (general) inferential models by Martin and Liu and of possibility theory, thereby creating a novel approach to statistics—in particular, to frequentist inference.

This thesis also focuses on the numerical implementation of possibilistic calculus, which is using both sampling-based and interval-based methods. Both methods are discussed in detail in chapter five regarding their advantages and disadvantages, and methods for implementing the extension principle, the Imprecise-Probability-to-Possibility-Transformation, and its variants are given. In summary, sampling methods are characterized by a high degree of flexibility and a straightforward implementation; the advantage of interval-based methods lies in conservative computations, which are particularly suitable for application to safety-critical systems.

In the sixth chapter, the theoretical and numerical results are applied to a practical example from control engineering, namely the filtering problem, in which the current state of a system is to be inferred from observations of a dynamic system and its mathematical description. For this purpose, a possibilistic particle filter is derived, exploiting the advantageous properties of the sampling-based implementation of the extension principle. Finally, a localization example from robotics is intended to demonstrate the properties and functionality of this filter.

The final chapter begins with some concluding remarks and ends with an outlook on potential future research topics in the area of possibility theory.

Chapter 1

Introduction

> Es geht um die Verlässlichkeit von Risikoannahmen und um die Verlässlichkeit von Wahrscheinlichkeitsanalysen. Denn diese Analysen bilden die Grundlage, auf der die Politik Entscheidungen treffen muss, ...
>
> Angela Merkel,
> Der Weg zur Energie der Zukunft (June 9, 2011)

Nobody is perfect. There is, perhaps, no better way to describe the raison d'être of the field of uncertainty quantification (UQ) than this century-old saying.

Throughout the entirety of their existence, humans have grappled with their ignorance producing such famous sayings as 'I know that I know nothing', which is attributed to Socrates and has found its way into our time. Socrates' method of choice for reducing his ignorance—and that of his contemporaries—was reportedly to ask many questions. In his view, the exchange of information and knowledge, and the subsequent questioning of one's own beliefs, was fundamental for reducing ignorance or uncertainty.[1]

On an abstract level, we might all be considered agents in a world whose past, present and future are, to a large extent, unknown to us. In order to successfully navigate it, we require mental models of this world that we are a part of—not least to anticipate the consequences of the decisions we make and, more importantly, of the actions we take. We calibrate these mental models with the evidence we obtain via our senses.

This evidence, however, may be misleading, and the resulting beliefs and predictions might turn out to be erroneous: The sky might have been bright and clear for the past four weeks leading one to predict that it will also be like this tomorrow—only for them having to cancel a hiking trip at the last minute. Alternatively, the available evidence might allow for several competing models to be true: A simple headache could be caused by dehydration,

[1] Hadot, P.: Philosophy as a Way of Life. Wiley, 1995.

a brain tumor, or something entirely else.[2] In other words, there is uncertainty of varying degrees associated with a process generating evidence, e.g. the weather, and there is uncertainty associated with the inferences this allows one to draw, e.g. a medical diagnosis.

Often, the human brain learns from evidence and accounts for some of the implied uncertainty by itself—even without much training. We know from experience that the weather of the past is not a definitive indicator of that of the future, and we are aware that not every headache is a death sentence. Nevertheless, with the ever-increasing efforts in developing artificially intelligent machines[3] and a future that could see robots making essential decisions without human supervision,[4] one should simultaneously devise ways to teach them how to reason under uncertainty—preferably in a comprehensive manner, such that human decision-making may also benefit. This bonus would not be a detriment, for there are many examples where the human perception of risk and uncertainty failed, e.g. in the Challenger disaster.[5]

Whereas humans typically do not have to rely on pure numbers for reasoning under uncertainty, mathematics is the traditional language of machines and computers. But how should we measure/quantify (un-)certainty, i.e. how should we put a number on it, and, more importantly, what conclusions should we draw concerning our beliefs about the past and present, and for our predictions of the future, i.e. how should we adjust that number, e.g., in the light of new evidence?

UQ can be summarized as answering precisely these questions. However, it requires a mathematical formalism, a language of uncertainty in which to ask these questions and formulate the answers.

1.1 Types of Uncertainty

The term 'uncertainty' generally refers to a property that is attached to hypotheses and predictions about the world's past, current or future state. It is usually not ontological but rather an expression of ignorance—perhaps with the notable exception of quantum physics. The reasons for uncertainty are manifold, and two are of main interest to the following discussion.

The most well-known type of uncertainty concerns the results of coin tosses and dice rolls, the winning numbers in the lottery, the exact number of cars that will pass through some

[2] `https://www.sciencedaily.com/terms/headache.htm` (accessed on December 21, 2021)

[3] `https://www.zeit.de/mobilitaet/2021-11/autonomes-fahren-kuenstliche-intelligenz-verkehrssicherheit-unfall` (accessed on December 21, 2021)

[4] `https://www.philomag.de/artikel/algorithmen-entscheiden-nichts` (accessed on December 21, 2021)

[5] `https://spectrum.ieee.org/the-space-shuttle-a-case-of-subjective-engineering` (accessed on December 21, 2021)

street tomorrow morning, or the exact amount of rain on a particular day in the future in Stuttgart. For instance, in the first case, it is not unimaginable that, if one had a very accurate model of the dice, its environment, the rolling motion, etc., one would be able to exactly predict the number of eyes shown. However, the level of scrutiny required to accurately predict this outcome is quasi-infeasible because it would require of level of modeling detail that is hardly achievable in reality. Instead, one may take a broader view and consider the dice to be subject to *irreducible* variability or *chance* expressing our inability to precisely predict how many eyes the dice will show next. This (possibly only perceived) randomness is often referred to as *aleatory uncertainty*. A useful description of this uncertainty would identify patterns or laws governing this randomness.

Still, there are more situations unrelated to chance, where it is not possible to make a definitive statement due to a *lack of knowledge*, also referred to as *epistemic uncertainty*. This lack of knowledge typically concerns the past or current state of the world, e.g., the percentage of the population with a certain blood type, with an income lower than a certain threshold or its percentage of vegetarians. Similar to the root cause of a headache one is experiencing on a particular day, these values do not necessarily exhibit variability: By surveying all members of the population, one would be able to arrive at the precise percentage but one may also estimate them with varying degrees of accuracy. Therefore, this type of uncertainty is often referred to as *reducible*, which should, however, not be mistaken for eliminability.

These two types of uncertainty are prevalent in engineering [HeltonJohnsonOberkampf04]; in practice, a mix of both these types of uncertainty is often present. As an example, consider the result of drawing a ball from an urn with an unknown number of red and blue balls, where some prior draws have been observed, or localization techniques based on noisy distance measurements and triangulation, as discussed later in this thesis. However, other types of uncertainty, stemming e.g. from the imprecision in human speech introduced by hedge words such as ‘very’, ‘few’, or ‘huge’, may also be considered.

In conclusion, several uncertainties of different shape and form may be attached to our models of the world. Due to these different appearances, in principle, one should be open to the idea of describing, e.g., aleatory and epistemic uncertainty with different languages [FersonGinzburg96]. Of course, a unification of these languages, i.e. a universal language of uncertainty, is a desirable goal, but the biggest concern should be a faithful description. Any mathematical formalism used to describe uncertainty should be judged by the fidelity and usefulness of the inferences it allows one to make—just like any other mathematical model. Indeed, the correct classification of uncertainties becomes somewhat insignificant if one can only find a good description.

1.2 Languages of Uncertainty

At the heart of UQ lies the question of which model ought to be used to describe uncertainty.Mathematics allows for the formulation of many different kinds of models, which might in its own right be referred to as the essence of science.[6] Of course, it is never expected that a model captures reality perfectly. However, different models may differ severely in quality, and all models should be judged by their fidelity, including both a faithful and an expressive representation.

In many disciplines, modeling is often restricted to deterministic representations of the system under consideration. Carefully applying the bequeathed formalisms typically yields a set of equations that describe the system under investigation. Moreover, manipulating and solving these equations enables one to analyze the system and make, e.g., predictions about its future behavior. Intuitively, it is clear that these results will never quite capture all of the facets of the entire system, which is owed to simplifications in the modeling process, e.g. linearization, a general lack of knowledge about specific sub-processes, e.g. damping, external disturbances, unforeseen loads, and so on. Most modelers will intuitively feel that one ought to allow for some disagreement between the model and reality. However, how much deviation from the predicted system response is acceptable? How should noisy measurements be included? Which value from a set of parameters that more or less produce the measured output is correct?

Statistical modeling may be considered an extension of deterministic modeling that addresses the mismatch between deterministic models and the world they are intended to described.

1.2.1 Sets and Intervals

The oldest and perhaps conceptually most straightforward uncertainty description is given by simple sets of possible values, particularly intervals. They constitute an elementary description of uncertainty since, e.g., in the interval case, they only provide bounds on the values of a variable. Interval analysis has found successful applications in a variety of fields, in particular in those fields close to robotics [JaulinEtAl01] and validated numerics [AuerKielRauh13], and will play an auxiliary role in this thesis.

Consider the floating-point representation of π on a rudimentary computer capable of representing only integers. On such a machine, e.g., the lower and upper bounds, $\pi \geq 3$ and $\pi \leq 4$, could be stored. A faithful program should communicate the resulting uncertainty concerning the area $A = \pi r^2$ of a circle with a given radius of $r = 2\,\mathrm{m}$ by returning the lower bound of $A \geq 12\,\mathrm{m}^2$ and an upper bound of $A \leq 16\,\mathrm{m}^2$. This would,

[6] `https://www.zeit.de/kultur/2021-05/wissenschaft-corona-krise-modelle-forschung-klimawandel` (accessed on December 21, 2021)

e.g., make it impossible to definitively evaluate the hypothesis $A \leq 15\,\mathrm{m}^2$. This situation is characterized by a (reducible) lack of knowledge relevant to the hypothesis: Even though a definitive answer exists, it cannot be identified. If, however, the program ran on a better computer capable of also representing the first decimal, it would be able to reduce the uncertainty about π by bracketing it with $\pi \in [3.1, 3.2]$ such that $A \in [12.4, 12.8]$ and arrive at the answer 'yes' regarding the hypothesis whether $A \leq 15\,\mathrm{m}^2$.

The example also shows that intervals have limited expressiveness because such descriptions do not contain additional information, e.g., about frequencies of occurrence in a process subject to variability. These properties eliminate the need to think about degrees of uncertainty and make intervals suitable to describe extreme/severe uncertainty when not much else can be said about a variable, making it particularly attractive to describe entirely epistemic uncertainty as in the above example.

1.2.2 Probability Theory

Probability is widely accepted as the appropriate description of variability and randomness. The description of chance, particularly of games of chance, was the driving force behind the development of probability theory, and the notions of chance and probability were used interchangeably by early scholars, such as Laplace and Bayes. The latter unequivocally stated that "by chance I mean the same as probability" [Bayes63, p. 376].

One of the earliest problems concerned the question of how to fairly split the money pool if a game consisting of several rounds had to be canceled prematurely, e.g., if, in a best of five between two opponents, where one would have to win at least three rounds, the game is canceled after player A has won two rounds, and player B has won one round only. The outcome of these rounds, and thus of the game, is subject to uncertainty because, of course, the outcomes of individual rounds, and of the game, cannot be predicted precisely. Experience suggests that a definitive answer would not reflect reality. If it did, the purpose of the game would arguably be defeated.

Probabilistically, one could describe this by the probability $p \in [0, 1]$ that player A wins the next round, and the corresponding probability $1 - p$ of the counter-event, i.e., player B wins the round. The minimal consensus is that $p = 1$ indicates total certainty that player A will win and, conversely, $p = 0$ that player A will lose. Especially degrees between zero and one can be interpreted in various ways, e.g., as limit frequencies or as expected earnings. In the first case, they denote the percentage of rounds player A will win in a hypothetically infinite sequence of rounds. In the second case, it is, the prize a gambler would be willing to pay for a bet returning one Euro if player A wins (and zero Euros otherwise). This prize can, of course, be subjective; therefore it is often referred to as the *subjectivist* interpretation of probability, whereas the former is usually referred to as the *frequentist* interpretation. Even though both interpretations can lead to the

same description, they differ in what they mean by probability. Moreover, this is not an exhaustive list of the various interpretations of probability that have been proposed throughout history [Hájek19].

1.2.3 Imprecise Probabilities

The claim of many probabilists, more precisely that of Bayesians, is that probability is the only language suitable to describe uncertainty, regardless of its (aleatory or epistemic) nature. When chance or randomness is the cause of uncertainty, it is widely accepted that probability is indeed the appropriate measure to describe the variability sufficiently. However, concerning epistemic uncertainty and personal belief, the matter is entirely different. For an exemplary discussion, refer to [Shafer90] and the follow-up papers in the same issue.

Sometimes it is not straightforward to find a precise probabilistic model if the relevant information is unavailable. For instance, in the above problem of the discontinued game of chance, the actual value of p is unknown.

The frequentist interpretation would require them to have played an infinite amount of rounds prior to this point in order to postulate a precise probabilistic model, i.e., in order to specify p, and ultimately, how to split the pot. But the only evidence is given in the form of three rounds (player A winning twice, player B winning once). Addressing this issue, frequentists would rely, e.g., on hypothesis tests and confidence intervals to express their uncertainty about p.

However, the subjectivist interpretation views probability as an expression of personal belief that everybody is allowed to have about virtually all hypotheses—as long as it adheres to the laws of probability. One way to infer this belief is by offering gambles to the so-called subject and asking them for a fair price. For instance, if one explained the above game of chance to a subjectivist, more precisely to a Bayesian, and then proposed a gamble that pays one Euro if player A wins and nothing otherwise, they would have to be able to state a price they would be willing to pay—under the condition that they would also have to be willing to sell that gamble for the same prize to somebody else. That is, they would be willing to buy the gamble for p Euros from someone else or sell it to them for p Euros. Seeing that the game's rules do not favor one player from the start, they would typically assume that $p = 0.5$ Euros is a fair price—both for selling and buying this gamble. They would expect neither to lose nor to win money by offering or taking this bet. Both outcomes (player A or B winning) are—in the absence of any evidence of the game being fixed or some player having a superior strategy—equally likely. Upon seeing the results of one or several rolls, they might then update their belief according to Bayes' theorem. Seeing that player A has won once more than player B would then alter the belief in A's favor, which would usually result in $p > 0.5$.

It has, however, been argued that human belief does not necessarily obey the laws of probability [Bradley19], i.e. it does not always adhere to its additive structure, and the Bayesian approach does not accurately describe how humans reason under uncertainty. A possible fix for this issue is a normative view of Bayesian belief: Rational decision-makers *should* form their beliefs and act according to the laws of probability (and of Bayes' Theorem). Failing to do so would make them vulnerable to a so-called Dutch book [Vineberg16] potentially losing them an infinite amount of money to a cleverer gambler.

Still, this leaves room for criticism, in particular of the role of the prior belief, ultimately leading to such undesirable phenomena as the False Confidence Theorem [BalchMartinFerson19], which will be explained in Chapter 4. Many more examples can be construed that point out possible deficiencies of probability as a representation of personal belief but shall not be repeated here.[7] As a result, many alternative theories have emerged in order to resolve these issues.

A less rigorous subjectivism is advocated by the Society for Imprecise Probabilities: Theory and Applications (SIPTA[8]). In his seminal book [Walley91], Walley, one of its forefathers, argues in favor of expressing one's lack of knowledge via *upper and lower previsions*, a generalization of expected values that can be inferred from not necessarily coinciding maximum acceptable buying prices and minimum acceptable selling prices for all kinds of gambles. These imply, amongst others, upper and lower bounds on probabilities, so-called *imprecise probabilities*.

A related theory enjoying broad appreciation in many communities[9] is the Dempster-Shafer Theory of Evidence. Building upon ideas of Dempster [Dempster67], Shafer [Shafer76] proposed a paradigm shift in statistics towards two measures of belief instead of just a single measure: The belief, which accounts for all the evidence that directly supports a proposition, and the plausibility, which accounts for all the evidence that does not explicitly rule this proposition out. The basic belief function is used to compute these measures, and by applying Demster's rule of combination (which he shows to be a generalization of Bayes' theorem), one can combine multiple pieces of evidence and refine one's personal belief.

While Shafer initially intended to develop a theory of belief, i.e. for quantifying epistemic uncertainty, it quickly became apparent that the framework he proposed could be used in order to also construct a more general theory of imprecise probabilities by viewing the belief measure as a bound for the lower probability and the plausibility measure as a bound for the upper probability. This inspired several other theo-

[7] A non-exhaustive collection of examples may be found under `https://en.wikipedia.org/wiki/Category:Probability_theory_paradoxes` (accessed December 24, 2021)

[8] `https://www.sipta.org/` (accessed December 24, 2021)

[9] Shafer himself discusses this success in [Shafer16].

ries, including, amongst others, probability boxes (p-boxes) [BeerFersonKreinovich13], monotone measures [BronevichKlir10], and clouds [Neumaier04], which are all related [KlirWierman99, DesterckeDuboisChojnacki08, TroffaesMirandaDestercke13]. The historical development can partially be traced in [YagerLiu08] and a more recent account thereof is to be found in [AugustinEtAl14].

In any case, the unique role of probability theory as the sole language of uncertainty, particularly of epistemic uncertainty, is disputed in the scientific community.

1.2.4 Possibility Theory

Possibility theory takes a unique role in the historical development, which is sketched above, as it originated from a different idea, namely that of fuzzy sets introduced by Zadeh in his seminal paper [Zadeh65]. Only later, it was discovered that certain aspects also fit into the framework of imprecise probabilities [DuboisPrade92].

Zadeh, troubled by entirely different issues than those that led Shafer to develop his theory of evidence, had proposed fuzzy sets as a natural way to model imprecision in human speech—yet another source of uncertainty distinct from aleatory and epistemic uncertainty. The fundamental idea is to admit gradual set membership instead of the classical notion of in or out. Therefore, he proposed a fuzzy set membership function, in essence, a generalization of the indicator function, which assigns degrees of membership to its elements. For instance, in the fuzzy set of 'warm temperatures', the temperatures of 0, 10, 20, 30 and 40 degrees Celsius might be associated with a membership degree of 0, 0.3, 0.8 1 and 0.5, respectively. This idea leads to a very different theory that has found many applications in logic [Zadeh88] but also engineering [Hanss05].

Again, the idea of a theory of possibility, viewing membership functions as the contour function of a possibility measure, put forth by Zadeh [Zadeh99], was not intended as a language for aleatory or epistemic uncertainty. In fact, tacing heavy opposition from statisticians saying that fuzzy set theory was superfluous and could not solve any problems manageable by probability theory, Zadeh always insisted on the distinct nature of uncertainty and imprecision in human speech, claiming that probability theory and fuzzy set theory are "complementary rather than competitive".[10] Nevertheless, this criticism also marks the beginning of the liaison between fuzzy sets and uncertainty theories.

Coming from an entirely different angle, namely that of statistics, Shafer had already expended an entire chapter of his book to the discussion of consonant belief functions, arguing that focal masses stemming from a simple piece of evidence should be nested [Shafer76, Chapter 10]. Moreover, he showed that, in this case, the corresponding plausibility measure

[10] An example of the scientific debate between the two factions can be found in [LavioletteEtAl95] and the follow-up articles, also by Zadeh [Zadeh95], in the same issue.

must be maxitive, i.e. the plausibility of the union of two events must be the maximum of their respective plausibilities, as opposed to the additivity of probability. Since the possibility measure proposed by Zadeh obeys the same law, the two measures are identical from a mathematical point of view. Subsequently, Dubois and Prade proposed viewing the possibility measure as an upper bound for the probability [DuboisPrade92] and advanced and formalized possibility theory with considerable effort and success as a theory of imprecise probabilities in numerous articles and books, e.g. [DuboisPrade88].

It should be noted that, nowadays, one can distinguish between two branches of possibility theory [DuboisPrade98]. To summarize, the qualitative branch is still very much connected to fuzzy sets, fuzzy logic and information theory, whereas the quantitative branch is concerned with UQ. This thesis focuses on the latter.

1.3 Possibilistic Uncertainty Quantification

Adopting this description-based point of view, this thesis does not discuss uncertainty and its correct quantification. It is a discussion of the language of possibility theory and what possibilistic models can describe. What possibility itself is, and whether it is suitable for quantifying uncertainty, is nicely summarized by

> "Probability is everything that fits the axioms."[11]

which is equally applicable to possibility. It is a language—a model that, under certain conditions, happens to describe the world.

To make this point even clearer: Partial differential equations, too, are a language. In mechanical engineering, they can be used to derive finite element methods that can successfully predict the behavior of mechanical systems. In other disciplines, they serve other purposes. The Maxwell equations describe electromagnetism, and the Schrödinger equation models quantum mechanical systems. Nevertheless, partial differential equations do not rely on these applications to exist. They can be postulated independently.

Consequently, this thesis derives possibility theory from some definitions, similar to the Kolmogorov axioms. It is, however, clear that an application to UQ is intended—but not in the usual way, as a description of chance of belief. It should be understood as an extension of probability theory, intended to describe uncertainty about probabilistic descriptions, i.e. to describe imprecise probabilities, and the inferences it enables one to draw, i.e. the *inferential models* that can be derived from such descriptions of imprecise probabilities. The precise meaning of 'possibility' in this thesis is entirely dependent on the meaning of probability. The only further prerequisite is the acceptance of the

[11] Scott Ferson (personal communication, 2020)

existence of distributional imprecision in statistical models. This will also yield two more interpretations of possibility: (data-based) confidence and predictions.

1.4 Composition and Contributions of this Thesis

Concluding this brief introduction and overview, the remainder of this thesis is organized as follows.

Chapter 2 presents possibility measures and probability measures as members of a larger class of capacities. This discussion results in the interpretation of possibility theory as a theory of imprecise probabilities under the fundamental concepts of consistency and specificity. Finally, the Imprecise-Probability-to-Possibility Transform, the fundamental tool for describing imprecise probabilities via possibilities in the remainder of this thesis, is derived from some principles. The most important conclusion of this chapter will be that possibility theory is well equipped to model the situation when the available information does not warrant a precise probabilistic model.

Chapter 3 further explores the role of possibilistic descriptions by some practical applications of this transform—in particular, how to reason with imprecise (random) variables that a possibility distribution can model. This includes the fundamental extension of such distributions (in a new generality) and their relation to stochastic dominance and imprecise expectations. Finally, dependency and interaction between such variables are investigated by reformulating some well-known dependency concepts from the theory of (imprecise) probabilities in a possibilistic manner.

Chapter 4 presents a theory of possibilistic statistics building upon the existing theory of inferential models. The different style of this chapter, including more direct quotes and philosophical arguments, is owed to the relative novelty of the topic. Possibilistic inferential models are, to the author's knowledge, little investigated; therefore, it aims at providing a rigorous theoretical basis, which, in turn, impedes an in-depth discussion of many advanced details. In this chapter, two well-known concepts from frequentist statistics, confidence and predictions, are revisited from a possibilistic point of view. It is shown how possibility theory offers a straightforward way of reasoning with them—suggesting that frequentist inference is, at heart, imprecise probabilistic. The novel tool for this purpose is the Pivotal Step (and its variants).

This thesis's most crucial finding is that reasoning with possibilities, confidence and predictions alike reduces to a few fundamental mathematical operations, most notably the membership transform and the extension principle. Chapter 5 discusses an efficient, sampling-based and a reliable, interval-based numerical strategy to implement them.

The accompanying examples throughout Chapters 2-5 are intended to show how the discussed concepts can be used in UQ.

Chapter 6 of this thesis concerns the application of possibilistic statistics to dynamical systems. The development of a (recursive) possibilistic (particle) filter in this chapter makes extensive use of the previous chapters. More importantly, it shows how possibility theory may be applied fruitfully to engineering, in this case to robotics, and how it can compete with existing Bayesian approaches in terms of expressiveness and ease of implementation as demonstrated by a final robot localization example.

The thesis concludes with a final discussion and outlook in Chapter 7.

Chapter 2

Possibilities and Imprecise Probabilities

> Wann immer einen die Dinge erschreckten, sei es eine gute Idee, sie zu messen.
>
> Daniel Kehlmann, *Die Vermessung der Welt*

This chapter is intended to provide an axiomatic approach to possibility theory and to introduce the basic notation, which has been adapted to be consistent with texts on probabilistic uncertainty quantification [Sullivan15]. A more thorough exposition of classical measure theory may be found in many standard works [Halmos13], and a more rigorous axiomatic approach to possibility theory is, e.g., provided by De Cooman [De Cooman97a, De Cooman97b, De Cooman97c].

2.1 Probability and Possibility Measures

In order to be able to relate probability and possibility measures to one another, initially, it is necessary to find a common framework. This is, e.g., accomplished through the notion of *capacities* [Choquet54], also known as *fuzzy measures* [KlirWierman99]. These are sufficiently general to yield both probability and possibility measures as special cases, thus, naturally lending themselves to this cause, and, at the same time, intuitive to understand. Other common frameworks, which would be equally suitable, include, e.g., random sets [Dempster67] and the Dempster-Shafer theory of evidence [Shafer76], or lower previsions [Walley91].

2.1.1 Measurable Spaces

Consider some unknown (past, present, or future) state of the world that one might wish to talk about. By definition, the *universal space*, *universe of discourse* or *sample space* Ω must be an exhaustive description thereof and contains a collection of all possible states, which can occur. These are the *elementary events* $\omega \in \Omega$. A so-called *event* is a subset $E \subseteq \Omega$ of the sample space. For instance, events concerning the external wind load on some structure can be expressed through subsets of the universal space $\Omega = \{\omega_l : l \in \mathbb{R}\}$ composed of the elementary events

$$\omega_l : \text{‘The wind load on the structure is } l \,\text{N.’} \tag{2.1}$$

for $l \in \mathbb{R}$. The event that the load exceeds 10 N is given by $E = \{\omega_l : l > 10\}$.

When Ω is finite or countable, one usually considers the set of all possible events, i.e. the power set $2^\Omega = \{E : E \subseteq \Omega\}$. However, when Ω is uncountably infinite, it makes sense to restrict oneself to only talk about a special class of subsets, so-called *measurable sets*. These measurable sets are collected in a *σ-algebra* $\Sigma \subseteq 2^\Omega$, which, by definition, contains the empty set, and is closed with respect to countable unions and complementation. This definition of a σ-algebra ensures that one can consider con- or disjunctive combinations of events and their opposites, i.e. the standard set operations, union, intersection and complement, on measurable sets and, again, obtain valid events, i.e. measurable sets. The tuple (Ω, Σ) is called a *measurable space*.

The *Borel σ-algebra* is the smallest σ-algebra on the universal set $\mathbb{R}$ and contains all open intervals (a, b) for $a, b \in \mathbb{R}$ with $a < b$. It includes many practically relevant sets: open, closed and one-sided intervals, unions and intersections thereof, single numbers, etc. It can, furthermore, easily be restricted to subsets $\mathbb{V} \subseteq \mathbb{R}$ or extended to higher dimensions, i.e. to $\mathbb{V} \subseteq \mathbb{R}^D$. The Borel σ-field corresponding to $\mathbb{V} \subseteq \mathbb{R}^D$ is denoted by $\mathbb{B}(\mathbb{V})$.

Furthermore, suppose that (Ω', Σ') is a second measurable space, e.g. $(\Omega', \Sigma') = (\mathbb{V}, \mathbb{B}(\mathbb{V}))$. A map $f : \Omega \to \Omega'$ is called a (Σ-Σ'-)*measurable function* if the pre-images of all $B \in \Sigma'$ are measurable with respect to the original measurable space (Ω, Σ), i.e. if $f^{-1}(B) \in \Sigma$. The concept of measurable functions is crucial for the definition of imprecise variables in Chapter 3. Most importantly, it enables one to express events of the form $E : f \in B$ which is an abuse of notation for $E = f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\}$. For instance, in the example above, the event concerning the external wind load can be expressed more concisely as $L > 10$ for the (measurable) function $L : \omega_l \mapsto l$.

The measurability of all sets and functions in this thesis is usually not considered explicitly unless it is essential, for it would seriously impede the readability of this thesis. In any case, albeit these definitions being very technical, they ensure a thorough definition of probability and possibility measures, and of imprecise variables.

2.1.2 Capacities

In order to be able to evaluate an event, its 'size' must be evaluated, a task that is accomplished by *capacities*, a generalization of *measures*.

A capacity is a function $\mathrm{M} : \Sigma \to \mathbb{R}$ on the measurable space (Ω, Σ) that assigns a numerical value to every measurable set in Σ. Well-known examples include the *cardinality*, i.e. the number of elements in a set, and the *Lebesgue measure*, i.e. the length, area, or volume of subsets of the Euclidean space $\mathbb{R}^D$. Such capacities may have a variety of properties.

Non-Negativity. Every measurable set $E \in \Sigma$ has a non-negative capacity $\mathrm{M}(E) \geq 0$.

Normality. The universal set has capacity $\mathrm{M}(\Omega) = 1$.

Completeness. The empty set has capacity $\mathrm{M}(\emptyset) = 0$.

Boundedness. The capacity of any measurable set $E \in \Sigma$ is bounded by $\mathrm{M}(E) < \infty$.

Monotonicity. For any two measurable sets $E_1, E_2 \in \Sigma$ with $E_1 \subseteq E_2$, their capacities satisfy $\mathrm{M}(E_1) \leq \mathrm{M}(E_2)$.

2-Monotonicity. For any two measurable sets $E_1, E_2 \in \Sigma$, their capacities satisfy $\mathrm{M}(E_1) + \mathrm{M}(E_2) \leq \mathrm{M}(E_1 \cup E_2) + \mathrm{M}(E_1 \cap E_2)$.

Self-Duality. The complement $\neg E$ of a measurable set $E \in \Sigma$ has capacity $\mathrm{M}(\neg E) = 1 - \mathrm{M}(E)$.

An example of a capacity that possesses all of these properties is the probability measure, but this thesis also considers other types of capacities. In any case, all investigated capacities will be non-negative, normal, complete, bounded and monotone.

These are intuitive properties, considering that capacities are intended to quantify the size of sets. Any set should be at least as big as any of its subsets; the set containing 'nothing', the empty set, should have size zero; and the size of the largest set, the universal set, should be normalized to one, such that the size of any other set is to be understood as a percentage of the size of Ω. In particular, this allows for intuitive comparisons both between measurable sets and between different capacities.

Notice that not all mnonotone capacities are 2-monotone. On the contrary, 2-monotonicity is a stronger form of monotonicity which follows by considering

$$\mathrm{M}(E_1) \leq \mathrm{M}(E_1 \cup E_2) + \mathrm{M}(E_1 \cap E_2) - \mathrm{M}(E_2) = \mathrm{M}(E_2) \tag{2.2}$$

since $E_1 = E_1 \cap E_2$ and $E_2 = E_1 \cup E_2$ for $E_1 \subseteq E_2$. Their generalization are so-called n-monotone and ∞-monotone measures [AugustinEtAl14]; the latter are also called *belief functions*.

What truly distinguishes capacities from each other, and, e.g., constitutes the fundamental difference between probability and possibility measures, is how monotonicity is achieved. In classical measure theory, all measures are σ-additive[12], yielding the probability measure in the normalized case. By relaxing the required additivity to monotonicity, and by generalizing the notion of a measure, other choices, such as (maxitive) possibility measures emerge. However, these measures are not generally self-dual anymore, as opposed to the probability measure, and dual measures, such as the (minitive) necessity measures, can be found. A recent in-depth discussion of capacities is provided by Denneberg [Denneberg94], and by Bronevich and Klir [BronevichKlir10] for 2-monotone measures in particular.

2.1.2.1 Interval Spaces

The perhaps most basic capacities are considered in interval analysis [JaulinEtAl01] where the information about the outcomes in Ω is simply modeled by a set $T \in \Sigma$ (usually an interval or a higher-dimensional box) to which the final outcome is certain to belong—without any further grading. The corresponding capacities $\Pi^{\text{Bool.}} : \Sigma \to [0,1]$ and $\mathrm{N}^{\text{Bool.}} : \Sigma \to [0,1]$, are the (Boolean) possibility and necessity measures, respectively—the same names that are given to the capacities in possibility theory, and indicate the close connection of the two theories. They are given by

$$\Pi^{\text{Bool.}}(E) = \begin{cases} 1 & \text{if } T \cap E \neq \emptyset \text{ and} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \mathrm{N}^{\text{Bool.}}(E) = \begin{cases} 1 & \text{if } T \subseteq E \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

for all $E \in \Sigma$. It is easy to check that these capacities are non-negative, normal, complete, bounded, monotone,[13] and dual in the sense that $\Pi^{\text{Bool.}}(E) + \mathrm{N}^{\text{Bool.}}(\neg E) = 1$ for all $E \in \Sigma$.

2.1.2.2 Probability Spaces

According to the well-known Kolmogorov axioms, a *probability measure* $\mathrm{P} : \Sigma \to [0,1]$ is a non-negative and normal capacity on the measurable space (Ω, Σ) that additionally exhibits

σ-Additivity. For unions of countable collections of pairwise disjunct measurable sets $E_k \in \Sigma$, their probability is given by $\mathrm{P}\left(\bigcup_{k\geq 1} E_k\right) = \sum_k \mathrm{P}\left(E_k\right)$.

If $\Omega = \{\omega_1, \omega_2, \ldots\}$ is a discrete space, then every probability measure is associated with a *probability mass function* $p : \Omega \to [0,1]$ with $\sum_{\omega_i \in \Omega} p(\omega_i) = 1$ and

$$\mathrm{P}(E) = \sum_{\omega_i \in E} p(\omega_i) \tag{2.4}$$

[12] In fact, 2-monotonicity is a weaker version of (σ-)subadditivity.
[13] The necessity measure is also 2-monotone.

for all $E \in \Sigma$. If Ω is continuous, then this concept may be generalized to a *probability density function*, see Section 3.1.4. The triple $(\Omega, \Sigma, \mathrm{P})$ is called a *probability space*.

From these axioms, the completeness, boundedness, (2-)monotonicity and self-duality of the probability measure follow immediately.[14] The space of all probability measures on (Ω, Σ) is denoted as $\mathbb{P}(\Omega, \Sigma)$.

The *conditional probability* of $E_1 \in \Sigma$ given $E_2 \in \Sigma$ with $\mathrm{P}(E_2) > 0$ is

$$\mathrm{P}(E_1|E_2) = \frac{\mathrm{P}(E_1 \cap E_2)}{\mathrm{P}(E_2)} \tag{2.5}$$

which can be interpreted as restricting the sample space Ω to the elementary events in E_2. The two events are called *stochastically independent* if $\mathrm{P}(E_1 \cap E_2) = \mathrm{P}(E_1)\mathrm{P}(E_2)$, which yields e.g. $\mathrm{P}(E_1|E_2) = \mathrm{P}(E_1)$. Knowledge or assumptions about the occurrence of E_2 do not change the probability of E_1, and vice versa.

This very basic introduction of Boolean possibility and necessity measures, and of probability measures will suffice for the purpose of this thesis. The following exposition of possibility theory may also be understood as a synthesis of interval analysis and probability theory, where probability mass is attached to nested collections of sets as opposed to the elementary events. A more general version thereof is the Dempster-Shafer Theory of Evidence [Shafer76] and the related theory of belief functions [DuboisPrade90, Cuzzolin13, MontesMirandaVicig19].

2.1.2.3 Possibility Spaces

The standard approach to possibility theory as discussed below well reflects that its roots lie in fuzzy set theory [Zadeh99] which is fundamentally based on so-called *membership functions* [Zadeh65, DuboisPrade20] that describe the degree of belonging to a fuzzy set.

Elementary Possibilities The *elementary possibility function* $\pi : \Omega \to [0, 1]$ on the measurable space (Ω, Σ), is the atom, i.e. the basic building block, of possibility theory. Its definition is commensurate with that of a fuzzy membership function: It need not satisfy many requirements, except that it be measurable and normal, i.e. that

$$\sup_{\omega \in \Omega} \pi(\omega) = 1. \tag{2.6}$$

In essence, all of (quantitative) possibility theory is connected to describing, inferring and combining these functions, reasoning with them, or manipulating them in more elaborate ways. They are also the primary objects of discussion in this thesis. The triple (Ω, Σ, π) is called a *possibility space*.

[14] The proofs of these properties appear in every standard text book on probability and statistics [Shao03].

Two special classes of elementary possibility functions deserve special mention: If an elementary possibility function is one everywhere on Ω, it is called *vacuous*; if it is exclusively $\{0,1\}$-valued it is called *quasi-vacuous*. Naturally, the latter definition includes the former, i.e., quasi-vacuous elementary possibility functions are also vacuous.

Possibility Measure Following standard expositions of possibility theory [DuboisPrade88, KlirWierman99], every elementary possibility function also induces a *possibility measure* $\Pi : \Sigma \to [0,1]$ on (Ω, Σ). The possibility of a measurable set $E \in \Sigma$ is simply defined as the supremum of its elementary possibilities, i.e.

$$\Pi(E) = \sup_{\omega \in E} \pi(\omega), \tag{2.7}$$

This definition is also consistent with the naming of the elementary possibility function; the possibility of an elementary event $\omega \in \Omega$ is given by $\Pi(\{\omega\}) = \pi(\omega)$.

The positivity, normality and boundedness of the possibility measure follow directly from the positivity, normality and boundedness of the elementary possibility function.

Regarding the completeness of the possibility measure it is easy to verify that, by convention,

$$\Pi(\emptyset) = \sup_{\omega \in \emptyset} \pi(\omega) = \sup \pi(\emptyset) = \sup \emptyset = 0. \tag{2.8}$$

The most fundamental observation about this capacity is that the possibility of the union $E_1 \cup E_2$ of two events $E_1, E_2 \in \Sigma$ is the maximum of their individual possibilities

$$\begin{aligned} \Pi(E_1 \cup E_2) &= \sup_{(\omega \in E_1) \vee (\omega \in E_2)} \pi(\omega) = \max\left(\sup_{\omega \in E_1} \pi(\omega), \sup_{\omega \in E_2} \pi(\omega) \right) \\ &= \max(\Pi(E_1), \Pi(E_2)). \end{aligned} \tag{2.9}$$

The general version of this property is the following.

Maxitivity. For countable collections of measurable sets $E_k \in \Sigma$, the possibility of their union is given by $\Pi\left(\bigcup_{k \geq 1} E_k\right) = \sup_{k \geq 1} \Pi(E_k)$.

Among others, maxitivity entails monotonicity: For $E_1, E_2 \in \Sigma$ with $E_1 \subseteq E_2$, it readily follows that

$$\Pi(E_1) \leq \max(\Pi(E_1), \Pi(E_2 \setminus E_1)) = \Pi(E_1 \cup (E_2 \setminus E_1)) = \Pi(E_2). \tag{2.10}$$

However, the possibility measure is not generally 2-monotone, but rather *2-alternating*: For any two sets $E_1, E_2 \in \Sigma$ (disjunct or not) it holds that $E_1 \cap E_2 \subseteq E_1$ and $E_1 \cap E_2 \subseteq E_2$, and from the monotonicity of Π it follows that $\Pi(E_1 \cap E_2) \leq \min(\Pi(E_1), \Pi(E_2))$ which, in turn, yields

$$\begin{aligned} \Pi(E_1 \cup E_2) + \Pi(E_1 \cap E_2) &\leq \max(\Pi(E_1), \Pi(E_2)) + \min(\Pi(E_1), \Pi(E_2)) \\ &= \Pi(E_1) + \Pi(E_2). \end{aligned} \tag{2.11}$$

Moreover, the possibility measure is not self-dual. Only knowing the possibility $\Pi(E)$ of an event $E \in \Sigma$, does not generally allow one to infer $\Pi(\neg E)$. The identity

$$1 = \Pi(\Omega) = \Pi(E \cup \neg E) = \max(\Pi(E), \Pi(\neg E)) \tag{2.12}$$

implies that either the possibility of E or the possibility of the counter-event $\neg E$ has to be one but nothing more. If $\Pi(E) = 1$, then nothing can be said about $\Pi(\neg E)$. It could lie anywhere between zero and one.

Necessity Measure The *necessity measure* $\mathrm{N} : \Sigma \to [0, 1]$ is dual to the possibility measure in the sense that the necessity of an event $E \in \Sigma$ is defined via the possibility of the complement $\neg E$ as

$$\mathrm{N}(E) = 1 - \Pi(\neg E) = \inf_{\omega \notin E} 1 - \pi(\omega). \tag{2.13}$$

Analogously, the possibility of E may be computed via the necessity of the complement as $\Pi(E) = 1 - \mathrm{N}(\neg E)$. However, the necessity measure does not yield additional information, it merely expresses the same (possibilistic) information differently. In fact, the properties of the necessity measure follow from the properties of the possibility measure.

Let $E \in \Sigma$. The boundedness of the possibility measure entails the positivity of the necessity measure

$$\mathrm{N}(E) = 1 - \underbrace{\Pi(\neg E)}_{\leq 1} \geq 0, \tag{2.14}$$

and the boundedness of the necessity measure follows from the positivity of the possibility measure, i.e.

$$\mathrm{N}(E) = 1 - \underbrace{\Pi(\neg E)}_{\geq 0} \leq 1. \tag{2.15}$$

Similarly, the normality of the possibility measure yields the completeness of the necessity measure

$$\mathrm{N}(\emptyset) = 1 - \underbrace{\Pi(\Omega)}_{=1} = 0, \tag{2.16}$$

and from the completeness of a possibility measure, the normality of the necessity measure

$$\mathrm{N}(\Omega) = 1 - \underbrace{\Pi(\emptyset)}_{=0} = 1, \tag{2.17}$$

is derived. Finally, by considering $\Pi\left(\neg\left(\bigcap_{k\geq 1} E_k\right)\right) = \Pi\left(\bigcup_{k\geq 1} \neg E_k\right) = \sup_{k\geq 1} \Pi(\neg E_k)$, the maxitivity of the possibility measure is expressed equivalently by the following property of the necessity measure.

Minitivity. For countable collections of events $E_k \in \Sigma$, the necessity of their intersection is given by $\mathrm{N}\left(\bigcap_{k\geq 1} E_k\right) = \inf_{k\geq 1} \mathrm{N}(E_k)$.

Contrary to the possibility measure, the necessity measure, is a 2-monotone capacity (and therefore also monotone): For arbitrary $E_1, E_2 \in \Sigma$, it is easy to verify that

$$\begin{aligned} \mathrm{N}(E_1 \cap E_2) + \mathrm{N}(E_1 \cup E_2) &\geq \min(\mathrm{N}(E_1), \mathrm{N}(E_2)) + \max(\mathrm{N}(E_1), \mathrm{N}(E_2)) \\ &= \mathrm{N}(E_1) + \mathrm{N}(E_2) \end{aligned} \tag{2.18}$$

which follows from the monotonicity of N and from $E_1, E_2 \subseteq E_1 \cup E_2$.

Evidently, the necessity measure is not self-dual, but a fundamental observation is that, since either $\Pi(E)$ or $\Pi(\neg E)$ are one, either $\Pi(E)$ must be one or $\mathrm{N}(E)$ must be zero.

Level Sets The elementary possibility function can also be interpreted as the contour function of certain level sets,[15] more precisely its *sublevel sets*

$$\mathcal{S}_\pi^\alpha = \{\omega : \pi(\omega) \leq \alpha\} \tag{2.19}$$

for all $\alpha \in [0, 1]$, and the dual *superlevel sets*

$$\mathcal{C}_\pi^\alpha = \neg \mathcal{S}_\pi^\alpha = \{\omega : \pi(\omega) > \alpha\}. \tag{2.20}$$

Conversely, if only a description of the sublevel sets is available, the elementary possibility of $\omega \in \Omega$ can be reconstructed via a variant of the Decomposition Theorem [Hanss05]

$$\pi(\omega) = \inf_{\alpha \in [0,1] : \omega \in \mathcal{S}_\pi^\alpha} \alpha. \tag{2.21}$$

The sub- and superlevel sets may very well be empty, as in $\mathcal{C}_\pi^1 = \emptyset$, or span all of Ω, as in $\mathcal{S}_\pi^1 = \Omega$. The superlevel set $\mathcal{C}_\pi^0$ is also referred to as the *support* of π in contrast to the hitherto unmentioned *core* which is simply the set of all elementary events with full possibility, i.e.

$$\mathrm{core}(\pi) = \{\omega \in \Omega : \pi(\omega) = 1\}. \tag{2.22}$$

Since the normality of the elementary possibility function merely guarantees a supremum (but not a maximum) of one, the core may very well be empty.

Evidently, the level sets are monotonously in- and decreasing, respectively, in the sense that from $0 \leq \alpha_1 \leq \alpha_2 \leq 1$ it follows that

$$\mathcal{S}_\pi^{\alpha_1} \subseteq \mathcal{S}_\pi^{\alpha_2} \quad \text{and} \quad \mathcal{C}_\pi^{\alpha_2} \subseteq \mathcal{C}_\pi^{\alpha_1}. \tag{2.23}$$

A different way of expressing this is by saying that the superlevel sets are *nested*.

The possibilities of $\mathcal{S}_\pi^\alpha$ and the necessities of $\mathcal{C}_\pi^\alpha$ are bounded by

$$\Pi(\mathcal{S}_\pi^\alpha) = \sup_{\omega \in \mathcal{S}_\pi^\alpha} \pi(\omega) \leq \alpha \qquad \text{and} \qquad \mathrm{N}\left(\mathcal{C}_\pi^{1-\alpha}\right) = 1 - \Pi(\mathcal{S}_\pi^{1-\alpha}) \geq \alpha \tag{2.24}$$

[15] The measurability of the elementary possibility function ensures that its sub- and superlevel sets are also measurable.

for all $\alpha \in [0,1]$, which follows directly from their definitions [CousoMontesGil01]. Below, these bounds shall prove to be of major importance for the imprecise probability interpretation of possibility measures. Equality in Eq. (2.24) can be achieved if there exists some $\omega \in \Omega$ with $\alpha = \pi(\omega)$. Then the possibility of the corresponding sublevel set $\mathcal{S}_\pi^\alpha$ is given by

$$\Pi\left(\mathcal{S}_\pi^\alpha\right) = \Pi\left(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}\right) = \pi(\omega) = \alpha \tag{2.25}$$

because, apart from Eq. (2.24), it holds that $\Pi(\mathcal{S}_\pi^\alpha) = \sup_{\zeta \in \mathcal{S}_\pi^\alpha} \pi(\zeta) \geq \pi(\omega)$.

Sublevel sets are an important concept in possibility theory as they provide a unique view of possibility measures: Consider any event $E^* \in \Sigma$ and its possibility $\alpha^* = \Pi(E^*)$. Obviously, $\pi(\omega) \leq \alpha^*$ obtains for all $\omega \in E^*$. Therefore, E^* must be a subset of $\mathcal{S}_\pi^{\alpha^*}$ which—similar to Eq. (2.25)—has the same possibility because

$$\begin{aligned}\Pi(\mathcal{S}_\pi^{\alpha^*}) &= \Pi\left(\left(\mathcal{S}_\pi^{\alpha^*} \cap E^*\right) \cup \left(\mathcal{S}_\pi^{\alpha^*} \cap \neg E^*\right)\right) \\ &= \max(\underbrace{\Pi\left(\mathcal{S}_\pi^{\alpha^*} \cap E^*\right)}_{=\Pi(E^*)=\alpha^*}, \underbrace{\Pi\left(\mathcal{S}_\pi^{\alpha^*} \cap \neg E^*\right)}_{\leq \alpha^*}) = \alpha^*.\end{aligned} \tag{2.26}$$

The information about E^* provided by Π is, thus, equivalent to the information about $\mathcal{S}_\pi^{\alpha^*}$, i.e. the smallest sublevel set it is contained in.

Similarly, for $\mathrm{n}^* = \mathrm{N}(E^*)$, one can show that E^* contains the superlevel set $\mathcal{C}_\pi^{1-\mathrm{n}^*}$ which has the same necessity $\mathrm{N}(\mathcal{C}_\pi^{1-\mathrm{n}^*}) = \mathrm{n}^*$. The information about E^* by N is equivalent to the information about $\mathcal{C}_\pi^{1-\mathrm{n}^*}$, i.e. the largest superlevel set it contains. Since either $\Pi(E) = 1$ or $\mathrm{N}(E) = 0$, it follows that either $\mathcal{S}_\pi^{\alpha^*} = \Omega$ or $\mathcal{C}_\pi^{1-\mathrm{n}^*} = \emptyset$.

Put differently, a possibilistic analysis can also be interpreted as the analysis of the level sets of π. As the sublevel sets can be computed from the superlevel sets and vice versa, it is, furthermore, sufficient to only analyze one of them—typically the superlevel sets.

To summarize this section, elementary possibility functions, possibility measures, necessity measures, and sub- and superlevel sets are all equivalent descriptions of possibilistic information. Depending on the application, it can make sense to alternate between these different *possibilistic structures* in order to simplify the theoretical considerations, a fact that will be exploited in later chapters. For instance, necessity measures fit into the framework of 2-monotone capacities and coherent lower prevision, and superlevel sets fit into the framework of random sets.

2.1.3 Inclusion

Given two capacities $\mathrm{M}^{(1)}$ and $\mathrm{M}^{(2)}$ on a measurable space (Ω, Σ), the capacity $\mathrm{M}^{(1)}$ is said to be *included* by $\mathrm{M}^{(2)}$ if and only if, for every $E \in \Sigma$, it holds that

$$\mathrm{M}^{(1)}(E) \leq \mathrm{M}^{(2)}(E) \tag{2.27}$$

which shall be denoted by $\mathrm{M}^{(1)} \preceq \mathrm{M}^{(2)}$.

The usefulness of the inclusion (pre-)order[16] $\preceq$ lies in the fact that it allows comparing all sorts of capacities which need not necessarily have the same structure [BronevichKlir10], e.g., maxitive possibility measures, minitive necessity measures and σ-additive probability measures.

For instance, one obtains the fundamental inclusion relation $\mathrm{N} \preceq \Pi$ between a possibility measure Π and its dual necessity measure N by considering that for all $E \in \Sigma$ the inequality $1 = \Pi(\Omega) = \max(\Pi(E), \Pi(\neg E)) \leq \Pi(E) + \Pi(\neg E)$ obtains, and therefore

$$\mathrm{N}(E) = 1 - \Pi(\neg E) \leq \Pi(E). \tag{2.28}$$

It is furthermore evident that the inclusion order does not naturally arise in an exclusively probabilistic framework: Suppose that for two probability measures $\mathrm{P}^{(1)}$ and $\mathrm{P}^{(2)}$ it holds that $\mathrm{P}^{(1)}(E) \leq \mathrm{P}^{(2)}(E)$ for some event $E \in \Sigma$. Then

$$\mathrm{P}^{(1)}(\neg E) = 1 - \mathrm{P}^{(1)}(E) \geq 1 - \mathrm{P}^{(2)}(E) = \mathrm{P}^{(2)}(\neg E) \tag{2.29}$$

which implies that—except in the case of equality $\mathrm{P}^{(1)} = \mathrm{P}^{(2)}$—two probability measures are generally incomparable with respect to $\preceq$.

In the remainder of this thesis, inclusion may also be understood with respect to the information that is encoded in a capacity. That is, all the information encoded in $\mathrm{M}^{(1)}$ is robustly accounted for by $\mathrm{M}^{(2)}$ since the output of the latter, e.g. the possibilities and necessities, always bounds the output of the former.

Based on this order, two key concepts may be derived: Inclusion reduces to the concept of consistency when comparing probability and possibility measures, and it reduces to the concept of specificity when comparing two possibility measures.

2.1.3.1 Consistency

The fundamental concept of probability-possibility *consistency* was first introduced by Delgado and Moral [DelgadoMoral87] and by Dubois and Prade [DuboisPrade92] in the form discussed here, although earlier attempts at linking probability theory and possibility theory were already made, e.g., by Zadeh [Zadeh99].

Let P be a probability on a measurable space (Ω, Σ), and let π be an elementary possibility function thereon. If P is included by the corresponding possibility measure Π, i.e. $\mathrm{P} \preceq \Pi$, then they are said to be *consistent*, which is also written as $\mathrm{P} \preceq \pi$. Naturally, if $\mathrm{P} \preceq \pi$, then the corresponding necessity N is also included by P: To see this, consider any event $E \in \Sigma$. Rearranging $1 - \mathrm{P}(E) = \mathrm{P}(\neg E) \leq \Pi(\neg E)$ immediately yields $\mathrm{N}(E) = 1 - \Pi(\neg E) \leq \mathrm{P}(E)$.

[16] The definition of a preorder requires that the relation $\preceq$ be reflexive and transitive. This follows directly from the respective properties of the total order $\leq$ on $[0, 1]$.

Example 1: Consistency

Consider the possibility measure Π defined via the elementary possibilities

$$\pi(\omega_1) = \frac{1}{2} \qquad \text{and} \qquad \pi(\omega_2) = 1$$

on the universe of discourse $\Omega = \{\omega_1, \omega_2\}$. The consistency $\mathrm{P}_\theta \preceq \pi$ of the probability measures P_θ given by

$$\mathrm{P}_\theta(\{\omega_1\}) = \theta \qquad \text{and} \qquad \mathrm{P}_\theta(\{\omega_2\}) = 1 - \theta$$

for $\theta \in [0, 1]$ follows by considering

$$\begin{array}{lllllll} \mathrm{N}(\emptyset) & = 0 & \leq & \mathrm{P}_\theta(\emptyset) & = 0 & \leq & \Pi(\emptyset) & = 0, \\ \mathrm{N}(\{\omega_1\}) & = 0 & \leq & \mathrm{P}_\theta(\{\omega_1\}) & = \theta & \leq & \Pi(\{\omega_1\}) & = \frac{1}{2}, \\ \mathrm{N}(\{\omega_2\}) & = \frac{1}{2} & \leq & \mathrm{P}_\theta(\{\omega_2\}) & = 1 - \theta & \leq & \Pi(\{\omega_2\}) & = 1, \quad \text{and} \\ \mathrm{N}(\Omega) & = 1 & \leq & \mathrm{P}_\theta(\Omega) & = 1 & \leq & \Pi(\Omega) & = 1. \end{array}$$

That is, $\mathrm{P}_\theta \preceq \pi$ if and only if $\theta \in [0, \frac{1}{2}]$.

In conclusion, the concept of consistency implies that possibility serves as an *upper probability*—and necessity as a *lower probability*. For every consistent probability measure, the respective probability of any event is bounded from above by its possibility and from below by its necessity. The elementary possibility function itself provides an upper bound on the elementary probabilities $\mathrm{P}(\{\omega\}) \leq \pi(\omega)$ for $\omega \in \Omega$.

Under the general inclusion formulation in Eq. (2.27), consistency is not trivial to verify, for, in principle, it requires one to assert $\mathrm{P}(E) \leq \Pi(E)$ for all events $E \in \Sigma$ which is a tedious task—especially in the case of infinite sample spaces. Recalling the arguments in Section 2.1.2.3, viewing a possibility measure as a measure of general events is not essential, and one generally does not have any more information about a set than, e.g., about the largest sublevel set it is contained in. Following this line of thought, consistency is expressible and—more importantly—verifiable by considering only the level sets. The following well-known lemma [DuboisEtAl04, DuboisPrade92, CousoMontesGil01] provides such a necessary and sufficient criterion.

Lemma 1. *The probability measure* P *and the elementary possibility function* π *are consistent if and only if*

$$\mathrm{P}(\mathcal{S}_\pi^\alpha) \leq \alpha \tag{2.30}$$

or equivalently if and only if

$$\mathrm{P}(\mathcal{C}_\pi^\alpha) \geq 1 - \alpha \tag{2.31}$$

for all $\alpha \in [0, 1]$.

Proof. For the "⇒"-direction, suppose that $\mathrm{P} \preceq \pi$ and let $\alpha \in [0,1]$. As discussed in Eq. (2.24), it holds that $\Pi(\mathcal{S}_\pi^\alpha) \leq \alpha$, and hence $\mathrm{P}(\mathcal{S}_\pi^\alpha) \leq \Pi(\mathcal{S}_\pi^\alpha) \leq \alpha$.
For the "⇐"-direction, suppose that Eq. (2.30) obtains. Moreover, let $E \in \Sigma$ and define $\alpha^* = \Pi(E)$. From the deliberations in Section 2.1.2.3, it follows that $E \subseteq \mathcal{S}_\pi^{\alpha^*}$. Utilizing the monotonicity of a probability measure, one obtains the inequality $\mathrm{P}(E) \leq \mathrm{P}\left(\mathcal{S}_\pi^{\alpha^*}\right) \leq \alpha^* = \Pi(E)$ proving the first part of the lemma.
The equivalence of Eq. (2.30) and Eq. (2.31) follows by considering that for all $\alpha \in [0,1]$ it holds that $\mathrm{P}\left(\mathcal{C}_\pi^{\alpha^*}\right) = \mathrm{P}\left(\neg\mathcal{S}_\pi^{\alpha^*}\right) = 1 - \mathrm{P}\left(\mathcal{S}_\pi^{\alpha^*}\right) \geq 1 - \alpha$. □

Many of the proofs in this thesis rely on this fundamental consistency criterion—often without explicit mention.

The expression of consistency on the basis of level sets is instructive and considerably simplifies any possibilistic analysis. Instead of checking consistency for all possible events individually, it suffices to check consistency for only the sublevel sets or the superlevel sets. For instance, applying Lemma 1 to Example 1 yields only one non-trivial condition, $\mathrm{P}_\theta(\{\omega_1\}) = \theta \leq \frac{1}{2}$. This supports the claims made in Section 2.1.2.3 that a possibilistic analysis is really only concerned with the analysis of the level sets of the elementary possibility function. Moreover, Lemma 1 points to a well-known interpretation of superlevel sets. In particular, the expression in Eq. (2.31) is a special case of a *prediction set* [Shao03], a well-known concept from statistics. For instance, the event $\mathcal{C}_\pi^0$ is predicted to happen (P-)almost surely since $\mathrm{P}(\mathcal{C}_\pi^0) = 1$ for all $\mathrm{P} \preceq \pi$, and the event $\mathcal{C}_\pi^{0.01}$ is predicted to occur with a probability greater than 99%. Conversely, $\mathcal{S}_\pi^{0.05}$ is guaranteed to occur with a probability of less than 5%, see also Section 4.3.3.

The following auxilliary lemma indicates that not all $\alpha \in [0,1]$ need to be checked in Lemma 1. It merely suffices to check Eq. (2.30) element-wise, i.e. for all $\alpha \in \{\pi(\omega) : \omega \in \Omega\}$.

Lemma 2. *The probability measure* P *and the elementary possibility function* π *are consistent if and only if*

$$\mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}) \leq \pi(\omega) \tag{2.32}$$

for all $\omega \in \Omega$.

Proof. Equivalence is shown by showing that Eq. (2.32) is equivalent to Eq. (2.30).
To show that Eq. (2.30) implies Eq. (2.32), let P be a probability measure on (Ω, Σ) that satisfies Eq. (2.30), let $\omega \in \Omega$ and define $\alpha = \pi(\omega)$. Then, the equality $\mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}) = \mathrm{P}(\mathcal{S}_\pi^\alpha) \leq \alpha = \pi(\omega)$ holds.
The reverse is shown by contradiction: Let P be a probability measure on (Ω, Σ). Furthermore, assume that Eq. (2.32) does not hold, i.e. that there exists an $\omega \in \Omega$ such that $\mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}) > \pi(\omega)$, and define $\alpha = \pi(\omega)$. In this case, the inequality $\alpha = \pi(\omega) < \mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}) = \mathrm{P}(\mathcal{S}_\pi^\alpha)$ holds, i.e. Eq. (2.30) fails, too. □

2.1.3.2 Specificity

The fundamental concept of *specificity* was first introduced by Dubois and Prade [DuboisPrade86a] and arises when comparing two possibility measures by the inclusion order.

Consider two elementary possibility functions $\pi^{(1)}$ and $\pi^{(2)}$ on the measurable space (Σ, Ω) with the corresponding possibility measures $\Pi^{(1)}$ and $\Pi^{(2)}$. If $\Pi^{(1)}$ is included by $\Pi^{(2)}$, then $\pi^{(1)}$ is said to be more *specific* than $\pi^{(2)}$, which is written as $\pi^{(1)} \preceq \pi^{(2)}$.

Example 2: Specificity

Consider the three elementary possibility functions $\pi^{(1)}$, $\pi^{(2)}$ and $\pi^{(3)}$ on the measurable space (Ω, Σ), consisting of the universe of discourse $\Omega = \{\omega_1, \omega_2, \omega_3\}$ and its powerset $\Sigma = 2^{\Omega}$. Their respective elementary possibilities are given by

$\omega \in \Omega$	$\pi^{(1)}(\omega)$	$\pi^{(2)}(\omega)$	$\pi^{(3)}(\omega)$
ω_1	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$
ω_2	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{3}{4}$
ω_3	1	1	1

By considering the corresponding possibilities

$E \in \Sigma = 2^{\Omega}$	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_3\}$	$\{\omega_2, \omega_3\}$	Ω
$\Pi^{(1)}(E)$	0	$\frac{1}{3}$	$\frac{1}{4}$	1	$\frac{1}{3}$	1	1	1
$\Pi^{(2)}(E)$	0	$\frac{1}{2}$	$\frac{1}{5}$	1	$\frac{1}{2}$	1	1	1
$\Pi^{(3)}(E)$	0	$\frac{2}{3}$	$\frac{3}{4}$	1	$\frac{3}{4}$	1	1	1

it is inferred that both $\pi^{(1)}$ and $\pi^{(2)}$ are more specific than $\pi^{(3)}$. However, $\Pi^{(1)}$ and $\Pi^{(2)}$ are incomparable with respect to $\preceq$, and, thus, neither $\pi^{(1)}$ nor $\pi^{(2)}$ is more specific than the other.

Naturally, this is also equivalent to the corresponding inclusion relation of the respective necessity measures, i.e., $\mathrm{N}^{(2)} \preceq \mathrm{N}^{(1)}$ follows from $\pi^{(1)} \preceq \pi^{(2)}$.

Similar to consistency, the general definition of specificity is tedious to check in most cases but can be replaced by the following lemma giving a necessary and sufficient criterion in terms of the elementary possibility function.[17]

Lemma 3. *The elementary possibility function $\pi^{(1)}$ is more specific than the elementary possibility function $\pi^{(2)}$ if and only if the former is elementwise lower than the latter, i.e.*

[17] This concept corresponds to fuzzy set inclusion.

if

$$\pi^{(1)}(\omega) \leq \pi^{(2)}(\omega) \tag{2.33}$$

for all $\omega \in \Omega$.

Proof. For the "⇒"-direction, suppose that $\pi^{(1)}$ is more specific than $\pi^{(2)}$ and let $\omega \in \Omega$. Then, it holds that $\pi^{(1)}(\omega) = \Pi^{(1)}(\{\omega\}) \leq \Pi^{(2)}(\{\omega\}) = \pi^{(2)}(\omega)$.
For the "⇐"-direction, suppose that $\pi^{(1)}(\omega) \leq \pi^{(2)}(\omega)$ for all $\omega \in \Omega$ and let $E \in \Sigma$. Then, it holds that $\Pi^{(1)}(E) = \sup_{\omega \in E} \pi^{(1)}(\omega) \leq \sup_{\omega \in E} \pi^{(2)}(\omega) = \Pi^{(2)}(E)$. □

This lemma, e.g., considerably simplifies the considerations in Example 2 as it suffices to compare only the elementary possibilities, which leads to the same result.

Similarly to Lemma 1, a necessary and sufficient condition for the inclusion order in terms of the level sets may also be given by the following lemma.

Lemma 4. *The elementary possibility function* $\pi^{(1)}$ *is more specific than the elementary possibility function* $\pi^{(2)}$ *if and only if*

$$\mathcal{S}^{\alpha}_{\pi^{(1)}} \supseteq \mathcal{S}^{\alpha}_{\pi^{(2)}}, \tag{2.34}$$

or, equivalently, if and only if

$$\mathcal{C}^{\alpha}_{\pi^{(1)}} \subseteq \mathcal{C}^{\alpha}_{\pi^{(2)}} \tag{2.35}$$

for all $\alpha \in [0, 1]$.

Proof. For the first criterion, it suffices to show the equivalence of Eqs. (2.33) and (2.34) as the latter is a necessary and sufficient condition for $\pi^{(1)}$ to be more specific than $\pi^{(2)}$ in Lemma 3.
For the "⇒"-direction, suppose that Eq. (2.33) obtains. Let $\alpha \in [0, 1]$ and let $\omega \in \mathcal{S}^{\alpha}_{\pi^{(2)}}$. Then $\pi^{(1)}(\omega) \leq \pi^{(2)}(\omega) \leq \alpha$, and therefore $\omega \in \mathcal{S}^{\alpha}_{\pi^{(1)}}$, i.e., Eq. (2.34) follows from Eq. (2.33).
For the "⇐"-direction, suppose that Eq. (2.34) obtains, let $\omega \in \Omega$ and define $\alpha^* = \pi^{(2)}(\omega)$. Then $\omega \in \mathcal{S}^{\alpha^*}_{\pi^{(2)}}$ and therefore also $\omega \in \mathcal{S}^{\alpha^*}_{\pi^{(1)}}$, i.e. $\pi^{(1)}(\omega) \leq \alpha^* = \pi^{(2)}(\omega)$, i.e., Eq. (2.33) follows from Eq. (2.34) and the first part of the proposition is proven.
The second part, i.e. the equivalence to Eq. (2.35), follows by a similar argument. □

2.2 Qualitative Possibility Theory

Pyt'ev suggests a qualitative view of possibility measures [Pyt'ev97], focusing only on ranking the possibilities of individual events. According to his Principle of Relativity, two possibility measures are equivalent if there exists a strictly increasing lower semi-continuous rescaling function between them. Consequently, the precise numerical values of

possibilities are irrelevant, only the relative degrees matter to make assertions [Zubyuk19] such as 'event $E_1 \in \Sigma$ is more possible than event $E_2 \in \Sigma$'.

On their own, the elementary possibilities $\pi(\omega)$ present such a qualitative assessment of the elementary outcomes $\omega \in \Omega$ by ranking them with respect to their *plausibility*, which is, indeed, intended to resemble the colloquial connotation of the word. It indicates how little one should be surprised [Neumaier03] by a certain outcome $\omega \in \Omega$—Dubois and Prade [DuboisPrade20] call this the *ordinal preference*—, and is closely connected to fuzzy set theory as a way to represent linguistic imprecision [Zadeh95].

More precisely, the elementary possibilities establish a *plausibility (pre-)order*[18] among the elements of Ω. The outcome $\omega_1 \in \Omega$ is said to be *less plausible* than $\omega_2 \in \Omega$ under π if and only if $\pi(\omega_1) \leq \pi(\omega_2)$, which is written as $\omega_1 \sqsubseteq_\pi \omega_2$.

A plausibility-order may be elicited from agents (so-called *experts*), e.g., by translating their assent or dissent to statements of the form 'ω_a is less plausible than ω_b', and can, subsequently, be used to construct elementary possibility functions.

Example 3: Plausibility Order

Three gladiators, Antonius, Brutus, and Cassius, are to compete in a three-way contest. Only one of them can win. Before the fight, the emperor is asked how plausible he thinks the three outcomes ω_A : 'Antonius wins.', ω_B : 'Brutus wins.', and ω_C : 'Cassius wins.' are. His subjective ranking of their plausibility of winning is given through the order

$$\omega_\mathrm{C} \sqsubseteq \omega_\mathrm{A} \sqsubseteq \omega_\mathrm{B},$$

i.e., he finds it most plausible that Brutus wins and least plausible that Cassius wins, which is e.g. reflected by the elementary possibilities

$$\pi^{(1)}(\omega_\mathrm{A}) = \frac{2}{3}, \qquad \pi^{(1)}(\omega_\mathrm{B}) = 1 \qquad \text{and} \qquad \pi^{(1)}(\omega_\mathrm{C}) = \frac{1}{3}$$

but also by

$$\pi^{(2)}(\omega_\mathrm{A}) = \frac{1}{2}, \qquad \pi^{(2)}(\omega_\mathrm{B}) = 1 \qquad \text{and} \qquad \pi^{(2)}(\omega_\mathrm{C}) = \frac{1}{2}.$$

In this thesis, a second elementary possibility function π' is said to be *plausibility-conform* to π if $\omega_1 \sqsubseteq_\pi \omega_2$ implies $\omega_1 \sqsubseteq_{\pi'} \omega_2$ for all $\omega_1, \omega_2 \in \Omega$. Notice that this property is not necessarily symmetric,[19] i.e., it does not imply that π is plausibility-conform to π'.

Finally, the N elementary possibility functions $\pi^{(1)}, \ldots, \pi^{(N)}$ are called *comonotone* [Bronevich01] if every one of them is plausibility-conform to their element-wise

[18] The definition of a preorder requires that the plausibility preorder be reflexive and transitive. This follows directly from the respective properties of the total order $\leq$ on $[0, 1]$.

[19] In Example 3, $\pi^{(2)}$ is plausibility-conform to $\pi^{(1)}$ but not vice versa.

minimum $\pi^{\min.}$ given by

$$\pi^{\min.}(\omega) = \min_{i=1,\ldots,N} \pi^{(i)}(\omega). \tag{2.36}$$

for all $\omega \in \Omega$. Comonotonicity may be interpreted as being able to find a common plausibility order given by $\pi^{\min.}$ to which none of the agents having stated the individual plausibility orders $\pi^{(i)}$ would be opposed.

Remark 5. Under the assumption of comonotonicity, the function $\pi^{\min.}$ must be an elementary possibility function. Its measurability follows from being the minimum of N measurable functions, and the normality obtains by the following consideration for $N = 2$. Let $\alpha = \sup_{\omega\in\Omega} \min\left(\pi^{(1)}(\omega), \pi^{(2)}(\omega)\right)$ be the supremum of both elementary possibility functions, which, for the sake of the counter-argument, is assumed to be subnormal, i.e. $\alpha < 1$. Then there exist $\omega_1, \omega_2 \in \Omega$ with $\pi^{(1)}(\omega_1) > \alpha$ but $\pi^{(2)}(\omega_1) \leq \alpha$, as well as $\pi^{(1)}(\omega_1) \leq \alpha$ but $\pi^{(2)}(\omega_2) > \alpha$. However, if $\pi^{(2)}(\omega_1) \leq \pi^{(1)}(\omega_2)$, then $\pi^{\min.}(\omega_1) \leq \pi^{\min.}(\omega_2)$ and $\pi^{(1)}$ is not plausibility-conform to $\pi^{\min.}$. Conversely, if $\pi^{(2)}(\omega_1) \leq \pi^{(1)}(\omega_2)$, then $\pi^{(2)}$ is not plausibility-conform to $\pi^{\min.}$. In conclusion, only $\alpha = 1$ is consistent with the assumption of comonotonicity, i.e., $\pi^{\min.}$ must be normal. A similar argument can be found for the minimum of N elementary possibility functions.

A deeper discussion of how to construct elementary possibility functions based on plausibility orders stated by agents, e.g. in the case of conflicting information[20] shall not be pursued, here, as it has extensively been treated already [DuboisPrade98, Chang81, ChameauSantamarina87].

The plausibility order implied by an elementary possibility function does, by itself, not yield any quantifiable, only qualitative, information. It is, therefore, of limited use to describe, e.g., uncertainty. Still, considering qualitative possibilities is not without its merits and will be revisited in the following exposition of quantitative possibility theory, e.g. when discussing the fundamental Principle of Plausibility.

2.3 Quantitative Possibility Theory

So far, probability measures and possibility measures have been introduced independently, and the concept of inclusion—more precisely, that of consistency—has been suggested for their comparison. Evidently, both capacities share similar properties, but especially the difference between σ-additivity and maxitivity results in considerable differences.

[20]Suppose an agent states an inconsistent plausibility order $\omega_a \sqsubseteq_\rho \omega_b$, $\omega_b \sqsubseteq_\rho \omega_c$ and $\omega_c \sqsubseteq_\rho \omega_a$. How should this conflict be resolved?

2.3.1 From Possibilities to Imprecise Probabilities

A straight-forward way of linking possibilities and probabilities is available via the concept of *imprecise probabilities* (IP), i.e. when considering families of probabilities $\mathfrak{P} = \{\mathrm{P}_1, \mathrm{P}_2, \ldots\}$ on the measurable space (Ω, Σ), which possibility theory offers a convenient way of describing. This allows for a quantitative interpretation of possibility measures—as opposed to the plausibility order-based qualitative interpretation of the elementary possibility function described above.

2.3.1.1 Credal Sets

The concept of probability-possibility consistency is the paradigm guiding the following development of possibility theory as a framework for reasoning with imprecise probabilities, and great emphasis is put on viewing possibilities as upper probabilities and on regarding possibilistic information as an expression of (un-)certainty about a probability measure.

By this line of reasoning, it is somewhat misleading to present possibility theory as an alternative to probability theory. For all intents and purposes, it is a mere extension based on the rejection of the assumption that one is always able to specify precise probabilities. It is still very much based and connected to probability theory, though, and it would be deprived of its meaning without probability.

As demonstrated in Example 1, consistency is not a one-to-one relation, and, generally, one can find infinitely many probability measures that are consistent with the elementary possibility function π on (Ω, Σ). The *credal set*

$$\mathfrak{C}(\pi) = \{\mathrm{P} \in \mathbb{P}(\Omega, \Sigma) : \mathrm{P} \preceq \pi\} \tag{2.37}$$

gathers all such consistent probability measures.

By definition [AntonucciCuzzolin10], credal sets must be convex. This is verified by considering two consistent probability measures $\mathrm{P}^{(1)}, \mathrm{P}^{(2)} \preceq \pi$ and a scalar $s \in [0, 1]$. The linearly combined probability measure $\mathrm{P}(\cdot) = s\mathrm{P}^{(1)}(\cdot) + (1 - s)\mathrm{P}^{(2)}(\cdot)$ is also consistent with π because for all $\alpha \in [0, 1]$

$$\mathrm{P}(\mathcal{S}_\pi^\alpha) = s\mathrm{P}^{(1)}(\mathcal{S}_\pi^\alpha) + (1 - s)\mathrm{P}^{(2)}(\mathcal{S}_\pi^\alpha) \leq s\alpha + (1 - s)\alpha = \alpha. \tag{2.38}$$

Finally, the concept of specificity, allows for a basic comparison of credal sets. By the transitivity of the inclusion preorder, it follows that $\mathfrak{C}(\pi^{(1)}) \subseteq \mathfrak{C}(\pi^{(2)})$ if $\pi^{(1)} \preceq \pi^{(2)}$. Then, it makes sense to write, e.g., $\mathrm{P} \preceq \pi^{(1)} \preceq \pi^{(2)}$ in order to show that P is consistent with $\pi^{(2)}$ because $\mathrm{P} \in \mathfrak{C}(\pi^{(1)}) \subseteq \mathfrak{C}(\pi^{(2)})$. In conclusion, specificity may be employed to measure the expressiveness of a possibility measure, or the 'size' of its corresponding credal set. This can also be used as a basis for measures of the information content of a possibility measure, for which many metrics have been proposed [KlirWierman99].

A detailed investigation of the credal sets of possibility measures is provided by Baudrit et al. [BaudritDubois06].

2.3.1.2 Possibilities as Upper Probabilities

Due to the very specific geometry of the credal sets of possibility measures, possibility theory provides a rather coarse framework for reasoning with imprecise probabilities. Possibilities are not able to exactly represent arbitrary families of probability measures but only those following a very specific structure as explained above. Most famously, a single probability measure can usually not precisely be represented—except for the trivial class of 'deterministic' measures to be discussed in Section 3.1.3. This coarseness is perhaps the biggest drawback of possibility theory. Possibilities simply provide upper (and lower) bounds of probabilities, making it a framework of ill-known or imprecise probabilities.

Two important properties of possibilistic descriptions of imprecise probabilities are listed below.

Avoiding Sure Loss The credal set of a possibility measure is never empty and one can always find at least one probability measure that is consistent with π. For instance, for every ω_c in the core of π, it is easy to verify that the probability measure P^{ω_c} given by

$$\mathrm{P}^{\omega_c}(E) = \begin{cases} 1 & \text{if } \omega_c \in E \\ 0 & \text{otherwise} \end{cases} \tag{2.39}$$

for all $E \in \Sigma$ is always contained in $\mathfrak{C}(\pi)$. In Example 1, this corresponds to $\omega_c = \omega_2$ and $\theta = 0$. Of course, the existence of such a ω_c is not always guaranteed because the core could be empty.

If no such core element exists, it is—due to the normality of the elementary possibility function—always possible to find a sequence $(\omega_i)_{i=1}^{\infty}$ in Ω such that $\pi(\omega_i) \to 1$ for $i \to \infty$. The probability mass function $p^{\in}$ with

$$p^{\in}(\omega_i^{\in}) = \Pi(\{\omega_1^{\in}, \ldots, \omega_i^{\in}\}) - \Pi(\{\omega_1^{\in}, \ldots, \omega_{i-1}^{\in}\}) \tag{2.40}$$

for $i = 1, 2, \ldots$ and $p^{\in}(\omega) = 0$ for all other $\omega \in \Omega$ induces a probability measure $\mathrm{P}^{\in}$ that certainly belongs to $\mathfrak{C}(\pi)$. This may be shown as follows. Let $\alpha \in [0, 1]$. The probability on the corresponding sublevel set depends on the probability masses of the $\omega_i^{\in} \in \mathcal{S}_\pi^\alpha$. Additionally, one may disregard the elements with probability mass zero, i.e., one may express the probability on $\mathcal{S}_\pi^\alpha$ as

$$\mathrm{P}^{\in}(\mathcal{S}_\pi^\alpha) = \sum_{\omega \in \Omega : \omega \in \mathcal{S}_\pi^\alpha} p(\{\omega_i^{\in}\}) = \sum_{i : \omega_i^{\in} \in \mathcal{S}_\pi^\alpha} p(\{\omega_i^{\in}\}) = \sum_{i : \omega_i^{\in} \in \mathcal{S}_\pi^\alpha \wedge p(\omega_i^{\in}) > 0} p(\{\omega_i^{\in}\}). \tag{2.41}$$

However, $0 < p(\omega_i^\in) = \Pi(\{\omega_1^\in, \dots, \omega_i^\in\}) - \Pi(\{\omega_1^\in, \dots, \omega_{i-1}^\in\})$ implies that

$$\begin{aligned}\Pi(\{\omega_1^\in, \dots, \omega_{i-1}^\in\}) < \Pi(\{\omega_1^\in, \dots, \omega_i^\in\}) &= \max\left(\Pi(\{\omega_1^\in, \dots, \omega_{i-1}^\in\}), \pi(\omega_i^\in)\right) \\ &= \pi(\omega_i^\in) \leq \alpha,\end{aligned} \tag{2.42}$$

i.e. $\omega_1^\in, \dots, \omega_i^\in \in \mathcal{S}_\pi^\alpha$. Then, the probability on $\mathcal{S}_\pi^\alpha$ can be upper bounded via the supremum of the (telescoping) sums over all these elements

$$\begin{aligned}\mathrm{P}^\in\left(\mathcal{S}_\pi^\alpha\right) &\leq \sup_{i\,:\,\omega_i^\in \in \mathcal{S}_\pi^\alpha \wedge p(\omega_i^\in)>0} \sum_{j=1}^{i} p(\{\omega_j^\in\}) \\ &= \sup_{i\,:\,\omega_i^\in \in \mathcal{S}_\pi^\alpha \wedge p(\omega_i^\in)>0} \Pi\left(\{\omega_1^\in, \dots, \omega_i^\in\}\right) - \Pi(\emptyset) \\ &= \sup_{i\,:\,\omega_i^\in \in \mathcal{S}_\pi^\alpha \wedge p(\omega_i^\in)>0} \pi(\omega_i^\in) - 0 \leq \alpha.\end{aligned} \tag{2.43}$$

Conversely, sub-normality implies $\Pi(\Omega) < 1$ and there cannot be any $\mathrm{P} \preceq \pi$ because, by definition, $\mathrm{P}(\Omega) = 1$ for all $\mathrm{P} \in \mathbb{P}(\Omega, \Sigma)$.
Alternatively, every probability measure $\mathrm{P}^\in$ constructed via

$$\mathrm{P}^\in(E) = \int_0^1 \frac{\mathrm{P}^0(\mathcal{C}_\pi^\beta \cap E)}{\mathrm{P}^0(\mathcal{C}_\pi^\beta)}\,\mathrm{d}\beta, \tag{2.44}$$

for all $E \in \Sigma$ is consistent with π. Therein, $\mathrm{P}^0 \in \mathbb{P}(\Omega, \Sigma)$ can be an arbitrary probability measure[21] on (Ω, Σ) that fulfills the (easily achievable) regularity assumption of non-zero probability on (μ-)almost all α-cuts. Under this definition, one can verify that $\mathrm{P}^\in$ is a consistent probability measure[22] by checking that

$$\mathrm{P}^\in(\mathcal{C}_\pi^\alpha) = \int_0^1 \frac{\mathrm{P}^0(\mathcal{C}_\pi^\alpha \cap \mathcal{C}_\pi^\beta)}{\mathrm{P}^0(\mathcal{C}^\beta)}\,\mathrm{d}\beta = \underbrace{\int_0^\alpha \frac{\mathrm{P}^0(\mathcal{C}_\pi^\alpha)}{\mathrm{P}^0(\mathcal{C}_\pi^\beta)}\,\mathrm{d}\beta}_{\geq 0} + \underbrace{\int_\alpha^1 \frac{\mathrm{P}^0(\mathcal{C}_\pi^\beta)}{\mathrm{P}^0(\mathcal{C}_\pi^\beta)}\,\mathrm{d}\beta}_{=1-\alpha} \geq 1 - \alpha \tag{2.45}$$

obtains for all $\alpha \in [0, 1]$.
In imprecise probabilistic terms, the non-emptiness of the credal set corresponds to *avoiding sure loss* [Walley91], and if $\mathrm{P}^0 = \lambda$ is a (normalized) Lebesgue measure, then this construction corresponds to the *Pignistic Transform* [Smets05], also called the *Shapley value* [Shapley53]. Finally, Eq. (2.44) describes a very general technique for *Possibility-to-Probability Transformations*. The converse (Imprecise-)Probability-to-Possibility Transformation shall be discussed in Section 2.3.2.2.

Coherence Fundamentally, every possibility measure is also a *coherent* upper probability [De CoomanAeyels99]. That is, for all events $E \in \Sigma$ it provides a tight upper bound

[21] In fact, it suffices if P^0 is a bounded, σ-additive measure that need not necessarily be normal.
[22] It follows from the positivity, normality and σ-additivity of P^0 (and from the linearity of the integral) that $\mathrm{P}^\in$ also possesses these properties and, too, is a probability measure.

on the probability $\mathrm{P}(E)$ via

$$\Pi(E) = \sup_{\mathrm{P} \preceq \pi} \mathrm{P}(E). \tag{2.46}$$

From the consistency criterion, it is clear that $\mathrm{P}(E) \leq \Pi(E)$ if $\mathrm{P} \preceq \pi$ but it is not trivial to see that, for all $E \in \Sigma$, a consistent probability measure $\mathrm{P}^* \preceq \pi$ can be found such that $\mathrm{P}^*(E) = \Pi(E)$.

Again, the arguments by Fetz and Oberguggenberger [FetzOberguggenberger04] are adapted. If $\Pi(E) = 1$, then, one can simply construct a sequence of elementary events in E with an appropriate probability mass function—similar to Eq. (2.40). Therefore, only the case $\Pi(E) < 1$ is left to be considered. Then, one may find a sequence of elementary events $(\omega_i^{\in})_{i \geq 1}$ in E such that $\pi(\omega_i^{\in}) \to \Pi(E)$ for $i \to \infty$, and a second sequence of elementary events $(\omega_j^{\notin})_{j \geq 1}$ in $\neg E$ such that $\pi(\omega_j^{\notin}) \to \Pi(\neg E) = 1$ for $j \to \infty$. The probability mass function p^* with

$$\begin{aligned} p^*(\omega_i^{\in}) &= \Pi(\{\omega_1^{\in}, \ldots, \omega_i^{\in}\}) - \Pi(\{\omega_1^{\in}, \ldots, \omega_{i-1}^{\in}\}) \qquad \text{and} \\ p^*(\omega_j^{\notin}) &= \Pi(E \cup \{\omega_1^{\notin}, \ldots, \omega_j^{\notin}\}) - \Pi(E \cup \{\omega_1^{\notin}, \ldots, \omega_{j-1}^{\notin}\}) \end{aligned} \tag{2.47}$$

for $i, j = 1, 2, \ldots$ and zero otherwise defines a consistent probability measure P^* achieving $\mathrm{P}^*(E) = \Pi(E)$. The consistency is shown similar to Eq. (2.43), and the equivalence follows from the telescoping sum

$$\mathrm{P}^*(E) = \sum_{i=1}^{\infty} p^*(\omega_i^{\in}) = \Pi(\{\omega_1^{\in}, \omega_2^{\in}, \ldots\}) - \Pi(\emptyset) = \Pi(E). \tag{2.48}$$

Alternatively, the fact that possibility measures are 2-alternating as pointed out in Eq. (2.11) suffices to show that they are special cases of coherent upper probabilities that avoid sure loss [BronevichRozenberg20]. Of course, the 2-monotone necessity is also a coherent lower probability, which avoids sure loss, i.e.

$$\mathrm{N}(E) = \inf_{\mathrm{P} \preceq \pi} \mathrm{P}(E). \tag{2.49}$$

Formally, coherence also corresponds to credal sets being closed with respect to the metric $d_{(\Omega,\Sigma)}$ given by $d_{(\Omega,\Sigma)}(\mathrm{P}^{(1)}, \mathrm{P}^{(2)}) = \sup_{E \in \Sigma} |\mathrm{P}^{(1)}(E) - \mathrm{P}^{(2)}(E)|$ for $\mathrm{P}^{(1)}, \mathrm{P}^{(2)} \in \mathbb{P}(\Omega, \Sigma)$.

Relation to Other Theories of Imprecise Probabilities A possibilistic description of an experiment is coarser than any precise probabilistic description. But, it is equally valid and arguably sometimes more appropriate. Moreover, this coarseness is compensated for by very efficient numerical implementations, which are only rivaled by interval analysis which provides even coarser approximations.[23] Concerning the connection of possibility theory to other prominent theories of imprecise probabilities, the following remarks can be made.

[23] In fact, many implementations from interval analysis can be extended to possibilistic analysis, see Chapter 5.

- The analysis of quasi-vacuous possibility measures, which are exclusively $\{0, 1\}$-valued, coincides with *interval analysis* [JaulinEtAl01].
- In order to arrive at the upper and lower probabilities induced by *random sets* as proposed by Dempster [Dempster67], consider the multivalued mapping $\alpha \mapsto \mathcal{C}_\pi^\alpha$ for the random variable α with standard uniform probability distribution.[24]
- The necessity measure is an example of Shafer's *consonant belief functions*, and conversely the possibility measure is a *consonant plausibility function* [Shafer76, DuboisPrade90].
- In the subjective framework of lower previsions [Walley91], the necessity $\mathrm{N}(E)$ of some event $E \in \Sigma$ is highest acceptable buying price for bet that pays 1 Euro if E occurs, and conversely the possibility $\Pi(E)$ is the lowest acceptable selling price if one had to offer this payout to someone else.

Of course, there exist many more links to other theories [DesterckeDuboisChojnacki08, AugustinEtAl14, Dubois82, Walley91].

2.3.2 From Imprecise Probabilities to Possibilities

Having demonstrated how credal sets naturally arise from the inclusion order, it remains to be addressed how an elementary possibility function can be chosen such that it describes a given set of probability measures, i.e. imprecise probabilities. This is essential because possibilistic models or descriptions do not arise naturally; instead, they must be inferred. For instance, one might say that 'the probability that this die shows a one is somewhere between 16% and 17%', or that 'the probability of rain in Stuttgart tomorrow is less than 5%', but these expressions hardly ever conform to a unequivocal possibilistic structure.

It is, unfortunately, not possible to express arbitrary sets of probability measures as the credal set of a single elementary possibility function. As an example, consider a 'fair' coin. Precisely expressing the corresponding probability measure with equal probabilities for every side cannot be achieved by a possibility measure as it would require, e.g., the upper elementary probabilities of both 'heads' and 'tails', i.e. the elementary possibilities, to be exactly $\frac{1}{2}$. But this would violate the normality requirement of the elementary possibility function. For this reason, one might ask whether this renders possibility theory useless. The remainder of this thesis is intended to convince skeptics of the opposite, but it must be conceded that possibility theory does not replace classical probability theory. If a precise, and well-warranted, stochastic description of an experiment is available, it is not recommended to perform a possibilistic analysis thereof as this would certainly yield too conservative results. If, however, there is good reason for doubting the precision of the

[24] See Section 3.1.1.

actual probabilistic description, it is worth considering possibilistic methods. These are meant to be employed in an imprecise probabilistic setting, i.e. if and only if one cannot conclusively rule out all probability measures but one.

From the specificity relation it is, furthermore, clear that some possibilistic structures possess larger and some possess smaller credal sets, and the fundamental question is how to 'best' express probabilistic imprecision in a possibilistic structure.

2.3.2.1 Fundamental Principles

In the following, three principles based on inclusion, i.e. consistency and specificity, and on plausibility conformity are proposed that, in the author's view, are simple, intuitive and straight-forward to apply.

The Principle of Representation The *Principle of Representation* can be summarized as 'what is probable must be [possible]'.[25] As discussed earlier, it is much more than a principle, it is what gives a quantitative meaning to the methods developed in this thesis, and, therefore, it possesses a special significance. A possibilistic structure must act as a placeholder for all consistent probability measures, and when reasoning with the former, the principal interest must be the analysis of the latter. Any result obtained by the possibilistic analysis must, therefore, serve as a summary of results obtained by the corresponding possibilistic analysis performed on the elements of the credal set. Only if this principle is obeyed in every step of the analysis, the upper probability interpretation of possibility is preserved, and only then possibilistic assertions are meaningful in the IP sense postulated in this thesis.

To build some intuition into the necessity of the further principles, consider the *vacuous possibility measure* given by the vacuous elementary possibilities $\pi^{\text{vac.}}(\omega) = 1$ for all $\omega \in \Omega$ on a measurable space (Ω, Σ). This possibility measure is consistent with all probability measures $\mathrm{P} \in \mathbb{P}(\Omega, \Sigma)$, and certainly adheres to the Principle of Representation under all given circumstances; but it is little expressive, as it bounds every event $E \in \Sigma$ with the trivial lower probability $\mathrm{N}^{\text{vac.}}(E) = 0$ and the trivial upper probability $\Pi^{\text{vac.}}(E) = 1$ with the exception of the case $E = \emptyset$ where $\Pi^{\text{vac.}}(\emptyset) = 0$ and the case $E = \Omega$ where $\mathrm{N}^{\text{vac.}}(\Omega) = 1$. Furthermore, no ranking in terms of the plausibility order is provided. Every elementary event $\omega \in \Omega$ is seen as equally plausible. The vacuous possibility measure does not provide any non-trivial insight into the experiment and does not account for additional information. It is most definitely never wrong to presuppose this model,[26] especially in

[25] A variant of 'what is probable must be plausible' [DuboisPrade88, p. 121] which itself is borrowed from the Laplace Principle of Insufficient Reason 'what is equipossible must be equiprobable'.

[26] In a logical sense, it is a tautology.

the case of *complete ignorance*[27] with respect to the true probabilities, yet it clearly lacks expressiveness.

The Principle of Expressiveness The *Principle of Expressiveness*, which is akin to the *Principle of Minimum/Maximum Specificity* postulated and advocated by Dubois et al. [Mauris09, DuboisPrade86a], states that 'what is [less] probable must be [less possible]'.[28] A possibility measure should be specified as specific as possible, i.e., given some constraints it has to satisfy, a most specific possibility measure should be selected among the possible choices. Among other things, this implies that the vacuous probability measure ought to be the last resort when modeling imprecise probabilities and that the elementary possibilities should be pointwise as low as possible, subject to the constraints imposed by the Principle of Representation.

For instance, given two possibility measures intended to describe the same information where one is more specific than the other, it follows from the inclusion property of credal sets that the credal set of the former is contained in the credal set of the latter. It includes fewer probability measures and can therefore provide tighter bounds on the probabilities, making it the preferred choice. Still, given that the overall paradigm is a robust analysis of the imprecise probabilities, especially near-robust approximations, even if they would drastically improve expressiveness, seem questionable.

Unfortunately, it is not generally possible to specify *the* unique most expressive and representative possibility measure as not all possibility measures are comparable with respect to $\preceq$. Reconsider, e.g., a possibilistic model of a 'fair' coin. The Principle of Representation postulates that both elementary possibilities ought to be at least as high as 50%, whereas the Principle of Expressiveness postulates that at least one of the elementary possibilities ought to be no higher than 50%; due to the normality criterion the other elementary possibility must be one. But, from these two principles alone, no specifications as to which elementary possibility ought to be chosen as $\frac{1}{2}$ and which one to be chosen as one can be inferred. Both choices are incomparable with respect to the specificity order. To resolve the issue of deciding between otherwise equivalent models, a plausibility order must also be provided.

The Principle of Plausibility The *Principle of Plausibility* states that 'what is [possible] must be plausible'.[29] The degrees of possibility, especially the elementary possibilities, ought to convey the general plausibility, and the plausibility order should represent a comprehensible assessment of the experiment. Ideally, the plausibility order of an objective

[27] Further discussion of the vacuous possibility measure as a representation of total ignorance will be provided in Section 3.1.2.

[28] See footnote 25.

[29] See footnote 25.

possibility measure ought to be based on some objective criterion, such as the likelihood; it is, however, permissible to base it on subjective criteria, such as an agents opinion, as this will not, generally, render the imprecise probability interpretation invalid. The Principle of Plausibility merely suggests how the elementary possibilities should be ordered.

In summary, a suitable possibility measure has to robustly, but tightly, account for all the information about the imprecisely known probabilities and should reflect a plausibility order. Quantitative possibility theory is the framework for imprecise probabilities based on such possibility measures [DuboisPrade98] and constitutes the core concept in this thesis. It adheres to all three fundamental principles of possibility theory.

These fundamental principles can be justified further: The theory of belief functions, which is one historic origin of possibility theory, emerged as a theory of statistical inference [Shafer76], not as a theory of imprecise probabilities. Only later, it was discovered that plausibility measures, and therefore possibility measures as well, could be interpreted as upper probabilities [WangKlir13, DuboisPrade92]. Consequently, possibility theory can also be derived, independently of the approach presented in this thesis, as a theory of statistical inference [LiuMartin20]. Incidentally, Martin and Liu [MartinLiu15] also present two fundamental principles, the Principles of Validity and Efficiency, describing what 'good' statistical inference ought to accomplish. These principles are completely analog to the Principles of Representation and Expressiveness, which describe what 'good' reasoning with imprecise probabilities ought to accomplish. As explained, the Principle of Plausibility is not a 'hard' constraint but merely a pointer toward sensible plausibility orderings and is, in essence, a generalization of the Likelihood Principle, which is crucial in statistical inference. The strong connection between Martin and Liu's *inferential models* and possibilistic imprecise probability models will further be investigated in the remainder of this thesis; nevertheless, the fact that this set of principles can be motivated in at least two superficially unrelated ways makes a compelling argument for their adequacy.

In order to avoid ambiguity, a clear distinction between qualitative and quantitative possibility measures must be made. Therefore, the elementary possibility function of a qualitative possibility measure will henceforth be called an *elementary plausibility function* denoted by ρ instead of π which is reserved for elementary possibility functions of quantitative possibility measures with an interpretation as an upper probability. The definition of an elementary plausibility function would, in theory, not require normality—in fact, one could even replace its image $[0, 1]$ by any metric space. But imposing the same definition guarantees that elementary possibilities can always also be interpreted as elementary plausibilities, and vice versa. Whenever a sub-normal elementary plausibility function ρ is provided in this thesis, this is done in order to avoid unnecessary technicalities such as defining the corresponding normalized elementary plausibility function $\bar{\rho}$ given by

$$\bar{\rho}(\omega) = \frac{\rho(\omega)}{\sup_{\zeta \in \Omega} \rho(\zeta)} \tag{2.50}$$

for all $\omega \in \Omega$; the results are usually independent of the assumed normality. Nevertheless, supernormal—or even unbounded—plausibility functions are not considered.

Equipped with these principles, the question at the beginning of this section, regarding the problem of how to encode knowledge in the form of imprecise probabilities by possibilities, can now be answered satisfactorily.

2.3.2.2 The Imprecise-Probability-to-Possibility Transform

Consider an arbitrary set of probability measures $\mathfrak{P} \subseteq \mathbb{P}(\Omega, \Sigma)$ defined on the measurable space (Ω, Σ) which, in this thesis, shall serve as the most general description of imprecise probabilities.[30] This set may be expressed in very different ways. For instance, it could be an arbitrary collection of individual probability measures $\mathfrak{P} = \{\mathrm{P}_1, \mathrm{P}_2, ...\}$, or it could be defined by an incomplete characterization, stating, e.g., that 'the probability of event $E \in \Sigma$ is twice the probability of event $E_2 \in \Sigma$', i.e. $\mathfrak{P} = \{\mathrm{P} \in \mathbb{P}(\Omega, \Sigma) : \mathrm{P}(E) = 2\,\mathrm{P}(E_2)\}$. It may even stem from a different description of imprecise probabilities, e.g., lower previsions [Walley91] or p-boxes [FersonEtAl15], or be vacuous, i.e., $\mathfrak{P} = \mathbb{P}(\Omega, \Sigma)$.

Now, the goal is to find an elementary possibility function describing this set $\mathfrak{P}$. If $\mathfrak{P} = \mathfrak{C}(\pi)$, then π is said to be an *exact possibilistic description* of $\mathfrak{P}$. Otherwise, if $\mathfrak{P} \subseteq \mathfrak{C}(\pi)$, it is an *outer possibilistic description/approximation*; if $\mathfrak{C}(\pi) \subseteq \mathfrak{P}$, it is an *inner possibilistic description/approximation*.

To this end, let a qualitative ranking of the elementary events $\omega \in \Omega$ be given through an elementary plausibility function $\rho : \Omega \to [0, 1]$. Below, a variety of suitable elementary plausibility functions for various imprecise probability descriptions will be discussed in detail.

The Principles of Representation, Expressiveness and Plausibility lead to a vaguely formulated mathematical problem for a possibilistic description of $(\mathfrak{P}, \rho)$. The goal is to find the most expressive possibility measure Π which includes all $\mathrm{P} \in \mathfrak{P}$, excludes as many $\mathrm{P} \notin \mathfrak{P}$ as possible, and where equiplausible elementary events under $\sqsubseteq_\rho$ also have the same possibility. In other words, the Principle of Representation and the Principle of Plausibility provide some inevitable constraints on the shape of π while the Principle of Expressiveness provides some optimality conditions.

Figuratively speaking, these principles correspond to the bounding problem illustrated in Figure 2.1 where the star-shaped family of probability measures $\mathfrak{P}$ is to be described by a possibility measure. Due to the very specific structure of these measures, their credal sets can only have certain geometries. For the purpose of illustration, say, e.g., that they can only be squares. The question is which square one ought to choose in order to robustly

[30] More general descriptions can, e.g., be achieved via the specification of desirable gambles in the theory of lower previsions [TroffaesMirandaDestercke13].

describe $\mathfrak{P}$. The Principle of Representation excludes the blue square for its inability to cover all of the star which would imply that some of the probability measures in $\mathfrak{P}$ are not actually contained in the credal set. The Principle of Expressiveness requires minimal size which would exclude the pink square for its unnecessary inclusion of far too many probability measures that do not not pertain to $\mathfrak{P}$. The green square would always be preferable to the pink one because the centers and orientations are equal, but the green credal set has a smaller volume without violating the Principle of Representation. Lastly, the Principle of Plausibility can be interpreted as a prescribed orientation of the square, enabling the decision whether to choose the orange or the green credal set which are otherwise equivalent, e.g., with respect to their size and center.

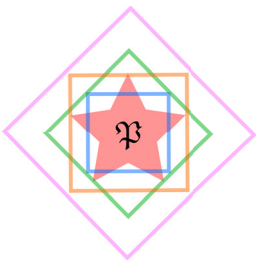


Figure 2.1: Illustration of the Imprecise-Probability-to-Possibility Transform.

The *Imprecise-Probability-to-Possibility Transform* (IP-Π-transform) [HoseHanss20, HoseHanss21c] states this mathematical program more precisely by defining the elementary function $\pi = \mathfrak{T}[\mathfrak{P}, \rho]$ whose values are given by

$$\pi(\omega) = \sup_{\mathrm{P} \in \mathfrak{P}} \mathrm{P}\left(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}\right) \tag{2.51}$$

for all $\omega \in \Omega$. If $\mathfrak{P} = \{\mathrm{P}\}$ is composed of one probability measure only, the IP-Π-transform is also referred to as the *Probability-to-Possibility Transform* (P-Π-transform) of P.

Before understanding the properties of the IP-Π-transform, some technical remarks are given.

Remark 6. Since ρ is measurable, the sublevel sets of ρ are also measurable, i.e., $\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\} \in \Sigma$, and the IP-Π-transform is well-defined.

Remark 7. The expression in Eq. (2.51) does not require ρ to be normal. In fact, every monotone rescaling of the elementary plausibility function would yield the same transformation. Thus, sub- or supernormal elementary plausibility functions are also admissible in the IP-Π-transform.

Remark 8. The probability $\mathrm{P}\left(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}\right)$ is monotonously increasing with respect to the value of $\rho(\omega)$ for all $\mathrm{P} \in \mathbb{P}(\Omega, \Sigma)$ due to the monotonicity of the probability measure. This is referred to as property $(*)$ in the proofs of Propositions 9, 11, 12, and 13.

First and foremost, it needs to be verified that the IP-Π-transform actually defines an elementary possibility function, i.e. that it is normal.

Proposition 9. *The* IP-Π*-transform defines an elementary possibility function.*

Proof. Trivially, $\pi(\omega) \in [0,1]$ for all $\omega \in \Omega$ because the probability measure is complete and bounded. In particular, this implies that $\sup_{\omega\in\Omega} \pi(\omega) \leq 1$. To prove the normality of π, let $\mathrm{P}^* \in \mathfrak{P}$. Due to the monotonicity property $(*)$, one can write $\sup_{\omega\in\Omega} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}) = \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \sup_{\omega\in\Omega} \rho(\omega)\})$ which is equivalent to $\mathrm{P}(\Omega) = 1$. Since $\sup_{\omega\in\Omega} \pi(\omega) \geq \sup_{\omega\in\Omega} \mathrm{P}^*(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}) = 1$, it is also clear that $\sup_{\omega\in\Omega} \pi(\omega) = 1$. □

Remark 10. In the proof of Proposition 9, the measurability of π is not mentioned, and without regularizing assumptions on the geometry of $\mathfrak{P}$, guarantees are hard to provide. Depending on the scenario, this must be addressed individually; however, for well-behaved, e.g. convex, $\mathfrak{P}$, this does not pose a serious issue.

The following propositions correspond to the different fundamental principles formulated above. Most importantly, it can be shown that the IP-Π-transform conforms to the Principle of Representation.

Proposition 11. *The elementary possibility function* π *defined by the* IP-Π*-transform is consistent with all* $\mathrm{P} \in \mathfrak{P}$.

Proof. Let $\mathrm{P}' \in \mathfrak{P}$ and $\alpha^* \in [0,1]$. Defining the maximum plausibility $\rho^* = \sup_{\omega\in\mathcal{S}_\pi^{\alpha^*}} \rho(\omega)$ and the corresponding sublevel set $E^* = \{\zeta \in \Omega : \rho(\zeta) \leq \rho^*\}$ yields $\mathcal{S}_\pi^{\alpha^*} \subseteq E^*$. From the monotonicity of the probability measure it follows that $\mathrm{P}'\left(\mathcal{S}_\pi^{\alpha^*}\right) \leq \mathrm{P}'(E^*)$, and from property $(*)$ in Remark 8 it follows that $\mathrm{P}'(E^*) = \sup_{\omega\in\mathcal{S}_\pi^{\alpha^*}} \mathrm{P}'(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\})$. Considering that $\mathrm{P}'(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}) \leq \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\})$, it is evident that $\mathrm{P}'(\mathcal{S}_\pi^{\alpha^*}) \leq \sup_{\omega\in\mathcal{S}_\pi^{\alpha^*}} \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\})$. By the definition in Eq. (2.51), $\pi(\omega) = \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\})$. Observing that $\sup_{\omega\in\mathcal{S}_\pi^{\alpha^*}} \pi(\omega) \leq \alpha^*$, finally, yields the consistency criterion $\mathrm{P}'(\mathcal{S}_\pi^{\alpha^*}) \leq \alpha^*$ and it is proven that $\mathrm{P}' \preceq \Pi$. □

Furthermore, it can be shown that the Principle of Plausibility is adhered to by the IP-Π-transform.

Proposition 12. *The elementary possibility function* π *defined by the* IP-Π*-transform is plausibility-conform to the elementary plausibility function* ρ.

Proof. Let $\omega_1, \omega_2 \in \Omega$ with $\rho(\omega_1) \leq \rho(\omega_2)$. Then, immediately, it follows by the monotonicity property $(*)$ that

$$\sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega_1)\}) \leq \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega_2)\})$$

i.e., $\pi(\omega_1) \leq \pi(\omega_2)$. □

A converse proposition may not be given. It is always possible that $\pi(\omega_1) = \pi(\omega_2)$ even though $\rho(\omega_1) < \rho(\omega_2)$ for some combination $\omega_1, \omega_2 \in \Omega$, and, therefore, that ρ is not plausibility-conform to π.

Lastly, also the Principle of Expressiveness is integrated into the IP-Π-transform.

Proposition 13. *Among all possibility densities which include all* $\mathrm{P} \in \mathfrak{P}$ *and are plausibility-conform to* ρ*, the elementary possibility function* π *defined by the* IP-Π*-transform is maximally specific.*

Proof. Let π' be an elementary possibility function that is plausibility-conform to ρ and consistent with all $\mathrm{P} \in \mathfrak{P}$, let $\omega \in \Omega$, and let $\alpha = \pi'(\omega)$. Consistency implies $\mathrm{P}(\mathcal{S}^{\alpha}_{\pi'}) \leq \alpha$ for all $\mathrm{P} \in \mathfrak{P}$, and therefore $\sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\mathcal{S}^{\alpha}_{\pi'}) \leq \alpha$. The plausibility-equivalence implies that for all $\zeta \in \Omega$ with $\rho(\zeta) \leq \rho(\omega)$ it follows that $\pi'(\zeta) \leq \pi'(\omega) = \alpha$, and therefore, $\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\} \subseteq \{\zeta \in \Omega : \pi'(\zeta) \leq \alpha\} = \mathcal{S}^{\alpha}_{\pi'}$. Finally, by considering the definition in Eq. (2.51), it follows that

$$\pi(\omega) = \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}\left(\{\zeta \in \Omega : \rho(\zeta) \leq \rho(\omega)\}\right) \leq \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}\left(\mathcal{S}^{\alpha}_{\pi'}\right) \leq \alpha^* = \pi'(\omega)$$

i.e., π is at least as specific as π'. □

The IP-Π-transform is fundamental to the remainder of this thesis. Its nature justifies calling possibility theory an Integrate-First approach because the (imprecise) probability measures $\mathfrak{P}$ are integrated on the sublevel sets of ρ before any further reasoning is performed, and these sublevel sets are the basis for the sublevel sets of π, which is ensured by the Principle of Plausibility. Put differently, the IP-Π-transform can also be understood as a monotone rescaling $\rho \mapsto \pi$ of the elementary plausibilities according to the imprecise probability mass contained in the corresponding sublevel sets of ρ. This observation would, in most cases, be enough to show the measurability of π.

The Optimal Imprecise-Probability-to-Possibility Transform The choice of $\mathfrak{P}$ is usually not debatable, as it must clearly describe the available (and only this!) information about the true probability distribution. The non-unique choice of the elementary plausibilities, however, indicates that there is not necessarily one 'best' way to perform a possibilistic analysis. In essence, it introduces infinitely many degrees of freedom to be fixed by the analyst. However, some choices seem more prudent than others, and suitable choices constitute a large portion of this thesis. A good default choice is the (potentially subnormal) *optimal elementary plausibility function* given by

$$\rho^{\mathrm{opt.}}_{\mathfrak{P}}(\omega) = \sup_{\mathrm{P}\in\mathfrak{P}} \mathrm{P}(\{\omega\}) \tag{2.52}$$

for $\omega \in \Omega$. The corresponding *Optimal* IP-Π*-transform* $\mathfrak{O}[\mathfrak{P}] = \mathfrak{T}[\mathfrak{P}, \rho^{\mathrm{opt.}}_{\mathfrak{P}}]$ yields the *optimal possibilistic description* of $\mathfrak{P}$. The term 'optimal' is owed to the close connection of

this transform to the Optimal P-Π-transform introduced by Dubois et al. [DuboisEtAl04], see Section 3.1.4.1.

The IP-Π-transform provides a constructive way of modeling and reasoning with imprecise probabilities. The generality of this result as well as its universal applicability shall be demonstrated in the following example and throughout the remainder of this thesis.

Example 4: IP-Π-transform

An urn contains 10 balls in total, out of which 6 are certainly red and 3 are certainly blue. The other ball is either blue or green. This information may be gathered in an imprecisely known (multinomial) probability measure P_θ on the sample space $\Omega = \{\omega_{\text{green}}, \omega_{\text{blue}}, \omega_{\text{red}}\}$ with the elementary probabilities

$$\mathrm{P}_\theta(\{\omega_{\text{green}}\}) = \theta, \quad \mathrm{P}_\theta(\{\omega_{\text{blue}}\}) = 0.4 - \theta \quad \text{and} \quad \mathrm{P}_\theta(\{\omega_{\text{red}}\}) = 0.6$$

for $\theta \in \{0, 0.1\}$. The corresponding family of multinomial probability measures is given by the (disconnected) set $\mathfrak{P} = \{\mathrm{P}_\theta : \theta \in \{0, 0.1\}\}$, and the optimal elementary plausibilities are given by

$$\rho_{\mathfrak{P}}^{\text{opt.}}(\omega_{\text{green}}) = 0.1, \qquad \rho_{\mathfrak{P}}^{\text{opt.}}(\omega_{\text{blue}}) = 0.4 \qquad \text{and} \qquad \rho_{\mathfrak{P}}^{\text{opt.}}(\omega_{\text{red}}) = 0.6.$$

Applying the Optimal IP-Π-transform yields $\pi = \mathfrak{O}[\mathfrak{P}]$ with the elementary possibilities

$$\pi(\omega_{\text{green}}) = 0.1, \qquad \pi(\omega_{\text{blue}}) = 0.4 \qquad \text{and} \qquad \pi(\omega_{\text{red}}) = 1.$$

Evidently, π is plausibility-conform to ρ and consistent with all $\mathrm{P} \in \mathfrak{P}$. Yet, it also admits probability measures which are not contained in $\mathfrak{P}$, such as

$$\mathrm{P}^{\notin}(\{\omega_{\text{green}}\}) = 0.1, \quad \mathrm{P}^{\notin}(\{\omega_{\text{blue}}\}) = 0.2 \quad \text{and} \quad \mathrm{P}^{\notin}(\{\omega_{\text{red}}\}) = 0.7.$$

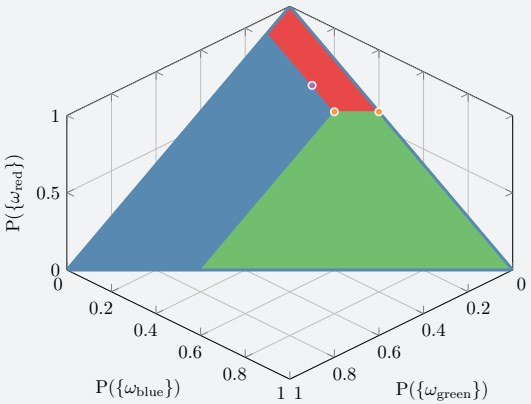
A different choice of the elementary plausibilities, e.g.

$$\rho'(\omega_{\text{green}}) = 0.5, \qquad \rho'(\omega_{\text{blue}}) = 1 \qquad \text{and} \qquad \rho'(\omega_{\text{red}}) = 0,$$

also yields different elementary possibilities, in this case

$$\pi'(\omega_{\text{green}}) = 0.7, \qquad \pi'(\omega_{\text{blue}}) = 1 \qquad \text{and} \qquad \pi'(\omega_{\text{red}}) = 0.6,$$

which does not include $\mathrm{P}^{\notin}$ but other probability measures that are not in $\mathfrak{P}$. Finally, consider the following illustration.



The triangular set of all probability measures $\mathbb{P}(\Omega, 2^{\Omega})$ is represented in blue and the (two-element) family of probability measures $\mathfrak{P}$ in orange. It constitutes a part of the edge and the vertices of both the optimal credal set $\mathfrak{C}(\pi)$ in red and the alternative credal set $\mathfrak{C}(\pi')$ in green. These active constraints are, generally, an indicator of maximum specificity. The probability measure $\mathrm{P}^{\notin}$ is depicted in purple. From the 'size' of the credal sets, it certainly seems advisable to prefer the Optimal IP-Π-transform over the IP-Π-transform under ρ'—not least because it better conforms with the Principle of Plausibility.

Contrary to the example above, it is not uncommon that an analytical evaluation of the (Optimal) IP-Π-transform is not possible. Then, robust outer, i.e. less specific, approximations must be provided in order not to violate the Principle of Representation, and expressiveness must be sacrificed for the sake of robustness. If this is the case, additional discussion regarding the actual loss of expressiveness shall be provided.

Finally, it remains to be established when a possibilistic description of $\mathfrak{P}$ via an elementary possibility function π is *exact*, i.e. when $\mathfrak{P} = \mathfrak{C}(\pi)$.

Invariance and Exact Possibilistic Descriptions A preliminary property of the Optimal IP-Π-transform is given by the following proposition stating that, if the family of probability distributions specifies a credal set already, the corresponding elementary possibility function can be recovered by the Optimal IP-Π-transform without any loss of expressiveness.

Proposition 14. *An elementary possibility function and its associated credal set are invariant under the Optimal* IP-Π*-transform.*

Proof. Let π be an elementary possibility function on the measurable space (Ω, Σ) and let $\mathfrak{C}(\pi)$ be its associated credal set. By Eq. (2.52), the corresponding optimal elementary plausibilities of $\omega \in \Omega$ are given by $\rho^{\text{opt.}}_{\mathfrak{C}(\pi)}(\omega) = \pi(\omega)$. Finally, the equality proving invariance $\pi(\omega) = \Pi(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{C}(\pi)}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{C}(\pi)}(\omega)\}) = \sup_{\mathrm{P} \in \mathfrak{C}(\pi)} \mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\})$ follows from the coherence of the possibility measure and from Eq. (2.25). □

This indicates that there is no loss of information, i.e., there is no inclusion $\mathrm{P} \in \mathfrak{C}(\mathfrak{O}[\mathfrak{P}])$ of undesired probability measures $\mathrm{P} \notin \mathfrak{P}$, under the Optimal IP-Π-transform if $\mathfrak{P} = \mathfrak{C}(\pi)$ possesses the shape of a possibilistic credal set already. Put differently, $\mathfrak{C}(\pi) = \mathfrak{C}(\mathfrak{O}[\mathfrak{C}(\pi)])$. If anything, this provides a good argument for the name of the Optimal IP-Π-transform. It is also possible to turn this result into a straightforward way of characterizing *exact possibilistic descriptions*.

Lemma 15. *The optimal plausibility function $\rho^{\text{opt.}}_{\mathfrak{P}}$ is an exact possibilistic description of $\mathfrak{P}$, i.e., $\mathfrak{C}(\rho^{\text{opt.}}_{\mathfrak{P}}) = \mathfrak{P}$, if and only if*

$$\sup_{\mathrm{P} \in \mathfrak{P}} \mathrm{P}\left(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{P}}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)\}\right) = \rho^{\text{opt.}}_{\mathfrak{P}}(\omega) \tag{2.53}$$

for all $\omega \in \Omega$.

Proof. The proof of the "⇒"-direction is similar to the proof of Proposition 14: Let π be an elementary possibility function on the measurable space (Ω, Σ). Then, the corresponding optimal elementary plausibilities of its credal set $\mathfrak{C}(\pi)$ are given by $\rho^{\text{opt.}}_{\mathfrak{P}}(\omega) = \pi(\omega)$ for all $\omega \in \Omega$ and, from the coherence of the possibility measure and from Eq. (2.25), it follows that

$$\sup_{\mathrm{P} \in \mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{P}}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)\}) = \sup_{\mathrm{P} \preceq \pi} \mathrm{P}(\{\zeta \in \Omega : \pi(\zeta) \leq \pi(\omega)\}) = \pi(\omega) = \rho^{\text{opt.}}_{\mathfrak{P}}(\omega).$$

To prove the "⇐"-direction, let $\mathfrak{P}$ be a family of probabilities with the optimal elementary plausibility function $\rho^{\text{opt.}}_{\mathfrak{P}}$ such that $\sup_{\mathrm{P} \in \mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{P}}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)\}) = \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)$ for all $\omega \in \Omega$. The evaluation of the corresponding Optimal IP-Π-transform (2.51) reads $\pi(\omega) = \sup_{\mathrm{P} \in \mathfrak{P}} \mathrm{P}(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{P}}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)\}) = \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)$ guaranteeing that $\mathfrak{P} \subseteq \mathfrak{C}(\rho^{\text{opt.}}_{\mathfrak{P}}) = \mathfrak{C}(\pi)$. Conversely, every $\mathrm{P} \in \mathfrak{C}(\rho^{\text{opt.}}_{\mathfrak{P}})$ is certainly contained in $\mathfrak{P}$ because $\sup_{\mathrm{P} \preceq \pi} \mathrm{P}(\{\zeta \in \Omega : \rho^{\text{opt.}}_{\mathfrak{P}}(\zeta) \leq \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)\}) = \rho^{\text{opt.}}_{\mathfrak{P}}(\omega)$ by Eq. (2.25) and the coherence of the possibility measure, and it can be concluded that $\mathfrak{P} = \mathfrak{C}(\rho^{\text{opt.}}_{\mathfrak{P}})$. □

This lemma is a good example where elementary plausibilities act as elementary possibilities and vice versa. That is, the exact possibilistic description is given by $\pi = \rho^{\text{opt.}}_{\mathfrak{P}}$.

The difference to earlier results about the IP-Π-transform is that the exact possibilistic description guarantees a loss-free Optimal IP-Π-transform of $\mathfrak{P}$. In the remainder of this thesis, this lemma shall be used extensively both to characterize and to construct exact possibilistic descriptions of various sets $\mathfrak{P}$ if the equality $\mathfrak{P} = \mathfrak{C}(\pi)$ cannot be shown directly.

2.3.3 Combinations of Possibilities

Finally, it is possible for the same experiment to be described by several elementary possibility functions $\pi^{(1)}, \ldots, \pi^{(N)}$ on (Ω, Σ). In the quantitative view of possibilities advocated in this thesis, the combination of several such possibilistic descriptions reduces to operations on the corresponding credal sets, i.e. their union and—more importantly—their intersection.

As demonstrated by Hose and Hanss [HoseHanss21c], several earlier results, e.g. by Bronevich et al. [Bronevich01, BronevichRozenberg18, BronevichRozenberg19], may be re-derived under the various IP-Π-transforms.

2.3.3.1 Disjunction

If the possibilistic descriptions are to be combined in a disjunctive manner, i.e., if either of the descriptions is assumed to be correct, their credal sets must be joined resulting in the family

$$\mathfrak{P}^{\text{disj.}} = \bigcup_{i=1}^{N} \mathfrak{C}(\pi^{(i)}), \tag{2.54}$$

which is written as $\pi^{(1)} \vee \ldots \vee \pi^{(N)}$. An optimal description $\pi^{\text{disj.}}$ thereof is easily obtained.

Lemma 16. *Let $\pi^{(1)}, \ldots, \pi^{(N)}$ be elementary possibility functions on (Ω, Σ). Then, $\pi^{\text{disj.}}$ given by*

$$\pi^{\text{disj.}}(\omega) = \max_{i=1,\ldots,N} \pi^{(i)}(\omega) \tag{2.55}$$

for all $\omega \in \Omega$ is an optimal possibilistic description of $\pi^{(1)} \vee \ldots \vee \pi^{(N)}$.

Proof. For all $E \in \Sigma$, the respective upper probability is easily re-written as $\sup_{\mathrm{P} \in \mathfrak{P}^{\text{disj.}}} \mathrm{P}(E) = \max_{i=1,\ldots,N} \sup_{\mathrm{P}_i \in \mathfrak{C}(\pi^{(i)})} \mathrm{P}_i(E) = \max_{i=1,\ldots,N} \Pi_i(E)$ which follows directly from the coherence of $\pi^{(1)}, \ldots, \pi^{(N)}$. In particular, the optimal plausibilities are $\rho^{\text{opt.}}_{\mathfrak{P}^{\text{disj.}}}(\omega) = \max_{i=1,\ldots,N} \pi^{(i)}(\omega)$ for all $\omega \in \Omega$, which, in turn, yields the optimal possibilities

$$\pi^{\text{disj.}}(\omega) = \max_{i=1,\ldots,N} \Pi_i(\{\zeta \in \Omega : \max_{j=1,\ldots,N} \pi_j(\zeta) \leq \max_{k=1,\ldots,N} \pi_k(\omega)\}) = \max_{i=1,\ldots,N} \pi^{(i)}(\omega).$$

□

Remark 17. As the maximum of N measurable and normal functions, $\pi^{\text{disj.}}$ is certainly measurable and normal.

Notice that this does not provide an exact possibilistic description of $\mathfrak{P}^{\text{disj.}}$ because $\pi^{\text{disj.}}$ also allows for convex combinations of probability measures from the individual credal sets, even though these combinations may not be in either of the credal sets themselves. On a

final note, this solution also coincides with the application of the most popular maximum-based s-norm for the union of N fuzzy sets [Hanss05] with the respective membership function $\pi^{(1)}, \ldots, \pi^{(N)}$.

2.3.3.2 Conjunction

If all of the possibilistic descriptions are deemed to be reliable, the credal sets ought to be intersected as

$$\mathfrak{P}^{\text{conj.}} = \bigcap_{i=1}^{N} \mathfrak{C}\left(\pi^{(i)}\right), \tag{2.56}$$

written as $\pi^{(1)} \wedge \ldots \wedge \pi^{(N)}$. A possibilistic description of this conjunction is less straightforward than that of their disjunction—not least because the intersection could very well be empty. The t-norm used for fuzzy set intersections [Hanss05] which is dual to the maximum-based s-norm (or t-conorm) from the previous discussion of the possibilistic disjunction is given by the possibilistic copula for non-interaction $\mathcal{J}^{\text{NI}}$, to be discussed in Section 3.5, which produces the element-wise minimum

$$\pi^{\text{min.}}(\omega) = \min_{i=1,\ldots,N} \pi^{(i)}(\omega) \tag{2.57}$$

for all $\omega \in \Omega$.

As a first step, consider the following lemma stating that comonotonicity is a sufficient condition for an exact possibilistic description of the possibilistic conjunction.

Lemma 18. *Let $\pi^{(1)}, \ldots, \pi^{(N)}$ be N elementary possibility functions on (Ω, Σ). If they are comonotone, then the element-wise minimum $\pi^{\text{min.}}$ is an exact possibilistic description of $\pi^{(1)} \wedge \ldots \wedge \pi^{(N)}$.*

Proof. In Remark 5, it has already been argued that $\pi^{\text{min.}}$ is an elementary possibility function, and, therefore, that its credal set is not empty. It remains to be shown that $\mathfrak{C}(\pi^{\text{min.}}) = \mathfrak{P}^{\text{conj.}}$.
Under Lemma 3, it follows directly that $\mathfrak{C}(\pi^{\text{min.}}) \subseteq \mathfrak{P}^{\text{conj.}}$. Since $\pi^{\text{min.}} \preceq \pi^{(i)}$, every $\mathrm{P} \in \mathfrak{C}(\pi^{\text{min.}})$ is also in $\mathfrak{C}(\pi^{(\mathrm{i})})$ for all $i = 1, \ldots, N$, and therefore also in $\mathfrak{P}^{\text{conj.}}$.
To see that $\mathfrak{P}^{\text{conj.}} \subseteq \mathfrak{C}(\pi^{\text{min.}})$, let $\mathrm{P} \in \mathfrak{P}^{\text{conj.}}$, i.e., $\mathrm{P} \preceq \pi^{(i)}$ for all $i = 1, \ldots, N$, and let $\omega \in \Omega$. Under Lemma 2, it suffices to show that $\mathrm{P}(\{\zeta \in \Omega : \pi^{\text{min.}}(\zeta) \leq \pi^{\text{min.}}(\omega)\}) \leq \pi^{\text{min.}}(\omega)$. Due to the comonotonicity, $\mathrm{P}(\{\zeta \in \Omega : \pi^{\text{min.}}(\zeta) \leq \pi^{\text{min.}}(\omega)\}) \leq \mathrm{P}(\zeta \in \Omega : \pi^{(\mathrm{i})}(\zeta) \leq \pi^{(\mathrm{i})}(\omega))$ for all $i = 1, \ldots, N$, and since $\mathrm{P} \preceq \pi^{(i)}$, it follows that $\mathrm{P}(\{\zeta \in \Omega : \pi^{(\mathrm{i})}(\zeta) \leq \pi^{(\mathrm{i})}(\omega)\}) \leq \pi^{(\mathrm{i})}(\omega)$. Therefore, $\mathrm{P}(\{\zeta \in \Omega : \pi^{\text{min.}}(\zeta) \leq \pi^{\text{min.}}(\omega)\}) \leq \min_{i=1,\ldots,N} \pi^{(\mathrm{i})}(\omega) = \pi^{\text{min.}}(\omega)$. □

Remark 19. As the minimum of N measurable and normal functions, $\pi^{\text{conj.}}$ is certainly measurable and normal.

The applicability of the element-wise minimum may also be derived when changing the assumption of comonotonicity to that of quasi-vacuousness.

Lemma 20. *Let $\pi^{(1)}, \ldots, \pi^{(N)}$ be N elementary possibility functions on (Ω, Σ), where $\pi^{(2)}, \ldots, \pi^{(N)}$ are quasi-vacuous. Then the element-wise minimum $\pi^{\min.}$ is an exact possibilistic description of $\pi^{(1)} \wedge \ldots \wedge \pi^{(N)}$.*

Proof. Under Lemma 3, it follows directly that $\mathfrak{C}(\pi^{\min.}) \subseteq \mathfrak{P}^{\text{conj.}}$. Since $\pi^{\min.} \preceq \pi^{(i)}$, every $\mathrm{P} \in \mathfrak{C}(\pi^{\min.})$ is also contained in $\mathfrak{C}(\pi^{(i)})$ for all $i = 1, \ldots, N$, and therefore also in $\mathfrak{P}^{\text{conj.}}$.
To see that $\mathfrak{P}^{\text{conj.}} \subseteq \mathfrak{C}(\pi^{\min.})$, let $\mathrm{P} \in \mathfrak{P}^{\text{conj.}}$, i.e., $\mathrm{P} \preceq \pi^{(i)}$ for all $i = 1, \ldots, N$, and let $\alpha \in [0, 1]$. Under Lemma 1, it suffices to show that $\mathrm{P}(\mathcal{S}^{\alpha}_{\pi^{\min.}}) = \mathrm{P}(\bigcup_{i=1}^{N} \mathcal{S}^{\alpha}_{\pi^{(i)}}) \leq \alpha$. Trivially, this is always fulfilled for $\alpha = 1$, therefore one may now consider the case $\alpha < 1$ which directly implies that $\mathcal{S}^{\alpha}_{\pi^{(i)}} = \mathcal{S}^{0}_{\pi^{(i)}}$ for all $i = 2, \ldots, N$ (but not for $i = 1$), and therefore, $\mathcal{S}^{\alpha}_{\pi^{\min.}} = \mathcal{S}^{\alpha}_{\pi^{(1)}} \cup (\bigcup_{i=2}^{N} \mathcal{S}^{0}_{\pi^{(i)}})$. Since $\mathrm{P} \preceq \pi^{(i)}$, i.e., $\mathrm{P}(\mathcal{S}^{0}_{\pi^{(i)}}) = 0$, it is clear that $\mathrm{P}(\bigcup_{i=2}^{N} \mathcal{S}^{0}_{\pi^{(i)}}) \leq \sum_{i=2}^{N} \mathrm{P}(\mathcal{S}^{0}_{\pi^{(i)}}) = 0$. Finally,

$$\mathrm{P}(\mathcal{S}^{\alpha}_{\pi^{\min.}}) \leq \mathrm{P}(\mathcal{S}^{\alpha}_{\pi^{(1)}}) + \mathrm{P}(\bigcup_{i=2}^{N} \mathcal{S}^{0}_{\pi^{(i)}}) = \mathrm{P}(\mathcal{S}^{\alpha}_{\pi^{(1)}}) \leq \alpha,$$

i.e., $\mathrm{P} \preceq \pi^{\min.}$, which concludes the proof. □

This result also holds true for the conjunction with totally vacuous elementary possibility functions as a special case of quasi-vacuous elementary possibility functions. The element-wise minimum, furthermore, produces an exact description of the conjunction if there exists a total specificity order among the elementary possibility functions.

Lemma 21. *Let $\pi^{(1)}, \ldots, \pi^{(N)}$ be N elementary possibility functions on (Ω, Σ) which are totally ordered with respect to the specificity order such that $\pi^{(1)} \preceq \ldots \preceq \pi^{(N)}$. Then, the element-wise minimum $\pi^{\min.} = \pi^{(1)}$ is an exact possibilistic description of $\pi^{(1)} \wedge \ldots \wedge \pi^{(N)}$.*

Proof. From the specificity ordering, it is evident that $\pi^{\min.}(\omega) = \pi^{(1)}(\omega) \leq \ldots \leq \pi^{(N)}(\omega)$ for all $\omega \in \Omega$ and that $\mathfrak{C}(\pi^{(1)}) \subseteq \ldots \subseteq \mathfrak{C}(\pi^{(N)})$. Therefore, $\mathfrak{P}^{\text{conj.}} = \mathfrak{C}(\pi^{(1)}) = \mathfrak{C}(\pi^{\min.})$. □

If $\pi^{(1)}, \ldots, \pi^{(N)}$ do not fulfill additional assumptions, such as comonotonicity, quasi-vacuousness or a total specificity order, an exact possibilistic description of the conjunction $\pi^{(1)} \wedge \ldots \wedge \pi^{(N)}$ cannot, generally, be provided. This is because the intersection of credal sets does not usually possess the shape of a credal set itself. Moreover, even the (Optimal) IP-Π-transform may only be evaluated approximately, i.e., only less specific outer descriptions can be found.

Regarding the optimal plausibilities of $\mathfrak{P}^{\text{conj.}}$, these are always bounded from above by the element-wise minimum since

$$\rho^{\text{opt.}}_{\mathfrak{P}^{\text{conj.}}}(\omega) = \sup_{\mathrm{P}\in\mathfrak{P}^{\text{conj.}}} P(\{\omega\}) \leq \min_{i=1,\dots,N} \Pi^{(i)}(\{\omega\}) = \min_{i=1,\dots,N} \pi^{(i)}(\omega) \tag{2.58}$$

for all $\omega \in \Omega$, which, incidentally, also constitutes the optimal plausibility function in the above lemmas. Based on these observations, the (potentially non-optimal) elementary plausibility function for the credal set intersection is chosen to be $\rho^{\text{conj.}} = \pi^{\text{min.}}$.

In the general case, this element-wise minimum does not necessarily describe an elementary possibility function since it may be subnormal; however, as the minimum of N measurable functions, it is still measurable and does constitute an admissible elementary plausibility function.

Lemma 22. *Let $\pi^{(1)}, \dots, \pi^{(N)}$ be N elementary possibility functions on (Ω, Σ). Then, the elementary possibility function $\pi^{\text{conj.}}$ given by*

$$\pi^{\text{conj.}}(\omega) = \min(1, N \cdot \rho^{\text{conj.}}(\omega)) \tag{2.59}$$

for all $\omega \in \Omega$ is an outer approximation of the IP-Π*-transform of $\mathfrak{P}^{\text{conj.}}$ under the elementary plausibility function $\rho^{\text{conj.}}$.*

Proof. Let $\mathrm{P} \in \mathfrak{P}^{\text{conj.}}$ and let $\omega \in \Omega$. The probability mass on the plausibility sublevel set $\{\zeta \in \Omega : \rho^{\text{conj.}}(\zeta) \leq \rho^{\text{conj.}}(\omega)\} = \bigcup_{i=1}^{N} \{\zeta \in \Omega : \pi^{(i)}(\omega) \leq \rho^{\text{conj.}}(\omega)\}$ is upper bounded by

$$\mathrm{P}(\{\zeta \in \Omega : \rho^{\text{conj.}}(\zeta) \leq \rho^{\text{conj.}}(\omega)\}) \leq \sum_{i=1}^{N} \mathrm{P}(\{\zeta \in \Omega : \pi^{(i)}(\omega) \leq \rho^{\text{conj.}}(\omega)\}) \leq N \cdot \rho^{\text{conj.}}(\omega)$$

or, trivially, by one, and therefore by $\pi^{\text{conj.}}(\omega) = \min(1, N \cdot \rho^{\text{conj.}}(\omega))$. That is, $\pi^{\text{conj.}}$ is less specific than the elementary possibility function resulting from the IP-Π-transform and, therefore, is an outer approximation. □

Remark 23. As the minimum of $N+1$ measurable functions, $\pi^{\text{conj.}}$ is certainly measurable.

This result has been derived by several authors in different contexts [Bronevich01, BronevichRozenberg18, BronevichRozenberg19, HoseHanss21c] and numerical experiments suggest that the outer approximation provided by this lemma is generally tight, i.e. the best possible.

As mentioned earlier, the intersection of the credal sets could potentially be empty. While this is impossible under the assumptions of Lemma 18 or Lemma 21 because the element-wise minimum certainly yields a normal elementary possibility function with a non-empty credal set, this possibility is not a-priori excluded in Lemmas 20 and 22, and may seem non-trivial to check. However, some general observations can be made:

- Lemma 20 may produce a sub-normal elementary possibility function. This immediately indicates that the intersection was empty to begin with because the element-wise minimum is an exact possibilistic description thereof. That is, the non-emptiness is easy to see.
- Lemma 22 provides an outer approximation which makes things less clear, and a sufficient criterion for determining whether the intersection is actually empty may not be given. Here, the normality of $\pi^{\text{conj.}}$ is only a necessary condition, and it could happen that the intersection is empty despite the normality.

Finally, other choices of the elementary plausibilities are certainly admissible and it does not seem impossible to find closed expressions of the exact and/or optimal elementary plausibilities and/or possibilities in further special cases. For instance, Bronevich and Karkishchenko derive other expressions under a different class of elementary plausibilities [Bronevich01] and Hose and Hanss [HoseHanss21c] generalize their results employing the IP-Π-transform, which is, however, not relevant for the purposes of this thesis.

The derived expressions for the disjunction and, more importantly, the conjunction of several elementary possibility functions shall prove useful in the remainder of this thesis.

Chapter 3

Imprecise Variables

> The Guide is definitive. Reality is frequently inaccurate.
>
> Douglas Adams,
> *The Restaurant at the End of the Universe*

Especially in engineering, the events of interest are usually expressed as some a-priori unknown quantity $\tilde{V}$ assuming a certain value or belonging to some set B, e.g. the measurement error of some sensor being lower than a given threshold. Depending on the outcome $\omega \in \Omega$ of the experiment, this quantity assumes the value $\tilde{V}(\omega) \in \mathbb{V}$. In order to measure the corresponding events $E = \{\omega \in \Omega : \tilde{V}(\omega) \in B\}$, they ought to be measurable; therefore, a $\mathbb{V}$-*valued imprecise variable*[31] is defined as a measurable function $\tilde{V} : \Omega \to \mathbb{V}$, and, by abuse of notation, the event E is expressed as $\tilde{V} \in B$.

The following discussion usually assumes *continuous real-valued imprecise variables*, i.e., $\mathbb{V} \subseteq \mathbb{R}^{D_{\mathbb{V}}}$, and Borel(-measurable) sets from the corresponding Borel σ-field $\mathbb{B}(\mathbb{V})$ as the events of interest, unless stated otherwise. Still, most results are easily transferable to other choices of $\mathbb{V}$ as well. If, for instance, $\mathbb{V}$ is a discrete space, then one may replace $\mathbb{B}(\mathbb{V})$ by $2^{\mathbb{V}}$.

3.1 Distributions

If M is a measure on the measurable space (Ω, Σ), the corresponding *pushforward* measure $\mathrm{M}_{\tilde{V}} : \mathbb{B}(\mathbb{V}) \to [0, 1]$ of M by the imprecise variable $\tilde{V}$, or simply the *distribution* of $\tilde{V}$,

[31] The more common term *random variable* is abandoned in order to emphasize the imprecise probabilistic approach in this thesis.

is defined on the measurable space $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ via

$$\mathrm{M}_{\tilde{V}}(B) = \mathrm{M}\left(\tilde{V} \in B\right) = \mathrm{M}\left(\{\omega \in \Omega : \tilde{V}(\omega) \in B\}\right) \tag{3.1}$$

for all $B \in \mathbb{B}(\mathbb{V})$. Notice that $\mathrm{M}_{\tilde{V}}$ merely re-expresses the information about (Ω, Σ) by confining Σ to (measurable) sets of the form $\tilde{V}^{-1}(B)$ for $B \in \mathbb{B}(\mathbb{V})$. That is, imprecise variables do nothing more than restricting Σ to the smaller σ-algebra

$$\Sigma|_{\tilde{V}} = \left\{\tilde{V}^{-1}(B) : B \in \mathbb{B}(\mathbb{V})\right\}. \tag{3.2}$$

Proposition 24. *Let $\mathrm{M}^{(1)}, \mathrm{M}^{(2)}$ be two measures on (Ω, Σ). Inclusion is preserved for the corresponding distributions, i.e. if $\mathrm{M}^{(1)} \preceq \mathrm{M}^{(2)}$ then $\mathrm{M}^{(1)}_{\tilde{V}} \preceq \mathrm{M}^{(2)}_{\tilde{V}}$.*

Proof. Let $B \in \mathbb{B}(\mathbb{V})$. Then $\mathrm{M}^{(1)}_{\tilde{V}}(B) = \mathrm{M}^{(1)}(\tilde{V} \in B) \leq \mathrm{M}^{(2)}(\tilde{V} \in B) = \mathrm{M}^{(2)}_{\tilde{V}}(B)$. □

A converse result may not be given.

3.1.1 Probability Distributions

Given a probability measure P on the measurable space (Ω, Σ), the pushforward probability measure $\mathrm{P}_{\tilde{V}}$ on $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ under a univariate $\mathbb{V}$-valued imprecise variable $\tilde{V}$, where $\mathbb{V} \subseteq \mathbb{R}$, is described by the *cumulative probability (distribution) function* (CPF) $F_{\tilde{V}} : \mathbb{V} \to [0, 1]$ given by

$$F_{\tilde{V}}(v) = \mathrm{P}\left(\tilde{V} \leq v\right) = \mathrm{P}_{\tilde{V}}\left((-\infty, v]\right) \tag{3.3}$$

for $v \in \mathbb{R}$. All CPFs are non-decreasing and right-continuous, and they provide an exhaustive representation of $\mathrm{P}_{\tilde{V}}$. In particular, the probability of an interval $(a, b]$ for $a \leq b$ can be obtained via $\mathrm{P}_{\tilde{V}}\left((a, b]\right) = F_{\tilde{V}}(b) - F_{\tilde{V}}(a)$, which is, in turn, sufficient to compute $\mathrm{P}_{\tilde{V}}(B)$ for arbitrary Borel sets $B \in \mathbb{B}(\mathbb{V})$ since these may be expressed as unions, intersections and complements of such intervals. Therefore, $\mathrm{P}_{\tilde{V}}$ and $F_{\tilde{V}}$ can be used interchangeably to denote a probability distribution.

The *complementary CPF* $\bar{F}_{\tilde{V}}$ is given by

$$\bar{F}_{\tilde{V}}(v) = \mathrm{P}\left(\tilde{V} \geq v\right) = \mathrm{P}_{\tilde{V}}\left([v, +\infty)\right) \tag{3.4}$$

for all $v \in \mathbb{R}$, and the (quasi-)inverse $F^{-1}_{\tilde{V}}$, given by

$$F^{-1}_{\tilde{V}}(\alpha) = \inf\left\{v \in \mathbb{V} : F_{\tilde{V}}(v) \geq \alpha\right\} \tag{3.5}$$

for all $\alpha \in [0, 1]$, is usually called the *quantile function*.

The generalization to CPFs of multivariate $(\mathbb{V}_1 \times \ldots \times \mathbb{V}_m)$-valued imprecise vectors $(\tilde{V}_1, \ldots, \tilde{V}_m)$ is analogously given by

$$\begin{aligned} F_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1, \ldots, v_m) &= \mathrm{P}\left(\tilde{V}_1 \leq v_1 \wedge \ldots \wedge \tilde{V}_m \leq v_m\right) \\ &= \mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left((-\infty, v_1] \times \ldots \times (-\infty, v_m]\right) \end{aligned} \tag{3.6}$$

for $(v_1, \ldots, v_m) \in \mathbb{R}^{D_\mathbb{V}}$.

In this thesis, every imprecise variable is assumed to follow some (perhaps only imprecisely known) probability distribution which is, depending on the available description, written as $\tilde{V} \sim \mathrm{P}_{\tilde{V}}$, or, equivalently, as $\tilde{V} \sim F_{\tilde{V}}$.

3.1.1.1 Uniform Distribution

The most important probability distribution in this thesis is the *uniform probability distribution* $\mathcal{U}(T)$ on the set $T \in \mathbb{B}(\mathbb{V})$ with the finite Lebesgue measure $\lambda(T) < \infty$, where T usually constitutes an interval or a higher-dimensional box. This probability distribution assigns to each $B \in \mathbb{B}(\mathbb{V})$ a probability proportional to the volume of the intersection with T, i.e.

$$\mathrm{P}_{\tilde{V}}^{\text{unif.}}(B) = \frac{\lambda(B \cap T)}{\lambda(T)}.$$

If $T = [a, b]$ is an interval with $a < b$, then the corresponding CPF is given by

$$F_{\tilde{V}}^{\text{unif.}}(v) = \begin{cases} 0 & \text{if } v < a, \\ \frac{v-a}{b-a} & \text{if } v \in [a, b] \text{ and} \\ 1 & \text{otherwise} \end{cases} \tag{3.7}$$

for all $v \in \mathbb{V}$.

Below, the *standard uniform distribution* $\mathcal{U}([0, 1])$ serves as a useful reference distribution.

3.1.2 Possibility Distributions

Given an elementary possibility function π on the measurable space (Ω, Σ) and a $\mathbb{V}$-valued imprecise variable $\tilde{V}$, the elementary possibility function $\pi_{\tilde{V}} : \mathbb{V} \to [0, 1]$ of $\tilde{V}$ is formally defined as

$$\pi_{\tilde{V}}(v) = \sup_{\omega \in \Omega \,:\, \tilde{V}(\omega) = v} \pi(\omega) \tag{3.8}$$

for all $v \in \mathbb{V}$. This definition is motivated by the following observation. The possibility distribution $\Pi_{\tilde{V}}$ on $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ defined by Eq. (3.1) can be computed from $\pi_{\tilde{V}}$ via

$$\Pi_{\tilde{V}}(B) = \sup_{\omega \in \Omega \,:\, \tilde{V}(\omega) \in B} \pi(\omega) = \sup_{v \in B,\, \omega \in \Omega \,:\, \tilde{V}(\omega) = v} \pi(\omega) = \sup_{v \in B} \pi_{\tilde{V}}(v) \tag{3.9}$$

for all $B \in \mathbb{B}(\mathbb{V})$. That is, $\pi_{\tilde{V}}$ is the elementary possibility function on $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ that corresponds to $\Pi_{\tilde{V}}$. For this reason, both $\pi_{\tilde{X}}$ and $\Pi_{\tilde{X}}$ can be used interchangeably to denote the *possibility distribution*. A plausibility distribution $\rho_{\tilde{V}}$ is defined analogously.

Additionally, $\pi_{\tilde{V}}$ also induces the pushforward necessity measure $\mathrm{N}_{\tilde{V}}$, the sublevel sets $\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}$ and the superlevel sets $\mathcal{C}^{\alpha}_{\pi_{\tilde{V}}}$ in the familiar way. A possibility distribution is said to be *convex* if the latter are convex.

By Proposition 24, pushforwards and distributions preserve consistency and specificity. In particular, the relation $\mathrm{P} \preceq \pi^{(1)} \preceq \pi^{(2)}$ implies $\mathrm{P}_{\tilde{V}} \preceq \pi^{(1)}_{\tilde{V}} \preceq \pi^{(2)}_{\tilde{V}}$, and the credal set $\mathfrak{C}(\pi_{\tilde{V}})$ of the possibility distribution $\pi_{\tilde{V}}$ is guaranteed to include the set of all admissible pushforward probability measures under π, which is given by $\mathfrak{P}_{\tilde{V}} = \{\mathrm{P}_{\tilde{V}} : \mathrm{P} \preceq \pi\}$. The following proposition strengthens the latter observation by showing equality.

Proposition 25. *The possibility distribution $\pi_{\tilde{V}}$ is an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}$.*

Proof. From the coherence $(*)$ of the possibility measure, it follows that for all $B \in \mathbb{B}(\mathbb{V})$

$$\sup_{\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}} \mathrm{P}_{\tilde{V}}(B) = \sup_{\mathrm{P} \preceq \pi} \mathrm{P}\left(\tilde{V} \in B\right) \overset{(*)}{=} \Pi\left(\tilde{V} \in B\right) = \Pi_{\tilde{V}}(B). \tag{3.10}$$

Letting $B = \{v\}$ in Eq. (3.10) yields $\rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(v) = \sup_{\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}} \mathrm{P}_{\tilde{V}}(\{v\}) = \Pi_{\tilde{V}}(\{v\}) = \pi_{\tilde{V}}(v)$ for all $v \in \mathbb{V}$, i.e., the optimal plausibilities of $\mathfrak{P}_{\tilde{V}}$ are given by $\rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}} = \pi_{\tilde{V}}$.
Similarly, considering $B = \{\xi \in \mathbb{V} : \rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(\xi) \leq \rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(v)\} = \{\xi \in \mathbb{V} : \pi_{\tilde{V}}(\xi) \leq \pi_{\tilde{V}}(v)\}$ for $v \in \mathbb{V}$ in Eq. (3.10) shows that, the requirements in Eq. (2.53) for applying Lemma 15 are met because

$$\begin{aligned} &\sup_{\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}} \mathrm{P}_{\tilde{V}}(\{\xi \in \mathbb{V} : \rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(\xi) \leq \rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(v)\}) = \Pi_{\tilde{V}}(\{\xi \in \mathbb{V} : \pi_{\tilde{V}}(\xi) \leq \pi_{\tilde{V}}(v)\}) \\ \overset{\text{Eq. (2.25)}}{=}\ &\pi_{\tilde{V}}(v) = \rho^{\text{opt.}}_{\mathfrak{P}_{\tilde{V}}}(v), \end{aligned}$$

and the proposition is proven. □

For this reason, possibility distributions also describe *imprecise probability distributions*. This is abbreviated as $\tilde{V} \sim \pi_{\tilde{V}}$, which must be understood as $\tilde{V} \sim \mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}$, i.e., the true (but possibly unknown) probability distribution $\mathrm{P}_{\tilde{V}}$ of $\tilde{V}$ is consistent with $\pi_{\tilde{V}}$.

It is, furthermore, possible to extend the notion of the CPF of an imprecise variable to *cumulative possibility distribution functions* (CΠF) and *cumulative necessity distribution functions* (CNF), via

$$\Pi(\tilde{V} \leq v) = \sup_{\xi \in \mathbb{V} : \xi \leq v} \pi_{\tilde{V}}(\xi) \qquad \text{and} \qquad \mathrm{N}(\tilde{V} \leq v) = \inf_{\xi \in \mathbb{V} : \xi > v} 1 - \pi_{\tilde{V}}(\xi) \tag{3.11}$$

for $v \in \mathbb{R}$, or the analog generalization in the multivariate case. Clearly, the CNF and the CΠF bound all CPFs of consistent probability distributions. However, they are not a sufficient description of the possibility distribution because $\pi_{\tilde{V}}$ cannot generally be reconstructed when knowing only the CNF and the CΠF [DesterckeDuboisChojnacki08].

Clearly, it is important to be able to rigorously define both pushforward probability and pushforward possibility measures in order to be able to talk about (imprecise) probability distributions. However, the typical starting point of a possibilistic analysis of some experiment or process, especially in engineering, is usually more pragmatic: Instead of introducing a measurable space (Ω, Σ) and defining a (possibility or probability) measure $\Pi, \mathrm{P} : \Sigma \to [0, 1]$ thereon, it is common and well-accepted to start by considering only

the imprecise variable $\tilde{V} : \Omega \to \mathbb{V}$ (or several variables) and provide its pushforwards $\Pi_{\tilde{V}}$ and $\mathrm{P}_{\tilde{V}}$ explicitly. Therefore, in the remainder of this thesis, the existence of the measurable space (Ω, Σ), of the corresponding possibility and probability measures Π and P thereon, and of the measurable space $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ is implicitly assumed and need not explicitly be mentioned, e.g., when stating $\tilde{V} \sim \pi_{\tilde{V}}$.

Below, some important possibility distributions and their connection to imprecise probability distributions are discussed. Depending on the available information about the true probability distribution, such possibility distributions are the basic building blocks of every possibilistic model. In this sense, this chapter is to be understood as a knowledge-based way of describing imprecisely known probability distributions, which contrasts the data-based construction of elementary possibility functions in Chapter 4. The fundamental tool to translate whatever knowledge is available into a possibility distribution is—again—the IP-Π-transform.

3.1.2.1 Vacuous Distribution

The *vacuous (possibility) distribution* $\mathcal{V}(\mathbb{V})$ corresponds to the pushforward of the vacuous possibility measure, which has already been introduced in Section 2.3.2.1. It is given by the vacuous elementary possibility function $\pi_{\tilde{V}}^{\mathrm{vac.}} \equiv 1$ on $\mathbb{V}$ and yields upper and lower probability bounds of $\Pi_{\tilde{V}}^{\mathrm{vac.}}(B) = 1$ and $\mathrm{N}_{\tilde{V}}^{\mathrm{vac.}}(B) = 0$ for all $B \in \mathbb{B}(\mathbb{V})$ except for $\Pi_{\tilde{V}}^{\mathrm{vac.}}(\emptyset) = 0$ and $\mathrm{N}_{\tilde{V}}^{\mathrm{vac.}}(\mathbb{V}) = 1$. In this sense, it does not provide any non-trivial insight, and corresponds to a complete lack of knowledge or *total ignorance* about the imprecise variable, and it is consistent with every probability distribution on $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$.

Proposition 26. *The vacuous possibility distribution $\mathcal{V}(\mathbb{V})$ is is an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}^{\mathrm{vac.}} = \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V}))$.*

Proof. The proposition follows directly from the trivial upper and lower probability bounds. □

Assuming that $\mathbb{V}$ is an exhaustive description of the possible values of the imprecise variable, it is most certainly always correct—albeit not at all expressive.

3.1.2.2 Quasi-Vacuous Distribution

As a slightly more expressive alternative to the vacuous possibility distribution, consider a set $T \in \mathbb{B}(\mathbb{V})$ to which the values of an imprecise variable are known to belong, usually an interval or higher-dimensional box, but no further indication about the actual probability distribution is available. The corresponding set of probability distributions

$$\mathfrak{P}_{\tilde{V}}^{\mathrm{quasi-vac.}} = \{\mathrm{P}_{\tilde{V}} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : \mathrm{P}_{\tilde{V}}(T) = 1\} \tag{3.12}$$

describes this *partial ignorance* by only restricting their supports to T.

The *quasi-vacuous (possibility) distribution* $\mathcal{Q}(T)$ possesses the $\{0,1\}$-valued elementary possibility function given by the indicator function $\mathcal{I}_T$ of T, i.e.

$$\pi_{\tilde{V}}^{\text{quasi-vac.}}(v) = \mathcal{I}_T(v) = \begin{cases} 1 & \text{if } v \in T \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{3.13}$$

for all $v \in \mathbb{V}$.

Proposition 27. *The quasi-vacuous possibility distribution $\mathcal{Q}(T)$ is an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}$.*

Proof. It is straight-forward to see that the optimal plausibilities of $\mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}$ are given by

$$\rho_{\tilde{V}}^{\text{quasi-vac.}}(v) = \sup_{\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}} \mathrm{P}_{\tilde{V}}(\{v\}) = \mathcal{I}_T(v) = \pi_{\tilde{V}}^{\text{quasi-vac.}}(v)$$

for all $v \in \mathbb{V}$. By Lemma 15, $\rho_{\tilde{V}}^{\text{quasi-vac.}}$ is an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}$ because

$$\mathrm{P}_{\tilde{V}}\left(\{\xi \in \mathbb{V} : \rho_{\tilde{V}}^{\text{quasi-vac.}}(\xi) \leq \rho_{\tilde{V}}^{\text{quasi-vac.}}(v)\}\right) = \begin{cases} \mathrm{P}_{\tilde{V}}(\mathbb{V}) = 1 & \text{if } v \in T \text{ and} \\ \mathrm{P}_{\tilde{V}}(\neg T) = 1 - \mathrm{P}_{\tilde{V}}(T) = 0 & \text{otherwise.} \end{cases}$$

In conclusion, it follows that

$$\sup_{\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}} \mathrm{P}_{\tilde{V}}(\{\xi \in \mathbb{V} : \rho_{\tilde{V}}^{\text{quasi-vac.}}(\xi) \leq \rho_{\tilde{V}}^{\text{quasi-vac.}}(v)\}) = \mathcal{I}_T(v)$$

for all $\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}$ and $v \in \mathbb{V}$. □

See Figure 3.1 for a visualization of a quasi-vacuous possibility distribution $\mathcal{Q}([a,b])$.

In essence, $\pi_{\tilde{V}}^{\text{quasi-vac.}}$ is simply the vacuous possibility distribution on the coarser measurable space $(T, \mathbb{B}(T))$. Its corresponding possibility and necessity measures coincide with the Boolean versions $\Pi_{\tilde{V}}^{\text{Bool.}} = \Pi_{\tilde{V}}^{\text{quasi-vac.}}$ and $\mathrm{N}_{\tilde{V}}^{\text{Bool.}} = \mathrm{N}_{\tilde{V}}^{\text{quasi-vac.}}$ presented in Section 2.1.2.1.

Another name for the quasi-vacuous possibility distribution is 'uniform possibility distribution', a name which due to possible confusion with the uniform probability distribution shall be avoided here; however, it is clear that the latter is consistent with the former, i.e., $\mathcal{U}(T) \preceq \mathcal{Q}(T)$, because $\mathrm{P}_{\tilde{V}}^{\text{unif.}}(T) = 1$, and therefore $\mathcal{U}(T) \in \mathfrak{P}_{\tilde{V}}^{\text{quasi-vac.}}$.

3.1.2.3 Triangular Distribution

The *triangular possibility distribution* $\Delta(a,b,c)$ is given by

$$\pi_{\tilde{V}}^{\text{tria.}}(v) = \begin{cases} 0 & \text{if } v < a, \\ \frac{v-a}{b-a} & \text{if } a \leq v < b, \\ \frac{c-v}{c-b} & \text{if } b \leq v < c \text{ and} \\ 0 & \text{if } v \geq c, \end{cases} \tag{3.14}$$

where $a \leq b \leq c$, see Figure 3.2. Its significance as an imprecise probability distribution is derived from the fact that all unimodal probability distributions with the mode located at b and a bounded support inside $[a, c]$ are consistent with it [DuboisEtAl04]. Moreover, the triangular possibility distribution has, owed to its piecewise linearity, established itself as a standard input distribution in many academic examples—similar to the uniform and the Gaussian probability distribution in probability theory. The particular geometric shape of this distribution function is sometimes also generalized to trapezoidal possibility distributions.

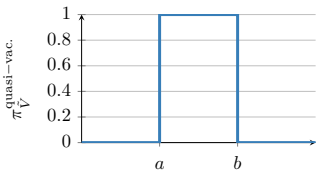


Figure 3.1: Quasi-Vacuous Distribution $\mathcal{Q}([a, b])$.

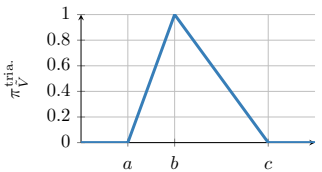


Figure 3.2: Triangular Distribution $\Delta(a, b, c)$.

Superuniform Distribution The *superuniform possibility distribution* of the $[0, 1]$-valued *superuniform imprecise variable* $\tilde{V}$ is given by $\pi_{\tilde{V}}(\alpha) = \alpha$ for all $\alpha \in [0, 1]$ and constitutes the special case of the triangular distribution, see Fig. 3.3. In this thesis, it is abbreviated as $\mathcal{A} = \Delta(0, 1, 1)$ or by simply saying that $\tilde{V}$ is superuniform. Reasons for the significance of the superuniform distribution will be provided at various points throughout the remainder of this thesis.

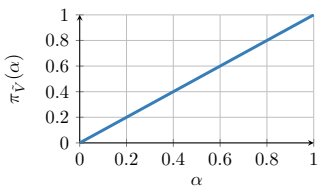


Figure 3.3: Superuniform Distribution $\mathcal{A}$.

3.1.3 Deterministic Distributions

A special class of possibility distributions includes only one unique probability distribution in its credal set. The *Deterministic Possibility Distribution* $\mathcal{D}(v^*) = \mathcal{Q}(\{v^*\})$ arises as a special case of the quasi-vacuous possibility distribution for $T = \{v^*\}$, where $v^* \in \mathbb{V}$.

It is the opposite of the vacuous possibility distribution in terms of expressiveness and corresponds to the *perfect knowledge* that $\tilde{V} \equiv v^*$ is actually precise, i.e. deterministic, and always assumes the value v^*. Its possibility distribution is given by

$$\pi_{\tilde{V}}^{\text{det.}}(v) = \begin{cases} 1 & \text{if } v = v^* \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

The upper and lower probabilities given by

$$\Pi_{\tilde{V}}^{\text{det.}}(B) = \mathrm{N}_{\tilde{V}}^{\text{det.}}(B) = \begin{cases} 1 & \text{if } v^* \in B \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

express that any event $B \in \mathbb{B}(\mathbb{V})$ that does not contain v^* has probability zero and can be disregarded. Similarly, any event including v^* certainly has probability one, and zero otherwise. Moreover, the upper and lower probabilities coincide and, therefore, also constitute a probability measure $\mathrm{P}_{\tilde{V}}^{\text{det.}} = \Pi_{\tilde{V}}^{\text{det.}} = \mathrm{N}_{\tilde{V}}^{\text{det.}}$, see Figure 3.4.

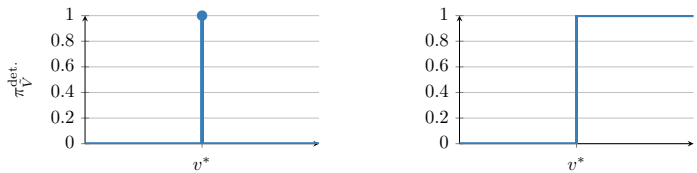


Figure 3.4: Deterministic distribution $\mathcal{D}(v^*)$ (left) and corresponding CPF, CΠF, and CNF (right).

Every possibility distribution $\pi_{\tilde{V}}$ with a non-empty core admits deterministic distributions for its most plausible values because, for all $c \in \text{core}(\pi_{\tilde{V}})$, it holds that $\mathcal{D}(c) \preceq \pi_{\tilde{V}}$. That is, imprecise variables may actually be deterministic variables $\tilde{V} \equiv c$; the vacuous distribution $\mathcal{V}(\mathbb{V})$ allowing this for all values in $\mathbb{V}$, and the quasi-vacuous possibility distribution $\mathcal{Q}(T)$ for all values in $T \in \mathbb{B}(\mathbb{V})$.

As a consequence, it is especially important to choose the core of a possibility distribution carefully. In fact, these considerations need already be taken into account when choosing the elementary plausibility distribution of $\tilde{V}$ before applying the IP-Π-transform since all values with maximum elementary plausibilities are guaranteed by their plausibility-conformity to also have unit elementary possibilities.

This peculiarity also justifies the term 'imprecise variable' for quantities described by a possibility distribution. The variable need not actually be random but can also be fixed/deterministic but unknown.

3.1.4 Probability-to-Possibility Transforms

If the true probability distribution is, indeed, known precisely, a variety of possibility distributions lends itself to encoding this knowledge via a Probability-to-Possibility Transform (P-Π-transform)—even though a precise description thereof can, as discussed above, usually not be achieved. In terms of the credal set, the question which possibility distribution to use depends on the other probability distributions one is willing to also admit under the resulting possibility distribution. It is worth noting that these probability-to-possibility transforms always loose much information and merely have theoretical rather than a practical relevance.

As a first step, it is useful to notice that every CPF $F_{\tilde{V}}$ of an imprecise variable $\tilde{V}$ can also act as a possibility distribution $\pi_{\tilde{V}}$. Both the normality and the measurability follow from its monotonicity. Most importantly, the possibility distribution defined by $\pi_{\tilde{V}} = F_{\tilde{V}}$ is consistent with the CPF $F_{\tilde{V}}$ itself because

$$\mathrm{P}_{\tilde{V}}\left(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}\right) = \mathrm{P}_{\tilde{V}}\left(v \in \mathbb{V} : F_{\tilde{V}}(v) \leq \alpha\right) = \mathrm{P}\left(F_{\tilde{V}}(\tilde{V}) \leq \alpha\right) = \alpha \tag{3.17}$$

for all $\alpha \in [0, 1]$. The last equality stems from the Probabilistic Universality of the Uniform to be discussed in Section 3.2.6. This is referred to as the *Cumulative* P-Π-*transform*. The same is true for choosing the complementary CPF as a possibility distribution $\pi_{\tilde{V}} = \bar{F}_{\tilde{V}}$, which is referred to as the *Complementary Cumulative* P-Π-*transform*.

3.1.4.1 The Optimal Probability-to-Possibility Transform

Alternatively, if $\mathrm{P}_{\tilde{V}}$ is induced by a *probability density function* $p_{\tilde{V}} : \mathbb{V} \to [0, +\infty)$ via

$$\mathrm{P}_{\tilde{V}}(B) = \int_B p_{\tilde{V}}(v)\,\mathrm{d}v \tag{3.18}$$

for $B \in \mathbb{B}(\mathbb{V})$, which is also written as $\tilde{V} \sim p_{\tilde{V}}$, Dubois et al. propose the *Optimal* P-Π-*transform* [DuboisEtAl04], wherein the probability density acts as the subjective plausibility $\rho_{\tilde{V}} = p_{\tilde{V}}$. This choice conforms best to the Principle of Plausibility since the probability density provides the likelihood, i.e. the relative plausibility, of each realization of $\tilde{V}$.

For instance, the well-known (multivariate) *Gaussian/normal probability distribution* $\mathcal{N}(\boldsymbol{m}, \boldsymbol{R})$ is given by the probability density

$$p^{\text{Gauss.}}_{\tilde{V}}(\boldsymbol{v}) = \frac{1}{\sqrt{2\pi\det(\boldsymbol{R})}} \exp\left(-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{m})^{\mathrm{T}}\boldsymbol{R}^{-1}(\boldsymbol{v}-\boldsymbol{m})\right) \tag{3.19}$$

for $\boldsymbol{v} \in \mathbb{R}^{D_{\mathbb{V}}}$, where $\boldsymbol{m} \in \mathbb{R}^{D_{\mathbb{V}}}$ is the mean vector and $\boldsymbol{R} \in \mathbb{R}^{D_{\mathbb{V}} \times D_{\mathbb{V}}}$ is the (positive definite) covariance matrix. See Figure 3.5 for a visualization of the possibility distributions resulting

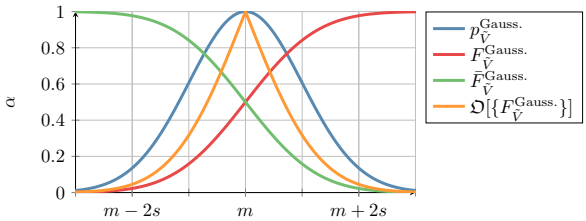


Figure 3.5: Possibilistic descriptions of a univariate Gaussian probability distribution with mean $m \in \mathbb{R}$ and variance $R = s^2 > 0$.

from the (Complementary) Cumulative and Optimal Transforms of the univariate Gaussian distribution.

Other choices of the subjective plausibility include symmetric plausibility distributions, and contour functions of hyper-boxes and -ellipsoids in higher dimensions, etc. [HoseHanss20, HoseHanss21c].

3.2 Extension

The *extension* of possibilistic descriptions of one imprecise variable to descriptions of a second imprecise variable via a connecting model is among the most basic techniques for reasoning with possibilities.

The overall goal is to solve those problems where the dependency between the $\mathbb{V}$-valued imprecise variable $\tilde{V}$ and the $\mathbb{Q}$-valued imprecise variable $\tilde{Q}$ is provided by the implicit, deterministic relationship

$$0 = \Xi\left(\tilde{V}(\omega), \tilde{Q}(\omega)\right) \tag{3.20}$$

for all $\omega \in \Omega$, which is simply written as $0 = \Xi(\tilde{V}, \tilde{Q})$, where $\Xi : \mathbb{V} \times \mathbb{Q} \to \mathbb{R}^D$ and 0 may be a vector of zeros. All that is assumed to be known is $\tilde{V} \sim \pi_{\tilde{V}}$, and the aim is to find a corresponding possibility distribution of $\tilde{Q}$. Generally, the solution to such problems is given by the implicit extension. But before considering this general problem, it is expedient to neglect Eq. (3.20) for a moment and consider the even simpler case of inferring the underlying possibility space (Ω, Σ, π) via the natural extension of $\pi_{\tilde{V}}$.

3.2.1 Natural Extension

In Section 3.1.2, it is assumed that the possibility distribution $\pi_{\tilde{V}}$ of the imprecise variable $\tilde{V}$ stems from an underlying possibility space consisting of the measurable space (Ω, Σ) and an elementary possibility function π thereon. This elementary possibility is generally

unknown, yet the key to solving the implicit extension problem above lies in inferring implicit possibilistic information about (Ω, Σ) via the natural extension of $\pi_{\tilde{V}}$ because, as Walley puts it,

> "the natural extension may be seen as the basic constructive step in statistical reasoning; it enables us to construct new [possibility distributions] from old ones." [Walley91, pp. 121–122]

In a first step, all admissible probability measures P on the underlying measurable space (Ω, Σ) that would yield a consistent probability distribution $\mathrm{P}_{\tilde{V}}$ of $\tilde{V}$ are gathered in the family of probabilities

$$\mathfrak{P}^{\mathrm{nat.}} = \{\mathrm{P} \in \mathbb{P}(\Omega, \Sigma) : \mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}\}. \tag{3.21}$$

This set may be viewed as the available information about (Ω, Σ), and it is to be described possibilistically. In particular, it seems unnatural to artificially exclude probability measures for the sake of expressiveness or regularity, which, nevertheless, is usual in Bayesian approaches that require a precise probabilistic specification of the available information.

The *natural extension* $\pi^{\mathrm{nat.}}$ of $\pi_{\tilde{V}}$ under $\tilde{V}$ is given by

$$\pi^{\mathrm{nat.}}(\omega) = \pi_{\tilde{V}}(\tilde{V}(\omega)) \tag{3.22}$$

for all $\omega \in \Omega$.

Proposition 28. *The natural extension $\pi^{\mathrm{nat.}}$ is an exact possibilistic description of $\mathfrak{P}^{\mathrm{nat.}}$.*

Proof. The equality $\mathfrak{P}^{\mathrm{nat.}} = \mathfrak{C}(\pi^{\mathrm{nat.}})$ is shown in two steps.
To see that $\mathfrak{P}^{\mathrm{nat.}} \subseteq \mathfrak{C}(\pi^{\mathrm{nat.}})$, let $\mathrm{P} \in \mathfrak{P}^{\mathrm{nat.}}$ and let $\alpha \in [0, 1]$. Then, the corresponding sublevel set $\mathcal{S}^{\alpha}_{\pi^{\mathrm{nat.}}} = \{\omega \in \Omega : \pi^{\mathrm{nat.}}_{\tilde{V}}(\tilde{V}(\omega)) \leq \alpha\}$ is a subset of $\tilde{V}^{-1}(\tilde{V}(\mathcal{S}^{\alpha}_{\pi^{\mathrm{nat.}}})) = \tilde{V}^{-1}(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}})$, and therefore $\mathrm{P}(\mathcal{S}^{\alpha}_{\pi^{\mathrm{nat.}}}) \leq \mathrm{P}(\tilde{V}^{-1}(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}})) = \mathrm{P}_{\tilde{V}}(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}) \leq \alpha$. The last inequality follows from the consistency in the definition of $\mathfrak{P}^{\mathrm{nat.}}$, of which P is an element, and it is concluded that $\mathrm{P} \in \mathfrak{C}(\pi^{\mathrm{nat.}})$.
Conversely, let $\mathrm{P} \preceq \pi^{\mathrm{nat.}}$ and let $\alpha \in [0, 1]$. It follows immediately that

$$\mathrm{P}_{\tilde{V}}\left(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}\right) = \mathrm{P}\left(\tilde{V}^{-1}\left(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}\right)\right) \leq \Pi^{\mathrm{nat.}}\left(\tilde{V}^{-1}\left(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}\right)\right) = \sup_{\omega \in \Omega\,:\,\tilde{V}(\omega) \in \mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}} \pi^{\mathrm{nat.}}(\omega) \leq \alpha.$$

That is, $\mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}$, and therefore $\mathfrak{C}(\pi^{\mathrm{nat.}}) \subseteq \mathfrak{P}^{\mathrm{nat.}}$ proving the proposition. □

The natural extension primarily serves a theoretical purpose in the sense that it allows one to extend the possibilistic description from a coarser measurable space $(\mathbb{V}, \mathbb{B}(\mathbb{V}))$ onto the finer space (Ω, Σ) without loss of information. With this preliminary result, it is now straightforward, to solve the implicit extension problem presented above.

3.2.2 Implicit Extension

The problem of extending the possibility distribution $\pi_{\tilde{V}}$ of the imprecise variable $\tilde{V}$ under the model in Eq. (3.20) onto the imprecise variable $\tilde{Q}$ may be solved in subsequent steps.

First, it is convenient to notice that Eq. (3.20) is basically only a restriction of the admissible values of $\omega \in \Omega$ which may be gathered in the feasible set

$$T = \left\{\omega \in \Omega : 0 = \Xi\left(\tilde{V}\left(\omega\right), \tilde{Q}\left(\omega\right)\right)\right\}. \tag{3.23}$$

In particular, this does not imply any constraints on the corresponding *implicit extension* $\pi^{\text{imp.}}$ on the underlying measurable space (Ω, Σ) other than that, apart from being confined to $\mathfrak{P}^{\text{nat.}}$ given by the natural extension $\pi^{\text{nat.}}$ of $\pi_{\tilde{V}}$, the admissible probabilities must also belong to the family of probabilities $\mathfrak{P}^{\text{feas.}} = \{\mathrm{P} \in \mathbb{P}(\Omega, \Sigma) : \mathrm{P}(T) = 1\}$. All probability mass must additionally be contained inside T, which is exactly described by the quasi-vacuous possibility distribution $\mathcal{Q}(T)$ with the elementary possibility function $\mathcal{I}_T$. In summary, the imprecise probabilistic information is gathered in the set

$$\mathfrak{P}^{\text{imp.}} = \mathfrak{P}^{\text{nat.}} \cap \mathfrak{P}^{\text{feas.}}. \tag{3.24}$$

The *implicit extension* $\pi^{\text{imp.}}$ of $\pi_{\tilde{V}}$ under $\tilde{V}$ and Ξ is defined via

$$\pi^{\text{imp.}}(\omega) = \min\left(\mathcal{I}_T(\omega), \pi^{\text{nat.}}(\omega)\right) = \begin{cases} \pi_{\tilde{V}}\left(\tilde{V}(\omega)\right) & \text{if } \omega \in T \text{ and} \\ 0 & \text{otherwise} \end{cases} \tag{3.25}$$

for all $\omega \in \Omega$.

Proposition 29. *The implicit extension $\pi^{\text{imp.}}$ of $\pi_{\tilde{V}}$ under $\tilde{V}$ and Ξ is an exact possibilistic description of $\mathfrak{P}^{\text{imp.}}$.*

Proof. By Propositions 27 and 28, $\mathfrak{P}^{\text{nat.}}$ and $\mathfrak{P}^{\text{feas.}}$ are exactly described by the natural extension of $\pi_{\tilde{V}}$ and the quasi-vacuous possibility distribution $\mathcal{Q}(T)$, respectively, and one may apply Lemma 20 for their conjunction. □

Second, the possibility distribution of $\tilde{Q}$, which, according to Proposition 25, exactly describes the set of admissible pushforwards under $\pi^{\text{imp.}}$, is computed from Eq. (3.8) via

$$\pi^{\text{imp.}}_{\tilde{Q}}(q) = \sup_{\omega \in \Omega\,:\,\tilde{Q}(\omega) = q} \pi^{\text{imp.}}(\omega) \tag{3.26}$$

for all $q \in \mathbb{Q}$. Given that $\pi^{\text{imp.}}(\omega)$ is only greater than zero if $\omega \in T$, this coincides with the more accessible formula

$$\pi^{\text{imp.}}_{\tilde{Q}}(q) = \sup_{\omega \in \Omega\,:\,0 = \Xi(\tilde{V}(\omega), q)} \pi_{\tilde{V}}\left(\tilde{V}(\omega)\right) = \sup_{v \in \mathbb{V}\,:\,0 = \Xi(v,q)} \pi_{\tilde{V}}(v), \tag{3.27}$$

which is referred to as the *(Implicit Possibilistic) Extension Principle*.

Proposition 30. *The implicit extension* $\pi_{\tilde{Q}}^{\text{imp.}}$ *of* $\pi_{\tilde{V}}$ *under* Ξ *is an exact possibilistic description of the set of all pushforwards of* $\tilde{Q}$ *that would be admissible under* $\pi_{\tilde{V}}$ *and* Ξ*, i.e. of* $\mathfrak{P}_{\tilde{Q}}^{\text{imp.}} = \{\mathrm{P}_{\tilde{Q}} : \mathrm{P} \in \mathfrak{P}^{\text{imp.}}\}$.

Proof. By Proposition 29, $\pi^{\text{imp.}}$ is an exact possibilistic description of $\mathfrak{P}^{\text{imp.}}$, and, by Proposition 25, $\pi_{\tilde{Q}}^{\text{imp.}}$ re-expresses this knowledge in the form of a possibility distribution of $\tilde{Q}$. □

To summarize the previous proposition, the implicit extension does not add or subtract information. It corresponds to the actual state of knowledge about the unknown probability distribution of $\tilde{Q}$, and, without introducing further assumptions, it is impossible to arrive at more expressive descriptions.

Equation (3.20) poses some potentially severe restrictions on all $\mathrm{P} \in \mathfrak{P}^{\text{imp.}}$ and special attention must be paid to this fact since Lemma 20 requires $\mathfrak{P}^{\text{imp.}}$ to be non-empty. Fortunately, this can be checked via the normality of $\pi_{\tilde{Q}}^{\text{imp.}}$. If $\pi_{\tilde{Q}}^{\text{imp.}}$ is subnormal, then $\pi^{\text{imp.}}$ must also be subnormal indicating that $\mathfrak{P}^{\text{imp.}}$ was empty to begin with. More practically expressed, for all $\alpha \in [0, 1)$, there must be at least one combination $(v, q) \in \mathbb{V} \times \mathbb{Q}$ such that $\pi_{\tilde{V}}(v) > \alpha$ and $\Xi(v, q) = 0$. In the simplest case, it suffices to find one such admissible combination (v_c, q_c) with $\Xi(v_c, q_c) = 0$ and maximum possibility $\pi_{\tilde{V}}(v_c) = 1$.

3.2.3 Explicit Extension

If $\Xi = \tilde{Q} - \phi(\tilde{V}) = 0$ describes an explicit dependence of $\tilde{Q}(\omega)$ on $\tilde{V}(\omega)$, written as

$$\tilde{Q} = \phi(\tilde{V}), \tag{3.28}$$

which is to be understood as the concatenation $\tilde{Q} = \phi \circ \tilde{V}$, then the implicit extension provides a general rule for propagating imprecise variables.

More generally, the pushforward $\mathrm{M}_{\tilde{Q}}$ of any capacity M under $\tilde{Q} = \phi(\tilde{V})$ can be computed directly from a given pushforward $\mathrm{M}_{\tilde{V}}$ via

$$\mathrm{M}_{\tilde{Q}}(B) = \mathrm{M}\left(\tilde{Q} \in B\right) = \mathrm{M}\left(\phi(\tilde{V}) \in B\right) = \mathrm{M}\left(\tilde{V} \in \phi^{-1}(B)\right) = \mathrm{M}_{\tilde{V}}\left(\phi^{-1}(B)\right) \tag{3.29}$$

for all $B \in \mathbb{B}(\mathbb{Q})$.

3.2.3.1 Possibility Propagation

The *explicit extension* $\pi_{\tilde{Q}}^{\text{exp.}}$ of the input distribution $\pi_{\tilde{V}}$ under ϕ arises as a special case of the implicit extension $\pi_{\tilde{Q}}^{\text{IE}}$ for this specific choice of Ξ in Eq. (3.27) and is given by

$$\pi_{\tilde{Q}}^{\text{exp.}}(q) = \sup_{v \in \mathbb{V} \,:\, q = \phi(v)} \pi_{\tilde{V}}(v) \tag{3.30}$$

for $q \in \mathbb{Q}$. This formulation not only constitutes a general variant of the Fuzzy Extension Principle [Zadeh75a, Hanss05] but also induces a possibility measure which coincides with the pushforward possibility measure $\Pi_{\tilde{Q}}$ given by Eq. (3.29). That is,

$$\Pi_{\tilde{Q}}(V) = \Pi_{\tilde{V}}\left(\phi^{-1}(V)\right) = \sup_{v \in \phi^{-1}(V)} \pi_{\tilde{V}}(v) = \sup_{q \in V} \sup_{v \in \mathbb{V}\,:\, q = \phi(v)} \pi_{\tilde{V}}(v) = \sup_{q \in V} \pi_{\tilde{Q}}^{\text{exp.}}(q) \tag{3.31}$$

for $V \in \mathbb{B}(\mathbb{Q})$.

Probability Propagation If $\mathrm{M}_{\tilde{V}} = \mathrm{P}_{\tilde{V}}$ is a probability measure, Eq. (3.29) reduces to the standard rule for propagating random variables [Sullivan15] stating that the probability distribution $\mathrm{P}_{\tilde{Q}} = \mathrm{P}_{\phi(\tilde{V})}$ is given by

$$\mathrm{P}_{\tilde{Q}}(B) = \mathrm{P}_{\tilde{V}}\left(\phi^{-1}(B)\right) \tag{3.32}$$

for all $B \in \mathbb{B}(\mathbb{Q})$, or, if ϕ is strictly increasing, that the CPF of $\tilde{Q}$ at $q \in \mathbb{Q}$ is given by

$$F_{\tilde{Q}}(q) = F_{\tilde{V}}\left((-\infty, \phi^{-1}(q)\right). \tag{3.33}$$

A possibilistic description of $\tilde{Q}$ should, then, be consistent with all probability distributions $\mathrm{P}_{\phi(\tilde{V})}$ that stem from a consistent probability distribution $\mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}$ under ϕ. These are gathered in the set

$$\mathfrak{P}_{\tilde{Q}}^{\text{exp.}} = \{\mathrm{P}_{\phi(\tilde{V})} : \mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}\}. \tag{3.34}$$

The well-known result [BronevichKlir10, BaudritCousoDubois07, HoseHanss19c] that the corresponding possibility distribution $\pi_{\tilde{Q}}^{\text{exp.}}$ as defined in Eq. (3.30) is also consistent with all $\mathrm{P}_{\tilde{Q}} \in \mathfrak{P}_{\tilde{Q}}^{\text{exp.}}$ is a less general version of Proposition 30, which implies that $\pi_{\tilde{Q}}^{\text{exp.}}$ is an exact possibilistic description of $\mathfrak{P}_{\tilde{Q}}^{\text{exp.}}$.

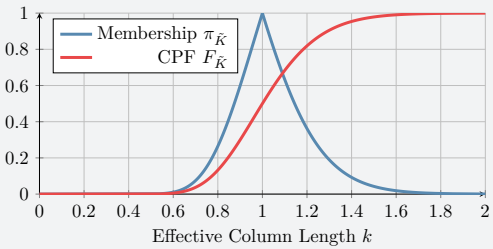
The explicit extension can, e.g., be used for risk propagation as the following example shows.

Example 5: Forward Extension

A wooden rod with length $L = 1\,\mathrm{m}$, elastic modulus $E = 11\,\mathrm{GPa}$, and the area moment of inertia $I = \frac{\pi r^2}{2}$ where the radius is $r = 0.01\,\mathrm{m}$, is subjected to the load F. For negative load margins

$$S = F - \frac{\pi^2 EI}{K^2 L^2} < 0,$$

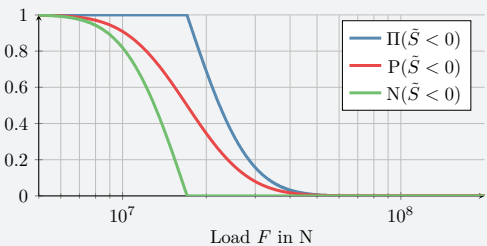
the structure is considered safe, i.e., the rod will not buckle.[32] Due to, e.g., imperfect boundary conditions, the column effective length factor K may not be determined exactly; instead, it is modeled as an imprecise input variable $\tilde{K}$ with the possibility distribution $\pi_{\tilde{K}}$ shown below. Amongst others, this possibility distribution is consistent with the probability distribution $F_{\tilde{K}}$, which is also shown.



The values of the possibility distribution $\pi_{\tilde{S}}$ of the output variable $\tilde{S} = \tilde{S}(\tilde{K})$ can then be computed by the explicit extension

$$\pi_{\tilde{S}}(s) = \max\left(\underbrace{\pi_{\tilde{K}}\left(-\sqrt{\frac{F-s}{EI}}\frac{L}{\pi}\right)}_{=0}, \underbrace{\pi_{\tilde{K}}\left(+\sqrt{\frac{F-s}{EI}}\frac{L}{\pi}\right)}_{\geq 0}\right) = \pi_{\tilde{K}}\left(\sqrt{\frac{F-s}{EI}}\frac{L}{\pi}\right)$$

for $s > 0$. Depending on the magnitude of the load F, the probability of stability is bounded from below and above by the necessity and possibility of $\tilde{S} < 0$, respectively, as depicted below. For reference, the probabilities obtained from the propagation of the (consistent) probability distribution are also shown.



In the remainder of this thesis, the explicit extension of $\pi_{\tilde{V}}$ under ϕ is often simply written as $\pi_{\phi(\tilde{V})}$.

3.2.3.2 Marginalization

A second special case of the explicit extension is *marginalization*. If $\mathrm{M}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ is a joint pushforward measure of the imprecise variables $\tilde{V}_1, \ldots, \tilde{V}_m$, then the marginal pushfor-

[32] Both in the expression for I and S, Archimedes' constant π is not to be confused with an elementary possibility function.

ward measure $\mathrm{M}_{\tilde{V}_i}$ of $\tilde{V}_i$ for $i = 1, \ldots, m$ is computed via the explicit extension under the function $\phi : \mathbb{V}_1 \times \ldots \times \mathbb{V}_m \to \mathbb{V}_i$ returning the i-th element $\phi(v_1, \ldots, v_m) = v_i$ for $(v_1, \ldots, v_m) \in \mathbb{V}_1 \times \ldots \times \mathbb{V}_m$, and therefore

$$\mathrm{M}_{\tilde{V}_i}(B_i) = \mathrm{M}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\mathbb{V}_1 \times \ldots \times \mathbb{V}_{i-1} \times B_i \times \mathbb{V}_{i+1} \times \ldots \times \mathbb{V}_m) \tag{3.35}$$

for all $B_i \in \mathbb{B}(\mathbb{V}_i)$.

In particular, if $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ is a joint possibility distribution, then the marginal distribution is obtained via

$$\pi_{\tilde{V}_i}(v_i) = \sup_{(v_1,\ldots,v_{i-1},v_{i+1},\ldots v_m) \in \mathbb{V}_1 \times \ldots \mathbb{V}_{i-1} \times \mathbb{V}_{i+1} \times \ldots \times \mathbb{V}_m} \pi_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1, \ldots, v_m) \tag{3.36}$$

for all $v_i \in \mathbb{V}_i$. Figuratively, this corresponds to the shadow that $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ casts onto the i-th coordinate axis. The reverse question of how to construct joint distributions from marginals shall be addressed in Section 3.5.

3.2.4 Inverse Extension

The implicit extension also covers the case of inverse dependencies, i.e., $\Xi = \tilde{V} - \psi(\tilde{Q}) = 0$, written as

$$\tilde{Q} = \psi^{-1}(\tilde{V}). \tag{3.37}$$

This is to be understood in the sense that $\tilde{V}$ is the output of some model $\psi : \mathbb{Q} \to \mathbb{V}$, and the corresponding implicit extension, which then reduces to the *inverse extension* of $\pi_{\tilde{V}}$ under ψ, seeks to provide a suitable input possibility distribution of $\tilde{Q}$.

Of course, a possibilistic description of $\tilde{Q}$ ought to be consistent with all admissible probability distributions $\mathrm{P}_{\tilde{Q}}$ that yield a consistent pushforward $\mathrm{P}_{\tilde{V}} = \mathrm{P}_{\psi(\tilde{Q})}$ under ψ. These are gathered in the set

$$\mathfrak{P}^{\text{inv.}}_{\tilde{Q}} = \{\mathrm{P}_{\tilde{Q}} \in \mathbb{P}(\mathbb{Q}, \mathbb{B}(\mathbb{Q})) : \mathrm{P}_{\psi(\tilde{Q})} \preceq \pi_{\tilde{V}}\}. \tag{3.38}$$

By Eq. (3.27), the corresponding implicit extension of $\pi^{\text{inv.}}_{\tilde{Q}}$ is given by

$$\pi^{\text{inv.}}_{\tilde{Q}}(q) = \pi_{\tilde{V}}\left(\psi(q)\right) \tag{3.39}$$

for all $q \in \mathbb{Q}$. This formulation has first been observed as a solution to inverse fuzzy arithmetic [HoseHanss18a], i.e. as an inverse operation to the Fuzzy Extension Principle, and a previous result [HoseHanss19a] states that (the credal set of) $\pi^{\text{inv.}}_{\tilde{Q}}$ provides a maximally specific outer approximation of $\mathfrak{P}^{\text{inv.}}_{\tilde{Q}}$. Again, Proposition 30 generalizes this by showing that $\pi^{\text{inv.}}_{\tilde{Q}}$ is an exact possibilistic description of $\mathfrak{P}^{\text{inv.}}_{\tilde{Q}}$.

The likeness between the formulations in Eq. (3.22) and Eq. (3.39) indicates that the inverse extension is closely connected to the natural extension. Both project the information that

is available on a smaller σ-algebra $\Sigma|_{\tilde{V}}$ onto the larger σ-algebras $\Sigma|_{\tilde{Q}}$ and Σ, respectively, where

$$\Sigma|_{\tilde{V}} \subseteq \Sigma|_{\tilde{Q}} \subseteq \Sigma. \tag{3.40}$$

This projection follows the same rules, which is, e.g., observed by considering that the natural extension of $\pi_{\tilde{V}}$ can, in this case, be computed directly from $\pi_{\tilde{V}}$ and $\tilde{V}$, or, equivalently, from $\pi_{\tilde{Q}}^{\text{inv.}}$ and $\tilde{Q}$ because for all $\omega \in \Omega$

$$\pi^{\text{nat.}}(\omega) = \pi_{\tilde{V}}\left(\tilde{V}(\omega)\right) = \pi_{\tilde{V}}\left(\psi(\tilde{Q}(\omega))\right) = \pi_{\tilde{Q}}^{\text{inv.}}\left(\tilde{Q}(\omega)\right). \tag{3.41}$$

This result can, e.g., be used in order to find possibility distributions of imprecise variables with partially specified quantiles as demonstrated in the following example. In particular, it eliminates the need for Maximum Entropy-based approaches [De MartinoDe Martino18], which are employed in various disciplines, e.g. in statistical mechanics [Jaynes57] where it is necessary to infer an input (a-priori) distribution of an unknown quantity, e.g. the position or velocity distribution of a particle swarm, from the a-posteriori distribution of a measurable output quantity, e.g. the energy. These inverse problems generally possess an infinite number of solutions, and regularizing assumptions, such as maximum entropy of the input distribution, are needed in order to obtain a unique solution $\mathrm{P}_{\tilde{Q}} \in \mathfrak{P}_{\tilde{Q}}^{\text{inv.}}$, which—at least, when considering that imprecise probabilistic techniques are readily available—seems rather questionable.

Example 6: Inverse Extension

Reconsider Example 5. A company produces wooden rods with length $L = 1\,\text{m}$, and the area moment of inertia $I = \frac{\pi r^2}{2}$ where the radius is $r = 0.01\,\text{m}$. The elastic modulus E of the produced rods is assumed to vary according to an unknown probability distribution $\mathrm{P}_{\tilde{E}}$. In order to identify faulty pieces, the last step of the manufacturing process involves a quality inspection where the rod is subjected to a load $F = 1\,\text{kN}$ in a controlled experiment with a known effective length factor $K = 1$. It is observed that less than 5% of the tested rods do not pass this test, i.e., they buckle, which is well-expressed by the possibility distribution

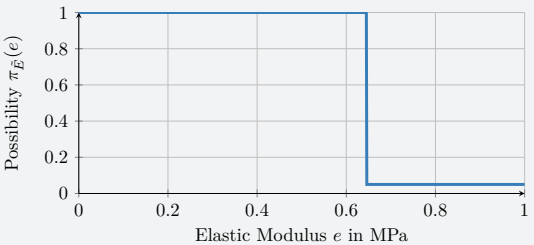
$$\pi_{\tilde{T}}(0) = 0.05 \qquad \text{and} \qquad \pi_{\tilde{T}}(1) = 1,$$

where $\tilde{T} = \psi(\tilde{E})$ and

$$\psi(e) = \begin{cases} 1 & \text{if } F - \frac{\pi^2 I}{K^2 L^2} e < 0 \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

That is, the variable $\tilde{T}$ is one if the rod does not buckle and zero otherwise. It is straightforward to show that $\pi_{\tilde{T}}$ is consistent with all probability distributions of $\mathrm{P}_{\tilde{T}}$ of $\tilde{T}$ that conform to the empirical observation $\mathrm{P}_{\tilde{T}}(0) \leq 0.05$.

This imprecise model allows to infer a possibility distribution $\pi_{\tilde{E}}$ of the imprecise elastic modulus $\tilde{E}$ via the inverse extension of $\pi_{\tilde{T}}$ under ψ. This possibility distribution is shown below. This distribution only states that the 95% quantile of $\tilde{E}$ is located at 0.645 GPa, and nothing more.



If the empirical observation is, indeed, correct, this possibility distribution is guaranteed to be consistent with the true probability distribution $\mathrm{P}_{\tilde{E}}$.

In the remainder of this thesis, the inverse extension of $\pi_{\tilde{V}}$ under ψ is usually written as $\pi_{\psi^{-1}(\tilde{V})}$.

3.2.4.1 Vacuous Extension

Contrary to marginalization, the *vacuous extension* formalizes the knowledge about the imprecise variables $\tilde{V}_1, \ldots, \tilde{V}_m$ if only the marginal distribution of $\tilde{V}_i$ for some $i = 1, \ldots, m$ is known. Clearly, nothing can be said about the joint distribution, except that its i-th marginal is $\pi_{\tilde{V}_i}$. This is expressed by the inverse extension under $\psi : \mathbb{V}_1 \times \ldots \times \mathbb{V}_m \to \mathbb{V}_i$ returning the i-th element $\psi(v_1, \ldots, v_m) = v_i$ for $(v_1, \ldots, v_m) \in \mathbb{V}_1 \times \ldots \times \mathbb{V}_m$, which, then, yields

$$\pi_{\tilde{Q}_1,\ldots,\tilde{Q}_m}(v_1, \ldots, v_m) = \pi_{\tilde{Q}_i}(v_i). \tag{3.42}$$

3.2.5 Inverse Explicit Extension and Explicit Inverse Extension

Finally, one may ask whether a repeated application of the implicit extension is able to recover the original distribution. For instance, an earlier discussion [HoseHanss18a] argues that the inverse extension is, indeed, inverse to the explicit extension, and vice versa, if ϕ and ψ possess certain properties.

Proposition 31. *Given an injective function $\phi : \mathbb{V} \to \mathbb{Q}$ and an input possibility distribution $\pi_{\tilde{V}}$, the inverse extension of the explicit extension of $\pi_{\tilde{V}}$ (both under ϕ) retrieves $\pi_{\tilde{V}}$.*

Proof. Let $v \in \mathbb{V}$. Then, $\pi_{\phi^{-1}(\phi(\tilde{V}))}(v) = \pi_{\phi(\tilde{V})}(\phi(v)) = \sup_{\xi \in \mathbb{V} : \phi(\xi) = \phi(v)} \pi_{\tilde{V}}(\xi) = \pi_{\tilde{V}}(v)$.

The last equality obtains since the injectivity of ϕ implies that $\phi(\xi) = \phi(v)$ is equivalent to $\xi = v$. □

Proposition 32. *Given a surjective function $\psi : \mathbb{Q} \to \mathbb{V}$ and an output possibility distribution $\pi_{\tilde{V}}$, the explicit extension of the inverse extension of $\pi_{\tilde{V}}$ (both under ψ) retrieves $\pi_{\tilde{V}}$.*

Proof. Let $v \in \mathbb{V}$. Then,

$$\pi_{\psi(\psi^{-1}(\tilde{V}))}(v) = \sup_{q \in \mathbb{Q} \,:\, \psi(q) = v} \pi_{\psi^{-1}(\tilde{Q})}(q) = \sup_{q \in \mathbb{Q} \,:\, \psi(q) = v} \pi_{\tilde{V}}(\psi(q)) = \pi_{\tilde{V}}(v).$$

The last equality obtains since surjectivity implies that for all $v \in \mathbb{V}$ there exists at least one $q \in \mathbb{Q}$ such that $\psi(q) = v$. □

Especially the surjectivity of ψ is a rather technical assumption. In essence, it requires one to verify that all of the support of $\pi_{\tilde{V}}$ may be reached under ψ and translates to an actual connection of $\tilde{V}$ and $\tilde{Q}$ under ψ. The adequacy of this requirement in a model should be self-evident.

Trivially, if ϕ is a bijective function with inverse ψ, both propositions apply. It is, furthermore, possible to generalize these observations in a general proposition about the involutiveness of the implicit extension. However, the necessary assumptions are quite restrictive, and individual considerations as discussed above seem favorable.

3.2.6 Elementary Extensions

By the definition of an imprecise variable, an elementary possibility function $\pi : \Omega \to [0, 1]$ is also a $[0, 1]$-valued imprecise variable $\tilde{\pi}$. This observation, in connection with the explicit and implicit extension, can be used to derive possibilistic analogues of two fundamental probabilistic concepts. These results constitute much of the theoretical significance of the superuniform distribution.

Notice that the probabilistic results they are inspired by are usually only applicable to real-valued imprecise variables, i.e. they are restricted to $\mathbb{V} \subseteq \mathbb{R}$. However, the possibilistic variants hold for general $\mathbb{V}$.

3.2.6.1 Possibilistic Universality of the Superuniform

Before discussing the Possibilistic Universality of the Superuniform, the corresponding concept from probability theory, the Probabilistic Universality of the Uniform, shall be explained briefly.

Probabilistic Universality of the Uniform Consider a continuous $\mathbb{V}$-valued imprecise variable $\tilde{V}$ that has CPF $F_{\tilde{V}}$. Additionally, assume that $F_{\tilde{V}}$ is invertible; however, this may also be shown for more general CPFs. Then, one may define the imprecise variable $\tilde{Q} = F_{\tilde{V}}(\tilde{V})$. The probability distribution $\mathrm{P}_{\tilde{Q}} = \mathrm{P}_{F_{\tilde{V}}(\tilde{V})}$ can be computed from Eq. (3.32), and also from Eq. (3.33) because $F_{\tilde{V}}$ is strictly increasing. In particular, it follows for all $q \in [0,1]$ that

$$F_{\tilde{Q}}(q) = \mathrm{P}\left(F_{\tilde{V}}(\tilde{V}) \leq q\right) = \mathrm{P}\left(\tilde{V} \leq F_{\tilde{V}}^{-1}(q)\right) = F_{\tilde{V}}\left(F_{\tilde{V}}^{-1}(q)\right) = q. \tag{3.43}$$

That is, $\tilde{Q}$ has a standard uniform probability distribution $\mathcal{U}([0,1])$. This result is known as the *Probability Integral Transform* [Angus94].

Similarly, one may define $\pi_{\tilde{V}}(\tilde{V})$ where $\tilde{V}$ is any (not necessarily continuous) $\mathbb{V}$-valued imprecise variable with possibility distribution $\pi_{\tilde{V}}$. Then, the explicit extension of $\pi_{\tilde{V}}$ (under itself) yields

$$\pi_{\pi_{\tilde{V}}(\tilde{V})}(\alpha) = \sup_{v \in \mathbb{V}\,:\,\pi_{\tilde{V}}(v)=\alpha} \pi_{\tilde{V}}(v) = \begin{cases} \alpha & \text{if } \exists v \in \mathbb{V} : \pi_{\tilde{V}}(v) = \alpha \\ 0 & \text{otherwise} \end{cases} \tag{3.44}$$

for all $s \in [0,1]$. If $\pi_{\tilde{V}}$ is surjective on $[0,1]$, equality holds everywhere, but $\pi_{\pi_{\tilde{V}}(\tilde{V})}(\alpha) \leq \alpha$ is always true for all $\alpha \in [0,1]$. Therefore, $\pi_{\pi_{\tilde{V}}(\tilde{V})}$ is more specific than the superuniform distribution $\mathcal{A}$, i.e.

$$\pi_{\pi_{\tilde{V}}(\tilde{V})} \preceq \mathcal{A}. \tag{3.45}$$

Furthermore, the properties of the explicit extension guarantee that $\pi_{\pi_{\tilde{V}}(\tilde{V})}$ is an exact possibilistic description of all admissible pushforward probability distributions $\mathrm{P}_{\pi_{\tilde{V}}(\tilde{V})}$ of $\pi_{\tilde{V}}(\tilde{V})$. In particular,

$$\mathrm{P}_{\pi_{\tilde{V}}(\tilde{V})} \preceq \pi_{\pi_{\tilde{V}}(\tilde{V})} \tag{3.46}$$

for all $\mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}$. Combining these two observations yields the *Possibilistic Universality of the Superuniform*, i.e. for all $\mathrm{P}_{\tilde{V}} \preceq \pi_{\tilde{V}}$

$$\mathrm{P}_{\pi_{\tilde{V}}(\tilde{V})} \preceq \mathcal{A}. \tag{3.47}$$

In other words, the consistency criterion provided by Eq. (2.30) reads

$$\mathrm{P}\left(\pi_{\tilde{V}}(\tilde{V}) \in \mathcal{S}_{\mathcal{A}}^{\alpha}\right) = \mathrm{P}\left(\pi_{\tilde{V}}(\tilde{V}) \leq \alpha\right) \leq \alpha, \tag{3.48}$$

which, considering that the sublevel sets of the superuniform distribution $\mathcal{A}$ for $\alpha \in [0,1]$ are given by $\mathcal{S}_{\mathcal{A}}^{\alpha} = [0,\alpha]$, reduces to the equivalent consistency criterion

$$F_{\pi_{\tilde{V}}(\tilde{V})}(\alpha) \leq \alpha \tag{3.49}$$

for all $\alpha \in [0,1]$. The probability distribution of the possibility distribution of $\tilde{V}$ must be superuniform: It must stochastically dominate the uniform probability distribution.[33]

[33] The concept of stochastic dominance is discussed in Section 3.3.

3.2.6.2 Inverse Possibility Transform

Before discussing the Inverse Possibility Transform, the analog concept from probability theory, the Inverse Probability Integral Transform, must also be explained.

Inverse Probability Integral Transform For a given CPF $F_{\tilde{Q}}$ and a standard uniform variable $\tilde{V} \sim \mathcal{U}([0,1])$, the imprecise variable $\tilde{Q} = F_{\tilde{Q}}^{-1}(\tilde{V})$ has the CPF $F_{\tilde{Q}}$, which follows by an argument similar to Eq. (3.43). This is known as the *Inverse Probability (Integral) Transform* or, more commonly, as *Inverse Transform Sampling*, and it is widely employed in Monte-Carlo simulations [Sullivan15] or Polynomial Chaos Expansions [Sudret15]. Furthermore, it allows casting every model into a standardized form with uniform input variables.

Similarly, for a given elementary possibility function $\pi_{\tilde{Q}}$ and a superuniform variable $\tilde{V} \sim \mathcal{A}$, the imprecise variable $\tilde{Q}$ defined via the inverse relationship $\tilde{Q} = \pi_{\tilde{Q}}^{-1}(\tilde{V})$ has the possibility distribution $\pi_{\tilde{Q}}$ because for all $q \in \mathbb{Q}$ the inverse extension of $\mathcal{A}$ under $\pi_{\tilde{Q}}$ is given by

$$\pi_{\pi_{\tilde{Q}}^{-1}(\tilde{V})}(q) = \pi_{\tilde{V}}\left(\pi_{\tilde{Q}}(q)\right) = \pi_{\tilde{Q}}(q). \tag{3.50}$$

Similar to the Inverse Probability Integral Transform, this allows to cast every imprecise model into a standardized form with superuniform input variables only. For instance, the implicit model in Eq. (3.20), may be re-written in standard form as

$$\Xi\left(\pi_{\tilde{V}}^{-1}(\tilde{A}), \tilde{Q}\right) = 0, \tag{3.51}$$

which depends only on the superuniform input variable $\tilde{A} \sim \mathcal{A}$.

3.3 Stochastic Dominance

Continuous possibility distributions on $\mathbb{V} \subseteq \mathbb{R}$—to which the discussion in this section and its subsections is, again, restricted—are intricately linked to the concept of *stochastic dominance* [Denœux09], which is usually introduced as a partial order between random variables. According to this definition, a random variable $\tilde{V}_1$ stochastically dominates the imprecise variable $\tilde{V}_2$ if the CPF of the former is lower than or equal to the latter, i.e. if

$$F_{\tilde{V}_1}(v) \leq F_{\tilde{V}_2}(v) \tag{3.52}$$

for all $v \in \mathbb{V}$. The direction of the inequality implies that $\tilde{V}_1$ has a higher probability of being high compared to $\tilde{V}_2$. Conversely, $\tilde{V}_2$ has a higher probability of being low compared to $\tilde{V}_1$, which justifies this name.

In this thesis, a partial order among CPFs—and not between imprecise variables—is more convenient. More precisely, if two admissible CPFs $F_{\tilde{V}}^{(1)}, F_{\tilde{V}}^{(2)}$ of an imprecise

variable $\tilde{V}$ (with unknown true probability distribution) fulfill

$$F_{\tilde{V}}^{(1)}(v) \leq F_{\tilde{V}}^{(2)}(v) \tag{3.53}$$

for all $v \in \mathbb{V}$, then $F_{\tilde{V}}^{(1)}$ is said to stochastically dominate $F_{\tilde{V}}^{(2)}$, which is written more concisely as $F_{\tilde{V}}^{(1)} \vdash F_{\tilde{V}}^{(2)}$. This is an equivalent concept of stochastic dominance—but with a focus on CPFs.

Now, reversing the arguments in Section 3.1.4, a non-decreasing possibility distribution may also act as a CPF.

Proposition 33. *If the possibility distribution $\pi_{\tilde{V}}$ of the imprecise variable $\tilde{V}$ is also a CPF, then it is an exact representation of the family of probability distributions whose CPFs stochastically dominate $\pi_{\tilde{V}}$, i.e. of $\mathfrak{P}_{\tilde{V}} = \{F_{\tilde{V}} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : F_{\tilde{V}} \vdash \pi_{\tilde{V}}\}$.*

Proof. Let $v \in \mathbb{V}$. As a first step, it is expedient to observe that, if $\pi_{\tilde{V}}$ is a CPF, so is $F_{\tilde{V}}^*$ defined by $F_{\tilde{V}}^*(\xi) = \pi_{\tilde{V}}(\xi)$ if $\xi \geq v$ and $F_{\tilde{V}}^*(\xi) = 0$ otherwise. Both the monotonicity and the right-continuity of $F_{\tilde{V}}^*$ follow immediately. Furthermore, this definition implies that $F_{\tilde{V}}^* \in \mathfrak{P}_{\tilde{V}}$, because $F_{\tilde{V}}^*(\xi) \leq \pi_{\tilde{V}}(\xi)$ for all $\xi \in \mathbb{V}$. Finally, $\mathrm{P}^*(\{v\}) = \pi_{\tilde{V}}(v)$ is the maximum possible value of $\mathrm{P}_{\tilde{V}}(\{v\}) \leq \mathrm{P}_{\tilde{V}}((-\infty, v]) = F_{\tilde{V}}(v) \leq \pi_{\tilde{V}}(v)$ for all $\mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}$, and therefore, the optimal plausibility of v is given by $\rho_{\tilde{V}}^{\mathrm{opt.}}(v) = \pi_{\tilde{V}}(v)$. In conclusion, the sublevel sets of $\pi_{\tilde{V}}$ constitute the sublevel sets of $\rho_{\tilde{V}}^{\mathrm{opt.}}$, and, from Eq. (2.25) and by the coherence of $\pi_{\tilde{V}}$, the proposition results as a consequence of Lemma 15. □

For instance, the superuniform possibility distribution is consistent with the standard uniform probability distribution $\mathcal{U}(0, 1)$ as it coincides with its CPF. By the Proposition 33, it is also clear that the superuniform distribution is also consistent with all probability distributions on $[0, 1]$ that stochastically dominate the uniform distribution, see Figure 3.6, providing further justification for its name.

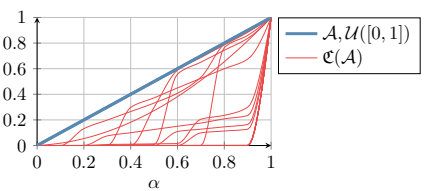


Figure 3.6: Credal Set of Superuniform Distribution.

Proposition 34. *If the possibility distribution $\pi_{\tilde{V}}$ of the imprecise variable $\tilde{V}$ is also a complementary CPF, then it is an exact representation of the family of probability distributions whose CPFs are stochastically dominated by $\pi_{\tilde{V}}$, i.e. of $\mathfrak{P} = \{F_{\tilde{V}} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : \pi_{\tilde{V}} \vdash F_{\tilde{V}}\}$.*

Proof. The proof is similar to that of Proposition 33. □

It is easy to see that these propositions are essentially variants of Lemma 3 for in- and decreasing possibility distributions. Put differently, specificity is an alternative concept of stochastic dominance.

Since the (Complementary) Cumulative P-Π-transform of $\mathrm{P}_{\tilde{V}}$ as discussed in Section 3.1.4 yields increasing (decreasing) possibility distributions $\pi_{\tilde{V}}$, it is also clear that its credal set then includes all probability distributions that stochastically dominate (are stochastically dominated by) $F_{\tilde{V}}$—not just $F_{\tilde{V}}$ itself. This observation is generalized in the following.

3.3.1 Cumulative Imprecise-Probability-to-Possibility Transforms

The *Cumulative* IP-Π-*transform* $\pi_{\tilde{V}}^{\mathrm{CPF}}$ and the *Complementary Cumulative* IP-Π-*transform* $\pi_{\tilde{V}}^{\mathrm{CCPF}}$ of a family of probabilities $\mathfrak{P}_{\tilde{V}}$ are given by

$$\pi_{\tilde{V}}^{\mathrm{CPF}}(v) = \sup_{F_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}} F_{\tilde{V}}(v) \qquad \text{and} \qquad \pi_{\tilde{V}}^{\mathrm{CCPF}}(v) = \sup_{F_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}} 1 - F_{\tilde{V}}(v), \tag{3.54}$$

for $v \in \mathbb{V}$, respectively. It follows immediately that this implies $F_{\tilde{V}} \vdash \pi_{\tilde{V}}^{\mathrm{CPF}}$ and $\pi_{\tilde{V}}^{\mathrm{CCPF}} \vdash F_{\tilde{V}}$, respectively, for all $F_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}$. As guaranteed by Propositions 33 and 34, this is a very straight-forward method of encoding such families of probabilities in a possibility distribution. For instance, the superuniform possibility distribution can be derived as the Cumulative IP-Π-transform of all probability distributions that stochastically dominate $\mathcal{U}([0,1])$.

Moreover, every IP-Π-transform can be viewed as a Cumulative IP-Π-transform of $\tilde{\rho} = \rho_{\tilde{V}}(\tilde{V})$, i.e. of $\mathfrak{P}_{\tilde{\rho}} = \{\mathrm{P}_{\rho_{\tilde{V}}(\tilde{V})} : \mathrm{P}_{\tilde{V}} \in \mathfrak{P}_{\tilde{V}}\}$, producing $\pi_{\tilde{\rho}}^{\mathrm{CPF}}$ in connection with the inverse extension under the map $\rho_{\tilde{V}}$ yielding

$$\pi_{\tilde{V}} = \pi_{\rho_{\tilde{V}}^{-1}(\tilde{\rho})} \tag{3.55}$$

This also justifies thinking of a possibility distributions as a generalization of CPFs.

Clearly, one cannot generally expect to obtain an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}$ by these transforms; however, the *(Complementary) Cumulative* IP-Π-*transform* allows for a possibilistic approach to a different well-known framework for imprecise probabilities.

3.3.2 Probability Boxes

Probability Boxes (p-boxes) are a convenient way of modeling stochastic dominance from below and above [FersonEtAl15]. To this end, a p-box is associated with an upper CPF $F_{\tilde{V}}^{+}$

and a lower CPF $F_{\tilde{V}}^-$, and its credal set contains precisely those CPFs $F_{\tilde{V}}$ that stochastically dominate the former and are stochastically dominated by the latter, i.e.

$$\mathfrak{P}_{\tilde{V}}^{\text{p-box}} = \{F_{\tilde{V}} : F_{\tilde{V}}^-(v) \leq F_{\tilde{V}}(v) \leq F_{\tilde{V}}^+(v) \; \forall \; v \in \mathbb{V}\}. \tag{3.56}$$

The relation between p-boxes and possibility distributions has been of interest to several scholars [DesterckeDuboisChojnacki08, TroffaesMirandaDestercke13, HoseHanss21c]. The main connections interpreted in the context of this thesis are given in the following.

P-boxes describe both types of stochastic dominance at once, whereas a possibility distribution can, at the most, describe one such phenomenon—but may also contain entirely different information if $\pi_{\tilde{V}}$ does not coincide with a (complementary) CPF. It is not surprising, then, that, due to their specific properties, it is not generally possible to transform a possibilistic description into a p-box description and vice versa without some loss of information. The credal set of a possibility distribution and a p-box hardly ever coincide—with a few notable exceptions.

- By Proposition 33, a possibility distribution $\pi_{\tilde{V}}$ which is also a CPF exactly describes the p-box with the trivial lower CPF with $F_{\tilde{V}}^-(v) = 0$ for all $v < \infty$ and $F_{\tilde{V}}^-(\infty) = 1$ and the upper CPF $F_{\tilde{V}}^+ = \pi_{\tilde{V}}$.
- By Proposition 34, a possibility distribution $\pi_{\tilde{V}}$ which is also a complementary CPF exactly describes the p-box with the lower CPF $F_{\tilde{V}}^-(v) = 1 - \pi_{\tilde{V}}(v)$ for all $v \in \mathbb{V}$ and the trivial upper CPF with $F^+(-\infty) = 0$ and $F_{\tilde{V}}^+(v) = 1$ for $v > -\infty$.
- Conversely, a p-box that is associated with at least one such trival upper or lower CPF, a so-called one-sided p-box,[34] coincides exactly with the credal set of the corresponding possibility distribution. For instance, if the upper CPF is trivial, then the appropriate possibility distribution is given by $\pi_{\tilde{V}}(v) = 1 - F_{\tilde{V}}^-(v)$ for all $v \in \mathbb{V}$.

In any case, one may always find conservative outer approximations of one description in the other framework.

3.3.2.1 From Probability Boxes to Possibilities

In general, every IP-Π-transform is admissible for converting a p-box description into a possibilistic one. Naturally, the Cumulative IP-Π-transform and the Complementary Cumulative IP-Π-transform lend themselves to this cause, if only the upper or the lower CPF is to be preserved since this is equivalent to relaxing the other CPF to be trivial. The more common *P-Box-to-Possibility Transform* is obtained by the IP-Π-transform

[34] An example of a possibility distribution that coincides with a one-sided p-box is given by the superuniform distribution.

under the elementary plausibilities $\rho_{\tilde{V}}(v) = \min(F^+_{\tilde{V}}(v), 1 - F^-_{\tilde{V}}(v))$ for $v \in \mathbb{V}$. This choice puts equal emphasis on both the lower and the upper tails of the p-box and yields the possibility distribution $\pi^{\text{p-box}}_{\tilde{V}}$ given by

$$\pi^{\text{p-box}}_{\tilde{V}}(v) = \min\left(1, 2F^+_{\tilde{V}}(v), 2\left(1 - F^-_{\tilde{V}}(v)\right)\right) \tag{3.57}$$

for all $v \in \mathbb{V}$, as shown by Hose and Hanss [HoseHanss21c]. A simple derivation of this result is that, as discussed in Propositions 33 and 34, Destercke et al. [DesterckeDuboisChojnacki08] correctly identify a p-box as the intersection of the credal sets of the two possibility measures $\pi^{(1)}_{\tilde{V}} = F^+_{\tilde{V}}$ and $\pi^{(2)}_{\tilde{V}} = 1 - F^-_{\tilde{V}}$ allowing for the application of Lemma 22, which yields the above possibility distribution.

Naturally, this is an outer approximation since p-boxes and possibility distributions express imprecise probabilistic information in fundamentally different ways and it is hardly ever possible to exactly describe one via the other. For instance, the CIIF and CNF resulting from the P-Box-to-Possibility Transform will usually be considerably higher than the upper and considerably lower than the lower CPF values of the original p-box, respectively, even though—under a loss-free transformation—these values would coincide.

When the p-box is degenerate, i.e. it is a precise probability distribution $F_{\tilde{V}} = F^-_{\tilde{V}} = F^+_{\tilde{V}}$, then the corresponding P-Box-to-Possibility Transform is also referred to as the *Symmetric Cumulative* P-Π-*transform*.

3.3.2.2 From Possibilities to Probability Boxes

In contrast to the non-unique choice for converting p-boxes into possibility distributions, the converse *Possibility-to-P-Box Transform* is intuitive and straight-forward. Due to the coherence of the possibility measure, the upper- and lower-most cumulative distributions in the credal set of an imprecise variable are given by the CIIF and the CNF, respectively, and choosing these to be the upper and lower CPF of the p-box, i.e.

$$F^+(v) = \Pi(\tilde{V} \leq v) \qquad \text{and} \qquad F^-(v) = \text{N}(\tilde{V} \leq v) \tag{3.58}$$

for all $v \in \mathbb{V}$, is an optimal choice, in the sense that every other outer approximation by a p-box would be strictly less expressive, i.e., would contain more probability distributions.

Nevertheless, this choice, too, does not coincide with the original credal set. This claim is verified via the example depicted in Fig. 3.7. The original triangular possibility distribution $\pi_{\tilde{V}} = \Delta(1, 4, 7)$ is outer approximated by the p-box associated with the upper and lower CPFs given by the former's CIIF $F^+_{\tilde{V}}$ and CNF $F^-_{\tilde{V}}$, respectively. Since the CPF $F^0_{\tilde{V}}$ is within the bounds specified by this p-box, it is a member thereof. However, this CPF is not contained in the credal set of the original possibility distribution. For $F^0_{\tilde{V}}$ to be a consistent probability distribution, e.g., the superlevel set $\mathcal{C}^{\frac{2}{3}}_{\pi_{\tilde{V}}} = (3, 5)$ would have to have a probability greater than or equal to $\frac{1}{3}$ but $\text{P}^0_{\tilde{V}}(\mathcal{C}^{\frac{2}{3}}_{\pi_{\tilde{V}}}) = F^0_{\tilde{V}}(5) - F^0_{\tilde{V}}(3) = 0$.

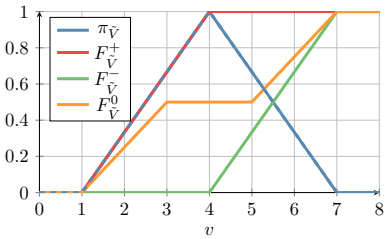


Figure 3.7: Inconsistent CPF $F^0_{\tilde{V}}$ included in the p-box resulting from Possibility-to-P-Box Transform of the triangular possibility distribution $\Delta(1, 4, 7)$ via corresponding CIIF $F^+_{\tilde{V}}$ and CNF $F^-_{\tilde{V}}$.

3.4 Imprecise Expectations and Moments

In probabilistic analyses, the *expected value* is a fundamental tool for the characterization of a real-valued imprecise variable. On a probability space, this expected value of the imprecise variable $\tilde{V} \sim \mathrm{P}_{\tilde{V}}$ is defined as the Lebesgue integral

$$\begin{aligned}\mathfrak{E}[\tilde{V}] &= \int_{\Omega} \tilde{V}(\omega)\,\mathrm{dP}(\omega) = \int_{\mathbb{V}} v\,\mathrm{dP}_{\tilde{V}}(v) \\ &= \int_{-\infty}^{0} \left(\mathrm{P}_{\tilde{V}}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) - 1\right) \mathrm{d}v + \int_{0}^{+\infty} \mathrm{P}_{\tilde{V}}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) \mathrm{d}v.\end{aligned} \tag{3.59}$$

and denotes the first moment of $\tilde{V}$, which is also abbreviated as $\mathfrak{E}[\tilde{V}]$. A very simple intuition behind the expected value comes from stochastic dominance. If the CPFs of the $\mathbb{V}$-valued imprecise variables $\tilde{V}_1$ and $\tilde{V}_2$ fulfill $F_{\tilde{V}_1} \vdash F_{\tilde{V}_2}$, then it follows immediately that

$$\begin{aligned}\mathfrak{E}[\tilde{V}_2] &= \int_{-\infty}^{0} \left(\mathrm{P}_{\tilde{V}_2}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) - 1\right) \mathrm{d}v + \int_{0}^{+\infty} \mathrm{P}_{\tilde{V}_2}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) \mathrm{d}v \\ &\leq \int_{-\infty}^{0} \left(\mathrm{P}_{\tilde{V}_1}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) - 1\right) \mathrm{d}v + \int_{0}^{+\infty} \mathrm{P}_{\tilde{V}_1}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) \mathrm{d}v \\ &= \mathfrak{E}[\tilde{V}_1].\end{aligned} \tag{3.60}$$

That is, $\tilde{V}_2$ is expected to be smaller than $\tilde{V}_1$ if the (CPF of the) former is stochastically dominated by the (CPF of the) latter.

3.4.1 From Possibilities to Moments

A general notion of *expectation* for *real-valued* imprecise variables $\tilde{V} : \Omega \to \mathbb{R}$ under capacities M on (Ω, Σ) can be defined via the Choquet integral [Choquet54]

$$\begin{aligned}\mathfrak{E}_{\mathrm{M}}[\tilde{V}] = &\int_{-\infty}^{0} \left(\mathrm{M}\left(\left\{\omega \in \Omega : \tilde{V}(\omega) \geq v\right\}\right) - \mathrm{M}\left(\Omega\right)\right) \mathrm{d}v \\ &+ \int_{0}^{+\infty} \mathrm{M}\left(\left\{\omega \in \Omega : \tilde{V}(\omega) \geq v\right\}\right) \mathrm{d}v.\end{aligned} \tag{3.61}$$

which is, for the purposes of this thesis, equivalent to

$$\mathfrak{E}_{\mathrm{M}}[\tilde{V}] = \int_{-\infty}^{0} \left(\mathrm{M}_{\tilde{V}}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) - 1\right) \mathrm{d}v + \int_{0}^{+\infty} \mathrm{M}_{\tilde{V}}\left(\{\xi \in \mathbb{R} : \xi \geq v\}\right) \mathrm{d}v. \tag{3.62}$$

If M is a possibility or necessity measure, then, according to Troffaes et al. [TroffaesCooman14, Proposition 7.14], this reduces to

$$\mathfrak{E}_{\mathrm{N}}[\tilde{V}] = \int_{0}^{1} \inf \mathcal{C}^{\alpha}_{\pi_{\tilde{V}}} \mathrm{d}\alpha \qquad \text{and} \qquad \mathfrak{E}_{\Pi}[\tilde{V}] = -\mathfrak{E}_{\mathrm{N}}[-\tilde{V}] = \int_{0}^{1} \sup \mathcal{C}^{\alpha}_{\pi_{\tilde{V}}} \mathrm{d}\alpha. \tag{3.63}$$

For instance, the superuniform variable $\tilde{V} \sim \mathcal{A}$ has a lower expectation of

$$\mathfrak{E}_{\mathrm{N}}[\tilde{V}] = \int_{0}^{1} 1 - v \, \mathrm{d}v = \frac{1}{2} \tag{3.64}$$

and an upper expectation of

$$\mathfrak{E}_{\Pi}[\tilde{V}] = \int_{0}^{1} 1 \, \mathrm{d}v = 1. \tag{3.65}$$

If $\mathrm{M} = \mathrm{P}$ is a probability measure, the corresponding expectation corresponds to the classical expected value [Denneberg94]. The generalization to expectations of multivariate imprecise variables is simply the elementwise computation of the marginal expectations.

If a possibilistic description of the imprecise variable $\tilde{V} \sim \pi_{\tilde{V}}$ is available, the expectations $\mathfrak{E}_{\Pi}[\tilde{V}]$ and $\mathfrak{E}_{\mathrm{N}}[\tilde{V}]$ provide upper and lower bounds [DuboisPrade87] of the true expected value $\mathfrak{E}_{\mathrm{P}}[\tilde{V}]$, i.e.

$$\mathfrak{E}_{\mathrm{N}}[\tilde{V}] = \inf_{\mathrm{P} \preceq \pi} \mathfrak{E}_{\mathrm{P}}[\tilde{V}] \qquad \text{and} \qquad \mathfrak{E}_{\Pi}[\tilde{V}] = \sup_{\mathrm{P} \preceq \pi} \mathfrak{E}_{\mathrm{P}}[\tilde{V}], \tag{3.66}$$

which constitutes a more general concept of coherence considering that under this definition the probability, possibility, necessity, etc. of any event $E \in \Sigma$ may be expressed as the corresponding expectation of the imprecise variable given by the indicator function $\mathcal{I}_E$. Therefore, they are called *upper* and *lower expected values* of $\tilde{V}$, or *imprecise expectations*. On the basis of such imprecise expectations, a very general framework for imprecise probabilities may also be formulated, which are then called upper and lower

previsions [Walley91, AugustinEtAl14]. However, further discussion of the relationship between possibilities and upper previsions would go far beyond the scope of this thesis.

Finally, the expected value is only the first of many *moments* of an imprecise variable. The k-th (absolute) moment of $\tilde{V} \sim \mathrm{P}_{\tilde{V}}$ is given by

$$\mathfrak{E}_{\mathrm{P}}[\tilde{V}^k] = \int_{\Omega} \tilde{V}^k(\omega)\,\mathrm{dP}(\omega) = \int_{\mathbb{V}} v^k\,\mathrm{dP}_{\tilde{V}}(v), \tag{3.67}$$

and the k-th central moment is $\mathfrak{E}_{\mathrm{P}}[(\tilde{V} - \mathfrak{E}_{\mathrm{P}}[\tilde{V}])^k]$. In particular, the second central moment is called the *variance* $\mathrm{Var}_{\mathrm{P}}[\tilde{V}] = \mathfrak{E}_{\mathrm{P}}[(\tilde{V} - \mathfrak{E}_{\mathrm{P}}[\tilde{V}])^2]$. By means of the Choquet integral in Eq. (3.61), the corresponding imprecise (absolute) moments may readily be computed—or equivalently as the expectation of the corresponding explicit extension; however, at least in possibility theory, higher-order moments of an imprecise variable generally bear little information.

3.4.2 From Moments to Possibilities

Conversely, based on well-known probabilistic inequalities, it is also possible to derive possibility distributions from distributional imprecision due to a limited number of known moments of an imprecise variable [DuboisEtAl04].

3.4.2.1 Markov Distribution

Suppose that the imprecise variable $\tilde{V}$ is known to be non-negative with expected value $\mathfrak{E}_{\mathrm{P}}[\tilde{V}] = m > 0$. In this case, the *Markov distribution* $\mathcal{M}(m)$ provides a possibilistic description of $\mathfrak{P}_{\tilde{V}}^{\mathrm{Markov}} = \{\mathrm{P}_{\tilde{V}} : \mathrm{P}(\tilde{V} \geq 0) = 1 \wedge \mathfrak{E}_{\mathrm{P}}[\tilde{V}] = m\}$. The well-known (tight) Markov inequality states that, under these assumptions,

$$\mathrm{P}\left(\tilde{V} \geq v\right) \leq \frac{m}{v} \tag{3.68}$$

for $v > 0$, and by the Complementary Cumulative IP-Π-transform, one immediately obtains the Markov distribution given by

$$\pi_{\tilde{V}}^{\mathrm{Markov}}(v) = \min\left(1, \frac{m}{v}\right) \tag{3.69}$$

for $v \geq 0$ and $\pi_{\tilde{V}}^{\mathrm{Markov}}(v) = 0$ for $v < 0$, see Fig. 3.8. Even though the Markov inequality is tight, and therefore the Markov distribution is maximally specific, this is not an exact possibilistic description of $\mathfrak{P}_{\tilde{V}}^{\mathrm{Markov}}$. For instance, the lower and upper expectations of $\tilde{V}$ under $\pi_{\tilde{V}}^{\mathrm{Markov}}$ are

$$\mathfrak{E}_{\mathrm{N}}[\tilde{V}] = 0 \qquad \text{and} \qquad \mathfrak{E}_{\Pi}[\tilde{V}] = \infty, \tag{3.70}$$

and therefore, $\mathfrak{P}_{\tilde{V}}^{\mathrm{Markov}} \neq \mathfrak{C}(\pi_{\tilde{V}}^{\mathrm{Markov}})$.

3.4.2.2 Chebychev Distribution

Similarly, the Chebychev inequality

$$\mathrm{P}\left(|\tilde{V} - m| \geq |v - m|\right) \leq \frac{s^2}{(v-m)^2} \tag{3.71}$$

for $v \in \mathbb{R}$ allows one to derive a possibilistic description of all probability distributions of the $\mathbb{R}$-valued imprecise variable $\tilde{V}$ with an expected value of $\mathfrak{E}_{\mathrm{P}}[\tilde{V}] = m \in \mathbb{R}$ and a variance of $\mathrm{Var}_{\mathrm{P}}[\tilde{V}] = s^2 > 0$. Under the IP-$\Pi$-transform with the symmetric plausibility distribution $\rho_{\tilde{V}}^{\text{opt.}}(v) = \exp(-v^2)$ for $v \in \mathbb{R}$, the *Chebychev distribution* $\mathcal{M}^2(m, s^2)$ is given by

$$\pi_{\tilde{V}}^{\text{Chebychev}}(v) = \min\left(1, \frac{s^2}{(v-m)^2}\right) \tag{3.72}$$

for $v \in \mathbb{R}$, see Fig. 3.9.

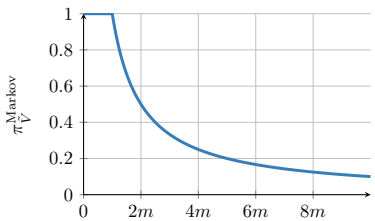


Figure 3.8: Markov Distribution.

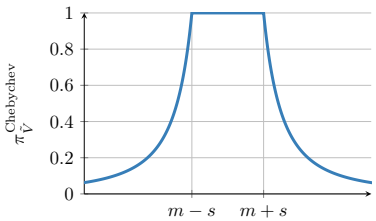


Figure 3.9: Chebychev Distribution.

Similarly, the *k-th order Chebychev Distribution* $\mathcal{M}^k(m, \xi^k)$, given by

$$\pi_{\tilde{V}}^{\text{Chebychev},k}(v) = \min\left(1, \frac{\xi^k}{(v-m)^k}\right) \tag{3.73}$$

for $v \in \mathbb{R}$, is obtained from the general Chebychev inequality for imprecise variables $\tilde{V}$ with expected value $\mathfrak{E}_{\mathrm{P}}[\tilde{V}] = m$ and k-th central moment $\mathfrak{E}_{\mathrm{P}}[(\tilde{V} - m)^k] = \xi^k$.

Many other probabilistic inequalities can be employed to derive such moment-based possibility distributions. Alternatively, also inequalities based on geometric properties, such as the Camp-Meidell inequality for unimodal and symmetric distributions may be exploited [DuboisEtAl04].

As a final remark, the rather large loss of expressiveness when converting partially known moments into possibility distributions, as indicated by the loss of information about the expected value in the Markov distribution, suggests that—contrary to very precise IP descriptions by means of stochastic dominance—possibility theory is not well suited to describe such distributional imprecision. Instead, distribution-free techniques [FersonGray21] for the analysis and propagation of such IP descriptions should be employed.

3.5 Dependency and Interaction

Amongst others, the extension formulations in Eqs. (3.27), (3.30) and (3.39) exhibit a clear commonality: They require knowledge of the possibility distribution $\pi_{\tilde{V}}$ of $\tilde{V}$. Yet, if $\tilde{V} = (\tilde{V}_1, \ldots, \tilde{V}_m)$ is composed of the marginal variables $\tilde{V}_1, \ldots, \tilde{V}_m$ on $\mathbb{V}_1, \ldots, \mathbb{V}_m$, respectively, such joint possibility distribution $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ cannot always be provided a-priori. More commonly, only the respective marginal possibility distributions $\pi_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m}$ are available, and it is not immediately clear how to construct the joint distributions $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ on $\mathbb{V} = \mathbb{V}_1 \times \ldots \times \mathbb{V}_m$. To be able to do so, one must provide additional dependency information [CousoMoralWalley00]; however, it is not impossible that one may not or only partially be able to specify the dependencies. This alone is a source of distributional imprecision—even when precise marginal probability distributions are available—because stochastic analyses generally require precise dependency models [WilliamsonDowns90].

In any case, the primary goal must be to find a possibilistic description of the family of admissible joint probability distributions under the given marginal and dependency information.

3.5.1 Copulae

Both in the probabilistic and the possibilistic case, dependency information may be expressed in the form of *copulae* that connect marginal and joint distributions.

A *stochastic/probability copula* $C : [0,1]^m \to [0,1]$ is the joint CPF of the m-dimensional imprecise vector $(F_{\tilde{V}_1}(\tilde{V}_1), \ldots, F_{\tilde{V}_m}(\tilde{V}_m))$, where—due to the Probability Integral Transform—the marginals must have a uniform probability distribution [Joe14]. This stochastic copula connects the marginal and joint CPFs of $\tilde{V}_1, \ldots, \tilde{V}_m$ via

$$F_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1, \ldots, v_m) = C\left(F_{\tilde{V}_1}(v_1), \ldots, F_{\tilde{V}_m}(v_m)\right) \tag{3.74}$$

for all $(v_1, \ldots, v_m) \in \mathbb{V}$, which is often written as

$$F_{\tilde{V}_1,\ldots,\tilde{V}_m} = C\left(F_{\tilde{V}_1}, \ldots, F_{\tilde{V}_m}\right). \tag{3.75}$$

Sklar's Theorem states that every multivariate CPF may be expressed via the corresponding marginal CPFs and a stochastic copula, i.e., the latter contains all the dependency information required to build the joint CPF.

Similarly, a *possibility copula* $\mathcal{J} : [0,1]^m \to [0,1]$ connects the marginal and joint possibility distributions via

$$\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1, \ldots, v_m) = \mathcal{J}\left(\pi_{\tilde{V}_1}(v_1), \ldots, \pi_{\tilde{V}_m}(v_m)\right) \tag{3.76}$$

for all $(v_1, \ldots, v_m) \in \mathbb{V}$, which is also written as $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m} = \mathcal{J}(\pi_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m})$. If a possibility copula (Π-copula) is employed to connect plausibility distributions, it is also said to be a

plausibility copula. The joint possibility distribution in Eq. (3.76) is said to be *produced under* $\mathcal{J}$.

This definition allows to state, e.g., the implicit extension in Eq. (3.27) under a given Π-copula $\mathcal{J}$ reading

$$\pi_{\tilde{Q}}(q) = \sup_{(v_1,\ldots,v_m)\in\mathbb{V}\,:\,0=\Xi(v_1,\ldots,v_m,q)} \mathcal{J}\left(\pi_{\tilde{V}_1}(v_1),\ldots,\pi_{\tilde{V}_m}(v_m)\right) \tag{3.77}$$

for all $q \in \mathbb{Q}$.

Of course, it is desirable that Π-copulae possess certain properties, namely measurability and normality, which are required so that the joint possibility distribution in Eq. (3.76) is also a valid elementary possibility function; this is the case for all Π-copulae proposed in the following. The most important aspect, however, is the dependency information it encodes.

3.5.2 Non-Interactivity

In the formulation of the original Extension Principle [Zadeh75a], Zadeh implicitly states that a Π-copula should be derived from some t-norm [Hanss05]. In practice, the t-norm is often chosen to be the minimum operator, which yields the so called *non-interactive* Π-*copula*

$$\mathcal{J}^{\mathrm{NI}}(\alpha_1,\ldots,\alpha_m) = \min_{i=1,\ldots,m} \alpha_i \tag{3.78}$$

for all $\alpha_1,\ldots,\alpha_m \in [0,1]$. This choice has already appeared informally in Section 2.3.3.2. Indeed, this is no coincidence as the conjunction of possibilistic descriptions and the construction of joint possibility distributions are closely related [TroffaesMirandaDestercke13].

Notably, the marginals of $\mathcal{J}^{\mathrm{NI}}(\pi_{\tilde{V}_1},\ldots,\pi_{\tilde{V}_m})$ always coincide with the original marginal possibility distributions $\pi_{\tilde{V}_1},\ldots,\pi_{\tilde{V}_m}$, which is verified by considering

$$\begin{aligned}
&\sup_{(\xi_1,\ldots,\xi_m)\in\mathbb{V}:\xi_i=v_i} \mathcal{J}^{\mathrm{NI}}\left(\pi_{\tilde{V}_1}(\xi_1),\ldots,\pi_{\tilde{V}_m}(\xi_m)\right) = \sup_{(\xi_1,\ldots,\xi_m)\in\mathbb{V}:\xi_i=v_i} \min_{j=1,\ldots,m} \pi_{\tilde{V}_j}(\xi_j) \\
&= \min\left(\underbrace{\sup_{\xi_1\in\mathbb{V}_1} \pi_{\tilde{V}_1}(\xi_1)}_{=1},\ldots,\underbrace{\sup_{\xi_{i-1}\in\mathbb{V}_{i-1}} \pi_{\tilde{V}_{i-1}}(\xi_{i-1})}_{=1},\ \pi_{\tilde{V}_i}(v_i),\right. \\
&\qquad\left.\underbrace{\sup_{\xi_{i+1}\in\mathbb{V}_{i+1}} \pi_{\tilde{V}_{i+1}}(\xi_{i+1})}_{=1},\ldots,\underbrace{\sup_{\xi_m\in\mathbb{V}_m} \pi_{\tilde{V}_m}(\xi_m)}_{=1}\right) = \pi_{\tilde{V}_i}(v_i)
\end{aligned} \tag{3.79}$$

for all $v_i \in \mathbb{V}_i$ and all $i = 1,\ldots,m$. By this observation, the term 'non-interactivity' is not ill-chosen when considering its origins in fuzzy set theory from a subjectivist point of

view. Therein, a membership function, which is—for the purposes of this thesis—the same as an elementary possibility function, is usually required to be provided by some expert. Assuming that this expert is in possession of a possibilistic description π of the entire universe Ω, within which all dependency information has already been accounted for, all of these membership functions, both the marginals $\pi_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m}$ and the joint $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$, would then be a corresponding pushforward of π. And non-interactivity expresses nothing other than the fact that the marginals have been derived as a pushforward of this universal description, and no further dependencies need to be considered. All dependency information is already encoded in the marginal possibility distributions warranting an application of the non-interactive Π-copula. Still, non-interactivity assumes that the expert is in possession of a single general possibilistic description of the entire universe, or at least the experiment—a claim that is, at least, debatable when using possibilities as a description of imprecise probabilities.

Therefore, it is little surprising, that Baudrit et al. point out that possibilistic calculus based on the non-interactive variant of the Extension Principle is not a conservative counterpart to probabilistic calculus in most cases [BaudritDuboisGuyonnet06] because the non-interactive joint possibility distribution rarely preserves consistency—especially under the popular assumption of stochastic independence. However, this finding directly violates the Principle of Representation.

Subsequently, more appropriate Π-copulae have been proposed by various scholars [DesterckeDuboisChojnacki09, TroffaesMirandaDestercke13, HoseHanss19c]. Very interestingly, these copulae, too, can be derived under the IP-Π-transform [HoseHanss21c]; however, an exact possibilistic description of the possibilistic information about the marginals and their dependencies is not generally possible, and the provided closed-form expressions are often only less specific outer approximations of the respective IP-Π-transform.

The sheer number of possible dependency information does not permit a comprehensive discussion of all types of (in-)dependence. Instead, only those with an intuitive interpretation that seem to be most applicable in most scenarios shall be investigated, namely unknown interaction and strong independence.

Due to the apparent similarities between elementary possibilities and p-values, which are to be explored in the following chapter, many of the obtained Π-copulae are based on methods for the combination of p-values, a topic which is further explored in Section 4.3.2.2.

3.5.3 Unknown Interaction

The most basic and hardly debatable assumption is *marginal consistency*, where all marginal probability distributions $\mathrm{P}_{\tilde{V}_1}, \ldots, \mathrm{P}_{\tilde{V}_m}$ ought to be consistent with the marginal possibility

distributions $\pi_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m}$, respectively. No further assumptions are imposed; in particular, no requirements on the interdependencies constraining the shape of the joint probability distributions $\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ are given, which is the most conservative dependency assessment and also referred to as the *Fréchet case*, the *general case*, or *unknown interaction*.

In the absence of additional information about dependencies and interactions, the admissible family of joint probability distributions is, thus, given by

$$\mathfrak{P}^{\mathrm{UI}}_{\tilde{V}_1,\ldots,\tilde{V}_m} = \left\{\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : \mathrm{P}_{\tilde{V}_1} \preceq \pi_{\tilde{V}_1}, \ldots, \mathrm{P}_{\tilde{V}_m} \preceq \pi_{\tilde{V}_m}\right\}. \tag{3.80}$$

and the *unknown-interaction* Π*-copula* $\mathcal{J}^{\mathrm{UI}}$ is defined as

$$\mathcal{J}^{\mathrm{UI}}(\alpha_1, \ldots, \alpha_m) = \min\left(1, m \cdot \min_{i=1,\ldots,m} \alpha_i\right) \tag{3.81}$$

for all $\alpha_1, \ldots, \alpha_m \in [0, 1]$.

Proposition 35. *The unknown-interaction* Π*-copula* $\mathcal{J}^{\mathrm{UI}}$ *produces an outer approximation of* $\mathfrak{P}^{\mathrm{UI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$.

Proof. The family $\mathfrak{P}^{\mathrm{UI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ may be re-written as the conjunction $\pi^{(1)}_{\tilde{V}_1,\ldots,\tilde{V}_m} \wedge \ldots \wedge \pi^{(m)}_{\tilde{V}_1,\ldots,\tilde{V}_m}$, where the $\pi^{(i)}_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m}$ are the vacuous extensions of $\pi^{(i)}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ for $i = 1, \ldots, m$, respectively. As they do not fulfill the assumptions of comonotonicity, quasi-vacuousness or a total specificity order, the most general Lemma 22 must be applied to evaluate the conjunction, which yields precisely the joint possibility distribution $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ that is produced under the unknown-interaction Π-copula. □

This technique for the construction of joint possibility distributions also coincides with Bonferrroni's method for combining p-values with an unknown dependency structure [CramerKamps20].

The advantage of this Π-copula is that it is always applicable because unknown interaction, requiring only marginal consistency, is the weakest sensible assumption. In this sense, the fundamental Principle of Representation is never violated and consistency is always guaranteed—even though its application may not always yield the most specific, admissible joint possibility distribution.

Naturally, under additional assumptions the joint possibility distribution may become more expressive, as in the following case, exhibiting an application of the non-interactive Π-copula.

Proposition 36. *The non-interactive* Π*-copula* $\mathcal{J}^{\mathrm{NI}}$ *produces an exact possibilistic description of* $\mathfrak{P}^{\mathrm{UI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ *if* $\pi_{\tilde{V}_1}, \ldots, \pi_{\tilde{V}_m}$ *are quasi-vacuous.*

Proof. The proof is the same as that of Proposition 35 except that, instead of Lemma 22, Lemma 20 can be applied, which then yields the joint possibility distribution that is produced under the non-interactive Π-copula. □

3.5.4 Strong Independence

Two imprecise variables $\tilde{V}_1$ and $\tilde{V}_2$ are—in accordance with the notion of stochastic independence introduced earlier—said to be stochastically independent if the joint probability distribution $\mathrm{P}_{\tilde{V}_1,\tilde{V}_2}$ is constructed from the marginal probability distributions $\mathrm{P}_{\tilde{V}_1}$ and $\mathrm{P}_{\tilde{V}_2}$ via

$$\mathrm{P}_{\tilde{V}_1,\tilde{V}_2}(B_1, B_2) = \mathrm{P}_{\tilde{V}_1}(B_1) \cdot \mathrm{P}_{\tilde{V}_2}(B_2) \tag{3.82}$$

for all $B_1 \in \mathbb{B}(\mathbb{V}_1)$ and $B_2 \in \mathbb{B}(\mathbb{V}_2)$. That is, the information or assumptions about one variable assuming certain values do not alter the probability distribution of the other variable; the conditional probability that $\tilde{V}_1$ is in B_1 under the condition that $\tilde{V}_2 \in B_2$ remains unchanged, i.e., $\mathrm{P}(\tilde{V}_1 \in B_1 | \tilde{V}_2 \in B_2) = \mathrm{P}(\tilde{V}_1 \in B_1)$.

The generalization to stochastic independence among m imprecise variables, where

$$\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}(B_1, \ldots, B_m) = \prod_{i=1}^{m} \mathrm{P}_{\tilde{V}_i}(B_i), \tag{3.83}$$

written as $\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} = \prod_{i=1}^{m} \mathrm{P}_{\tilde{V}_i}$, for all $B_1 \in \mathbb{B}(\mathbb{V}_1), \ldots, B_m \in \mathbb{B}(\mathbb{V}_m)$, is expressed by the *independence P-copula* $C^{\text{ind.}} = \prod$.

Stochastic independence is fundamental to many probabilistic models, and a de-facto standard assumption in engineering. However, stochastic independence fans out into various non-equivalent concepts for imprecise probabilities—such as strong independence, random set independence, or epistemic irrelevance and independence, etc. — that all reduce to stochastic independence when considering the degenerate case of precise probabilistic knowledge [AugustinEtAl14, CousoMoralWalley00].

The perhaps most intuitive generalization of stochastic independence to an IP concept is *strong independence*. It assumes that the joint probability distributions may only be constructed from the marginal probability distributions in the credal sets of the marginal possibility distributions under the assumption of stochastic independence, and are, therefore, gathered in

$$\begin{aligned}\mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m} = \{\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : \mathrm{P}_{\tilde{V}_1} \preceq \pi_{\tilde{V}_1}, \ldots, \mathrm{P}_{\tilde{V}_m} \preceq \pi_{\tilde{V}_m}, \\ \mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} = \prod_{i=1}^{m} \mathrm{P}_{\tilde{V}_i}\}.\end{aligned} \tag{3.84}$$

3.5.4.1 Minimum-Based Description

Throughout this thesis, the minimum-operator has been used extensively to compute elementary possibilities and plausibilities from several individual elementary possibilities, e.g. in the previous Propositions 35 and 36. If only for this continuity, it is worth to investigate this choice also in the case of strong independence, i.e. the IP-Π-transform

of $\mathfrak{P}^{\mathrm{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ under the joint plausibility distribution $\rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ produced under the minimum-based plausibility copula $\mathcal{J}^{\min.\mathrm{pl.}} = \mathcal{J}^{\mathrm{UI}}$. The resulting *minimum-based strong-independence* Π-*copula* $\mathcal{J}^{\min.\mathrm{SI}}$ is

$$\mathcal{J}^{\min.\mathrm{SI}}(\alpha_1,\ldots,\alpha_m) = \min_{i=1,\ldots,m} 1-(1-\alpha_i)^m \tag{3.85}$$

for all $\alpha_1,\ldots,\alpha_m \in [0,1]$.

The following proposition is an earlier result of Hose and Hanss [HoseMäckHanss19a, HoseHanss19c, HoseHanss21c].

Proposition 37. *The minimum-based strong-independence* Π-*copula* $\mathcal{J}^{\mathrm{SI},\min.}$ *produces an outer approximation of* $\mathfrak{P}^{\mathrm{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$.

Proof. The proof is based on choosing the unknown-interactivity Π-copula as the plausibility copula $\mathcal{J}^{\min.\mathrm{pl.}} = \mathcal{J}^{\mathrm{UI}}$ and evaluating the IP-Π-transform $\mathfrak{T}[\mathfrak{P}^{\mathrm{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}, \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}]$ under the joint plausibility distribution $\rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ produced by $\mathcal{J}^{\min.\mathrm{pl.}}$.
Let $v_1 \in \mathbb{V}_1,\ldots,v_N \in \mathbb{V}_m$, and define $\alpha_1 = \pi_{\tilde{V}_1}(v_1),\ldots,\alpha_m = \pi_{\tilde{V}_1}(v_1)$ as well as $\rho^* = \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m) = \min_{i=1,\ldots,m}\alpha_i$. For all $\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} \in \mathfrak{P}^{\mathrm{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ it follows that

$$\begin{aligned}
&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) \leq \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m)\right\}\right)\\
=&1-\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) > \rho^*\right\}\right),
\end{aligned}$$

where $\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) > \rho^*\right\} = \mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_1}} \times \ldots \times \mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_m}}$, and, under the assumption of strong independence, this may be re-written as

$$\begin{aligned}
&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) > \rho^*\right\}\right)\\
=&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_1}} \times \ldots \times \mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_m}}\right) = \prod_{i=1}^{m}\mathrm{P}_{\tilde{V}_i}\left(\mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_i}}\right).
\end{aligned}$$

From the marginal consistency, it follows that $\mathrm{P}_{\tilde{V}_i}\left(\mathcal{C}^{\rho^*}_{\pi_{\tilde{V}_i}}\right) \geq 1-\rho^*$ for all $i = 1,\ldots,m$, and therefore

$$\begin{aligned}
&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) \leq \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m)\right\}\right)\\
\leq&1-(1-\rho^*) = \mathcal{J}^{\min.\mathrm{SI}}(\alpha_1,\ldots,\alpha_m).
\end{aligned}$$

Thus, $\mathfrak{T}[\mathfrak{P}^{\mathrm{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}, \rho^{\min.}_{\tilde{V}_1,\ldots,\tilde{V}_m}]$ is less specific than the joint possibility distribution produced by the minimum-based strong-independence Π-copula $\mathcal{J}^{\min.\mathrm{SI}}$ and the proposition is proven. □

The same Π-copula can be used in the case of *random set independence* [DesterckeDuboisChojnacki09], which shall not be discussed in this thesis.

Still, it is possible to find alternative Π-copulae for describing strong independence.

3.5.4.2 Maximum-Based Description

For instance, one may also define a maximum-based plausibility copula

$$\mathcal{J}^{\text{max.pl.}}(\alpha_1,\ldots,\alpha_m) = \max_{i=1,\ldots,m} \alpha_i \tag{3.86}$$

for all $\alpha_1,\ldots,\alpha_m \in [0,1]$, which can be employed to produce a maximum-based description of strong independence. The resulting *maximum-based strong-independence* Π*-copula* $\mathcal{J}^{\text{max.SI}}$, given by

$$\mathcal{J}^{\text{max.SI}}(\alpha_1,\ldots,\alpha_m) = \left(\max_{i=1,\ldots,m} \alpha_i\right)^m \tag{3.87}$$

for all $\alpha_1,\ldots,\alpha_m \in [0,1]$, produces a joint possibility distribution that has also been shown to be adequate in the case of *epistemic independence* [TroffaesMirandaDestercke13], which is not discussed here.

Proposition 38. *The maximum-based strong-independence* Π*-copula* $\mathcal{J}^{\text{max.SI}}$ *produces an outer approximation of* $\mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$.

Proof. The proof is based on the maximum-based plausibility copula $\mathcal{J}^{\text{max.pl.}}$ and the subsequent evaluation of the IP-Π-transform $\mathfrak{T}[\mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}, \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}]$ under the joint plausibility distribution $\rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$ produced by $\mathcal{J}^{\text{max.pl.}}$.
Let $v_1 \in \mathbb{V}_1,\ldots,v_N \in \mathbb{V}_m$, and define $\alpha_1 = \pi_{\tilde{V}_1}(v_1),\ldots,\alpha_m = \pi_{\tilde{V}_1}(v_1)$ as well as $\rho^* = \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m) = \max_{i=1,\ldots,m}\alpha_i$. Since

$$\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) \leq \rho^*\right\} = \mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_1}} \times \ldots \times \mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_m}},$$

and under the assumption of strong independence, it follows that

$$\begin{aligned}&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) \leq \rho^*\right\}\right)\\ =&\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_1}} \times \ldots \times \mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_m}}\right) = \prod_{i=1}^{m} \mathrm{P}_{\tilde{V}_i}\left(\mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_i}}\right)\end{aligned}$$

for all $\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m} \in \mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$. From the marginal consistency, it follows that $\mathrm{P}_{\tilde{V}_i}\left(\mathcal{S}^{\rho^*}_{\pi_{\tilde{V}_i}}\right) \leq \rho^*$ for all $i = 1,\ldots,m$, and therefore

$$\begin{aligned}\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m}\left(\left\{\xi_1 \in \mathbb{V}_1,\ldots,\xi_N \in \mathbb{V}_m : \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}(\xi_1,\ldots,\xi_m) \leq \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m)\right\}\right)&\\ \leq (\rho^*)^m = \mathcal{J}^{\text{max.SI}}(\alpha_1,\ldots,\alpha_m).&\end{aligned}$$

Thus, $\mathfrak{T}[\mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}, \rho^{\text{max.}}_{\tilde{V}_1,\ldots,\tilde{V}_m}]$ is less specific than the joint possibility distribution produced by the maximum-based strong-independence Π-copula $\mathcal{J}^{\text{max.SI}}$, and the proposition is proven. □

3.5.4.3 Other Descriptions

The minimum and maximum constitute well-known t- and s-norms from fuzzy set theory. It is, therefore, conjectured that similar strong independence Π-copulae may also be derived for other t- and s-norms. The proofs of Propositions 37 and 38 are instructive as to how this may be accomplished: For any joint plausibility distribution produced under the plausibility copula $\mathcal{J}^{\text{pl.}}$, the corresponding strong independence Π-copula $\mathcal{J}^{\text{SI}}$ is given by the Lebesgue measure, i.e. volume, of the set $\{z_1 \in [0,1], \ldots, z_N \in [0,1] : \mathcal{J}^{\text{pl.}}(z_1, \ldots, z_m) \leq \mathcal{J}^{\text{pl.}}(\alpha_1, \ldots, \alpha_m)\}$, i.e. by the integral

$$\begin{aligned}\mathcal{J}^{\text{SI}}(\alpha_1, \ldots, \alpha_m) &= \lambda\left(\{z_1 \in [0,1], \ldots, z_N \in [0,1] : \mathcal{J}^{\text{pl.}}(z_1, \ldots, z_m) \leq \mathcal{J}^{\text{pl.}}(\alpha_1, \ldots, \alpha_m)\}\right) \\ &= \int_{\substack{z_1 \in [0,1], \ldots, z_N \in [0,1]: \\ \mathcal{J}^{\text{pl.}}(z_1, \ldots, z_m) \leq \mathcal{J}^{\text{pl.}}(\alpha_1, \ldots, \alpha_m)}} \mathrm{d}z_1 \ldots \mathrm{d}z_m\end{aligned} \tag{3.88}$$

for all $\alpha_1, \ldots, \alpha_m \in [0,1]$. Figures 3.10, and 3.11 visualize this integration rule for the minimum- and maximum-based plausibility copulae for $m = 2$, respectively.

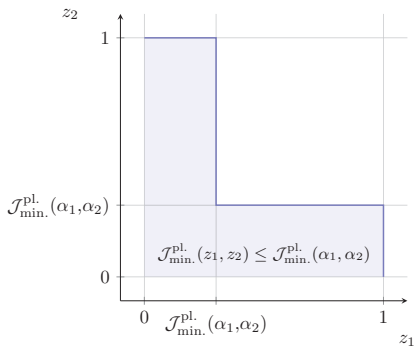


Figure 3.10: Integration rule for the minimum-based strong-independence Π-copula.

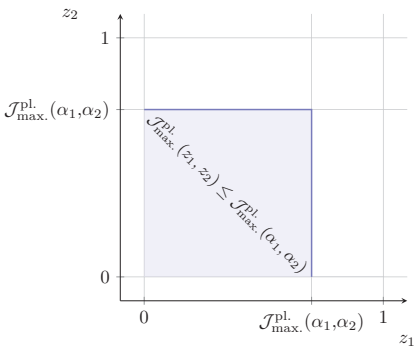


Figure 3.11: Integration rule for the maximum-based strong-independence Π-copula.

As an example, applying the integration rule to the product-based plausibility copula

$$\mathcal{J}^{\text{prod.pl.}}(\alpha_1, \ldots, \alpha_m) = \prod_{i=1}^{m} \alpha_i \tag{3.89}$$

for all $\alpha_1, \ldots, \alpha_m \in [0,1]$ yields the *product-based strong-independence* Π*-copula* $\mathcal{J}^{\text{prod.SI}}$ given by

$$\mathcal{J}^{\text{prod.SI}}(\alpha_1, \ldots, \alpha_m) = \left(\prod_{i=1}^{m} \alpha_i\right) \cdot \sum_{k=0}^{m-1} \frac{\left(-\log\left(\prod_{i=1}^{m} \alpha_i\right)\right)^k}{k!}, \tag{3.90}$$

for all $\alpha_1, \ldots, \alpha_m \in [0,1]$, see Fig. 3.12. Incidentally, the perhaps most popular technique

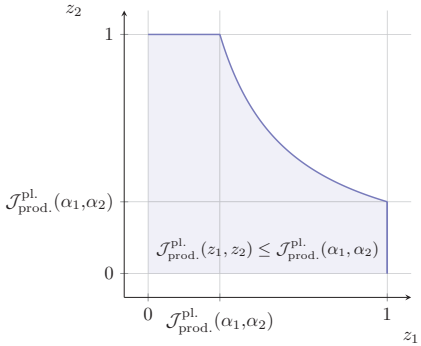


Figure 3.12: Integration rule for the product-based strong-independence Π-copula.

for combining independent p-values is Fisher's method [Fisher46] and its various extensions under different types of dependency [KostMcDermott02, Brown75]. Whereas the original method by Fisher is based on the χ^2-distribution, which needs to be approximated numerically, Jost[35] is able to find an explicit formula for this combination, which coincides precisely with Eq. (3.90).

In conclusion, the minimum-, maximum- and (presumably) product-based strong-independence Π-copulae may be used to model strong independence, and it cannot generally be said that one description is preferable over another. The 'best' choice is always context-specific. Numerical experiments suggest that all of the above Π-copulae produce joint possibility distributions that are as specific as possible—without violating the Principle of Representation—but all of them also include joint probability distributions in their credal sets, which are not in $\mathfrak{P}^{\text{SI}}_{\tilde{V}_1,\ldots,\tilde{V}_m}$.

Nevertheless, the minimum-based strong-independence Π-copula $\mathcal{J}^{\text{min.SI}}$ possesses certain advantageous properties listed in the following that make it the preferred choice in the remainder of this thesis.

Due to the monotonicity of $\mathcal{J}^{\text{min.SI}}$ with respect to all arguments, the joint possibility distribution produced under the minimum-based strong-independence Π-copula may also be re-written as

$$\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}(v_1,\ldots,v_m) = 1 - \left(1 - \min_{i=1,\ldots,m} \pi_{\tilde{V}_i}(v_i)\right)^m \tag{3.91}$$

for all $v_1 \in \mathbb{V}_1, \ldots, v_m \in \mathbb{V}_m$, and the superlevel sets of this joint distribution reduce to

$$\mathcal{C}^{\alpha}_{\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}} = \mathcal{C}^{1-\sqrt[m]{1-\alpha}}_{\pi_{\tilde{V}_1}} \times \ldots \times \mathcal{C}^{1-\sqrt[m]{1-\alpha}}_{\pi_{\tilde{V}_1}} \tag{3.92}$$

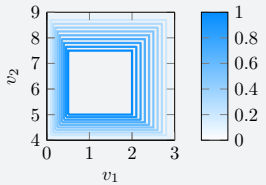
[35] http://www.loujost.com/StatisticsandPhysics/SignificanceLevels/CombiningPValues.htm (accessed on June 30th, 2021)

for all $\alpha \in [0, 1]$. The joint superlevel sets are a Cartesian product of the marginal superlevel sets. Moreover, if the marginal superlevel sets of $\pi_{\tilde{V}_i}$ are convex on all α-levels, then the joint superlevel sets of $\pi_{\tilde{V}_1,\ldots,\tilde{V}_m}$ are also convex, which is computationally desirable and shall prove to be useful in Chapter 5. This makes the minimum-based strong-independence Π-copula $\mathcal{J}^{\text{min.SI}}$ the preferred Π-copula when modeling strong independence. For the sake of brevity it is, henceforth, simply be called the SI-Π-copula $\mathcal{J}^{\text{SI}}$. Notice that these convexity observations also obtain for all other minimum-based Π-copulae, such as the unknown-interaction and the non-interactive Π-copula.

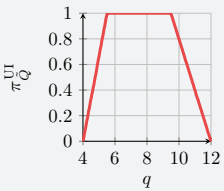
The following example illustrates how the various Π-copulae produce different joint possibility distributions and how these may then be extended.

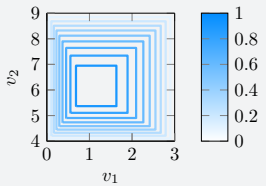
Example 7: Π**-Copulae**

Consider the two imprecise input variables $\tilde{V}_1 \sim \Delta(0, 1, 3)$ and $\tilde{V}_2 \sim \Delta(4, 6, 9)$ with triangular possibility distributions and the imprecise output variable $\tilde{Q} = \tilde{V}_1 + \tilde{V}_2$. Depending on the available dependency information, different joint possibility distributions of $\tilde{V}_1$ and $\tilde{V}_2$ are obtained and the resulting output possibility distribution of $\tilde{Q}$ changes accordingly. In the following figures, a contour plot of $\pi_{\tilde{V}_1,\tilde{V}_2}$ is shown on the left and $\pi_{\tilde{Q}}$ is shown on the right for some exemplary choices of the Π-copula.

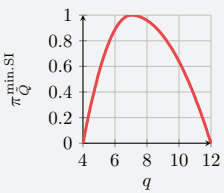


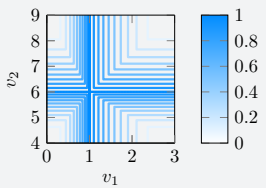
Unknown-Interaction Π-copula $\mathcal{J}^{\text{UI}}$



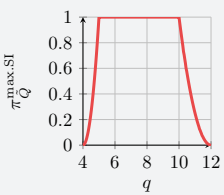


Minimum-Based Strong-Independence Π-copula $\mathcal{J}^{\text{min.SI}}$/ SI-Π-copula $\mathcal{J}^{\text{SI}}$





Maximum-Based Strong-Independence Π-copula $\mathcal{J}^{\text{max.SI}}$



In general, the presented Π-copulae generally loose much of the information contained in the original family of joint probability distributions, which leads to the heuristic recommendation that one should employ as few of these Π-copulae as possible when formulating a possibilistic model. If, for instance, the precise probability distributions of two independent imprecise variables are known, then it is not recommended to first perform, e.g., the Optimal P-Π-transform to obtain marginal possibility distributions and then produce the corresponding joint possibility distribution under the SI-Π-copula. Instead, one should first assemble the precise joint probability distribution and then perform, e.g., the Optimal P-Π-transform to obtain a joint possibility distribution. Both methods are admissible, but the latter, generally, yields more expressive results. Likewise, if three imprecise variables with available marginal possibility distributions are, e.g., to be joined under the assumption of unknown interaction, then there are four conceivable ways in order to accomplish this. They may be expressed as

$$\begin{aligned}
\pi^{(1)}_{\tilde{V}_1,\tilde{V}_2,\tilde{V}_3} &= \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_1}, \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_2}, \pi_{\tilde{V}_3}\right)\right), \\
\pi^{(2)}_{\tilde{V}_1,\tilde{V}_2,\tilde{V}_3} &= \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_2}, \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_1}, \pi_{\tilde{V}_3}\right)\right), \\
\pi^{(3)}_{\tilde{V}_1,\tilde{V}_2,\tilde{V}_3} &= \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_3}, \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_2}, \pi_{\tilde{V}_3}\right)\right) \text{ and} \\
\pi^{(4)}_{\tilde{V}_1,\tilde{V}_2,\tilde{V}_3} &= \mathcal{J}^{\text{UI}}\left(\pi_{\tilde{V}_1}, \pi_{\tilde{V}_2}, \pi_{\tilde{V}_3}\right).
\end{aligned} \tag{3.93}$$

Whereas the first three options require two applications of the unknown interaction Π-copula, the fourth option only requires one application. Therefore, the latter is generally expected to yield the most expressive results.

Of course, Proposition 36 also obtains in the case of strong independence. By not assuming any dependency information, it is universally applicable if only the quasi-vacuousness of all but one marginal possibility distributions is ensured. Since, in this case, the non-interactive Π-copula produces a more specific joint possibility distribution than the SI-Π-copula, it is expedient to employ the former whenever all marginal possibility distributions but one are quasi-vacuous—irrespective of possible dependency information.

A final observation, which is in line with the loss of expressiveness by Π-copulae, is that the two most important Π-copulae presented in this thesis, namely the SI- and the UI-Π-copula, produce joint distributions that converge to quasi-vacuous distributions

for large m. More precisely, interpreting these Π-copulae as a rescaling of the marginal possibility distributions as explained in Eq. (3.91), all non-zero joint possibilities converge pointwise to one because the minima of the marginal possibilities are rescaled by a strictly increasing function of m, see Figs. 3.13 and 3.14. If many marginal imprecise variables are present in a possilistic model, this becomes an argument for switching to interval calculus as an easier version of possibilistic calculus with minimal loss of expressiveness.

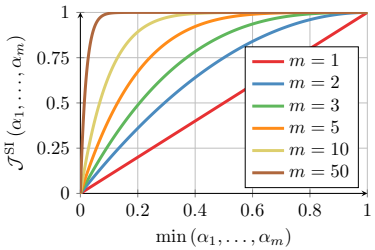


Figure 3.13: Marginal-to-joint possibility rescaling under SI-Π-copula for various values of m.

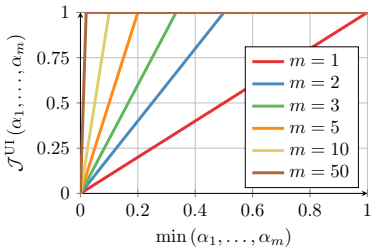


Figure 3.14: Marginal-to-joint possibility rescaling under UI-Π-copula for various values of m.

Further discussion regarding copulae in an IP context is given by Schmelzer [Schmelzer15a, Schmelzer15b], and the connections between probability and possibility copulae are also investigated by Gray et al. [GrayEtAl21].

Remark 39. The issue of measurability has not been addressed in this chapter, and, indeed, this point is of little concern. A thorough investigation of the measurability of the various possibility distributions cannot be postponed indefinitely and should at some point be considered; still, the purpose of this thesis is a discussion of possibilistic methods for applied reasoning with imprecise probabilities, but such highly theoretical investigations—potentially including many technical assumptions, which are, however, easily satisfied in most realistic scenarios—would seriously impede its readability. Therefore, a general recommendation that is expected to prevent the vast majority of potential issues is that, of course, any possibility distribution stated by a modeler should be measurable. Apart from this trivial requirement, no additional precautions are of essential importance, and the issue of measurability shall not be addressed in the remainder of this thesis.

Chapter 4

Possibilistic Statistics

> Truth is a compliment we give to our successful beliefs.
>
> Louis Menand, on William James' view of truth[36]

The description of imprecise probabilities is not the only quantitative interpretation of possibilities; in fact, many other ways of assigning degrees of possibilities have been proposed, and this section discusses some of these. In particular, it proposes two interpretations related to (frequentist) statistical inference. The aim is to show that these interpretations can be tightly linked to possibilistic descriptions of (imprecise) probabilities—in fact, the former can be derived from the latter—, and that possibility theory provides a powerful framework for statistical inference, which is easily established and involves mostly straightforward computations involving variants of the IP-Π-transform and, to a lesser extent, the Extension Principle.

Statistical inference typically revolves around the question of what may be learned about a model from data, particularly about the model's properties, mainly in the form of its parameters, and its future outputs. Yet, instead of trying to describe this model from the given information as done in Chapter 3, an experiment is performed, and data is collected and used as evidence. This clear emphasis on data and what to learn from them is the fundamental feature of statistical inference and enables a connection from measure theories, such as probability and possibility theory, to observations of the real world. In this sense, every quantitative theory of statistical inference constitutes a 'mathematical theory of evidence' [Shafer76].

The question of how to form and express one's belief about the (past, current or future) state of the world, or simply the properties and output of a model, requires some prior

[36] AoM Podcast #752: The Metaphysical Club, `https://www.artofmanliness.com/character/knowledge-of-men/podcast-752-the-metaphysical-club/` (accessed on November 15, 2021).

deliberations of a rather philosophical nature. The ongoing debate between Bayesians, frequentists, etc. shall not be repeated here; instead, the argument by Reid and Cox that

> "[...] it is unacceptable if a procedure yielding regions of high [belief] would, if used repeatedly, give systematically misleading conclusions"
>
> [ReidCox15, p. 295]

seems to be an agreeable starting point for most statisticians, though one may encounter various meanings and definitions of the term 'belief' in the literature. Reid and Cox speak of "probability in the sense of representing uncertain knowledge", which many, e.g. Shafer [Shafer76], would call a 'Bayesian belief'.

The fundamental Bayesian assumption is that all belief must adhere to the Kolmogorov axioms of probability theory; it must be described by probabilities. While it is not immediately apparent why one should wish to manipulate belief according to the rules of probability theory, Balch et al. even provide compelling arguments in the form of the False Confidence Theorem that Bayesian beliefs may sometimes fail to satisfy seemingly trivial requirements, such as the one in the above quote [BalchMartinFerson19].

Having already presented possibility theory as a framework for imprecise probabilities, which rejects the singular position of probability in measure theory (but acknowledges its saliency), it is only logical also to investigate possibilistic belief, i.e. belief adhering to the axioms of possibility theory. Indeed, the point of this chapter is to show that it is useful to consider (possibilistic) alternatives to Bayesian belief frameworks. This course of action also continues the axiomatic approach of this thesis and does not necessarily side with anyone in the philosophical debate, even though the following discussion bears a strong frequentist (and somewhat fiducial) flavor with many connections to the theory of confidence procedures.

To summarize, in this chapter, a theory of statistical inference based on possibility theory is developed that can be understood as a measure-theoretic basis for frequentist confidence procedures. It is strongly connected to inferential models [MartinLiu15]—in fact, the derived methodology will constitute special instances thereof—but relies on a slightly simpler notation at the expense of less generality. For this reason, much of the notation and terminology follows Martin et al. but is also consistent with many textbooks on mathematical statistics [Shao03]. Finally, the discussion is restricted to only certain aspects of statistical inference required to ultimately derive a theory of inference for dynamical systems in Chapter 6. Many topics, such as non-parametric inference, point estimation and decision-making, are not treated—despite their importance—and remain to be investigated.

4.1 Statistical Setup

The traditional setup in many statistical problems may be described as follows.

An experiment on the probability space $(\Omega, \Sigma, \mathrm{P})$ is performed, and $\omega \in \Omega$ is observed. Here, the data are considered to be realizations $\tilde{Q}(\omega) = q$ of the $\mathbb{Q}$-valued imprecise variable $\tilde{Q}$ with the pushforward probability distribution $\mathrm{P}_{\tilde{Q}}$, which is also called the *population*.

Formally, the population is *known* if the value of $\mathrm{P}(\tilde{Q} \in B) = \mathrm{P}_{\tilde{Q}}(B)$ may be specified precisely for all $B \in \mathbb{B}(\mathbb{Q})$. Then, $\mathrm{P}_{\tilde{Q}}$ is perfectly characterized. In any statistical setting, however, the population is assumed to be (at least partially) unknown. The goal of statistical inference is to learn $\mathrm{P}_{\tilde{Q}}$, or at least to deduce some of its properties, such as, e.g., means, moments, quantiles, etc.

To this end, a family of candidate probability distributions $\mathfrak{S} \subseteq \mathbb{P}(\mathbb{Q}, \mathbb{B}(\mathbb{Q}))$ is considered, to which the population is assumed to belong, i.e., $\mathrm{P}_{\tilde{Q}} \in \mathfrak{S}$. The following discussion assumes *parametric statistical models*, i.e.

$$\mathfrak{S} = \left\{ \mathrm{P}_{\tilde{Q}|\theta} \in \mathbb{P}(\mathbb{Q}, \mathbb{B}(\mathbb{Q})) : \theta \in \Theta \right\}. \tag{4.1}$$

The actual population $\mathrm{P}_{\tilde{Q}} = \mathrm{P}_{\tilde{Q}|\theta^*}$ is associated with the (unknown) *true parameter* value $\theta^* \in \Theta$ in the *parameter space* $\Theta \subseteq \mathbb{R}^{D_\Theta}$, which describes the space of possible parameter combinations, i.e., it describes all conceivable experimental setups. This justifies thinking of the true parameter $\theta^* = \hat{\theta}(\mathrm{P}_{\tilde{Q}})$ as the population-dependent output of the parameter function $\hat{\theta} : \mathbb{P}(\mathbb{Q}, \mathbb{B}(\mathbb{Q})) \to \Theta$, as done in some textbooks [Shao03]—similar to the definition of imprecise variables.

The goal of statistical inference is, in a certain sense, inverse to that of the descriptive problems considered in earlier chapters. Instead of predicting the values of $\tilde{Q}$ when θ^* is available, i.e. when the population is known, the general problem of statistical inference reduces to the problem of inferring θ^* from the observations $\tilde{Q} = q$.

Often, $\tilde{Q} = (\tilde{Q}_1, \ldots, \tilde{Q}_m)$ is composed of m observations $\tilde{Q}_1 \sim \mathrm{P}_{\tilde{Q}_1|\theta}, \ldots, \tilde{Q}_m \sim \mathrm{P}_{\tilde{Q}_m|\theta}$ whose interdependency is described by some P-copula C, i.e., the statistical model is simply given by

$$F_{\tilde{Q}|\theta} = C\left(F_{\tilde{Q}_1|\theta}, \ldots, F_{\tilde{Q}_m|\theta}\right). \tag{4.2}$$

If $C = C^{\mathrm{I}}$ and $\mathrm{P}_{\tilde{Q}_1|\theta} = \ldots = \mathrm{P}_{\tilde{Q}_m|\theta}$, the observations are said to be *independent and identically distributed* (iid).

This setup is not the most general formulation of *parametric inference*, but it suffices for the purposes of this thesis.

4.2 Inferential Models

It is seldom the case that θ^* may be inferred precisely and that all other values $\Theta \setminus \{\theta^*\}$ can be ruled out conclusively. By the nature of a statistical model, an inverse map $q \mapsto \theta^*$ from the observations $\tilde{Q} = q$ to the parameters does not exist—just as little as the knowledge of the population and/or θ^* would allow for a precise prediction of the next realization of $\tilde{Q}$. If two populations can, in principle, produce the same realizations, how should one be able to decide which one did?

Instead, one may ask questions in the form of hypotheses $\theta^* \in T$ for some set $T \in \mathbb{B}(\Theta)$. The answer to such questions may uniquely be 'yes' or 'no' but can only be stated with full confidence, i.e. without (even educated) guessing, if the population, and therefore θ^*, is known.

A fundamental idea of (frequentist) inference is to let $T = T(\tilde{Q})$ depend on the observations $\tilde{Q}$ in such a way that probabilities may be specified for the (in-)correct acceptance or rejection of the hypothesis that $\theta^* \in T(\tilde{Q})$. In this sense, frequentist inference constitutes an *a-priori* approach to statistical inference because such probabilities $\mathrm{P}(\theta^* \in T(\tilde{Q}))$ can only be computed before seeing the data; the *a-posteriori* probability $\mathrm{P}(\theta^* \in T(q))$ after observing $\tilde{Q} = q$ is either zero or one. That is, an a-priori success probability is associated with the strategy, or *procedure*, that generates $T(\tilde{Q})$ and must not be interpreted as the probability that $\theta^* \in T(q)$. Deciding to follow the proposed strategy only implies a certain probability of success.

This point of view is also the fundamental difference to Bayesian inference, where, broadly speaking, the unknown parameter is considered to be a Θ-valued imprecise variable $\tilde{\theta}$, whose (posterior) probability distribution $\mathrm{P}_{\tilde{\theta}|q}$ can be computed via Bayes' theorem in order to specify the probability of $\tilde{\theta} \in T$ given some prior, where T is a fixed set and independent of the data [Sullivan15].

Including not only the frequentist and the Bayesian approach but also many others, every map $\mathfrak{M}$ from $\mathfrak{S}$ and the observations $\tilde{Q} = q$ (and possible other information, such as, e.g., a prior) to a data-dependent capacity $\mathfrak{M}(\mathfrak{S}, q, \ldots) = \mathrm{M}_{\tilde{\theta}|q}$ on the measurable space $(\Theta, \mathbb{B}(\Theta))$ constitutes an *inferential model* (IM) in the sense of Martin and Liu [MartinLiu15]. The belief and disbelief in the hypothesis $\theta^* \in T$ for $T \in \mathbb{B}(\Theta)$ are then expressed by means of the values of $\mathrm{M}_{\tilde{\theta}|q}(T)$ and $\mathrm{M}_{\tilde{\theta}|q}(\neg T)$. Again, the restriction to hypotheses based on Borel sets $\mathbb{B}(\Theta)$ is unnecessarily restrictive but suffices for this thesis.

Such data-dependent capacities are not restricted to any specific shape. They may, in principle, be probability, possibility or necessity measures, but also upper/lower probabilities induced by random sets, belief or plausibility measures in the sense of Shafer, lower previsions, etc. This thesis is restricted to the investigation of IMs producing possibility measures, which have, in broad sense, been investigated by sev-

eral scholars [Fériet82, DuboisPrade86b, Smets90, DuboisPrade90, DuboisDencœux10, Cuzzolin13, BronevichRozenberg19, DuboisMoralPrade97, DuboisPrade98, Dubois06, MassonDencœux06, DuboisPrade16, LiuMartin20, HoseHanss20, HoseHanss21c]—and are now to be connected to the previously established theory of possibilistic descriptions of imprecise probabilities.

The details of the construction of IMs as proposed by Martin et al., including the association, prediction and combination step, shall not be discussed, since the construction methods advocated here follow a slightly different route.

4.2.1 Fundamental Properties

Having found a general definition of IMs, it remains to be established what distinguishes 'good' from 'bad' statistical inference and what constitutes functional IMs. This question requires a set of fundamental principles on which to base such properties—similar to Section 2.3.2.1. These can, e.g., be found in the original literature on IMs [MartinLiu15, LiuMartin20, Martin21] but also in the works of others [ReidCox15]. Some intuition regarding the prudence of these principles shall be provided throughout the remainder of this chapter.

4.2.1.1 Validity

Validity addresses the issue of calibrating an IM $\mathfrak{M}$ in such a way that the error rate when rejecting and accepting hypotheses in the proposed manner is well-controlled.

Consider the hypothesis of θ^* pertaining to $T \in \mathbb{B}(\Theta)$. Martin et al. speak of validity if all possible a-priori CPFs of $\mathrm{M}_{\hat{\theta}|\tilde{Q}}(T)$ stochastically dominate the uniform distribution, i.e. if

$$\sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta} \left(\{ q \in \mathbb{Q} : \mathrm{M}_{\hat{\theta}|q}(T) \leq \alpha \} \right) \leq \alpha \tag{4.3}$$

for all $\alpha \in [0, 1]$ and for all $T \in \mathbb{B}(\Theta)$. In words, the (upper) probability of the capacity assigning a certain value to the hypothesis must never be greater than that value. This tells a statistician that, if their IM yielded $\alpha = \mathrm{M}_{\hat{\theta}|q}(T)$ for the observation $\tilde{Q} = q$, all population candidates included by T had a maximum a-priori probability of α to produce this result.

This definition of validity, which may seem unintuitive at first glance, guarantees that $\mathfrak{M}$ provides practical information because it avoids 'systematically misleading conclusions'. If a valid IM returns a low value $\alpha = \mathrm{M}_{\hat{\theta}|q}(T)$, one should tend to reject the hypothesis $\theta^* \in T$ because it is guaranteed to have had a low a-priori probability of occurring. If one repeatedly accepted hypotheses with $\alpha \ll 1$, one would have to expect to be wrong with probability $1 - \alpha$, which is then close to one—an unacceptable systematic error in the words of Reid and Cox.

Balch et al. show that IMs producing additive belief measures, i.e. Bayesian posterior probability distributions, are never valid if the prior is not genuine, i.e. perfectly calibrated. This result is captured in their False Confidence Theorem [BalchMartinFerson19].

As for all non-self-dual monotone measures, validity does not necessarily imply that high values of α should immediately lead to an acceptance of the hypothesis. Rather, one should accept the hypothesis if one is willing to reject its alternative $\theta^* \notin T$ according to the above considerations, which is expressed by the dual capacity.

Balch rephrases the principle of validity by requiring that "inferential belief be represented in terms of Neyman-Pearson confidence" [Balch12, p. 1017]. He, furthermore, acknowledges that it "is not enough to define a unique method of inference" [ibid.] and that an additional measure of precision is needed. This observation leads to the following concept of efficiency.

4.2.1.2 Efficiency

The concept of *efficiency* establishes a partial order among IMs and can be used to identify when one IM is preferable over a second one.

Subject to the validity constraint, more precise evaluations of a hypothesis are clearly preferable to wider ones. For instance, an IM that always produces the vacuous measure $\mathrm{M}_{\hat{\theta}|q}$ with $\mathrm{M}_{\hat{\theta}|q}(T) = 1$ for all $q \in \mathbb{Q}$ and all $T \in \mathbb{B}(\Theta)$ is certainly valid but little useful because it never leads to the acceptance or rejection of any hypothesis.

Consequently, a simple and sufficient (but not necessary) criterion for better efficiency is

$$\mathrm{M}^{(1)}_{\hat{\theta}|q}(T) \leq \mathrm{M}^{(2)}_{\hat{\theta}|q}(T) \tag{4.4}$$

for all $q \in \mathbb{Q}$ and all $T \in \mathbb{B}(\Theta)$; however, this requirement is quite strong, and a slightly weaker version is proposed.

Paraphrasing Martin and Liu, the IM $\mathfrak{M}^{(1)}$ is said to be *more efficient with respect to some* $T \in \mathbb{B}(\Theta)$ than the IM $\mathfrak{M}^{(2)}$ if

$$\sup_{\theta \in T} \mathrm{P}_{Q|\theta}\left(\{q \in \mathbb{Q} : \mathrm{M}^{(1)}_{\hat{\theta}|q}(T) \leq \alpha\}\right) \geq \sup_{\theta \in T} \mathrm{P}_{Q|\theta}\left(\{q \in \mathbb{Q} : \mathrm{M}^{(2)}_{\hat{\theta}|q}(T) \leq \alpha\}\right) \tag{4.5}$$

for all $\alpha \in [0, 1]$. In words, $\mathfrak{M}^{(1)}$ is more useful than $\mathfrak{M}^{(2)}$ because the former allows for more informed evaluations of the hypothesis $\theta^* \in T$ and should clearly be preferred over the latter. If $\mathfrak{M}^{(1)}$ is more efficient than the IM $\mathfrak{M}^{(2)}$ with respect to all $T = \{\theta\}$ for $\theta \in \Theta$, it is said to be *element-wise more efficient*.

Additional intuition concerning the appropriateness of these definitions shall be provided in Section 4.2.2.2.

4.2.2 Possibilistic Inferential Models

The above-mentioned concepts of IMs are strongly related to those of possibility theory.

- Consistency and validity calibrate the obtained descriptions of the family of probabilities and the statistical model, respectively. These concepts guarantee that information, e.g. about upper or a-priori probabilities, is encoded robustly in these descriptions.
- Specificity and efficiency establish a partial order among possibility distributions and IMs, respectively by ranking them according to their expressiveness. Subject to the consistency and validity constraints, this makes certain possibilistic descriptions or IMs preferable to others, thus avoiding over-conservatism. For instance, both frameworks allow for the existence of a vacuous measure that is 'never wrong' but contains no information.

Not least due to these similarities, it makes sense to also investigate a special class of IMs, namely *possibilistic inferential models* (Π-IMs), written as $\gamma_{\hat{\theta}|\tilde{Q}}$. Depending on the observation $\tilde{Q} = q$, they produce an elementary possibility function $\gamma_{\hat{\theta}|q} : \Theta \to [0, 1]$, which, in turn, induces a data-dependent capacity, namely the possibility measure $\overline{\Gamma}_{\hat{\theta}|q} : \mathbb{B}(\Theta) \to [0, 1]$. Here, the normality requirement is dropped. That is, in principle, $\gamma_{\hat{\theta}|q}$ need only be an elementary plausibility function; the validity property, however, requires at least some of them to be normal. The issue of sub-normality is also revisited further below.

In order to avoid confusion with elementary possibility and/or plausibility functions of imprecise variables, $\gamma_{\hat{\theta}|q}$ shall, henceforth, be called an *elementary confidence function*. The corresponding possibility measure $\overline{\Gamma}_{\hat{\theta}|q}$ is then called an *upper confidence measure* and the corresponding necessity measure $\underline{\Gamma}_{\hat{\theta}|q}$ a *lower confidence measure*. Since the information content is the same, any of these objects may be referred to as a *confidence distribution*.[37] If a clearer distinction between the various types of possibility distributions is deemed to be necessary, possibility distributions of imprecise variables, as considered in the previous chapters, are also called *(IP-)description/descriptive distributions*.

The difference between the two types of distributions is straightforward. Whereas IP-description distributions describe the possible set of probability distributions of an imprecise variable, confidence distributions grade the possible values of an unknown but precise parameter.

In this manner, even though (possibilistic) IMs constitute an inherently frequentist approach to statistical inference, it is still possible to specify a distribution of the unknown parameter. The fact that this is not a probability distribution only requires an independent investigation

[37] The proposed definition of a confidence distribution is different from other definitions [ClopperPearson34, Balch12], but related.

of the appropriate calculus, which turns out to be mathematically equivalent to that of IP-description distributions.

As for all possibility distributions, the lower and upper confidence measures must satisfy either $\underline{\Gamma}_{\hat{\theta}|q}(T) = 0$ or $\overline{\Gamma}_{\hat{\theta}|q}(T) = 1$. That is, the lower and upper confidence values of the hypothesis $\theta^* \in T$ are given by either

$$\left(\underline{\Gamma}_{\hat{\theta}|q}(T) = 0, \overline{\Gamma}_{\hat{\theta}|q}(T) = \alpha\right) \qquad \text{or} \qquad \left(\underline{\Gamma}_{\hat{\theta}|q}(T) = \beta, \overline{\Gamma}_{\hat{\theta}|q}(T) = 1\right) \tag{4.6}$$

for $\alpha, \beta \in [0, 1]$. In the first case, a low upper confidence value α obtained from a valid Π-IM should lead to a rejection of the hypothesis $\theta^* \in T$; in the second case, a high lower confidence value β should favor its acceptance, as discussed previously. A high α or a low β, however, should lead to neither a rejection nor an acceptance. In such cases, the evidence simply does not support these inferences.

The additional feature of Π-IMs, as opposed to classical IMs, is that they are entirely *consonant*[38] and do not exhibit any internal conflict. For instance, in the first case, the corresponding lower confidence value implies no evidence in favor of the hypothesis by a lower confidence value of $\underline{\Gamma}_{\hat{\theta}|q}(T) = 0$ because the confidence values of the alternative hypothesis $\theta^* \notin T$ read

$$\underline{\Gamma}_{\hat{\theta}|q}(\neg T) = 1 - \overline{\Gamma}_{\hat{\theta}|q}(T) = 1 - \alpha \qquad \text{and} \qquad \overline{\Gamma}_{\hat{\theta}|q}(\neg T) = 1 - \underline{\Gamma}_{\hat{\theta}|q}(T) = 1, \tag{4.7}$$

That is, a Π-IM "can be described as pointing in a single direction; it is heterogeneous only in that it varies in the precision of its focus" [Shafer76, p. 219]. With some reason, Shafer even goes as far as claiming that consonance is natural in statistical inference because inferential evidence should be regarded as consonant, which further justifies the investigation of Π-IMs.

On a final note, it is possible to evaluate arbitrary hypotheses of the form $\theta^* \in T$ for any set $T \in \mathbb{B}(\Theta)$; however, by the argument that possibility measures are technically only concerned with the sub- and superlevel sets of the corresponding elementary possibility function, meaningful hypotheses about θ^* will typically only be based on the level sets of the confidence distribution, i.e.

$$T = T(q) = \mathcal{S}^{\alpha}_{\gamma_{\hat{\theta}|q}} \qquad \text{or} \qquad T = T(q) = \mathcal{C}^{\alpha}_{\gamma_{\hat{\theta}|q}} \tag{4.8}$$

for $\alpha \in [0, 1]$ and $q \in \mathbb{Q}$.

In the following, the inferential concepts of validity, efficiency and likelihood are investigated for Π-IMs.

[38] In the Dempster-Shafer Theory of Evidence, the possibility and necessity measure are called 'consonant' plausibility and belief functions, respectively [Shafer76].

4.2.2.1 Confidence and Validity

Validity bears a striking resemblance to consistency, and, indeed, a criterion that appears to be very similar to Lemma 1 based on elementary confidence can be deduced.

Lemma 40. *A* Π*-IM* $\gamma_{\hat{\theta}|\tilde{Q}}$ *is valid if and only if*

$$\mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}\right) \leq \alpha \tag{4.9}$$

for all $\theta \in \Theta$ *and all* $\alpha \in [0,1]$.

Proof. Let $\alpha \in [0,1]$. Equation (4.3) follows from Eq. (4.9) due to the inequality

$$\begin{aligned}\sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \overline{\Gamma}_{\hat{\theta}|q}(T) \leq \alpha\}\right) &= \sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \sup_{\theta' \in T} \gamma_{\hat{\theta}|q}(\theta') \leq \alpha\}\right) \\ &\leq \sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}\right) \\ &\leq \alpha,\end{aligned}$$

which obtains for all $T \in \mathbb{B}(\Theta)$. Conversely, Eq. (4.9) follows from Eq. (4.3) if $T = \{\theta\}$ for any $\theta \in \Theta$ by considering

$$\begin{aligned}\alpha &\geq \sup_{\theta' \in T} \mathrm{P}_{\tilde{Q}|\theta'}\left(\{q \in \mathbb{Q} : \overline{\Gamma}_{\hat{\theta}|\tilde{Q}}(T) \leq \alpha\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \overline{\Gamma}_{\hat{\theta}|\tilde{Q}}(\{\theta\}) \leq \alpha\}\right) = \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}\right).\end{aligned}$$

□

Martin describes ways of constructing (general) IMs from confidence procedures and hypothesis tests [Martin21], which indicates a close connection between IMs and frequentist inference. Upon closer inspection of Eq. (4.9), it becomes evident that Π-IMs, too, are directly linked to these well-known frequentist concepts.

Confidence Sets The superlevel sets $\mathcal{C}^{\alpha}_{\gamma_{\hat{\theta}|q}}$ of a confidence distribution $\gamma_{\hat{\theta}|q}$ are *confidence sets* [Shao03] of the unknown parameter because they possess an a-priori *coverage probability/significance level* of $1-\alpha$, which is verified by considering

$$\begin{aligned}\inf_{\theta \in \Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \theta \in \mathcal{C}^{\alpha}_{\gamma_{\hat{\theta}|q}}\}\right) &= \inf_{\theta \in \Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) > \alpha\}\right) \\ &= 1 - \sup_{\theta \in \Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}\right) \\ &\geq 1 - \alpha\end{aligned} \tag{4.10}$$

for all $\alpha \in [0,1]$, where the last inequality follows from the validity criterion in Eq. (4.9). These sets are also the smallest sets with a lower confidence greater than or equal to $1-\alpha$.

If any of these confidence sets happen to be empty, then a subnormal confidence distribution has been obtained. Contrary to IP-description distributions, which then lose their coherence property and do not avoid sure loss anymore, this is not unheard of for confidence distributions [BallBrittonO'neill02]. Even though a universally accepted interpretation and treatment of empty confidence sets has not been agreed upon, this issue does not appear to yield any practical problems.

Hypothesis Testing and P-Values A hypothesis testing problem may be described as finding an observation-based rule for deciding whether $\theta^* \in T$ for some $T \in \mathbb{B}(\Theta)$ is true or not [Shao03]. For this purpose, two hypotheses are formulated, namely

$$H_0 : \theta^* \in T \qquad \text{vs.} \qquad H_1 : \theta^* \notin T. \tag{4.11}$$

The former is called the *null hypothesis*, whereas the latter is the *alternative hypothesis*, and the rule for deciding between either of them is based on an imprecise variable, the *test statistic* $\tilde{S} : \mathbb{Q} \to \{0, 1\}$, in a straight-forward manner: After $\tilde{Q} = q$ is observed, the hypothesis $H_{\tilde{S}(q)}$ is accepted. Under such a test, two types of errors can occur: The *type I error* when H_0 is true but rejected in favor of H_1, or the *type II error* when H_0 is false yet accepted instead of H_1. Generally, it is not possible to simultaneously control both types of errors when formulating a testing procedure, and in practice, a test is commonly designed such that the probability of a type I error is bounded by some *level of significance* $\alpha \in [0, 1]$.
Such a test can easily be based on a confidence distribution by defining the test statistic $\tilde{S}$ via

$$\tilde{S}(q) = \begin{cases} 0 & \text{if } \overline{\Gamma}_{\hat{\theta}|q}(T) > \alpha \text{ and} \\ 1 & \text{if } \overline{\Gamma}_{\hat{\theta}|q}(T) \leq \alpha \end{cases} \tag{4.12}$$

because, then, the type I error is bounded by the significance level $\alpha \in [0, 1]$ via

$$\begin{aligned} &\sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta} \left(\{q \in \mathbb{Q} : \tilde{S}(q) = 1\} \right) = \sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta} \left(\{q \in \mathbb{Q} : \overline{\Gamma}_{\hat{\theta}|q}(T) \leq \alpha\} \right) \\ =& \sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta} \left(\{q \in \mathbb{Q} : \sup_{\theta \in T} \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\} \right) \leq \sup_{\theta \in T} \mathrm{P}_{\tilde{Q}|\theta} \left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\} \right) \leq \alpha, \end{aligned} \tag{4.13}$$

where the last inequality follows from Eq. (4.9). By this observation, $\overline{\Gamma}_{\hat{\theta}|q}(T)$ fulfills the definition of a *p-value*, the upper probability of obtaining $\tilde{S} = 1$ under the condition that the null hypothesis holds in reality: If this value is small, an event with low a-priori probability must have occurred in order for the null hypothesis to actually be true. Therefore, the null hypothesis is accepted if the p-value is strictly greater than the level of significance, i.e. if $\overline{\Gamma}_{\hat{\theta}|q}(T) > \alpha$. If $T = \{\theta\}$ for some $\theta \in \Theta$, the elementary confidence $\gamma_{\hat{\theta}|q}(\theta)$ is simply the *elementary p-value* of θ with respect to the hypothesis $H_0 : \theta^* = \theta$ given the observation $\tilde{Q} = q$. To summarize, validity as defined in Eq. (4.9) states that all p-values

must stochastically dominate the uniform distribution. For the probability of the type II error, also called the *power* of the hypothesis test,

$$\begin{aligned}\sup_{\theta\notin T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\tilde{S}(q)=0\}\right) &= \sup_{\theta\notin T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\overline{\Gamma}_{\hat{\theta}|q}(T)>\alpha\}\right)\\ &= \sup_{\theta\notin T} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\sup_{\theta'\in T}\gamma_{\hat{\theta}|q}(\theta')>\alpha\}\right),\end{aligned} \tag{4.14}$$

one cannot find an equivalent general bound for Π-IMs.

The advantage of confidence distributions over confidence sets and hypothesis tests is that its information content is richer, and a significance level need not be fixed in advance. Instead, one may reason, e.g., with all confidence sets or p-values at once as explained in the remainder of this chapter.

4.2.2.2 Confidence and Efficiency

Efficiency is the inferential concept that is similar to the descriptive concept of specificity. It allows to compare two Π-IMs $\gamma^{(1)}_{\hat{\theta}|\tilde{Q}}$ and $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$ associated with the same statistical model with respect to their expressiveness.

Re-expressing the former definition in Eq. (4.5) for Π-IMs, the Π-IM $\gamma^{(1)}_{\hat{\theta}|\tilde{Q}}$ is said to be element-wise more efficient than the Π-IM $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$ if

$$\sup_{\theta\in\Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\gamma^{(1)}_{\hat{\theta}|q}(\theta)\leq\alpha\}\right) \geq \sup_{\theta\in\Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\gamma^{(2)}_{\hat{\theta}|q}(\theta)\leq\alpha\}\right) \tag{4.15}$$

for all $\alpha\in[0,1]$.

Similar to credal subsets, it follows immediately that, if the Π-IM $\gamma^{(1)}_{\hat{\theta}|\tilde{Q}}$ is valid and more efficient than $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$, then $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$ is also valid. However, the former should be preferred as it produces more expressive confidence bounds.

Since (element-wise) efficiency is not as sharp of a concept as specificity, it is not easy to find alternative, more tangible formulations. By the validity property, both expressions on the left- and the right-hand side in Eq. (4.15) are bounded from above by α. 'Good' (possibilistic) IMs aim at making the confidence distributions as efficient as possible, which translates to formulating hypothesis tests with desirable statistical properties that make the p-value 'as uniformly distributed as possible' and formulating confidence procedures that yield as tight as possible confidence sets.

Generally, it is not possible to find a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ that achieves

$$\mathrm{P}_{\tilde{Q}|\theta}\left(\{q\in\mathbb{Q}:\gamma_{\hat{\theta}|q}(\theta)\leq\alpha\}\right)=\alpha \tag{4.16}$$

for all $\alpha\in[0,1]$ and all $\theta\in\Theta$. Instead, a Π-IM is said to be *(element-wise) maximally efficient* if the supremum of all produced elementary p-values is uniformly distributed, i.e.

if

$$\sup_{\theta\in\Theta} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}\right) = \alpha \tag{4.17}$$

for all $\alpha \in [0,1]$. In this formulation, maximal efficiency naturally entails validity, and the goal of statistical inference reduces to finding a maximally specific Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$. Maximal efficiency is, however, not a unique property. That is, different Π-IMs may simultaneously be maximally efficient with respect to the same statistical model.

Some guidance on how to choose a unique maximally efficient Π-IM may be drawn from the likelihood principle.

4.2.2.3 Confidence and Likelihood

Similar to IP-description distributions, a confidence distribution produced by a Π-IM also establishes a plausibility order on Θ. If the p-values of $\theta, \theta' \in \Theta$ exhibit the relationship $\gamma_{\hat{\theta}|q}(\theta) \leq \gamma_{\hat{\theta}|q}(\theta')$ for some $q \in \mathbb{Q}$, one would be inclined to reject the hypothesis $\theta^* = \theta$ before rejecting $\theta^* = \theta'$, which is consistent with saying that one finds θ' more plausible than θ under the given data. Indeed, Martin gives the somewhat non-traditional interpretations that p-values, i.e. elementary confidence values, are "a measure of how plausible the null hypothesis is" and that confidence sets, i.e., the superlevel sets of a confidence distribution, are "a set of sufficiently plausible parameter values" [Martin21, p. 2].

Conversely, suppose a statistical model $\mathrm{P}_{\tilde{Q}|\theta}$ be induced by a parameter-dependent probability mass (density) function $p_{\tilde{Q}|\theta} : \Omega \to \mathbb{R}$ in the discrete (continuous) case, and the value $\tilde{Q} = q$ is observed. The function $\ell_{\hat{\theta}|q} : \Theta \to [0, \infty)$ defined by

$$\ell_{\hat{\theta}|q}(\theta) = p_\theta(q) \tag{4.18}$$

for all $\theta \in \Theta$ was coined the *likelihood function* by Fisher; yet, early on it was also known as the *relative plausibility function*. It is widely agreed that this function establishes a very sensible plausibility order on Θ. For instance, in his seminal book, Shafer writes that "we feel that q renders $\theta \in \Theta$ more plausible than $\theta' \in \Theta$ whenever $p_\theta(q) > p_{\theta'}(q)$. It is but a short step from this intuition to the idea that q should lend plausibility to a singleton"[39] [Shafer76, p. 238]. The idea of this plausibility order can, furthermore, be found in *likelihood-ratio tests* or the in the *law of likelihood* [Shao03].

In conclusion, both the confidence distribution and the likelihood function provide plausibility orders on the parameter space Θ, respectively. The lower the plausibility or likelihood of some hypothesis is, the less one expects it to be true. As they do so with equally good reasons, it would be desirable to merge both plausibility orders.[40]

[39] In this quote, the mathematical notation has been altered in order to fit the notation of this thesis.

[40] In personal communications, Ryan Martin also affirmed that the plausibility contour of the data-dependent capacity should, in the best case, be commensurate with the likelihood.

The simplest way to achieve this is by considering the Π-IMs that return precisely the likelihood function as a confidence distribution. In fact, one of the first interpretations of possibilities in a statistical context states that likelihood functions ought to be viewed as elementary possibility functions [DuboisMoralPrade97, Aickin00, DuboisDenœux10, Denoeux14, Cattaneo17], a procedure that was already proposed by Shafer [Shafer76, Ch. 10]. As mentioned earlier, he observes that consonance is natural in statistical inference, which is substantiated by Wassermann and Denœux, who—in contrast to Shafer—also provide axiomatic justifications: Wassermann argues in favor of likelihoods as an elementary belief based on the axioms that a flat likelihood function should lead to a vacuous belief function, that increasing the likelihood in one point should not decrease the belief in that same point, and that belief-function inference should be consistent with Bayesian inference when a Bayesian prior is available [Wasserman90]. Similarly, Denœux bases his justification on the likelihood principle, compatibility with Bayesian updating, and a principle of least commitment[41] [Denoeux14].

These arguments all seem compelling, but, unfortunately, Π-IMs that produce likelihood functions as confidence distributions do not generally exhibit validity. To see this, consider the statistical model $\mathrm{P}_{\tilde{Q}|\theta}$ of the $\{0,1\}$-valued imprecise variable $\tilde{Q}$, given by the probability mass function

$$p_{\tilde{Q}|\theta}(q) = \begin{cases} 1-\theta & \text{if } q = 0 \text{ and} \\ \theta & \text{if } q = 1, \end{cases} \tag{4.19}$$

which also constitutes the likelihood function $\ell_{\hat{\theta}|q}$. That is, the proposed likelihood Π-IM produces the confidence distributions $\gamma_{\hat{\theta}|q}(\theta) = \ell_{\hat{\theta}|q}(\theta) = p_{\tilde{Q}|\theta}(q)$ for all $\theta \in [0,1]$ and all $q \in \{0,1\}$. But, letting, e.g., $\theta = \frac{1}{2}$, which implies that $p_{\tilde{Q}|\theta}(q) = \frac{1}{2}$ for both $q = 0$ and $q = 1$, one finds that

$$\begin{aligned} \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \gamma_{q|\theta}(\theta) \leq \tfrac{1}{2}\}\right) &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \mathbb{Q} : \ell_{q|\theta}(\theta) \leq \tfrac{1}{2}\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{q \in \{0,1\} : p_{\tilde{Q}|\theta}(q) \leq \tfrac{1}{2}\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{0,1\}\right) = 1 > \tfrac{1}{2}, \end{aligned} \tag{4.20}$$

which directly violates Lemma 40. The proposed Π-IM is not valid.[42] Moreover, using the un-scaled likelihood function $\ell_{\hat{\theta}|q}(\theta) = p_{\tilde{Q}|\theta}(q)$ may produce (inadmissible) supernormal elementary confidence functions if $\mathbb{Q}$ is continuous.

The latter issue can, in principle, be obviated by interpreting only relative likelihoods

$$\Lambda_{\hat{\theta}|q}(\theta) = c(q) \cdot p_{\tilde{Q}|\theta}(q) \tag{4.21}$$

as elementary confidence for $\theta \in \Theta$ given $\tilde{Q} = q$. Therein, the normalization constant $c(q) = (\sup_{\theta\in\Theta} p_{\tilde{Q}|\theta}(q))^{-1}$ is given by the inverse of the maximum likelihood and

[41] The least commitment principle by Denœux may be thought of as a principle of minimum specificity or expressiveness.

[42] A valid Π-IM for this example will be discussed in Ex. 8 further below.

depends on q only. In fact, Savage and others repeatedly emphasize that the likelihood function is only defined up to a multiplicative constant, and parameter likelihoods for different data may only be compared by ratios, not by their absolute values, which guarantees comparability for different pieces of evidence [Shafer76, DuboisMoralPrade97, Cattaneo17].

In principle, letting $\gamma_{\hat{\theta}|q}(\theta) = \Lambda_{\hat{\theta}|q}(\theta)$ for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$ does, indeed, specify a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$. Nevertheless, this Π-IM is still not guaranteed to be valid. For instance, in the above counter-example, the normalization constant is $c(q) = 1$ both in the case of observing $q = 0$ and in the case of observing $q = 1$. It must be conceded that (relative) likelihood by itself does not imply validity.

To conclude this section, a very weak Π-IM principle for achieving compatibility between the plausibility orders of the confidence distribution and the likelihood function is formulated, which is inspired by the role of the elementary plausibility function in the IP-Π-transform: The confidence distributions produced by a Π-IM ought to be plausibility-conform to the likelihood function, i.e., it must be *likelihood-conform*. However, this principle does not appear to be as easily satisfiable as the Principle of Plausibility in the IP-Π-transform, see, e.g., Ex. 9.

Contrary to Wassermann and Denœux, this thesis does, however, not argue that the principles of validity, efficiency and likelihood conformity only permit Π-IMs. These principles merely serve as a guideline—just like the principles of quantitative possibility theory.

The following example discusses the principles of validity, efficiency and likelihood conformity for the most basic statistical model in detail.

Example 8: Bernoulli Π-IM

Similar to above, consider the statistical model of a Bernoulli experiment, e.g. a coin toss, given by the probability mass function

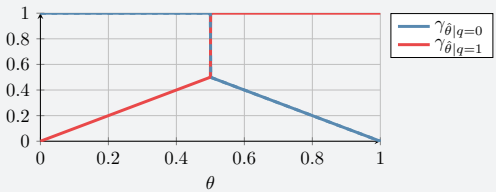
$$p_{\tilde{Q}|\theta}(q) = \begin{cases} 1-\theta & \text{if } q = 0 \text{ and} \\ \theta & \text{if } q = 1, \end{cases}$$

where $\tilde{Q} = 0$ for 'tails' and $\tilde{Q} = 1$ for 'heads' and $\theta \in [0, 1]$ denotes the unknown probability of observing the latter. What can be learned from one coin toss only?

A possibilistic IM $\gamma_{\hat{\theta}|\tilde{Q}}$ is given by the values of $\gamma_{\hat{\theta}|q}(\theta)$ in the following table.

$\gamma_{\hat{\theta}\|q}(\theta)$	$\theta \leq \frac{1}{2}$	$\theta > \frac{1}{2}$
$q = 0$	1	$1 - \theta$
$q = 1$	θ	1

If a single experiment is performed (the coin is tossed once) and $\tilde{Q} = q$ is observed, the Π-IM produces the confidence distributions shown below.



This Π-IM is valid and likelihood-conform but not maximally specific.

Validity can be checked via Lemma 40. For this, let $\theta \in [0,1]$, $\alpha \in [0,1]$ and denote by

$$B(\theta, \alpha) = \{q \in \{0,1\} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}$$

the corresponding set-valued argument in Eq. (4.9), whose composition is detailed in the following.

a) If $\theta \leq \frac{1}{2}$, then $0 \in B(\theta, \alpha)$ if and only if $\alpha = 1$, and $1 \in B(\theta, \alpha)$ if and only if $\alpha \geq \theta$. Therefore,

$$\mathrm{P}_{\tilde{Q}|\theta}(B(\theta, \alpha)) = \begin{cases} 0 & \text{if } \alpha \in [0, \theta), \\ \theta & \text{if } \alpha \in [\theta, 1) \text{ and} \\ 1 & \text{if } \alpha = 1. \end{cases}$$

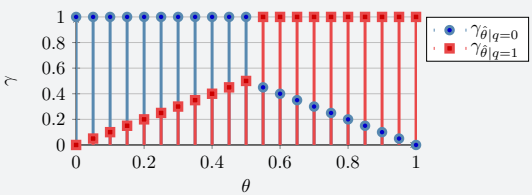
b) If $\theta > \frac{1}{2}$, then $1 \in B(\theta, \alpha)$ if and only if $\alpha = 1$, and $0 \in B(\theta, \alpha)$ if and only if $\alpha \geq 1 - \theta$. Therefore,

$$\mathrm{P}_{\tilde{Q}|\theta}(B(\theta, \alpha)) = \begin{cases} 0 & \text{if } \alpha \in [0, 1-\theta), \\ 1-\theta & \text{if } \alpha \in [1-\theta, 1) \text{ and} \\ 1 & \text{if } \alpha = 1. \end{cases}$$

In conclusion $\mathrm{P}_{\tilde{Q}|\theta}(B(\theta, \alpha)) \leq \alpha$ for all $\theta, \alpha \in [0,1]$, and the validity of the proposed Π-IM is shown under Lemma 40.

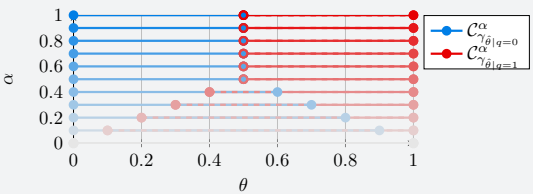
This admits a two-fold interpretation of the obtained confidence distributions.

- The elementary p-values for $\theta \in [0,1]$ are given by $\gamma_{\hat{\theta}|q}(\theta)$ as depicted below.



For instance, the elementary p-values of $\theta = 1$ are given by $\gamma_{\hat{\theta}|q=0}(1) = 0$ and by $\gamma_{\hat{\theta}|q=1}(1) = 1$, depending on the observations $q \in \{0, 1\}$. This is commensurate with intuition: If 'heads' has been observed, it is certainly plausible that the probablity of 'heads' is one; conversely, if 'tails' has been observed, it is entirely implausible for 'heads' to have probability one, i.e. to be (P-almost) sure to occur. The statistical model does not allow for this possibility, and this hypothesis should certainly be rejected.

- The confidence sets for $\alpha \in [0, 1]$ are given by the superlevel sets of $\gamma_{\hat{\theta}|q}$, which are illustrated in the following.



The 100%-confidence sets

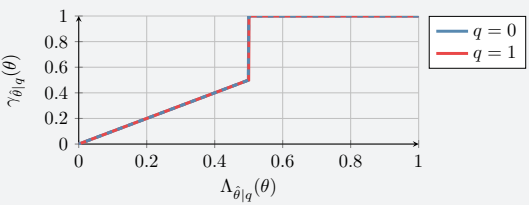
$$\mathcal{C}^{0}_{\gamma_{\hat{\theta}|q=0}} = \mathcal{C}^{0}_{\gamma_{\hat{\theta}|q=1}} = [0, 1]$$

coincide; however, e.g. the 50%-confidence sets

$$\mathcal{C}^{\frac{1}{2}}_{\gamma_{\hat{\theta}|q=0}} = [\frac{1}{2}, 1] \qquad \text{and} \qquad \mathcal{C}^{\frac{1}{2}}_{\gamma_{\hat{\theta}|q=1}} = [0, \frac{1}{2}]$$

only intersect at $\theta = \frac{1}{2}$.

Likelihood Conformity is best inspected visually, see below.



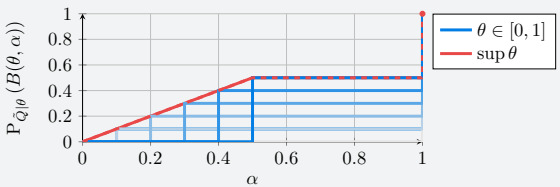
Both in the case $q = 0$ and $q = 1$, the elementary confidence functions are increasing functions of the likelihood, which demonstrates likelihood conformity.

Maximal Efficiency is not generally achieved. Only for $\alpha \in [0, \frac{1}{2}]$ or $\alpha = 1$, the equality in Eq. (4.17) is attained by letting, e.g., $\theta = \alpha$. However, if $\alpha \in (\frac{1}{2}, 1)$,

one cannot find a $\theta \in [0,1]$ such that the equality $\mathrm{P}_{\tilde{Q}|\theta}(B(\theta,\alpha)) = \alpha$ holds because, by the above considerations, this would imply either of the following contradictions.

- If $\theta > \frac{1}{2}$, then $1-\theta < \frac{1}{2} < \alpha$ and $B(\theta,\alpha) = \{0\}$. But this implies the contradiction $\mathrm{P}_{\tilde{Q}|\theta}(B(\theta,\alpha)) = 1-\theta < \frac{1}{2} < \alpha$.
- If $\theta \leq \frac{1}{2} < \alpha$, then $B(\theta,\alpha) = \{1\}$, which implies the contradiction $\mathrm{P}_{\tilde{Q}|\theta}(B(\theta,\alpha)) = \theta \leq \frac{1}{2}$.

In conclusion, $\mathrm{P}_{\tilde{Q}|\theta}(B(\theta,\alpha)) \leq \frac{1}{2}$ for all $\theta \in [0,1)$ and, e.g., $\theta = \frac{1}{2}$ yields the maximum $\mathrm{P}_{\tilde{Q}|\theta=\frac{1}{2}}(B(\theta=\frac{1}{2},\alpha)) = \frac{1}{2}$ for all $\alpha > \frac{1}{2}$. The supremum of the elementary p-values is also visualized below.



Maximally efficient Π-IMs for this particular statistical model are, e.g., obtained in Ex. 9 under both Cumulative P-Γ-transforms for $m = 1$.

The above example demonstrates that designing a Π-IM and checking validity, expressiveness and likelihood conformity can be tedious without some 'recipe'. In the general framework of IMs, such a guideline consists of the successive application of the a-, p- and c-step, which provides guarantees concerning validity and expressiveness. An arguably more straightforward derivation of Π-IMs from a given statistical model is discussed in Section 4.3.1.

4.2.3 Possibilistic Predictor Models

It is often desirable to make predictions about future observations $\tilde{V}$ based on past observations $\tilde{Q} = q$ if the statistical models

$$\tilde{V} \sim \mathrm{P}_{\tilde{V}|\theta} \qquad \text{and} \qquad \tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta} \tag{4.22}$$

both depend on the same set of unknown parameters $\theta \in \Theta$. The application of IMs to such prediction problems has, e.g., been investigated by Cella and Martin [CellaMartin22].

The discussions in Chapters 2 and 3 fall short regarding the consideration of information obtained from data. IP-description distributions are used to model unperturbed, certain

information, and the IP-Π-transform does not take the likelihood of the different probability measures in the family $\mathfrak{P}$ into account. The following exposition gives some brief intuition on how this may be still be achieved in a manner that is consistent with the previous discussion of possibilistic descriptions and Π-IMs.

Similar to a Π-IM, a *Possibilistic Predictor Model* (Π-PM) $\kappa_{\tilde{V}|\tilde{Q}}$ of $\tilde{V}$ given $\tilde{Q}$ is any map that produces an elementary possibility function $\kappa_{\tilde{V}|q} : \mathbb{V} \to [0,1]$, a so-called *elementary prediction function* of $\tilde{V}$, given an observation $\tilde{Q} = q$. The induced necessity and possibility measures $\underline{K}_{\tilde{V}|q} : \mathbb{B}(\mathbb{V}) \to [0,1]$ and $\overline{K}_{\tilde{V}|q} : \mathbb{B}(\mathbb{V}) \to [0,1]$ are also data-dependent capacities and are called *lower* and *upper prediction measures*, respectively. Either of these objects may also be called a *prediction distribution* of $\tilde{V}$.

These prediction distributions are not to be confused with those of, e.g., Lawless and Fredette [LawlessFredette05], whose predictor models produce probabilistic prediction distributions adhering to the Kolmogorov axioms. Nevertheless, they are related by what they intend to describe.

4.2.3.1 Perceptiveness

Similar to the validity criterion in Eq. (4.9)—and with similar reasons—, a Π-PM is said to be *perceptive* if

$$\sup_{\mathrm{P}_{\tilde{V},\tilde{Q}} \in \mathfrak{S}} \mathrm{P}_{\tilde{V},\tilde{Q}} \left(\left\{ (v,q) \in \mathbb{V} \times \mathbb{Q} : \kappa_{\tilde{V}|q}(v) \leq \alpha \right\} \right) \leq \alpha \tag{4.23}$$

for all $\alpha \in [0,1]$. Perceptiveness can be related to the concept of type-2 validity introduced by Cella and Martin [CellaMartin22], which—applied to Π-PMs—translates to

$$\sup_{\mathrm{P}_{\tilde{V},\tilde{Q}} \in \mathfrak{S}} \mathrm{P}_{\tilde{V},\tilde{Q}} \left(\left\{ (v,q) \in \mathbb{V} \times \mathbb{Q} : \overline{K}_{\tilde{V}|q}(B) \leq \alpha \wedge v \in B \right\} \right) \leq \alpha \tag{4.24}$$

for all $B \in \mathbb{B}(\mathbb{V})$ and all $\alpha \in [0,1]$.

Proposition 41. *A perceptive* Π*-PM* $\kappa_{\tilde{V}|\tilde{Q}}$ *is type-2 valid.*

Proof. To show that Eq. (4.23) implies Eq. (4.24), let $\alpha \in [0,1]$, let $B \in \mathbb{B}(\mathbb{V})$, and consider any $v \in B$ and $q \in \mathbb{Q}$ with $\overline{K}_{\tilde{V}|q}(B) \leq \alpha$. Then, $\kappa_{\tilde{V}|q}(v) \leq \sup_{v' \in B} \kappa_{\tilde{V}|q}(v') = \overline{K}_{\tilde{V}|q}(B) \leq \alpha$ implies that

$$\begin{aligned} &\sup_{\mathrm{P}_{\tilde{V},\tilde{Q}} \in \mathfrak{S}} \mathrm{P}_{\tilde{V},\tilde{Q}} \left(\left\{ (v,q) \in \mathbb{V} \times \mathbb{Q} : \overline{K}_{\tilde{V}|q}(B) \leq \alpha \wedge v \in B \right\} \right) \\ \leq &\sup_{\mathrm{P}_{\tilde{V},\tilde{Q}} \in \mathfrak{S}} \mathrm{P}_{\tilde{V},\tilde{Q}} \left(\left\{ (v,q) \in \mathbb{V} \times \mathbb{Q} : \kappa_{\tilde{V}|q}(v) \leq \alpha \right\} \right) \leq \alpha. \end{aligned}$$

□

The converse, i.e. that type-2 validity implies perceptiveness, does not appear to be generally true—even though, as of yet, it cannot be ruled out conclusively.

It must be stressed that the interpretation of the prediction distribution $\kappa_{\tilde{V}|q}$ is very different from that of a simple IP description $\pi_{\tilde{V}}$; on the contrary, the information contained in a prediction distribution is much more similar to the information contained in a confidence distribution than to that in an IP description.

Whereas, e.g., the possibility measure corresponding to $\pi_{\tilde{V}}$ provides an actual upper bound $\Pi_{\tilde{V}}(B)$ on the probability $\mathrm{P}_{\tilde{V}}(B)$ of the event $\tilde{V} \in B$ for some $B \in \mathbb{B}(\mathbb{V})$, the bound $\overline{K}_{\tilde{V}|q}(B)$ induced by $\kappa_{\tilde{V}|q}$ is not to be understood as rigorously. Instead, it is—again—based on a-priori probabilities that rule out 'unacceptable systematic errors'. If $\kappa_{\tilde{V}|q}(v)$ is small, then the event leading to the combination (v, q) must—under any admissible population candidate—be somewhat improbable, and it is reasonable to regard it as implausible or surprising for $\tilde{Q} = q$ to also occur after having seen $\tilde{V} = v$.
In other words, the type-2 validity implied by the perceptiveness of the prediction distribution states that, for any $\alpha \in [0, 1]$, the probability that the probability of the events $\overline{K}_{\tilde{V}|\tilde{Q}}(B) \leq \alpha$ and $\tilde{V} \in B$ occuring simultaneously is bounded by Eq. (4.24). Agreeing on the strategy of predicting that the next realization of $\tilde{V}$ will fall into B if the observation $\tilde{Q} = q$ yields $\overline{K}_{\tilde{V}|q}(B) \leq \alpha$ can be expected to be a successful prediction with at most probability α. Conversely, following, e.g., the strategy of predicting $\tilde{V} \notin B$ when $\overline{K}_{\tilde{V}|q}(B)$ is close to zero, is guaranteed to be successful with a probability close to one. Nevertheless, all these probabilities are obtained before seeing the data and make statements about the particular prediction strategy based on the Π-PM only. The actual probability $\mathrm{P}_{\tilde{V}}(B) = \mathrm{P}(\tilde{V} \in B)$ will usually be entirely different from $\overline{K}_{\tilde{V}|q}(B)$.

Prediction Sets Similar to confidence sets, the superlevel sets $\mathcal{C}^{\alpha}_{\kappa_{\tilde{V}|q}}$ produced by a perceptive Π-PM $\kappa_{\tilde{V}|\tilde{Q}}$ also fulfill the traditional (frequentist) definition of a prediction set [Shao03], namely that they have a guaranteed coverage probability of

$$\begin{aligned} &\inf_{\theta\in\Theta} \mathrm{P}_{\tilde{V},\tilde{Q}|\theta}\left(\left\{(v,q) \in \mathbb{V}\times\mathbb{Q} : v \in \mathcal{C}^{\alpha}_{\kappa_{\tilde{V}|q}}\right\}\right) \\ =&\inf_{\theta\in\Theta} \mathrm{P}_{\tilde{V},\tilde{Q}|\theta}\left(\left\{(v,q) \in \mathbb{V}\times\mathbb{Q} : \kappa_{\tilde{V}|q}(v) > \alpha\right\}\right) \\ =&1 - \sup_{\theta\in\Theta} \mathrm{P}_{\tilde{V},\tilde{Q}|\theta}\left(\left\{(v,q) \in \mathbb{V}\times\mathbb{Q} : \kappa_{\tilde{V}|q}(v) \leq \alpha\right\}\right) \geq 1-\alpha \end{aligned} \tag{4.25}$$

for all $\alpha \in [0, 1]$. Again, this bound is not to be understood as an a-posteriori probability of $\tilde{V} \in \mathcal{C}^{\alpha}_{\kappa_{\tilde{V}|q}}$ after seeing $\tilde{Q} = q$. It is the a-priori probability that $\tilde{V}$ will fall into the prediction distributions' superlevel sets before making observations. If these sets are constructed according to the Π-PM, the strategy of predicting that the realization $\tilde{V} = v$ is in $\mathcal{C}^{\alpha}_{\kappa_{\tilde{V}|q}}$, before seeing $\tilde{Q} = q$, has a success probability of at least $1 - \alpha$.

While prediction distributions are not IP-description distributions, the assertions made through an IP description $\pi_{\tilde{V}}$ also possess a predictive interpretation. Under the assmuption

that $\mathrm{P}_{\tilde{V}|\theta} \preceq \pi_{\tilde{V}}$ for all $\theta \in \Theta$, predicting that $\tilde{V} \in B$ if $\Pi_{\tilde{V}}(B) \leq \alpha$, has a success probability of at most α. Therefore, $\pi_{\tilde{V}} = \kappa_{\tilde{V}|\emptyset}$ is a special type of Π-PM for $\tilde{Q} = \emptyset$ because then Eq. (4.23) reads

$$\sup_{\theta\in\Theta} \mathrm{P}_{\tilde{V}|\theta}\left(\left\{v \in \mathbb{V} : \kappa_{\tilde{V}|\emptyset}(v) \leq \alpha\right\}\right) = \sup_{\theta\in\Theta} \mathrm{P}_{\tilde{V}|\theta}\left(\mathcal{S}^{\alpha}_{\pi_{\tilde{V}}}\right) \leq \alpha \tag{4.26}$$

for all $\alpha \in [0,1]$, which follows from $\mathrm{P}_{\tilde{V}|\theta} \preceq \pi_{\tilde{V}}$ for all $\theta \in \Theta$. The converse is, however, not true. By observing $\tilde{Q} = q$, additional evidence with respect to the confidence in the various parameter values $\theta \in \Theta$ can be gathered, which naturally allows for more expressive predictions of the future values of $\tilde{V}$, but it is certainly false to presume that $\mathrm{P}_{\tilde{V}} \preceq \kappa_{\tilde{V}|q}$ for all $q \in \mathbb{Q}$.

Put differently, perceptiveness is a (weaker) concept for prediction distributions, whereas consistency is a (stronger) descriptive concept that serves a theoretical purpose and can only indirectly be connected to prediction distributions, e.g., via the methods described in Section 4.3.3.

4.3 Possibilistic Inference

This section presents novel methods for the construction of Π-IMs and Π-PMs via possibilistic descriptions of a statistical model $\mathfrak{S}$. The origins of this approach may be found in [HoseHanss21a, HoseHanss21b, HoseHanss21c]. Alternative methods have been described by Martin et al. [Martin15, Martin18, CahoonMartin19, CahoonMartin21]

4.3.1 Parametric Inference

In the following, some prototypes for Π-IMs are discussed.

Suppose a parameter-dependent IP-description distribution $\pi_{\tilde{Q}|\theta} : \mathbb{Q} \to [0,1]$ of $\mathrm{P}_{\tilde{Q}|\theta}$ is available, i.e., $\mathrm{P}_{\tilde{Q}|\theta} \preceq \pi_{\tilde{Q}|\theta}$ for all $\theta \in \Theta$. Then $\pi_{\tilde{Q}|\theta}$ is said to be a *parameter-dependent (possibilistic) description* of $\mathfrak{S}$, of $\mathrm{P}_{\tilde{Q}|\theta}$ or of $\tilde{Q}$.

If θ^* were known, then $\pi_{\tilde{Q}|\theta^*}$ would constitute a consistent possibilistic description of $\tilde{Q}$. Yet, the actual parameter is unknown, and instead $\tilde{Q} = q$ is observed, where $\tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta^*}$. This section discusses what information the parameter-dependent possibilistic description contains regarding the possible parameter values.

4.3.1.1 The Pivotal Step

What may be learned about θ^* from the values of $\pi_{\tilde{Q}|\theta}(q)$ for the possible parameter values $\theta \in \Theta$ given an observation $\tilde{Q} = q$?

To answer this question, a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ based on the parameter-dependent possibilistic description is introduced: Under the so-called *Pivotal Step* its elementary confidence distributions are given by

$$\gamma_{\hat{\theta}|q}(\theta) = \pi_{\tilde{Q}|\theta}(q) \tag{4.27}$$

for all $\theta \in \Theta$ and all $q \in \mathbb{Q}$. This simple shift of indices and arguments is the fundamental technique that enables possibilistic inference with a close resemblance to the construction of likelihood functions from a probability mass/density function. Appreciating that one of the simplest descriptions of $\mathrm{P}_{\tilde{Q}|\theta}$ is, e.g., given by the Cumulative P-Π-transform $\pi_{\tilde{Q}|\theta} = F_{\tilde{Q}|\theta}$, it could also be argued that it constitutes a likelihood concept in a generalized cumulative space—instead of the mass/density space.

Since, by construction, these confidence distributions—and, hence, the corresponding upper and lower confidence measures, too—depend on the statistical model and the observations only, the Pivotal Step, indeed, produces a Π-IM. Moreover, it yields valuable information for hypotheses about θ^*, for it adheres to the core principles of Π-IMs, which will be demonstrated in the following.

The following proposition may be summarized by saying that consistency implies validity under the Pivotal Step.

Proposition 42. *The Pivotal Step in Eq.* (4.27) *yields a valid* Π*-IM* $\gamma_{\hat{\theta}|\tilde{Q}}$ *if and only if* $\mathrm{P}_{\tilde{Q}|\theta} \preceq \pi_{\tilde{Q}|\theta}$ *for all* $\theta \in \Theta$.

Proof. The proposition follows immediately by verifying that, under the respective assumptions, $\mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}) = \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \pi_{\tilde{Q}|\theta}(q) \leq \alpha\}) \leq \alpha$ for all $\alpha \in [0,1]$ and all $\theta \in \Theta$, where the inequality follows from the consistency of $\pi_{\tilde{Q}|\theta}$ and $\mathrm{P}_{\tilde{Q}|\theta}$ or the validity of $\gamma_{\hat{\theta}|\tilde{Q}}$. □

Additionally, more specific parameter-dependent descriptions also imply element-wise more efficient Π-IMs.

Proposition 43. *Let* $\pi^{(1)}_{\tilde{Q}|\theta}$ *and* $\pi^{(2)}_{\tilde{Q}|\theta}$ *be two parameter-dependent descriptions of* $\mathfrak{S}$ *with the corresponding* Π*-IMs* $\gamma^{(1)}_{\hat{\theta}|\tilde{Q}}$ *and* $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$ *obtained in the Pivotal Step in Eq.* (4.27)*. Then,* $\gamma^{(1)}_{\hat{\theta}|\tilde{Q}}$ *is element-wise more efficient than* $\gamma^{(2)}_{\hat{\theta}|\tilde{Q}}$ *if and only if* $\pi^{(1)}_{\tilde{Q}|\theta} \preceq \pi^{(2)}_{\tilde{Q}|\theta}$ *for all* $\theta \in \Theta$.

Proof. The proposition follows immediately by verifying that, under the given assumptions,

$$\begin{aligned}\mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \gamma^{(1)}_{\hat{\theta}|q}(\theta) \leq \alpha\}) &= \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \pi^{(1)}_{\tilde{Q}|\theta}(q) \leq \alpha\}) \\ &\geq \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \pi^{(2)}_{\tilde{Q}|\theta}(q) \leq \alpha\}) = \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \gamma^{(2)}_{\hat{\theta}|q}(\theta) \leq \alpha\})\end{aligned}$$

for all $\alpha \in [0,1]$ and all $\theta \in \Theta$, where the inequality follows from the specificity or the efficiency relation. □

Regarding the likelihood conformity of the obtained Π-IM, general results may not be given.

The Reverse Pivotal Step The 'if and only if' in Propositions 42 and 43 suggests that, given a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ of some statistical model $\mathrm{P}_{\tilde{Q}|\theta}$, it is also possible to obtain a parameter-dependent description $\pi_{\tilde{Q}|\theta}$ thereof under a *Reverse Pivotal Step* by defining

$$\pi_{\tilde{Q}|\theta}(q) = \gamma_{\hat{\theta}|q}(\theta) \tag{4.28}$$

for all $\theta \in \Theta$ and all $q \in \mathbb{Q}$, which reads exactly like the Pivotal Step in Eq. (4.27). By the above-mentioned propositions, the obtained descriptions $\pi_{\tilde{Q}|\theta}$ are, e.g., consistent with $\mathrm{P}_{\tilde{Q}|\theta}$ for all $\theta \in \Theta$ if the Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ is valid, and they retain their efficiency properties in the form of specificity.

4.3.1.2 Probability-to-Confidence Transforms

The IP-Π-transform, which embodies the principles of quantitative possibility theory, generally produces consistent and maximally specific IP-description distributions with respect to some plausibility order. Therefore, it is justified to investigate the parameter-dependent descriptions of $\mathfrak{S}$ that may be obtained by the IP-Π-transform because these properties directly translate to the validity and efficiency of the corresponding Π-IMs resulting from the subsequent application of the Pivotal Step. Moreover, this also allows to choose the remaining degree of freedom in the IP-Π-transform, the elementary plausibility function, based on the likelihood function.

Employing the IP-Π-transform, a statistical model $\mathfrak{S}$ can be converted into an IP-description distribution $\mathcal{T}[\mathfrak{S}, \rho_{\tilde{Q}}]$ under some plausibility distribution $\rho_{\tilde{Q}}$ in order to obtain a possibilistic description of $\mathfrak{S}$. As this would both marginalize the statistical model, i.e. the parameter dependency, and entirely neglect the available data $\tilde{Q} = q$ and the information they contain, it would certainly miss the point of statistical inference; but, considering the advantageous properties of the Pivotal Step, it is expedient to consider the parameter-dependent possibilistic description $\pi_{\tilde{Q}|\theta} : \mathbb{Q} \to [0,1]$ of $\tilde{Q}$ obtained via the P-Π-transform of $\mathrm{P}_{\tilde{Q}|\theta}$ under some parameter-dependent plausibility distribution $\rho_{\tilde{Q}|\theta} : \mathbb{Q} \to [0,1]$. The corresponding expression reads

$$\pi_{\tilde{Q}|\theta}(q) = \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : \rho_{\tilde{Q}|\theta}(\xi) \leq \rho_{\tilde{Q}|\theta}(q)\}\right) \tag{4.29}$$

for $q \in \mathbb{Q}$ and $\theta \in \Theta$. This may be thought of as an intermediary result of the IP-Π-transform of $\mathfrak{S}$, for it is easily verified that the disjunction

$$\pi_{\tilde{Q}}(q) = \sup_{\theta \in \Theta} \pi_{\tilde{Q}|\theta}(q) \tag{4.30}$$

provides a possibilistic description of $\mathfrak{P} = \mathfrak{S}$, where the parameter dependency has been marginalized. Nevertheless, $\pi_{\tilde{Q}|\theta}$ still contains all the relevant information.

Inserting Eq. (4.29) into Eq. (4.27), one obtains the *Probability-to-Confidence Transform* (P-Γ-transform) which reads

$$\gamma_{\hat{\theta}|q}(\theta) = \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : \rho_{\tilde{Q}|\theta}(\xi) \leq \rho_{\tilde{Q}|\theta}(q)\}\right) \tag{4.31}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$.

Under Propositions 11 and 42, the Π-IM obtained by the P-Γ-transform is clearly valid; and, under Propositions 13 and 43, it is element-wise more efficient than any other valid Π-IM $\gamma'_{\hat{\theta}|\tilde{Q}}$ produced under the Pivotal Step from a parameter-dependent possibilistic description $\pi'_{\tilde{Q}|\theta}$ that is consistent with $\mathrm{P}_{\tilde{Q}|\theta}$ and plausibility-conform to $\rho_{\tilde{Q}|\theta}$ for all $\theta \in \Theta$, respectively. Especially the latter property is quite technical and intangible; therefore, it is paraphrased by the following result.

Proposition 44. *The* Π*-IM* $\gamma_{\hat{\theta}|\tilde{Q}}$ *obtained by the* P-Γ*-transform of* $\mathrm{P}_{\tilde{Q}|\theta}$ *under a parameter-dependent elementary plausibility function* $\rho_{\tilde{Q}|\theta}$ *is maximally efficient if the imprecise variable* $\tilde{V} = \rho_{\tilde{Q}|\theta}(\tilde{Q})$ *has a continuous CPF* $F_{\tilde{V}|\theta}$ *for all* $\theta \in \Theta$.

Proof. Let $\theta \in \Theta$ and let $\alpha \in [0, 1]$. Since $\tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta}$, the P-Π-transform of $\mathrm{P}_{\tilde{Q}|\theta}$ reads

$$\pi_{\tilde{Q}|\theta}(q) = \mathrm{P}_{\tilde{Q}|\theta}\left(\{q' \in \mathbb{Q} : \rho_{\tilde{Q}|\theta}(q') \leq \rho_{\tilde{Q}|\theta}(q)\}\right) = F_{\tilde{V}|\theta}\left(\rho_{\tilde{Q}|\theta}(q)\right)$$

for all $q \in \mathbb{Q}$. The proposition follows directly from the Probability Integral Transform of $\tilde{V}$. More precisely, it is easily verified that

$$\begin{aligned}\alpha &= \mathrm{P}_{\tilde{V}|\theta}\left(\left\{v \in \mathbb{V} : F_{\tilde{V}|\theta}(v) \leq \alpha\right\}\right) = \mathrm{P}_{\tilde{Q}|\theta}\left(\left\{q \in \mathbb{Q} : F_{\tilde{V}|\theta}\left(\rho_{\tilde{Q}|\theta}(q)\right) \leq \alpha\right\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\left\{q \in \mathbb{Q} : \pi_{\tilde{Q}|\theta}(q) \leq \alpha\right\}\right) = \mathrm{P}_{\tilde{Q}|\theta}\left(\left\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\right\}\right).\end{aligned}$$

The elementary p-values $\gamma_{\hat{\theta}|\tilde{Q}}(\theta)$ are uniformly distributed for all $\theta \in \Theta$, which indicates maximal efficiency. □

The required continuity of $F_{\tilde{V}|\theta}$ immediately implies that Π-IMs of discrete statistical models are usually not maximally specific, which can, e.g., be observed in Ex. 8. However, even for discrete distributions, the Π-IM obtained by the P-Γ-transform is certainly valid and often 'as efficient as possible'.

The P-Γ-transform constitutes a simple prototype for a family of Π-IMs, and—depending on the choice of the parameter-dependent plausibility distribution of $\tilde{Q}$—several special cases can be derived.

Cumulative Probability-to-Confidence Transforms If $\tilde{Q}$ is an $\mathbb{R}$-valued observation, then the (Complementary) Cumulative P-Π-transform in connection with the Pivotal Step yield the *(Complementary) Cumulative* P-Γ-*transform* producing the Π-IMs $\gamma^{\mathrm{CPF}}_{\hat{\theta}|\tilde{Q}}$ and $\gamma^{\mathrm{CCPF}}_{\hat{\theta}|\tilde{Q}}$, respectively. The obtained confidence distributions

$$\gamma^{\mathrm{CPF}}_{\hat{\theta}|q}(\theta) = F_{\tilde{Q}|\theta}(q) \qquad \text{and} \qquad \gamma^{\mathrm{CCPF}}_{\hat{\theta}|q}(\theta) = \bar{F}_{\tilde{Q}|\theta}(q) \tag{4.32}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$ typically correspond to one-sided hypothesis tests and yield one-sided confidence intervals. The *Symmetric Cumulative* P-Γ-*transform*

$$\gamma^{\mathrm{SCPF}}_{\hat{\theta}|q}(\theta) = \min\left(1, 2 \cdot \min\left(F_{\tilde{Q}|\theta}(q), 1 - F_{\tilde{Q}|\theta}(q)\right)\right) \tag{4.33}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$, which is based on the Symmetric Cumulative P-Π-transform often yields standard two-sided confidence intervals.

Likelihood-to-Confidence Transforms Based on the likelihood function, the *likelihood principle* is formulated, whose fundamental importance has been widely recognized by many statisticians. The original name goes back to Savage [SavageEtAl62, p. 17], who observes that "[the likelihood function] constitutes the entire evidence of the experiment, that is, it tells all that the experiment has to tell." He formalizes this by saying that, if the same experiment is performed twice yielding $\tilde{Q}_1 = q_1$ and $\tilde{Q}_2 = q_2$, respectively, and if the ratio $\frac{\ell_{\hat{\theta}|q_1}(\theta)}{\ell_{\hat{\theta}|q_2}(\theta)}$ is constant for all $\theta \in \Theta$, "then each of the two data [...] have exactly the same thing to say about the [parameter] values" [ibid.]. The immediate conclusion is that, if a likelihood function is available, a Π-IM accounting for all information in the data can be built from the likelihood function alone. This conclusion is similar to Axiom (A2) of Wassermann [Wasserman90] and Axiom (L) of Denœux [Denoeux14]; however, by itself, it does not necessarily require confidence and likelihood to coincide, which is what Wassermann and Denœux conclude.

Recalling Shafer's argument "that q renders $\theta \in \Theta$ more plausible than $\theta' \in \Theta$ whenever $p_\theta(q) > p_{\theta'}(q)$." [Shafer76, p. 238], it is, indeed, a short step from this insight to the idea that the likelihood function establishes a very sensible plausibility order among the elements of Θ and using the likelihood function as a parameter-dependent plausibility distribution

$$\rho_{\tilde{Q}|\theta}(q) = \ell_{\hat{\theta}|q}(\theta) = p_\theta(q) \tag{4.34}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$ is consistent with intuition. It also constitutes a Reverse Pivotal Step because, similar to the Pivotal Step, the indices and arguments are simply switched, and the plausibility order is transferred onto $\mathbb{Q}$. Finally, employing the resulting parameter-dependent plausibility distribution $\rho_{\tilde{Q}|\theta}$ in Eq. (4.31) constitutes the P-Γ-transform based on the likelihood, or the *Absolute Likelihood-to-Confidence Transform* producing the

Π-IM $\gamma^{\ell}_{\hat{\theta}|\tilde{Q}}$ given by the confidence distributions

$$\begin{aligned}\gamma^{\ell}_{\hat{\theta}|q}(\theta) &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : \ell_{\hat{\theta}|\xi}(\theta) \leq \ell_{\hat{\theta}|q}(\theta)\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : p_{\tilde{Q}|\theta}(\xi) \leq p_{\tilde{Q}|\theta}(q)\}\right)\end{aligned} \tag{4.35}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$. This is also the Π-IM that is obtained by a parameter-dependent Optimal P-Π-transform of $\mathrm{P}_{\tilde{Q}|\theta}$ and the subsequent Pivotal Step and has, e.g., been employed in Ex. 8 in order to obtain the proposed confidence distribution.
However, this choice is only almost commensurate with the likelihood principle and the suggestion of Shafer, who proposes to use the relative likelihood function $\Lambda_{\hat{\theta}|q}$ as an elementary plausibility function, which arguably generates more comparability among likelihood functions, in particular among the induced plausibility orders, for different pieces of evidence, i.e. for different sets of data. The P-Γ-transform based on the relative likelihood, the *Relative Likelihood-to-Confidence Transform*, produces the Π-IM $\gamma^{\Lambda}_{\hat{\theta}|\tilde{Q}}$ given by

$$\begin{aligned}\gamma^{\Lambda}_{\hat{\theta}|q}(\theta) &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : \Lambda_{\hat{\theta}|\xi}(\theta) \leq \Lambda_{\hat{\theta}|q}(\theta)\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\left\{\xi \in \mathbb{Q} : \frac{p_{\tilde{Q}|\theta}(\xi)}{\sup_{\theta' \in \Theta} p_{\tilde{Q}|\theta'}(\xi)} \leq \frac{p_{\tilde{Q}|\theta}(q)}{\sup_{\theta' \in \Theta} p_{\tilde{Q}|\theta'}(q)}\right\}\right)\end{aligned} \tag{4.36}$$

for all $q \in \mathbb{Q}$ and all $\theta \in \Theta$. By commensurability with the likelihood principle, they can often be expected to lead to optimal results—not only in terms of validity and efficiency.

In particular, the formulation in Eq. (4.36) is 'powerful' in a well-defined manner because, formally, it corresponds to *likelihood ratio (hypothesis) tests*, which are *uniformly most powerful*, i.e., they minimize the probability of the type II error in Eq. (4.14), under a variety of assumptions—not least because the likelihood ratio depends on the data only through a sufficient statistic, if there is one. For instance, according to the well-known Neyman-Pearson Lemma, the likelihood ratio test is uniformly most powerful for simple hypotheses, i.e. if the parameter space $\Theta = \{\theta_1, \theta_2\}$ consists of two alternatives only. For general (composite) hypotheses, without restrictions on the parameter space, it is also uniformly most powerful if the likelihood ratio is monotone with respect to some statistic of the data [Shao03, Ch. 6.1].

In principle, every confidence distribution produced by a Π-IM bears an apparent resemblance to the likelihood function; yet, the two must not be confused. Generally speaking, whereas the likelihood is precisely the probability $\mathrm{P}_{\theta}(\tilde{Q} = q)$ of obtaining the data q given some parameter value $\theta \in \Theta$, the likelihood-based confidence of that parameter is the probability $\mathrm{P}_{\theta}(\rho_{\hat{\theta}|\tilde{Q}}(\theta) \leq \rho_{\hat{\theta}|q}(\theta))$ of obtaining at the most the given likelihood (ratio) of the data q given some parameter value $\theta \in \Theta$ under the statistical model. The implied rare events leading to errors of the first type are precisely those that produce lower relative likelihoods. If the data are, by chance, so misleading that the relative likelihood of the

actual parameter value is very low, then it should not come as a surprise that a true hypothesis is rejected.

It must be pointed out that the formulations in Eqs. (4.35) and (4.36) do not necessarily guarantee likelihood conformity. The produced confidence distribution $\gamma_{\hat{\theta}|q}$ are often plausibility-conform to its likelihood distribution $\ell_{\hat{\theta}|q}$, but not always.

As a final remark, Balch observes that in his "theory of confidence structures [which is compatible with IMs], satisfying the likelihood principle is, at most, a secondary goal" [Balch12, p. 1018]. That is, the likelihood principle is only to be considered once the more important principles of validity and efficiency have been satisfied—similar to how the elementary plausibility function resolves the final issue of non-uniqueness for the IP-Π-transform under the principal constraints of consistency and maximal specificity. Moreover, the Relative Likelihood-to-Confidence Transform also inherits some rather severe shortcomings of likelihood ratio tests, especially that they can be tedious to compute. Since, in particular for vectors of parameters, uniformly most powerful tests often do not exist, it is therefore expedient to also consider other types of P-Γ-transforms, e.g. based on other P-Π-transforms. Nevertheless, likelihood-based P-Γ-transforms possess an intuitive appeal and can be very useful—especially as a starting point for the analysis of a statistical model.

Example 9: Binomial Π-IM

Consider the problem of estimating the success probability of a Bernoulli experiment as discussed in Ex. 8, but now the experiment, e.g. the coin toss, is repeated m times, and the outcomes $\tilde{Q}_1, \ldots, \tilde{Q}_m$ are observed, where $\tilde{K} = \sum_{i=1}^{m} \tilde{Q}_i$ is the number of observed successes, which may be anywhere between 0 and m. The appropriate statistical model for $\tilde{K} \sim \mathrm{P}_{\tilde{K}|\theta}$ is the binomial sampling model given by the probability mass function

$$p_{\tilde{K}|\theta}(k) = \binom{m}{k} \theta^k (1-\theta)^{m-k}$$

for $k = 0, \ldots, m$.

This problem is elementary to mathematical statistics and has been addressed countless times [ClopperPearson34, Vollset93, AgrestiCoull98, BrownCaiDasGupta01, Rigollet15, Balch20]. In fact, the ongoing debate regarding its solution is perhaps the most telling evidence for the non-existence of a universally accepted theory of statistical inference.

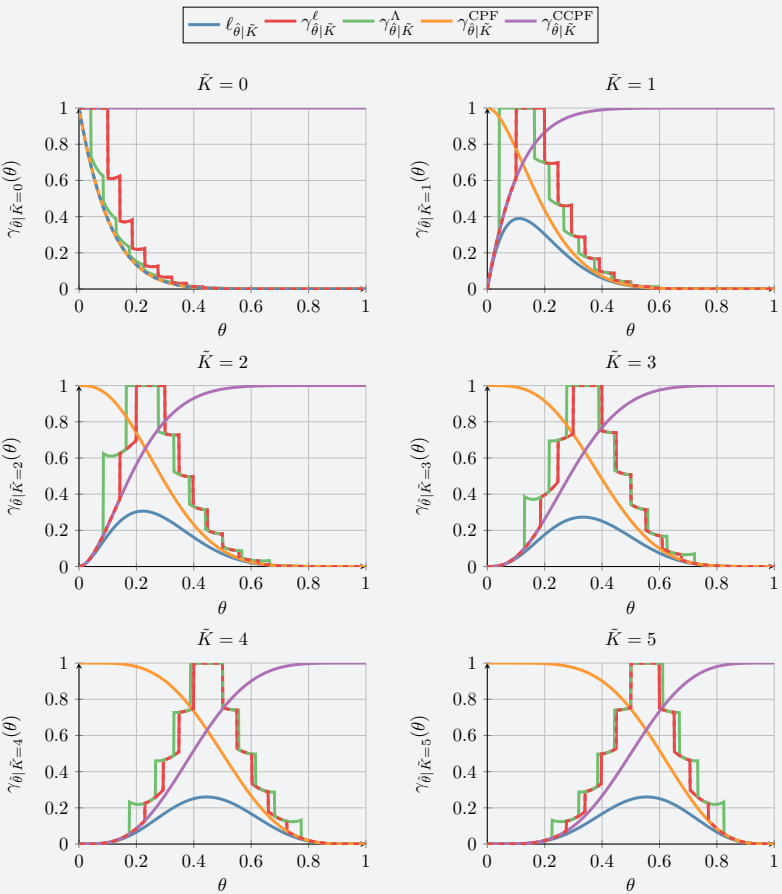
Four Π-IMs producing confidence distributions of the success probability from an observation $\tilde{K} = k$ are considered, namely those resulting from the Absolute and Relative Likelihood-to-Confidence Transforms, and from the Cumulative and

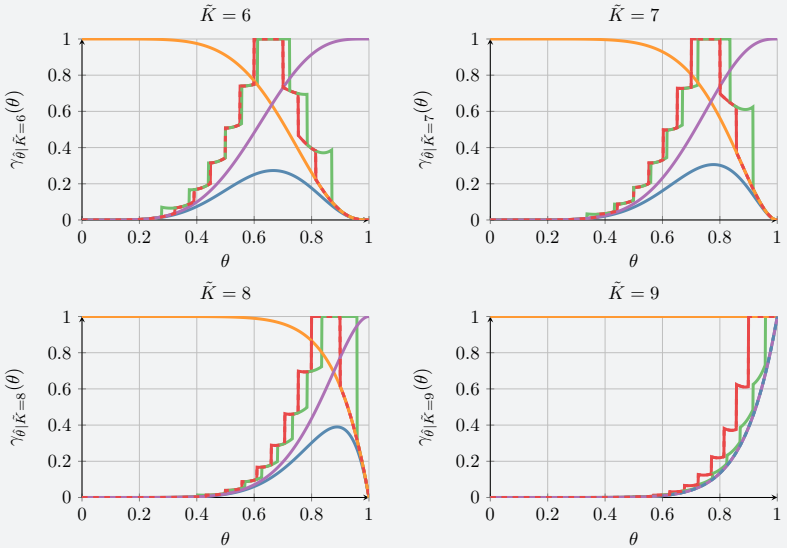
Complementary Cumulative P-Γ-transforms, given by

$$\gamma^{\ell}_{\hat{\theta}|k}(\theta) = \sum_{i=0,\dots,m\,:\,p_{\tilde{K}|\theta}(i)\le p_{\tilde{K}|\theta}(k)} p_{\tilde{K}|\theta}(i), \qquad \gamma^{\Lambda}_{\hat{\theta}|k}(\theta) = \sum_{i=0,\dots,m\,:\,\frac{p_{\tilde{K}|\theta}(i)}{p_{\tilde{K}|\frac{k}{m}}(i)}\le\frac{p_{\tilde{K}|\theta}(k)}{p_{\tilde{K}|\frac{k}{m}}(k)}} p_{\tilde{K}|\theta}(i),$$

$$\gamma^{\mathrm{CPF}}_{\hat{\theta}|k}(\theta) = \sum_{i=0}^{k} p_{\tilde{K}|\theta}(i) \qquad \text{and} \qquad \gamma^{\mathrm{CCPF}}_{\hat{\theta}|k}(\theta) = \sum_{i=k}^{m} p_{\tilde{K}|\theta}(i)$$

for $\theta \in [0,1]$, respectively. Below, they are depicted for $m = 9$.

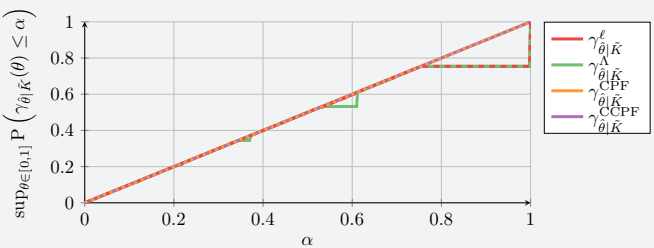




The Symmetric Cumulative P-Γ-transform corresponds to classical two-sided Pearson-Clopper confidence intervals for this problem but is generally less expressive than the likelihood-based P-Γ-transforms and, therefore, not shown.
By Proposition 42, all of these Π-IMs are valid, i.e., the probability of producing an elementary p-value less than or equal to $\alpha \in [0, 1]$ is bounded by

$$\mathrm{P}(\gamma_{\hat{\theta}|\tilde{K}}(\theta) \leq \alpha) = \mathrm{P}_{\tilde{K}|\theta^*}\left(\{k \in \{0, \ldots, m\} : \gamma_{\hat{\theta}|k}(\theta) \leq \alpha\}\right) = \sum_{k\,:\,\gamma_{\hat{\theta}|k}(\theta) \leq \alpha} p_{\tilde{K}|\theta^*}(k) \leq \alpha$$

for all $\hat{\theta} \in [0, 1]$. This is also observed in the following figure.



However, it is also apparent that only the Cumulative P-Γ-transforms, which coincide with the one-sided Pearson-Clopper confidence intervals, produce maximally efficient

Π-IMs; the elementary p-values obtained by the suprema of the likelihood-based P-Γ-transforms are not uniformly distributed because, in this scenario, these P-Γ-transforms do not fulfill the continuity assumption in Proposition 44. On the other hand, the Cumulative P-Γ-transform produces a vacuous, i.e. entirely uninformative, confidence distribution when $\tilde{K} = 9$, whereas the Complementary Cumulative P-Γ-transform produces a very expressive confidence distribution—and vice versa for $\tilde{K} = 0$.
The above claim that both Cumulative P-Γ-transforms produce one-sided confidence intervals is supported by this example: The confidence sets obtained by the Cumulative P-Γ-transform always have a lower bound of zero. The confidence sets obtained by the Complementary Cumulative P-Γ-transform always have an upper bound of one. If this one-sidedness is not desirable, the Likelihood-Based P-Γ-transforms are good alternatives, but a clear ranking in terms of their expressiveness is not possible. It is also not immediately clear which of the two to prefer because the produced confidence distributions tend to lie closely together, but neither is generally more expressive than the other.
The most conspicuous features of the Likelihood-Based P-Γ-transforms are the local maxima exhibited by the multi-modal confidence distributions that, e.g., imply non-convex confidence sets, a phenomenon that is called the 'spikiness problem' by Balch, who also provides some additional discussion [Balch20] and remedies. For this thesis, this peculiarity does not pose a serious issue, but, depending on how the confidence distributions are to be used later, this non-convexity may be disadvantageous for some implementations of membership computations, see Chapter 5.
Finally, it is evident, e.g., from the 'spikes', that most of the resulting confidence distributions do not achieve likelihood conformity. This property seems to be especially challenging to achieve in discrete statistical models.

In conclusion, all of these Π-IMs have certain (dis-)advantages, and a general recommendation which one to prefer cannot be given. Their respective appropriateness depends on the given circumstances.

4.3.1.3 Extending Possibilities to Confidence

The following *Possibilistic Inference Principle* first described by Hose and Hanss [HoseHanss21b, HoseHanss21c] provides a method of statistical inference that is perhaps most useful to practitioners. It allows deriving confidence distributions for a large class of (imprecisely stated) statistical models and is related to statistical inference based on pivotal quantities.

Especially in larger statistical models, it is often the case that $\tilde{Q}$ may be connected to a $\mathbb{V}$-valued imprecise variable $\tilde{V}$, a pivotal quantity with a known IP-description distribution $\pi_{\tilde{V}}$ that is independent of $\theta \in \Theta$, e.g. a superuniform variable, via an implicit relationship $\Xi = \Xi_\theta$ defined as in Section 3.2, which contains the entire parameter dependency of the statistical model. That is, the corresponding statistical model is described by

$$0 = \Xi_\theta(\tilde{V}, \tilde{Q}) \tag{4.37}$$

for $\theta \in \Theta$ and $\tilde{V} \sim \pi_{\tilde{V}}$. This specification allows to compute a parameter-dependent possibilistic description $\pi_{\tilde{Q}|\theta}$ of $\tilde{Q}$ via the Implicit Extension Principle—or via its explicit and inverse variants—ultimately specifying a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ that produces the confidence distributions

$$\gamma_{\hat{\theta}|q}(\theta) = \pi_{\tilde{Q}|\theta}(q) = \sup_{v \in \mathbb{V} \,:\, 0 = \Xi_\theta(v,q)} \pi_{\tilde{V}}(v) \tag{4.38}$$

for all $\theta \in \Theta$ and all $q \in \mathbb{Q}$ under the Pivotal Step. This expression, allowing to compute confidence distributions by the extension of description distributions, is, arguably, more in line with the fiducial approach to statistical inference and corresponds to the a-, p- and c-step of Martin and Liu [MartinLiu15].

This technique can also be applied when a precise probability distribution of the pivotal quantity $\tilde{V}$ is available, which can then be described possibilistically under any of the available P-Π-transforms.

4.3.2 Reasoning with Confidence

Often, obtaining a confidence distribution of some set of parameters will only be an intermediate step in solving a larger problem.

For instance, the unknown parameter θ often corresponds to some of the system's properties when analyzing a system. Having found a confidence distribution thereof by the successful application of a Π-IM, it could then be of interest to compute other features δ of this system that can be computed from these properties. Moreover, these features might also depend on other system properties θ', of which confidence distributions might have been obtained in a separate experiment and which may or may not overlap with θ.

Some fundamental tools are needed to compute such features, namely the extension of confidence distributions and their combination and updating.

4.3.2.1 Confidence Extension

In the following, assume that a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ depending on the $\mathbb{Q}$-valued observation $\tilde{Q}$ is available, and that—similarly to Eq. (3.20)—a *feature* $\delta \in \mathbb{D}$ in the *feature space* $\mathbb{D} \subseteq \mathbb{R}^{D_\mathbb{D}}$,

depends on $\theta \in \Theta$ via the implicit relationship

$$0 = \Xi(\theta, \delta), \tag{4.39}$$

where $\Xi : \Theta \times \mathbb{D} \to \mathbb{R}^D$, and 0 may be a vector of zeros.

For notational convenience, this feature $\delta = \hat{\delta}(\mathrm{P}_{\tilde{Q}|\theta})$ is also taken to be the output of the feature function $\hat{\delta} : \mathbb{P}(\mathbb{Q}, \mathbb{B}(\mathbb{Q})) \to \mathbb{D}$.

The fundamental question in this section is how $\gamma_{\hat{\theta}|\tilde{Q}}$ may be extended onto the feature space $\mathbb{D}$ in order to assess hypotheses $\delta^* \in K$ about $\delta^* = \hat{\delta}(\mathrm{P}_{\tilde{Q}})$ for some $K \in \mathbb{B}(\mathbb{D})$.

Indeed, the given information may be used to define a feature Π-IM $\gamma_{\hat{\delta}|\tilde{Q}}$, which is given by

$$\gamma_{\hat{\delta}|q}(\delta) = \sup_{\theta \in \Theta \,:\, 0 = \Xi(\theta, \delta)} \gamma_{\hat{\theta}|q}(\theta) \tag{4.40}$$

for all $\delta \in \mathbb{D}$ and all $q \in \mathbb{Q}$.

Defining a Π-IM of a feature in this manner requires some elaboration of how to read the expression $\gamma_{\hat{\delta}|\tilde{Q}}$. In particular, this notation does not imply that $\tilde{Q}$ be distributed according to $\mathrm{P}_{\tilde{Q}|\delta}$, which one might be tempted to assume since this is how one could read $\gamma_{\hat{\theta}|\tilde{Q}}$. On the contrary, $\tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta}$ is still implied in this notation, and the subscript $\hat{\delta}$ does not make any statements about the distribution of $\tilde{Q}$, except that θ and δ are related via Eq. (4.39); it simply denotes the parameter/feature that the Π-IM belongs to.

For instance, expressed in terms of Eq. (4.9), validity for the Π-IM $\gamma_{\hat{\delta}|\tilde{Q}}$ reads as

$$\mathrm{P}_{\tilde{Q}|\theta}\Big(\{q \in \mathbb{Q} : \gamma_{\hat{\delta}|q}(\delta) \leq \alpha\}\Big) \leq \alpha \tag{4.41}$$

for all $\alpha \in [0, 1]$, and for all $\theta \in \Theta$ and $\delta \in \mathbb{D}$ satisfying Eq. (4.39). Likewise, $\gamma_{\hat{\delta}|\tilde{Q}}$ is most efficient if the outer inequality is an equality.

Finally, the validity of $\gamma_{\hat{\delta}|\tilde{Q}}$ is inherited from that of $\gamma_{\hat{\theta}|\tilde{Q}}$—similar to how consistency is preserved under pushforwards.

Proposition 45. *If $\gamma_{\hat{\theta}|\tilde{Q}}$ is valid, then $\gamma_{\hat{\delta}|\tilde{Q}}$ is valid.*

Proof. The proof follows immediately from the validity of $\gamma_{\hat{\theta}|\tilde{Q}}$ by verifying that

$$\begin{aligned}\mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \gamma_{\hat{\delta}|q}(\delta) \leq \alpha\}) &= \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \sup_{\theta' \in \Theta \,:\, 0 = \Xi(\theta', \delta)} \gamma_{\hat{\theta}|q}(\theta') \leq \alpha\}) \\ &\leq \mathrm{P}_{\tilde{Q}|\theta}(\{q \in \mathbb{Q} : \gamma_{\hat{\theta}|q}(\theta) \leq \alpha\}) \leq \alpha\end{aligned}$$

for all $\alpha \in [0, 1]$, and for all $\theta \in \Theta$ and $\delta \in \mathbb{D}$ satisfying Eq. (4.39). □

The preservation of efficiency, on the other hand, seems to depend on the regularity of Ξ. For instance, if $\gamma_{\hat{\theta}|\tilde{Q}}$ is maximally efficient and if $\Xi(\theta, \delta) = \delta - \phi(\theta) = 0$ describes an

explicit relationship $\delta = \phi(\theta)$, where $\phi : \Theta \to \mathbb{D}$ is a bijective function, it is easy to see that the inequality in the proof of the above proposition becomes an equality, i.e., maximal efficiency is preserved.

In an analog manner, one may show that, from (valid) Π-IMs of certain features, one may also compute (valid) Π-IMs of other features under the Confidence Extension Principle.

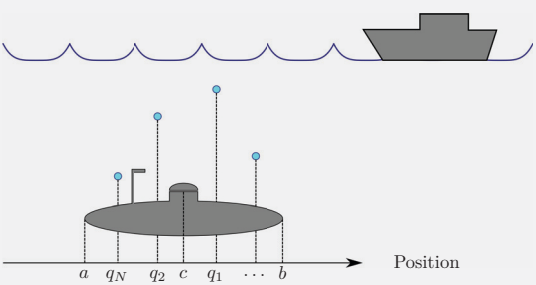
Nuisance Parameters Marginalization via the Confidence Extension Principle may be used to eliminate the dependency of the confidence distribution on nuisance parameters. For instance, if the Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ produces confidence distributions of $\theta = (\delta, \eta)$ composed of the parameters of interest $\delta \in \mathbb{D}$ and the nuisance parameters $\eta \in \mathbb{E}$, a Π-IM $\gamma_{\hat{\delta}|\tilde{Q}}$ of only the feature is given by

$$\gamma_{\hat{\delta}|q}(\delta) = \sup_{\eta \in \mathbb{E}} \gamma_{\hat{\theta}|q}(\delta, \eta) \tag{4.42}$$

for all $\delta \in \mathbb{D}$ and all $q \in \mathbb{Q}$. The dependency on the nuisance parameter η has been eliminated.

Example 10: Uniform Π-IM

The following example, in its many variations, enjoys wide popularity in the statistical literature, refer, e.g., to Morey et al. [MoreyEtAl16] and references therein.[43]
A submersible of unknown length has malfunctioned and is now floating helplessly in deep water, unable to ascend on its own. Even worse, it has lost contact with its support vessel at the surface. In case of such an emergency, the support crew has been instructed to drop a rescue line to the submersible's emergency hatch that is located halfway along its length in order to evacuate the aquanauts inside. The rescue team does not know the exact position of the submersible, but they can observe bubbles that form along the submersible's length, independently, with equal probability, and float to the surface. The situation is depicted below.



Because time matters and consecutive attempts would take too long, the support team wants to ensure that the rescue line be as close as possible to the hatch on the

first try.

The corresponding statistical model reads as follows. The positions $q_1, \ldots, q_m$ of the observed bubbles are realizations of the m iid variables $\tilde{Q}_1, \ldots, \tilde{Q}_m \sim \mathcal{U}(a, b)$ following a uniform probability distribution depending on the start and end points $a, b \in \mathbb{R}$ of the submersible, where $a \leq b$. These points constitute the vector of unknown parameters $\theta = [a,\ b]^{\mathrm{T}}$, and the position of the hatch can be expressed as the feature $c = \frac{1}{2}(a + b)$. The corresponding likelihood-based Π-IM is derived in the following.

The joint probability distribution of $\tilde{Q}_1, \ldots, \tilde{Q}_m \sim \mathcal{U}(a, b)$ is given by the probability density function

$$p_{\tilde{Q}_1,\ldots,\tilde{Q}_m|a,b}(q_1, \ldots, q_m) = \begin{cases} \frac{1}{(b-a)^m} & \text{for } a \leq q_1, \ldots, q_m \leq b \text{ and} \\ 0 & \text{otherwise} \end{cases}$$

for all $a, b \in \mathbb{R}$ with $a \leq b$. The relative likelihood variable is given by

$$\tilde{V} = \Lambda_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}(a, b) = \left(\frac{\tilde{Q}_{(m)} - \tilde{Q}_{(1)}}{b - a} \right)^m$$

if $a \leq \tilde{Q}_1, \ldots, \tilde{Q}_m \leq b$, where

$$\tilde{Q}_{(1)} = \min_{i=1,\ldots,m} \tilde{Q}_i \qquad \text{and} \qquad \tilde{Q}_{(m)} = \max_{i=1,\ldots,m} \tilde{Q}_i$$

are the first and last order statistic of $\tilde{Q}_1, \ldots, \tilde{Q}_m$, respectively. A well-known result states that the difference between the order statistics obtained from standard uniform variables is beta distributed [DavidNagaraja04], i.e., $\frac{\tilde{Q}_{(m)} - \tilde{Q}_{(1)}}{b-a} \sim \text{beta}(m-1, 2)$. Therefore, the CPF of the relative likelihood is given by

$$F_{\tilde{V}|a,b}(v) = \mathrm{P}(\tilde{V} \leq v) = \mathrm{P}\left(\frac{\tilde{Q}_{(m)} - \tilde{Q}_{(1)}}{b - a} \leq \sqrt[m]{v} \right) = F_{\text{beta}(m-1,1)}\left(\sqrt[m]{v} \right)$$

for all $v \in [0, 1]$, and the resulting Relative Likelihood-to-Confidence Transform reads

$$\gamma^{\Lambda}_{\hat{\theta}|q_1,\ldots,q_m}(a, b) = \begin{cases} F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)} - q_{(1)}}{b-a} \right) & \text{for } a \leq q_1, \ldots, q_m \leq b \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

These confidence distributions can be extended to the marginals $\delta_1 = a$ and $\delta_2 = b$,

and the feature $\delta_3 = c$ yielding the marginal/feature confidence distributions

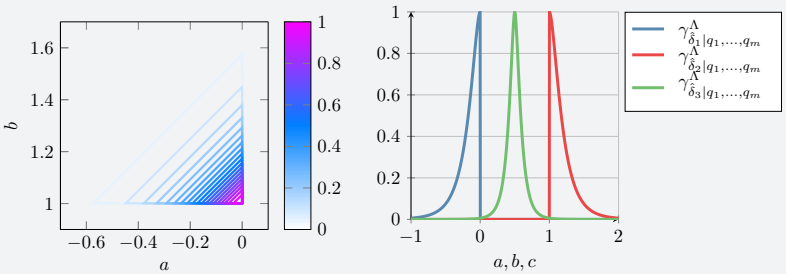
$$\gamma^{\Lambda}_{\hat{\delta}_1|q_1,\ldots,q_m}(a) = \begin{cases} F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)}-q_{(1)}}{q_{(m)}-a}\right) & \text{for } a \le q_1,\ldots,q_m \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

$$\gamma^{\Lambda}_{\hat{\delta}_2|q_1,\ldots,q_m}(b) = \begin{cases} F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)}-q_{(1)}}{b-q_{(1)}}\right) & \text{for } q_1,\ldots,q_m \le b \text{ and} \\ 0 & \text{otherwise,} \end{cases}$$

$$\begin{aligned} \gamma^{\Lambda}_{\hat{\delta}_3|q_1,\ldots,q_m}(c) &= \sup_{a,b\in\mathbb{R}\,:\,c=\frac{a+b}{2}} \gamma^{\Lambda}_{\hat{\theta}|q_1,\ldots,q_m}(a,b) = \sup_{a\in\mathbb{R}} \gamma^{\Lambda}_{\hat{\theta}|q_1,\ldots,q_m}(a,2c-a) \\ &= \sup_{a\in\mathbb{R}\,:\,a\le q_{(1)}\,\wedge\,2c-a\le q_{(m)}} F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)}-q_{(1)}}{2(c-a)}\right) \\ &= \begin{cases} F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)}-q_{(1)}}{2(q_{(m)}-c)}\right) & \text{if } c \le \frac{q_{(1)}+q_{(m)}}{2} \text{ and} \\ F_{\text{beta}(m-1,1)}\left(\frac{q_{(m)}-q_{(1)}}{2(c-q_{(1)})}\right) & \text{otherwise} \end{cases} \end{aligned}$$

after some analysis.

The following figures show the contour lines of the joint confidence distribution $\gamma^{\Lambda}_{\hat{\theta}|q_1,\ldots,q_m}$ (left), and its marginal/feature confidence distributions $\gamma^{\Lambda}_{\hat{\delta}_1|q_1,\ldots,q_m}$, $\gamma^{\Lambda}_{\hat{\delta}_2|q_1,\ldots,q_m}$ and $\gamma^{\Lambda}_{\hat{\delta}_3|q_1,\ldots,q_m}$ (right), all obtained after observing $q_i = \frac{i-1}{m-1}$ for $m = 11$.

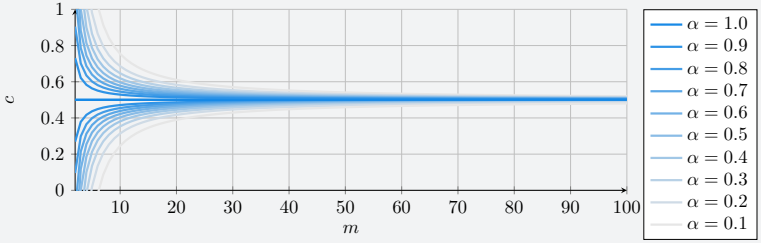


Joint Confidence Distribution Marginal/Feature Confidence Distributions

Alternatively, one could have formulated the statistical model of the observations $\tilde{Q}_1,\ldots,\tilde{Q}_m \sim \mathcal{U}(c-\frac{l}{2}, c+\frac{l}{2})$ for the vector of unknown parameters $\theta = [c,\, l]^{\mathrm{T}}$ composed of the hatch position c and the submersible's length $l = b - a$. The confidence distribution of c would then have been obtained by marginalizing over the nuisance parameter l.

It is also expedient to consider how the confidence distribution $\gamma^{\Lambda}_{\hat{\delta}_3|q_1,\ldots,q_m}$ gains expressiveness for increasing numbers of m, assuming that the first- and last-order statistics remain at $q_{(1)} = 0$ and $q_{(m)} = 1$, respectively, see below, where the upper

and lower limits of the corresponding confidence intervals are depicted as a function of m.

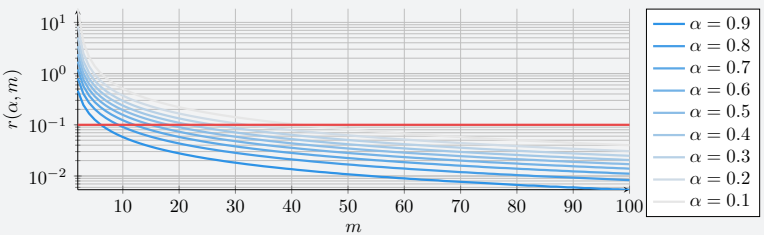


Confidence Interval Evolution

Depending on the desired level of upper or lower confidence and the allowed width

$$r(\alpha, m) = \left(\sup_{c \in \mathbb{R} : \gamma^{\Lambda}_{\delta_3 | q_1, \ldots, q_m}(c) > \alpha} c \right) - \left(\inf_{c \in \mathbb{R} : \gamma^{\Lambda}_{\delta_3 | q_1, \ldots, q_m}(c) > \alpha} c \right)$$

of the respective confidence intervals for $\alpha \in [0, 1]$ as shown below, one may then decide for how many bubbles to wait before dispatching the rescue line.



Confidence Interval Width r

If, e.g., the rescue team wants to ensure that the position of the hatch is located in a 90%-confidence interval with a length of no more than 0.1 distance units (red line), then the team could act after approximately 40 observed bubbles.

[43] Unfortunately, the scope of this thesis does not permit a detailed address of Morey et al.'s criticism of confidence procedures, but the framework of Π-IMs, in particular the Likelihood-to-Confidence Transform, avoids many of the described fallacies.

4.3.2.2 Confidence Combination

The second tool that is required for reasoning with confidence distributions concerns the construction of combined confidence distributions from individual ones that have been produced under different Π-IMs.

Consider, e.g., the case where two (or more) different Π-IMs for the same statistical model have been derived. Intuitively, it is not a recommendable procedure to see which confidence distributions they produce and then choose the more expressive one, or their minimum, because this would not preserve validity [MoreyEtAl16]. A combined Π-IM must be derived before seeing the data—just like a Bayesian prior must be chosen before seeing the data.

More generally, suppose that $\gamma_{\hat{\theta}_1|\tilde{Q}_1}, \ldots, \gamma_{\hat{\theta}_m|\tilde{Q}_m}$ are m Π-IMs corresponding to the m statistical models of $\tilde{Q}_1 \sim \mathrm{P}_{\tilde{Q}_1|\theta_1}, \ldots, \tilde{Q}_m \sim \mathrm{P}_{\tilde{Q}_m|\theta_m}$, respectively. The following discussion is restricted to stochastically independent imprecise observations, but the individual parameters may coincide, overlap or be interdependent, i.e., $\theta_1 = \varphi_1(\theta), \ldots, \theta_m = \varphi_m(\theta)$ are simply features of the 'overall' parameter vector θ, of which a combined Π-IM $\gamma_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ for the combined statistical model

$$(\tilde{Q}_1, \ldots, \tilde{Q}_m) \sim \mathrm{P}_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta} = C^{\mathrm{I}}\left(\mathrm{P}_{\tilde{Q}_1|\theta_1}, \ldots, \mathrm{P}_{\tilde{Q}_m|\theta_m}\right) \tag{4.43}$$

must first be constructed. The procedure to do just this, which is described in the following, is similar to the procedures described in Sections 2.3.3.2 and 3.5.

In the spirit of a natural extension, the inverse extension of the features $\theta_1, \ldots, \theta_m$ to the parameter θ under $\varphi_1, \ldots, \varphi_m$, respectively, yields the Π-IMs $\gamma_{\hat{\theta}|\tilde{Q}_i}$ given by

$$\gamma_{\hat{\theta}|q_i}(\theta) = \gamma_{\hat{\theta}_i|q_i}(\varphi_i(\theta)) \tag{4.44}$$

for all $q_i \in \mathbb{Q}_i$, all $\theta \in \Theta$ and all $i = 1, \ldots, m$; and under the Reverse Pivotal Step, the corresponding Π-IMs $\gamma_{\hat{\theta}_1|\tilde{Q}_1}, \ldots, \gamma_{\hat{\theta}_m|\tilde{Q}_m}$ define the m parameter-dependent descriptive distributions $\pi_{\tilde{Q}_i|\theta}$ given by

$$\pi_{\tilde{Q}_i|\theta}(q_i) = \gamma_{\hat{\theta}|q_i}(\theta) = \gamma_{\hat{\theta}_i|q_i}(\varphi_i(\theta)). \tag{4.45}$$

From these, it is possible to derive a combined parameter-dependent possibilistic description $\pi_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}$. More precisely, the combination of the above Π-IMs follows directly from the construction of joint IP-description distributions under the strong independence assumption as discussed in Section 3.5.4. If $\tilde{Q}_1, \ldots, \tilde{Q}_m$ are assumed to be stochastically independent, then the application of the SI-Π-copula $\mathcal{J}^{\mathrm{SI}}$ is appropriate and yields

$$\pi_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}(q_1, \ldots, q_m) = \mathcal{J}^{\mathrm{SI}}\left(\pi_{\tilde{Q}_1|\theta}(q_1), \ldots, \pi_{\tilde{Q}_m|\theta}(q_m)\right) \tag{4.46}$$

for all $\theta \in \Theta$ and all $q_1 \in \mathbb{Q}_1, \ldots, q_m \in \mathbb{Q}_m$. And under the Pivotal Step, this—again—defines a combined Π-IM $\gamma_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ via

$$\gamma_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}(\theta) = \pi_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}(q_1, \ldots, q_m) = \mathcal{J}^{\mathrm{SI}}\left(\gamma_{\hat{\theta}|q_1}(\theta), \ldots, \gamma_{\hat{\theta}|q_m}(\theta)\right) \tag{4.47}$$

for all $\theta \in \Theta$ and all $q_1 \in \mathbb{Q}_1, \ldots, q_m \in \mathbb{Q}_m$.

Proposition 46. *If $\gamma_{\hat{\theta}_1|\tilde{Q}_1}, \ldots, \gamma_{\hat{\theta}_m|\tilde{Q}_m}$ are valid with respect to the corresponding statistical models $\mathrm{P}_{\tilde{Q}_1|\theta_1}, \ldots, \mathrm{P}_{\tilde{Q}_m|\theta_m}$, then $\gamma_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ is valid with respect to $\mathrm{P}_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}$.*

Proof. By Proposition 45, the Π-IMs $\gamma_{\hat{\theta}|\tilde{Q}_1}, \ldots, \gamma_{\hat{\theta}|\tilde{Q}_m}$ are valid with respect to the statistical models $\mathrm{P}_{\tilde{Q}_1|\theta}, \ldots, \mathrm{P}_{\tilde{Q}_m|\theta}$ under the given assumptions; but then, by Proposition 42, the IP-description distributions $\pi_{\tilde{Q}_1|\theta}, \ldots, \pi_{\tilde{Q}_m|\theta}$ obtained by the Reverse Pivotal Step are also consistent with $\mathrm{P}_{\tilde{Q}_1|\theta}, \ldots, \mathrm{P}_{\tilde{Q}_m|\theta}$ for all $\theta \in \Theta$. And finally, by Proposition 37, $\pi_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}$ obtained under the SI-Π-copula is consistent with $\mathrm{P}_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}$ for all $\theta \in \Theta$. Therefore, $\gamma_{\hat{\theta}|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ is valid with respect to $\mathrm{P}_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta}$ by application of Proposition 42 to the Pivotal Step in Eq. (4.47). □

Furthermore, notice that the SI-Π-copula is non-decreasing in all arguments, and therefore efficiency, too, is preserved. That is, more efficient individual Π-IMs will always yield more efficient joint Π-IMs.

In summary, the SI-Π-copula is not only applicable for the construction of joint IP-description distributions under the assumption of strong independence but also for the construction of combined confidence distributions under the assumption of stochastic independence of the data, and the preservation of consistency for the former immediately implies the preservation of validity for the latter—just like the specificity preservation implies the preservation of the efficiency order.

This general rule for the combination of combined Π-IMs encompasses several special setups.

Joint Confidence Most obviously, the combined Π-IM can further be (explicitly) extended under $\varphi : \theta \mapsto (\varphi_1(\theta), \ldots, \varphi_m(\theta))$ ultimately yielding the joint Π-IM $\gamma_{\hat{\theta}_1,\ldots,\hat{\theta}_m|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ given by

$$\begin{aligned} & \gamma_{\hat{\theta}_1,\ldots,\hat{\theta}_m|q_1,\ldots,q_m}(\theta_1, \ldots, \theta_m) \\ = & \sup_{\theta \in \Theta : \varphi(\theta) = (\theta_1,\ldots,\theta_m)} \mathcal{J}^{\mathrm{SI}}\left(\gamma_{\hat{\theta}|q_1}(\varphi_1(\theta)), \ldots, \gamma_{\hat{\theta}|q_m}(\varphi_m(\theta))\right) \\ = & \mathcal{J}^{\mathrm{SI}}\left(\gamma_{\hat{\theta}_1|q_1}(\theta_1), \ldots, \gamma_{\hat{\theta}_m|q_m}(\theta_m)\right) \end{aligned} \tag{4.48}$$

for all $\theta_1 \in \Theta_1, \ldots, \theta_m \in \Theta_m$ and all $q_1 \in \mathbb{Q}_1, \ldots, q_m \in \mathbb{Q}_m$. In this manner, a joint confidence distribution is computed from marginal ones.

Confidence Conjunction If $\theta = \theta_1 = \ldots = \theta_m$, the Π-IM combination rule describes how to compute the conjunction of the individual Π-IMs $\gamma_{\hat{\theta}|\tilde{Q}_1}, \ldots, \gamma_{\hat{\theta}|\tilde{Q}_m}$, i.e. the 'intersection'

$$\gamma_{\hat{\theta}|q_1,\ldots,q_m}(\theta_1, \ldots, \theta_m) = \mathcal{J}^{\mathrm{SI}}\left(\gamma_{\hat{\theta}|q_1}(\theta), \ldots, \gamma_{\hat{\theta}|q_m}(\theta)\right) \tag{4.49}$$

of the respective confidence distributions for all $\theta \in \Theta$ and all $q_1 \in \mathbb{Q}_1, \ldots, q_m \in \mathbb{Q}_m$. If, e.g., m iid experiments are performed, for which marginal statistical models and the corresponding (coinciding) Π-IMs $\gamma_{\hat{\theta}|\tilde{Q}_1} = \ldots = \gamma_{\hat{\theta}|\tilde{Q}_m}$ are available, but the joint statistical model is not available or difficult to evaluate, then one may simply employ Eq. (4.49) in order to construct the joint Π-IM instead.

Confidence Updating Similar to the conjunction of confidence, the combination of Π-IMs also provides a simple way of recursively 'updating' a confidence distribution $\gamma_{\hat{\theta}|q_1,\ldots,q_{m-1}}$ depending on 'old' data $\tilde{Q}_1 = q_1, \ldots, \tilde{Q}_{m-1} = q_{m-1}$ once 'new' data $\tilde{Q}_m = q_m$ become available, resulting in

$$\gamma_{\hat{\theta}|q_1,\ldots,q_m}(\theta_1, \ldots, \theta_m) = \mathcal{J}^{\mathrm{SI}}\left(\gamma_{\hat{\theta}|q_1,\ldots,q_{m-1}}(\theta), \gamma_{\hat{\theta}|q_m}(\theta)\right) \tag{4.50}$$

for all $\theta \in \Theta$ and all $q_1 \in \mathbb{Q}_1, \ldots, q_m \in \mathbb{Q}_m$. This updating technique shall prove useful in the remainder of this thesis, in particular in Chapter 6.

Naturally, one may also construct combined Π-IMs from other strong-independence Π-copulae, e.g. for the maximum- and the product-based ones. Very curiously, the application of the latter leads to the well-known formula for the combination of independent p-values described by Fisher and refined by Jost as explained in Section 3.5.4. Otherwise, if the independence of the data cannot be guaranteed, then the UI-Π-copula is applicable and yields Bonferroni's rule for the combination of p-values without any dependency assumptions [CramerKamps20].

4.3.3 Predictive Inference

This section shows a simple method of constructing Π-PMs based on possibilistic IP descriptions and Π-IMs.

4.3.3.1 The Semi-Pivotal Step

A simple method of finding a perceptive Π-PM $\kappa_{\tilde{V}|\tilde{Q}}$ of $\tilde{V}$ given $\tilde{Q}$ is finding a possibilistic description $\pi_{\tilde{V},\tilde{Q}}$ of the statistical model $\mathfrak{S}$ in combination with the *Semi-Pivotal Step* producing the prediction distributions given by

$$\kappa_{\tilde{V}|q}(v) = \pi_{\tilde{V},\tilde{Q}}(v, q) \tag{4.51}$$

for all $v \in \mathbb{V}$ and all $q \in \mathbb{Q}$ because, then, the perceptiveness in Eq. (4.23) follows directly from the consistency $\mathrm{P}_{\tilde{V},\tilde{Q}|\theta} \preceq \pi_{\tilde{V},\tilde{Q}}$ for all $\theta \in \Theta$.

The possibilistic description $\pi_{\tilde{V},\tilde{Q}}$ may stem from the disjunction of the parameter-dependent descriptions $\pi_{\tilde{V},\tilde{Q}|\theta}$, i.e.

$$\pi_{\tilde{V},\tilde{Q}}(v, q) = \sup_{\theta \in \Theta} \pi_{\tilde{V},\tilde{Q}|\theta}(v, q), \tag{4.52}$$

for all $v \in \mathbb{V}$ and all $q \in \mathbb{Q}$. Moreover, the joint distribution $\pi_{\tilde{V},\tilde{Q}|\theta}$ can be constructed from two (independently derived) marginal possibilistic descriptions $\pi_{\tilde{V}|\theta}$ and $\pi_{\tilde{Q}|\theta}$ of $\tilde{V}$ and $\tilde{Q}$, respectively, i.e. via

$$\pi_{\tilde{V},\tilde{Q}|\theta}(v,q) = \mathcal{J}\left(\pi_{\tilde{V}|\theta}(v), \pi_{\tilde{Q}|\theta}(q)\right) \tag{4.53}$$

for all $\theta \in \Theta$ and an appropriate Π-copula $\mathcal{J}$. Usually, if $\tilde{V}$ and $\tilde{Q}$ are assumed to be independent, $\mathcal{J} = \mathcal{J}^{\mathrm{SI}}$ is an appropriate choice. However, if either of the involved elementary possibility functions is vacuous, then—by Proposition 36—one may also choose $\mathcal{J} = \mathcal{J}^{\mathrm{NI}}$ yielding more specific results.

Ultimately, this yields the *Possibility-to-Prediction Transform* given by

$$\kappa_{\tilde{V}|q}(v) = \sup_{\theta\in\Theta} \mathcal{J}\left(\pi_{\tilde{V}|\theta}(v), \pi_{\tilde{Q}|\theta}(q)\right) \tag{4.54}$$

for all $v \in \mathbb{V}$ and all $q \in \mathbb{Q}$.

4.3.3.2 From Confidence to Predictions

The Possibility-to-Prediction Transform in Eq. (4.54) implies that—under the Reverse Pivotal Step—the confidence distributions obtained from a valid Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ may be used for the construction of a perceptive Π-PM $\kappa_{\tilde{V}|\tilde{Q}}$ producing the prediction distributions

$$\kappa_{\tilde{V}|q}(v) = \sup_{\theta\in\Theta} \mathcal{J}\left(\pi_{\tilde{V}|\theta}(v), \gamma_{\hat{\theta}|q}(\theta)\right) \tag{4.55}$$

for all $v \in \mathbb{V}$ and all $q \in \mathbb{Q}$.

In conclusion, it is straightforward to predict future realizations of the observed variable $\tilde{V}$ by Eq. (4.55) if a valid Π-IM of θ and a parameter-dependent description of $\tilde{V}$ are available. A general procedure for constructing such Π-PMs is illustrated by the following example.

Example 11: Exponential Π-PM

Suppose that, e.g., the overall reliability of a batch of identically constructed system components is to be evaluated. For this purpose, $m = 5$ components are selected at random from this batch, they are subjected to an endurance test, and their life spans are documented. Assuming that aging processes are negligible for the overall reliability of the component, it is reasonable to assume that the life spans are m iid imprecise variables $\tilde{Q}_1, \ldots, \tilde{Q}_m \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ following an exponential distribution given by the probability density function

$$p_{\tilde{Q}_i|\theta}(q_i) = \frac{1}{\theta}\exp\left(-\frac{q_i}{\theta}\right)$$

with the unknown mean $\theta > 0$ for all $q_i \geq 0$ and $i = 1, \ldots, m$.
The confidence distribution $\gamma^{\Lambda}_{\hat{\theta}|q_1,\ldots,q_m}$ obtained by the Relative Likelihood-to-

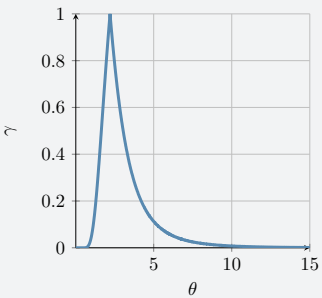
Confidence Transform after observing[44]

$$q_1 = 2.16, \quad q_2 = 0.447, \quad q_3 = 8.13, \quad q_4 = 0.006 \quad \text{and} \quad q_5 = 0.279$$

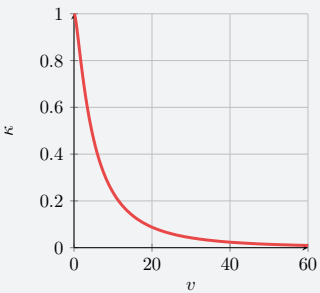
is shown below on the left.
In order to infer how long a sixth component from this batch will endure, a parameter-dependent predictive distribution of $\tilde{V} = \tilde{Q}_6 \sim \mathcal{E}(\theta)$ is computed under the Complementary Cumulative P-Π-transform yielding

$$\pi_{\tilde{V}|\theta}(v) = \exp\left(-\frac{v}{\theta}\right)$$

for all $v \geq 0$, which also coincides with the Optimal P-Π-transform. Finally, evaluating Eq. (4.55) under the assumption of independence produces the prediction distribution $\kappa_{\tilde{V}|q_1,\ldots,q_5}$ shown below on the right.



Confidence Distribution $\gamma^{\Lambda}_{\tilde{\theta}|q_1,\ldots,q_5}$



Feature Distribution $\kappa_{\tilde{V}|q_1,\ldots,q_5}$

Regarding the interpretation of $\kappa_{\tilde{V}|q_1,\ldots,q_5}$, much has been said already. It is, e.g., clear that for $\tilde{V}$ to be greater than, e.g. 60, some rare event with less than 1% probability would have to be in the process of ocurring—it has not ocurred already because $\tilde{V} = v$ has not yet been observed—since $\overline{K}_{\tilde{V}|q}((60, +\infty)) < 0.01$. Conversely, it is entirely plausible that $\tilde{V} \leq 10$ because $\overline{K}_{\tilde{V}|q}([0, 10]) = 1$, but it is in no way guaranteed since $\underline{K}_{\tilde{V}|q}([0, 10]) < 0.24$, i.e., the a-priori probability that $\tilde{Q} = q$ and $\tilde{V} \in [0, 10]$ is less than 24%.

Being based on Π-copulae, a certain degree of conservatism, i.e. a loss of efficiency compared to other approaches, such as [MartinLingham16], is to be expected.

[44] These data points correspond to problem #5 of Ferson's exemplar problems for risk assessments from sparse information available under `https://sites.google.com/site/epistemicunc/exemplar-problems` (accessed September 7, 2021).

4.3.4 Reasoning with Predictions

Finally, it is straightforward to show that, due to their very similar nature, reasoning with prediction distributions produced by perceptive Π-PMs, i.e. extending and/or combining them, is similar to reasoning with confidence distributions produced by valid Π-IMs. Therefore, the discussion Section 4.3.2 need not be repeated for Π-PMs and only the main results are stated.

The extension of perceptive Π-PMs to other Π-PMs is treated in the following proposition.

Proposition 47. *Let $\kappa_{\tilde{V}|\tilde{Q}}$ be a perceptive Π-PM of the imprecise variable $\tilde{V}$ given the observation $\tilde{Q}$, and let the $\mathbb{T}$-valued imprecise variable $\tilde{T}$ be connected to $\tilde{V}$ via the implicit relationship $0 = \Xi(\tilde{T}, \tilde{V})$. Then, the Π-PM $\kappa_{\tilde{T}|\tilde{Q}}$ that produces the corresponding prediction distributions $\kappa_{\tilde{T}|q} : \mathbb{T} \to [0,1]$, which are computed by extending $\kappa_{\tilde{V}|q}$ via*

$$\kappa_{\tilde{T}|q}(t) = \sup_{v \in \mathbb{V} \,:\, 0 = \Xi(t,v)} \kappa_{\tilde{V}|q}(v) \tag{4.56}$$

for all $q \in \mathbb{Q}$ and all $t \in \mathbb{T}$, is a perceptive Π-PM of $\tilde{T}$ given $\tilde{Q}$.

Proof. Similar to the discussion in Section 3.2.2, the statistical model is the set of all pushforward probability measures $\mathfrak{S} = \{\mathrm{P}_{\tilde{T},\tilde{V},\tilde{Q}|\theta} : \mathrm{P}_\theta \in \mathfrak{S}^0\}$ under $\tilde{T}$, $\tilde{V}$ and $\tilde{Q}$ from the underlying statistical model $\mathfrak{S}^0 = \{\mathrm{P}_\theta \in \mathbb{P}(\Omega, \Sigma) : \theta \in \Theta\}$ on (Ω, Σ). Additionally, the imprecise variables $\tilde{T}$, and $\tilde{V}$ satisfy the implicit relationship $0 = \Xi(\tilde{T}(\omega), \tilde{V}(\omega))$ for all $\omega \in \Omega$. The proposition follows by verifying that, under the perceptiveness of $\kappa_{\tilde{V}|\tilde{Q}}$,

$$\begin{aligned}
& \mathrm{P}_{\tilde{T},\tilde{Q}|\theta}\left(\left\{(t,q) \in \mathbb{T} \times \mathbb{Q} : \kappa_{\tilde{T}|q}(t) \leq \alpha\right\}\right) \\
= \;& \mathrm{P}_{|\theta}\left(\left\{\omega \in \Omega : \kappa_{\tilde{T}|\tilde{Q}(\omega)}\left(\tilde{T}(\omega)\right) \leq \alpha\right\}\right) \\
= \;& \mathrm{P}_{|\theta}\left(\left\{\omega \in \Omega : \sup_{v \in \mathbb{V} \,:\, 0 = \Xi\left(\tilde{T}(\omega), v\right)} \kappa_{\tilde{V}|\tilde{Q}(\omega)}(v) \leq \alpha\right\}\right) \\
\leq \;& \mathrm{P}_{|\theta}\left(\left\{\omega \in \Omega : \kappa_{\tilde{V}|\tilde{Q}(\omega)}\left(\tilde{V}(\omega)\right) \leq \alpha\right\}\right) \\
= \;& \mathrm{P}_{\tilde{V},\tilde{Q}|\theta}\left(\left\{(v,q) \in \mathbb{V} \times \mathbb{Q} : \kappa_{\tilde{V}|q}(v) \leq \alpha\right\}\right) \\
\leq \;& \alpha
\end{aligned}$$

for all $\theta \in \Theta$ and all $\alpha \in [0,1]$. □

Of course, this also allows, e.g., for a marginalization of 'nuisance' variables.

The decomposition of statistical inference into smaller subproblems and the subsequent combination of the obtained Π-PMs is enabled by the following proposition.

Proposition 48. *Let $\kappa^{(1)}_{\tilde{V}|\tilde{Q}_1}$ and $\kappa^{(2)}_{\tilde{V}|\tilde{Q}_2}$ be two perceptive Π-PMs of the imprecise variable $\tilde{V}$ given the observations $\tilde{Q}_1$ and $\tilde{Q}_2$, respectively. Then, the Π-PM $\kappa_{\tilde{V}|\tilde{Q}_1,\tilde{Q}_2}$ producing the*

prediction distributions $\kappa_{\tilde{V}|q_1,q_2} : \mathbb{V} \to [0,1]$*, which are computed by combining* $\kappa^{(1)}_{\tilde{V}|q_1}$ *and* $\kappa^{(2)}_{\tilde{V}|q_2}$ *via*

$$\kappa_{\tilde{V}|q_1,q_2} = \mathcal{J}^{\mathrm{UI}}\left(\kappa^{(1)}_{\tilde{V}|q_1}, \kappa^{(2)}_{\tilde{V}|q_2}\right) \tag{4.57}$$

for all $q_1 \in \mathbb{Q}_1$*, all* $q_2 \in \mathbb{Q}_2$*, and all* $v \in \mathbb{V}$ *is a perceptive* Π*-PM of* $\tilde{V}$ *given* $\tilde{Q}_1$ *and* $\tilde{Q}_2$*.*

Proof. The proof is similar to those of Propositions 35, 36, 37, and 38. The perceptiveness of $\kappa_{\tilde{V}|\tilde{Q}_1,\tilde{Q}_2}$ follows directly from that of $\kappa^{(1)}_{\tilde{V}|\tilde{Q}_1}$ and $\kappa^{(2)}_{\tilde{V}|\tilde{Q}_2}$ by verifying that

$$\begin{aligned}
& \mathrm{P}_{\tilde{V},\tilde{Q}_1,\tilde{Q}_2|\theta}\left(\left\{(v,q_1,q_2) \in \mathbb{V}\times\mathbb{Q}_1\times\mathbb{Q}_2 : \kappa_{\tilde{V}|q_1,q_2}(v) \leq \alpha\right\}\right) \\
= \; & \mathrm{P}_{\tilde{V},\tilde{Q}_1,\tilde{Q}_2|\theta}\left(\left\{(v,q_1,q_2) \in \mathbb{V}\times\mathbb{Q}_1\times\mathbb{Q}_2 : \mathcal{J}^{\mathrm{UI}}\left(\kappa^{(1)}_{\tilde{V}|q_1}(v), \kappa^{(2)}_{\tilde{V}|q_2}(v)\right) \leq \alpha\right\}\right) \\
\leq \; & \mathrm{P}_{\tilde{V},\tilde{Q}_1,\tilde{Q}_2|\theta}\left(\left\{(v,q_1,q_2) \in \mathbb{V}\times\mathbb{Q}_1\times\mathbb{Q}_2 : 2\kappa^{(1)}_{\tilde{V}|q_1}(v) \leq \alpha\right\}\right) \\
& \qquad + \mathrm{P}_{\tilde{V},\tilde{Q}_1,\tilde{Q}_2|\theta}\left(\left\{(v,q_1,q_2) \in \mathbb{V}\times\mathbb{Q}_1\times\mathbb{Q}_2 : 2\kappa^{(2)}_{\tilde{V}|q_2}(v) \leq \alpha\right\}\right) \\
= \; & \mathrm{P}_{\tilde{V},\tilde{Q}_1|\theta}\left(\left\{(v,q_1) \in \mathbb{V}\times\mathbb{Q}_1\times : \kappa^{(1)}_{\tilde{V}|q_1}(v) \leq \frac{\alpha}{2}\right\}\right) \\
& \qquad + \mathrm{P}_{\tilde{V},\tilde{Q}_2|\theta}\left(\left\{(v,q_2) \in \mathbb{V}\times\mathbb{Q}_2 : \kappa^{(2)}_{\tilde{V}|q_2}(v) \leq \frac{\alpha}{2}\right\}\right) \\
\leq \; & \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha
\end{aligned}$$

for all $\theta \in \Theta$ and all $\alpha \in [0,1]$. □

Without additional assumptions, this result cannot be tightened with a more specific Π-copula, such as $\mathcal{J}^{\mathrm{SI}}$, because, e.g., stochastic independence between $(\tilde{V},\tilde{Q}_1)$ and $(\tilde{V},\tilde{Q}_2)$ is impossible to have.

4.3.5 Subjectivist, Fiducial and Likelihoodist Aspects

As a final remark, Eq. (4.55) suggests an alternative interpretation of descriptive possibility distributions by establishing a deeper relationship between description, prediction and confidence distributions.

If all that is known about the true parameter is $\theta^* \in T$ for some $T \in \mathbb{B}(\Theta)$, an appropriate possibilistic description of $\tilde{V}$ is given by the disjunction of the parameter-dependent possibilistic descriptions $\pi_{\tilde{V}|\theta}$, which is just another way to write the IP-Π-transform (2.51), and reads

$$\begin{aligned}
\pi_{\tilde{V}}(v) = \sup_{\theta\in T} \pi_{\tilde{V}|\theta}(v) &= \sup_{\theta\in T} \min\left(\pi_{\tilde{V}|\theta}(v), 1\right) \\
&= \sup_{\theta\in\Theta} \mathcal{J}^{\mathrm{NI}}\left(\pi_{\tilde{V}|\theta}(v), \mathcal{I}_T(\theta)\right) = \kappa_{\tilde{V}|\emptyset}(v)
\end{aligned} \tag{4.58}$$

for all $v \in \mathbb{V}$. The application of the non-interactive Π-copula is warranted by Lemma 20. This expression corresponds to that of a prediction distribution produced by the Π-PM $\kappa_{\tilde{V}|\tilde{Q}}$,

in the absence of data, written as $\tilde{Q} = \emptyset$. That is, the (quasi-)vacuous distribution $\mathcal{Q}(T)$ is the appropriate way to model the knowledge that $\theta^* \in T$; and, Eq. (4.55) simply constitutes a generalization of the disjunction of parameter-dependent possibilistic descriptions.

Put differently, one may say that description distributions are a particular case of prediction distributions in the absence of data. Conversely, prediction distributions can be viewed as description distributions conditional on some data/observations.

Therefore, one can argue that possibility theory unifies several ostensibly opposing points of view. Adopting the frequentist perspective, which is arguably closer to the concept of IMs, the lack of data $\tilde{Q} = \emptyset$ would be modeled via the confidence distribution $\gamma_{\hat{\theta}|\emptyset} = \mathcal{I}_T$. Nevertheless, it might equally well be said in a subjectivist framework that this should be modeled via a description distribution $\pi_{\tilde{\theta}|\emptyset} = \mathcal{I}_T$ of an imprecise variable $\tilde{\theta} \sim \mathcal{Q}(T)$ of which only the support T is known. Both result in the same elementary possibility function and are manipulated in the same way—via the Extension Principle. In this sense, the border between frequentism and subjectivism blurs in possibility theory.

As mentioned already, the Possibilistic Inference Principle in Eq. (4.38) also appears to be a fiducial technique since it allows to compute confidence distributions from description distributions via a simple extension, and the Pivotal Step resembles a possibilistic likelihood concept. That is, possibilistic inference weakly includes many of the predominant theories of uncertainty; however, not in the original sense. Moreover, it excludes Bayesianism, which is not to be confused with subjectivism, e.g., by rejecting Bayes' rule as a way of updating beliefs. This rule is replaced by the various conjunction rules.

It is interesting to observe that possibilistic statistics can also be applied to 'deterministic' statistical models. For instance, if both $\pi_{\tilde{Q}|\theta} = \mathrm{P}_{\tilde{Q}|\theta} = \mathcal{D}(\theta)$ are deterministic, indicating that $\tilde{Q} = \theta$ for $\theta \in \Theta$, the Pivotal Step yields

$$\gamma_{\hat{\theta}|q}(\theta) = \pi_{\tilde{Q}|\theta}(q) = \begin{cases} 1 & \text{if } q = \theta \text{ and} \\ 0 & \text{otherwise.} \end{cases} \tag{4.59}$$

That is, $\gamma_{\hat{\theta}|q} = \mathcal{D}(q)$ is given by the deterministic distribution, too, which is consistent with intuition and with the arguments in Sections 3.1.3. Moreover, it is consistent with elementary arithmetical concepts. Suppose, for instance, that $\tilde{Q} = \tilde{V}^2$, where $\tilde{V} \sim \mathcal{D}(\theta)$ for some $\theta \in \mathbb{R}$ and $\tilde{Q} = q \geq 0$ is observed. Then, the Possibilistic Inference Principle in Eq. (4.38) states that a confidence distribution of θ is given by

$$\gamma_{\hat{\theta}|q}(\theta) = \begin{cases} 1 & \text{if } \theta^2 = q \\ 0 & \text{otherwise.} \end{cases} \tag{4.60}$$

for all $\theta \in \mathbb{R}$. This is nothing other than stating that, in this case, both $\theta = +\sqrt{q}$ and $\theta = -\sqrt{q}$ are possible.

Chapter 5

Membership Computations

The purpose of computing is insight, not numbers.

Richard Hamming [Hamming12]

The preceding chapters have, above all else, shown that the calculus of possibility theory applies to IP-description distributions, confidence distributions, and prediction distributions alike and is, furthermore, restricted to a few fundamental operations that allow for the solution of a large variety of problems. The following chapter is intended to showcase some numerical strategies for the various Transforms described thus far and for the description of membership functions and their manipulation, most importantly via extension.

The prefix 'μ', e.g., in the following definition of the μ-transform, will usually indicate that functions and classes corresponding to these concepts should also be present in a numerical implementation.

5.1 Membership Transforms

At the core, the (I)P-Π-Trafo and the P-Γ-transform require one to be able to evaluate the *Membership Transform* producing the (parameter-dependent) descriptions $\pi_{\tilde{V}|\theta} : \mathbb{V} \to [0,1]$ given by

$$\pi_{\tilde{V}|\theta}(v) = \mathrm{P}_{\tilde{V}|\theta}\left(\left\{\xi \in \mathbb{V} : \rho_{\tilde{V}|\theta}(\xi) \leq \rho_{\tilde{V}|\theta}(v)\right\}\right) \tag{5.1}$$

under a given parameter-dependent plausibility distribution $\rho_{\tilde{V}|\theta} : \mathbb{V} \to [0,1]$ for all $v \in \mathbb{V}$ and all $\theta \in \Theta$. The reasons for its name shall become clear in Section 5.2.

These parameter-dependent description distributions constitute the basic building block of any possibilistic analysis because, from them, one may derive

IP Description Distributions $\pi_{\tilde{V}} : \mathbb{V} \to [0,1]$ resulting from the IP-Π-transform of a

parametric family of probability distributions $\mathfrak{P}_{\tilde{V}} = \{\mathrm{P}_{\tilde{V}|\theta} \in \mathbb{P}(\mathbb{V}, \mathbb{B}(\mathbb{V})) : \theta \in \Theta\}$ under the (parameter-independent) plausibility distribution $\rho_{\tilde{V}|\theta} = \rho_{\tilde{V}}$ for all $\theta \in \Theta$, which may be expressed as the disjunction

$$\pi_{\tilde{V}}(v) = \sup_{\theta \in \Theta} \pi_{\tilde{V}|\theta}(v) \tag{5.2}$$

for all $v \in \mathbb{V}$,

Confidence Distributions $\gamma_{\hat{\theta}|q} : \Theta \to [0, 1]$, which are produced by a Π-IM $\gamma_{\hat{\theta}|\tilde{Q}}$ given some observation $\tilde{Q} = q$, and are obtained under the Pivotal Step

$$\gamma_{\hat{\theta}|q}(\theta) = \pi_{\tilde{Q}|\theta}(q) \tag{5.3}$$

for all $\theta \in \Theta$, and

Prediction Distributions $\kappa_{\tilde{V}|q} : \mathbb{V} \to [0, 1]$, which are produced by a Π-PM $\kappa_{\tilde{V}|\tilde{Q}}$ under some observation $\tilde{Q} = q$, and are, e.g., given by

$$\kappa_{\tilde{V}|q}(v) = \sup_{\theta \in \Theta} \mathcal{J}\left(\pi_{\tilde{V}|\theta}(v), \pi_{\tilde{Q}|\theta}(q)\right) \tag{5.4}$$

for all $v \in \mathbb{V}$ and an appropriate Π-copula $\mathcal{J}$.

Henceforth, Eq. (5.1) will also be referred to as the *μ-transform* of $\mathrm{P}_{\tilde{V}|\theta}$ under $\rho_{\tilde{V}|\theta}$.

The fact that one may potentially need to evaluate the μ-transform for all $\theta \in \Theta$ and/or all $v \in \mathbb{V}$ necessitates an efficient evaluation thereof—especially when the implied probabilities are not directly available. In principle, every numerical quadrature rule may be used to compute the respective probabilities, but especially Monte-Carlo integration methods naturally lend themselves to this cause due to their straightforward numerical implementation.

5.1.1 Approximate Membership Transform

Monte-Carlo methods are widely applied in probabilistic analyses [RobertCasella13]. As Sullivan observes, "[they] are, in essence, an application of the Law of Large Numbers" [Sullivan15, p. 178], which states that, if $\tilde{V}_1, \ldots, \tilde{V}_m$ are m iid imprecise variables following the distribution of the imprecise variable $\tilde{V} \sim \mathrm{P}_{\tilde{V}}$, and if $\varphi : \mathbb{V} \to \mathbb{R}$ is a measurable function, such that the expectation $\mathfrak{E}_{\mathrm{P}_{\tilde{V}}}[\varphi(\tilde{V})]$ is finite, then the average of $\varphi(\tilde{V}_1), \ldots, \varphi(\tilde{V}_m)$ converges to this expectation both in a weak and a strong sense [Jennrich69, Sullivan15]. The basic 'Vanilla Monte-Carlo' algorithm for numerical quadrature simply approximates the expected value of $\varphi(\tilde{V})$ via the average

$$\mathfrak{E}_{\mathrm{P}_{\tilde{V}}}[\varphi(\tilde{V})] \approx \frac{1}{m} \sum_{j=1}^{m} \varphi(\tilde{V}_j). \tag{5.5}$$

Applied to the μ-transform in Eq. (5.1), $\pi_{\tilde{V}|\theta}(v)$ may be understood as the expected value of $\varphi_{v,\theta}(\tilde{V})$ for the function $\varphi_{v,\theta} : \mathbb{V} \to \{0,1\}$ given by

$$\varphi_{v,\theta}(\xi) = \begin{cases} 1 & \text{if } \rho_{\tilde{V}|\theta}(\xi) \leq \rho_{\tilde{V}|\theta}(v) \quad \text{and} \\ 0 & \text{otherwise,} \end{cases} \tag{5.6}$$

That is, if a sampling procedure—e.g. inverse transform sampling, rejection sampling, or a Markov-Chain Monte-Carlo method—generates the m iid realizations $\tilde{V}_1, \ldots, \tilde{V}_m$ of $\tilde{V} \sim \mathrm{P}_{\tilde{V}|\theta}$, the Vanilla Monte-Carlo algorithm can be used to approximate $\pi_{\tilde{V}|\theta}(v)$ via the average

$$I_m^{\mathrm{MC}}(v,\theta) = \frac{1}{m} \sum_{j=1}^{m} \varphi_{v,\theta}(\tilde{V}_j) \tag{5.7}$$

for all $\theta \in \Theta$ and all $v \in \mathbb{V}$, i.e. via the empirical frequency of realizations producing a lower plausibility than v. Being bounded, the $\{0,1\}$-valued imprecise variable $\varphi_{v,\theta}(\tilde{V})$ certainly has a finite expectation $\mathfrak{E}_{\mathrm{P}_{\tilde{V}|\theta}}[\varphi_{v,\theta}(\tilde{V})] \in [0,1]$, and therefore, the approximation

$$\pi_{\tilde{V}|\theta}(v) = \mathfrak{E}_{\mathrm{P}_{\tilde{V}|\theta}}\left[\varphi_{v,\theta}(\tilde{V})\right] \approx I_m^{\mathrm{MC}}(v,\theta) \tag{5.8}$$

is justified.

The idea for such sample-based constructions of possibility distributions was first described by Hanselowski et al. [HanselowskiIhrleHanss15], who termed the resulting superlevel sets *percentage sets*. A conceptually similar approach had also been proposed earlier by Masson and Denœux [MassonDenœux06]. While the latter method was intentionally developed for encoding the sampling (probability) distribution, the former was not initially developed for a possibilistic IP theory. Only later, it was also discussed with respect to the proposed framework of quantitative possibility theory by Hose and Hanss [HoseHanss19d], who reframed it as a method for approximating the P-Π-transform [HoseHanss21c] and proposed a reliable variant [HoseHanss20].

The need for a reliable variant may be explained via the main impediment of the Approximate μ-transform, the fact that general reliability guarantees regarding the approximation of $\pi_{\tilde{V}|\theta}(v)$ via $I_m^{\mathrm{MC}}(v,\theta)$ are difficult to give. Of course, for large sample numbers m, the estimator $I_m^{\mathrm{MC}}(v,\theta)$ is guaranteed to converge to $\pi_{\tilde{V}|\theta}(v)$, and, from the Chebychev inequality in Eq. (3.71), it follows that

$$\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m|\theta}\left(\left|\pi_{\tilde{V}|\theta}(v) - I_m^{\mathrm{MC}}(v,\theta)\right| \geq t\right) \leq \frac{1}{nt^2}\mathrm{Var}_{\mathrm{P}_{\tilde{V}|\theta}}\left[\varphi_{v,\theta}(\tilde{V})\right]. \tag{5.9}$$

Yet, it is impossible to determine an appropriate sample size m in order to guarantee a certain distance $t > 0$ for a given reliability level $\gamma = \frac{1}{nt^2}\mathrm{Var}_{\mathrm{P}_{\tilde{V}|\theta}}[\varphi_{v,\theta}(\tilde{V})]$ because the moments of $\varphi_{v,\theta}(\tilde{V})$ are generally unknown. A remedy for this shortcoming is addressed in the following section.

5.1.2 Reliable Membership Transform

Recalling that an elementary possibility function that is less specific than some other elementary possibility function also encodes all the information contained in the latter — the credal set of the former also contains that of the latter—it is sensible to try and find a reliable upper bound of $\pi_{\tilde{V}|\theta}(v)$ in the μ-transform in Eq. (5.1). This idea is pursued in the following.

The imprecise variable $\varphi_{v,\theta}(\tilde{V})$ is Bernoulli distributed with the unknown success probability $\pi_{\tilde{V}|\theta}(v)$. Therefore, the imprecise variable

$$\tilde{Q} = \tilde{Q}(v,\theta) = \sum_{j=1}^{m} \varphi_{v,\theta}(\tilde{V}_j), \tag{5.10}$$

counting the number of successes where $\rho_{\tilde{V}|\theta}(\tilde{V}_j) \leq \rho_{\tilde{V}|\theta}(v)$, follows a binomial distribution and all of the Π-IMs described in Ex. 9 are also appropriate for finding a confidence distribution of the success probability $\pi_{\tilde{V}|\theta}(v)$. since it is of particular interest to find an upper bound, the Cumulative P-Γ-transform yields the most suitable Π-IM $\gamma_{\hat{\pi}_{\tilde{V}|\theta}(v)|\tilde{Q}}$. Inverting the resulting confidence distribution

$$\gamma^{\mathrm{CPF}}_{\hat{\pi}_{\tilde{V}|\theta}(v)|q} = \sum_{k=0}^{q} \binom{m}{q} \left(\pi_{\tilde{V}|\theta}(v)\right)^{k} \left(1 - \pi_{\tilde{V}|\theta}(v)\right)^{m-k} \tag{5.11}$$

for an observed number of successes $\tilde{Q}(v,\theta) = q$ via the inverse of the CPF of the beta distribution at a fixed confidence level $\gamma \in [0,1]$—which is then called the *reliability level*—gives the upper bound

$$I^{\mathrm{rel.}}_{m,\gamma}(v,\theta) = \begin{cases} \mathrm{betainv}(\gamma, q+1, m-q) & \text{if } q < m \text{ and} \\ 1 & \text{otherwise,} \end{cases} \tag{5.12}$$

of the confidence interval corresponding to the confidence level γ. Therefore, the value $I^{\mathrm{rel.}}_{m,\gamma}(v,\theta)$ can be used as a reliable approximation of $\pi_{\tilde{V}|\theta}(v)$.

The Π-IM-based derivation of $I^{\mathrm{rel.}}_{m,\gamma}(v,\theta)$ guarantees that

$$\mathrm{P}_{\tilde{V}_1,\ldots,\tilde{V}_m|\theta}\left(\pi_{\tilde{V}|\theta}(v) \leq I^{\mathrm{rel.}}_{m,\gamma}(v,\theta)\right) \geq \gamma. \tag{5.13}$$

True to the (Π-)IM framework, this can be interpreted as $I^{\mathrm{rel.}}_{m,\gamma}(v,\theta)$ having a guaranteed a-priori probability γ of not exceeding $\pi_{\tilde{V}|\theta}(v)$ with respect to the statistical model of $\tilde{V}_1,\ldots,\tilde{V}_m$. However, this reliability is to be understood with respect to the Monte-Carlo sampling procedure; in particular, it is not a general probability of overall lower specificity because it only holds pointwise. Still, future investigations may provide additional insight into this matter because numerical experiments suggest favorable properties concerning this global specificity attribute.

The great advantage of the Reliable μ-transform is that, even for very small sample sizes (as low as $m = 1$), the reliability guarantee in Eq. (5.13) always holds, whereas the Approximate μ-transform requires a potentially large number of samples $m \gg 1$ in order to yield a reliable estimator of $\pi_{\tilde{V}|\theta}(v)$. This is especially favorable if the sampling model of $\tilde{V}$ is non-trivial and every realization $\tilde{V}_j$ requires a significant computational effort by itself.

The following example highlights the price for this guaranteed reliability, the lower specificity of the Reliable μ-transform compared to the Approximate μ-transform.

Example 12: Implementation of Gaussian Π-IM

Consider the m iid realizations $\tilde{Q}_1, \ldots, \tilde{Q}_m$ of an imprecise variable $\tilde{Q} \sim \mathcal{N}(\theta_1, \theta_2)$ following a Gaussian distribution with unknown mean $\theta_1 \in \mathbb{R}$ and variance $\theta_2 > 0$. In the following, a very basic Π-IM for $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ is constructed. In particular, this Π-IM does not rely on existing results from statistics about suitable pivotal quantities in such inferential problems [CramerKamps20, pp. 287–289]. Instead, consider the cost function

$$J(q_1, \ldots, q_m, \theta_1, \theta_2) = (\theta_1 - \bar{m}(q))^2 + \left(\theta_2 - \bar{s}^2(q)\right)^2$$

for $q_1, \ldots, q_m \in \mathbb{R}$, $\theta_1 \in \mathbb{R}$ and $\theta_2 > 0$, where $\bar{m}(q) = \frac{1}{m}\sum_{i=1}^m q_i$ is the sample mean and $\bar{s}^2(q) = \frac{1}{m}\sum_{i=1}^m (\bar{m}(q) - q_i)^2$ is the (biased) sample variance. One may rescale this cost function into a normalized plausibility function, e.g. via

$$\rho_{\tilde{Q}_1,\ldots,\tilde{Q}_m|\theta_1,\theta_2}(q_1, \ldots, q_m) = \frac{1}{1 + J(q_1, \ldots, q_m, \theta_1, \theta_2)},$$

which is consistent with intuition: The cost function J is close to zero if the sample mean and the sample variance are close to their theoretical counterparts θ_1 and θ_2, respectively, yielding a plausibility value close to one. Otherwise, a higher cost value leads to plausibility values closer to zero. In conclusion, the plausibility resembles the agreement between the observations $q_1, \ldots, q_m$ and the parameters θ_1 and θ_2 without any need, e.g., for computing relative or absolute likelihoods.

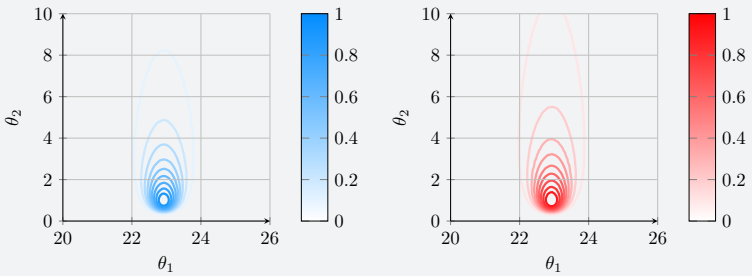
Applying the μ-transform to the corresponding Π-IM $\gamma_{\hat{\theta}_1,\hat{\theta}_2|\tilde{Q}_1,\ldots,\tilde{Q}_m}$ for $m = 10^3$ under this plausibility function produces the joint confidence distributions

$$\gamma_{\hat{\theta}_1,\hat{\theta}_2|q_1,\ldots,q_m}(\theta_1, \theta_2) = I_m(q_1, \ldots, q_m, \theta_1, \theta_2),$$

which are shown below for the observations[45]

$$\begin{array}{llllllll} q_1 & = 24.544, & q_2 & = 22.516, & q_3 & = 21.081, & q_4 & = 23.483, \\ q_5 & = 22.342, & q_6 & = 23.596, & q_7 & = 22.449, & q_8 & = 23.378. \end{array}$$

Evidently, the Reliable μ-transform produces a confidence distribution with much 'bulkier' superlevel sets, i.e., it is less specific, than the Approximate μ-transform.



Approximate μ-transform Reliable μ-transform ($\gamma = 0.99$)

In conclusion, whenever the μ-transform is invoked below, $\pi_{\tilde{V}|\theta}(v)$ may be replaced by the Approximate μ-transform $I_m^{\mathrm{MC}}(v,\theta)$ or by the *Reliable μ-transform* $I_{m,\gamma}^{\mathrm{rel.}}(v,\theta)$. Either of the two μ-transform implementations may also be referred to as $I_m(v,\theta)$. Since the reliability guarantee is very valuable and allows one to use arbitrarily low sample numbers m, the incurred loss of specificity is deemed to be more than acceptable, such that, in the following, the reliable implementation of the μ-transform is usually applied (typically with a reliability level of $\gamma = 0.99$)—unless indicated otherwise.

5.1.3 Cost Functions as Plausibilities

This section explores the role of cost functions in possibility theory following the idea of Example 12, more specifically the role of the cost function played therein—penalizing a mismatch between the sample moments and the theoretical moments implied by the parameters—and how a plausibility function is defined from this.

In many STEM (science, technology, engineering, mathematics) fields, a (positive) *cost function* $J : \mathbb{Q} \times \mathbb{Y} \to [0, +\infty)$ is usually intended to describe the fit between an observation $q \in \mathbb{Q}$ and the output of an *explanatory model* $y : \mathbb{T} \to \mathbb{Y}$ that is intended to make predictions $y(\theta)$ about $\tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta}$ for $\theta \in \mathbb{T}$, where $\mathbb{Y} \subseteq \mathbb{Q}$. A simple explanatory model is, e.g., the expectation $y(\theta) = \mathfrak{E}_{\mathrm{P}_{\tilde{Q}|\theta}}[\tilde{Q}]$. Often, this fit

$$J(q, y) = ||q - y|| \tag{5.14}$$

[45]These data points correspond to problem #7 of Ferson's exemplar problems for risk assessments from sparse information available under `https://sites.google.com/site/epistemicunc/exemplar-problems` (accessed November 8, 2021).

is defined with respect to some norm $||\cdot||$ of the difference between the observation and the model output (on $\mathbb{Q}$); popular examples include the quadratic cost functions employed, e.g., in regression and/or the training of neural networks [HastieTibshiraniFriedman01]. Usually, the minimizer

$$\theta^{\min} = \arg\min_{\theta\in\mathbb{T}} J(q, y(\theta)) \tag{5.15}$$

of such functions is sought, which is then used as an estimator of the ground truth θ^*.

If $J(q, y(\theta))$ is close to zero, θ is said to be in agreement with q. Conversely, very large values of $J(q, y(\theta))$ imply incompatibility of q and θ. Put differently, $J(q, y(\theta))$ is a graded fitness value lending plausibility to the combination (q, θ). Continuing the idea of plausibilities being derived from a cost function, one—but not the only—convenient way of defining a plausibility distribution $\rho_{\tilde{Q}|\theta} : \mathbb{Q} \to [0, 1]$ from a cost function is proposed in Example 12, namely by letting

$$\rho_{\tilde{Q}|\theta}(q) = \frac{1}{1 + J(q, y(\theta))} \tag{5.16}$$

for all $q \in \mathbb{Q}$ and $\theta \in \mathbb{T}$, such that the corresponding μ-transform reads

$$\begin{aligned} \pi_{\tilde{Q}|\theta}(q) &= \mathrm{P}_{\tilde{Q}|\theta}\left(\left\{\xi \in \mathbb{Q} : \rho_{\tilde{Q}|\theta}(\xi) \leq \rho_{\tilde{Q}|\theta}(q)\right\}\right) \\ &= \mathrm{P}_{\tilde{Q}|\theta}\left(\{\xi \in \mathbb{Q} : J\left(q, y(\theta)\right) \leq J\left(\xi, y(\theta)\right)\}\right). \end{aligned} \tag{5.17}$$

By this definition, the elementary confidence $\gamma_{\tilde{\theta}|q}(\theta) = \pi_{\tilde{Q}|\theta}(q)$ of $\theta \in \mathbb{T}$ is the a-priori probability of obtaining a worse cost/fit under $\tilde{Q} \sim \mathrm{P}_{\tilde{Q}|\theta}$ having observed $\tilde{Q} = q$.

5.2 Membership Arithmetic

The most widely applicable operation of possibilistic calculus is the extension principle including the extension of descriptive distributions to descriptive distributions in Eq. (3.27), the extension of descriptive distributions to confidence distributions in Eq. (4.38), the extension of confidence distributions to feature distributions in Eq. (4.40), and the extension of prediction distributions to prediction distributions in Eq. (4.56).

Every elementary possibility function being a *membership function*,[46] the following discussion will use the latter term to denote (parameter-dependent) elementary plausibility functions, (parameter-dependent) elementary IP-description functions, elementary confidence functions or elementary prediction functions alike. This general term is warranted because the formulations of all of the above operations are mathematically equivalent to the Extension Principle discussed in Section 3.2 and independent of the type of membership functions that are involved in the formulation.

[46] This name is owed to the close connection to this very concept in fuzzy set theory, see Section 2.1.2.3.

5.2.1 Membership Extensions

By a slight abuse of notation, one may, then, consider general extensions of membership functions $\mu_{\mathring{T}} : \mathbb{T} \to [0,1]$ of $\mathbb{T}$-valued fuzzy variables $\mathring{T}$, where $\mathbb{T} \subseteq \mathbb{R}^{D_\mathbb{T}}$. These fuzzy variables $\mathring{T}$ may either be an imprecise variable $\tilde{T} : \Omega \to \mathbb{T}$ or a parameter/feature function $\hat{T} : \mathbb{P}(\Omega, \mathbb{B}(\Omega)) \to \mathbb{T}$, and one can simply write $\mathring{T} \sim \mu_{\mathring{T}}$ in order to indicate that its membership function is given by $\mu_{\mathring{T}} : \mathbb{T} \to [0,1]$. These membership functions will typically be available in a closed-form expression or via the μ-transform. A general formulation of the Extension Principle, then, reads as follows.

Theorem 49. *Let $\mathring{T} \sim \mu_{\mathring{T}}$ be a $\mathbb{T}$-valued fuzzy variable, and let $\mathring{Z}$ be a second $\mathbb{Z}$-valued fuzzy variable that is connected to $\mathring{T}$ via the implicit relationship*

$$0 = \Xi(\mathring{T}, \mathring{Z}), \tag{5.18}$$

where $\Xi : \mathbb{T} \times \mathbb{Z} \to \mathbb{R}^D$ and 0 may be a vector of zeros. Then, the corresponding membership function $\mu_{\mathring{Z}} : \mathbb{Z} \to [0,1]$ of $\mathring{Z}$ is given by

$$\mu_{\mathring{Z}}(z) = \sup_{t \in \mathbb{T} \,:\, 0 = \Xi(t,z)} \mu_{\mathring{T}}(t) \tag{5.19}$$

for $z \in \mathbb{Z}$.

Theorem 49 will, henceforth, simply be referred to as the *(Implicit) Membership Extension Principle*, and the proposed numerical strategies for its solution, which are discussed next, can be applied to all the special cases mentioned above. Furthermore, it forms the basis for a *membership arithmetic* applying to descriptive, confidence and prediction distributions alike.

According to the formulation of the Implicit Membership Extension Principle, a potentially non-convex and arbitrarily complex optimization problem needs to be solved in order to compute the membership of every argument $z \in \mathbb{Z}$. Except for simple cases, where this may be accomplished analytically, or well-behaved extension problems [HoseHanss21a], this may quickly become an infeasible problem in a numerical implementation, and alternative solution techniques must be found.

As a first remark, reformulating an implicit extension as an explicit or inverse membership extension may dramatically improve the computability of Eq. (5.19). These formulations are discussed next.

5.2.1.1 Explicit Membership Extension

The most common formulation of membership arithmetic, the *explicit membership extension* is the basic principle for the propagation and marginalization of fuzzy variables, i.e.

when $0 = \Xi(\mathring{T}, \mathring{Z})$ describes an explicit dependency $\mathring{Z} = \phi(\mathring{T})$ for the function $\phi : \mathbb{T} \to \mathbb{Z}$ given $\mathring{T} \sim \mu_{\mathring{T}}$. Under this explicit dependency, Eq. (5.19) reduces to

$$\mu_{\mathring{Z}}(z) = \sup_{t \in \mathbb{T} \,:\, z = \phi(t)} \mu_{\mathring{T}}(t) \tag{5.20}$$

for $z \in \mathbb{Z}$, which is mathematically equivalent to the explicit extension in Eq. (3.30), the marginalization of descriptive distributions in Eq. (3.36), the marginalization of nuisance parameters in Eq. (4.42) and the construction of predictive distributions from confidence and descriptive distributions in Eq. (4.55).

The explicit membership extension is, furthermore, similar to (forward) fuzzy arithmetic, i.e. to the fuzzy Extension Principle [Zadeh75a, Zadeh75b, Zadeh75c], such that only minor modifications to available fuzzy arithmetical methods [Hanss02, Hanss05, Walz16, MäckHanss21] are needed in order to implement explicit possibilistic membership arithmetic. More precisely, classical fuzzy arithmetic is typically only concerned with the propagation of fuzzy variables under the non-interactive Π-copula, which ought to be exchanged for more suitable alternatives in a possibilistic context.

Alternatively, the explicit membership extension can also be formulated on the basis of the superlevel sets of membership functions, the *α-cuts*[47] of a fuzzy variable, similar to fuzzy arithmetic [Hanss05].

Proposition 50. *The α-cuts of $\mathring{Z} = \phi(\mathring{T})$ are the images of the corresponding α-cuts of $\mathring{T}$ under ϕ, i.e., $\mathcal{C}^{\alpha}_{\mu_{\mathring{Z}}} = \phi(\mathcal{C}^{\alpha}_{\mu_{\mathring{T}}})$ for all $\alpha \in [0,1]$.*

Proof. Let $\alpha \in [0,1]$ and let $z \in \phi(\mathcal{C}^{\alpha}_{\mu_{\mathring{T}}})$, i.e., $\phi^{-1}(\{z\}) \cap \mathcal{C}^{\alpha}_{\mu_{\mathring{T}}} \neq \emptyset$. Then, one finds that $\mu_{\mathring{Z}}(z) = \sup_{t \in \mathbb{T} \,:\, z = \phi(t)} \mu_{\mathring{T}}(t) = \sup_{t \in \phi^{-1}(\{z\})} \mu_{\mathring{T}}(t) > \alpha$, and therefore $z \in \mathcal{C}^{\alpha}_{\mu_{\mathring{Z}}}$.
Conversely, let $\alpha \in [0,1]$ and let $z \in \mathbb{Z}$ such that $z \notin \phi(\mathcal{C}^{\alpha}_{\mu_{\mathring{T}}})$. Then, $\phi^{-1}(\{z\}) \cap \mathcal{C}^{\alpha}_{\mu_{\mathring{T}}} = \emptyset$ and $\mu_{\mathring{Z}}(z) = \sup_{t \in \mathbb{T} \,:\, z = \phi(t)} \mu_{\mathring{T}}(t) = \sup_{t \in \phi^{-1}(\{z\})} \mu_{\mathring{T}}(t) \leq \alpha$. That is, $z \notin \mathcal{C}^{\alpha}_{\mu_{\mathring{Z}}}$.
In conclusion, $z \in \mathcal{C}^{\alpha}_{\mu_{\mathring{Z}}}$ if and only if $z \in \phi(\mathcal{C}^{\alpha}_{\mu_{\mathring{T}}})$. □

This formulation of fuzzy arithmetic is often used as a basis for interval arithmetical solutions of the extension principle [Hanss05].

5.2.1.2 Inverse Membership Extension

If the relationship $0 = \Xi(\mathring{T}, \mathring{Z})$ is equivalent to an inverse dependency $\mathring{T} = \psi(\mathring{Z})$, also written as $\mathring{Z} = \psi^{-1}(\mathring{T})$, for some function $\psi : \mathbb{Z} \to \mathbb{T}$ given the fuzzy variable $\mathring{T} \sim \mu_{\mathring{T}}$, the formulation of the *inverse membership extension* reads

$$\mu_{\mathring{Z}}(z) = \mu_{\mathring{T}}(\psi(z)) \tag{5.21}$$

[47] See footnote 46.

for $z \in \mathbb{Z}$, which is, e.g., equivalent to the inverse extension in Eq. (3.39). This inversion is another common operation on membership functions—and perhaps also the simplest because it amounts to a mere concatenation $\mu_{\hat{Z}} = \mu_{\hat{T}} \circ \psi$. Therefore, it does not necessarily require a particular implementation.

In the following, implicit and explicit extension implementations are discussed in the famework of membership Graphs, and algorithms are provided where necessary. More precisely, two fundamentally different paradigms, each with its benefits and drawbacks, will be investigated: interval- and sample-based implementations.

In order to do so, it is prudent to first discuss a unified strategy of representing membership functions numerically.

5.2.2 Membership Graphs

In principle, every membership function $\mu_{\hat{T}} : \mathbb{T} \to [0,1]$ of a fuzzy variable $\hat{T}$ can be represented via its graph $\{(t, \mu_{\hat{T}}(t)) : t \in \mathbb{T}\}$, which is a precise representation thereof but will usually possess an infinite number of elements. Such graphs can generally not be represented exactly on a computer.

A *μ-cover* $K = (S, \alpha)$ is a tuple consisting of a *μ-set* $S \subseteq \mathbb{T}$ and the corresponding *μ-level* $\alpha \in [0,1]$. In the following, only interval μ-sets will be considered, as well as their D-dimensional generalizations, box μ-sets, which will be discussed in Section 5.2.3.1. This μ-set shape can be represented very efficiently on a computer, and it includes singleton μ-sets, i.e. degenerate interval μ-sets composed of a single point, such as $\{0\} = [0,0]$.

A finite set of μ-covers $\mathcal{G} = \{K_1, \ldots, K_m\}$ is said to be a *μ-graph* on $\mathbb{T}$.

5.2.2.1 Membership Functions

The concept of μ-graphs naturally lends itself to the description of membership functions.

Keeping the convention that the maximum of the empty set is zero, every μ-graph $\mathcal{G}$ on $\mathbb{T}$ induces a membership function $\mu_{\mathcal{G}} : \mathbb{T} \to [0,1]$, the so-called *$\mu$-function* of $\mathcal{G}$, given by the maximum μ-level of all the μ-covers covering a certain point, i.e.

$$\mu_{\mathcal{G}}(t) = \max_{(S,\alpha) \in \mathcal{G} \,:\, t \in S} \alpha \tag{5.22}$$

for all $t \in \mathbb{T}$. This μ-function is locally constant everywhere (except at the boundaries of the corresponding μ-sets), cf. Fig. 5.1, and it is justified to think of the μ-level of a μ-cover as a lower bound on the μ-function on the corresponding μ-set.

A crucial aspect of the proposed representation of membership functions via μ-graphs is that the corresponding possibility and necessity measures, level sets, cumulative distribution functions, and expectations are very efficiently evaluated.

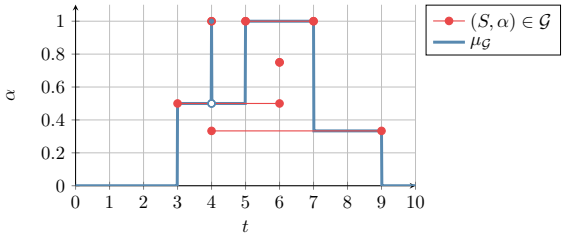


Figure 5.1: μ-covers of the μ-graph $\mathcal{G} = \{([3,6], \frac{1}{2}), (\{4\}, 1), ([4,9], \frac{1}{3}), ([5,7], 1), (\{6\}, \frac{3}{4})\}$ and its induced μ-function $\mu_{\mathcal{G}}$.

5.2.2.2 Possibilities and Necessities

Computing the possibility $\Pi_{\mathcal{G}}(B)$ of some event $B \in \mathbb{B}(\mathbb{T})$ from a μ-graph $\mathcal{G}$ on $\mathbb{T}$ is elementary. From the definitions of a possibility measure and a μ-function, the possibility

$$\Pi_{\mathcal{G}}(B) = \sup_{t \in B} \mu_{\mathcal{G}}(t) = \sup_{t \in B} \max_{(S,\alpha) \in \mathcal{G} \,:\, t \in S} \alpha = \max_{(S,\alpha) \in \mathcal{G} \,:\, B \cap S \neq \emptyset} \alpha \tag{5.23}$$

is simply the maximum μ-level of all μ-sets that have a non-empty intersection with B.

From this observation, one may, furthermore, deduce that the necessity

$$\mathrm{N}_{\mathcal{G}}(B) = 1 - \Pi_{\mathcal{G}}(\neg B) = 1 - \max_{(S,\alpha) \in \mathcal{G} \,:\, \neg B \cap S \neq \emptyset} \alpha = \min_{(S,\alpha) \in \mathcal{G} \,:\, S \not\subseteq B} 1 - \alpha \tag{5.24}$$

of $B \in \mathbb{B}(\mathbb{T})$ is derived from only the μ-covers in $\mathcal{G}$ whose μ-sets are not a subset of B.

5.2.2.3 Level Sets

The superlevel sets of the μ-function $\mu_{\mathcal{G}}$ induced by a μ-graph $\mathcal{G}$ on $\mathbb{T}$ are

$$\mathcal{C}^{\alpha}_{\mu_{\mathcal{G}}} = \{t \in \mathbb{T} : \mu_{\mathcal{G}}(t) > \alpha\} = \bigcup_{(S,\alpha') \in \mathcal{G} \,:\, \alpha' > \alpha} S \tag{5.25}$$

for $\alpha \in [0,1]$, i.e., the superlevel sets consist of all those μ-sets from the original μ-graph whose associated μ-levels are strictly greater than α. By complementation of the superlevel set, one obtains

$$\mathcal{S}^{\alpha}_{\mu_{\mathcal{G}}} = \neg \mathcal{C}^{\alpha}_{\mu_{\mathcal{G}}} = \bigcap_{(S,\alpha') \in \mathcal{G} : \alpha' > \alpha} \neg S. \tag{5.26}$$

for the sublevel sets. However, in contrast to the superlevel sets, this expression may be difficult to describe numerically when the μ-sets are intervals/boxes or singletons since, e.g., the intersection of complements of intervals can assume quite arbitrary shapes.

5.2.2.4 Cumulative Distribution Functions

Consider the one-dimensional[48] μ-graph $\mathcal{G}$ on $\mathbb{T} \subseteq \mathbb{R}$.

The corresponding CΠF $\mathcal{F}^+ : \mathbb{R} \to [0,1]$ of $\mu_{\mathcal{G}}$, the so-called μ-CΠF, is given by

$$\mathcal{F}^+(t) = \Pi_{\mathcal{G}}(\{\tau \in \mathbb{R} : \tau \leq t\}) = \max_{(S,\alpha)\in\mathcal{G}\,:\,\{\tau\in\mathbb{R}:\tau\leq t\}\cap S\neq\emptyset} \alpha = \max_{(S,\alpha)\in\mathcal{G}\,:\,\inf S\leq t} \alpha \tag{5.27}$$

for $t \in \mathbb{T}$, and the corresponding CNF $\mathcal{F}^- : \mathbb{R} \to [0,1]$, the so-called μ-CNF, is given by

$$\mathcal{F}^-(t) = \mathrm{N}_{\mathcal{G}}(\{\tau \in \mathbb{R} : \tau \leq t\}) = \min_{(S,\alpha)\in\mathcal{G}\,:\,S\not\subseteq\{\tau\in\mathbb{R}:\tau\leq t\}} 1-\alpha = \min_{(S,\alpha)\in\mathcal{G}\,:\,\sup S>t} 1-\alpha. \tag{5.28}$$

Notice that this also constitutes an implementation of the Possibility-to-P-Box Transform.

Similarly, the complementary μ-CΠF $\bar{\mathcal{F}}^+ : \mathbb{R} \to [0,1]$ and μ-CNF $\bar{\mathcal{F}}^- : \mathbb{R} \to [0,1]$ of $\mathcal{G}$ are given by

$$\bar{\mathcal{F}}^+(t) = \Pi_{\mathcal{G}}(\{\tau \in \mathbb{R} : \tau \geq t\}) = \max_{(S,\alpha)\in\mathcal{G}\,:\,\sup S\geq t} \alpha \tag{5.29}$$

and

$$\bar{\mathcal{F}}^-(t) = \mathrm{N}_{\mathcal{G}}(\{\tau \in \mathbb{R} : \tau \geq t\}) = \min_{(S,\alpha)\in\mathcal{G}\,:\,\inf S<t} 1-\alpha \tag{5.30}$$

for $t \in \mathbb{T}$, respectively.

5.2.2.5 Expectations

Finally, it is also possible to derive the upper and lower expectations of a one-dimensional[49] imprecise variable $\tilde{T}$, whose possibility distribution is available in the form of a μ-function $\mu_{\mathcal{G}_{\tilde{T}}}$ induced by the μ-graph $\mathcal{G}_{\tilde{T}}$ on $\mathbb{T} \subseteq \mathbb{R}$.

The upper expectation implied by $\mathcal{G}_{\tilde{T}}$ can be computed from the corresponding Choquet integral in Eq. (3.63) and reads

$$\mathfrak{E}_{\Pi_{\mathcal{G}_{\tilde{T}}}}(\tilde{T}) = \int_0^1 \sup \mathcal{C}^{\alpha}_{\mu_{\mathcal{G}_{\tilde{T}}}} \mathrm{d}\alpha = \int_0^1 \sup_{(S,\alpha')\in\mathcal{G}_{\tilde{T}}:\alpha'>\alpha} \sup S \,\mathrm{d}\alpha. \tag{5.31}$$

For the lower expectation, one can apply the same formalism yielding

$$\mathfrak{E}_{\mathrm{N}_{\mathcal{G}_{\tilde{T}}}}(\tilde{T}) = \int_0^1 \inf \mathcal{C}^{\alpha}_{\mu_{\mathcal{G}_{\tilde{T}}}} \mathrm{d}\alpha = \int_0^1 \inf_{(S,\alpha')\in\mathcal{G}_{\tilde{T}}:\alpha'>\alpha} \inf S \,\mathrm{d}\alpha. \tag{5.32}$$

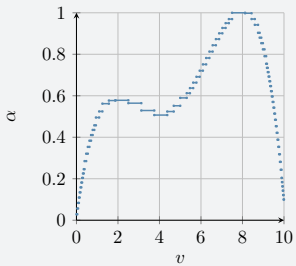
These values can, e.g., be approximated via a Monte-Carlo quadrature.

[48] This observation is only valid for μ-graphs on a one-dimensional space $\mathbb{T} \subseteq \mathbb{R}$. If, instead, a D-dimensional μ-graph is available, one must first perform the corresponding marginalization in order to obtain a one-dimensional μ-graph.

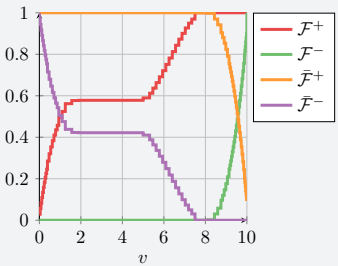
[49] See footnote 48.

Example 13: Membership Evaluations

Consider the μ-graph $\mathcal{G}_{\tilde{V}}$ of the $[0,10]$-valued imprecise variable $\tilde{V}$ shown below on the left.[50] The corresponding cumulative μ-CIIFs and μ-CNFs $\mathcal{F}^+$, $\mathcal{F}^-$, $\bar{\mathcal{F}}^+$ and $\bar{\mathcal{F}}^-$ are shown below on the right.



μ-Graph $\mathcal{G}_{\tilde{V}}$



Cumulative μ-Partitions

The expected-value bounds implied by this μ-graph are approximated from $m = 10^5$ Monte-Carlo samples yielding

$$\mathfrak{E}_{\mathrm{N}_{\mathcal{G}_{\tilde{V}}}}(\tilde{V}) \approx 2.8 \qquad \text{and} \qquad \mathfrak{E}_{\Pi_{\mathcal{G}_{\tilde{V}}}}(\tilde{V}) \approx 9.6.$$

5.2.2.6 Combination

Consider a number of membership functions in the form of μ-functions $\mu_{\mathcal{G}_1}, \ldots, \mu_{\mathcal{G}_m}$ induced by the μ-graphs $\mathcal{G}_1, \ldots, \mathcal{G}_m$ on $\mathbb{T}$. If these membership functions have been constructed from different sources of information—i.e. knowledge and/or data—, their disjunctive and conjunctive combinations must be considered to correctly account for all the available information in a possibilistic model.

Disjunction The disjunctive combination $\mathcal{G}^{\text{disj.}}_{1,\ldots,m} = \mathcal{G}_1 \vee \ldots \vee \mathcal{G}_m$ of m μ-graphs $\mathcal{G}_1, \ldots, \mathcal{G}_m$ on $\mathbb{T}$, more precisely of their μ-functions $\mu_{\mathcal{G}_1}, \ldots, \mu_{\mathcal{G}_m}$, is, according to Lemma 16, given by

$$\mu^{\text{disj.}}(t) = \max_{i=1,\ldots,m} \mu_{\mathcal{G}_i}(t) = \max_{i=1,\ldots,m} \max_{(S,\alpha)\in\mathcal{G}_i \,:\, t\in S} \alpha = \max_{(S,\alpha)\in\bigcup_{i=1}^m \mathcal{G}_i \,:\, t\in S} \alpha = \mu_{\mathcal{G}^{\text{disj.}}_{1,\ldots,m}}(t) \tag{5.33}$$

for all $t \in \mathbb{T}$, where $\mathcal{G}^{\text{disj.}}_{1,\ldots,m} = \bigcup_{i=1}^m \mathcal{G}_i$. That is, $\mu^{\text{disj.}} = \mu_{\mathcal{G}^{\text{disj.}}_{1,\ldots,m}}$ also has the form of a μ-function induced by $\mu_{\mathcal{G}^{\text{disj.}}_{1,\ldots,m}}$, the union of the individual μ-graphs, which is straightforward

[50] The μ-partition $\mathcal{G}_{\tilde{V}} \leftarrow \texttt{fsivia}\,([\mu], [0,10], 0.02)$ has been obtained under the FSIVIA Algorithm 1 to be discussed in Section 5.2.3.2 for the natural inclusion function of the membership function $\mu : [0,10] \to [0,1]$ given by $\mu(v) = \frac{1}{1024}\left(-3v^4 + 56v^3 - 336v^2 + 768v\right)$ for $v \in [0,10]$.

to implement—but also of little practical relevance.

The following implementation of the conjunction is arguably more important.

Conjunction The conjunctive combination $\mathcal{G}^{\text{conj.}}_{1,\dots,m} = \mathcal{G}_1 \wedge \ldots \wedge \mathcal{G}_m$ of m μ-graphs $\mathcal{G}_1, \dots, \mathcal{G}_m$ on $\mathbb{T}$, more precisely of their μ-functions $\mu_{\mathcal{G}_1}, \dots, \mu_{\mathcal{G}_m}$, includes the various credal set conjunctions in Lemmas 18, 20, 21 and 22, and can also be related to the construction of joint IP-description distributions in Propositions 35, 36 and 37, and the combination of confidence and prediction distributions in Propositions 46 and 48, which are all achieved by the application of either the UI-, SI-, or the NI-Π-copula given in Eqs. (3.78), (3.81) and (3.85). Since the conjunctions under the SI- or the UI-Π-copula are, essentially, rescaled variants of the conjunction under the NI-Π-copula, the latter shall primarily be discussed. In a general formulation, it can be expressed as

$$\begin{aligned}\mu^{\text{conj.}}(t) &= \mathcal{J}^{\text{NI}}\left(\mu_{\mathcal{G}_1}(t), \dots, \mu_{\mathcal{G}_m}(t)\right) = \min_{i=1,\dots,m} \mu_{\mathcal{G}_i}(t) \\ &= \min_{i=1,\dots,m} \max_{(S,\alpha)\in\mathcal{G}_i\,:\,t\in S} \alpha = \max_{(S,\alpha)\in\mathcal{G}^{\text{conj.}}_{1,\dots,m}\,:\,t\in S} \alpha = \mu_{\mathcal{G}^{\text{conj.}}_{1,\dots,m}}(t)\end{aligned} \tag{5.34}$$

for all $t \in \mathbb{T}$, where

$$\mathcal{G}^{\text{conj.}}_{1,\dots,m} = \left\{ \left(\bigcap_{i=1}^{m} S_i, \min_{i=1,\dots,m} \alpha_i \right) : (S_1, \alpha_1) \in \mathcal{G}_1, \dots, (S_m, \alpha_m) \in \mathcal{G}_m \right\}. \tag{5.35}$$

More generally, it holds that

$$\mathcal{G}^{\text{conj.}}_{1,\dots,m} = \left\{ \left(\bigcap_{i=1}^{m} S_i, \mathcal{J}(\alpha_1, \dots, \alpha_m) \right) : (S_1, \alpha_1) \in \mathcal{G}_1, \dots, (S_m, \alpha_m) \in \mathcal{G}_m \right\} \tag{5.36}$$

for $\mathcal{J} \in \{\mathcal{J}^{\text{NI}}, \mathcal{J}^{\text{UI}}, \mathcal{J}^{\text{SI}}\}$. Similar to the disjunction, the conjunction $\mu^{\text{conj.}} = \mu_{\mathcal{G}^{\text{conj.}}_{1,\dots,m}}$ also has the form of a μ-function induced by $\mathcal{G}^{\text{conj.}}_{1,\dots,m}$, which is slightly more complex but still straightforward to implement.

For the construction of joint IP-description distributions in Propositions 35, 36 and 37, one can assume that $\mathcal{G}_i$ are the vacuous extensions of the μ-graphs $\mathcal{G}'_i$ on $\mathbb{T}_i$ for $i = 1, \dots, m$, respectively. More precisely, the vacuous extensions onto $\mathbb{T} = \mathbb{T}_1 \times \ldots \times \mathbb{T}_m$ are induced by the μ-graphs

$$\mathcal{G}_i = \{([\mathbb{T}_1] \times \ldots \times [\mathbb{T}_{i-1}] \times S_i \times [\mathbb{T}_{i+1}] \times \ldots \times [\mathbb{T}_m], \alpha) : (S_i, \alpha) \in \mathcal{G}'_i\} \tag{5.37}$$

and Eq. (5.36) reduces to

$$\mathcal{G}^{\text{conj.}}_{1,\dots,m} = \{(S_1 \times \ldots \times S_m, \mathcal{J}(\alpha_1, \dots, \alpha_m)) : (S_1, \alpha_1) \in \mathcal{G}'_1, \dots, (S_m, \alpha_m) \in \mathcal{G}'_m\} \tag{5.38}$$

for $\mathcal{J} \in \{\mathcal{J}^{\text{NI}}, \mathcal{J}^{\text{UI}}, \mathcal{J}^{\text{SI}}\}$. That is, $\mathcal{G}^{\text{conj.}}_{1,\dots,m}$ resembles an extended Cartesian product of the μ-graphs $\mathcal{G}'_1, \dots, \mathcal{G}'_m$, i.e. the Cartesian product of their μ-sets along with an application of the Π-copula to the corresponding μ-levels.

A similar discussion can be applied to the various combinations of confidence distributions under Proposition 46.

In the following, the notation $\hat{T} \sim \mathcal{G}_{\hat{T}}$ indicates that the membership function $\mu_{\hat{T}}$ of $\hat{T}$ is the μ-function $\mu_{\mathcal{G}_{\hat{T}}}$.

Next, interval- and sample-based strategies for implementing membership extensions are discussed.

5.2.3 Interval-Based Extension Methods

The numerical strategies described in this section are fundamentally based on ideas from interval analysis. The justification for membership arithmetic via interval analysis instead of more efficient, sample-based implementations is the lack of computational schemes that can provide rigorous solutions, against which the latter (approximate solutions) can be validated. The interval-arithmetical methods proposed below can typically only be applied to smaller academic examples.

5.2.3.1 Interval Analysis

The following discussion commences with a brief introduction of interval analysis that is essential for understanding the remainder of this thesis. It is heavily based on the book by Jaulin et al. [JaulinEtAl01], which provides more detailed explanations of the various concepts employed in this chapter.

Intervals have already been used extensively in the previous chapters, but, for the sake of having a formal definition, an *interval* $[t]$ is a connected subset of $\mathbb{R}$. Here, it is also assumed to be bounded, and the lower and upper bounds of $[t]$ are given by

$$[t]^- = \inf [t] \qquad \text{and} \qquad [t]^+ = \sup [t]. \tag{5.39}$$

Therefore, one may write $[t] = [[t]^-, [t]^+]$, as done in the previous chapters. The interval's *center* and *radius* are given by

$$[t]^{\mathrm{c}} = \frac{[t]^+ + [t]^-}{2} \qquad \text{and} \qquad [t]^{\mathrm{r}} = \frac{[t]^+ - [t]^-}{2}, \tag{5.40}$$

and one may equivalently write $[t] = [t]^{\mathrm{c}} \pm [t]^{\mathrm{r}}$.

The following discussion applies both to intervals and to D-dimensional *boxes*, i.e. the Cartesian products

$$[\boldsymbol{t}] = [t_1] \times \ldots \times [t_D] \tag{5.41}$$

of D intervals, and will usually speak of the latter. Then, the definitions of lower and upper bound, center and radius are to be understood element-wise and yield vectors. The

(D-)volume of $[\boldsymbol{t}]$ is given by the Lebesgue measure

$$\lambda\left([\boldsymbol{t}]\right) = 2^D \cdot [t_1]^{\mathrm{r}} \cdot \ldots \cdot [t_D]^{\mathrm{r}}. \tag{5.42}$$

The space of all boxes in $\mathbb{T} \subseteq \mathbb{R}^D$ is written as $\mathbb{I}(\mathbb{T})$, and, given any set $B \subseteq \mathbb{T}$, its *interval hull*

$$[t] = \left[\inf_{\boldsymbol{t}\in B} t_1, \sup_{\boldsymbol{t}\in B} t_1\right] \times \ldots \times \left[\inf_{\boldsymbol{t}\in B} t_D, \sup_{\boldsymbol{t}\in B} t_D\right] \tag{5.43}$$

is the smallest box in $\mathbb{I}(\mathbb{T})$ that contains B. Notice that $\mathbb{I}(\mathbb{T}) \subseteq \mathbb{B}(\mathbb{T})$.

One of the simplest methods for determining if the two boxes $[\boldsymbol{t}] = [t_1] \times \ldots \times [t_D]$ and $[\boldsymbol{\tau}] = [\tau_1] \times \ldots \times [\tau_D]$ intersect, i.e. if $[\boldsymbol{t}] \cap [\boldsymbol{\tau}] \neq \emptyset$, is to check whether

$$[t_i]^+ \leq [\tau_i]^- \qquad \text{and} \qquad [\tau_i]^+ \leq [t_i]^- \tag{5.44}$$

for all $i = 1, \ldots, D$. Subset inclusion $[\boldsymbol{t}] \subseteq [\boldsymbol{\tau}]$ is verified via

$$[\tau_i]^- \leq [t_i]^- \qquad \text{and} \qquad [t_i]^+ \leq [\tau_i]^+ \tag{5.45}$$

for all $i = 1, \ldots, D$.

Finally, bisecting a box produces the lower and upper part

$$\begin{aligned} [\boldsymbol{t}]^{\mathrm{l}} &= [t_1] \times \ldots \times [t_{i^{\max}-1}] \times [[t_{i^{\max}}]^-, [t_{i^{\max}}]^{\mathrm{c}}] \times [t_{i^{\max}+1}] \times \ldots \times [t_D] \quad \text{and} \\ [\boldsymbol{t}]^{\mathrm{u}} &= [t_1] \times \ldots \times [t_{i^{\max}-1}] \times [[t_{i^{\max}}]^{\mathrm{c}}, [t_{i^{\max}}]^+] \times [t_{i^{\max}+1}] \times \ldots \times [t_D], \end{aligned} \tag{5.46}$$

respectively, where $i^{\max} = \arg\max_{i=1,\ldots,D} [t_i]^{\mathrm{r}}$ is the index of the interval constituent with maximum radius.

Interval Arithmetic A function $f : \mathbb{T} \to \mathbb{Z}$, evaluated with the box-valued argument $[\boldsymbol{t}]$ returns a general set $f([\boldsymbol{t}])$ that is not necessarily a box. The interval function $[f] : \mathbb{I}(\mathbb{Z}) \to \mathbb{I}(\mathbb{T})$ is an *inclusion function* of f if

$$f([\boldsymbol{t}]) \subseteq [f]([\boldsymbol{t}]) \tag{5.47}$$

for all $[\boldsymbol{t}] \in \mathbb{I}(\mathbb{T})$, i.e. if it is guaranteed to produce a box containing the true image of $[\boldsymbol{t}]$ under f. Inclusion functions are an important concept in interval analysis and may exhibit various properties, which shall, however, not be the topic of this discussion;[51] however, stated in the language of possibility theory, considering the indicator functions of $f([\boldsymbol{t}])$ and $[f]([\boldsymbol{t}])$, i.e. their $\{0,1\}$-valued membership functions, the goal is generally to find an inclusion function that produces a maximally specific indicator function $\mathcal{I}_{[f]([\boldsymbol{t}])}$ with $\mathcal{I}_{f([\boldsymbol{t}])} \preceq \mathcal{I}_{[f]([\boldsymbol{t}])}$. Clearly, an optimal inclusion function always achieves $[f]([\boldsymbol{t}]) = [f([\boldsymbol{t}])]$; however, it is not always possible to find such functions without

[51] Refer, e.g., to Jaulin et al. [JaulinEtAl01] and the many references therein.

undue computational effort. An efficient implementation of interval arithmetic is, e.g., described by Herrero et al. [HerreroEtAl12].

The inclusion functions employed in this thesis are primarily *natural inclusion functions* unless indicated otherwise. Natural inclusion functions are among the most basic techniques for constructing useful inclusion functions. They can be built from virtually all regular functions f composed of only elementary operations, such as addition, subtraction, multiplication, and division, and elementary functions (exp, sin, cos, tan, polynomials and more) by replacing these with the corresponding interval operations and functions. This approach to interval arithmetic is often called 'standard interval arithmetic' [Hanss05] and often leads to good convergence properties of the constructed inclusion function [JaulinEtAl01, Theorem 2.2]. Therefore, it is recommendable in many scenarios—the only impediment being the 'problem of repeated variables' [Hanss05] that may produce unnecessarily conservative inclusion functions. Moreover, it can often be implemented by operator-overloading, requiring only minimal changes to existing code based on computations with precise numbers.

Inclusion functions of the membership functions obtained under the μ-transform and of μ-functions are derived in Appendices A and B.

5.2.3.2 Membership Partitions

A μ-graph $\mathcal{P}$ composed of non-overlapping interval/box μ-sets covering all of $\mathbb{T}$ is said to be a *μ-partition* on $\mathbb{T}$. That is, the interiors of all contained μ-sets are disjoint (but they may share boundaries), and the union of all μ-sets covers $\mathbb{T} \subseteq \bigcup_{(S,\alpha)\in\mathcal{P}} S$.

A natural question is how one may obtain a μ-partition from a given membership function. To this end, suppose an inclusion function $[\mu] : \mathbb{I}(\mathbb{T}) \to \mathbb{I}([0,1])$ of a given membership function $\mu : \mathbb{T} \to [0,1]$ is available. Then, the Fuzzy Set Inversion Via Interval Analysis (FSIVIA) Algorithm 1 that is adapted from Set Inversion Via Interval Analysis (SIVIA) [JaulinEtAl01, Table 3.1] and was first presented by Hose and Hanss [HoseHanss20] can be used to compute an outer approximation in the form of a μ-partition.

The gradedness, i.e. the continuity, of the membership function[52] enables bisections according to the membership range and generalizes the bisections based on the simple indicator function in the classical SIVIA. The consideration of discrete distributions may require slight modifications to the proposed algorithms.

In its recursive variant, the FSIVIA Algorithm 1 computes the membership range of the box that is to be evaluated. If this range is small enough, the box and the maximum

[52]Here, all membership functions are assumed to be continuous, which is deemed more useful for practitioners.

Algorithm 1: `fsivia`

input : Membership Inclusion Function $[\mu]$, Box $[\boldsymbol{t}]$, Tolerance ε_α
output : μ-Partition $\mathcal{P}$

$[\alpha] \leftarrow [\mu]([\boldsymbol{t}])$ `// membership range in box`
if $[\alpha]^{\mathrm{r}} < \varepsilon_\alpha$ **then** `// termination criterion`
  $\mathcal{P} \leftarrow \{([\boldsymbol{t}], [\alpha]^{+})\}$ `// store` μ`-cover in` μ`-partition`
else
  $\mathcal{P} \leftarrow \texttt{fsivia}\left([\mu], [\boldsymbol{t}]^{\mathrm{l}}, \varepsilon_\alpha\right) \cup \texttt{fsivia}\left([\mu], [\boldsymbol{t}]^{\mathrm{u}}, \varepsilon_\alpha\right)$ `// FSIVIA on bisection`
end

membership in the range are added as a μ-cover to the μ-partition. Otherwise, the box is bisected, the FSIVIA algorithm is called on the two resulting (lower and upper) parts, and the results are joined. This procedure returns a μ-partition that is, above all else, composed of μ-covers where the original membership function is almost constant on the corresponding μ-set—and guaranteed to always be lower than the respective μ-level in order to enable a faithful, i.e. less specific, representation of the original membership function via the one induced by the resulting μ-partition.

Notice that `fsivia` only works with bounded input boxes.

In summary, FSIVIA is fundamental to many of the following implementations because it allows to compute a μ-partition $\mathcal{P}$ on $\mathbb{T}$ via the initial call

$$\mathcal{P} \leftarrow \texttt{fsivia}([\mu], [\mathbb{T}], \varepsilon_\alpha) \tag{5.48}$$

for a desired membership resolution $\varepsilon_\alpha \in [0, 1]$ from any membership function $\mu : \mathbb{T} \to [0, 1]$ of which an inclusion function $[\mu] : \mathbb{I}(\mathbb{T}) \to \mathbb{I}([0, 1])$ can be found. By construction of the algorithm, the μ-function of $\mathcal{P}$ is guaranteed to be less specific than μ because the upper bound of the membership range is taken to be the μ-level of the new μ-cover, which guarantees that the information contained in μ is robustly accounted for by the slightly less expressive outer approximation $\mu_{\mathcal{P}}$. For instance, if μ describes a family of imprecise probabilities, the credal set of $\mu \preceq \mu_{\mathcal{P}}$ is a subset of that of $\mu_{\mathcal{P}}$.

Especially if μ exhibits jumps and discontinuities, it is sensible to consider other termination criteria, such as a fixed number of maximum bisections or a minimum box volume, to guarantee a successful termination. It may also be advantageous to require a maximum box radius that must not be exceeded.

Example 14: FSIVIA of a Triangular Membership Function

Below, the μ-partitions resulting from a call to `fsivia` on a triangular membership function $\mu^{\text{tria.}} = \Delta(0, 1, 4)$,

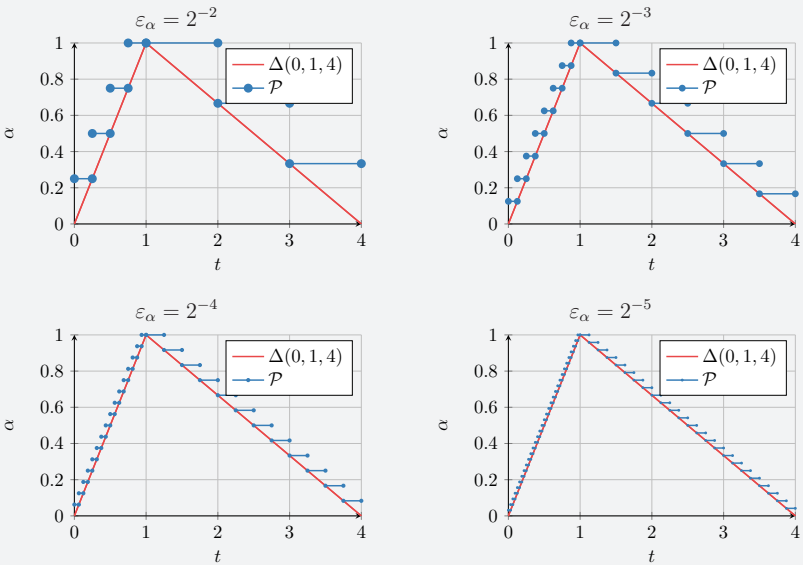
$$\mathcal{P} \leftarrow \texttt{fsivia}([\Delta(0, 1, 4)], [0, 4], \varepsilon_\alpha),$$

for various tolerance levels ε_α can be observed. Therein, the membership inclusion function is given by the lower bound

$$[\mu^{\text{tria.}}]^-([t]) = \min\left(\mu^{\text{tria.}}([t]^-), \mu^{\text{tria.}}([t]^+)\right)$$

and the upper bound

$$[\mu^{\text{tria.}}]^+([t]) = \begin{cases} 1 & \text{if } c \in [t] \\ \max\left(\mu^{\text{tria.}}([t]^-), \mu^{\text{tria.}}([t]^+)\right) & \text{otherwise.} \end{cases}$$



It is evident that ε_α can be used very efficiently to control the resolution of the resulting μ-partition and that the induced μ-function is always less specific than $\Delta(0,1,4)$.

Considering the inverse membership extension in Eq. (5.21), FSIVIA is directly applicable if an inclusion function of the concatenation $[\mu_{\tilde{T}}] \circ [\psi] : \mathbb{I}(\mathbb{Z}) \to \mathbb{I}([0,1])$ of the inclusion functions of $\mu_{\tilde{T}}$ and ψ is available.

Example 15: μ Transform via Membership Inversion

Consider two standard Gaussian variables $\tilde{Q}_1, \tilde{Q}_2 \sim \mathcal{N}(0,1)$. In order to find a joint possibility distribution, it is usually necessary to perform the μ-transform under some plausibility function as explained above. An arguably simpler technique can

be derived under the membership inversion.

The imprecise variable

$$\tilde{V} = \psi(\tilde{Q}_1, \tilde{Q}_2) = \tilde{Q}_1^2 + \tilde{Q}_2^2 \sim \chi^2(2)$$

follows a χ^2-distribution with two degrees of freedom. Such one-dimensional imprecise variables can be transformed very efficiently via the various (Complementary/Symmetric) Cumulative Probability-to-Possibility Transforms

$$\begin{aligned} \pi_{\tilde{V}}^{\mathrm{CPF}}(v) &= F_{\chi^2(2)}(v), \\ \pi_{\tilde{V}}^{\mathrm{CCPF}}(v) &= 1 - F_{\chi^2(2)}(v) \qquad \text{and} \\ \pi_{\tilde{V}}^{\mathrm{SCPF}}(v) &= 2 \cdot \min\left(F_{\chi^2(2)}(v), 1 - F_{\chi^2(2)}(v)\right), \end{aligned}$$

where $F_{\chi^2(2)}$ is the CPF of the χ^2-distribution with two degrees of freedom. Since CPFs are non-decreasing with respect to their argument, a tight inclusion function thereof can be constructed via

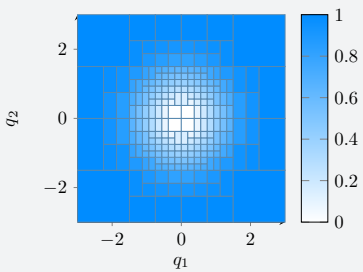
$$\left[F_{\chi^2(2)}\right]^-([v]) = F_{\chi^2(2)}([v]^-) \qquad \text{and} \qquad \left[F_{\chi^2(2)}\right]^+([v]) = F_{\chi^2(2)}([v]^+)$$

for $[v] \in \mathbb{I}([0, \infty))$ even though an explicit formula of this CPF is not available and, e.g., a natural inclusion function may not be constructed. From this inclusion function, (quasi-)natural inclusion functions of $\pi_{\tilde{V}}^{\mathrm{CPF}}$, $\pi_{\tilde{V}}^{\mathrm{CCPF}}$, $\pi_{\tilde{V}}^{\mathrm{SCPF}}$ and their concatenations with $[\psi]$ can be constructed.

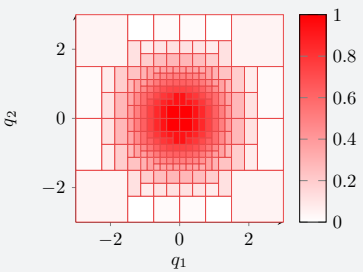
Under the membership inversion, the corresponding membership functions of $(\tilde{Q}_1, \tilde{Q}_2)$ are, e.g., induced by the μ-partitions obtained under

$$\begin{aligned} \mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}^{\mathrm{CPF}} &\leftarrow \texttt{fsivia}([\pi_{\tilde{V}}^{\mathrm{CPF}}] \circ [\psi], [-3,3] \times [-3,3], 0.1), \\ \mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}^{\mathrm{CCPF}} &\leftarrow \texttt{fsivia}([\pi_{\tilde{V}}^{\mathrm{CCPF}}] \circ [\psi], [-3,3] \times [-3,3], 0.1) \qquad \text{and} \\ \mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}^{\mathrm{SCPF}} &\leftarrow \texttt{fsivia}([\pi_{\tilde{V}}^{\mathrm{SCPF}}] \circ [\psi], [-3,3] \times [-3,3], 0.1), \end{aligned}$$

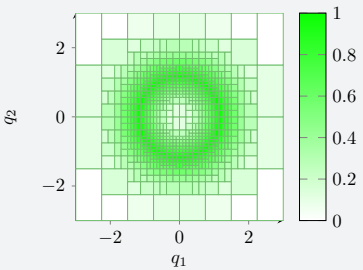
which are also shown below.



μ-Partition $\mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}^{\mathrm{CPF}}$



μ-Partition $\mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}^{\mathrm{CCPF}}$



μ-Partition $\mathcal{P}^{\mathrm{SCPF}}_{\tilde{Q}_1,\tilde{Q}_2}$

The rectangles correspond to the μ-sets, and the colors indicate the μ-levels of the μ-clusters pertaining to the various $\mathcal{P}_{\tilde{Q}_1,\tilde{Q}_2}$. Notice that $\mathcal{P}^{\mathrm{CCPF}}_{\tilde{Q}_1,\tilde{Q}_2}$ resembles the Optimal μ-transform of $\mathrm{P}_{\tilde{Q}_1,\tilde{Q}_2}$.

5.2.3.3 Interval-Based Implicit Extension

In order to implement the general implicit membership extension as given in Eq. (5.19), let $[\Xi] : \mathbb{I}(\mathbb{T}) \times \mathbb{I}(\mathbb{Z}) \to \mathbb{I}(\mathbb{R}^D)$ be an inclusion function of Ξ, and let $[\mu_{\hat{T}}] : \mathbb{I}(\mathbb{T}) \times \mathbb{I}([0,1])$ be an inclusion function of $\mu_{\hat{T}}$; let $[\boldsymbol{t}] \in \mathbb{I}(\mathbb{T})$ and $[\boldsymbol{z}] \in \mathbb{I}(\mathbb{Z})$, and define $[\boldsymbol{x}] = [\Xi]([\boldsymbol{t}], [\boldsymbol{z}])$ and $[\alpha] = [\mu_{\hat{T}}]([\boldsymbol{t}])$. Then, four cases may be distinguished.

1. If 0 is not an element of $[\boldsymbol{x}]$, then it is impossible that any combination $(\boldsymbol{t}, \boldsymbol{z}) \in [\boldsymbol{t}] \times [\boldsymbol{z}]$ achieves $\Xi(\boldsymbol{t}, \boldsymbol{z}) = 0$, and $[\boldsymbol{t}]$ and $[\boldsymbol{z}]$ may be discarded.

In the following, $0 \in [\boldsymbol{x}]$ is assumed.

2. If both $[\boldsymbol{x}]$ and $[\alpha]$ meet some tolerance constraints—e.g. if their radii are no wider than ε_0 and ε_α, respectively—then one may add the μ-cover $([\boldsymbol{z}], [\alpha]^+)$ to the μ-graph $\mathcal{P}_{\hat{Z}}$. The upper bound $[\alpha]^+$ guarantees that this μ-level is higher than the membership $\mu_{\hat{T}}(\boldsymbol{t})$ of any $\boldsymbol{t} \in [\boldsymbol{t}]$ for which some $\boldsymbol{z} \in [\boldsymbol{z}]$ exists such that $\Xi(\boldsymbol{t}, \boldsymbol{z}) = 0$.

3. If $[\boldsymbol{x}]$ does not meet some tolerance constraint—e.g. if its radius is wider than ε_0—then one may quadrisect $[\boldsymbol{t}] \times [\boldsymbol{z}]$ and inspect the combinations $([\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{l}})$, $([\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{u}})$, $([\boldsymbol{t}]^{\mathrm{u}}, [\boldsymbol{z}]^{\mathrm{l}})$ and $([\boldsymbol{t}]^{\mathrm{u}}, [\boldsymbol{z}]^{\mathrm{u}})$ individually.

4. If only $[\alpha]$ does not meet some tolerance constraint—e.g. if its radius is wider than ε_α—then one may bisect $[\boldsymbol{t}]$ and inspect the combinations $([\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}])$ and $([\boldsymbol{t}]^{\mathrm{u}}, [\boldsymbol{z}])$ individually. Bisecting $[\boldsymbol{z}]$ is not necessary because it does not influence $[\alpha]$.

The recursive Membership Extension via Interval Analysis (MEVIA) Algorithm 2 is a generalization of the FSIVIA Algorithm 1, following these deliberations.

Algorithm 2: `mevia`

input : Membership Inclusion Function $[\mu_{\mathring{T}}]$, Dependency Inclusion Function $[\Xi]$, Input Box $[\boldsymbol{t}]$, Output Box $[\boldsymbol{z}]$, Tolerance ε_α, Tolerance ε_0

output : μ-Graph $\mathcal{G}_{\mathring{Z}}$

$[\boldsymbol{x}] \leftarrow [\Xi]([\boldsymbol{t}], [\boldsymbol{z}])$ `// compute image range of boxes`
$[\alpha] \leftarrow [\mu_{\mathring{T}}]([\boldsymbol{t}])$ `// compute membership range in box`
if $0 \notin [\boldsymbol{x}]$ **then** `// case 1: discard`
  $\mathcal{G}_{\mathring{Z}} \leftarrow \emptyset$
else if $[\alpha]^{\mathrm{r}} < \varepsilon_\alpha$ **and** $[\boldsymbol{x}]^{\mathrm{r}} < \varepsilon_0$ **then** `// case 2: store`
  $\mathcal{G}_{\mathring{Z}} \leftarrow \{([\boldsymbol{z}], [\alpha]^+)\}$
else if $[\boldsymbol{x}]^{\mathrm{r}} \geq \varepsilon_0$ **then** `// case 3: quadrisect`
  $\mathcal{G}_1 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{l}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_2 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{u}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_3 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{u}}, [\boldsymbol{z}]^{\mathrm{l}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_4 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{u}}, [\boldsymbol{z}]^{\mathrm{u}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_{\mathring{Z}} \leftarrow \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3 \cup \mathcal{G}_4$
else `// case 4: bisect`
  $\mathcal{G}_1 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{l}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_2 \leftarrow \texttt{mevia}([\mu], [\Xi], [\boldsymbol{t}]^{\mathrm{l}}, [\boldsymbol{z}]^{\mathrm{u}}, \varepsilon_\alpha, \varepsilon_0)$
  $\mathcal{G}_{\mathring{Z}} \leftarrow \mathcal{G}_1 \cup \mathcal{G}_2$
end

Again, the call

$$\mathcal{G}_{\mathring{Z}} \leftarrow \texttt{mevia}\,([\mu], [\Xi], [\mathbb{T}], [\mathbb{Z}], \varepsilon_\alpha, \varepsilon_0) \tag{5.49}$$

for some tolerance constraints $\varepsilon_\alpha \in [0, 1]$ and $\varepsilon_0 > 0$ guarantees a robust outer approximation of $\mu_{\mathring{Z}} \preceq \mu_{\mathcal{G}_{\mathring{Z}}}$ via the μ-function induced by the μ-graph $\mathcal{G}_{\mathring{Z}}$. Of course, additional termination criteria may be inserted into this algorithm.

Example 16: Membership Extension

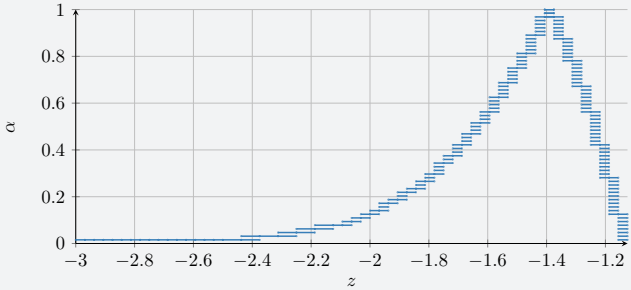
Consider the input fuzzy variable $\mathring{T} \sim \Delta(0, 1, 2)$ and the output fuzzy variable $\mathring{Z}$ that is connected to $\mathring{T}$ via the implicit relationship

$$0 = \Xi\left(\mathring{T}, \mathring{Z}\right) = \left(\mathring{Z} + 1\right)\left(\mathring{T} - 2\right) - \exp\left(\mathring{T} - \mathring{Z}^2\right).$$

Notice that this relationship is difficult to invert analytically with respect to either $\mathring{T}$ or $\mathring{Z}$. Still, an outer approximation of $\mu_{\mathring{Z}}$ can be computed in the form of the μ-

graph $\mathcal{P}_{\mathring{Z}}$ shown below by calling MEVIA via

$$\mathcal{P}_{\mathring{Z}} \leftarrow \texttt{mevia}\left([\Delta(0,1,2)]\,,[\Xi]\,,[0,2],[-3,-1],0.05,0.05\right).$$



μ-Graph $\mathcal{P}_{\mathring{Z}}$

Notice that, in contrast to the FSIVIA Algorithm 1, the MEVIA Algorithm 2 does not return a μ-partition but a μ-graph instead. However, a μ-partition can always be obtained by applying FSIVIA to an inclusion function of the μ-function $\mu_{\mathcal{G}_{\mathring{Z}}}$, see Appendix B.

Nevertheless, the MEVIA Algorithm 2 is computationally demanding, and, if one can rewrite an implicit extension as an explicit or inverse extension, this is generally recommendable. For instance the membership inversion under the FSIVIA Algorithm 1 avoids the additional bisections of $[\mathbb{T}]$ under MEVIA and/or first computing a μ-graph of $\mathring{T}$—even though, under Eq. (5.54), it implies

$$\mathcal{P}_{\mathring{T}} = \{([\psi]([\boldsymbol{z}]),\alpha) : ([\boldsymbol{z}],\alpha) \in \mathcal{P}_{\mathring{Z}}\}\,. \tag{5.50}$$

An efficient interval-based strategy for membership propagations is discussed next.

5.2.3.4 Interval-Based Membership Propagation

As mentioned above, the *α-cut arithmetic* enabled by Proposition 50 is often considered to be the single basis for fuzzy/membership arithmetic based on interval arithmetic [Hanss05]. According to this formulation, a fuzzy input variable $\mathring{T}$ must be decomposed into a finite number of α-cuts $(\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{T}}})_{i=1}^{m}$ for the corresponding membership levels $\alpha_{(i)} \in [0,1]$ for $i = 1,\ldots,m$ producing the m input tuples $(\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{T}}},\alpha_{(i)})_{i=1}^{m}$. Finally, the output α-cuts under ϕ can be bounded bounded via

$$\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{Z}}} = \phi\left(\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{T}}}\right) \subseteq [\phi]\left(\left[\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{T}}}\right]\right), \tag{5.51}$$

i.e., the output fuzzy variable $\mathring{Z}$ is numerically approximated via the tuples $([\phi]([\mathcal{C}^{\alpha_{(i)}}_{\mu_{\mathring{T}}}]),\alpha_{(i)})_{i=1}^{m}$. Classical fuzzy arithmetic based on interval arithmetic

only considers fuzzy numbers, i.e. marginal fuzzy variables whose α-cuts are intervals, which—under the minimum-based Π-copulae—lead to joint input α-cuts in the form of boxes [Hanss05], but the above propagation rule holds for any input α-cut shape.

The conservatism of this approximation critically depends on the quality of the inclusion function $[\phi]$—e.g. on its ability to avoid overestimation due to repeated variables—and on the initial box-shapedness of $\mathcal{C}_{\mu_{\mathring{T}}}^{\alpha_{(i)}}$. This guarantees the robust approximation of the output α-cuts on the given levels $\alpha_{(i)}$, but it is not trivial to guarantee robustness on the α-cuts in-between.

Membership arithmetic based on μ-graphs composed of inverval/box μ-sets is an alternative to α-cut arithmetic, and the membership propagation for fuzzy variables $\mathring{T} \sim \mathcal{G}_{\mathring{T}}$ shall now be discussed. Of course, the μ-graph $\mathcal{G}_{\mathring{T}}$ on $\mathbb{T}$ may have, e.g., resulted from FSIVIA or MEVIA, and, if $\mathring{T} = (\mathring{T}_1, \ldots, \mathring{T}_m)$ is composed of the marginal variables $\mathring{T}_1 \sim \mathcal{G}_{\mathring{T}_1}, \ldots, \mathring{T}_m \sim \mathcal{G}_{\mathring{T}_m}$, then one can obtain the joint μ-graph $\mathcal{G}_{\mathring{T}} = \mathcal{G}_{\mathring{T}_1,\ldots,\mathring{T}_m}$ under the Π-copula $\mathcal{J}$ via the conjunctive combination of their vacuous extensions as described in Section 5.2.2.6, i.e.

$$\begin{aligned}\mathcal{G}_{\mathring{T}} = \big\{ &([\boldsymbol{t}]_1 \times \ldots \times [\boldsymbol{t}]_m, \mathcal{J}(\alpha_1, \ldots, \alpha_m)) : \\ &([\boldsymbol{t}]_1, \alpha_1) \in \mathcal{G}_{\mathring{T}_1}, \ldots, ([\boldsymbol{t}]_m, \alpha_m) \in \mathcal{G}_{\mathring{T}_m} \big\}.\end{aligned} \tag{5.52}$$

Given an input fuzzy variable $\mathring{T} \sim \mathcal{G}_{\mathring{T}}$, the membership function of the output fuzzy variable $\mathring{Z} = \phi(\mathring{T})$ on some box $[\boldsymbol{z}] \in \mathbb{I}(\mathbb{Z})$ is, according to the explicit membership extension in Eq. (5.20), given by

$$\begin{aligned}\mu_{\mathring{Z}}(\boldsymbol{z}) &= \sup_{\boldsymbol{t}\in\mathbb{T}\,:\,\boldsymbol{z}=\phi(\boldsymbol{t})} \mu_{\mathcal{G}_{\mathring{T}}}(\boldsymbol{t}) = \sup_{\boldsymbol{t}\in\mathbb{T}\,:\,\boldsymbol{z}=\phi(\boldsymbol{t})} \max_{([\boldsymbol{t}],\alpha)\in\mathcal{G}_{\mathring{T}}\,:\,\boldsymbol{t}\in[\boldsymbol{t}]} \alpha \\ &= \max_{([\boldsymbol{t}],\alpha)\in\mathcal{G}_{\mathring{T}}\,:\,\boldsymbol{z}\in\phi([\boldsymbol{t}])} \alpha \leq \max_{([\boldsymbol{t}],\alpha)\in\mathcal{G}_{\mathring{T}}\,:\,\boldsymbol{z}\in[\phi]([\boldsymbol{t}])} \alpha = \max_{([\boldsymbol{z}],\alpha)\in\mathcal{G}_{\mathring{Z}}\,:\,\boldsymbol{z}\in[\boldsymbol{z}]} \alpha = \mu_{\mathcal{G}_{\mathring{Z}}}(\boldsymbol{z})\end{aligned} \tag{5.53}$$

for all $\boldsymbol{z} \in [\boldsymbol{z}]$. Therein, the inequality is potentially an equality, depending on the inclusion function of ϕ. In any case, the possible loss of specificity generates a robust outer approximation of the information contained in $\mu_{\mathring{Z}}$. More importantly, the expression on the right-hand side of the inequality can be understood as the μ-function induced by the μ-graph

$$\mathcal{G}_{\mathring{Z}} = \{([\phi]([\boldsymbol{t}]), \alpha) : ([\boldsymbol{t}], \alpha) \in \mathcal{G}_{\mathring{T}}\}. \tag{5.54}$$

That is, the μ-graph of $\mathring{Z}$ is computed by taking the μ-covers $([\boldsymbol{t}], \alpha) \in \mathcal{G}_{\mathring{T}}$, propagating the associated interval/box μ-sets $[\boldsymbol{t}]$ through the inclusion function of ϕ, and storing the resulting μ-sets $[\phi]([\boldsymbol{t}])$ along with the original μ-levels α in $\mathcal{G}_{\mathring{Z}}$.

Example 17: Membership Propagation

Consider the propagation of the fuzzy input variable $\mathring{T} \sim \mathcal{P}_{\mathring{T}}$ whose μ-partition, which is shown below on the left, has been obtained from the Chebychev distribution

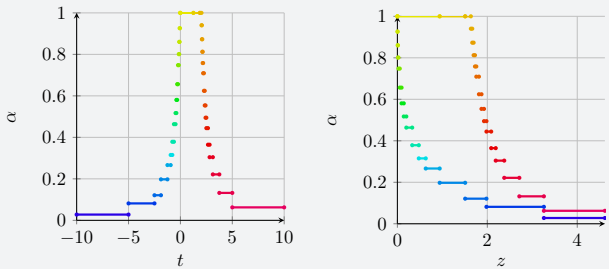
via

$$\mathcal{P}_{\mathring{T}} \leftarrow \texttt{fsivia}\left([\mathcal{M}^2(1,1)], [-10,10], 0.05\right).$$

The resulting membership function of the fuzzy output variable

$$\mathring{Z} = \phi(\mathring{T}) = \log(\mathring{T}^2 + 1)$$

can then be approximated by computing the μ-graph $\mathcal{G}_{\mathring{Z}}$ in Eq. (5.54) under ϕ, which is shown below on the right.



Input μ-Partition $\mathcal{P}_{\mathring{T}}$ Output μ-Partition $\mathcal{G}_{\mathring{Z}}$

In both figures, the corresponding μ-covers, $([t], \alpha) \in \mathcal{P}_{\mathring{T}}$ and $([\phi]([t]), \alpha) \in \mathcal{G}_{\mathring{Z}}$, have the same color in order to trace the propagation procedure. The μ-sets differ, whereas the μ-levels remain the same. Evidently, the μ-covers stemming from $t < 0$ do not contribute to the μ-function of $\mathcal{G}_{\mathring{Z}}$

Notice that, e.g. in the example above, even though the input may be a μ-partition it cannot be guaranteed that the output is also a μ-partition. It may very well only be a μ-graph.

This approach corresponds to classical α-cut arithmetic if the μ-graph $\mathcal{G}_{\mathring{T}} = \{(\mathcal{C}_{\mu_{\mathring{T}}}^{\alpha^{(i)}}, \alpha^{(i)})\}$ is composed of the α-cuts of $\mathring{T}$. That is, the proposed interval-based propagation method constitutes a generalization of classical α-cut arithmetic.

On a final note, the marginalization of a μ-graph reduces to a simple elimination of the respective dimensions $i = 1, \ldots, m$ from the associated μ-sets in Eq. (5.54), e.g. via

$$\mathcal{P}_{\mathring{T}_i} = \{([\boldsymbol{t}]_i, \alpha) : ([\boldsymbol{t}], \alpha) \in \mathcal{P}_{\mathring{T}}\}. \tag{5.55}$$

As mentioned above, interval-based membership extension methods are computationally demanding, restricting their application to smaller problems. Additionally, they require inclusion functions of the relationships Ξ, ϕ, and ψ, which usually implies these models to be 'white-box' models that can intrusively be modified in order to apply interval arithmetic. Their great benefit is the guaranteed outer, i.e. conservative, approximation of

the information about the model output $\mathring{Z}$.

Next, an alternative, sample-based extension method shall briefly be discussed, which addresses these drawbacks. In particular, it is non-intrusive, i.e., it can be applied to 'black-box' models; yet, the computed solutions are always inner, i.e. non-robust, approximations.

5.2.4 Sample-Based Extension Methods

Membership propagation based on samples loosely corresponds to Vanilla Monte-Carlo methods. It is guided by the idea that any membership function can be represented reasonably well by a sufficiently large number of samples. However,—in contrast to Monte-Carlo techniques—it is not necessary for the samples' densities to be representative of a population because, instead, they are equipped with a membership value—similar to statistical sample weights. The general idea is to choose these samples such that the original membership function may be reconstructed from the sample positions and their respective memberships. The samples should generally be densely distributed in areas where the approximated membership function has large gradients, discontinuities, etc., and more sparsely distributed in areas where it is effectively constant.

This sampling idea can be cast into the framework of μ-graphs.

5.2.4.1 Membership Clusters

A μ-graph $\mathcal{K}$ composed of μ-covers containing only singleton μ-sets is said to be a *μ-cluster*, and its μ-covers are usually referred to as *μ-nodes*. For convenience, and if no ambiguity is possible, such μ-nodes may simply be written as a tuple $(\boldsymbol{t}, \alpha) \in \mathcal{K}$ for the singleton $\boldsymbol{t} \in \mathbb{T}$ and the μ-level $\alpha \in [0, 1]$, which would—strictly speaking—have to be expressed as $(\{\boldsymbol{t}\}, \alpha) \in \mathcal{K}$.

Given a membership function $\mu_{\hat{T}} : \mathbb{T} \to [0, 1]$, an approximation via a μ-cluster can be achieved in a variety of ways, e.g. via

- regular sampling, e.g. on equidistant regular grids, or following specific procedures, such as the Transformation Method [Hanss02],
- unstructured/random (uniform, Latin hypercube, Sobol, etc.) sampling, or
- structured sampling, such as sparse-grid sampling [Walz16], or sequentially weighted sampling [MäckHanss21].

Of course, this list is not exhaustive, and other methods are certainly viable.

Generally speaking, all of these methods have the commonality that they produce a set of positions $(\boldsymbol{t}_{(i)})_{i=1}^{m}$, for which the corresponding membership is given by $\alpha_{(i)} = \mu_{\hat{T}}(\boldsymbol{t}_{(i)})$, and,

together, they form the μ-nodes $\mathcal{K}_{\mathring{T}} = \{(\boldsymbol{t}_{(1)}, \alpha_{(1)}), \ldots, (\boldsymbol{t}_{(m)}, \alpha_{(m)})\}$. The corresponding μ-function yields the specificity relation $\mu_{\mathcal{K}_{\mathring{T}}} \preceq \mu_{\mathring{T}}$, i.e.

$$\mu_{\mathcal{K}_{\mathring{T}}}(\boldsymbol{t}) = \left\{ \begin{array}{rl} \mu_{\mathring{T}}(\boldsymbol{t}) & \text{if } \boldsymbol{t} = \boldsymbol{t}_{(i)} \text{ for some } i = 1, \ldots, m \text{ and} \\ 0 & \text{otherwise} \end{array} \right\} \leq \mu_{\mathring{T}}(\boldsymbol{t}) \tag{5.56}$$

for all $\boldsymbol{t} \in \mathbb{T}$. The μ-cluster $\mathcal{K}_{\mathring{T}}$ always constitutes an inner approximation of $\mu_{\mathring{T}}$, and the underestimation depends on the resolution of its μ-nodes.

5.2.4.2 Sample-Based Membership Propagation

Given an input fuzzy variable $\mathring{T} \sim \mathcal{K}_{\mathring{T}}$ whose membership function is represented via a μ-cluster $\mathcal{K}_{\mathring{T}}$, the membership function of the output fuzzy variable $\mathring{Z} = \phi(\mathring{T})$ is, according to the explicit membership extension in Eq. (5.20), given by

$$\begin{aligned} \mu_{\mathring{Z}}(\boldsymbol{z}) = \sup_{\boldsymbol{t} \in \mathbb{T} \,:\, \boldsymbol{z} = \phi(\boldsymbol{t})} \mu_{\mathcal{K}_{\mathring{T}}}(\boldsymbol{t}) &= \sup_{\boldsymbol{t} \in \mathbb{T} \,:\, \boldsymbol{z} = \phi(\boldsymbol{t})} \max_{(\boldsymbol{t}', \alpha) \in \mathcal{K}_{\mathring{T}} \,:\, \boldsymbol{t} = \boldsymbol{t}'} \alpha \\ &= \max_{(\boldsymbol{t}', \alpha) \in \mathcal{K}_{\mathring{T}} \,:\, \boldsymbol{z} = \phi(\boldsymbol{t}')} \alpha = \max_{(\boldsymbol{z}', \alpha) \in \mathcal{K}_{\mathring{Z}} \,:\, \boldsymbol{z} = \boldsymbol{z}'} \alpha = \mu_{\mathcal{K}_{\mathring{Z}}}(\boldsymbol{z}) \end{aligned} \tag{5.57}$$

for all $\boldsymbol{z} \in \mathbb{Z}$. This membership function, too, can be described as the μ-function $\mu_{\mathcal{K}_{\mathring{Z}}}$ induced by the μ-cluster

$$\mathcal{K}_{\mathring{Z}} = \{(\phi(\boldsymbol{t}), \alpha) : (\boldsymbol{t}, \alpha) \in \mathcal{K}_{\mathring{T}}\}, \tag{5.58}$$

which corresponds to Eq. (5.54). The μ-node positions are simply propagated under ϕ, which—in this case—does not require any intrusive modifications of ϕ, and the memberships are carried along. Thus, the propagation rule can be applied to black-box models without much computational effort.

Similarly, the marginalization of $\mathcal{K}_{\mathring{T}}$ onto the i-th dimension is straightforward to compute via

$$\mathcal{K}_{\mathring{T}_i} = \{(t_i, \alpha) : (\boldsymbol{t}, \alpha) \in \mathcal{K}_{\mathring{T}}\}. \tag{5.59}$$

Of course, if $\mathcal{K}_{\mathring{T}}$ is a sampled approximation of $\mu_{\mathring{T}}$, then $\mathcal{K}_{\mathring{Z}}$ merely provides an inner approximation of the corresponding membership extension $\mu_{\mathring{Z}}$ because, then, by Eq. (5.56) it holds that

$$\mu_{\mathcal{K}_{\mathring{Z}}}(\boldsymbol{z}) = \sup_{\boldsymbol{t} \in \mathbb{T} \,:\, \boldsymbol{z} = \phi(\boldsymbol{t})} \mu_{\mathcal{K}_{\mathring{T}}}(\boldsymbol{t}) \leq \sup_{\boldsymbol{t} \in \mathbb{T} \,:\, \boldsymbol{z} = \phi(\boldsymbol{t})} \mu_{\mathring{T}}(\boldsymbol{t}) = \mu_{\mathring{Z}}(\boldsymbol{z}) \tag{5.60}$$

for all $\boldsymbol{z} \in \mathbb{Z}$. Whereas it is non-trivial to exactly reconstruct the output membership function $\mu_{\mathring{Z}}$ from $\mathcal{K}_{\mathring{Z}}$—Walz proposes a binning approach [Walz16]—, good approximations, e.g. of the output possibility measure $\Pi_{\mathring{Z}}$, can be found via Eq. (5.23) yielding

$$\Pi_{\mathring{Z}}(B) \approx \Pi_{\mathcal{K}_{\mathring{Z}}}(B) = \max_{(\boldsymbol{z}, \alpha) \in \mathcal{K}_{\mathring{Z}} \,:\, \boldsymbol{z} \in B} \alpha = \max_{(\boldsymbol{t}, \alpha) \in \mathcal{K}_{\mathring{T}} \,:\, \phi(\boldsymbol{t}) \in B} \alpha \tag{5.61}$$

for $B \in (\mathbb{Z})$. Compared to Vanilla Monte-Carlo methods, which sum over the samples—or rather over their weights—that map into B under ϕ, one maximizes over their memberships instead. As a rule of thumb, the quality of this approximation grows with the number of samples that fall into B. If this number is zero, it is difficult to judge whether the possibility of B is zero, too, or whether the initial sampling was defective. Naturally, precision increases for large sample numbers and for larger, more inclusive sets B.

Similar observations apply to the approximation of necessities, cumulative distribution functions and expectations.

Example 18: Π-IM for Ratio of Gaussian Means

Suppose $\tilde{Q}_1 \sim \mathcal{N}(\theta_1, 1)$ and $\tilde{Q}_2 \sim \mathcal{N}(\theta_2, 1)$ are two independent observations from Gaussian distributions with unknown means $\theta_1, \theta_2 \in \mathbb{R}$ and unit variance and zero covariance. Using the same technique for the μ-transform as in Example 15, the joint confidence distribution of θ_1 and θ_2 obtained under the Complementary Cumulative μ-transform of the $\chi^2(2)$-distributed imprecise variable $\tilde{V} = (\tilde{Q}_1 - \theta_1)^2 + (\tilde{Q}_2 - \theta_2)^2$ is given by

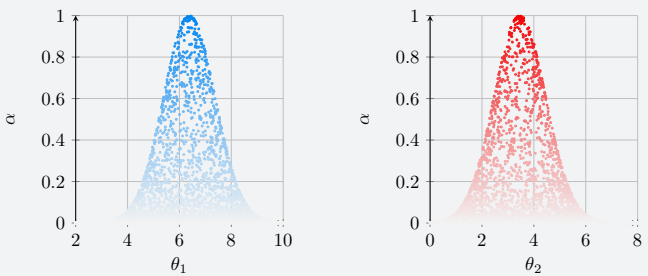
$$\gamma_{\hat{\theta}_1,\hat{\theta}_2|q_1,q_2}(\theta_1, \theta_2) = 1 - F_{\chi^2(2)}\left((q_1 - \theta_1)^2 + (q_2 - \theta_2)^2\right)$$

for $\theta_1, \theta_2 \in \mathbb{R}$ and the observations $\tilde{Q}_1 = q_1$ and $\tilde{Q}_2 = q_2$.

The μ-clusters $\mathcal{K}_{\hat{\theta}_1}$ and $\mathcal{K}_{\hat{\theta}_2}$ obtained by the marginalization of the μ-cluster $\mathcal{K}_{\hat{\theta}_1,\hat{\theta}_2}$ approximating $\gamma_{\hat{\theta}_1,\hat{\theta}_2|q_1,q_2}$ via $m = 5 \cdot 10^3$ Latin hypercube samples on $[2, 10] \times [0, 8]$ for the observations[53]

$$q_1 = 6.38 \qquad \text{and} \qquad q_2 = 3.46$$

are shown below.



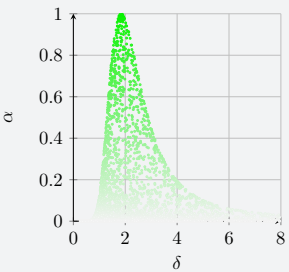
Marginal Parameter μ-Cluster $\mathcal{K}_{\hat{\theta}_1}$ Marginal Parameter μ-Cluster $\mathcal{K}_{\hat{\theta}_2}$

The feature μ-cluster $\mathcal{K}_{\hat{\delta}}$ for $\delta = \phi(\theta_1, \theta_2) = \frac{\theta_1}{\theta_2}$ obtained by the propagation of $\mathcal{K}_{\hat{\theta}_1,\hat{\theta}_2}$

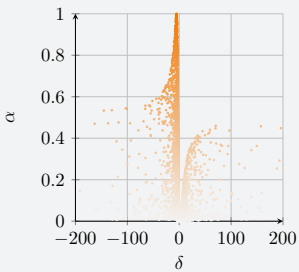
under ϕ is shown below on the left. Additionally, the propagation output $\mathcal{K}'_{\hat{\delta}}$ for the observations[54]

$$q'_1 = 5.98 \qquad \text{and} \qquad q'_2 = -1.16$$

from $m' = 5 \cdot 10^3$ initial Latin hypercube samples on $[3, 9] \times [-4, 2]$ is shown below on the right.

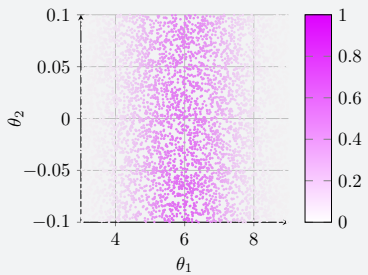


Feature μ-Cluster $\mathcal{K}_{\hat{\delta}}$

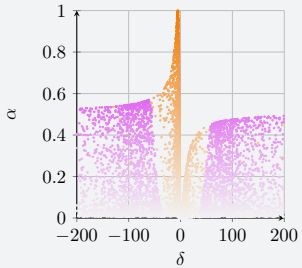


Feature μ-Cluster $\mathcal{K}_{\hat{\delta}}{}'$

Notice that, in order to better resolve the edge of the implied μ-function $\mu_{\mathcal{K}'_{\hat{\delta}}}$, especially in the second figure, it is necessary to generate more samples of θ_2 around the origin since these points are mainly contributing to the μ-function far away from zero. For instance, adding $m'' = 5 \cdot 10^3$ Latin hypercube samples on $[3, 9] \times [-0.1, 0.1]$, which produces the μ-cluster $\mathcal{K}''_{\hat{\theta}_1,\hat{\theta}_2}$, shown below on the left, results in the additional feature μ-cluster $\mathcal{K}''_{\hat{\delta}}$ shown on the right, and achieves the desired resolution of the approximated feature distribution $\gamma_{\hat{\delta}|q'_1,q'_2}$.



Parameter μ-Cluster $\mathcal{K}''_{\hat{\theta}_1,\hat{\theta}_2}$



Feature μ-Clusters $\mathcal{K}'_{\hat{\delta}}$ and $\mathcal{K}''_{\hat{\delta}}$

The problem of finding confidence intervals of ratios of means is known as the *Fieller-Creasy Problem* [KappenmanGeisserAntle70], for which the Fieller Theorem [Fieller54] provides a theoretical solution. The results obtained in the II-IM framework are similar to the membership functions obtained by Martin and Liu[55] for

this problem; when replacing the above μ-transform by the Symmetric Cumulative μ-transform of the normally distributed imprecise variable

$$\tilde{V} = \frac{\tilde{Q}_1 - \frac{\theta_1}{\theta_2}\tilde{Q}_2}{\sqrt{1 + \left(\frac{\theta_1}{\theta_2}\right)^2}} \sim \mathcal{N}(0,1)$$

they coincide. As Martin and Liu observe, "the [second output μ-cluster] has a very unusual shape" [MartinLiu15, p. 134] indicating confidence regions that are composed of two disjoint intervals, which further confirms the results of Fieller.

In conclusion, both the membership transform and the membership extension are straightforward to implement via interval- and sampling-based methods. Especially the propagation of fuzzy variables, the most common membership extension operation, is very efficiently implemented by the latter, whereas general implicit extensions may only be computed through less efficient but rigorous interval arithmetic. The following chapter will extensively use these numerical strategies and showcase some applications to the analysis of dynamical systems.

[53] These data points correspond to those considered by Martin and Liu [MartinLiu15, Chapter 7.3.2].
[54] See footnote 53.
[55] Martin and Liu call this a marginal plausibility function, which is—albeit related—not to be confused with the plausibility function defined here.

Chapter 6

Possibilistic Filtering

> 'As simple as that? You didn't use magic?'
> 'Only common sense. It's a lot more reliable in the long run.'
>
> Terry Pratchett, *Mort*

This chapter is intended to showcase an application of the derived theory to dynamical systems, namely dynamic filtering.

Possibilisitic analyses have repeatedly been applied to dynamical systems, e.g. to crash simulations [BiehlerEtAl19, MäckHanss19], predictions of the vibratory properties of helicopters [FröhlichEtAl22], the stability analysis of machining systems [HamannEtAl18], and controller design [HofmannHanss16, HoseMäckHanss19b]. While these analyses are, for the most part, restricted to *uncertainty propagation*, i.e. to the propagation of fuzzy variables through a dynamic model, considerably fewer have attempted to perform statistical inference, e.g. for the parameter identification of machining systems [HoseEtAl18] or simple airplane models [HoseHanss19d].

The main statistical tasks in the analysis of dynamical systems are *decision-making*, including, e.g., the system design and control, and the identification of model parameters and the estimation of the current dynamic states, which may be subsumed under the term *inference*. An exhaustive discussion of these tasks is beyond the scope of this thesis; however, thorough overviews of classical concepts of the former and the latter are, e.g., given by Skogestad and Postlethwaite [SkogestadPostlethwaite07] and Tangirala [Tangirala14], respectively.

The *filtering problem* to be considered next can be summarized as the task of quantifying the belief surrounding an estimate of the current state of a dynamical system. This is, e.g., motivated by the need to be later able to make robust decisions, e.g. to find appropriate control inputs. Preliminary concepts of a possibilistic filter [BenferhatDuboisPrade00,

HoseHanss21a, HoseHanss21b] serve as a basis for the following discussion, which is also inspired by set-membership filtering techniques [MilaneseNovara11, LeongNair16].

The presented filtering methodology can be understood as a rigorous alternative to Bayesian filters [Särkkä13], where—apart from the general unsuitedness of additive belief structures inherent to the latter—the common, but problematic, modeling of process noise by a single probability measure is avoided. As explained in more detail below, this perturbation is usually an expression of missing insight into the actual system dynamics—rather than of actual random behavior of the system; in fact, the system may very well be entirely deterministic and only imperfectly described, a property that is common to the majority of mathematical abstractions and models. By modeling this uncertainty with a precise probability distribution, the actual information is not accurately described, and any interpretability of the corresponding belief measure, something engineers should seek to provide, is eliminated.

Next, the statistical filtering setup shall be discussed, and two formulations of a possibilistic filter shall be derived.

6.1 Formulation

Consider a time-discrete nonlinear system with the state space description consisting of the *dynamic model*

$$\boldsymbol{x}_k = f\left(\boldsymbol{x}_{k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{n}_{k-1}\right), \tag{6.1}$$

and the *observation/measurement model*

$$\boldsymbol{y}_k = g\left(\boldsymbol{x}_k\right) + \boldsymbol{w}_k, \tag{6.2}$$

where $\boldsymbol{x}_k \in \mathbb{X} \subseteq \mathbb{R}^{D_\mathbb{X}}$ is the *system state*, including, e.g., the pose (position and orientation) and (translational and rotatory) velocities of a robot. Furthermore, $\boldsymbol{u}_k \in \mathbb{U} \subseteq \mathbb{R}^{D_\mathbb{U}}$ is the *system input*, $\boldsymbol{y}_k \in \mathbb{Y} \subseteq \mathbb{R}^{D_\mathbb{Y}}$ is the *system output*, $\boldsymbol{n}_k \in \mathbb{N} \subseteq \mathbb{R}^{D_\mathbb{N}}$ is the *process noise* and $\boldsymbol{w}_k \in \mathbb{W} \subseteq \mathbb{R}^{D_\mathbb{W}}$ is the *additive measurement error* at time instants $k = 1, \ldots, K$, respectively. The *initial state* is denoted by $\boldsymbol{x}_0 \in \mathbb{X}$.

The input is also said to be the *control input* because it describes an exogenous actuation from an operator, controller, etc. Conversely, the process noise is considered to be an *uncontrollable*, unknown input to the system such that its time evolution, i.e. the sequence of states $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_k$ is not entirely predictable. Additionally, the measurement model describes the fact that the states may only be indirectly observable[56] and perturbed by the (random) measurement error. The additivity of the latter is usually justified as a credible uncertainty description of noisy sensor output and considerably justifies the

[56] A thorough discussion and treatment of the concept of observability cannot be provided here, see e.g. [SkogestadPostlethwaite07, Chapter 4].

following deliberations. It is certainly possible to extend the derived filtering scheme to non-additive measurement noise, but the increased generality would come at the expense of a significantly higher computational effort, and it would prohibit the particle-based implementation proposed below.

Only if $\boldsymbol{x}_0$ and $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{k-1}$ were known in addition to $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}$, one would be able to precisely predict the true value of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Furthermore, knowing $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$, would enable one to also predict $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$. Yet, precisely $\boldsymbol{x}_0$, $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{k-1}, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ are unknown, and the fundamental question is, what information about the states is gained from the measurements $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$ that are observed instead. Consequently, the aim of a filter for the system in Eqs. (6.1) and (6.2) is to infer the current system state $\boldsymbol{x}_k$ given the past and current observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, as well as the dynamic and measurement model and the past inputs $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}$.

This task may be cast into the framework of statistical inference, such that the problem becomes that of deriving an appropriate IM for the current state given suitable uncertainty models of the initial state, process noise and measurement error. More precisely, a Π-PM of $\boldsymbol{x}_k$ shall be derived in the following; but, before doing this, a statistical model of the dynamic and measurement model must be formulated, i.e., the uncertainties contained therein must be addressed.

6.1.1 Statistical Model

Beginning with the least debatable uncertainty model, the measurement errors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ are generally assumed to be observations of the k imprecise variables $\tilde{W}_1, \ldots, \tilde{W}_k$ that are iid realizations of the imprecise variable $\tilde{W} \sim \mathrm{P}_{\tilde{W}}$. Usually, $\tilde{W}$ is modeled as a multivariate Gaussian variable $\tilde{W} \sim \mathcal{N}(0, \boldsymbol{R})$ with zero mean and known covariance matrix $\boldsymbol{R}$, which can, e.g., be justified because the error exhibited by most sensors is approximately zero-mean Gaussian. In any case, one may assume that $\tilde{W}$ is (robustly) described by an available (I)P description distribution $\pi_{\tilde{W}}$, such that the possibilistic uncertainty model of the measurement error is $\tilde{W}_1, \ldots, \tilde{W}_k \sim \pi_{\tilde{W}}$. These marginal variables are often assumed to be independent, but for the sake of robustness, one may also assume unknown interaction.

From this discussion alone, one could formulate the statistical model

$$\tilde{Y}_k = g(\boldsymbol{x}_k) + \tilde{W}_k \tag{6.3}$$

which, under the membership inversion and the Pivotal Step in Eq. (4.27), implies a Π-IM $\gamma_{\hat{\boldsymbol{x}}_k|\tilde{Y}_k}$ producing the confidence distributions

$$\gamma_{\hat{\boldsymbol{x}}_k|\boldsymbol{y}_k}(\boldsymbol{x}_k) = \pi_{\tilde{Y}_k|\boldsymbol{x}_k}(\boldsymbol{y}_k) = \pi_{\tilde{W}}\left(\boldsymbol{y}_k - g\left(\boldsymbol{x}_k\right)\right) \tag{6.4}$$

for the unknown current state. This would, however, entirely neglect any information that may have been obtained from the previous observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}$ with respect to the

previous state sequence $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{k-1}$, which is connected to $\boldsymbol{x}_k$ via the dynamic model in Eq. (6.1), and may improve the inference. Yet, the inclusion of such information requires additional considerations.

The process noise $\boldsymbol{n}_0, \ldots, \boldsymbol{n}_{k-1}$, too, is often considered to be composed of k iid realizations of a multivariate Gaussian variable. This approach follows the early formulations, e.g. of the Kalman filter [Kalman60], and allows for intuitive and straightforward recursive formulations of Bayesian filters [Särkkä13, ShaoHuangLee10], which do, however, suffer from probability dilution as shown by the False Confidence Theorem [BalchMartinFerson19]. While one may argue that, as a model for the superposition of many random effects, every realization $\boldsymbol{n}_i$ itself may indeed be approximately Gaussian, it is much more challenging to fix the appropriate mean and covariance in advance—and it is undoubtedly questionable to assume these to be iid. More precisely, one may argue that the process noise is primarily an effect of unmodeled or ill-understood system dynamics, including but not limited to such effects as, e.g., the integration error made by transforming a time-continuous into a time-discrete system. They might exhibit very little random behavior and, instead, obey largely deterministic laws—explaining the success that, e.g., the identification of hidden system dynamics via Gaussian Process models has seen recently [EschmannEbelEberhard21]. In other words, modeling the process noise to follow a precise probability distribution may fail to describe the dynamics accurately, and IP-based uncertainty models are preferable.

To avoid unwarranted assumptions, it is always safe to assume the extreme case that the probability distribution of $\tilde{N}$ is confined to $\mathbb{N}$, and nothing more. This lack of knowledge can be described exactly by a (quasi-)vacuous[57] IP description distribution on $\mathbb{N}$. Additionally, it includes a purely deterministic dynamic model as explained in Section 4.3.5 because distinguishing between (quasi-)vacuous imprecise variables and unknown parameters in the absence of data is, to a certain extent, arbitrary; the resulting membership function is always given by the (quasi-)vacuous distribution $\mathcal{Q}(\mathbb{N})$. Alternatively,—but only with good reason—one could also justify an alternative (I)P description distribution $\pi_{\tilde{N}}$ if genuine imprecise probabilistic information on the distribution of $\tilde{N}$ is available; therefore, the following discussion assumes a general possibilistic uncertainty model $\tilde{N}_0, \ldots, \tilde{N}_{k-1} \sim \pi_{\tilde{N}}$, but $\pi_{\tilde{N}} = \mathcal{Q}(\mathbb{N})$ should usually be assumed. In any case, any kind of independence assumption would, in most cases, be questionable, and unknown interaction between these marginals should usually be assumed.

Similar arguments justify modeling the information about the initial state $\boldsymbol{x}_0$ via the quasi-vacuous distribution $\mathcal{Q}(\mathbb{X}_0)$ on $\mathbb{X}_0 \subseteq \mathbb{X}$, but here the more general possibilistic uncertainty model $\tilde{X}_0 \sim \pi_{\tilde{X}_0}$ is considered.

Finally, if $\tilde{N}_0, \ldots, \tilde{N}_{k-1}, \tilde{W}_1, \ldots, \tilde{W}_k, \tilde{X}_0$ are assumed to be imprecise variables, the sequence of states also constitutes a sequence of imprecise variables $\tilde{X}_1, \ldots, \tilde{X}_k$ that are correlated

[57] The distribution is vacuous if the reference space is $\mathbb{N}$ and quasi-vacuous if the reference space is $\mathbb{R}^{D_{\mathbb{N}}}$.

via the statistical dynamic model

$$\begin{aligned} \tilde{X}_1 &= f\left(\tilde{X}_0, \boldsymbol{u}_0, \tilde{N}_0\right) \\ &\vdots \\ \tilde{X}_k &= f\left(\tilde{X}_{k-1}, \boldsymbol{u}_{k-1}, \tilde{N}_{k-1}\right), \end{aligned} \tag{6.5}$$

and the measurements are modeled as imprecise variables $\tilde{Y}_1, \ldots, \tilde{Y}_k$ given by the statistical measurement model

$$\begin{aligned} \tilde{Y}_1 &= g\left(\tilde{X}_1\right) + \tilde{W}_1 \\ &\vdots \\ \tilde{Y}_k &= g\left(\tilde{X}_k\right) + \tilde{W}_k \end{aligned} \tag{6.6}$$

for $k = 1, \ldots, K$.

To summarize the above discussion, Eqs. (6.5) and (6.6) constitute the implicit relationship

$$\begin{aligned} 0 &= \Xi(\tilde{N}_0, \ldots, \tilde{N}_{k-1}, \tilde{W}_1, \ldots, \tilde{W}_k, \tilde{X}_0, \ldots, \tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k) \\ &= \begin{pmatrix} \tilde{X}_1 - f\left(\tilde{X}_0, \boldsymbol{u}_0, \tilde{N}_0\right) \\ \vdots \\ \tilde{X}_k - f\left(\tilde{X}_{k-1}, \boldsymbol{u}_{k-1}, \tilde{N}_{k-1}\right) \\ \tilde{Y}_1 - g\left(\tilde{X}_1\right) - \tilde{W}_1 \\ \vdots \\ \tilde{Y}_k - g\left(\tilde{X}_k\right) - \tilde{W}_k \end{pmatrix} \end{aligned} \tag{6.7}$$

between the imprecise variables that can be described possibilistically, i.e. between the process noise $\tilde{N}_0, \ldots, \tilde{N}_{k-1} \sim \pi_{\tilde{N}}$, the measurement error $\tilde{W}_1, \ldots, \tilde{W}_k \sim \pi_{\tilde{W}}$ and the initial state $\tilde{X}_0 \sim \pi_{\tilde{X}_0}$, and the past and current states $\tilde{X}_1, \ldots, \tilde{X}_k$ and measurements $\tilde{Y}_1, \ldots, \tilde{Y}_k$ with unknown possibility distribution, where $\tilde{Y}_1 = \boldsymbol{y}_1, \ldots, \tilde{Y}_k = \boldsymbol{y}_k$ is observed for a given input sequence $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}$. Altogether, this constitutes the *statistical filtering model*.

6.1.2 Filter Formulation

The statistical filtering model in Eq. (6.7) itself is not explicitly determined by any unknown population parameters; in particular, the current state $\boldsymbol{x}_k$ itself, as the unknown quantity of interest in filtering, should not be considered to be a population parameter but rather an imprecise variable that must be predicted by a Π-PM $\kappa_{\tilde{X}_k|\tilde{Y}_1,\ldots,\tilde{Y}_k}$. The statistical model in Eq. (6.7) forms the basis to construct at least two different filtering Π-PMs of $\tilde{X}_k$, which shall be discussed in the following.

6.1.2.1 Batch Formulation

Following earlier ideas by Hose and Hanss [HoseHanss21a], a batch filter is straightforward to derive. To this end, a joint possibility distribution of the marginal variables $\tilde{N}_0, \ldots, \tilde{N}_{k-1}, \tilde{W}_1, \ldots, \tilde{W}_k, \tilde{X}_0$ must be found. Here, it is constructed step-wise.

A joint possibility distribution of $\tilde{W}_1, \ldots, \tilde{W}_k$ can usually be obtained directly by the μ-transform of the joint probability distribution

$$\mathrm{P}_{\tilde{W}_1,\ldots,\tilde{W}_k} = C\left(\mathrm{P}_{\tilde{W}_1}, \ldots, \mathrm{P}_{\tilde{W}_k}\right), \tag{6.8}$$

e.g. for $C = C^{\text{ind.}}$ if the measurements are considered to be stochastically independent—which is often applicable when noisy sensor output is modeled by the measurement error—, yielding $\pi_{\tilde{W}_1,\ldots,\tilde{W}_k}$.

In contrast, for the process noise $\tilde{N}_0, \ldots, \tilde{N}_{k-1}$, the UI-Π-copula $\mathcal{J} = \mathcal{J}^{\mathrm{UI}}$ must usually be applied to find the joint possibility distribution

$$\pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1}} = \mathcal{J}\left(\pi_{\tilde{N}_0}, \ldots, \pi_{\tilde{N}_{k-1}}\right) \tag{6.9}$$

from the IP descriptive marginal distributions $\pi_{\tilde{N}_0}, \ldots, \pi_{\tilde{N}_{k-1}}$ due to the process noise possibly describing unmodeled system dynamics, making $\tilde{N}_0, \ldots, \tilde{N}_{k-1}$ highly correlated. However, assuming quasi-vacuousness marginal distributions this always results in the quasi-vacuous joint distribution $\pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1}} = \mathcal{Q}(\bigtimes_k \mathbb{N})$ on the k-fold Cartesian product of $\mathbb{N}$ with itself.

Finally, the relationship between the remaining three marginals $(\tilde{N}_0, \ldots, \tilde{N}_{k-1})$, $(\tilde{W}_1, \ldots, \tilde{W}_k)$ and $\tilde{X}_0$ can be assumed to be well described by stochastic independence, such that the SI-Π-copula $\mathcal{J}^{\mathrm{SI}}$ can usually be applied for the construction of the joint possibility distribution

$$\pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1},\tilde{W}_1,\ldots,\tilde{W}_k,\tilde{X}_0} = \mathcal{J}\left(\pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1}}, \pi_{\tilde{W}_1,\ldots,\tilde{W}_k}, \pi_{\tilde{X}_0}\right). \tag{6.10}$$

However, if both $(\tilde{N}_0, \ldots, \tilde{N}_{k-1})$ and $\tilde{X}_0$ are described by quasi-vacuous possibility distributions, one may, by Proposition 36, apply the non-interactive Π-copula $\mathcal{J}^{\mathrm{NI}}$.

In principle, a possibilistic description $\pi_{\tilde{X}_1,\ldots,\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ of the past and current states $\tilde{X}_1, \ldots, \tilde{X}_k$ and measurements $\tilde{Y}_1, \ldots, \tilde{Y}_k$ can directly be computed from the implicit

extension of $\pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1},\tilde{W}_1,\ldots,\tilde{W}_k,\tilde{X}_0}$ under Ξ, yielding the *batch description*

$$\begin{aligned}
&\pi_{\tilde{X}_1,\ldots,\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k)\\
&= \sup_{\substack{\boldsymbol{x}_0\in\mathbb{X},\\ \boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1}\in\mathbb{N},\\ \boldsymbol{w}_1,\ldots,\boldsymbol{w}_k\in\mathbb{W}:\\ 0=\Xi(\,\boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1},\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k,\\ \boldsymbol{x}_0,\ldots,\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k\,)}} \pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1},\tilde{W}_1,\ldots,\tilde{W}_k,\tilde{X}_0}(\boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1},\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k,\boldsymbol{x}_0)\\
&= \sup_{\substack{\boldsymbol{x}_0\in\mathbb{X},\\ \boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1}\in\mathbb{N},\\ \boldsymbol{w}_1,\ldots,\boldsymbol{w}_k\in\mathbb{W}:\\ \boldsymbol{x}_1=f(\boldsymbol{x}_0,\boldsymbol{u}_0,\boldsymbol{n}_0),\\ \vdots\\ \boldsymbol{x}_k=f(\boldsymbol{x}_{k-1},\boldsymbol{u}_{k-1},\boldsymbol{n}_{k-1}),\\ \boldsymbol{y}_1=g(\boldsymbol{x}_1)+\boldsymbol{w}_1,\\ \vdots\\ \boldsymbol{y}_k=g(\boldsymbol{x}_k)+\boldsymbol{w}_k}} \pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1},\tilde{W}_1,\ldots,\tilde{W}_k,\tilde{X}_0}(\boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1},\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k,\boldsymbol{x}_0)
\end{aligned} \tag{6.11}$$

for all $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k\in\mathbb{X}$ and all $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k\in\mathbb{Y}$.

From this expression, one could already construct a Π-PM of $\tilde{X}_1,\ldots,\tilde{X}_k$ given the observations $\tilde{Y}_1=\boldsymbol{y}_1,\ldots,\tilde{Y}_k=\boldsymbol{y}_k$. However, since only the current state is of interest, and the past states are considered to be 'nuisance predictions', a further marginalization yields the marginal description $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ given by

$$\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}}(\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k)=\sup_{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{k-1}\in\mathbb{X}}\pi_{\tilde{X}_1,\ldots,\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) \tag{6.12}$$

for all $\boldsymbol{x}_k\in\mathbb{X}$.

Under the Semi-Pivotal Step in Eq. (4.51), this allows one to define a Π-PM $\kappa_{\tilde{X}_k|\tilde{Y}_1,\ldots,\tilde{Y}_k}$ of $\tilde{X}_k$ given $\tilde{Y}_1,\ldots,\tilde{Y}_k$. For the observations $\tilde{Y}_1=\boldsymbol{y}_1,\ldots,\tilde{Y}_k=\boldsymbol{y}_k$, this produces the prediction distributions

$$\begin{aligned}
&\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}(\boldsymbol{x}_k)=\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k)\\
&= \sup_{\substack{\boldsymbol{x}_0,\ldots,\boldsymbol{x}_{k-1}\in\mathbb{X},\\ \boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1}\in\mathbb{N},\\ \boldsymbol{w}_1,\ldots,\boldsymbol{w}_k\in\mathbb{W}:\\ \boldsymbol{x}_1=f(\boldsymbol{x}_0,\boldsymbol{u}_0,\boldsymbol{n}_0),\\ \vdots\\ \boldsymbol{x}_k=f(\boldsymbol{x}_{k-1},\boldsymbol{u}_{k-1},\boldsymbol{n}_{k-1}),\\ \boldsymbol{y}_1=g(\boldsymbol{x}_1)+\boldsymbol{w}_1,\\ \vdots\\ \boldsymbol{y}_k=g(\boldsymbol{x}_k)+\boldsymbol{w}_k}} \pi_{\tilde{N}_0,\ldots,\tilde{N}_{k-1},\tilde{W}_1,\ldots,\tilde{W}_k,\tilde{X}_0}(\boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1},\boldsymbol{w}_1,\ldots,\boldsymbol{w}_k,\boldsymbol{x}_0)
\end{aligned} \tag{6.13}$$

for all $\boldsymbol{x}_k\in\mathbb{X}$, given the input and measurement sequences $\boldsymbol{u}_0,\ldots,\boldsymbol{u}_{k-1}$ and $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k$, respectively.

This *possibilistic batch filtering formulation* in the form of a mathematical program, including possibly non-linear and/or non-convex constraints and objective functions, is

not trivial and can, generally, only be solved efficiently via numerical optimization in special cases, such as linear time-invariant (LTI) systems in connection with concave and/or (quasi-)vacuous input possibility distributions [HoseHanss21a].

LTI Batch Filtering Consider the LTI system dynamics and the measurement model given by

$$f(\boldsymbol{x}, \boldsymbol{u}, \boldsymbol{n}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}_u\boldsymbol{u} + \boldsymbol{B}_n\boldsymbol{n} \qquad \text{and} \qquad g(\boldsymbol{x}) = \boldsymbol{C}\boldsymbol{x} \tag{6.14}$$

for the system matrix $\boldsymbol{A} \in \mathbb{R}^{D_\mathbb{X} \times D_\mathbb{X}}$, the control input matrix $\boldsymbol{B}_u \in \mathbb{R}^{D_\mathbb{X} \times D_\mathbb{U}}$, the noise input matrix $\boldsymbol{B}_n \in \mathbb{R}^{D_\mathbb{X} \times D_\mathbb{N}}$, and the observation matrix $\boldsymbol{C} \in \mathbb{R}^{D_\mathbb{Y} \times D_\mathbb{X}}$. Let the process noise and initial state descriptions be (quasi-)vacuous on $\mathbb{N}$ and $\mathbb{X}_0$, respectively, and let the measurement error description, e.g., be given by

$$\pi_{\tilde{W}_1,\ldots,\tilde{W}_k}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k) = 1 - F_{\chi^2(k \cdot D_\mathbb{W})}\left(\sum_{i=1}^{k} \boldsymbol{w}_i^\mathrm{T} \cdot \boldsymbol{R}^{-1}\boldsymbol{w}_i\right) \tag{6.15}$$

for $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k \in \mathbb{W}$. This possibility distribution is—by the inversion of the Complementary Cumulative μ-transform of the χ^2-distributed variable $\tilde{V} = \sum_{i=1}^{k} \tilde{W}_i^\mathrm{T} \cdot \boldsymbol{R}^{-1}\tilde{W}_i$ with $k \cdot D_\mathbb{W}$ degrees of freedom—consistent with the joint probability distribution $\mathrm{P}_{\tilde{W}_1,\ldots,\tilde{W}_k}$ of k iid multivariate Gaussian variables $\tilde{W}_1, \ldots, \tilde{W}_k \sim \mathcal{N}(0, \boldsymbol{R})$ with covariance matrix $\boldsymbol{R}$, cf. Examples 15 and 18. Moreover, it coincides with the Optimal μ-transform of $\mathrm{P}_{\tilde{W}_1,\ldots,\tilde{W}_k}$. Then, the possibilistic batch filtering formulation in Eq. (6.13) reduces to

$$\begin{aligned}
\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}(\boldsymbol{x}_k) = & \sup_{\substack{\boldsymbol{x}_0 \in \mathbb{X}_0,\\ \boldsymbol{x}_1,\ldots,\boldsymbol{x}_{k-1} \in \mathbb{X},\\ \boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1} \in \mathbb{N},\\ \boldsymbol{w}_1,\ldots,\boldsymbol{w}_k \in \mathbb{W}:\\ \boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{B}_u\boldsymbol{u}_0 + \boldsymbol{B}_n\boldsymbol{n}_0,\\ \vdots\\ \boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{B}_u\boldsymbol{u}_{k-1} + \boldsymbol{B}_n\boldsymbol{n}_{k-1},\\ \boldsymbol{y}_1 = \boldsymbol{C}\boldsymbol{x}_1 + \boldsymbol{w}_1,\\ \vdots\\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{w}_k}} \pi_{\tilde{W}_1,\ldots,\tilde{W}_k}(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k)\\
= & \sup_{\substack{\boldsymbol{x}_0 \in \mathbb{X}_0,\\ \boldsymbol{x}_1,\ldots,\boldsymbol{x}_{k-1} \in \mathbb{X},\\ \boldsymbol{n}_0,\ldots,\boldsymbol{n}_{k-1} \in \mathbb{N},\\ \boldsymbol{w}_1,\ldots,\boldsymbol{w}_k \in \mathbb{W}:\\ \boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{B}_u\boldsymbol{u}_0 + \boldsymbol{B}_n\boldsymbol{n}_0,\\ \vdots\\ \boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{B}_u\boldsymbol{u}_{k-1} + \boldsymbol{B}_n\boldsymbol{n}_{k-1},\\ \boldsymbol{y}_1 = \boldsymbol{C}\boldsymbol{x}_1 + \boldsymbol{w}_1,\\ \vdots\\ \boldsymbol{y}_k = \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{w}_k}} 1 - F_{\chi^2(k \cdot D_\mathbb{W})}\left(\sum_{i=1}^{k} \boldsymbol{w}_i^\mathrm{T} \cdot \boldsymbol{R}^{-1}\boldsymbol{w}_i\right)
\end{aligned} \tag{6.16}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$, given the in- and output sequences $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, respectively. This expression describes a basic mathematical program with linear equality constraints and a non-convex, but well-behaved, objective function that can efficiently be solved by standard optimization algorithms building upon an abundance of existing theory [NocedalWright06].

An implementation of this possibilistic LTI batch filter based on polytopes and linear programming is described by Hose and Hanss [HoseHanss21a]. On the one hand, this implementation guarantees a numerically exact solution of the mathematical program in Eq. (6.16); on the other hand, it is computationally demanding and quickly suffers from numerical infeasibility problems for large in- and output sequences, i.e. for large k. In order to enable the real-time applicability of this batch filtering technique, it would be recommendable to consider moving horizon-like batch filtering [RaoRawlingsMayne03], where the in- and output sequences are to consist of only the last k' data points, where $k' \ll k$, such that the optimization variables in the mathematical program are kept at a fixed number and do not grow infinitely. While this implies a loss of information, it may result in a considerable speed-up and robustify the inference process by discarding older—possibly misleading—data and, thus, avoid numerical infeasibilities. This does, however, still not solve the issue of the restriction to LTI systems, which is quite limiting when, especially in engineering, system descriptions tend to be linear—not least due to the increased computational power that is available nowadays, and allows to simulate nonlinear dynamics, and due to the sensor models moving away from direct to more indirect observations. Therefore, a different, application-oriented filtering formulation shall be pursued in the following.

6.1.2.2 Recursive Formulation

An arguably computationally more tractable filtering technique is obtained when deriving a recursive formulation based on iterative predictions, (inversions) and updates—similar to the steps commonly found in other filter formulations [Kalman60, Särkkä13, ShaoHuangLee10, MilaneseNovara11, LeongNair16]. The general feasibility of such schemes in a possibilistic setting has so far been considered by few scholars, e.g. from a perspective of possibilistic conditioning [BenferhatDuboisPrade00]. However, Hose and Hanss [HoseHanss21b] describe a filtering Π-IM, treating the current state as an unknown parameter. Even though this philosophy disagrees with its interpretation as an imprecise variable here, their derivation still serves as a guideline to the following inductive derivation in k.

The statistical model in Eq. (6.7) admits the decomposition

$$\begin{aligned} 0 &= \Xi\left(\tilde{N}_0, \dots, \tilde{N}_{k-1}, \tilde{W}_1, \dots, \tilde{W}_k, \tilde{X}_0, \dots, \tilde{X}_k, \tilde{Y}_1, \dots, \tilde{Y}_k\right) \\ &= \begin{pmatrix} \Xi_1(\tilde{N}_0, \dots, \tilde{N}_{k-2}, \tilde{W}_1, \dots, \tilde{W}_{k-1}, \tilde{X}_0, \dots, \tilde{X}_k, \tilde{Y}_1, \dots, \tilde{Y}_k) \\ \Xi_2(\tilde{N}_{k-1}, \tilde{X}_{k-1}, \tilde{X}_k) \\ \Xi_3(\tilde{W}_k, \tilde{X}_k, \tilde{Y}_k) \end{pmatrix} \end{aligned} \tag{6.17}$$

as visualized in Figure 6.1. Put simply, submodel Ξ_1 describes the relation between the past states and the past measurements, whereas Ξ_2 is the relation between the past and current state. Finally, submodel Ξ_3 connects the current state to the current measurement.

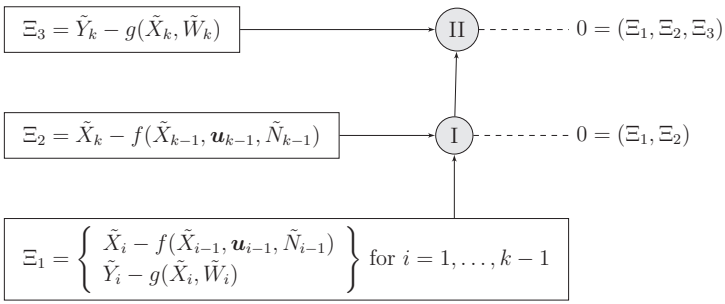


Figure 6.1: Two-fold decomposition of $0 = \Xi = (\Xi_1, \Xi_2, \Xi_3)$.

As a preliminary remark, observe that $0 = \Xi_1$ is equivalent to the full statistical model $0 = \Xi$ at time step $k-1$. Assuming that a description distribution $\pi_{\tilde{X}_{k-1}, \tilde{Y}_1, \ldots, \tilde{Y}_{k-1}}$ describes the statistical model implied by $0 = \Xi_1$, i.e. $0 = \Xi$ at the previous time step $k-1$, it remains to be shown that, therefrom, one can construct a description distribution $\pi_{\tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k}$ under $0 = \Xi$ at the current time step k.

This assumption is fulfilled for $k = 1$, where only the possibilistic description of the initial state $\tilde{X}_0$ in the absence of measurements is available. Reiterating the argument from Section 4.3.5 that a prediction distribution in the absence of data coincides with a description distribution, the corresponding initial filtering Π-PM is

$$\kappa_{\tilde{X}_0|\emptyset} = \pi_{\tilde{X}_0}. \tag{6.18}$$

If one can show the sufficiency of the existence of such a possibilistic description $\pi_{\tilde{X}_{k-1}, \tilde{Y}_1, \ldots, \tilde{Y}_{k-1}}$ for the purpose of deriving a possibilistic description $\pi_{\tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k}$, the admissibility of the recursive filter formulation follows by induction in k.

To this end, the following general strategy is proposed: The two remaining statistical submodels Ξ_2 and Ξ_3 are considered individually, leading to three distinct possibilistic descriptions stemming from different information in the statistical model implied by $0 = \Xi$. Following Figure 6.1, the two-fold decomposition also requires a two-fold re-combination of the information/descriptions, which shall now be approached step-wise.

1. In a preliminarly step, the vacuous extension of the previous description $\pi_{\tilde{X}_{k-1}, \tilde{Y}_1, \ldots, \tilde{Y}_{k-1}}$ from $(\tilde{X}_{k-1}, \tilde{Y}_1, \ldots, \tilde{Y}_k)$ onto $(\tilde{X}_{k-1}, \tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k)$ is formally defined via Eq. (3.42) yielding

   $$\pi^{(1)}_{\tilde{X}_{k-1}, \tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k}(\boldsymbol{x}_{k-1}, \boldsymbol{x}_k, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_k) = \pi_{\tilde{X}_{k-1}, \tilde{Y}_1, \ldots, \tilde{Y}_{k-1}}(\boldsymbol{x}_{k-1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}) \tag{6.19}$$

   for all $\boldsymbol{x}_{k-1}, \boldsymbol{x}_k \in \mathbb{X}$ and all $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathbb{Y}$.

The reason for this technicality is the intended combination with submodels Ξ_2 and Ξ_3 at nodes I and II, respectively. Such combinations of possibilistic descriptions must happen either via the conjunction on the same sample space, or via the construction of a joint distribution on entirely disjoint sample spaces. Since submodels Ξ_1 and Ξ_2 share a dependency on $\tilde{X}_{k-1}$, the conjunction on a shared sample space promises to yield a more specific description.

2. The (implicit) extension of $\pi_{\tilde{N}}$ under Ξ_2 produces the description $\pi_{\tilde{X}_{k-1},\tilde{X}_k}$ given by

$$\begin{aligned}\pi_{\tilde{X}_{k-1},\tilde{X}_k}(\boldsymbol{x}_{k-1},\boldsymbol{x}_k) &= \sup_{\substack{\boldsymbol{n}_{k-1}\in\mathbb{N}:\\ 0=\Xi_2(\boldsymbol{n}_{k-1},\boldsymbol{x}_{k-1},\boldsymbol{x}_k)}} \pi_{\tilde{N}}\left(\boldsymbol{n}_{k-1}\right) \\ &= \sup_{\substack{\boldsymbol{n}_{k-1}\in\mathbb{N}:\\ \boldsymbol{x}_k=f(\boldsymbol{x}_k,\boldsymbol{u}_{k-1},\boldsymbol{n}_{k-1})}} \pi_{\tilde{N}}\left(\boldsymbol{n}_{k-1}\right)\end{aligned} \tag{6.20}$$

   for all $\boldsymbol{x}_{k-1},\boldsymbol{x}_k \in \mathbb{X}$.

3. For the conjunction with $\pi^{(1)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ at node I, this description, too, must be extended onto $(\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k)$, producing the vacuous extension $\pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ given by

$$\pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_{k-1},\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) = \pi_{\tilde{X}_{k-1},\tilde{X}_k}(\boldsymbol{x}_{k-1},\boldsymbol{x}_k) \tag{6.21}$$

   for all $\boldsymbol{x}_{k-1},\ldots,\boldsymbol{x}_k \in \mathbb{X}$ and all $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1} \in \mathbb{Y}$.

Now, both $\pi^{(1)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ and $\pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ describe $(\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k)$ and can be combined.

4. For the conjunction of $\pi^{(1)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ and $\pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$, Lemma 22 must generally be applied, which implies a combination under the UI-Π-copula, i.e. $\mathcal{J}^{(1,2)} = \mathcal{J}^{\mathrm{UI}}$ in

$$\pi^{(1,2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k} = \mathcal{J}^{(1,2)}\left(\pi^{(1)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}, \pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}\right). \tag{6.22}$$

   However, if $\pi_{\tilde{N}}$ is (quasi-)vacuous—which has been argued to often be applicable –, then so are $\pi_{\tilde{X}_{k-1},\tilde{X}_k}$ and $\pi^{(2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$. In this case, Lemma 20 allows the application of the more specific NI-Π-copula, i.e. $\mathcal{J}^{(1,2)} = \mathcal{J}^{\mathrm{NI}}$.

For the re-combination of $0 = (\Xi_1,\Xi_2)$ and $0 = \Xi_3$ at node II, again, a shared sample space is sought.

5. Since Ξ_3 does not depend on $\tilde{X}_{k-1}$, which is also not required in the final description $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$, one may marginalize this dependency from $\pi^{(1,2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$, i.e.

$$\pi^{(1,2)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) = \sup_{\boldsymbol{x}_{k-1}\in\mathbb{X}} \pi^{(1,2)}_{\tilde{X}_{k-1},\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_{k-1},\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) \tag{6.23}$$

   for all $\boldsymbol{x}_k \in \mathbb{X}$ and all $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k \in \mathbb{Y}$.

6. The (inverse) extension of $\pi_{\tilde{W}_k}$ under Ξ_3 produces $\pi_{\tilde{X}_k,\tilde{Y}_k}$ given by

$$\begin{aligned}\pi_{\tilde{X}_k,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_k) &= \sup_{\substack{\boldsymbol{w}_k\in\mathbb{W},:\\ 0=\Xi_3(\boldsymbol{w}_k,\boldsymbol{x}_k,\boldsymbol{y}_k)}} \pi_{\tilde{W}}(\boldsymbol{w}_k) \\ &= \sup_{\substack{\boldsymbol{w}_k\in\mathbb{W},:\\ \boldsymbol{y}_k=g(\boldsymbol{x}_k)+\boldsymbol{w}_k}} \pi_{\tilde{W}}(\boldsymbol{w}_k) = \pi_{\tilde{W}_k}(\boldsymbol{y}_k - g(\boldsymbol{x}_k))\end{aligned} \tag{6.24}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$ and all $\boldsymbol{y}_k \in \mathbb{Y}$.

7. This inverse extension is also extended onto $(\tilde{X}_k, \tilde{Y}_1, \ldots, \tilde{Y}_k)$, producing the vacuous extension $\pi^{(3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ given by

$$\pi^{(3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k) = \pi_{\tilde{X}_k,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_k) \tag{6.25}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$ and all $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k \in \mathbb{Y}$.

8. Finally, the desired possibilistic description

$$\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k} = \pi^{(1,2,3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k} = \mathcal{J}^{\mathrm{UI}}\left(\pi^{(1,2)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}, \pi^{(3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}\right) \tag{6.26}$$

is obtained under the conjunction of $\pi^{(12)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ and $\pi^{(3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ at node II. Here, Lemma 22 must be applied, because neither $\pi^{(12)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ nor $\pi^{(3)}_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ warrant the application of a tighter Π-copula under Lemmas 18, 20 or 21.

From this point on, the procedure is the same as for the derivation of the batch filter formulation in Section 6.1.2.1.

9. Define the Π-PM $\kappa_{\tilde{X}_k|\tilde{Y}_1,\ldots,\tilde{Y}_k}$ via the prediction distributions obtained under the Semi-Pivotal Step

$$\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}(\boldsymbol{x}_k) = \pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\boldsymbol{x}_k,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k), \tag{6.27}$$

given the input and output sequences $\boldsymbol{u}_0, \ldots, \boldsymbol{u}_{k-1}$ and $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k$, respectively.

The final possibility distribution $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ describes the statistical model in Eq. (6.7) in the same way that the batch description in Eq. (6.11) does. In particular, $\kappa_{\tilde{X}_k|\tilde{Y}_1,\ldots,\tilde{Y}_k}$ is guaranteed to be perceptive. Nevertheless, they may look very different when fully evaluated. The reason for this is the non-uniqueness of possibilistic descriptions of an experiment, which is, e.g., indicated by the virtually arbitrary choice of the plausibility function in the IP-Π-transform. Here, the difference stems from the step-wise application of the various Π-copulae instead of the batch application in the previous section, which are both admissible as the discussion of Eq. (3.93) has revealed earlier.

Fundamentally, the prediction distribution $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}$ depends on the previous description $\pi_{\tilde{X}_{k-1},\tilde{Y}_1,\ldots,\tilde{Y}_{k-1}}$ only through $\kappa_{\tilde{X}_{k-1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}$, $\pi_{\tilde{N}}$ and $\pi_{\tilde{W}}$. In particular, no other values of $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_{k-1}}(\cdot,\boldsymbol{y}'_1,\ldots,\boldsymbol{y}'_{k-1})$ or $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}(\cdot,\boldsymbol{y}'_1,\ldots,\boldsymbol{y}'_k)$, where $\boldsymbol{y}'_i \neq \boldsymbol{y}_i$ for any $i = 1,\ldots,k$, are required throughout Steps 1-9, even though the (technical) vacuous extensions may appear to suggest otherwise. The measurement sequence $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k$ is fixed and may simply be carried along from one time step to the next, making the distinction between the description $\pi_{\tilde{X}_k,\tilde{Y}_1,\ldots,\tilde{Y}_k}$ and the prediction distribution $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}$ in this particular instance, to a certain extent, negligible.

An implementation of the recursive procedure outlined above would be characterized by the repeated evaluation of Steps 1-9. The fundamental question is, what has been gained from this decomposition. After all, both the batch and the recursive Π-PMs describe the same information, just in slightly different ways. The answer to this question lies in the computability of the two expressions. More precisely, the two possibilistic descriptions $\pi^{(2)}_{\tilde{Q}}$ and, in particular, $\pi^{(3)}_{\tilde{Q}}$ are much easier to compute than the first description $\pi^{(1)}_{\tilde{Q}}$. The former constitute basic extension problems with a fixed number of involved variables; the latter, however, is, upon first sight, the same formulation as Eq. (6.13), only for $k = k-1$, and can, therefore, be assumed to be available from the previous time step. This is crucial for the following recursive formulation.

The above, step-wise description helps demonstrate the feasibility of recursive filters; however, a much simpler formulation can be found, which is more in line with standard formulations, such as the Kalman filter or more advanced filtering techniques [Tangirala14]. More precisely, the recursive formulation can be summarized as an iterative succession of (initialization,) predictions, inversions and updates as depicted in Figure 6.2.

Initialization By Eq. (6.18), the filter is initialized with

$$\kappa_{\tilde{X}_0|\emptyset}(\boldsymbol{x}_0) = \pi_{\tilde{X}_0}(\boldsymbol{x}_0) \tag{6.28}$$

for all $\boldsymbol{x}_0 \in \mathbb{X}$ in the *initial step*. No measurements have been obtained and, in the absence of other information, the prediction distribiution of $\tilde{X}_0$ coincides with its description distribution.

Prediction Summarizing Steps 1-5, the *prediction step* projects the old information about the previous state $\tilde{X}_{k-1}$ onto the current state yielding the intermediate membership function $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}$ that can be interpreted as a prediction distribution of $\tilde{X}_k$ given only the past measurements $\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}$. To this end, $\tilde{X}_{k-1}$ is propagated according to the system dynamics taking the effect of the process noise $\tilde{N}_{k-1}$, which potentially affects the predicted state, into account, but not yet incorporating the information obtained from the

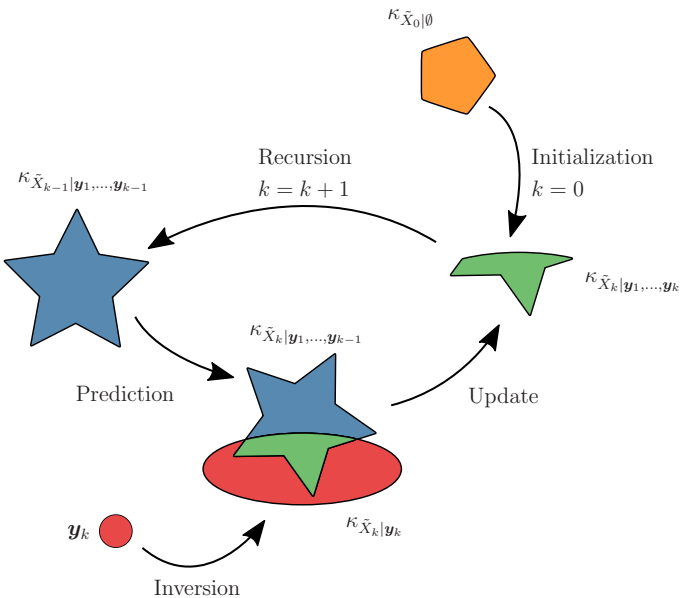


Figure 6.2: Sequence of the recursive filtering process.

next measurement $\boldsymbol{y}_k$. That is, $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}$ is given by

$$\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}(\boldsymbol{x}_k) = \sup_{\substack{\boldsymbol{x}_{k-1}\in\mathbb{X},\\ \boldsymbol{n}_{k-1}\in\mathbb{N}:\\ \boldsymbol{x}_k=f(\boldsymbol{x}_{k-1},\boldsymbol{u}_{k-1},\boldsymbol{n}_{k-1})}} \mathcal{J}^{(1,2)}\left(\kappa_{\tilde{X}_{k-1}|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}(\boldsymbol{x}_{k-1}), \pi_{\tilde{N}}(\boldsymbol{n}_{k-1})\right) \tag{6.29}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$. This expression is a basic propagation of a combination of two input membership functions.

Inversion Step 6 constitutes the *inversion step*, and is concerned with projecting the information obtained from the current measurement $\boldsymbol{y}_k$ onto the state space in order to derive the intermediate membership function $\kappa_{\tilde{X}_k|\boldsymbol{y}_k}$ that can be interpreted as a prediction distribution of $\tilde{X}_k$ given only the current measurement, and is given by

$$\kappa_{\tilde{X}_k|\boldsymbol{y}_k}(\boldsymbol{x}_k) = \pi_{\tilde{W}}(\boldsymbol{y}_k - g(\boldsymbol{x}_k)) \tag{6.30}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$, respectively. This expression is a basic membership inversion.

Update Finally, the *updating* or *correction step* summarizes Steps 7-9 by combining both projections obtained under the prediction and inversion step. The resulting prediction

distribution $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}$ accounting for both the past and current measurements implied by Steps 1-9 can then be expressed as

$$\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_k}(\boldsymbol{x}_k) = \mathcal{J}^{\mathrm{UI}}\left(\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}(\boldsymbol{x}_k), \kappa_{\tilde{X}_k|\boldsymbol{y}_k}(\boldsymbol{x}_k)\right) \tag{6.31}$$

for all $\boldsymbol{x}_k \in \mathbb{X}$. This expression is nothing more than combining the two intermediate prediction functions.

Equations (6.28), (6.29), (6.30) and (6.31) constitute the four basic steps of the recursive possibilistic filter, and any implementation must provide details on their computational evaluation.

6.2 Implementation

In the following, a practical strategy for implementing a possibilistic filter is discussed.

Hose and Hanss [HoseHanss21b] arrive at the same recursive formulation treating the current state as an unknown parameter, which can be explained with the quasi-non-existent distinction between different types of fuzzy variables and membership functions in the calculus of possibility theory. Therefore, Hose and Hanss' idea of a particle-based implementation can be adopted in a one-to-one fashion.

6.2.1 Particle Filter

The recursive filtering formulation enables an implementation based on μ-clusters and sample-based membership arithmetic. Therefore, in the following, all of the involved (intermediate) prediction distributions are assumed to be μ-functions induced by μ-clusters. For instance, the (intermediate) prediction distribution of $\tilde{X}_k$ given the measurements $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{k-1}$ is assumed to be described by the μ-cluster $\mathcal{K}_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}$ on $\mathbb{X}$, which is, for convenience, simply written as $\mathcal{X}_{k|1,\ldots,k-1} = \mathcal{K}_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}}$. That is, $\kappa_{\tilde{X}_k|\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{k-1}} = \mu_{\mathcal{X}_{k|1,\ldots,k-1}}$.

The μ-covers in all such μ-clusters, e.g. $(\boldsymbol{x}^{(i)}_{k|1,\ldots,k-1}, \alpha^{(i)}_{k|1,\ldots,k-1}) \in \mathcal{X}_{k|1,\ldots,k-1}$, shall be referred to as *particles*, and their number $m = |\mathcal{X}_{k|1,\ldots,k-1}| = |\mathcal{X}_{k|k}| = |\mathcal{X}_{k|1,\ldots,k}|$ shall be kept constant, i.e. $i = 1, \ldots, m$.

6.2.1.1 Initialization

In the initialization step, the initial μ-cluster $\mathcal{X}_0$ is constructed from Eq. (6.28). Following the suggestions in Section 5.2.4.1, regular, random or structured sampling strategies may be employed for this purpose. In practice, especially random sampling approaches, such

as Latin hypercube sampling and Sobol sampling, exhibit advantageous properties—not least because they avoid systematic sampling errors. More precisely, the initial μ-cluster

$$\mathcal{X}_0 = \left\{ \left(\boldsymbol{x}_0^{(i)}, \alpha_0^{(i)} \right) : i = 1, \ldots, m \right\} \tag{6.32}$$

is composed of the m Latin hypercube samples $\boldsymbol{x}_0^{(1)}, \ldots, \boldsymbol{x}_0^{(m)}$ on $[\mathbb{X}_0]$ constituting the particle positions and of the corresponding particle memberships computed from Eq. (6.28), i.e.

$$\alpha_0^{(i)} = \pi_{\tilde{X}_0} \left(\boldsymbol{x}_0^{(i)} \right) \tag{6.33}$$

for $i = 1, \ldots, m$. Notice that, despite the general applicability of this assumption, this implementation does not rely, e.g., on the (quasi-)vacuousness of $\pi_{\tilde{X}_0}$ but, instead, works for arbitrary possibility distributions.

Subsequently, the prediction, inversion and updating step must be performed recursively for every subsequent time step. The particles $\mathcal{X}_{k-1|1,\ldots,k-1}$ can be assumed to be available from the previous time step $k-1$ for $k > 1$, or by $\mathcal{X}_{0|\emptyset} = \mathcal{X}_0$ for $k = 1$.

6.2.1.2 Prediction

In the prediction step, the previous μ-cluster $\mathcal{X}_{k-1|1,\ldots,k-1}$ is propagated under the system dynamics f given the previous input $\boldsymbol{u}_{k-1}$ according to Eq. (6.29).

First and foremost, this also requires a μ-cluster representation $\mathcal{K}_{\tilde{N}}$ of the process noise $\tilde{N}$, more precisely of $\pi_{\tilde{N}}$. This description can be obtained in the same manner as in the above initialization step. That is,

$$\mathcal{K}_{\tilde{N}} = \left\{ \left(\boldsymbol{n}_{k-1}^{(i)}, \alpha_{k-1|\emptyset}^{(i)} \right) : i = 1, \ldots, m \right\} \tag{6.34}$$

is composed of the m Latin hypercube samples $\boldsymbol{n}_{k-1}^{(1)}, \ldots, \boldsymbol{n}_{k-1}^{(m)}$ on $[\mathbb{N}]$ and of the corresponding μ-levels

$$\alpha_{k-1|\emptyset}^{(i)} = \pi_{\tilde{N}} \left(\boldsymbol{n}_{k-1}^{(i)} \right) \tag{6.35}$$

for $i = 1, \ldots, m$. This μ-cluster should be constructed anew in every time step in order to avoid systematic errors.

In principle, one would, then, have to compute the combination of $\mathcal{X}_{k-1|1,\ldots,k-1}$ and $\mathcal{K}_{\tilde{N}}$ under the appropriate Π-copula $\mathcal{J}^{(1,2)}$ under Eq. (5.38) leading to a total of m^2 μ-nodes in the resulting μ-cluster. In order to keep the number of particles fixed to m, one would then have to select m out of these m^2 μ-nodes to proceed with. However, as long as $\mathcal{K}_{\tilde{N}}$ is obtained via random sampling and the selection is performed randomly (with equal sampling weights), it is equivalent to consider the joint μ-cluster

$$\mathcal{K}_{\tilde{X}_{k-1},\tilde{N}} = \left\{ \left(\left(\boldsymbol{x}_{k-1|1,\ldots,k-1}^{(i)}, \boldsymbol{n}_{k-1}^{(i)} \right), \mathcal{J}^{(1,2)} \left(\alpha_{k-1|1,\ldots,k-1}^{(i)}, \alpha_{k-1|\emptyset}^{(i)} \right) \right) : i = 1, \ldots, m \right\} \tag{6.36}$$

consisting of m μ-nodes only.

Finally, this joint μ-cluster is propagated according to Eqs. (6.29). By Eq. (5.58), this results in the prediction μ-cluster

$$\mathcal{X}_{k|1,\ldots,k-1} = \left\{ \left(\boldsymbol{x}^{(i)}_{k|1,\ldots,k-1}, \alpha^{(i)}_{k|1,\ldots,k-1} \right) : i = 1, \ldots, m \right\}, \tag{6.37}$$

where the new particle positions are given by

$$\boldsymbol{x}^{(i)}_{k|1,\ldots,k-1} = f\left(\boldsymbol{x}^{(i)}_{k-1|1,\ldots,k-1}, \boldsymbol{u}_{k-1}, \boldsymbol{n}^{(i)}_{k-1} \right), \tag{6.38}$$

and the corresponding particle memberships are given by

$$\alpha^{(i)}_{k|1,\ldots,k-1} = \mathcal{J}^{(1,2)}\left(\alpha^{(i)}_{k-1|1,\ldots,k-1}, \alpha^{(i)}_{k-1|\emptyset} \right) \tag{6.39}$$

for $i = 1, \ldots, m$, respectively. Again, despite the general applicability of this assumption, this implementation does not rely, e.g., on the (quasi-)vacuousness of $\pi_{\tilde{N}}$, which would, however, lead to the admissibility of the NI-Π-copula $\mathcal{J}^{(1,2)} = \mathcal{J}^{\mathrm{NI}}$.

6.2.1.3 Inversion

The inversion step is straightforward to implement. For the computation of the inversion μ-cluster

$$\mathcal{X}_{k|k} = \left\{ \left(\boldsymbol{x}^{(i)}_{k|k}, \alpha^{(i)}_{k|k} \right) : i = 1, \ldots, m \right\}, \tag{6.40}$$

it is expedient to make use of the already computed particle positions from the prediction step, and let

$$\boldsymbol{x}^{(i)}_{k|k} = \boldsymbol{x}^{(i)}_{k|1,\ldots,k-1} \tag{6.41}$$

for $i = 1, \ldots, m$. Furthermore, evaluating Eq. (6.30) at theses particle locations directly yields the corresponding particle memberships

$$\alpha^{(i)}_{k|k} = \pi_{\tilde{W}}\left(\boldsymbol{y}_k - g\left(\boldsymbol{x}^{(i)}_{k|k} \right) \right) \tag{6.42}$$

for $i = 1, \ldots, m$.

6.2.1.4 Update

Since the particle positions obtained in the prediction and inversion step are the same by construction, the updating step is implemented with little effort. The prediction and inversion μ-clusters are combined into the updated μ-cluster

$$\mathcal{X}_{k|1,\ldots,k} = \left\{ \left(\boldsymbol{x}^{(i)}_{k|1,\ldots,k}, \alpha^{(i)}_{k|1,\ldots,k} \right) : i = 1, \ldots, m \right\} \tag{6.43}$$

under the UI-Π-copula according to Eq. (6.31) yielding the same particle positions

$$\boldsymbol{x}^{(i)}_{k|1,\ldots,k} = \boldsymbol{x}^{(i)}_{k|1,\ldots,k-1} = \boldsymbol{x}^{(i)}_{k|k} \tag{6.44}$$

and the updated particle memberships

$$\alpha^{(i)}_{k|1,\dots,k} = \mathcal{J}^{\mathrm{UI}}\left(\alpha^{(i)}_{k|1,\dots,k-1}, \alpha^{(i)}_{k|k}\right) \tag{6.45}$$

for $i = 1, \dots, m$, respectively.

This constitutes a basic implementation of the recursive possibilistic filter.

6.2.1.5 Resampling

After the membership correction in the updating step, many particles with poor measurement agreement indicated by low inversion memberships will exhibit updated memberships that are also (close to) zero. Such particles bear little information—not least because their membership can be expected to remain small throughout the remainder of the filtering process. Therefore, they are, at most, helpful in bounding the support of the prediction distribution of $\tilde{X}$. Conversely, if the inversion memberships are very high (close to one) due to good agreement with the present measurement, the updated membership of such particles will increase compared to the previous and predicted membership. Such particles bear much information as they constitute the prediction regions of $\tilde{X}$.

It is, therefore, recommendable to include a resampling step, in which the former particles are eliminated from the μ-cluster, and the latter are kept. One would, however, be ill-advised to only care about the particles with maximal membership; instead, equal emphasis should be put on all particles with non-zero membership.

A basic, but—from experience—very effective, resampling procedure that shall be used below is given by a weighted sampling from all particles with their respective sampling weights

$$\lambda^{(i)} = \begin{cases} 1 & \text{if } \alpha^{(i)}_{k|1,\dots,k} > \varepsilon_1 \text{ and} \\ \lambda_0 & \text{otherwise.} \end{cases} \tag{6.46}$$

for $i = 1, \dots, m$, more precisely of their normalized values. That is, all particles whose memberships exceed a certain threshold $\varepsilon_1 \ll 1$ are assigned a high resampling weight. The smaller weight $\lambda_0 \ll 1$ assigned to the remaining particles is included for technical reasons; otherwise, the resampling procedure might fail if no particles with high membership exist, which can happen especially in the first few time steps of the filtering process.

The flow chart of the recursive possibilistic particle filter (RPPF) containing the fundamental equations for its implementation in the correct order is also visualized in Fig. 6.3.

Therein, $\texttt{lhs}([\boldsymbol{a}], m)$ is assumed to generate m Latin hypercube samples on $[\boldsymbol{a}] \in \mathbb{I}(\mathbb{R}^D)$. This could, however, be replaced with a call, e.g., to Sobol sampling. Since, at every time step, the prediction, inversion and updated particle positions are the same, they need only be computed once (in the prediction step) and are denoted by

$$\boldsymbol{x}^{(i)}_k = \boldsymbol{x}^{(i)}_{k|1,\dots,k} = \boldsymbol{x}^{(i)}_{k|1,\dots,k-1} = \boldsymbol{x}^{(i)}_{k|k} \tag{6.47}$$

for $i = 1, \ldots, m$.

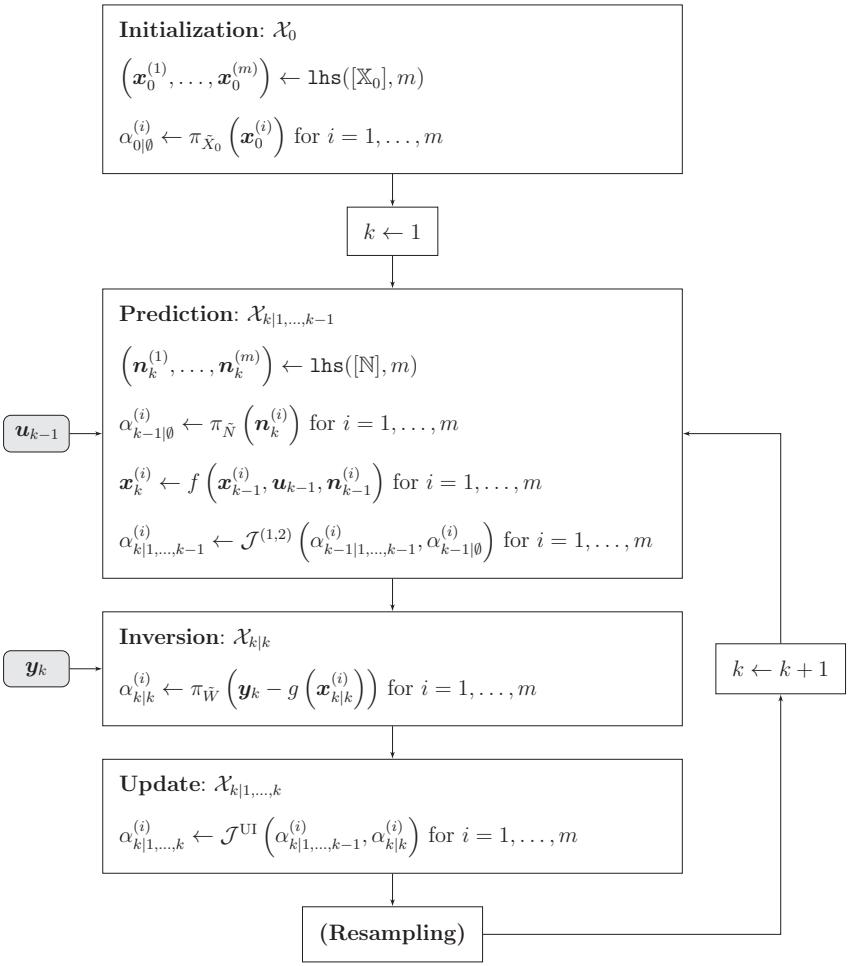


Figure 6.3: RPPF Flowchart.

6.2.2 Other Implementations

Of course, other implementations of the (recursive) possibilistic filter are conceivable, as demonstrated by earlier results of Hose and Hanss [HoseHanss21a]. In particular,

one could easily replace the μ-cluster-based implementation of the particle filter by a μ-partition-based implementation. In theory, this is achieved by exchanging the calls to `lhs` for a call to `fsivia` and replacing all standard floating-point operations of the RPPF, e.g. in the flow chart in Fig. 6.3, by their interval-arithmetical counterparts—if they are available. In any case, the excellent results obtained with the RPPF cast doubt on the actual need for such interval-based implementations—except for the obvious use case in safety-critical systems, where a robust approach is undoubtedly preferable. Nevertheless, experimental studies have shown that the required computational effort is far too large to achieve the desirable real-time applicability, which is fundamental in filtering applications, combined with insightful results in the near future. Its discussion is, therefore, left to future scholars.

In the following, the RPPF shall be applied to an examplary dynamical system.

6.3 Application: Robot Localization

A basic application of filters can be found in robotics. In order to successfully navigate its environment, every robot needs to be aware of its pose, i.e. its position and orientation. This is commonly referred to as 'localization', and various techniques to achieve this goal, often combined with mapping the unknown environment [Thrun07], are available. Here, a landmark-based localization technique is proposed, following the general setup put forth by Huang and Dissanayake [HuangDissanayake99]. The preceding chapters' notation is slightly over-written to conform to standard notation in robotics.

Apart from this example, the RPPF has successfully been applied to a two-state batch reaction already [HoseHanss21b].

6.3.1 Setup

To begin, consider Fig. 6.4. The depicted differential drive robot is to be located with respect to the inertial coordinate system $\{0, x^0, y^0, z^0\}$. More precisely, both the robots position $\boldsymbol{r}^{\text{robot}} = (x^{\text{robot}}, y^{\text{robot}}, z^{\text{robot}})$ and its heading ϕ^{robot} need to be inferred, where the latter is simply the rotation (about the z-axis) of the relative coordinate system $\{0, x^{\text{r}}, y^{\text{r}}, z^{\text{r}}\}$ whose x-axis points in the robots' direction of travel $\boldsymbol{v}^{\text{robot}}$. The problem is considered to be planar, i.e. all z-coordinates are zero, and there is no rolling about the x^{r}-axis or pitching about the y^{r}-axis—only yawing about the z-axis.

Assuming no-slip conditions, the (forward Euler) discretized system dynamics f of the

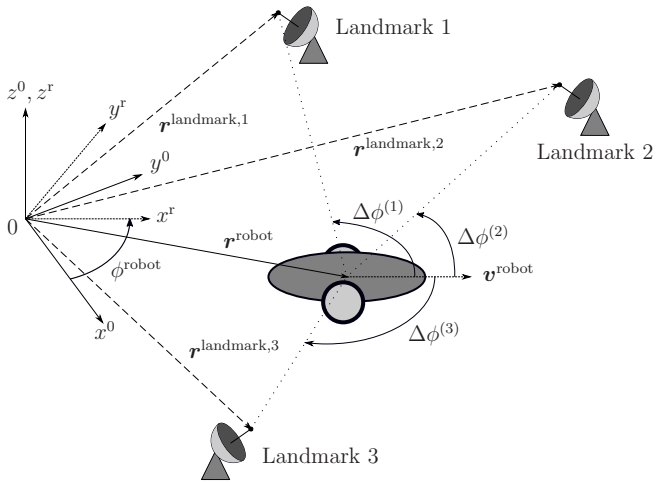


Figure 6.4: Localization Setup.

differential drive robot with pose $\boldsymbol{x}_k = (x_k^{\mathrm{robot}}, y_k^{\mathrm{robot}}, \phi_k^{\mathrm{robot}})$ are given by

$$\begin{aligned} x_k^{\mathrm{robot}} &= x_{k-1}^{\mathrm{robot}} + \Delta t \left(v_{k-1} + \delta v_{k-1}\right) \cos\left(\phi_{k-1}^{\mathrm{robot}}\right), \\ y_k^{\mathrm{robot}} &= y_{k-1}^{\mathrm{robot}} + \Delta t \left(v_{k-1} + \delta v_{k-1}\right) \sin\left(\phi_{k-1}^{\mathrm{robot}}\right) \text{ and} \\ \phi_k^{\mathrm{robot}} &= \phi_{k-1}^{\mathrm{robot}} + \Delta t \left(\omega_{k-1} + \delta\omega_{k-1}\right), \end{aligned} \tag{6.48}$$

where Δt is the time step width. The forward velocity v_k and the angular velocity ω_k are gathered in the contol input $\boldsymbol{u}_k = (v_k, \omega_k)$, and their respective disturbances are gathered in the process noise $\boldsymbol{n}_k = (\delta v_k, \delta\omega_k)$.

The localization is to be achieved via the L landmarks $1, \ldots, L$ with known position $\boldsymbol{r}^{\mathrm{landmark},j} = (x^{\mathrm{landmark},j}, y^{\mathrm{landmark},j}, z^{\mathrm{landmark},j})$ to which the relative orientation—but not the distance—can be measured. This is often applicable, when the sensors mounted on the robot in order to detect these landmarks are simple cameras instead of, e.g., more expensive radar systems. Then, the relative orientation is given by the corresponding measurement model g, which reads

$$\Delta\phi_k^{(j)} = \mathrm{mod}\left(\arctan\left(\frac{y^{\mathrm{landmark},j} - y_k^{\mathrm{robot}}}{x^{\mathrm{landmark},j} - x_k^{\mathrm{robot}}}\right) - \phi_k^{\mathrm{robot}} + \delta\phi_k^{(j)}, 2\pi\right) \tag{6.49}$$

for $j = 1, \ldots, L$, where $w_k = (\delta\phi_k^{(1)}, \ldots, \delta\phi_k^{(L)})$ constitutes the vector of all measurement errors. The modulo function mod ensures that the measured angle is always in $[0, 2\pi)$.

This application certainly favors a mostly deterministic dynamic model because robotic systems generally exhibit a high degree of reproducibility without any significant variation—

except for the measurement error—rendering random dynamic effects implausible. Typically, the process noise is an expression of missing insight into the actual system dynamics rather than of actual random behavior of the system. Still, the process noise is usually assumed to follow a Gaussian probability distribution in standard localization techniques. Avoiding such a precise description, it is arguably more suitable to model the process noise by a (quasi-)vacuous possibility distribution $\tilde{N} \sim \mathcal{Q}([\mathbb{N}])$ on a given support $[\mathbb{N}]$. This could be interpreted both as an expression of the effect of the explicit Euler discretization of the continuous-time dynamics and of the mismatch between the given and the actual input that may, e.g., stem from unmodeled second-order system dynamics. Due to this choice, one can set $\mathcal{J}^{(1,2)} = \mathcal{J}^{\mathrm{NI}}$ to be the NI-Π-copula.

Similarly, the initial value $\tilde{X}_0 \sim \mathcal{Q}([\mathbb{X}_0])$ is modeled by a (quasi-)vacuous possibility distribution on a given support $[\mathbb{X}_0]$, which may be wider or smaller—depending on the precision with which one can specify the starting pose of the robot.

Finally, the measurement error is usually well-described by an uncorrelated, zero-mean multivariate Gaussian distribution $\mathcal{N}(0, \boldsymbol{R})$ with the diagonal covariance matrix

$$\boldsymbol{R} = s^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.50}$$

for the standard deviation $s > 0$. A consistent possibility distribution $\pi_{\tilde{W}}$ is given by the Complementary Cumulative Probability-to-Possibility Transform of Eq. (3.19), i.e.

$$\pi_{\tilde{W}}(\boldsymbol{w}_k) = 1 - F_{\chi^2(L)}\left(\boldsymbol{w}_k^{\mathrm{T}} \cdot \boldsymbol{R}^{-1} \boldsymbol{w}_k\right) = 1 - F_{\chi^2(L)}\left(\sum_{j=1}^{L} \left(\frac{\delta\phi_k^{(j)}}{s}\right)^2\right) \tag{6.51}$$

for all $\boldsymbol{w}_k \in \mathbb{R}^L$, where $F_{\chi^2(L)}$ is the CPF of the χ^2-distribution with L degrees of freedom.

6.3.2 Simulation

In the following, a simulation with the remaining specifications given in Table 6.1 is discussed. Notice that the supports of the initial values and the process noise are chosen to be very wide in order to model severe uncertainty about the dynamic model. The initial value distribution, for instance, is vacuous with respect to the initial bearing. The measurement error with a standard deviation of five degrees, too, implies a rather rudimentary goniometry.

In order to achieve an eight-shaped trajectory, the forward and angular velocity are chosen as

$$v_k = 2 + \sin\left(\frac{2\pi k}{25}\right) \qquad \text{and} \qquad \omega_k = \begin{cases} \frac{\pi}{25} & \text{if } k \bmod 100 < 50 \text{ and} \\ -\frac{\pi}{25} & \text{otherwise} \end{cases} \tag{6.52}$$

Table 6.1: Technical Specifications of Localization Simulation.

Time Step Width	$\Delta t = 1$
True Initial Position	$\boldsymbol{r}_0^{\text{robot}} = (0, 0, 0)$
True Initial Orientation	$\phi_0^{\text{robot}} = \frac{\pi}{4}$
Landmark Position 1	$\boldsymbol{r}^{\text{landmark},1} = (-20, -20, 0)$
Landmark Position 2	$\boldsymbol{r}^{\text{landmark},2} = (0, 10, 0)$
Landmark Position 3	$\boldsymbol{r}^{\text{landmark},3} = (20, 0, 0)$
Initial Value Support	$[\mathbb{X}_0] = [-1, +1] \times [-1, +1] \times [-\pi, +\pi]$
Process Noise Support	$[\mathbb{N}] = [-\frac{1}{2}, +\frac{1}{2}] \times [-\frac{\pi}{36}, +\frac{\pi}{36}]$
Measurement Error Standard Deviation	$s = \frac{\pi}{36}$
Particle Number	$m = 2 \cdot 10^3$
Resampling Weight	$\lambda_0 = 10^{-3}$
Resampling Tolerance	$\varepsilon_1 = 10^{-3}$

for all $k \geq 0$. A realization of this trajectory, shown in Figure 6.5, and the corresponding output signals (additionally perturbed by the measurement error) are shown in Figure 6.6. For this simulation, the system dynamics are perturbed by the systematic error $n_k = (-\frac{1}{10}, \frac{\pi}{180})$ for all $k \geq 0$ in order to showcase the ability of the RPPF to deal with such extreme cases of imprecision; any Gaussian probability distribution would undoubtedly be ill-suited to model this kind of process noise, but as long as the actual realizations of the process noise reside in $[\mathbb{N}]$, the RPPF will be perceptive. The unperturped reference trajectory is also shown in order to demonstrate the detrimental effect this noise has on the trajectory. Finally, the measurement error is simulated from a Gaussian distribution $\mathcal{N}(0, \boldsymbol{R})$ with the assumed covariance matrix.

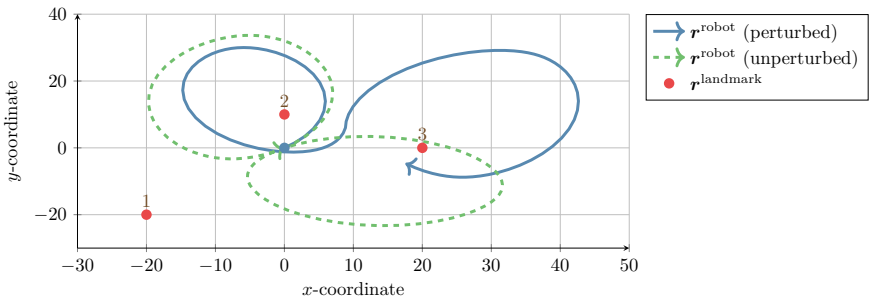


Figure 6.5: Robot Trajectory.

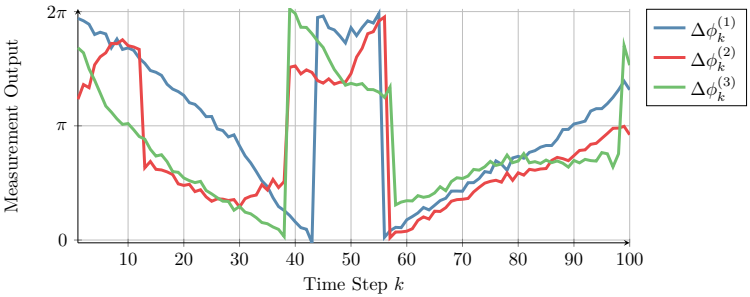


Figure 6.6: Output Signal.

Finally, the RPPF is applied to this in- and output sequence.

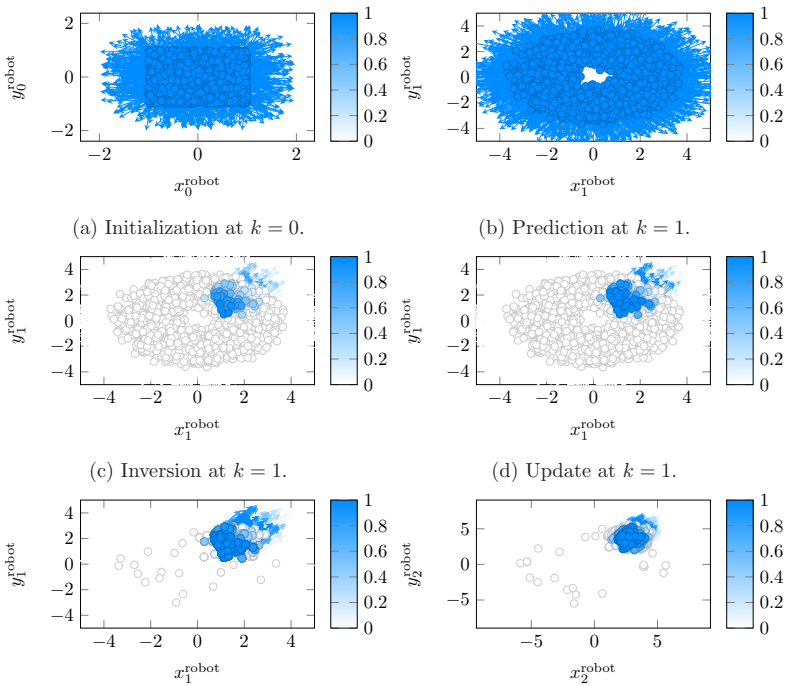
The results of the initialization at $k = 0$ and the prediction, inversion, updating and resampling step at $k = 1$ and $k = 2$ are exemplarily shown in Fig. 6.7. The x- and y-coordinates of the particles indicate the robot position, and the arrows indicate the robot bearing. The particle membership is encoded by its color.

The RPPF behaves as expected: In particular, the inversion step identifies those particles resulting from the prediction of the initial Latin hypercube samples that produce outputs commensurate with the first measurement. The updating step combines the two obtained memberships and assigns higher memberships to these particles and quasi-zero memberships to all other particles with poor measurement agreement. Finally, the resampling step selects the particles with a high updated membership for re-use in the following prediction.

The full RPPF iteration up to $K = 100$ at ten selected time steps is visualized in Fig. 6.8. The RPPF can extract the information about the robot's current pose that is contained in the statistical model and the in- and output sequence, and it can give reliable estimates of the current robot position. In particular, regions with high membership typically contain the true system state but, the farther the robot moves away from the landmarks, the worse the estimates become because, there, imprecisions in angular measurements have a more detrimental effect, and vice versa.

For reference, the—in practice, unknown—true pose of the robot shown in Fig. 6.5 is also evaluated, i.e. it is added to the μ-clusters as the $(m + 1)$-th particle. Its memberships are shown in Fig. 6.9. The true system pose does not always exhibit high memberships, insteady they are rather evenly distributed on $[0, 1]$ with a tendency towards higher values as indicated by the empirical CPF shown in Fig. 6.10, which must—due to the guaranteed perceptiveness of the RPPF—be approximately (super-)uniform. The exhibited 'slack' indicates a certain degree of conservatism in this approach, i.e. a lack of efficiency and may be explained by the very general methods employed for constructing joint possibility

distributions and the conservative uncertainty descriptions. Nevertheless, it successfully proves the general feasibility and, more importantly, the applicability of the RPPF.



(a) Initialization at $k = 0$.

(b) Prediction at $k = 1$.

(c) Inversion at $k = 1$.

(d) Update at $k = 1$.

(e) Resampling at $k = 1$.

(f) Prediction at $k = 2$.

Figure 6.7: Basic filtering steps at the beginning of the localization simulation.

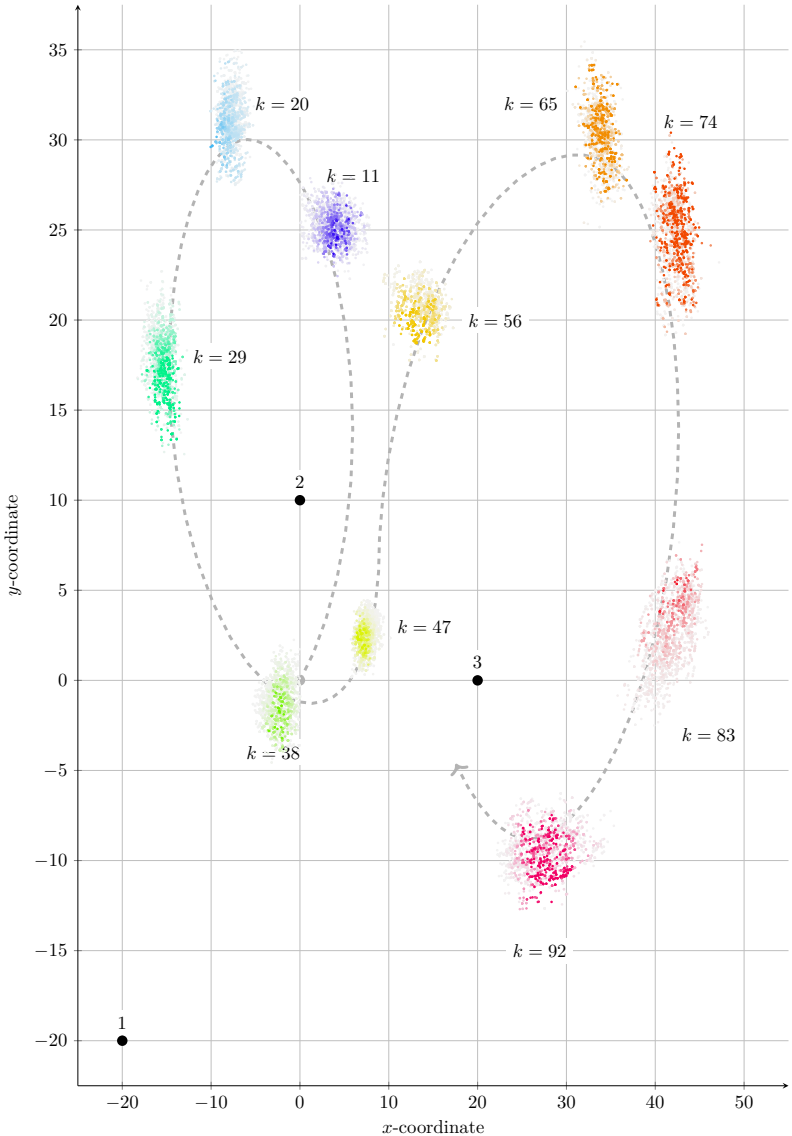


Figure 6.8: Particle Positions in RPPF.

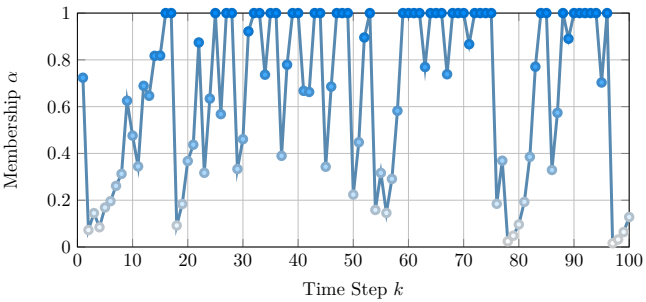


Figure 6.9: Memberships of True System Pose.

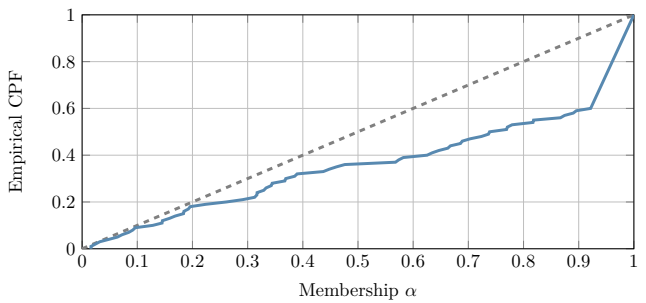


Figure 6.10: Empirical Membership CPF of True System Pose.

Chapter 7

Conclusion and Outlook

> The whole future lies in uncertainty: live immediately.
>
> Seneca, *On the Shortness of Life*

To summarize this thesis, possibility theory is a powerful and intuitive framework for reasoning with imprecise probabilities. This includes not only their description and propagation but also statistical inference in the form of Π-IMs. A possibilistic analysis corresponds to an analysis of nested random sets, or α-cuts, whose membership levels can encode various types of information, e.g. IP descriptions and data-dependent confidence or predictions. The membership function is the fundamental tool in any possibilistic analysis, and the μ-transform provides a general rule for obtaining it. Its correct manipulation also constitutes most of the calculus of possibility theory and is one of the core topics of this thesis, in particular the membership extension. The numerical implementation of these operations is straightforward and allows for a simple application to filtering problems.

Possibility theory can also be summarized differently. Considering that consistency may be understood as a general concept of stochastic dominance of CPFs and that the possibilistic calculus fundamentally depends on the computation of suprema, one can express the tenet of possibility theory as 'integrate first, then optimize'. The first part of this statement is a direct reference, e.g., to the μ-transform, which integrates (imprecise) probability distributions on the sublevel sets of the plausibility function in order to project them onto the possibility space. The second part refers to the extension principle, which is—in essence—a mathematical program where the information, in the form of data and (implicit, explicit, or inverse) relationships is written into the constraints, and the possibilistic description constitutes the objective function to be maximized.

It is conjectured that this principle also obtains in a more general form, e.g. including other types of information—a claim that is also rooted in the observation that the same

calculus applies to imprecise variables and unknown parameters alike and can, potentially, also be extended to other mathematical objects.

The goal of this thesis has been to provide a theoretical foundation of possibility theory and possibilistic statistics. Nevertheless, much remains to be explored to better understand its full potential and for it to mature into a universal framework for uncertainty quantification. Some key questions that have not been answered in this thesis, but appear to be promising directions of further research, shall briefly be outlined in the following.

1. Based on Neumaier's idea of clouds [Neumaier04], Destercke et al. observe that a p-box is essentially equivalent to the intersection of the credal sets of two possibility distributions [DesterckeDuboisChojnacki08]. This idea may be spun further when considering conjunctions of three, four, or more possibility distributions. Similar to how convex sets can be approximated with increasing precision by the intersection of more and more half-spaces, this would allow one to describe arbitrary credal sets with increasing specificity. The fundamental tool for such *possibility boxes* would be a vector-valued elementary possibility function $\boldsymbol{\pi} : \Omega \to [0,1]^D$. The possibilistic calculus would still apply on each individual possibility function $\pi_1, \ldots, \pi_D$ contained therein, but this could potentially lead to tighter bounds on upper and lower probabilities and expectations.

2. Unfortunately, the reliability level in the μ-transform cannot be interpreted as a probability of lower specificity or consistency, but numerical experiments exhibit favorable properties in this respect. By considering multinomial Π-IMs instead of individual binomial Π-IMs, it could be possible to obtain better theoretical guarantees regarding the consistency of the obtained percentage sets.

3. The RPPF provides an entry to a rich field of statistical analysis in dynamical systems. Preliminary results suggest that it is straightforward to use the information contained therein, e.g. for predicting collisions or avoiding them. This leads directly to the statistical discipline of decision-making, which this thesis has not covered. Nevertheless, the RPPF may be connected to existing possibilistic controller synthesis methods [MäckHoseHanss17, HoseMäckHanss18, HoseMäckHanss19b].

Finally, the field of possibilistic statistics offers much room for further exploration.

4. This thesis has revealed many connections between possibilistic and frequentist statistics and some similarities to fiducial and likelihoodist inference. What remains to be investigated is the link of possibilistic statistics to Bayesian inference. For instance, it seems promising to investigate whether Bayesian updating under certain types of priors is accounted for by a Π-IM.

5. Most of the results in this thesis are exact or conservative, which is certainly advantageous. Since possibilistic statistics, as presented here, is a very young field with many discoveries to be made, one should, primarily, be interested in finding more results of this kind. It is, however, conjectured that approximate methods could further simplify and speed up possibilistic inference if exact or conservative solutions prove to be problematic or difficult to obtain. To this end, surrogate models could be used to derive approximate Π-IMs, and modeling the approximation error by suitable (quasi-)vacuous possibility distributions would still allow for valid inference. A natural choice for such surrogates is Interval Predictor Models [CampiCalafioreGaratti09] and model order reduction techniques with appropriate error estimators [GrunertFehrHaasdonk20]. Alternatively, possibilistic bootstrapping methods seem feasible, or standard results, such as Wilk's theorem, could be used to approximately evaluate the μ-transform.

6. Nonparametric Π-IMs have not been investigated in this thesis, though they have been elsewhere [CellaMartin22]. One can, without difficulty, derive a Π-IM, e.g. of a distribution mean from the Markov distribution, and it is likely that one may likely be able to use other probabilisitic inequalities to obtain similar nonparametric Π-IMs. Similarly, the formulation of nonparametric Π-PMs could be achieved with some inspiration, e.g. from Hill's assumption [Hill68] and previous discussions of the 'predicting the next observation'-problem in the statistical community [Seidenfeld95].

Possibility theory is a powerful tool for uncertainty quantification, and it can be expected to be useful in many future applications. It is also a rich field with many exciting discoveries to be made and should, therefore, be pursued rigorously. Ultimately, however, it must also be formulated in a more accessible fashion in order to attract practitioners. While this goal has not yet been reached, this thesis is intended to be a step in this direction.

Appendix A

Inclusion Functions of the Membership Transform

In order to find an inclusion function of the μ-transform in Eq. (5.1) with respect to both $\boldsymbol{\theta}$ and $\boldsymbol{v}$, assume that the (strictly positive and bounded) probability density function $p_{\tilde{V}|\boldsymbol{\theta}}$ corresponding to $\mathrm{P}_{\tilde{V}|\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in \mathbb{T}$ is available, as well as its inclusion function with respect to $\boldsymbol{\theta}$. Furthermore, assume that an inclusion function of the elementary plausibility function is available, let $[\boldsymbol{\theta}] \in \mathbb{I}(\mathbb{T})$ and $[\boldsymbol{v}] \in \mathbb{I}(\mathbb{V})$, and define $[\beta] = [\rho]_{\tilde{V}|[\boldsymbol{\theta}]}([\boldsymbol{v}])$.

In order to derive an upper bound of the μ-transform, it suffices to compute an upper bound of $\mathrm{P}_{\tilde{V}|\boldsymbol{\theta}}(B^+)$, where $B^+ = \{\boldsymbol{\xi} \in \mathbb{V} : \rho_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{\xi}) \leq [\beta]^+\}$, since

$$\sup_{\boldsymbol{\theta}\in[\boldsymbol{\theta}],\boldsymbol{v}\in[\boldsymbol{v}]} \pi_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{v}) \leq \sup_{\boldsymbol{\theta}\in[\boldsymbol{\theta}]} \mathrm{P}_{\tilde{V}|\boldsymbol{\theta}}\left(B^+\right) \tag{A.1}$$

for $\boldsymbol{\theta} \in [\boldsymbol{\theta}]$. The inequality holds because the μ-transform is non-decreasing with respect to $\rho_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{v})$.

Following a reweighting-based idea for the simulation of imprecise (random) variables proposed by Fetz and Oberguggenberger [FetzOberguggenberger16], consider

$$\begin{aligned}
\mathrm{P}_{\tilde{V}|\boldsymbol{\theta}}\left(B^+\right) &= \mathfrak{E}_{\mathrm{P}_{\tilde{V}}|\boldsymbol{\theta}}\left[\mathcal{I}_{B^+}\left(\tilde{V}\right)\right] = \int_{\mathbb{V}} \mathcal{I}_{B^+}(\boldsymbol{v})\, p_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} \\
&= \int_{\mathbb{V}} \mathcal{I}_{B^+}(\boldsymbol{v})\, \frac{p_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{v})}{p_{\tilde{V}|\boldsymbol{\theta}^*}(\boldsymbol{v})} p_{\tilde{V}|\boldsymbol{\theta}^*}(\boldsymbol{v})\, \mathrm{d}\boldsymbol{v} = \mathfrak{E}_{\mathrm{P}_{\tilde{V}}|\boldsymbol{\theta}^*}\left[\mathcal{I}_{B^+}\left(\tilde{V}\right) \frac{p_{\tilde{V}|\boldsymbol{\theta}}\left(\tilde{V}\right)}{p_{\tilde{V}|\boldsymbol{\theta}^*}\left(\tilde{V}\right)}\right] \\
&\approx \frac{1}{m}\sum_{j=1}^{m} \mathcal{I}_{B^+}\left(\tilde{V}_j^*\right) \frac{p_{\tilde{V}|\boldsymbol{\theta}}\left(\tilde{V}_j^*\right)}{p_{\tilde{V}|\boldsymbol{\theta}^*}\left(\tilde{V}_j^*\right)}
\end{aligned} \tag{A.2}$$

The approximation is, again, justified by Borel's law of large numbers for the m iid realizations $\tilde{V}_1^*, \ldots, \tilde{V}_m^*$ of $\tilde{V}^* \sim \mathrm{P}_{\tilde{V}|\boldsymbol{\theta}^*}$ for all $\boldsymbol{\theta}^* \in \mathbb{T}$, e.g. for $\boldsymbol{\theta}^* = [\boldsymbol{\theta}]^{\mathrm{c}}$. This expression

depends on $\boldsymbol{\theta}$ only through $p_{\tilde{V}|\boldsymbol{\theta}}(\tilde{V}_j^*)$, which can be replaced by the corresponding inclusion function and box, ultimately, yielding

$$\left[I_m^{\mathrm{MC}}\right]^+([\boldsymbol{v}],[\boldsymbol{\theta}]) = \frac{1}{m}\sum_{j=1}^{m} \frac{\mathcal{I}_{B^+}\left(\tilde{V}_j^*\right)}{p_{\tilde{V}|\boldsymbol{\theta}^*}\left(\tilde{V}_j^*\right)} [p]^+_{\tilde{V}|[\boldsymbol{\theta}]}\left(\tilde{V}_j^*\right). \tag{A.3}$$

By a similar argument, one obtains

$$\left[I_m^{\mathrm{MC}}\right]^-([\boldsymbol{v}],[\boldsymbol{\theta}]) = \frac{1}{m}\sum_{j=1}^{m} \frac{\mathcal{I}_{B^-}\left(\tilde{V}_j^*\right)}{p_{\tilde{V}|\boldsymbol{\theta}^*}\left(\tilde{V}_j^*\right)} [p]^-_{\tilde{V}|[\boldsymbol{\theta}]}\left(\tilde{V}_j^*\right), \tag{A.4}$$

where $B^- = \{\boldsymbol{\xi} \in \mathbb{V} : \rho_{\tilde{V}|\boldsymbol{\theta}}(\boldsymbol{\xi}) \leq [\beta]^-\}$.

These expressions define the inclusion function $\left[I_m^{\mathrm{MC}}\right] : \mathbb{I}(\mathbb{V}) \times \mathbb{I}(\mathbb{T}) \to \mathbb{I}([0,1])$ of the Approximate μ-transform in Eq. (5.7).

Appendix B

Inclusion Functions of Membership Functions

In order to find an inclusion function $[\mu_{\mathcal{G}}] : \mathbb{I}(\mathbb{T}) \to \mathbb{I}([0,1])$ of the μ-function $\mu_{\mathcal{G}} : \mathbb{T} \to [0,1]$ corresponding to a μ-graph $\mathcal{G}$ on $\mathbb{T}$ that tightly bounds the minimum and maximum value of $\mu_{\mathcal{G}}$ on a box $[\boldsymbol{t}] \in \mathbb{I}(\mathbb{T})$, the following is considered.

The function $[\mu_{\mathcal{G}}]^+ : \mathbb{I}(\mathbb{T}) \to [0,1]$ is straightforward to establish. The upper bound of $\mu_{\mathcal{G}}([\boldsymbol{t}])$ for some $[\boldsymbol{t}] \in \mathbb{I}(\mathbb{T})$ is found to be the maximum of all μ-levels of those μ-clusters whose μ-sets intersect with $[\boldsymbol{t}]$. That is,

$$[\mu_{\mathcal{G}}]^+([\boldsymbol{t}]) = \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G} : [\boldsymbol{t}] \cap [\boldsymbol{\tau}] \neq \emptyset} \alpha' = \sup_{\boldsymbol{t} \in [\boldsymbol{t}]} \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G} : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' = \sup_{\boldsymbol{t} \in [\boldsymbol{t}]} \mu_{\mathcal{G}}(\boldsymbol{t}) \tag{B.1}$$

produces the upper bound of $\mu_{\mathcal{G}}([\boldsymbol{t}])$. Finding the lower bound

$$\inf \mu_{\mathcal{G}}([\boldsymbol{t}]) = \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \mu_{\mathcal{G}}(\boldsymbol{t}) = \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G} : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' \tag{B.2}$$

is slightly more involved because two scenarios are conceivable.

- If there is at least one μ-cluster $([\boldsymbol{\tau}], \alpha')$ in $\mathcal{G}$ such that both $[\boldsymbol{t}] \cap [\boldsymbol{\tau}] \neq \emptyset$ and $[\boldsymbol{t}] \cap \neg[\boldsymbol{\tau}] \neq \emptyset$, further refinement is needed. One can bisect $[\boldsymbol{t}]$, and it follows that

$$\inf \mu_{\mathcal{G}}([\boldsymbol{t}]) = \min \left(\inf \mu_{\mathcal{G}} \left([\boldsymbol{t}]^{\mathrm{l}} \right), \inf \mu_{\mathcal{G}} \left([\boldsymbol{t}]^{\mathrm{u}} \right) \right). \tag{B.3}$$

- Otherwise, the μ-sets $[\boldsymbol{\tau}]$ in $\mathcal{G}$ must either satisfy $[\boldsymbol{t}] \cap \neg[\boldsymbol{\tau}] = \emptyset$, i.e they are superlevel sets of $[\boldsymbol{t}]$, or $[\boldsymbol{t}] \cap [\boldsymbol{\tau}] = \emptyset$, i.e. they are disjoint. Gathering the former in $\mathcal{G}' = \{([\boldsymbol{\tau}], \alpha') : ([\boldsymbol{\tau}], \alpha') \in \mathcal{G} \wedge [\boldsymbol{t}] \subseteq [\boldsymbol{\tau}]\}$, one obtains

$$\begin{aligned} \inf \mu_{\mathcal{G}}([\boldsymbol{t}]) &= \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G} : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' \\ &= \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G}' : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' = \max_{([\boldsymbol{\tau}],\alpha') \in \mathcal{G}' : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha', \end{aligned} \tag{B.4}$$

  i.e. the lower bound is the minimum of all μ-levels of the μ-clusters whose μ-sets are supersets of $[\boldsymbol{t}]$. If no such superset exists, the lower bound is, of course, zero.

A corresponding procedure to recursively compute $[\mu_{\mathcal{G}}]([\boldsymbol{t}])$ is formalized in Algorithm 3, which follows the above deliberations. In order to avoid infinite recursions and guarantee convergence, a technical termination criterion, the volume of $[\boldsymbol{t}]$ becoming too small in comparison to that of the μ-sets in $\mathcal{G}^{\cap}$, needs to also be introduced. The conservative bound employed in this case, is always true because

$$\begin{aligned} \inf \mu_{\mathcal{G}}([\boldsymbol{t}]) = \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \mu_{\mathcal{G}}(\boldsymbol{t}) &= \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \max_{([\boldsymbol{\tau}], \alpha') \in \mathcal{G} : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' \\ &\geq \inf_{\boldsymbol{t} \in [\boldsymbol{t}]} \min_{([\boldsymbol{\tau}], \alpha') \in \mathcal{G} : \boldsymbol{t} \in [\boldsymbol{\tau}]} \alpha' = \min_{([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap}} \alpha'. \end{aligned} \tag{B.5}$$

Algorithm 3: `mu`

input : μ-Partitions $\mathcal{G}$, Input Box $[\boldsymbol{t}]$, Tolerance ϵ_λ

output : Membership Box $[\mu]$

$\mathcal{G}^{\cap} \leftarrow \{([\boldsymbol{\tau}], \alpha') \in \mathcal{G} : [\boldsymbol{t}] \cap [\boldsymbol{\tau}] \neq \emptyset\}$ `// find intersecting` μ`-clusters`

if $\frac{\lambda([\boldsymbol{t}])}{\max\{\lambda([\boldsymbol{\tau}]) : ([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap}\}} < \epsilon_\lambda$ **then** `// technical termination criterion`

  $\alpha^{\min} \leftarrow \min\{\alpha' : ([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap}\}$ `// conservative lower bound`

  $\alpha^{\max} \leftarrow \max\{\alpha' : ([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap}\}$ `// tight upper bound`

else if $\exists([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap} : [\boldsymbol{t}] \cap \neg[\boldsymbol{\tau}] \neq \emptyset$ **then** `// find non-super-`μ`-set`

  $[\mu]_1 \leftarrow \texttt{mu}\left(\mathcal{G}^{\cap}, [\boldsymbol{t}]^{\mathrm{l}}\right)$ `// recursive call on lower subset`

  $[\mu]_2 \leftarrow \texttt{mu}\left(\mathcal{G}^{\cap}, [\boldsymbol{t}]^{\mathrm{u}}\right)$ `// recursive call on upper subset`

  $\alpha^{\min} \leftarrow \min\left([\mu]_1^-, [\mu]_2^-\right)$ `// lower bound is min. of lower subset bounds`

  $\alpha^{\max} \leftarrow \max\left([\mu]_1^+, [\mu]_2^+\right)$ `// upper bound is max. of upper subset bounds`

else `// termination criterion`

  $\alpha^{\min}, \alpha^{\max} \leftarrow \max\{\alpha' : ([\boldsymbol{\tau}], \alpha') \in \mathcal{G}^{\cap}\}$ `// tight lower and upper bound`

end

$[\mu] \leftarrow [\alpha^{\min}, \alpha^{\max}]$ `// assembly of membership box`

Bibliography

[AgrestiCoull98] Agresti, A.; Coull, B.A.: Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician, Vol. 52, No. 2, pp. 119–126, 1998.

[Aickin00] Aickin, M.: Connecting Dempster–Shafer belief functions with likelihood-based inference. Synthese, Vol. 123, No. 3, pp. 347–364, 2000.

[Angus94] Angus, J.E.: The probability integral transform and related results. SIAM review, Vol. 36, No. 4, pp. 652–654, 1994.

[AntonucciCuzzolin10] Antonucci, A.; Cuzzolin, F.: Credal sets approximation by lower probabilities: application to credal networks. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 716–725, Springer, 2010.

[AuerKielRauh13] Auer, E.; Kiel, S.; Rauh, A.: A verified method for solving piecewise smooth initial value problems. International Journal of Applied Mathematics and Computer Science, Vol. 23, No. 4, pp. 731–747, 2013.

[AugustinEtAl14] Augustin, T.; Coolen, F.P.; De Cooman, G.; Troffaes, M.C.: Introduction to Imprecise Probabilities. John Wiley & Sons, 2014.

[Balch12] Balch, M.S.: Mathematical foundations for a theory of confidence structures. International Journal of Approximate Reasoning, Vol. 53, No. 7, pp. 1003–1019, 2012.

[Balch20] Balch, M.S.: New Two-Sided Confidence Intervals for Binomial Inference Derived Using Walley's Imprecise Posterior Likelihood as a Test Statistic. International Journal of Approximate Reasoning, Vol. 123, pp. 77–98, 2020.

[BalchMartinFerson19] Balch, M.S.; Martin, R.; Ferson, S.: Satellite conjunction analysis and the false confidence theorem. Proceedings of the Royal Society A, Vol. 475, No. 2227, p. 20180565, 2019.

[BallBrittonO'neill02] Ball, F.G.; Britton, T.; O'neill, P.D.: Empty confidence sets for epidemics, branching processes and Brownian motion. Biometrika, Vol. 89, No. 1, pp. 211–224, 2002.

[BaudritCousoDubois07] Baudrit, C.; Couso, I.; Dubois, D.: Joint propagation of probability and possibility in risk analysis: Towards a formal framework. International Journal of Approximate Reasoning, Vol. 45, No. 1, pp. 82–105, 2007.

[BaudritDubois06] Baudrit, C.; Dubois, D.: Practical representations of incomplete probabilistic knowledge. Computational Statistics & Data Analysis, Vol. 51, No. 1, pp. 86–108, 2006.

[BaudritDuboisGuyonnet06] Baudrit, C.; Dubois, D.; Guyonnet, D.: Joint propagation and exploitation of probabilistic and possibilistic information in risk assessment. IEEE Transactions on Fuzzy Systems, Vol. 14, No. 5, pp. 593–608, 2006.

[Bayes63] Bayes, T.: An essay towards solving a problem in the doctrine of chances. Philosophical transactions of the Royal Society of London, Vol. 53, pp. 370–418, 1763.

[BeerFersonKreinovich13] Beer, M.; Ferson, S.; Kreinovich, V.: Imprecise probabilities in engineering analyses. Mechanical Systems and Signal Processing, Vol. 37, No. 1, pp. 4–29, 2013.

[BenferhatDuboisPrade00] Benferhat, S.; Dubois, D.; Prade, H.: Kalman-like filtering and updating in a possibilistic setting. In Proceedings of the 14th European Conference on Artificial Intelligence, pp. 8–12, IOS Press, 2000.

[BiehlerEtAl19] Biehler, J.; Mäck, M.; Nitzler, J.; Hanss, M.; Koutsourelakis, P.S.; Wall, W.A.: Multifidelity approaches for uncertainty quantification. GAMM-Mitteilungen, Vol. 42, No. 2, p. e201900008, 2019.

[Bradley19] Bradley, S.: Imprecise Probabilities. In Zalta, E.N. (Ed.): The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Spring 2019.

[Bronevich01] Bronevich, A.G.: The logical inferences in possibility theory based on the concept of upper (lower) probability. In Proceedings of the 10th IEEE International Conference on Fuzzy Systems, Vol. 1, pp. 192–194, IEEE, 2001.

[BronevichKlir10] Bronevich, A.; Klir, G.J.: Measures of uncertainty for imprecise probabilities: An axiomatic approach. International Journal of Approximate Reasoning, Vol. 51, No. 4, pp. 365 – 390, 2010.

[BronevichRozenberg18] Bronevich, A.G.; Rozenberg, I.N.: Modelling uncertainty with generalized credal sets: application to conjunction and decision. International Journal of General Systems, Vol. 47, No. 1, pp. 67–96, 2018.

[BronevichRozenberg19] Bronevich, A.G.; Rozenberg, I.N.: The contradiction between belief functions: its description, measurement, and correction based on generalized credal sets. International Journal of Approximate Reasoning, Vol. 112, pp. 119–139, 2019.

[BronevichRozenberg20] Bronevich, A.G.; Rozenberg, I.N.: Canonical sequences of monotone measures. Fuzzy Sets and Systems, Vol. 379, pp. 1–19, 2020.

[Brown75] Brown, M.B.: A method for combining non-independent, one-sided tests of significance. Biometrics, Vol. 31, No. 4, pp. 987–992, 1975.

[BrownCaiDasGupta01] Brown, L.D.; Cai, T.T.; DasGupta, A.: Interval estimation for a binomial proportion. Statistical Science, Vol. 16, No. 2, pp. 101–133, 2001.

[CahoonMartin19] Cahoon, J.; Martin, R.: Generalized inferential models for meta-analyses based on few studies. arXiv preprint arXiv:1910.00533, 2019.

[CahoonMartin21] Cahoon, J.; Martin, R.: Generalized inferential models for censored data. International Journal of Approximate Reasoning, Vol. 137, pp. 51–66, 2021.

[CampiCalafioreGaratti09] Campi, M.C.; Calafiore, G.; Garatti, S.: Interval predictor models: Identification and reliability. Automatica, Vol. 45, No. 2, pp. 382–392, 2009.

[Cattaneo17] Cattaneo, M.E.: The likelihood interpretation as the foundation of fuzzy set theory. International Journal of Approximate Reasoning, Vol. 90, pp. 333–340, 2017.

[CellaMartin22] Cella, L.; Martin, R.: Validity, consonant plausibility measures, and conformal prediction. International Journal of Approximate Reasoning, Vol. 141, pp. 110–130, 2022.

[ChameauSantamarina87] Chameau, J.L.; Santamarina, J.C.: Membership functions I: Comparing methods of measurement. International Journal of Approximate Reasoning, Vol. 1, No. 3, pp. 287–301, 1987.

[Chang81] Chang, W.: Ranking of fuzzy utilities with triangular membership functions. In Proceedings of the International Conference of Policy Analysis and Information Systems, pp. 263–272, 1981.

[Choquet54] Choquet, G.: Theory of capacities. In Annales de l'institut Fourier, Vol. 5, pp. 131–295, 1954.

[ClopperPearson34] Clopper, C.J.; Pearson, E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika, Vol. 26, No. 4, pp. 404–413, 1934.

[CousoMontesGil01] Couso, I.; Montes, S.; Gil, P.: The necessity of the strong α-cuts of a fuzzy set. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 9, No. 2, pp. 249–262, 2001.

[CousoMoralWalley00] Couso, I.; Moral, S.; Walley, P.: A survey of concepts of independence for imprecise probabilities. Risk, Decision and Policy, Vol. 5, No. 2, pp. 165–181, 2000.

[CramerKamps20] Cramer, E.; Kamps, U.: Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. Springer, 2020.

[Cuzzolin13] Cuzzolin, F.: L_p consonant approximations of belief functions. IEEE Transactions on Fuzzy Systems, Vol. 22, No. 2, pp. 420–436, 2013.

[DavidNagaraja04] David, H.A.; Nagaraja, H.N.: Order Statistics. John Wiley & Sons, 2004.

[De Cooman97a] De Cooman, G.: Possibility theory I: The measure-and integral-theoretic groundwork. International Journal of General Systems, Vol. 25, No. 4, pp. 291–323, 1997.

[De Cooman97b] De Cooman, G.: Possibility theory II: Conditional possibility. International Journal of General Systems, Vol. 25, No. 4, pp. 325–351, 1997.

[De Cooman97c] De Cooman, G.: Possibility theory III: Possibilistic independence. International Journal of General Systems, Vol. 25, No. 4, pp. 353–371, 1997.

[De CoomanAeyels99] De Cooman, G.; Aeyels, D.: Supremum preserving upper probabilities. Information Sciences, Vol. 118, No. 1-4, pp. 173–212, 1999.

[De MartinoDe Martino18] De Martino, A.; De Martino, D.: An introduction to the maximum entropy approach and its application to inference problems in biology. Heliyon, Vol. 4, No. 4, p. e00596, 2018.

[DelgadoMoral87] Delgado, M.; Moral, S.: On the concept of possibility-probability consistency. Fuzzy Sets and Systems, Vol. 21, No. 3, pp. 311–318, 1987.

[Dempster67] Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. Annals of Mathematical Statistics, Vol. 38, pp. 325–339, 1967.

[Denneberg94] Denneberg, D.: Non-additive measure and integral. Springer Science & Business Media, 1994.

[Denoeux14] Denoeux, T.: Likelihood-based belief function: Justification and some extensions to low-quality data. International Journal of Approximate Reasoning, Vol. 55, No. 7, pp. 1535–1547, 2014.

[Denœux09] Denœux, T.: Extending stochastic ordering to belief functions on the real line. Information Sciences, Vol. 179, No. 9, pp. 1362–1376, 2009.

[DesterckeDuboisChojnacki08] Destercke, S.; Dubois, D.; Chojnacki, E.: Unifying practical uncertainty representations. Part I: Generalized p-boxes. International Journal of Approximate Reasoning, Vol. 49, No. 3, pp. 649–663, 2008.

[DesterckeDuboisChojnacki09] Destercke, S.; Dubois, D.; Chojnacki, E.: A consonant approximation of the product of independent consonant random sets. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 17, No. 6, pp. 773–792, 2009.

[Dubois82] Dubois, D.: On Several Presentations of uncertain body of evidence. In Madan M. Gupta, E.S. (Ed.): Fuzzy Information and Decision Processes, pp. 167–189. North-Holland, Amsterdam, 1982.

[Dubois06] Dubois, D.: Possibility theory and statistical reasoning. Computational Statistics & Data Analysis, Vol. 51, No. 1, pp. 47–69, 2006.

[DuboisDenœux10] Dubois, D.; Denœux, T.: Statistical inference with belief functions and possibility measures: A discussion of basic assumptions. In Borgelt, C.; González-Rodríguez, G.; Trutschnig, W.; Lubiano, M.A.; Gil, M.Á.; Grzegorzewski, P.; Hryniewicz, O. (Eds.): Combining Soft Computing and Statistical Methods in Data Analysis, pp. 217–225. Springer, 2010.

[DuboisEtAl04] Dubois, D.; Foulloy, L.; Mauris, G.; Prade, H.: Probability-possibility transformations, triangular fuzzy sets, and probabilistic inequalities. Reliable Computing, Vol. 10, No. 4, pp. 273–297, 2004.

[DuboisMoralPrade97] Dubois, D.; Moral, S.; Prade, H.: A semantics for possibility theory based on likelihoods. Journal of Mathematical Analysis and Applications, Vol. 205, No. 2, pp. 359–380, 1997.

[DuboisPrade86a] Dubois, D.; Prade, H.: The principle of minimum specificity as a basis for evidential reasoning. In Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, pp. 75–84, Springer, 1986.

[DuboisPrade86b] Dubois, D.; Prade, H.: A set-theoretic view of belief functions logical operations and approximations by fuzzy sets. International Journal Of General Systems, Vol. 12, No. 3, pp. 193–226, 1986.

[DuboisPrade87] Dubois, D.; Prade, H.: The mean value of a fuzzy number. Fuzzy Sets and Systems, Vol. 24, No. 3, pp. 279 – 300, 1987.

[DuboisPrade88] Dubois, D.; Prade, H.: Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, 1988.

[DuboisPrade90] Dubois, D.; Prade, H.: Consonant approximations of belief functions. International Journal of Approximate Reasoning, Vol. 4, No. 5, pp. 419 – 449, 1990.

[DuboisPrade92] Dubois, D.; Prade, H.: When upper probabilities are possibility measures. Fuzzy Sets and Systems, Vol. 49, No. 1, pp. 65–74, 1992.

[DuboisPrade98] Dubois, D.; Prade, H.: Possibility theory: Qualitative and quantitative aspects. In Smets, P. (Ed.): Quantified Representation of Uncertainty and Imprecision, pp. 169–226, Springer, 1998.

[DuboisPrade16] Dubois, D.; Prade, H.: Practical methods for constructing possibility distributions. International Journal of Intelligent Systems, Vol. 31, No. 3, pp. 215–239, 2016.

[DuboisPrade20] Dubois, D.; Prade, H.: Membership Functions. In Fuzzy Approaches for Soft Computing and Approximate Reasoning: Theories and Applications, pp. 5–20, Springer, 2020.

[EschmannEbelEberhard21] Eschmann, H.; Ebel, H.; Eberhard, P.: Data-based model of an omnidirectional mobile robot using Gaussian processes. IFAC-PapersOnLine, Vol. 54, No. 7, pp. 13–18, 2021.

[Fériet82] de Fériet, J.K.: Interpretation of membership functions of fuzzy sets in terms of plausibility and belief. Fuzzy Information and Decision Processes, Vol. 1, pp. 93–98, 1982.

[FersonEtAl15] Ferson, S.; Kreinovich, V.; Grinzburg, L.; Myers, D.; Sentz, K.: Constructing probability boxes and Dempster-Shafer structures. Tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2015.

[FersonGinzburg96] Ferson, S.; Ginzburg, L.R.: Different methods are needed to propagate ignorance and variability. Reliability Engineering & System Safety, Vol. 54, No. 2, pp. 133–144, 1996.

[FersonGray21] Ferson, S.; Gray, A.: Distribution-Free Uncertainty Propagation. In Proceedings of the 9th International Workshop on Reliable Engineering Computing in Taormina (Italy), pp. 395–407, 2021.

[FetzOberguggenberger04] Fetz, T.; Oberguggenberger, M.: Propagation of uncertainty through multivariate functions in the framework of sets of probability measures. Reliability Engineering & System Safety, Vol. 85, No. 1-3, pp. 73–87, 2004.

[FetzOberguggenberger16] Fetz, T.; Oberguggenberger, M.: Imprecise random variables, random sets, and Monte Carlo simulation. International Journal of Approximate Reasoning, Vol. 78, pp. 252–264, 2016.

[Fieller54] Fieller, E.C.: Some problems in interval estimation. Journal of the Royal Statistical Society: Series B (Methodological), Vol. 16, No. 2, pp. 175–185, 1954.

[Fisher46] Fisher, R.A.: In Statistical methods for research workers: pp. 66–70. Springer, 10. Edn., 1946.

[FröhlichEtAl22] Fröhlich, B.; Hose, D.; Dieterich, O.; Hanss, M.; Eberhard, P.: Uncertainty quantification of large-scale dynamical systems using parametric model order reduction. Mechanical Systems and Signal Processing, Vol. 171, p. 108855, 2022.

[GrayEtAl21] Gray, A.; Hose, D.; De Angelis, M.; Hanss, M.; Ferson, S.: Dependent possibilistic arithmetic using copulas. In Proceedings of Machine Learning Research, 147, pp. 169–179, 2021.

[GrunertFehrHaasdonk20] Grunert, D.; Fehr, J.; Haasdonk, B.: Well-scaled, a-posteriori error estimation for model order reduction of large second-order mechanical systems. Zeitschrift für Angewandte Mathematik und Mechanik, Vol. 100, No. 8, p. e201900186, 2020.

[Halmos13] Halmos, P.R.: Measure theory, Vol. 18. Springer, 2013.

[HamannEtAl18] Hamann, D.; Walz, N.P.; Fischer, A.; Hanss, M.; Eberhard, P.: Fuzzy arithmetical stability analysis of uncertain machining systems. Mechanical Systems and Signal Processing, Vol. 98, pp. 534–547, 2018.

[Hamming12] Hamming, R.: Numerical methods for scientists and engineers. Courier Corporation, 2012.

[HanselowskiIhrleHanss15] Hanselowski, A.; Ihrle, S.; Hanss, M.: A fuzzy model updating technique motivated by bayesian inference. In Proceedings of the 1st ECOMMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering, pp. 548–559, 2015.

[Hanss02] Hanss, M.: The transformation method for the simulation and analysis of systems with uncertain parameters. Fuzzy Sets and Systems, Vol. 130, No. 3, pp. 277–289, 2002.

[Hanss05] Hanss, M.: Applied Fuzzy Arithmetic – An Introduction with Engineering Applications. Springer, 2005.

[HastieTibshiraniFriedman01] Hastie, T.; Tibshirani, R.; Friedman, J.: The elements of statistical learning, Vol. 1. Springer series in statistics, New York, 2001.

[HeltonJohnsonOberkampf04] Helton, J.C.; Johnson, J.D.; Oberkampf, W.L.: An exploration of alternative approaches to the representation of uncertainty in model predictions. Reliability Engineering & System Safety, Vol. 85, No. 1-3, pp. 39–71, 2004.

[HerreroEtAl12] Herrero, P.; Georgiou, P.; Toumazou, C.; Delaunay, B.; Jaulin, L.: An Efficient Implementation of the SIVIA Algorithm in a High-Level Numerical Programming Language. Reliable Computing, Vol. 16, pp. 239–251, 2012.

[Hill68] Hill, B.M.: Posterior distribution of percentiles: Bayes' theorem for sampling from a population. Journal of the American Statistical Association, Vol. 63, No. 322, pp. 677–691, 1968.

[Hájek19] Hájek, A.: Interpretations of Probability. In Zalta, E.N. (Ed.): The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Fall 2019.

[HofmannHanss16] Hofmann, A.; Hanss, M.: In A Fuzzy Arithmetical Strategy for Robust Controller Design in the Presence of Model Uncertainties. 2016.

[HoseEtAl18] Hose, D.; Hamann, D.; Hanss, M.; Eberhard, P.: A Data-Driven Possibilistic Approach to the Identification of Uncertain Stability Lobe Diagrams. PAMM, Vol. 18, No. 1, p. e201800087, 2018.

[HoseHanss18a] Hose, D.; Hanss, M.: On inverse fuzzy arithmetical problems in uncertainty analysis. In Proceedings of the 7th International Conference on Uncertainties in Structural Dynamics in Leuven (Belgium), 2018.

[HoseHanss18b] Hose, D.; Hanss, M.: Possibilistic Identification of Reliable Finite Impulse Response Models. In Proceedings of the 8th International Workshop on Reliable Engineering Computing in Liverpool (United Kingdom), 2018.

[HoseHanss19a] Hose, D.; Hanss, M.: Consistent Inverse Probability and Possibility Propagation. In Proceedings of the 2019 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology, Atlantis Press, 2019.

[HoseHanss19b] Hose, D.; Hanss, M.: Fuzzy linear least squares for the identification of possibilistic regression models. Fuzzy Sets and Systems, Vol. 367, pp. 82–95, 2019.

[HoseHanss19c] Hose, D.; Hanss, M.: Possibilistic calculus as a conservative counterpart to probabilistic calculus. Mechanical Systems and Signal Processing, Vol. 133, p. 106290, 2019.

[HoseHanss19d] Hose, D.; Hanss, M.: Towards a General Theory for Data-Based Possibilistic Parameter Inference. In Papadrakakis, M.; Papadopoulos, V.; Stefanou, G. (Eds.): Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering in Crete (Greece), pp. 114–126, ECCOMAS Proceedia, 2019.

[HoseHanss20] Hose, D.; Hanss, M.: On data-based estimation of possibility distributions. Fuzzy Sets and Systems, Vol. 399, pp. 77–94, 2020.

[HoseHanss21a] Hose, D.; Hanss, M.: On the Solution of Forward and Inverse Problems in Possibilistic Uncertainty Quantification for Dynamical Systems. In Proceedings of the 9th International Workshop on Reliable Engineering Computing in Taormina (Italy), pp. 295–307, 2021.

[HoseHanss21b] Hose, D.; Hanss, M.: A recursive formulation of possibilistic filters. In Proceedings of Machine Learning Research, 147, pp. 180–190, 2021.

[HoseHanss21c] Hose, D.; Hanss, M.: A universal approach to imprecise probabilities in possibility theory. International Journal of Approximate Reasoning, Vol. 133, pp. 133–158, 2021.

[HoseMäckHanss18] Hose, D.; Mäck, M.; Hanss, M.: A Possibilistic Approach to the Optimization of Uncertain Systems. In Proceedings of the 3rd International Conference on Vulnerability and Risk Analysis and Management, and the 7th International Symposium on Uncertainty Modeling and Analysis in Florianópolis (Brazil), 2018.

[HoseMäckHanss19a] Hose, D.; Mäck, M.; Hanss, M.: On Probability-Possibility Consistency in High-Dimensional Propagation Problems. In Papadrakakis, M.; Papadopoulos, V.; Stefanou, G. (Eds.): Proceedings of the 3rd International Conference on Uncertainty Quantification in Computational Sciences and Engineering in Crete (Greece), pp. 127–136, ECCOMAS Proceedia, 2019.

[HoseMäckHanss19b] Hose, D.; Mäck, M.; Hanss, M.F.: Robust optimization in possibility theory. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, Vol. 5, No. 4, 2019.

[HuangDissanayake99] Huang, S.; Dissanayake, G.: Robot localization: An introduction. Wiley Encyclopedia of Electrical and Electronics Engineering, pp. 1–10, 1999.

[JaulinEtAl01] Jaulin, L.; Kieffer, M.; Didrit, O.; Walter, E.: Applied interval analysis: with examples in parameter and state estimation, robust control and robotics, Vol. 1. Springer Science & Business Media, 2001.

[Jaynes57] Jaynes, E.T.: Information theory and statistical mechanics. Physical Review, Vol. 106, No. 4, p. 620, 1957.

[Jennrich69] Jennrich, R.I.: Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics, Vol. 40, No. 2, pp. 633–643, 1969.

[Joe14] Joe, H.: Dependence Modeling with Copulas. CRC press, 2014.

[Kalman60] Kalman, R.E.: A new approach to linear filtering and prediction problems. Journal of Fluids Engineering, Vol. 82, No. 1, pp. 35–45, 1960.

[KappenmanGeisserAntle70] Kappenman, R.F.; Geisser, S.; Antle, C.E.: Bayesian and fiducial solutions for the Fieller-Creasy problem. The Indian Journal of Statistics, Series B, pp. 331–340, 1970.

[KlirWierman99] Klir, G.; Wierman, M.: Uncertainty-Based Information: Elements of Generalized Information Theory, Vol. 15. Springer Science & Business Media, 1999.

[KostMcDermott02] Kost, J.T.; McDermott, M.P.: Combining dependent p-values. Statistics & Probability Letters, Vol. 60, No. 2, pp. 183–190, 2002.

[LavioletteEtAl95] Laviolette, M.; Seaman, J.W.; Barrett, J.D.; Woodall, W.H.: A probabilistic and statistical view of fuzzy methods. Technometrics, Vol. 37, No. 3, pp. 249–261, 1995.

[LawlessFredette05] Lawless, J.; Fredette, M.: Frequentist prediction intervals and predictive distributions. Biometrika, Vol. 92, No. 3, pp. 529–542, 2005.

[LeongNair16] Leong, P.H.; Nair, G.N.: Set-membership filtering using random samples. In Proceedings of the 19th International Conference on Information Fusion (FUSION), pp. 1087–1094, IEEE, 2016.

[LiuMartin20] Liu, C.; Martin, R.: Inferential models and possibility measures. arXiv preprint arXiv:2008.06874, 2020.

[MäckHanss19] Mäck, M.; Hanss, M.: Efficient Possibilistic Uncertainty Analysis of a Car Crash Scenario Using a Multifidelity Approach. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, Vol. 5, No. 4, 2019. 041015.

[MäckHanss21] Mäck, M.; Hanss, M.: An advanced sampling technique for possibilistic uncertainty propagation. Mechanical Systems and Signal Processing, Vol. 147, p. 107064, 2021.

[MäckHoseHanss17] Mäck, M.; Hose, D.; Hanss, M.: On using fuzzy arithmetic in optimization problems with uncertain constraints. PAMM, Vol. 17, No. 1, pp. 57–58, 2017.

[Martin15] Martin, R.: Plausibility functions and exact frequentist inference. Journal of the American Statistical Association, Vol. 110, No. 512, pp. 1552–1561, 2015.

[Martin18] Martin, R.: On an inferential model construction using generalized associations. Journal of Statistical Planning and Inference, Vol. 195, pp. 105–115, 2018.

[Martin21] Martin, R.: An imprecise-probabilistic characterization of frequentist statistical inference. arXiv preprint arXiv:2112.10904, 2021.

[MartinLingham16] Martin, R.; Lingham, R.T.: Prior-free probabilistic prediction of future observations. Technometrics, Vol. 58, No. 2, pp. 225–235, 2016.

[MartinLiu15] Martin, R.; Liu, C.: Inferential Models: Reasoning with Uncertainty, Vol. 145. CRC Press, 2015.

[MassonDenœux06] Masson, M.H.; Denœux, T.: Inferring a possibility distribution from empirical data. Fuzzy Sets and Systems, Vol. 157, No. 3, pp. 319–340, 2006.

[Mauris09] Mauris, G.: The principle of possibility maximum specificity as a basis for measurement uncertainty expression. In Proceedings of IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement, pp. 5–9, IEEE, 2009.

[MilaneseNovara11] Milanese, M.; Novara, C.: Unified set membership theory for identification, prediction and filtering of nonlinear systems. Automatica, Vol. 47, No. 10, pp. 2141–2151, 2011.

[MontesMirandaVicig19] Montes, I.; Miranda, E.; Vicig, P.: Outer approximating coherent lower probabilities with belief functions. International Journal of Approximate Reasoning, Vol. 110, pp. 1–30, 2019.

[MoreyEtAl16] Morey, R.D.; Hoekstra, R.; Rouder, J.N.; Lee, M.D.; Wagenmakers, E.J.: The fallacy of placing confidence in confidence intervals. Psychonomic Bulletin & Review, Vol. 23, No. 1, pp. 103–123, 2016.

[Neumaier03] Neumaier, A.: Fuzzy modeling in terms of surprise. Fuzzy Sets and Systems, Vol. 135, No. 1, pp. 21–38, 2003.

[Neumaier04] Neumaier, A.: Clouds, fuzzy sets, and probability intervals. Reliable Computing, Vol. 10, No. 4, pp. 249–272, 2004.

[NocedalWright06] Nocedal, J.; Wright, S.: Numerical Optimization. Springer Science & Business Media, 2006.

[Pyt'ev97] Pyt'ev, Y.P.: Methods of the theory of possibilities in the problems of optimal estimation and decision making, part 1. Pattern Recognition and Image Analysis, Vol. 7, No. 3, pp. 338–346, 1997.

[RaoRawlingsMayne03] Rao, C.V.; Rawlings, J.B.; Mayne, D.Q.: Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. IEEE Transactions on Automatic Control, Vol. 48, No. 2, pp. 246–258, 2003.

[ReidCox15] Reid, N.; Cox, D.R.: On some principles of statistical inference. International Statistical Review, Vol. 83, No. 2, pp. 293–308, 2015.

[Rigollet15] Rigollet, P.: Confidence Intervals for the binomial parameter p. Lecture notes in MIT mathematics course 18.650 "Fundamentals of Statistics", 2015.

[RobertCasella13] Robert, C.; Casella, G.: Monte Carlo Statistical Methods. Springer Science & Business Media, 2013.

[Särkkä13] Särkkä, S.: Bayesian Filtering and Smoothing. 3. Cambridge University Press, 2013.

[SavageEtAl62] Savage, L.J.; Barnard, G.; Cornfield, J.; Bross, I.; Good, I.; Lindley, D.; Clunies-Ross, C.; Pratt, J.W.; Levene, H.; Goldman, T.: On the foundations of statistical inference: Discussion. Journal of the American Statistical Association, Vol. 57, No. 298, pp. 307–326, 1962.

[Schmelzer15a] Schmelzer, B.: Joint distributions of random sets and their relation to copulas. International Journal of Approximate Reasoning, Vol. 65, pp. 59–69, 2015.

[Schmelzer15b] Schmelzer, B.: Sklar's theorem for minitive belief functions. International Journal of Approximate Reasoning, Vol. 63, pp. 48–61, 2015.

[Seidenfeld95] Seidenfeld, T.: Jeffreys, Fisher, and Keynes: Predicting the third observation, given the first two. History of Political Economy, Vol. 27, pp. 39–52, 1995.

[Shafer76] Shafer, G.: A Mathematical Theory of Evidence, Vol. 42. Princeton University Press, 1976.

[Shafer90] Shafer, G.: The unity and diversity of probability. Statistical Science, Vol. 5, No. 4, pp. 435–444, 1990.

[Shafer16] Shafer, G.: A mathematical theory of evidence turns 40. International Journal of Approximate Reasoning, Vol. 79, pp. 7–25, 2016.

[Shao03] Shao, J.: Mathematical Statistics. Springer, 2. Edn., 2003.

[ShaoHuangLee10] Shao, X.; Huang, B.; Lee, J.M.: Constrained Bayesian state estimation–A comparative study and a new particle filter based approach. Journal of Process Control, Vol. 20, No. 2, pp. 143–157, 2010.

[Shapley53] Shapley, L.S.: A value for n-person games. Contributions to the Theory of Games, Vol. 2, No. 28, pp. 307–317, 1953.

[SkogestadPostlethwaite07] Skogestad, S.; Postlethwaite, I.: Multivariable Feedback Control: Analysis and Design, Vol. 2. Wiley, 2007.

[Smets90] Smets, P.: In Constructing the Pignistic Probability Function in a Context of Uncertainty: Uncertainty in Artificial Intelligence, Vol. 10 of Machine Intelligence and Pattern Recognition, pp. 29 – 39. North-Holland, 1990.

[Smets05] Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. International Journal of Approximate Reasoning, Vol. 38, No. 2, pp. 133 – 147, 2005.

[Sudret15] Sudret, B.: Polynomial chaos expansions and stochastic finite element methods. Risk and Reliability in Geotechnical Engineering, pp. 265–300, 2015.

[Sullivan15] Sullivan, T.J.: Introduction to Uncertainty Quantification. Springer, 2015.

[Tangirala14] Tangirala, A.K.: Principles of System Identification: Theory and Practice. CRC Press, 2014.

[Thrun07] Thrun, S.: Simultaneous localization and mapping. In Robotics and Cognitive Approaches to Spatial Mapping, pp. 13–41, Springer, 2007.

[TroffaesCooman14] Troffaes, M.C.M.; de Cooman, G.: Lower Previsions. Wiley Series in Probability and Statistics. John Wiley & Sons, 2014.

[TroffaesMirandaDestercke13] Troffaes, M.C.; Miranda, E.; Destercke, S.: On the connection between probability boxes and possibility measures. Information Sciences, Vol. 224, pp. 88–108, 2013.

[Vineberg16] Vineberg, S.: Dutch Book Arguments. In Zalta, E.N. (Ed.): The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, Spring 2016.

[Vollset93] Vollset, S.E.: Confidence intervals for a binomial proportion. Statistics in Medicine, Vol. 12, No. 9, pp. 809–824, 1993.

[Walley91] Walley, P.: Statistical Reasoning with Imprecise Probabilities. Chapman and Hall, 1991.

[Walz16] Walz, N.P.: Fuzzy Arithmetical Methods for Possibilistic Uncertainty Analysis. Schriften aus dem Institut für Technische und Numerische Mechanik der Universität Stuttgart. Shaker, 2016. Ph.D. Thesis.

[WangKlir13] Wang, Z.; Klir, G.J.: Fuzzy Measure Theory. Springer Science & Business Media, 2013.

[Wasserman90] Wasserman, L.A.: Belief functions and statistical inference. Canadian Journal of Statistics, Vol. 18, No. 3, pp. 183–196, 1990.

[WilliamsonDowns90] Williamson, R.C.; Downs, T.: Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds. International Journal of Approximate Reasoning, Vol. 4, No. 2, pp. 89–158, 1990.

[YagerLiu08] Yager, R.R.; Liu, L.: Classic Works of the Dempster-Shafer Theory of Belief Functions, Vol. 219. Springer, 2008.

[Zadeh65] Zadeh, L.A.: Fuzzy sets. Information and Control, Vol. 8, No. 3, pp. 338–353, 1965.

[Zadeh75a] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning - I. Information Sciences, Vol. 8, No. 3, pp. 199–249, 1975.

[Zadeh75b] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-II. Information Sciences, Vol. 8, No. 4, pp. 301–357, 1975.

[Zadeh75c] Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning-III. Information Sciences, Vol. 9, No. 1, pp. 43–80, 1975.

[Zadeh88] Zadeh, L.A.: Fuzzy logic. Computer, Vol. 21, No. 4, pp. 83–93, 1988.

[Zadeh95] Zadeh, L.A.: Discussion: Probability theory and fuzzy logic are complementary rather than competitive. Technometrics, Vol. 37, No. 3, pp. 271–276, 1995.

[Zadeh99] Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, Vol. 100, pp. 9–34, 1999.

[Zubyuk19] Zubyuk, A.: A new approach to specificity in possibility theory: Decision-making point of view. Fuzzy Sets and Systems, Vol. 364, pp. 76–95, 2019.

Nomenclature and Notation

The following overview lists the most important symbols, mathematical conventions and abbreviations in this thesis.

Latin Minuscules

f	system dynamics
g	system output equation
ℓ	absolute likelihood function
$\boldsymbol{n}$	process noise
p	probability mass/density function
q	observation
$\boldsymbol{u}$	system input
$\boldsymbol{w}$	measurement noise
$\boldsymbol{x}$	system state
$\boldsymbol{y}$	system output

Latin Capitals

C	stochastic/probability copula
D	dimension
F	cumulative probability distribution function
I	Membership Transform
J	cost function
M	capacity
N	necessity measure
P	probability measure
$\tilde{Q}$	imprecise output variable
$\tilde{V}$	imprecise variable
Var	variance

Fraktur Latin Capitals

$\mathfrak{C}$	credal set
$\mathfrak{E}$	expectation
$\mathfrak{M}$	inferential model
$\mathfrak{O}$	Optimal Imprecise-Probability-to-Possibility Transform
$\mathfrak{P}$	set of probability measures
$\mathfrak{S}$	statistical model
$\mathfrak{T}$	Imprecise-Probability-to-Possibility Transform

Blackboard Bold Latin Capitals

$\mathbb{B}$	Borel σ-field
$\mathbb{D}$	feature space
$\mathbb{E}$	nuisance parameter space
$\mathbb{I}$	interval/box space
$\mathbb{N}$	process noise space
$\mathbb{P}$	space of probability measures
$\mathbb{Q}$	observation space
$\mathbb{R}$	Euclidean space
$\mathbb{U}$	input space
$\mathbb{V}$	imprecise variable space
$\mathbb{W}$	measurement error space
$\mathbb{X}$	state space
$\mathbb{Y}$	output space

Calligraphic Latin Capitals

$\mathcal{A}$	superuniform possibility distribution
$\mathcal{C}$	superlevel set, α-cut
$\mathcal{D}$	deterministic distribution
$\mathcal{E}$	exponential probability distribution
$\mathcal{G}$	μ-graph
$\mathcal{I}$	indicator function
$\mathcal{J}$	possibility copula
$\mathcal{K}$	μ-cluster
$\mathcal{M}$	Markov/Chebychev possibility distribution
$\mathcal{N}$	normal/Gaussian probability distribution
$\mathcal{P}$	μ-partition
$\mathcal{Q}$	quasi-vacuous possibility distribution
$\mathcal{S}$	sublevel set
$\mathcal{U}$	uniform probability distribution
$\mathcal{V}$	vacuous possibility distribution

Greek Minuscules

γ	elementary confidence function
δ	population feature
ε	tolerance
η	nuisance parameter
θ	population parameter
κ	elementary prediction function
λ	Lebesgue measure
μ	membership function
π	elementary possibility/ (IP-)description distribution
ρ	elementary plausibility function
ϕ	explicit relationship
ψ	inverse relationship

Greek Capitals

Δ	triangular possibility distribution	Ξ	implicit relationship
Θ	parameter space	Π	possibility measure
Λ	relative likelihood function	Σ	σ-field
		Ω	universal set

Relations

$\sqsubseteq$	plausibility order	$\sim$	distributed as
$\preceq$	inclusion order	$\vee$	disjunction
$\vdash$	stochastic order	$\wedge$	conjunction

Mathematical Conventions

A function $f : \mathbb{X} \to \mathbb{Y}$ with a set-valued argument $X \subseteq \mathbb{X}$ is to be understood as the set of all images $f(X) = \{f(x) : x \in X\}$ and, similarly, $f^{-1}(Y) = \{x \in \mathbb{X} : f(x) \in Y\}$ for $Y \subseteq \mathbb{Y}$.

The definite integral of a real function $f : \mathbb{R} \to \mathbb{R}$, where the bounds are reversed is understood as $\int_b^a f(x)\mathrm{d}x = -\int_a^b f(x)\mathrm{d}x$ for $a, b \in \mathbb{R}$ and $a < b$.

The complement of the set $X \subseteq \mathbb{X}$ is indicated by $\neg X = \{x \in \mathbb{X} : x \notin X\}$.

The maximum and supremum of the empty set $\emptyset$ are defined to be $\max \emptyset = \sup \emptyset = 0$.

Vectors $\boldsymbol{x} \in \mathbb{R}^D$ and matrices $\boldsymbol{M} \in \mathbb{R}^{D \times D}$ are written in bold letters and their transpose is denoted by $\boldsymbol{x}^{\mathrm{T}}$ and $\boldsymbol{M}^{\mathrm{T}}$, respectively, and the i-th element of $\boldsymbol{x}$ is denoted by x_i for $i = 1, \ldots, D$.

Abbreviations

CNF	cumulative necessity distribution function
CPF	cumulative probability distribution function
CΠF	cumulative possibility distribution function
FSIVIA	fuzzy set inversion via interval analysis
iid	independent and identically distributed
IM	inferential model
IP	imprecise probabilities
IP-Π-transform	Imprecise-Probability-to-Possibility Transform
LTI	linear time-invariant
MEVIA	membership extension via interval analysis
NI	non-interaction
p-box	probability box
P-Π-transform	Probability-to-Possibility Transform
P-Γ-transform	Probability-to-Confidence Transform
RPPF	recursive possibilistic particle filter
SI	strong independence
UI	unknown interaction
μ-transform	Membership Transform
Π-copula	possibility copula
Π-IM	possibilistic inferential model
Π-PM	possibilistic predictor model

Mischief managed.

Joanne K. Rowling,
Harry Potter and the Prisoner of Azkaban