UMEÅ UNIVERSITY

Probabilistic Metric Space for Machine Learning: Data and Model Spaces

Mariam Taha

Doctoral Thesis, May 2025
Department of Computing Science
Umeå University
Sweden

Department of Computing Science
Umeå University
SE-901 87 Umeå, Sweden

mariamt@cs.umu.se

ISBN 978-91-8070-680-3 (print)
ISBN 978-91-8070-681-0 (digital)
ISSN 0348-0542
UMINF 25.05
Cover illustrated by Mariam Taha and Marlene Lahti

Printed by Scandinavian Print Group, Umeå University, 2025

Dedicated to my father, whose memory has been a constant source of strength

Abstract

Machine learning models are inherently shaped by the data used to train them. Understanding the relationship between datasets and the models they generate is essential for tasks such as model selection, privacy metrics, and robustness evaluation. This thesis presents a rigorous mathematical framework for comparing machine learning models and algorithms by formalizing the interaction between two fundamental spaces: the database space, which captures possible datasets, and the model space, which contains the models or hypotheses derived from those datasets.

A central motivation stems from the observation that different datasets can lead to the same or highly similar models. Such recurrent models—which arise frequently across diverse data sources—are particularly significant in privacy-sensitive applications. Their recurrence suggests reduced dependence on any specific data point or subgroup, thus offering inherent privacy and generalization benefits. By quantifying the relationship between models and their generating data, this work enables principled evaluation of a model's robustness and disclosure risk.

To formalize relationships between the two spaces, the thesis develops a family of probabilistic metric space constructions tailored to different aspects of the data–model interaction. The first contribution models database evolution as a Markov process and defines probabilistic distances between models based on the likelihood of transitioning between their generating datasets. The second contribution introduces F-space, a framework based on fuzzy measures that captures richer structural properties of the data—such as redundancy, synergy, and overlap among subsets. Building on this, the third contribution applies the F-space theory in practical machine learning scenarios. It demonstrates how fuzzy measures can be used to compare different linear regression algorithms trained over structured subsets of real datasets. The final contribution further generalizes the framework through Generalized F-spaces, where the model space itself is endowed with probabilistic structure—allowing uncertainty in both the datasets and the model outputs to be captured simultaneously.

Together, these constructions offer a principled alternative to traditional model comparison metrics. Rather than relying solely on pointwise loss or accuracy, the proposed framework incorporates the diversity, dynamics, and internal structure of the data that underlies each model—enabling more robust and privacy-aware assessments.

Sammanfattning

Maskininlärningsmodeller formas i grunden av den data de tränas på. Att förstå relationen mellan datamängder och de modeller som genereras från dem är avgörande för uppgifter såsom modellval, sekretessmätningar och robusthetsanalys. Denna avhandling presenterar ett rigoröst matematiskt ramverk för att jämföra maskininlärningsmodeller och algoritmer genom att formalisera samspelet mellan två grundläggande omfång: databasrummet, som representerar möjliga datamängder, och modellrummet, som innehåller de modeller eller hypoteser som härrör från dessa datamängder.

Ett centralt motiv är observationen att olika datamängder kan leda till samma eller mycket liknande modeller. Sådana återkommande modeller — som ofta uppstår över varierande datakällor — är särskilt betydelsefulla i integritetskänsliga tillämpningar. Återkommandet antyder ett minskat beroende av enskilda datapunkter eller undergrupper, vilket ger fördelar vad gäller både integritet och generaliserbarhet. Genom att kvantifiera relationen mellan modeller och deras genererande data möjliggör detta arbete en principbaserad utvärdering av en modells robusthet och risk för avslöjande.

För att formalisera relationen mellan de två omfången introducerar avhandlingen en familj av probabilistiska metriska rum, anpassade för olika aspekter av samspelet mellan data och modeller. Det första bidraget modellerar databasers utveckling som en Markovprocess och definierar probabilistiska avstånd mellan modeller baserat på sannolikheten att övergå mellan deras genererande datamängder. Det andra bidraget introducerar F-rum (F-space), ett ramverk baserat på fuzzy-mått som fångar rikare strukturella egenskaper hos data— såsom redundans, synergi och överlappning mellan delmängder. Det tredje bidraget tillämpar F-rum-teorin i praktiska maskininlärningsscenarier. Det visar hur fuzzy-mått kan användas för att jämföra olika linjära regressionsalgoritmer tränade på strukturerade delmängder av verkliga datamängder. Det fjärde och sista bidraget generaliserar ramverket ytterligare genom Generaliserade F-rum, där även modellrummet ges en probabilistisk struktur — vilket möjliggör att osäkerhet i både datamängden och modellutdata fångas samtidigt. Tillsammans erbjuder dessa konstruktioner ett principiellt alternativ till traditionella jämförelsemått för modeller. I stället för att enbart förlita sig på punktvisa fel eller noggrannhet beaktar det föreslagna ramverket datans mångfald, dynamik och inre struktur — vilket möjliggör mer robusta och integritetsmedvetna analyser.

Preface

This thesis is based on the following papers:

Paper I — Vicenç Torra, **Mariam Taha**, and Guillermo Navarro-Arribas. *The space of models in machine learning: using Markov chains to model transitions.* Progress in Artificial Intelligence, **10**, 321–332 (2021).
https://doi.org/10.1007/s13748-021-00242-6

Paper II — Yasuo Narukawa, **Mariam Taha**, and Vicenç Torra. *On the definition of probabilistic metric spaces by means of fuzzy measures.* Fuzzy Sets and Systems, **465**, Article 108528 (2023).
https://doi.org/10.1016/j.fss.2023.108528

Paper III — **Mariam Taha** and Vicenç Torra. *Measuring the distance between machine learning models using F-space.* In: Proceedings of the 13th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2023) and the 12th International Summer School on Aggregation Operators (AGOP 2023), Palma de Mallorca, Spain, September 4–8, 2023.
Nominated for Best Student Paper Award

Paper IV — **Mariam Taha** and Vicenç Torra. *Generalized F-spaces through the lens of fuzzy measures.* Fuzzy Sets and Systems, **507**, Article 109317 (2025).
https://doi.org/10.1016/j.fss.2025.109317

The thesis was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by Knut and Alice Wallenberg Foundation.

Acknowledgements

This journey has been one of transformation, both academically and personally. The path to completing a Ph.D. degree is rarely walked alone, and I feel deeply fortunate for the guidance, encouragement, and companionship I received throughout. What follows is a heartfelt expression of gratitude to those who made this experience meaningful, inspiring, and possible.

First and foremost, I would like to express my heartfelt gratitude to my supervisor, **Vicenç Torra**, for his steady guidance, constant support, and humble leadership throughout this journey. Starting my Ph.D. journey in 2020 during global uncertainty, his trust and encouragement gave me clarity and strength when I needed it most. I am also grateful to my co-supervisor, **Lili Jiang**, for making my early transition smoother and for the chance to contribute as a teaching assistant in her course—an experience I quietly valued.

Being the first member of the Privacy-aware transparent decisions group (NAUSICA), I had the privilege of watching it grow from the very beginning. With every new face, something special was added—something that enriched me personally and made the Ph.D. journey feel less solitary. Thank you, **Sudipta**, for your enthusiasm and energy, which always lifted the space. Being my batchmate, we shared many moments—courses, seminars, and chats—that made the journey easier and more connected. Thank you, **Ayush**, for your sharp wit and perfectly timed sarcasm that brought life and laughs to our group discussions. Thank you **Sonakshi**, for your calm presence and wise thoughts. I wish the three of you every success in your upcoming defenses. I am grateful to **Sargam**, for her joyful energy and contagious laughter, I'm especially grateful for her kindness during the course we studied together—at a time when I needed extra support, she took the lead and made sure we crossed the finish line. I am also grateful for our former researchers **Aso**, **Saloni**, **Shekhar**, and **Kayode**— for being part of the roots that shaped this group.

I would also like to thank our postdoctoral researchers for simplifying things, reassuring us that everything would work out in the end. To **Fatemeh**, I am grateful for the cultural connection we shared and the meaningful conversations that always brought a sense of familiarity and comfort. Thank you **Zuzana**— for your invaluable help in proofreading my thesis on such short notice. Your clarity, precision, and generous support made a real difference. And to **Jed**, the latest PhD member, I wish you a fulfilling journey ahead. I would also like to thank our former visiting researchers—**Ezgi**, **Toful**, and **Maria**—for the warmth and knowledge you brought. To **Manuel**, thank you for the quiet support, thoughtful feedback, and for sharing your magic with us during your time in the group.

To **Sonakshi**, I want to thank you once again—this time not just for your wisdom, but for the companionship we shared through the final year. From step-by-step processes to deep discussions, your support and advice were incredibly grounding. Sharing this final stretch with you made the journey less daunting and far more meaningful.

Throughout these four years, I'm grateful for the international spirit and joyful moments that made NAUSICA special. When I look back, it's often the small things I remember most—the daily routines, quiet conversations, and spontaneous laughter that brought a sense of rhythm and belonging to our shared days. With members from across the world, we shared not just research, but cultures, creativity, and laughter. From fika breaks and recipe swaps to lighthearted talents and souvenirs—thank you for the warmth and joy you brought. Your presence truly lightened the journey. I also appreciated your thoughtful planning that made it easier for me to stay involved despite other responsibilities. I would like to extend my gratitude to the people behind the scenes—IT support, HR, and administrative staff—whose efficiency, kindness, creativity, and patience made our lives easier, even when their work often went unseen.

In a world that often reduces success to academic achievements, my family reminded me that life is multidimensional. Balancing various aspects alongside research was challenging, but it added depth to the journey and helped me see things from new perspectives. To my mother—no words are enough. Your love and sacrifices are the foundation of everything I am. To my sisters—my lifelong safe space, thank you for always being there and for listening through it all. To my children—thank you for the joy you bring and your sweet little hugs, and for loving me even when I wasn't always present. To my husband, **Ahmed**—thank you for being my anchor. You saw the invisible weight and never let me bear it alone. Your belief, patience, and presence carried me through.

I have to admit that my family have always been the core of my support system. From them, I found the balance and resilience to navigate everything that came my way, and the grounding that helped protect my mental space and keep me going. They reminded me that life is more than deadlines and academic pressure. They loved me beyond titles and achievements, and for that, I am endlessly grateful.

Mariam Taha
April, 2025.

Contents

Chapter 1

Introduction

Machine learning has experienced rapid advancements over the past few decades, establishing itself as a transformative field within computer science. Driven by increased computational power, data availability, and algorithmic innovation, machine learning has found widespread applications across various domains, including healthcare, finance, and environmental science [JM21; LBH15]. For example, in healthcare, machine learning assists in medical diagnosis and predictive analytics [Est+17]; in finance, it powers fraud detection systems and algorithmic trading [HPW17]; and in environmental science, it contributes to climate modeling and resource optimization [Rei+19].

At its core, machine learning involves leveraging data to extract patterns and build models that generalize from past observations [Mur12]. Given a dataset, the objective is to develop models capable of making accurate predictions or informed decisions. Various machine learning models address different problem domains, including decision trees [Qui96], support vector machines [CV95], neural networks [GBC16], and ensemble techniques such as random forests [Bre01] and gradient boosting machines [Fri01].

Traditionally, model selection in machine learning has been primarily based on accuracy metrics, where models are evaluated based on their predictive performance on test datasets [Bis06]. However, recent advancements and societal considerations have broadened evaluation criteria beyond accuracy alone. Modern machine learning models must also be explainable, interpretable, unbiased, and privacy-preserving [DK17; Lip18]. Explainability refers to the ability to understand and articulate how a model arrives at its decisions, which is crucial for trust and regulatory compliance in high-stakes applications. Interpretability ensures that model outputs are human-comprehensible and can be linked to domain knowledge, while bias-free models aim to prevent unfair or discriminatory outcomes [Meh+21].

One of the most pressing concerns in machine learning is data privacy. As machine learning models increasingly rely on large-scale datasets containing sensitive information—ranging from medical records and financial transactions to behavioral data—the risk of exposing private details grows. Moreover, most real-world data are dynamic and subject to regular updates. This dynamic nature affects the consistency of aggregations and inferences drawn from the data unless models are continuously updated. For example, a machine learning model built on an evolving data source must be regularly updated to stay aligned with its underlying dataset. Changes in training data can lead to model transformations, and an adversary with access to auxiliary information might exploit these changes to infer sensitive details [Sal+19; TN16].

Among the key privacy-preserving models that have been developed are differential privacy and integral privacy. Differential privacy ensures that the inclusion or exclusion of any individual data point in a dataset does not significantly alter the model's output [Dwo06]. Integral privacy [TN16], on the other hand, emphasizes generating recurrent models from distinct datasets to prevent any single dataset from being the sole source of learning.

1.1 Motivation

Despite the growing emphasis on privacy, interpretability, and fairness in machine learning, existing approaches often overlook the fundamental relationship between datasets and the models they produce. As data evolves, machine learning models must be updated accordingly, raising critical questions about privacy risks, model stability, and selection criteria. Understanding how dataset changes influence model behavior is essential for designing robust and privacy-preserving machine learning systems. Another crucial yet underexplored aspect is the comparison of models concerning the similarity of the databases that have generated them. Our aim is to provide tools to analyse the relationship between the space of data and the space of models, specifically in the context of privacy-preserving machine learning models. To the best of our knowledge, aside from [TN18], this aspect has not been explored in the literature, highlighting the need for further research in this direction to better understand model similarities in relation to the data that generate them.

Machine learning models are inherently dependent on data, which evolves over time. This creates a fundamental interaction between two spaces: the database space (the space of datasets) and the model space (the space of trained machine learning models). Changes in datasets—whether due to new information, updates—impact model construction, necessitating model updates. Understanding this interaction is crucial for several reasons:

- Privacy Considerations: If a privacy-preserving model undergoes updates due to dataset modifications, it is essential to ensure that these updates do not inadvertently reveal sensitive information.

- Model Stability and Robustness: Small changes in training data can lead to significant shifts in model behavior. Quantifying these shifts helps assessing the reliability and robustness of machine learning models.

- Model Selection: Machine learning can be seen as a selection process, where the goal is to choose models that achieve high accuracy, avoid overfitting, remain resistant to membership inference attacks, and exhibit similarity to models generated from related datasets.

A key challenge in addressing these concerns is establishing a theoretical framework for comparing machine learning models and algorithms while explicitly accounting for the datasets on which they are trained. Traditional similarity measures—such as comparing model parameters, architectures, or performance metrics—fail to capture the nuanced impact of dataset evolution on model behavior. Instead, a principled approach is needed that explicitly integrates the structure of both the database space and the model space to quantify model relationships meaningfully.

1.2 Approach: Probabilistic Metric Space

To analyze the relationship between data and models and to compare models and algorithms, it is essential to define distances and metrics for the spaces. Since model comparison is based on the sets of generators, these metrics must be defined on sets rather than on individual elements.

Metric spaces [Fré06] provide a rigorous mathematical foundation for defining distances, consisting of a non-empty set and a distance function (or metric) that satisfies three fundamental properties: non-negativity, symmetry, and the triangle inequality. However, extending a metric from individual elements to sets of elements is not straightforward, as it requires a principled way of aggregating pairwise distances while preserving the essential properties of a metric.

Several set-based distance measures exist, including the Hausdorff distance [Hau14], the sum of minimum distances [Nii87], and the Surjection distance [Odd79]. However, these measures often fail to capture the overall structural relationships within the sets and, as a result, do not satisfy all the properties required for the distance to be a metric. To address this limitation, [EM97] introduced a metric for sets that is based on finding an optimal path between the elements of the two sets, providing a more robust approach to defining distances in set spaces.

Classical metric spaces assign a single numerical value to distances, which may not fully capture the inherent uncertainties in model comparisons. In contrast, probabilistic metric spaces (PMS) [SS83] generalize the notion of a metric by defining distances as distribution functions rather than fixed numbers. The axioms of PMS correspond to those of classical metric spaces, with adaptations to account for uncertainty. In particular, the positive definiteness and symmetry axioms remain, while the triangle inequality is reformulated in terms of a triangle function. This triangle function, which is crucial to PMS, is often constructed using t-norms. T-norms—binary operations that generalize the logical conjunction in fuzzy logic—are defined on the unit interval and satisfy properties such as commutativity, associativity, monotonicity, and having

1 as the neutral element. See, e.g. the reference books by Alsina et al. and Klement et al. [AFS06; KMP00]). The choice of a particular t-norm directly influences the strictness of the triangle inequality, thereby affecting the overall structure of the space.

A noteworthy subclass of PMS is the Menger space, where the triangle function is directly induced by a t-norm. In Menger spaces, the generalized triangle inequality is enforced in a manner that closely parallels the classical metric case. This concept was originally introduced by Menger [Men42] and later developed by Schweizer and Sklar [SS83], providing a rigorous framework for modeling distances under uncertainty.

Probabilistic metric spaces provide a natural framework for model comparison by encoding distances as distribution functions rather than fixed numerical values. This formulation inherently captures uncertainty in the distance measure. Moreover, when models are generated from datasets, the uncertainty embedded in the datasets can be inherited by the PMS framework, thereby offering a comprehensive tool for comparing models under realistic conditions.

1.3 Research Questions and Problem

The central objective of this thesis is to establish a theoretical framework that formalizes the interaction between the database space and the model space (see Figure 1.1), providing mathematical tools for quantifying algorithmic and model distances.

We formulate our problem as follows: Let Ω be the space of databases (the base space), and let G denote the set of algorithms, where each algorithm $g \in G$ maps elements from Ω to the model space M (the target space). For any given model $m \in M$, let $Gen(m)$ represent the set of all databases that can generate m. The objective is to compare models m_1, m_2 and algorithms g_1, g_2 by constructing distances based on the sets of datasets that produce them. In other words distances based on $Gen(m_1)$ and $Gen(m_2)$.

This research is guided by the following key questions:

RQ1: How can models m_1 and m_2 in M be compared while accounting for transformations in the database space Ω?

RQ2: How can we construct distances and metrics for machine learning algorithms in G that capture complex interactions of the databases?

RQ3: Which characterizations can be provided for the metrics we propose?

1.4 Thesis Contributions

To address (RQ1), we utilize Markov chains and transition matrices to model transformations within the database space. Specifically, we introduce two def-

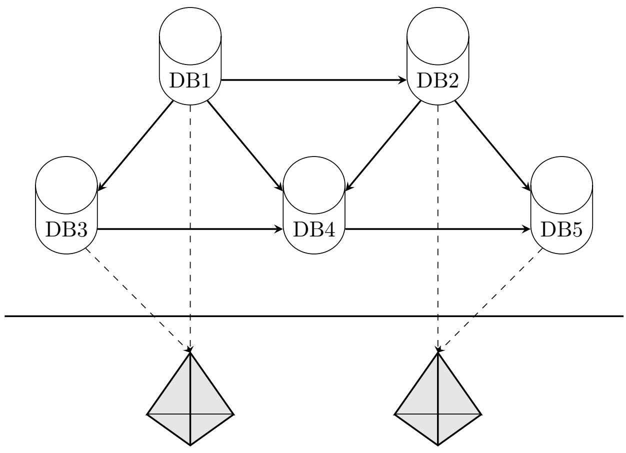

Figure 1.1: Graphical representation of databases (cylinders) and machine learning models (3D pyramids)

initions of probabilistic metric spaces for databases, both grounded in transition matrices and Markov chains. The first definition quantifies the distance between two databases based on the probability of one being transformed into the other. This formulation constructs the probabilistic metric space exclusively from the transition matrices. We present both symmetric and asymmetric definitions for the distance distribution functions, providing a structured approach to measuring database similarity. We refer to this type of space as the Visited Database-Based Probabilistic Metric Space (VD-PMS). The second definition, in contrast, evaluates the distance between two databases in terms of their evolution over time. Instead of considering direct transformations, this approach examines whether two databases will exhibit similarity as time progresses. We term this approach Database Distance-based Probabilistic Metric Space (DD-PMS) These resulting metrics are then extended to define distances between models (Paper 1).

To address (RQ2), we introduce (in Paper 2) a specialized type of probabilistic metric space, called F-space, where the base space is structured as a measurable space using fuzzy measures. This framework enables the modeling of dependencies and interactions within datasets, allowing for a more nuanced representation of data relationships. F-space facilitates the computation of distances between functions and algorithms that map from the base space to the target space. Specifically, it evaluates sets of elements whose distances do not exceed a given threshold when mapped to the target space. We demonstrate (in Paper 3) how to apply F-space in machine learning using the Sugeno λ-measure. To answer (RQ3) for F-spaces, we analyze the properties of the measure on the database space that yields a probabilistic metric space satisfying a generalized triangle inequality. Our theoretical results found in (Paper

2) demonstrate that when employing a stronger t-norm, fewer constraints are imposed on the measure to ensure that the space retains the desired metric properties.

In our initial formulation, models in F-space were represented in a metric space with fixed distances. However, practical machine learning scenarios demand a framework that captures uncertainties arising from dataset shifts, noise, and randomness in training processes. To address this, we extend the F-space construction by generalizing the target space from a metric space to a probabilistic metric space (PMS). This extension allows distances to be represented as distribution functions, thereby capturing the inherent uncertainty and variability present in real-world machine learning scenarios. We characterize mathematically the conditions under which the induced probabilistic metric space satisfies a generalized triangle inequality. In particular, we analyze how properties of the underlying measure defined on the database space (e.g., supermodularity, inclusion-based properties) interact with the choice of triangular norms (t-norms), and when it can yield to a proper Menger space. Therefore, we answer for these generalized F-spaces (RQ3).

More concretely, our theoretical results in (Paper 4) confirm that employing a stronger t-norm relaxes the constraints imposed on the underlying measure while still ensuring that the space retains its desired metric properties. These findings reveal a clear relationship between the type of probabilistic metric space and the t-norm used: the stronger the t-norm, the fewer restrictions are required on the measure. For instance, when using the drastic t-norm, the measure only needs to satisfy minimal conditions, whereas employing a less strict t-norm, such as the minimum, necessitates more constraints on the measure (e.g., unanimity measure). Overall, our results are fully consistent with the previous framework (F-space), collectively demonstrating that a careful selection of t-norms allows for a broader class of measures to be used without compromising the generalized metric structure.

We provide practical validation of our theoretical findings by applying the results to the comparison of machine learning algorithms (RQ2). By modeling the database space as a measurable space equipped with fuzzy measures and the model space as a PMS, our approach allows for a more nuanced assessment of model similarities that accounts for complex interactions among datasets and uncertainties arising from factors such as dataset shifts, noise, and randomness in training.

1.5 Thesis Outline

The remainder of this thesis is structured as follows: Chapter 2 explores the foundational principles of probabilistic metric spaces, their mathematical structures, and their connections to classical metric spaces. Chapter 3 presents fuzzy sets and their properties, followed by fuzzy measures, including Sugeno λ-measures for non-additive aggregation. Chapter 4 presents the main research contributions, while Chapter 5 concludes the thesis and outlines potential directions for future work. The appendix includes four research papers that form the basis of this work.

Chapter 2

Probabilistic Metric Space

The 19th century marked the beginning of the modern scientific era, characterized by significant advancements in measurement and mathematical formalism. These developments paved the way for major breakthroughs but also highlighted the inevitable presence of errors in measurement processes. Early in the 20th century, it was believed that meticulous design and large datasets could reduce measurement errors to arbitrarily small levels. However, the advent of quantum mechanics fundamentally challenged this belief. Heisenberg's uncertainty principle [Hei27] demonstrated that uncertainties are intrinsic to the measurement process itself and cannot be completely eliminated. This marked a paradigm shift, revealing the limitations of determinism and introducing probabilistic methods as an essential framework.

By the mid-20th century, the recognition of inherent uncertainties became a central theme across disciplines such as psychometrics, communication theory, and pattern recognition [Sha48; Bri56; DH73]. This perspective profoundly influenced mathematical frameworks like cluster analysis and interval analysis [Jan78; She80]. Despite these advancements, many mathematical models continued to assume idealized, rigid reference frames for measurements, overlooking the distributed nature of uncertainties in real-world systems.

The term "metric" originates from the Greek word *metron*, meaning "measure". The modern concept of metric spaces was introduced by Maurice Fréchet in his seminal Ph.D. thesis [Fré06]. This work laid the foundation for systematically quantifying distances in mathematics and science. The concept of an abstract metric space offers a unifying framework applicable to diverse constructs, from points and functions to sets and even subjective experiences like sensations [Blu70]. Metric spaces elegantly associate a non-negative real number with each ordered pair of elements in a set, governed by axioms reflecting the intuitive properties of physical distances.

However, real-world applications often reveal that assigning a single value to represent the distance between two elements is an oversimplification. For instance, measuring physical length frequently involves averaging multiple observations, making it more accurate to treat distance as a statistical measure rather than a deterministic quantity. To address this limitation, K. Menger [Men42] introduced the concept of a statistical metric space in 1942. Instead of defining distance as a single numerical value $d(p, q)$, he proposed a distribution function $F_{p,q}(x)$ that represents the probability of the distance being less than x. This probabilistic generalization allowed for modeling systems with inherent uncertainties, broadening the classical notion of metric spaces to stochastic settings [SS83].

Shortly thereafter, A. Wald [Wal43] critiqued Menger's generalized triangle inequality and proposed an alternative formulation that refined the framework. This alternative formed the basis of a theory of betweenness, offering certain advantages in practical applications. Menger [Men51] expanded the theory with additional examples, further solidifying the foundations of statistical metric spaces and exploring new directions for their application.

This chapter focuses on the foundational aspects of probabilistic metric spaces together with their mathematical structures. It is organized into four main sections: **Menger Space**, **Developments on Probabilistic Metric Spaces**, and **On Some Specific Cases**, and **Random Metric Spaces**. The first section introduces the core definitions and properties of Menger spaces. The second section explores refinements and advancements in PM-Space theory, including critiques and alternative formulations. The third section presents some specific cases and examples of PM-Spaces. The final section introduces E-Spaces, a significant construction within PM-Space theory introduced by Sherwood [She69]. This section outlines the use of measurable functions and probability spaces in defining E-Spaces and highlights their connection to classical metric spaces via the Lebesgue measure.

Topics such as fixed-point theory or the topology of probabilistic metric spaces are not included in this chapter, as the focus remains on establishing a foundational understanding and tracing key developments within PM-Space theory.

2.1 Menger Space

The notion of a distance introduced by Frechet was later given the name Metric Space by F. Hausdorff [Hau14] in 1914. Metric space is an ordered pair (S, d) where S is an abstract set and d is a mapping of $S \times S$ into the real numbers.

Definition 1. *Let* $d : S \times S \to \mathbb{R}_0^+$, *then* d *is called a metric on* S *if the following properties hold for* $a, b, c \in S$*:*

(1) $d(a, b) \geq 0$ *with equality if and only if* $a = b$ *(non-negativity property),*

(2) $d(a, b) = d(b, a)$ *(symmetry property), and*

(3) $d(a, b) \leq d(a, c) + d(c, b)$ *(triangle inequality property).*

When the distance function does not satisfy the symmetry condition, the space (S, d) is called a quasimetric space. If the distance function does not fulfill the triangle inequality, the space (S, d) is identified as a semimetric space.

When defining probabilistic metric space (briefly PM-Space), we use the notion of distribution functions as follows.

Definition 2. *A* distribution function *is a non-decreasing function* F *defined on* $\mathbb{R}$, *satisfying* $F(-\infty) = 0$ *and* $F(+\infty) = 1$.
If F *is defined on* $\mathbb{R}^+$ *and satisfies the following conditions:*

- $F(0) = 0$, $F(\infty) = 1$,

- F *is left-continuous on* $(0, \infty)$,

then F *is referred to as a* distance distribution function.

Distribution functions are commonly associated with probabilities, where $F(x)$ represents the probability that the distance between two elements p and q is less than or equal to x for $p, q \in S$. We denote the distribution function of F for p and q as $F_{p,q}$. The collection of all distance distribution functions is denoted by Δ^+, while the distance distribution function corresponding to a classical distance equal to a is denoted by ϵ_a and is defined as follows:

$$\epsilon_a(x) = \begin{cases} 0, & 0 \leq x \leq a, \\ 1, & a < x \leq \infty. \end{cases} \tag{2.1}$$

To generalize the metric space to a *probabilistic metric space*, the function F must satisfy the following properties:

(1) $F_{p,q}(x) \leq F_{p,q}(y)$ whenever $x \leq y$.

(2) If $p = q$, then $F_{p,q}(x) = 1$ for all $x > 0$.

(3) If $p \neq q$, then $F_{p,q}(x) \leq 1$ for some $x > 0$.

(4) $F_{p,q} = F_{q,p}$.

Although distribution functions effectively generalize the first two axioms of metric spaces, extending axiom (3) in Definition 1 poses challenges. This has led to the study of alternative triangle inequalities, which remain a central topic in the development of probabilistic metric spaces.

The weakest generalization of the triangle inequality is the one given by Schweizer and Skalar [SS83], defined as follows:

Definition 3. *A* PM-space *is an ordered pair* (S, F)*, where* S *is a non-empty set and* F *is a map* $F : S \times S \to \Delta^+$*, satisfying the following properties:*

PM-1 $F_{p,q}(x) = 1 \; \forall x > 0 \iff p = q$*,*

PM-2 $F_{p,q}(0) = 0$*,*

PM-3 $F_{p,q} = F_{q,p}$*,*

PM-4 *If* $F_{p,q}(x) = 1$ *and* $F_{q,r}(y) = 1$*, then* $F_{p,r}(x + y) = 1$*.*

Inequality **PM-4** means that If it is certain that the distance of p and q is less than x, and likewise certain that the distance of q and r is less than y, then it is certain that the distance of p and r is less than $x + y$.

Remark 1. *In view of the condition **PM-2**, which obviously implies that* $F_{p,q}(x) = 0$ *for* $x \leq 0$*, condition **PM-2** is equivalent to the statement*

$$p = q \Leftrightarrow F_{p,q} = \epsilon_0.$$

Remark 2. ***PM-4*** *can be seen as a minimum generalization of the ordinary triangle inequality, however, it is vacuous in all space in which functions* $F_{p,q}$*, for* $p \neq q$ *never attains the value 1.*

Remark 3. *Every Metric space can be viewed as PM space if we set*

$$F_{p,q}(x) = \epsilon_{d(p,q)}(x)$$

The condition **PM-4** of the above definition is always satisfied in metric space where it reduces to ordinary triangle inequality.

In 1942, Menger [Men42] introduced a generalization of the triangle inequality and defined a *statistical metric space* as a set S equipped with a family of distribution functions F. He formulated the generalized triangle inequality, also known as the *Menger triangle inequality*, as follows $\forall p, q, r \in S$ and $\forall x, y \in \mathbb{R}$:

$$\textbf{PM-5:} \quad F_{p,r}(x + y) \geq T(F_{p,q}(x), F_{q,r}(y))$$

Where T is a binary operation that satisfies the conditions described below.

Definition 4. *Let* $T : [0, 1] \times [0, 1] \to [0, 1]$*, be a binary operation, where* T *satisfies the following properties for all* $a, b, c, d \in [0, 1]$*:*

T-1 $0 \leq T(a, b) \leq 1$*,*

T-2 $T(a, b) = T(b, a)$*,*

T-3 $T(a, b) \leq T(c, d)$ *whenever* $a \leq c$ *and* $b \leq d$*,*

T-4 $T(1,1) = 1$,

T-5 $T(a,1) > 0 \; \forall a > 0$.

Definition 5. *A triangle inequality is said to hold universally in a PM-space iff it holds for all triples of points, not necessarily distinct.*

Menger's triangle inequality suggests that our knowledge of the third side of a triangle depends monotonically on the probabilities of the other two sides. This interpretation can be further clarified by specifying T as a particular function. Below, there are some examples:

- **Minimum:** $\top(a,b) = \min(a,b)$, denoted by $M(a,b)$.

- **Algebraic product:** $\top(a,b) = ab$, denoted by $\Pi(a,b)$.

- **Bounded difference**
  $\top(a,b) = \max(0, a+b-1)$, denoted by $W(a,b)$.

- **Maximum:** $\top(a,b) = \max(a,b)$, denoted by $M^*(a,b)$.

For example, if we choose $T = \Pi$, then the probability that the distance from p to r is less than $x + y$ is at least as large as the joint probability, independently for the distance from p to q is less than x, and the distance from q to r is less than y.

Remark 4. *Given condition **T-4**, it can be observed that **PM-5** encompasses condition **PM-4** as a special case.*

A metric space emerges as a specific case of a Menger space when d is defined as a function from $S \times S \to [0, \infty)$, such that:

$$F_{p,q}(x) = \epsilon_{d(p,q)} \tag{2.2}$$

Therefore, the Menger space (S, F, T) with the above definition of F, is a classical metric space.
Naturally, we have to prove only the classical triangle inequality, since all other properties hold trivially. If we have $p, q, r \in S$ with $d(p,q) < x$ and $d(q,r) < y$, for some $x, y > 0$ then from Equation 2.2 it follows that $F_{p,q}(x) = 1$ and $F_{q,r}(y) = 1$. Then by **PM-5** and the boundary property of T, we obtain $d(p,r) < x + y$, which gives the desired inequality. If we begin with a metric space (S, d), then by taking $F_{p,q}$ as defined in 2.2 we find the functions $F_{p,q}$ are probability distribution functions satisfying all the conditions in Definition 3 and **PM-5** under any T.

2.2 Developments on Probabilistic Metric Space

Schweizer and Sklar further refined Menger's triangle function through the introduction of t-norms, which were inspired by three foundational results. These advancements provided a more generalized and flexible framework for modeling probabilistic metric spaces.

Lemma 1. *If a PM-space contains two distinct points, then the condition* ***PM-5*** *can not hold universally in the space under the choice* $T = M^*$.

Proof. Let p and q be two distinct points of space and let x and y satisfy $0 < y < x$. Suppose that **PM-5** holds universally with $T = M^*$. Then we have

$$F_{p,q}(x) \geq max(F_{p,q}(x-y), F_{q,q}(y)) = 1.$$

Since x can be any positive number, condition **PM-1** implies $p = q$, which contradicts the assumption.

□

Lemma 2. *If in a nonmetric PM-space,* ***PM-5*** *holds universally for some choice of* T *satisfying the conditions* ***T-1*** *to* ***T-5****, then the function* T *has the property that there exists a number* $a, 0 < a < 1$*, such that* $T(a, 1) \leq a$.

Proof. If PM-space is not a metric space, then there is a point $(p, q) \in S \times S$ in which $F_{p,q}$ assumes values other than 0 or 1. Since F is left continuous and monotonic, this means there is an open interval (x, y) on which $0 < F_{p,q} < 1$. Now, let us assume that $T(a, 1) = a + \Phi(a)$, where $\Phi(a) \geq 0$, for $0 < a < 1$. Let $z \in (x, y)$ and take $t > 0$. Then we have

$$F_{p,q}(z+t) \geq T(F_{p,q}(z), F_{q,q}(t)) = T(F_{p,q}(z), 1) = F_{p,q}(z) + \Phi(F_{p,q}(z))$$

Now, let $t \to 0$, we have

$$F_{p,q}(z) \geq F_{p,q}(z) + \Phi(F_{p,q}(z)) \geq F_{p,q}(z)$$

This proves discontinuity of $F_{p,q}$ at z, and therefore at every point of (x, y). However, this is a contradiction as non-decreasing functions can be discontinuous at only denumerable points. This ends the proof. □

Theorem 3. *If a PM-space* (S, F, T) *where*

(1) ***PM-5*** *holds universally.*

(2) T *is continuous, then for any* $x > 0$

$$T(F_{p,q}(x), 1) \leq F_{p,q}(x)$$

Proof. Let $p, q \in S$, and let $x > 0$ be given. Choose y such that $0 < y < x$. Then, we have:

$$F_{p,q}(x) \geq T(F_{p,q}(x-y), F_{q,q}(y)) = T(F_{p,q}(x-y), 1).$$

Letting $y \to 0^+$, we obtain:

$$F_{p,q}(x) \geq \lim_{y \to 0^+} T(F_{p,q}(x-y), 1).$$

Using the assumed continuity of $F_{p,q}$, we can write:

$$\lim_{y \to 0^+} T(F_{p,q}(x-y), 1) = T\left(\lim_{y \to 0^+} F_{p,q}(x-y), 1\right).$$

By the left continuity of $F_{p,q}$, we know:

$$\lim_{y \to 0^+} F_{p,q}(x-y) = F_{p,q}(x).$$

Substituting this back, we find:

$$\lim_{y \to 0^+} T(F_{p,q}(x-y), 1) = T(F_{p,q}(x), 1).$$

Therefore, $F_{p,q}(x) \geq T(F_{p,q}(x), 1)$. This completes the proof. □

Motivated by these lemmas and the observation that there are three weak functions in T satisfying $T(a, 1) = a$, Sklar and Schweizer [SS83] redefined the concept of triangle inequality in Definition 4, introducing what is now known as t-norms. In this redefinition, the conditions **T-1**, **T-4**, and **T-5** are replaced with the following:

- **T-6** : $T(a, 1) = a$ and $T(0, 0) = 0$.
- **T-7** : The associative condition, $T(T(a, b), c) = T(a, T(b, c))$.

This modification enables the extension to a polygonal inequality. With these adjustments, we are now in a position to introduce the following definition.

Definition 6. *A triangular norm (t-norm) is a binary function* $T : [0, 1] \times [0, 1] \to [0, 1]$ *that satisfies* ***T-2****,* ***T-3****,* ***T-6****, and* ***T-7****.*

If, for two t-norms T_1 and T_2, the inequality $T_1(x, y) \leq T_2(x, y)$ holds $\forall (x, y) \in [0, 1]^2$, then we say T_1 is weaker than T_2 or, equivalently, that T_2 is stronger than T_1.

Remark 5. *The strongest t-norm is the minimum t-norm,* $T(a, b) = \min(a, b)$*. On the other hand, the weakest t-norm is the drastic product, defined as:*

$$T_D(a, b) = \begin{cases} \min(a, b), & \textit{if } \max(a, b) = 1, \\ 0, & \textit{otherwise.} \end{cases}$$

Remark 6. *In respect of a t-norm T, an element $x \in [0,1]$ with $T(x,x) = x$ is called an idempotent element of T. It is immediate that 0 and 1 are idempotent elements (which are termed as trivial idempotent elements) for every t-norm.*

In addition to t-norms, another fundamental concept in fuzzy set theory and probabilistic metric spaces is the **triangular conorm (t-conorm)**, which serves as the dual operation to t-norms.

Definition 7. *A triangular conorm (t-conorm), also called an s-norm, is a binary function $S : [0,1] \times [0,1] \to [0,1]$ that satisfies the following conditions:*

- ***S-1**: Commutativity, $S(a,b) = S(b,a)$.*
- ***S-2**: Associativity, $S(S(a,b),c) = S(a, S(b,c))$.*
- ***S-3**: Monotonicity, $a \leq b \Rightarrow S(a,c) \leq S(b,c)$.*
- ***S-4**: Boundary conditions, $S(a,0) = a$ and $S(1,1) = 1$.*

Similar to t-norms, if for two t-conorms S_1 and S_2, the inequality $S_1(x,y) \geq S_2(x,y)$ holds for all $(x,y) \in [0,1]^2$, we say that S_1 is stronger than S_2 (or equivalently, S_2 is weaker than S_1).

Remark 7. *The strongest t-conorm is the maximum t-conorm, given by $S(a,b) = \max(a,b)$. The weakest t-conorm is the probabilistic sum, defined as:*

$$S_P(a,b) = a + b - ab.$$

Having established the fundamental operations of t-norms, we now revisit the concept of Menger spaces, incorporating t-norms as the underlying function for probabilistic metric structures.

Definition 8. *A Menger space is a probabilistic metric space (PMS) in which the condition **PM-5** holds universally for a function T that satisfies **T-2**, **T-3**, **T-6**, and **T-7**.*

The following lemma establishes that, in determining whether a PM-space is a Menger PM-space, it is sufficient to consider only triplets of distinct points.

Lemma 4. *If the points p, q, r are not all distinct, then the condition **PM-5** holds for the triple p, q, r under any choice of T satisfying **T-2**, **T-3**, **T-6**, and **T-7**.*

Proof. We only need to consider the case where $T = \min$.
- If $p = r$, then $F_{p,r} = \epsilon_0$, and the conclusion is immediate.
- If $p = q \neq r$, then for any $x, y \geq 0$, we can write:

$$\min(F_{p,q}(x), F_{q,r}(y)) = \min(\epsilon_0(x), F_{q,r}(y)) \leq F_{q,r}(y),$$

and since $F_{q,r}(y) \leq F_{q,r}(x+y)$, it follows that:

$$\min(F_{p,q}(x), F_{q,r}(y)) \leq F_{q,r}(x+y).$$

Hence, the condition holds. □

The other triangle inequality, attributed to Wald [Wal43], is introduced below.

Definition 9. *Wald triangle inequality is defined as:*

$$\textbf{\textit{(PM-4)}} \quad F_{p,r}(x) \geq [F_{p,q} * F_{q,r}](x), \quad \forall x \geq 0,$$

where $*$ denotes convolution. Specifically,

$$[F_{p,q} * F_{q,r}](x) = \int_{-\infty}^{+\infty} F_{p,q}(x-y)\, dF_{q,r}(y).$$

Since $F_{p,q}(x-y) = 0$ for $y \geq x$ and $F_{q,r}(y) = 0$ for $y \leq 0$, we may evidently write

$$[F_{p,q} * F_{q,r}](x) = \int_{0}^{x} F_{p,q}(x-y)\, dF_{q,r}(y)$$

as the convolution of the distribution functions of two independent random variables gives the distribution function of their sum.

Definition 10. *A PM space (S, F, T) where T is a convolution is called a Wald space.*

Using the equality $\epsilon_a * \epsilon_b = \epsilon_{a+b}$ and $F_{p,q} = \epsilon_{d(p,q)}$ where $d : [S \times S] \to [0, \infty)$, one can show the triple $(S, F, *)$ is a Wald space if and only if (S, d) is the classical metric space.

Theorem 5. *Every Wald space is a Menger PM-space under the choice $T = \Pi$.*

Proof. In a Wald space, for any $x, y \geq 0$, we have:

$$F_{p,r}(x+y) \geq \int_0^{x+y} F_{p,q}(x+y-z)\, dF_{q,r}(z).$$

Expanding the expression:

$$F_{p,r}(x+y) \geq \int_0^{x+y} \left[\int_0^{x+y-z} dF_{p,q}(t) \right] dF_{q,r}(z).$$

Using Fubini's theorem:

$$F_{p,r}(x+y) \geq \iint_{t,z \geq 0, t+z \leq x+y} dF_{p,q}(t)\, dF_{q,r}(z).$$

Now, observe that:

$$\iint_{t,z \geq 0, t+z \leq x+y} dF_{p,q}(t)\, dF_{q,r}(z) \geq \iint_{0 \leq t \leq x, 0 \leq z \leq y} dF_{p,q}(t)\, dF_{q,r}(z),$$

since $\{(t,z) \mid 0 \leq t \leq x, 0 \leq z \leq y\} \subset \{(t,z) \mid t,z \geq 0, t+z \leq x+y\}$ and F is non-decreasing.

Now, for the integral over the subset:

$$\iint_{0 \leq t \leq x, 0 \leq z \leq y} dF_{p,q}(t)\, dF_{q,r}(z) = \int_0^x \int_0^y dF_{p,q}(t)\, dF_{q,r}(z).$$

Simplifying further:

$$\int_0^x \int_0^y dF_{p,q}(t)\, dF_{q,r}(z) = \int_0^x dF_{p,q}(t) \cdot \int_0^y dF_{q,r}(z) = F_{p,q}(x) \cdot F_{q,r}(y).$$

Therefore, by combining the inequalities, we obtain:

$$F_{p,r}(x+y) \geq F_{p,q}(x) \cdot F_{q,r}(y).$$

This inequality is indeed **PM-5** under the t-norm product. □

Corollary 1. *If the Wald inequality (**PM-4**) holds, then the inequality **PM-4** also holds.*

Proof. Since a Wald space is a Menger PM-space in which **PM-4** holds, we have:

$$F_{p,r}(x+y) \geq F_{p,q}(x) \cdot F_{q,r}(y).$$

If $F_{p,q}(x) = 1$ and $F_{q,r}(y) = 1$, then:

$$F_{p,r}(x+y) \geq 1 \cdot 1 = 1,$$

which implies $F_{p,r}(x+y) = 1$. Therefore, the inequality **PM-4** holds. □

Lemma 6. *If the points p, q, r are not all distinct, then the condition **PM-4** holds for the triple p, q, r.*

Proof. Consider the following cases:

- If $p = r$, then $F_{p,r} = \epsilon_0$, and the condition **PM-4** is satisfied.
- If $p = q \neq r$, then for $x \geq 0$:

$$F_{p,r}(x) = F_{q,r}(x) = \int_0^x dF_{q,r}(y).$$

  Expanding further:

$$F_{p,r}(x) = \int_0^x \epsilon_0(x-y)\, dF_{q,r}(y).$$

  Rewriting:

$$F_{p,r}(x) = \int_0^x F_{p,q}(x-y)\, dF_{q,r}(y) \geq [F_{p,q} * F_{q,r}](x).$$

- The case $p \neq q = r$ follows similarly by interchanging p and r.

This concludes the proof. □

Theorem 7. *If in a PM-space, the condition **PM-5** holds for all triples of distinct points under $T = M^*$, then the space is a Wald space.*

Proof. Let p, q, r be distinct points. For any $x \geq 0$, we have:

$$F_{p,r}(x) \geq \max(F_{p,q}(x), F_{q,r}(x)).$$

Expanding further:

$$F_{p,r}(x) \geq \int_0^x dF_{q,r}(y).$$

By the definition of convolution and the fact that $0 \leq F_{p,q}(x-y) \leq 1$, we can write:

$$F_{p,r}(x) \geq \int_0^x F_{p,q}(x-y)\, dF_{q,r}(y).$$

Therefore, the condition **PM-4** holds for all triples of distinct points in the space. This implies that the space is a Wald space. □

In 1962, Šerstnev [Šer63] proposed a generalized triangle inequality that encompasses all previously defined inequalities as special cases. To formally define PM-spaces in the sense of Šerstnev, the notion of a triangle function is introduced as follows:

Definition 11. *A triangle function T is a binary operation on Δ^+ that satisfies the following properties for any $F, G, H, K \in \Delta^+$:*

(1) $T(F, \epsilon_0) = F$*,*

(2) $T(F, G) = T(G, F)$ *(commutativity),*

(3) $T(F, G) \leq T(H, K)$ *whenever* $F \leq H$ *and* $G \leq K$ *(monotonicity),*

(4) $T(T(F, G), H) = T(F, T(G, H))$ *(associativity).*

Definition 12. *A triangle function* T_1 *is stronger than a triangle function* T_2 *if for all* $F, G \in \Delta^+$*, and all* $x \in \mathbb{R}^+$*,* $T_2(F, G)(x) \leq T_1(F, G)(x)$*.*

Example 1. *Let* T *be a left continuous t-norm. Then the function* $T : \Delta^+ \times \Delta^+ \to \Delta^+$ *defined by*

$$T(F, G)(x) = T(F(x), G(x))$$

is a triangle function.

Example 2. *The maximal triangle function* $T_M(F, G)(x) = min(F(x), G(x))$*. For any triangle function* T *we have:*

$$\begin{aligned} T(F, G) &\leq T(F, \epsilon_0) = F, \\ T(F, G) &\leq T(\epsilon_0, G) = G, \\ &\textit{Hence} \\ T(F, G)(x) &\leq min(F(x), G(x)) = T_M(F, G)(x). \end{aligned}$$

Example 3. *If* T *is a left-continuous t-norm, then* T_τ *defined by*

$$T_\tau(F, G)(x) = \sup\{T(F(u), G(v)) \mid u + v = x\}$$

is a triangle function.

Definition 13. *Let* (S, F, T) *be a PM-space where*

$$T_\tau(F, G)(x) = \sup\{T(F(u), G(v)) \mid u + v = x\}.$$

Then (S, F, T) *is called a Menger space, which will be denoted by* (S, F, T) *in the sequel.*

Remark 8. *If the t-norm* T *is left-continuous, then* T_τ *in Definition 13 is a triangle function. Thus, we have*

$$F_{p,r}(x + y) \geq T(F_{p,q}(x), F_{q,r}(y)) \quad \forall p, q, r \in X \text{ and } x, y \in \mathbb{R}.$$

Definition 14. *A PM-space (in the sense of Šerstnev [Šer63]) is a triple* (S, F, τ)*, where:*

– S *is a non-empty set,*

- $F : S \times S \to \Delta^+$,
- τ *is a triangle function,*

such that the following conditions are satisfied for all $p, q, r \in X$*:*

(1) $F_{p,p} = \epsilon_0$*,*

(2) $F_{p,q} \neq \epsilon_0$ *for* $p \neq q$*,*

(3) $F_{p,q} = F_{q,p}$ *(symmetry),*

(4) $F_{p,r} \geq \tau(F_{p,q}, F_{q,r})$ *(**PM-6**, triangle inequality).*

Given a probabilistic metric space (S, F, τ), we say that (S, F) is a probabilistic metric space under τ. A probabilistic pseudometric space (PPM space) (S, F, τ) is defined as above but not requiring condition (2). When all conditions above apply but (4) is not required we have a probabilistic semimetric space. When all conditions apply but condition (3) is not required we have a probabilistic quasimetric space.

Remark 9. *If* $\tau(\epsilon_a, \epsilon_b) \geq \inf\{\epsilon_c \mid c < a+b\}$ *for all* $a, b > 0$*, then the inequality reduces to **PM-4**. If* τ *is a convolution, then **PM-6** reduces to **PM-4**.*

Definition 15. *Let* (S, F, τ) *be a PM-space. The space* (S, F, τ) *is called **proper** if:*

$$\tau(\epsilon_a, \epsilon_b) \geq \epsilon_{a+b}, \quad \forall a, b \in \mathbb{R}^+.$$

2.3 On some specific cases

The simplest metric spaces are discrete spaces, often referred to as equilateral spaces, where the metric d is defined as

$$d(p,q) = \begin{cases} a, & p \neq q \\ 0, & p = q \end{cases}$$

where a is a positive constant. Similarly, a PM-space is termed equilateral if it satisfies the following property for a specific distribution function G where G(0)=0,

$$F_{p,q}(x) = \begin{cases} G(x), & p \neq q \\ \epsilon_0(x), & p = q \end{cases}$$

It can be easily confirmed that all the conditions (**PM-1** through **PM-4**) required for PM-spaces are fulfilled.

Theorem 8. *In an equilateral PM-space, the Menger triangle inequality (i.e., **PM-5**) holds for any triple of distinct points under* $T = M^*$ *and universally under* $T = M$*.*

Proof. Since G is non decreasing,

$$\begin{aligned} G(x+y) &\geq max(G(x), G(y)) \\ &\geq min(G(x), G(y)) \\ and & \\ G(x+y) &\geq min(G(x), 1). \end{aligned}$$

□

Next, we provide examples demonstrating the existence of equilateral PM-spaces where the generalized triangle inequality (**PM-5**) is satisfied under a t-norm stronger than $T = M^*$.

Example 4. *Let*

$$G(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 1 \\ 1 & 1 \leq x. \end{cases}$$

*For any triple of distinct points in this space, the condition **PM-5** holds under $T = W$, as in all cases, we have*

$$G(x+y) \geq min(G(x) + G(y), 1).$$

Example 5. *Let*

$$G(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x}, & x \geq 0. \end{cases}$$

*For any triple of distinct points in this space, condition **PM-5** holds under $T = (a+b) - ab$. This is evident in the view of the fact that $e^{-x}e^{-y} = e^{-(x+y)}$.*

A more interesting class of PM-spaces, compared to equilateral PM-spaces, can be defined using the concept of a specific distribution as follows:

Consider (X, d) as a metric space and let G represent a distribution function distinct from ϵ_x such that $G(0) = 0$. For each pair of points $p, q \in X$, the distribution function $F_{p,q}$ is defined as follows.

$$F_{p,q}(x) = \begin{cases} G\left(\frac{x}{d(p,q)}\right), & p \neq q \\ \epsilon_0(x), & p = q. \end{cases} \tag{2.3}$$

Definition 16. *A PM-space $(S, \mathcal{F})$ is said to be a simple space iff there exists a metric d on S and a distribution function G satisfying $G(0)=0$, such that for every point $p, q \in S$, $F_{p,q}$ is given by Equation 2.3. Furthermore, we say that $(S, \mathcal{F})$ is a simple space generated by the metric space (S, d) and the distribution function G.*

Theorem 9. *A simple space is a Menger PM-space under any choice of T satisfying **T-2**, **T-3**, **T-6**, and **T-7**.*

Proof. It is sufficient to show that the condition **PM-5** holds universally under $T = M$, since this is the strongest choice of T possible. From Lemma 4, we have only to show that for p, q, r are distinct

$$G\left(\frac{x+y}{d(p,r)}\right) \geq min\left(G\left(\frac{x}{d(p,q)}\right), G\left(\frac{y}{d(q,r)}\right)\right) \tag{2.4}$$

Since d is an ordinary metric, therefore

$$d(p,r) \leq d(p,q) + d(q,r).$$

Which in turn yields that

$$\frac{x+y}{d(p,r)} \geq \frac{x+y}{d(p,q)+d(q,r)} \tag{2.5}$$

Furthermore, since $d(p,q)$ and $d(q,r)$ are positive real numbers, therefore

$$\begin{aligned} max\left(\frac{x}{d(p,q)}, \frac{y}{d(q,r)}\right) &\geq \frac{x+y}{d(p,q)+d(q,r)} \\ &\geq min\left(\frac{x}{d(p,q)}, \frac{y}{d(q,r)}\right) \end{aligned} \tag{2.6}$$

with the equality of either side iff

$$\frac{x}{d(p,q)} = \frac{y}{d(q,r)}$$

Consequently, from inequalities 2.5 and 2.6, we have

$$\tfrac{x+y}{d(p,r)} \geq min\left(\tfrac{x}{d(p,q)}, \tfrac{y}{d(q,r)}\right)$$

Since G is non decreasing, it implies 2.4, which completes the proof. □

Corollary 2. *If $G(x) = \epsilon_0(x-1)$, then the generated PM-space reduces to the generating metric space.*

Proof. Consider the function:

$$F_{p,q}(x) = \epsilon_0\left(\frac{x}{d(p,q)} - 1\right) = \epsilon_{d(p,q)}.$$

In most simple spaces, using the t-norm $T = M^*$ will be too restrictive. This is evident from inequality 2.6, which shows that for a triple of distinct points p, q, r satisfying

$$d(p, r) = d(p, q) + d(q, r),$$

the condition **PM-5** fails under the t-norm M^*. □

2.4 Random Metric Space

In this section, we introduce a family of PM-spaces called E-Spaces, which were introduced by Sherwood [She69]. To define E-Spaces, we first need to establish the foundational concept of probability spaces.

2.4.1 Probability Spaces

Definition 17. *Let Ω be a non-empty set. A sigma-field (or σ-field) $\mathcal{A}$ on Ω is a collection of subsets that satisfies the following properties:*

- $\Omega \in \mathcal{A}$*, i.e., the sample space is included,*
- *If $A \in \mathcal{A}$, then $\Omega \setminus A \in \mathcal{A}$ (closure under complements),*
- *If $A_1, A_2, A_3, \ldots \in \mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ (closure under countable unions).*

Definition 18. *A probability space is a triple $(\Omega, \mathcal{A}, P)$, where:*

- Ω *is a non-empty set,*
- $\mathcal{A}$ *is a sigma-field on Ω,*
- P *is a function $P : \mathcal{A} \to [0, 1]$ satisfying:*
  (1) $P(\Omega) = 1$ and $P(\emptyset) = 0$,
  (2) If $\{A_n\}_{n=1}^{\infty}$ is a sequence of pairwise disjoint sets in $\mathcal{A}$, then:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Lemma 10. *Let $(\Omega, \mathcal{A}, P)$ be a probability space. Then for any $A, B \in \mathcal{A}$, the following properties hold:*

(1) If $A \subset B$, then:

$$P(B) = P(A) + P(B \setminus A),$$

and hence:

$$P(A) \leq P(B), \tag{2.7}$$

which shows that P is non-decreasing on $\mathcal{A}$.

(2) The complement satisfies:

$$P(\Omega \setminus A) = 1 - P(A). \tag{2.8}$$

(3) The union and intersection satisfy:

$$P(A \cup B) + P(A \cap B) = P(A) + P(B).$$

(4) For triangle norms W and M, the following inequality holds:

$$W(P(A), P(B)) \leq P(A \cap B) \leq M(P(A), P(B)). \tag{2.9}$$

Proof. The first three properties follow directly from the axioms of probability. We now prove the inequality in (2.9). The right-hand inequality follows from (2.7), which is a property of probabilities. The left-hand inequality arises from the fact that $P(A \cap B) \geq 0$. Furthermore:

$$P(A \cap B) = P(A) + P(B) - P(A \cup B),$$

and since $P(A \cup B) \leq P(\Omega) = 1$, it follows that:

$$P(A \cap B) \geq P(A) + P(B) - P(\Omega).$$

Substituting $P(\Omega) = 1$, we obtain:

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

This completes the proof. □

Remark 10. *If Ω is the unit interval I and F is the identity function on I with $F(0) = 0$ and $F(1) = 1$, then F defines a unique probability measure called Lebesgue measure denoted by λ, and the corresponding space is denoted by (I, λ)*

E-Space

E-Spaces are one family of PM-spaces [She69; Ste68] defined through the use of measurable functions and probability spaces. These spaces provide a framework for quantifying the measure of points where the distance between two functions does not exceed a given value x. The construction of E-Spaces is based on probability measures, which extend the traditional application of the Lebesgue measure on the interval $I = [0, 1]$, as highlighted by Schweizer and Sklar [SS83].

Formally, let $p, q : I \to M$, where (M, d) is a metric space with a distance function $d(a, b) = |a - b|$. Using the Lebesgue measure λ, the distance distribution function for $\Omega = I = [0, 1]$ is defined as:

$$F_{p,q}(x) = \lambda \left(\{ t \in I \mid |p(t) - q(t)| < x \} \right).$$

E-Spaces generalize this construction by incorporating a probability space $(\Omega, \mathcal{A}, P)$ rather than limiting the scope to the Lebesgue measure (I, λ). This generalization allows for a broader class of measurable spaces. Furthermore, E-Spaces make use of $L_1^+(\Omega)$, which represents the set of all positive, almost everywhere finite, Lebesgue-measurable functions on Ω.

Definition 19. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, let (M, d) be a metric space, let S be a set of functions from Ω into M and let $\mathcal{F}$ be a mapping from $S \times S$ into Δ^+. Then, $(S, \mathcal{F})$ is an E-space with base $(\Omega, \mathcal{A}, P)$ and target (M, d) if*

- *(i) For all $p, q \in S$ and all $x \in \mathbb{R}^+$ the set*

$$\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}$$

  belongs to $\mathcal{A}$; i.e., the composite function $d(p, q)$ from Ω into $\mathbb{R}^+$ is P-measurable and therefore in $L_1^+(\Omega)$.

- *(ii) For all $p, q \in S$, $\mathcal{F}(p, q) = F_{pq}$ defined by*

$$F_{p,q}(x) = P(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}). \tag{2.10}$$

Equation 2.10 implies that F satisfies Properties (1) and (3) in Definition 14. If F also satisfies Property (2), then (S, F) is a canonical E-space.

Theorem 11. *Let (S, F) be an E-space. Then (S, F) is a PPM space under τ_W. If (S, F) is canonical, then it is a Menger space under W.*

Proof. We need only establish **PM-6**. For any $p, q, r \in S$ and any $x \in \mathbb{R}^+$, let $u, v \in \mathbb{R}^+$ such that $u + v = x$. Define the sets A, B, C as follows:

$$A = \{w \in \Omega \mid d(p(w), q(w)) < u\},$$

$$B = \{w \in \Omega \mid d(q(w), r(w)) < v\},$$

$$C = \{w \in \Omega \mid d(p(w), r(w)) < x\}.$$

Since d satisfies the triangle inequality, it follows that $A \cap B \subseteq C$. Hence 2.9 yields:

$$P(C) \geq P(A \cap B) \geq W(P(A), P(B)).$$

By 2.10, we have:

$$P(A) = F_{p,q}(u), \quad P(B) = F_{q,r}(v), \quad P(C) = F_{p,r}(x).$$

Thus:

$$F_{p,r}(x) \geq W(F_{p,q}(u), F_{q,r}(v)).$$

□

Chapter 3

Fuzzy Measures

Uncertainty is an inherent aspect of human cognition and decision-making, manifesting in various forms such as imprecision and vagueness. While these terms are often used interchangeably, they have distinct meanings: *imprecision* refers to a lack of exactness in numerical or quantitative data, whereas *vagueness* arises in qualitative or linguistic contexts. For example, describing an image as having "good quality" is inherently vague, as its interpretation depends on subjective and context-dependent factors. The ability to model and quantify such uncertainty is crucial, particularly in domains where precise boundaries or strict definitions are difficult to establish [KY95].

Traditional mathematical frameworks, such as classical set theory and probability measures, have long been employed to address uncertainty. However, these models rely on binary and additive principles that often fail to accommodate the gradual transitions and overlapping categories characteristic of real-world phenomena [Zad65]. This limitation has led to the development of alternative approaches, including fuzzy set theory and its extension, fuzzy measure theory, which provides a more flexible and expressive means of representing and reasoning about uncertainty [Sug74; TNS14; TN07].

The foundations of fuzzy set theory were introduced by Lotfi Zadeh in 1965, marking a paradigm shift in the mathematical treatment of vagueness and imprecision. Unlike classical set theory, which enforces strict membership rules, fuzzy set theory allows elements to have partial membership, quantified by a degree between zero and one [Zad65]. This fundamental concept laid the groundwork for fuzzy measures, introduced by Sugeno in 1974, which generalize classical measures by relaxing the requirement of additivity [Sug74]. Instead, fuzzy measures rely on properties such as monotonicity and continuity, enabling them to model uncertainty in a way that better aligns with human reasoning and real-world complexity. For instance, whereas classical measures assign a fixed numerical value to the area of a shape, fuzzy measures can express the degree to which a region belongs to a particular category, such as estimating the proportion of blackness in an image based on reflected light intensity [WK92].

Since their introduction, fuzzy measures have been further developed through the contributions of researchers such as Wang and Klir [WK92], Ralescu and Adams [RA78], Grabisch [Gra96], and Pap [Pap95], who have refined their mathematical properties and expanded their applicability. Unlike classical probability measures, which require strict additivity, fuzzy measures provide non-additive models of uncertainty including plausibility, possibility, and belief functions [Sha76]. These concepts allow fuzzy measures to handle both probabilistic and non-probabilistic uncertainty, broadening their utility across various disciplines. For example, belief functions quantify the degree of support for a given hypothesis, while possibility measures assess the feasibility of events under incomplete information [DP88].

Fuzzy measure theory has been widely applied in artificial intelligence, decision analysis, data mining, and information fusion. In artificial intelligence, fuzzy measures facilitate multi-attribute evaluation and data aggregation, supporting tasks such as pattern recognition, clustering, and classification [Zad78]. In decision-making, they provide a structured framework for aggregating expert opinions and evaluating alternatives, particularly in multi-criteria decision analysis [MS89]. The adoption of fuzzy measures in these domains has been instrumental in improving the interpretability and robustness of computational models, particularly in contexts where traditional probability-based approaches struggle to handle uncertainty effectively [Yag81].

A significant application of fuzzy measures is the concept of fuzzy integrals, which generalize classical integrals to accommodate non-additive measures. Notable examples include the Choquet and Sugeno integrals, which provide alternative methods for aggregating information in decision-making and data analysis. The Choquet integral is particularly valuable for modeling interactions between attributes in multi-criteria decision-making [Cho54], while the Sugeno integral is useful in scenarios where max-min aggregation is preferable, such as in robust decision models [Sug74].

Beyond theoretical advancements, fuzzy measures have been applied to a range of real-world problems involving uncertainty and imprecision. Examples include assessing the severity of cerebral damage using fuzzy classification models [Blo96], modeling the spatial extent of vague geographical regions [Rob03], and optimizing resource allocation in complex decision-making environments [Gra97]. These applications highlight the adaptability of fuzzy measures in addressing challenges that require reasoning under uncertainty.

This chapter provides a brief introduction to fuzzy measures, beginning with the fundamental concepts of fuzzy set theory, including its motivation, formal definitions, and key operations. It then presents fuzzy measures as an extension of fuzzy sets, outlining their key properties such as monotonicity, continuity, and non-additivity. Additionally, it introduces Sugeno λ-measures, a specific class of fuzzy measures that incorporates an interaction parameter to model non-additive aggregation.

3.1 Fuzzy Sets

Fuzzy set theory extends classical (crisp) set theory by allowing elements to have varying degrees of membership, rather than being strictly included or excluded. In classical set theory, also referred to as *crisp set theory*, a set A defined on a universe X contains an element x if and only if x is a full member of A. This means that every element is either fully included ($x \in A$) or completely excluded ($x \notin A$), with no intermediate states.

Fuzzy set theory generalizes this concept by introducing *partial membership*, where elements belong to a set with a degree of membership that continuously ranges between 0 and 1. This is captured by a *membership function*, which assigns a real value in the interval $[0, 1]$ to each element in X, reflecting its degree of association with the set. Higher values indicate stronger membership, allowing for a more refined representation of uncertainty and vagueness in mathematical modeling.

Definition 20. *A fuzzy set A on the universe X is defined by a membership function:*

$$\mu_A : X \to [0, 1] \tag{3.1}$$

where $\mu_A(x)$ represents the membership degree of element x in A.

One of the key applications of fuzzy sets is in modeling concepts that lack precise boundaries, such as the notion of *expensiveness*.

Example 6. *Consider the concept of an expensive car. In classical set theory, a car would either be classified as expensive or not, based on a strict threshold. However, in fuzzy set theory, expensiveness is viewed as a spectrum, allowing cars to belong to the category of expensive cars with varying degrees of membership.*

Suppose we consider a selection of cars: Ferrari, Rolls Royce, Mercedes, BMW, Honda, Fiat, and Renault. Some, such as Ferrari and Rolls Royce, are unquestionably expensive, while others, like Fiat and Renault, are considered relatively inexpensive. Using a fuzzy set, we can model the concept of expensive cars as follows:

$$A = \begin{cases} (\textit{Ferrari}, 1) \\ (\textit{Rolls Royce}, 1) \\ (\textit{Mercedes}, 0.8) \\ (\textit{BMW}, 0.7) \\ (\textit{Honda}, 0.4) \\ (\textit{Fiat}, 0.2) \\ (\textit{Renault}, 0.2) \end{cases}$$

Here, Ferrari and Rolls Royce have a membership value of 1, indicating that they are fully considered expensive. Mercedes and BMW have lower membership *values of 0.8 and 0.7, respectively, reflecting their relative cost. Honda, Fiat, and Renault have even lower values, indicating that they are perceived as less expensive.*

Fuzzy sets are widely used to represent linguistic terms such as *low*, *medium*, and *high*. Such terms describe variables that transition smoothly rather than abruptly. A variable that follows this principle is referred to as a *fuzzy variable*.

The importance of fuzzy variables lies in their ability to model gradual changes and handle measurement uncertainty. For instance, temperature can be described as "cold", "warm", or "hot," but these categories do not have strict boundaries. Instead of enforcing sharp divisions, fuzzy sets allow for a smooth transition between these states, enabling a more flexible and intuitive representation of imprecise concepts.

In some cases, defining membership functions with exact values may not be feasible due to inherent uncertainty. Instead of assigning a single precise membership value, an *interval-valued fuzzy set* represents membership as a closed interval within $[0, 1]$. This approach provides greater flexibility in modeling uncertainty and capturing variations in data.

Definition 21. *An interval-valued fuzzy set A on the universe X is defined by a membership function:*

$$\mu_A : X \to \mathcal{P}([0, 1]) \tag{3.2}$$

where $\mathcal{P}([0, 1])$ represents the set of all closed subintervals within $[0, 1]$.

By allowing interval-based membership values, these sets accommodate a broader range of uncertainties that may arise in real-world applications. Compared to standard fuzzy sets, interval-valued fuzzy sets provide greater flexibility by eliminating the need for exact membership values. This characteristic makes them especially valuable in scenarios where data is imprecise or uncertain. However, this flexibility introduces greater computational complexity, making processing and analysis more demanding.

3.1.1 Notation and Classical Sets

A fuzzy set on a universe X is a mapping from X to the interval $[0, 1]$. The collection of all fuzzy sets on X is denoted by:

$$\mathcal{F}(X) = \{A \mid A : X \to [0, 1]\}.$$

For a finite universe $X = \{x_1, x_2, \ldots, x_n\}$, a fuzzy set A can be represented as:

$$A = \{a_1/x_1, a_2/x_2, \ldots, a_n/x_n\}, \quad a_i \in (0, 1].$$

Elements with zero membership ($A(x) = 0$) are typically omitted for brevity.

Several fundamental classical sets are associated with fuzzy sets. The *support* of a fuzzy set consists of all elements with nonzero membership, capturing the range of elements that contribute to the fuzzy set. The *α-cut* of a fuzzy set represents the subset of elements whose membership degree is at least

Definition 22. *The support of a fuzzy set $A \in \mathcal{F}(X)$ is the set of elements with positive membership:*

$$Supp(A) = \{x \in X \mid A(x) > 0\}. \tag{3.3}$$

Definition 23. *The α-cut of A for a given threshold $\alpha \in [0,1]$ is defined as:*

$$A_\alpha = \{x \in X \mid A(x) \geq \alpha\}. \tag{3.4}$$

The α-cut produces a crisp subset of X that includes all elements whose membership value in A meets or exceeds α. A fundamental property of α-cuts is their nesting behavior:

$$A_\alpha \supseteq A_\beta, \quad \text{for } \alpha \leq \beta. \tag{3.5}$$

This means that as α increases, the corresponding α-cut becomes a smaller subset of X, reflecting a stricter inclusion criterion.

3.1.2 Convexity of Fuzzy Sets

An important property of fuzzy sets is their convexity, which generalizes the classical notion of convexity in crisp sets.

Definition 24. *A fuzzy set A on $\mathbb{R}$ is convex if and only if for all $x_1, x_2 \in X$ and for all $\lambda \in [0,1]$, the following inequality holds:*

$$A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{A(x_1), A(x_2)\}. \tag{3.6}$$

Proof.

1. Assume that A is convex. Let $\alpha = A(x_1) \leq A(x_2)$. Then, $x_1, x_2 \in A_\alpha$, and by the convexity of A, we have:

$$A(\lambda x_1 + (1-\lambda)x_2) \geq \alpha = A(x_1) = \min\{A(x_1), A(x_2)\}.$$

2. Assume that A satisfies (3.6). We need to show that for any $\alpha \in (0,1)$, A_α is convex. Since $x_1, x_2 \in A_\alpha$, meaning $A(x_1) \geq \alpha$ and $A(x_2) \geq \alpha$, using (3.6), we get:

$$A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{A(x_1), A(x_2)\} \geq \alpha.$$

This implies that $\lambda x_1 + (1-\lambda)x_2 \in A_\alpha$, thus proving convexity.

□

3.1.3 Operations on Fuzzy Sets

Similar to classical set theory, fundamental operations such as intersection, union, and complement can be extended to fuzzy sets. These operations are governed by appropriate aggregation functions, specifically t-norms for intersection and t-conorms for union. The following definitions formalize these operations.

Definition 25. *Let $A, B \in \mathcal{F}(X)$, and let T and S denote a t-norm and a t-conorm, respectively. Then, the basic operations on fuzzy sets are defined as follows:*

- *The intersection of A and B, denoted as $C = A \cap B$, is a fuzzy set $C \in \mathcal{F}(X)$ with the membership function given by:*

$$C(x) = T(A(x), B(x)), \quad \forall x \in X. \tag{3.7}$$

- *The union of A and B, denoted as $D = A \cup B$, is a fuzzy set $D \in \mathcal{F}(X)$ with the membership function given by:*

$$D(x) = S(A(x), B(x)), \quad \forall x \in X. \tag{3.8}$$

- *The complement of A, denoted as $\bar{A}$, is defined by:*

$$\bar{A}(x) = 1 - A(x), \quad \forall x \in X. \tag{3.9}$$

A commonly used choice for these operations are T=min and S= max.

These standard definitions ensure that fuzzy set operations generalize classical set operations while accommodating degrees of membership, providing a flexible framework for handling imprecise information.

3.1.4 Fuzzy Relations

Introduced in Zadeh's seminal work on fuzzy sets [Zad65], fuzzy relations extend classical (crisp) relations by allowing elements to be associated with varying degrees of membership rather than a strict binary classification. Unlike crisp relations, which define a strict inclusion or exclusion of element pairs, fuzzy relations are represented as fuzzy sets over pairs (or n-tuples) of objects, enabling a more flexible and nuanced representation of relationships.

Let $A \in \mathcal{F}(X)$ and $B \in \mathcal{F}(Y)$. Their *Cartesian product* is a fuzzy set $A \times B \in \mathcal{F}(X \times Y)$ with the membership function:

$$(A \times B)(x, y) = A(x) \wedge B(y), \quad \forall x \in X,\ y \in Y. \tag{3.10}$$

In (3.10), the minimum operator $\wedge$ is used to compute the membership degree. This is the most usual operator for conjunction. However, a more general formulation allows for the use of an arbitrary t-norm T, leading to the generalized Cartesian product:

$$(A \times_T B)(x, y) = T(A(x), B(y)), \quad \forall x \in X,\ y \in Y. \tag{3.11}$$

Definition 26. *An n-ary fuzzy relation R is a fuzzy set defined over the Cartesian product $X_1 \times \cdots \times X_n$ of n universes. When all domains are identical, i.e., $X_1 = \cdots = X_n = X$, the relation is referred to as an n-ary fuzzy relation on X.*

The membership function $R(x_1, \ldots, x_n)$ quantifies the degree to which the elements $x_i \in X_i$, for $i = 1, \ldots, n$, belong to the relation R. Notably, the Cartesian product of two fuzzy sets corresponds to a special case of a binary fuzzy relation.

Since a fuzzy relation is a fuzzy set defined on a Cartesian product of crisp sets, its fundamental operations—intersection, union, and complement—are naturally inherited from fuzzy set theory.

3.2 Fuzzy Measures

Fuzzy sets and fuzzy measures are closely related but fundamentally different in how they represent and handle uncertainty. A fuzzy set assigns a membership degree to each individual element, indicating the extent to which it belongs to the set. This degree of membership typically ranges between 0 and 1, allowing for partial membership and a smooth transition between inclusion and exclusion.

In contrast, fuzzy measures do not evaluate individual elements directly but instead assign values to entire subsets. Rather than indicating the membership of a single element, they quantify the overall belief or confidence in the classification of a subset, often based on available evidence. This transition from an element-wise classification to a subset-level evaluation enables fuzzy measures to provide a more comprehensive representation of uncertainty.

To illustrate this distinction, consider the problem of assessing an individual's guilt in a legal context. A fuzzy set would assign a degree of guilt directly to the individual, reflecting the level of certainty about their culpability. Conversely, a fuzzy measure would evaluate the strength of evidence supporting the classification of a subset of individuals as guilty. This broader perspective allows fuzzy measures to incorporate multiple sources of information and express varying degrees of confidence more effectively.

Fuzzy measures are also known as *capacities* or *non-additive measures*, as they generalize classical probability measures while relaxing the additivity property. Their formal definition is as follows:

Definition 27. *Let X be a universal set and $\mathcal{A}$ a nonempty family of subsets of X. A function $\mu : \mathcal{A} \to [0, 1]$ is called a* ***fuzzy measure*** *if it satisfies the following conditions:*

FM-1

$$\mu(\emptyset) = 0 \quad \textit{and} \quad \mu(X) = 1$$

The boundary condition ensures that the empty set has no evidence assigned to it, and the entire universal set receives full certainty.

FM-2

$$\forall A, B \in \mathcal{A}, \quad A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$$

This property, called Monotonicity, states that as a set expands, the measure should not decrease, ensuring consistency in uncertainty representation.

3.2.1 Properties of Fuzzy Measures

Fuzzy measures can be classified based on the following continuity properties.

FM-3

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i)$$

This ensures that the fuzzy measure behaves consistently for increasing sequences of sets.

FM-4

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i)$$

This requirement applies to decreasing sequences of sets, ensuring stability in convergence.

Specifically, a measure μ is termed *lower semicontinuous* if it satisfies conditions **FM-1**, **FM-2**, and **FM-3**, whereas it is considered *upper semicontinuous* if it meets conditions **FM-1**, **FM-2**, and **FM-4**. A measure that satisfies both lower and upper semicontinuity is referred to as a *continuous fuzzy measure*.

Beyond continuity, fuzzy measures exhibit various fundamental properties that characterize how they aggregate information over sets. One crucial aspect is the *additivity condition*, which determines how the measure behaves when applied to disjoint sets. Classical probability measures are strictly *additive*, meaning that the measure of a union of disjoint sets equals the sum of their individual measures. However, in many real-world applications, uncertainty cannot be precisely modeled using strict additivity, leading to the need for *non-additive measures* such as *subadditive* and *superadditive* measures.

Definition 28. *Let μ be a non-additive measure on the measurable space $(X, \mathcal{A})$. The measure μ satisfies the following properties:*

- *Additivity: μ is additive if, for sets $A, B \in \mathcal{A}$,*

$$\mu(A \cup B) = \mu(A) + \mu(B), \quad \textit{when } A \cap B = \emptyset. \tag{3.12}$$

- *Superadditivity: μ is superadditive if:*

$$\mu(A \cup B) \geq \mu(A) + \mu(B), \quad \textit{when } A \cap B = \emptyset. \tag{3.13}$$

- *Subadditivity:* μ *is subadditive if:*

$$\mu(A \cup B) \leq \mu(A) + \mu(B), \quad \text{when } A \cap B = \emptyset. \tag{3.14}$$

- *Submodularity:* μ *is submodular if:*

$$\mu(A) + \mu(B) \geq \mu(A \cup B) + \mu(A \cap B). \tag{3.15}$$

- *Supermodularity:* μ *is supermodular if:*

$$\mu(A) + \mu(B) \leq \mu(A \cup B) + \mu(A \cap B). \tag{3.16}$$

- *Symmetry:* μ *is symmetric if, for a finite set* X*, whenever* $|A| = |B|$*, then:*

$$\mu(A) = \mu(B). \tag{3.17}$$

Additionally, a *supermodular measure* inherently satisfies the *superadditivity* property, while a *submodular measure* implies *subadditivity*. These relationships highlight the structural dependencies between different properties of fuzzy measures.

3.2.2 Examples on Fuzzy Measures

Example 7. *Let* μ *be the Dirac measure on* $(X, \mathcal{A})$*, i.e., for any* $E \in \mathcal{A}$*,*

$$\mu(E) = \begin{cases} 1, & x_0 \in E, \\ 0, & x_0 \notin E, \end{cases}$$

where x_0 *is a fixed point in* X*. This set function* μ *is a probability measure and a fuzzy measure.*

Example 8. *Let* $X = \{1, 2, \ldots, n\}$*, and let* $\mathcal{A} = \mathcal{P}(X)$*, where* $\mathcal{P}(X)$ *denotes the power set of* X*. Define the function* $\mu : \mathcal{A} \to [0, 1]$ *by:*

$$\mu(E) = \left(\frac{|E|}{n} \right)^2,$$

where $|E|$ *is the number of elements in* E*, then* μ *is a fuzzy measure. Since* X *is finite, continuity (both from above and below) is naturally satisfied.*

Example 9. *Let* $X_0 = \{1, 2, \ldots\}$*, and define* $X = X_0 \times X_0$*. If* $E \in \mathcal{P}(X)$*, define*

$$\mu(E) = |Proj\ E|,$$

where

$$Proj\ E = \{x \mid (x, y) \in E\}.$$

μ *satisfies the conditions* ***(FM1)****,* ***(FM2)****, and* ***(FM3)****, but is not continuous from above.*

Example 10. *Let $f(x)$ be a nonnegative, extended real-valued function defined on $X = (-\infty, \infty)$. If*

$$\mu(E) = \sup_{x \in E} f(x), \quad \forall E \in \mathcal{P}(X),$$

*then μ satisfies **(FM-1)**, **(FM-2)**, and **(FM-3)**, but is not necessarily continuous from above. Thus, μ is a lower semicontinuous fuzzy measure on $(X, \mathcal{P}(X))$.*

Example 11. *Let the measurable space $(X, \mathcal{P}(X))$ be the same as in the previous example. Let $f : X \to [0, 1]$ is such that*

$$\inf_{x \in X} f(x) = 0,$$

then the set function μ is given by

$$\mu(E) = \inf_{x \in E} f(x).$$

3.2.3 Sugeno λ-Measures

Sugeno λ-measures are a class of fuzzy measures that generalize classical probability measures by introducing an interaction parameter λ. Unlike additive probability measures, which assume independent contributions from disjoint subsets, Sugeno λ-measures allow for interactions between elements, making them particularly useful for *non-additive aggregation* in applications such as *decision-making, information fusion, and uncertainty modeling.*

A key property of Sugeno λ-measures is their λ-*decomposability*, which defines how the measure of a union of two disjoint sets is computed based on their individual measures. This decomposability is given by a specific functional form that accounts for potential synergy or redundancy between subsets.

Definition 29. *A fuzzy measure μ is called a Sugeno λ-measure if there exists a fixed parameter $\lambda \in \left(\frac{-1}{\sup \mu}, \infty\right)$, where $\sup \mu = \sup_{E \in \mathcal{A}} \mu(E)$, such that for all sets $A, B \in \mathcal{A}$, with $A \cup B \in \mathcal{A}$ and $A \cap B = \emptyset$, the following equation holds:*

$$\mu(A \cup B) = \mu(A) + \mu(B) + \lambda \mu(A)\mu(B). \tag{3.18}$$

This equation expresses the λ-*decomposability property*, meaning that the measure of a union is *not strictly additive* but rather incorporates an *interaction term* $\lambda \mu(A)\mu(B)$. The value of λ determines whether the measure exhibits *synergistic* ($\lambda > 0$), *neutral* ($\lambda = 0$), or *redundant* ($\lambda < 0$) aggregation effects:

- If $\lambda > 0$, the measure is *superadditive*, meaning that the whole is greater than the sum of its parts, modeling reinforcement or synergy.

- If $\lambda = 0$, the measure reduces to standard additivity, behaving like a classical probability measure.
- If $\lambda < 0$, the measure is *subadditive*, implying redundancy, where the whole is less than the sum of its parts.

An important consequence of this formulation is that a *Sugeno λ-measure is fully determined* by specifying the measure values for all *singletons* in the universe and the parameter λ. The following proposition formalizes this result and provides a general expression for computing Sugeno λ-measures over arbitrary subsets.

Proposition 1. *[Sug74] Let $v : X \to [0,1]$ and $\lambda > -1$ be such that*

$$\frac{1}{\lambda}\left(\prod_{x_i \in X}[1+\lambda v(x_i)] - 1\right) = 1 \quad \text{if } \lambda \neq 0$$

$$\sum_{x_i \in X} v(x_i) = 1 \quad \text{if } \lambda = 0;$$

then, the fuzzy measure defined by

$$\mu(A) = \begin{cases} v(x_i) & \text{if } A = \{x_i\} \\ \frac{1}{\lambda}\left(\prod_{x_i \in A}[1+\lambda v(x_i)] - 1\right) & \text{if } |A| \neq 1 \text{ and } \lambda \neq 0 \\ \sum_{x_i \in A} v(x_i) & \text{if } |A| \neq 1 \text{ and } \lambda = 0 \end{cases}$$

is a Sugeno λ-measure.

Constructing a λ-fuzzy measure is a significant and interesting issue. Consider a finite set X, and let $\mathcal{F}$ consist of X and all singletons. Suppose μ is known on the singleton subsets, i.e., $\mu(\{x_i\})$ for all $x_i \in X$, with the condition that $\mu(\{x_i\}) < \mu(X)$. A λ-fuzzy measure on $\mathcal{F}$ satisfies the equation:

$$\mu(X) = \frac{1}{\lambda}\left[\prod_{i=1}^{n}(1+\lambda\mu(\{x_i\})) - 1\right] \tag{3.19}$$

An important result establishes that the measure values assigned to the singletons uniquely determine the parameter λ. Consequently, defining a measure within this family requires only $|X|$ values, making it computationally efficient. The following theorem provides a formulation for computing λ.

Theorem 12. *[Sug74] Under Equation 3.19, the parameter λ is uniquely determined by the following equation:*

$$1 + \lambda\mu(X) = \prod_{i=1}^{n}(1+\lambda\mu(\{x_i\})).$$

Moreover, the sign and magnitude of λ are characterized as follows:

(1) $\lambda > 0$ *if* $\sum_{i=1}^{n} \mu(\{x_i\}) < \mu(X)$*, indicating a superadditive measure.*

(2) $\lambda = 0$ *if* $\sum_{i=1}^{n} \mu(\{x_i\}) = \mu(X)$*, reducing to an additive measure.*

(3) $-\frac{1}{\mu(X)} < \lambda < 0$ *if* $\sum_{i=1}^{n} \mu(\{x_i\}) > \mu(X)$*, corresponding to a subadditive measure.*

Proof. Define $\mu(X) = a_0$ and $\mu(\{x_i\}) = a_i$ for $i = 1, 2, \ldots, n$. Consider the function:

$$f_k(\lambda) = (1 + a_k\lambda) f_{k-1}(\lambda).$$

By differentiation, we obtain:

$$f_k'(\lambda) = a_k f_{k-1}(\lambda) + (1 + a_k\lambda) f_{k-1}'(\lambda).$$

which leads to:

$$f_n'(\lambda) = \sum_{i=1}^{n} a_i.$$

Since $f_n(\lambda)$ is concave and $\lim_{\lambda\to\infty} f_n(\lambda) = \infty$, it follows that $f_n(\lambda)$ has a unique intersection with $f(\lambda) = 1 + a_0\lambda$, establishing the uniqueness of λ.

Thus, solving a polynomial equation of degree $n-1$ yields the unique valid λ. □

Example 12. *Consider the finite set* $X = \{a, b, c\}$ *with* $\mu(X) = 1$ *and the following singleton values:*

$$\mu(\{a\}) = 0.3, \quad \mu(\{b\}) = 0.25, \quad \mu(\{c\}) = 0.15.$$

Using Equation 3.19, we obtain:

$$1 = \frac{(1 + 0.3\lambda)(1 + 0.25\lambda)(1 + 0.15\lambda) - 1}{\lambda}.$$

Expanding the expression:

$$0.01125\lambda^2 + 0.7\lambda - 0.5 = 0.$$

Solving the quadratic equation:

$$\lambda = \frac{-0.7 \pm \sqrt{0.49 + 0.045}}{0.0225}.$$

$$\lambda = \frac{-0.7 \pm 0.72}{0.0225}.$$

$$\lambda = 1.70 \quad \text{or} \quad -15.70.$$

Since $\lambda = -15.70$ *is outside the feasible range* $\lambda > -1$*, the unique valid solution is* $\lambda = 1.70$*.*

Chapter 4

Summary of Contributions

This thesis formalized the interaction between the space of datasets (database space) and the space of trained machine learning models (model space), with the goal of developing principled, uncertainty-aware distances for comparing models and learning algorithms. Grounded in probabilistic metric spaces, the thesis develops a flexible theoretical framework that supports multiple scenarios—ranging from evolving datasets, to structured data interactions, to uncertainty within the model space—using tools such as fuzzy measures and transformation-based modeling to assess model similarities.

The central research problem was defined as: How can we meaningfully compare machine learning models and algorithms by explicitly accounting for the datasets that generate them? This problem was articulated through three key research questions:

RQ1: How can models m_1 and m_2 in M be compared while accounting for transformations in the database space Ω?

RQ2: How can we construct distances and metrics for machine learning algorithms in G that capture complex interactions of the databases?

RQ3: Which characterizations can be provided for the metrics we propose?

Answering these questions required the development of new mathematical tools to quantify uncertainty in model comparisons. Collectively, the four papers present a coherent framework for understanding and measuring distances between machine learning models, grounded in probabilistic metric space theory and shaped by dataset dynamics and data-driven interactions.

This chapter provides an overview of the papers included in the thesis, outlining their individual contributions and how they address the research questions.

Summary of Paper I

Vicenç Torra, Mariam Taha, & Guillermo Navarro-Arribas. The space of models in machine learning: using Markov chains to model transitions. *Progress in Artificial Intelligence, 10, 321–332 (2021).* https://doi.org/10.1007/s13748-021-00242-6

Paper I develops a probabilistic metric space framework to characterize how transformations within the database space affect the resulting model space. To reflect the evolving nature of real-world datasets, the paper models database transitions using Markov chains and transition matrices. This approach enables the definition of a distance metric between models, grounded in the probability of transitioning between the datasets from which they are generated. Two forms of probabilistic metric spaces are introduced:

- Visited Database-Based Probabilistic Metric Space (VD-PMS): Measures the probability that one database transitions into another within a given number of steps.

- Database Distance-Based Probabilistic Metric Space (DD-PMS): Defines distances between databases based on their long-term evolution rather than immediate transitions.

To approximate distances efficiently and reduce computational costs, the paper examines the minimum number of steps needed to transition between databases, ensuring that probability values remain zero before a specific threshold. Additionally, the approach allows for non-modifying transitions, increasing the number of valid paths while preserving theoretical consistency. This leads to a probabilistic formulation where the probability of transitioning in a given number of steps follows a recursive structure.

The paper also proposes approximating transition probabilities by considering only a random subset of valid transition paths, leading to a computationally feasible lower bound for the distance function. This ensures that any decision based on applying a threshold to these approximated probabilities—for example, determining whether two databases or models are sufficiently close—remains valid even when the full set of transition paths is considered. Furthermore, the use of a reference database to approximate distances emerges naturally from the triangle inequality property of the space, allowing efficient estimation without full pairwise computations across all database states.

These definitions extend to machine learning models by associating each model with the set of databases that have generated it. The model distance function is derived by averaging the probabilistic distances over all generating databases, ensuring that similarity comparisons reflect dataset evolution. This approach accounts for dataset-driven transformations. Furthermore, when lower bounds of database distances are used, the model distance function retains these lower bounds. This formulation directly contributes to answering

RQ1: How can models be compared while accounting for transformations in the database space? The contributions of this paper provide a theoretical foundation for model similarity that integrates dataset evolution, paving the way for robust model selection and privacy-preserving analysis in dynamic data environments.

Summary of Paper II

Yasuo Narukawa, Mariam Taha, & Vicenç Torra. On the definition of probabilistic metric spaces by means of fuzzy measures. *Fuzzy Sets and Systems, Vol. 465, Article ID 108528, 2023.*https://doi.org/10.1016/j.fss.2023.108528

Paper II presents a new framework for constructing probabilistic metric spaces based on fuzzy measures, referred to as F-spaces. Unlike other approaches such as E-spaces that rely on additive probability measures, F-spaces enable the modeling of complex interactions among subsets of a base space—such as redundancy, synergy, or overlap—by leveraging non-additive set functions. This theoretical advancement provides a more expressive foundation for measuring similarities between models or functions that operate over input domains composed of interacting or overlapping subsets.

The core idea is to assess how similarly two functions behave across groups of inputs that are considered meaningful according to a fuzzy measure. This yields a distance distribution function, which captures how much of the input space—weighted by importance—satisfies a given similarity threshold.

To ensure that the space satisfies the properties of a probabilistic metric space, the paper investigates the compatibility between classes of fuzzy measures (such as Sugeno λ-measures and indicator-based measures) and various t-norms (e.g., minimum, product, Lukasiewicz).

The results of this investigation are presented through a sequence of formal propositions and theorems. For instance, Theorem 2 proves that when the fuzzy measure is supermodular, the induced F-space satisfies the triangle inequality under the bounded difference t-norm, thus forming a probabilistic pseudometric space. Proposition 3 extends this to convex distortions of probability measures, showing how distorted probabilities also yield valid F-spaces. This illustrates that moving from a classical E-space to an F-space is conceptually and technically straightforward: by applying a convex transformation to a probability measure, one can construct a valid fuzzy measure that preserves the desired probabilistic metric properties under suitable t-norms. Theorem 3 confirms that canonical F-spaces form proper Menger spaces, while Theorem 4 shows that even 1^--measures result in valid pseudometric structures under the drastic t-norm. Theorem 5 further demonstrates that unanimity measures still yield valid F-spaces under the minimum t-norm. Finally, Proposition 4 connects F-spaces with the Choquet integral framework, showing how new valid fuzzy measures can be constructed from existing ones, expanding the versatility of the framework.

Collectively, these results highlight a key insight: the stronger the t-norm used, the fewer constraints are needed on the fuzzy measure to preserve the triangle inequality. This analysis directly addresses RQ3, identifying the conditions under which fuzzy-measure-based distances remain well-structured. This theoretical foundation is particularly relevant in statistical databases, where functions such as arithmetic or harmonic means are commonly used to summarize subsets of data. F-spaces provide a principle way to compare these statistical functions by evaluating how similarly they behave across subsets considered important under the fuzzy measure—enabling nuanced comparisons that align with real-world data analysis needs. This is demonstrated in the application section of the paper.

Summary of Paper III

Mariam Taha & Vicenç Torra. Measuring the distance between machine learning models using F-space. *In: Proceedings of the 13th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2023) and the 12th International Summer School on Aggregation Operators (AGOP 2023), Palma de Mallorca, Spain, September 4–8, 2023.*
Nominated for Best Student Paper Award

Paper III applies the F-space framework to the comparison of machine learning algorithms, leveraging the structure of their generating data as the foundation for defining similarity. This directly contributes to addressing RQ2, by showing how distances between models can incorporate complex interactions and dependencies within the training data.

Each model is treated as a function mapping a dataset to a trained output (e.g., model parameters), and distances between models are computed by evaluating how closely their outputs align across multiple databases.

The study examines both additive measures—recovering the classical E-space formulation—and non-additive ones such as Sugeno λ-measures and unanimity measures. Experiments involve comparing linear regression, Huber regression, and Ridge regression models trained on sampled subsets of a real dataset. The resulting model distances are analyzed using F-space constructions, illustrating how the choice of a fuzzy measure affects whether the induced space satisfies the triangle inequality.

The results confirm that supermodular fuzzy measures—such as Sugeno λ-measures with positive λ—lead to valid probabilistic pseudometric spaces, whereas submodular measures may violate key metric properties. The findings also confirm theoretical results on how different t-norms, when combined with specific fuzzy measures, preserve the triangle inequality.

Paper III provides practical validation that F-spaces capture nuanced differences between models beyond classical metrics. By incorporating fuzzy notions of subset importance, the approach supports more robust comparisons.

Summary of Paper IV

Mariam Taha & Vicenç Torra. Generalized F-spaces through the lens of fuzzy measures. *Fuzzy Sets and Systems, Vol. 507, Article ID 109317, 2025.* https://doi.org/10.1016/j.fss.2025.109317

Paper IV introduces a further generalization of the F-space framework by extending the target space from a metric space to a probabilistic metric space, resulting in what we term a Generalized F-space. This construction allows both the base space and the target space to account for uncertainty: the base is equipped with a fuzzy (non-additive) measure, while the target employs distance distribution functions instead of fixed distances. This dual-layered probabilistic reasoning enables more expressive modeling of both data structure and uncertainty in learned models.

The key contribution of this work is the formal definition and analysis of Generalized F-spaces. It proves that under specific conditions—particularly when the fuzzy measure is supermodular and the triangle function in the target space is proper—the resulting space satisfies the axioms of a probabilistic pseudometric space. Several theorems confirm that different classes of fuzzy measures (e.g., Sugeno λ-measures and indicator-based 1-measures) yield valid Generalized F-spaces under suitable t-norms (e.g., bounded difference, drastic, or minimum).

An important theoretical insight is that Generalized F-spaces preserve the structural benefits of F-spaces while allowing the target distances themselves to be probabilistic, which captures uncertainty in model behavior due to data variability, randomness in training, or architectural differences.

To demonstrate practical relevance, the paper applies this framework to machine learning models—specifically, to estimate distances between classifiers trained on varying subsets of data. By viewing the database space as the base and the model space as a probabilistic metric space (constructed using performance differences), the paper shows how generalized F-spaces allow nuanced, uncertainty-aware comparisons between classifiers. Experiments using the IRIS dataset and three classifiers—Logistic Regression, Random Forest, and SVM—confirm the theoretical results. The use of both Sugeno λ-measures and indicator-based fuzzy measures illustrates the flexibility of the approach under different structural assumptions.

This work directly contributes to addressing RQ2 and RQ3: it expands the class of models and spaces in which distances can be defined, and establishes conditions under which these distances form valid probabilistic metrics. As such, it pushes the boundary of the F-space framework, enabling its application to more complex model evaluation scenarios involving uncertainty and variability in both data and model behavior.

Chapter 5

Conclusion and Future Directions

This thesis develops a mathematical framework for comparing machine learning models by incorporating the influence and structure of the datasets that generate them. Motivated by challenges in data privacy, generalization, and uncertainty quantification, it addresses the fundamental question: How can we compare models in a way that reflects their dependence on data? Traditional evaluation methods often overlook how variations in training data affect model behavior. In contrast, this work formalizes the relationship between the database space and the model space, proposing that model comparison must account for dataset diversity, evolution, and internal structure.

The approach is grounded in probabilistic metric spaces, using Markov chains to model dataset evolution and fuzzy measures to capture redundancy, synergy, and importance among data subsets. These tools lead to distribution-valued distances, which quantify model differences while expressing uncertainty about their origin. The contributions span four papers. The first models data transitions to define distances between models based on transformation probabilities. The second introduces F-spaces, a class of probabilistic metric spaces based on fuzzy measures. The third applies this framework to compare regression models trained on real data. The fourth generalizes the model space itself, allowing uncertainty to be represented in both data and models.

Together, these contributions offer a principled and flexible foundation for models comparison. The resulting framework supports privacy-aware, robust model comparison that goes beyond traditional scalar metrics.

5.1 Future Directions

While the thesis establishes a theoretical framework for comparing machine learning models using probabilistic metric spaces, several promising directions remain for future investigation.

A key area for future work is the scalability of the proposed distance functions to real-size databases. While the current framework has been validated on illustrative examples and small datasets, applying it to large-scale, high-dimensional datasets presents both computational and modeling challenges. Techniques such as sampling, approximation of transition paths, or dimensionality reduction could help make distance computation feasible in practical settings. Second, there is a need for refined strategies to approximate distances especially in settings where the set of possible databases becomes prohibitively large. Future research may explore lumpability in Markov chains as a principled way to aggregate states while preserving transition dynamics, or adopt clustering-based state aggregation to group databases with similar behavior as a more flexible approximation. Additional directions include exploring boundary conditions, path-based approximations, and selecting optimized subsets of transition chains to reduce computational complexity while maintaining theoretical guarantees.

Another line of investigation involves extending the model selection framework to incorporate privacy and robustness considerations. The metric structures developed in this thesis allow for comparing models not only by accuracy but also by their sensitivity to training data variations. In privacy contexts, these distances can be used to formulate disclosure risks measures. This is especially relevant to integral privacy, where private models are defined as those with multiple possible generators; the proposed distances provide a natural way to assess how far a given model is from the private ones. Future work could integrate these metrics into visualization tools and decision-support systems, helping practitioners identify models that are not only accurate but also generalizable, robust, and privacy-preserving.

There is also interest in exploring more expressive fuzzy measures tailored to specific machine learning tasks. While the thesis explored classical examples like Sugeno λ-measures and indicator-type measures, future directions include learning fuzzy measures from data, or defining them based on privacy risk, fairness criteria, or domain-specific importance.

In the theoretical realm, further exploration is warranted regarding the role of associativity in t-norms, particularly how it affects the construction of probabilistic metric spaces. More work is also needed to understand how F-spaces and Generalized F-spaces behave under various assumptions on the measure space and on the model outputs—especially in the case of non-deterministic functions or randomized models, which were outside the scope of this thesis.

Finally, this framework opens opportunities for advancing privacy auditing and fairness assessment in machine learning. Because model distances are defined in terms of their generating datasets, the theory provides a natural lens for evaluating how sensitive a model is to specific training data. Future extensions could include developing formal privacy bounds or fairness diagnostics based on the distance geometry induced by the framework.

Bibliography

[AFS06] C. Alsina, M. J. Frank, and B. Schweizer. *Associative Functions: Triangular Norms and Copulas*. World Scientific, 2006.

[Bis06] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[Blo96] Isabel Bloch. "Information Combination Operators for Data Fusion: A Comparative Review with Classification". In: *IEEE Transactions on Systems, Man, and Cybernetics* 26.1 (1996), pp. 52–67.

[Blu70] Leonard M. Blumenthal. *Theory and Applications of Distance Geometry*. Oxford University Press, 1970.

[Bre01] Leo Breiman. *Random Forests*. Springer, 2001.

[Bri56] Léon Brillouin. *Science and Information Theory*. Academic Press, 1956.

[Cho54] G. Choquet. "Theory of Capacities". In: *Annales de l'Institut Fourier* 5 (1954), pp. 131–295.

[CV95] Corinna Cortes and Vladimir Vapnik. "Support-Vector Networks". In: *Machine Learning* 20.3 (1995), pp. 273–297.

[DH73] Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis*. Wiley, 1973.

[DK17] Finale Doshi-Velez and Been Kim. "Towards A Rigorous Science of Interpretable Machine Learning". In: *arXiv preprint arXiv:1702.08608* (2017). URL: `https://arxiv.org/abs/1702.08608`.

[DP88] D. Dubois and H. Prade. *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. Springer Science Business Media, 1988.

[Dwo06] Cynthia Dwork. "Differential Privacy". In: *Proceedings of the International Colloquium on Automata, Languages, and Programming* (2006).

[EM97] Thomas Eiter and Heikki Mannila. "Distance measures for point sets and their computation". In: *Acta Informatica* 34 (1997), pp. 109–133.

[Est+17] Andre Esteva et al. “Dermatologist-level classification of skin cancer with deep neural networks”. In: *Nature* 542 (Jan. 2017). DOI: `10.1038/nature21056`.

[Fré06] Maurice Fréchet. “Sur quelques points du calcul fonctionnel”. PhD thesis. University of Paris, 1906.

[Fri01] Jerome H. Friedman. “Greedy Function Approximation: A Gradient Boosting Machine”. In: *Annals of Statistics* 29.5 (2001), pp. 1189–1232.

[GBC16] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. `http://www.deeplearningbook.org`. MIT Press, 2016.

[Gra96] M. Grabisch. “The Application of Fuzzy Measures in Decision Making”. In: *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 4.3 (1996), pp. 245–262.

[Gra97] M. Grabisch. “Fuzzy Measures and Integrals in Multicriteria Decision Making”. In: *European Journal of Operational Research* 92.3 (1997), pp. 613–626.

[Hau14] Felix Hausdorff. *Grundzüge der Mengenlehre*. Leipzig: Veit & Comp., 1914.

[Hei27] Werner Heisenberg. “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik”. In: *Zeitschrift für Physik* 43.3–4 (1927), pp. 172–198.

[HPW17] J.B. Heaton, N.G. Polson, and J.H. Witte. “Deep Learning for Finance: Deep Portfolios”. In: *Applied Stochastic Models in Business and Industry* 33.1 (2017), pp. 3–12.

[Jan78] Martin F. Janowitz. “Cluster Analysis Algorithms”. In: *Mathematical Biosciences* 41 (1978), pp. 49–95.

[JM21] Michael I. Jordan and Tom M. Mitchell. “Machine Learning: Trends, Perspectives, and Prospects”. In: *Science* 349.6245 (2021), pp. 255–260.

[KMP00] E. P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, 2000.

[KY95] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Prentice Hall, 1995.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. *Deep Learning*. Vol. 521. 7553. Nature, 2015, pp. 436–444.

[Lip18] Zachary C. Lipton. “The Mythos of Model Interpretability”. In: *Proceedings of the 2018 ICML Workshop on Human Interpretability in Machine Learning (WHI)*. 2018. URL: `https://arxiv.org/abs/1606.03490`.

[Meh+21] Ninareh Mehrabi et al. "A Survey on Bias and Fairness in Machine Learning". In: *ACM Computing Surveys (CSUR)* 54.6 (2021), pp. 1–35. DOI: 10.1145/3457607. URL: https://dl.acm.org/doi/10.1145/3457607.

[Men42] Karl Menger. "Statistical Metrics". In: *Proceedings of the National Academy of Sciences* 28.12 (1942), pp. 535–537.

[Men51] Karl Menger. "Probabilistic geometry". In: *Proceedings of the National Academy of Sciences of the United States of America* 37 (1951), pp. 226–229.

[MS89] T. Murofushi and M. Sugeno. "An Interpretation of Fuzzy Measures and the Choquet Integral as an Integral with Respect to a Fuzzy Measure". In: *Fuzzy Sets and Systems* 29.2 (1989), pp. 201–227.

[Mur12] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective.* MIT Press, 2012.

[Nii87] Ilkka Niiniluoto. *Truthlikeness.* Dordrecht, Holland: D. Reidel Publishing Company, 1987.

[Odd79] Graham Oddie. "Verisimilitude and Distance in Logical Space". In: *The Logic and Epistemology of Scientific Change.* Ed. by Ilkka Niiniluoto and Raimo Tuomela. Vol. 30. Acta Philosophica Fennica. North-Holland, 1979, pp. 243–264.

[Pap95] E. Pap. *Null-Additive Set Functions.* Kluwer Academic Publishers, 1995.

[Qui96] J. Ross Quinlan. "Improved Use of Continuous Attributes in C4.5". In: *Journal of Artificial Intelligence Research* 4 (1996), pp. 77–90.

[RA78] D. A. Ralescu and C. Adams. "The Fuzzy Integral". In: *Journal of Mathematical Analysis and Applications* 62.1 (1978), pp. 12–24.

[Rei+19] Markus Reichstein et al. "Deep learning and process understanding for data-driven Earth system science". In: *Nature* 566.7743 (2019), pp. 195–204.

[Rob03] P. Robinson. "Fuzzy Measures and Geographical Information Systems". In: *Geographical and Environmental Modelling* 7.2 (2003), pp. 175–198.

[Sal+19] Ahmed Salem et al. "Updates Leak: Data Set Inference and Reconstruction Attacks in Online Learning". In: *Proceedings of the 29th USENIX Security Symposium.* USENIX Association, 2019, pp. 1291–1308. URL: https://www.usenix.org/conference/usenixsecurity20/presentation/salem.

[Šer63] A. N. Šerstnev. "On certain generalizations of the triangle inequality in probabilistic metric spaces". In: *Siberian Mathematical Journal* 4 (1963), pp. 112–124.

[Sha48] Claude E. Shannon. "A Mathematical Theory of Communication". In: *Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[She69] H. Sherwood. "On E-spaces and their relation to other classes of probabilistic metric spaces". In: *Journal of the London Mathematical Society* 44 (1969), pp. 441–448.

[She80] Roger N. Shepard. "Multidimensional Scaling, Tree-Fitting, and Clustering". In: *Science* 210.4468 (1980), pp. 390–398.

[SS83] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*. Elsvier-North-Holland, 1983.

[Ste68] S. S. Stevens. "Metrically generated probabilistic metric spaces". In: *Fundamenta Mathematicae* 61 (1968), pp. 259–269.

[Sug74] M. Sugeno. "Theory of Fuzzy Integrals and Its Applications". PhD thesis. Tokyo Institute of Technology, 1974.

[TN07] V. Torra and Y. Narukawa. *Modelling Decision: Information Fusion and Aggregation Operators*. Springer, 2007.

[TN16] V. Torra and G. Navarro-Arribas. "Integral Privacy". In: *Cryptology and Network Security: 15th International Conference, CANS 2016, Milan, Italy, November 14-16, 2016, Proceedings*. Vol. 10052. Lecture Notes in Computer Science. Springer, 2016, pp. 389–399. DOI: 10.1007/978-3-319-48965-0_44. URL: https://link.springer.com/chapter/10.1007/978-3-319-48965-0_44.

[TN18] V. Torra and G. Navarro-Arribas. "Probabilistic Metric Spaces for Privacy by Design Machine Learning Algorithms: Modeling Database Changes". In: *Data Privacy Management, Cryptocurrencies and Blockchain Technology: ESORICS 2018 International Workshops, DPM 2018 and CBT 2018, Barcelona, Spain, September 6-7, 2018, Proceedings*. Vol. 11025. Lecture Notes in Computer Science. Springer, 2018, pp. 422–430. DOI: 10.1007/978-3-030-00305-0_30. URL: https://link.springer.com/chapter/10.1007/978-3-030-00305-0_30.

[TNS14] V. Torra, Y. Narukawa, and M. Sugeno. *Non-Additive Measure—Theory and Applications. Studies in Fuzziness and Soft Computing*. Vol. 310. Springer, 2014.

[Wal43] Abraham Wald. "On Some Systems of Equations of Mathematical Economics". In: *Econometrica* 11.4 (1943), pp. 367–403.

[WK92] Z. Wang and G. J. Klir. *Fuzzy Measure Theory*. Springer Science Business Media, 1992.

[Yag81] R. R. Yager. "A Procedure for Ordering Fuzzy Subsets of the Unit Interval". In: *Information Sciences* 24.2 (1981), pp. 143–161.

[Zad65] L. A. Zadeh. "Fuzzy Sets". In: *Information and Control* 8.3 (1965), pp. 338–353.

[Zad78] L. A. Zadeh. "Fuzzy Sets as a Basis for a Theory of Possibility". In: *Fuzzy Sets and Systems* 1.1 (1978), pp. 3–28.