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

Copyright © 2025 by Mariam Taha

Except:

Paper I
The Authors, published by Springer under CC BY 4.0 License, 2021.

Paper II
The Authors, published by Elsevier B.V. under CC BY 4.0 License, 2023.

Paper III
The Authors, under exclusive license to Springer Nature Switzerland AG, 2023.

Paper IV
The Authors, published by Elsevier B.V., under CC BY 4.0 License, 2025.

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-

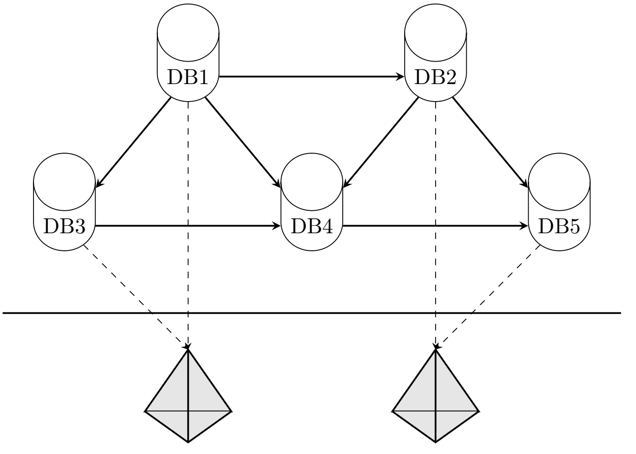


Figure 1.1: Graphical representation of databases (cylinders) and machine learning models (3D pyramids)

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:*

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

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:*

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

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. $\square$

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

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

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 $A, B \in \mathcal{A}$,*

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

- *Superadditivity: μ is superadditive if:*

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

- *Subadditivity: μ is subadditive if:*

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

- *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*

$$\mu(E) = |\mathit{Proj}\ E|,$$

where

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

*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*

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

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

then, the fuzzy measure defined by

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

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:

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

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

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

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.

Paper I

The space of models in machine learning: using Markov chains to model transitions

Vicenç Torra, Mariam Taha, and Guillermo Navarro-Arribas.

Progress in Artificial Intelligence, 10, 321-332 (2021).

Progress in Artificial Intelligence (2021) 10:321–332
https://doi.org/10.1007/s13748-021-00242-6

REGULAR PAPER

The space of models in machine learning: using Markov chains to model transitions

Vicenç Torra[1,2,3] · Mariam Taha[4] · Guillermo Navarro-Arribas[5]

Received: 4 April 2020 / Accepted: 26 March 2021 / Published online: 12 April 2021
© The Author(s) 2021

Abstract
Machine and statistical learning is about constructing models from data. Data is usually understood as a set of records, a database. Nevertheless, databases are not static but change over time. We can understand this as follows: there is a space of possible databases and a database during its lifetime transits this space. Therefore, we may consider transitions between databases, and the database space. NoSQL databases also fit with this representation. In addition, when we learn models from databases, we can also consider the space of models. Naturally, there are relationships between the space of data and the space of models. Any transition in the space of data may correspond to a transition in the space of models. We argue that a better understanding of the space of data and the space of models, as well as the relationships between these two spaces is basic for machine and statistical learning. The relationship between these two spaces can be exploited in several contexts as, e.g., in model selection and data privacy. We consider that this relationship between spaces is also fundamental to understand generalization and overfitting. In this paper, we develop these ideas. Then, we consider a distance on the space of models based on a distance on the space of data. More particularly, we consider distance distribution functions and probabilistic metric spaces on the space of data and the space of models. Our modelization of changes in databases is based on Markov chains and transition matrices. This modelization is used in the definition of distances. We provide examples of our definitions.

Keywords Machine and statistical learning models · Space of data · Space of models · Hypothesis space · Probabilistic metric spaces

1 Introduction

Machine and statistical learning can be seen as a search problem. That is, we have a state space corresponding to possible models and we want to find the one that better represents our data. Here, *better* can correspond to the one with best accuracy. Different definitions of *better* as well as different search strategies to find a good solution can be considered. From this perspective, we consider operators that permit to transform one model into another one. In this case, a transformation is usually to improve accuracy (i.e., *better* = *better accuracy)*. Examples of these transformations include operators that expand a node in a decision tree, update weights in a deep learning model, or operators that mutate a solution in genetic algorithms.

In this paper, we consider a different perspective. We consider the space of models taking into account the space of data that have generated these models.

When we learn a model from an actual database, the database is just a database from the space of data. When databases change, we are traversing the space of data through a particular path. Different databases in this particular path can lead to different machine learning models.

Vicenç Torra
vtorra@ieee.org

Mariam Taha
mariamt@cs.umu.se

Guillermo Navarro-Arribas
guillermo.navarro@uab.cat

1 Department Computing Science, Umeå University, Umeå, Sweden

2 Hamilton Institute, Maynooth University, Maynooth, Ireland

3 School of Informatics, University of Skövde, Skövde, Sweden

4 Department Computing Science, Umeå University, Umeå, Sweden

5 Dept. Information and Communications Engineering – CYBERCAT, Universitat Autònoma de Barcelona, Bellaterra, Catalonia, Spain

We claim that it is necessary to study how the space of data interacts with the space of models. More particularly, that any decision on the space of models has to take into account relationships between the space of data that generate these models.

We consider that this perspective is of great interest in the following areas.

- *In model selection for statistical and machine learning* In this area the goal is to select models that better generalize data and avoid overfitting. The study of the interaction between the space of models and the space of data can increase our understanding on the models themselves, and on the methods that generate the models (comparing their respective mappings between the two spaces). In particular, we think it is fundamental to understand the concepts of generalization and overfitting. This also relates to the effect of outliers and influential points in learning. It is important to understand generalization and overfitting in terms of the relationship between the space of models and the space of data.
- *In privacy preserving data mining and machine learning* The need to study the relationship between the two spaces was first proposed in [14] in the context of integral privacy [10,11]. In short, a model is integrally private if it can be generated by a large number of databases, and these databases are sufficiently different (e.g., they do not share records). This is to avoid some type of privacy attacks on machine learning models as, e.g., membership attacks [12].

We are interested in knowing when two models are similar, where similar does not correspond to a syntactic similarity of the models (e.g., if two decision trees have the same structure), nor on a semantic similarity of the models (e.g., if two models have the same accuracy). We are interested in knowing when models are similar because they have been generated from similar databases.

There are naturally different ways of understanding the similarity between databases. For example, one database may be similar because it is a noisy version of the other (e.g., an anonymized version of the original database). Here, we focus on changes in databases due to the natural processes a database suffers in a company. That is, we consider a database that is updated, as time passes, by means of e.g., adding and removing records. These types of changes are usual when databases are in production. In addition, these types of changes are also relevant in the framework of data privacy [13] with the right to be forgotten and the right to amend (under the GDPR).

In [14], the authors proposed the use of probabilistic metric spaces [9] for modeling the similarity between models. Informally, these spaces are defined in terms of distance distribution functions. That is, distance between pairs of objects are not a real number but a distribution on these numbers. This approach permits us to define a distance between pairs of models taking into account the distance between the set of databases that have generated these models.

In this paper, we propose the use of Markov chains and transition matrices to represent, respectively, sequences of changes in databases and the probability of changes taking place. This representation permits the definition of probabilistic metric spaces on the space of data. We use them later to define distance distribution functions for the space of models in terms of the databases that have generated them. This is a much simpler approach than the one introduced in [14].

The structure of this paper is as follows. In Sect. 2, we introduce the definitions that are needed later in the paper. In particular, we introduce Markov chains and probabilistic metric spaces. In Sect. 3, we introduce two definitions of metric spaces for databases based on Markov chains and prove some results. In Sect. 4, we introduce definitions for distance distribution functions for models based on the probabilistic metric spaces introduced in Sect. 3. We provide some examples of how these distances can be actually computed. The paper finishes with a discussion and lines for future work.

2 Preliminaries

In this section, we review some concepts that are needed later. We begin with Markov chains and transition matrices. We also discuss probabilistic metric spaces and distances for sets of elements.

2.1 Markov chains

In this paper, we will use Markov transition matrices and Markov chains to model the space of databases. Because of that, we will review in this section a few concepts that we need later. See e.g., [7] for details.

We consider a state space S finite or enumerable. We will use

$$S = \{DB_1, DB_2, DB_3, \ldots\}$$

to denote the space of possible databases. Thus, in our case, a finite although extremely huge set.

We will consider chains defined on the state space S. That is, $(Z_n)_{n\in\mathbb{N}}$ taking values in S, i.e., $Z_n \in S$. More particularly, we consider Markov chains. This corresponds to chains in which the probability distribution on Z_{n+1} depends only on the process Z_n at time n and not on previous values of Z. In other words, there is no memory on previous transitions.

Formally,

$$P(Z_{n+1} = DB_j | Z_n = DB_i, Z_{n-1} = DB_{n-1}, \ldots, Z_0 = DB_0) \\ = P(Z_{n+1} = DB_j | Z_n = DB_i).$$

We consider time-homogeneous Markov chains. That is, the probability of transition does not depend on time. This is expressed mathematically as

$$P(Z_{n+1} = DB_j | Z_n = DB_i) = P(Z_{m+1} = DB_j | Z_m = DB_i)$$

for any n, m.

As the probability does not depend on n, we will not use this index unless required. Then, we will use P_{ij} to denote $P(Z_{n+1} = DB_j | Z_n = DB_i)$ (for any n). For the sake of simplicity, we will also use $P(Z_{n+1} = j | Z_n = i)$ when no confusion arises.

From the explanation above, it is clear that transition depends only on the probabilities P_{ij}. These probabilities for all states i and j define a matrix. It is known as transition matrix. Formally, a transition matrix P is a $S \times S$ matrix with values in [0, 1] such that (for any n)

$$\sum_{DB_j \in S} P_{ij} = \sum_{DB_j \in S} P(Z_{n+1} = DB_j | Z_n = DB_i) = 1.$$

We can prove that given a probability distribution π on S for time 0, say probabilities $P(Z_0 = i)$ for $i \in S$, the probability distribution for time 1, say probabilities $P(Z_1 = j)$ for $j \in S$, can be expressed in matrix notation as πP. Let us denote by P^n the transition matrix defined by $P^n_{ij} = P(Z_{m+n} = DB_j | Z_m = DB_i)$. Naturally, the computation of P^n_{ij} does not depend on m. We can prove that

$$P^{r+t}_{ij} = \sum_{k \in S} P^r_{ik} P^t_{kj},$$

or in matrix form $P^{r+t} = P^r P^t$. This is called the Chapman–Kolmogorov equation.

2.2 Probabilistic metric spaces

Probabilistic metric spaces [9] are a generalization of metric spaces in which a distance distribution function replaces the role of distance functions. That is, instead of considering $d(a, b)$ as a real number, it is a distribution function on the real numbers.

Recall that metric spaces are defined in terms of sets (a non-empty set) and a distance or metric for pairs of elements in this set. Formally, we denote a metric space by (S, d), where S is the set and d for $a, b \in S$ the distance. The function d is required to satisfy the following properties: (i) positiveness, (ii) symmetry, and (iii) triangle inequality (formally, $d(a, b) \leq d(a, c) + d(c, b)$ for any a, b, c in S). Also, it is usual to require that if a and b are different then the distance should be strictly positive. Special names are given when some of these conditions fail. For example, when the distance does not satisfy the symmetry condition, we say that (S, d) is a quasimetric space; and when the distance does not satisfy the triangle inequality, we say that (S, d) is a semimetric space.

Probabilistic metric spaces are a generalization of metric spaces. As stated above, we can informally consider that we replace the function $d(a, b)$ by a distribution function $F(a, b)$ defined on $\mathbb{R}$. These functions are known as distance distribution functions. Their definition follows.

Definition 1 [9] A nondecreasing function F defined on $\mathbb{R}^+$ that satisfies (i) $F(0) = 0$; (ii) $F(\infty) = 1$, and (iii) that is left continuous on $(0, \infty)$ is a distance distribution function. Δ^+ denotes the set of all distance distribution functions.

The following interpretation is usual for these functions: $F(x)$ corresponds to the probability that the distance is less than or equal to x. Note that this definition is a generalization of a distance.

In particular, we use ϵ_a to denote the distance distribution function that represents the classical distance a. This ϵ_a function is just a step function at a. Its definition follows.

Definition 2 [9] For any a in $\mathbb{R}$, we define ϵ_a as the function given by

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

In order to define probabilistic metric spaces we need to consider the set of distance distribution functions, and we need to define a condition on triples of functions in this set analogous to the triangle equality in metric spaces. This condition given below is based on triangle functions. Let us start defining the triangle functions.

Definition 3 [9] Let Δ^+ be defined as above, then a binary operation on Δ^+ is a triangle function if it is commutative, associative, and nondecreasing in each place, and has ϵ_0 as the identity.

Using triangle functions we can establish the definition of probabilistic metric spaces.

Definition 4 [9] Let $(S, \mathcal{F}, \tau)$ be a triple where S is a nonempty set, $\mathcal{F}$ is a function from $S \times S$ into Δ^+, τ is a triangle function; then $(S, \mathcal{F}, \tau)$ is a probabilistic metric space if the following conditions are satisfied for all p, q, and r in S:

- (i) $\mathcal{F}(p, p) = \epsilon_0$
- (ii) $\mathcal{F}(p, q) \neq \epsilon_0$ if $p \neq q$
- (iii) $\mathcal{F}(p, q) = \mathcal{F}(q, p)$
- (iv) $\mathcal{F}(p, r) \geq \tau(\mathcal{F}(p, q), \mathcal{F}(q, r))$.

As usual in this field, we will use F_{pq} instead of $\mathcal{F}(p, q)$. This permits to express the value of the distance distribution function at x by means of the expression: $F_{pq}(x)$.

2.3 Metrics for sets of objects

In order to define the distance between pairs of models, we will consider the set of databases that have generated these models. This permits to define the distance in terms of the distance between these sets. Let us first consider the classical setting with a standard distance.

Let G be an algorithm that given a database generates a model, then, the set of generators of a model m is defined by $Gen_m = \{DB|G(DB) = m\}$. In this case, given two models m_1 and m_2 we define the distance between m_1 and m_2 in terms of Gen_{m_1} and Gen_{m_2}. In order to do so, we need to extend the distance for databases to sets of databases.

Nevertheless, given a metric space (S, d), its extension to a set of elements of S is not trivial. This is so because although several distances have been defined on sets, not all of them satisfy the triangle inequality. This implies that they are not valid to define a metric.

In [14], different types of functions are considered. The discussion includes the Hausdorff distance and the sum of minimum distance (which do not satisfy the triangle inequality) and the definition by Eiter and Mannila [3] that is indeed a valid definition of a distance and leads to a metric space. Nevertheless, this is a very complex function to compute.

3 Probabilistic metric spaces from Markov chains

We consider that transition matrices are a suitable approach to model changes on databases. In other words, we consider that for a given database there is some probability that this database is transformed by means of a modification to another database. For the sake of simplicity, we consider in this work time-homogeneous Markov chains. That is, as explained in Sect. 2 that changes on a database only depend on what is currently available in the database and that it does not depend on its previous values (history of the database). This assumption can be considered simplistic, as e.g., the probability of adding a record may depend on how many times has been already present in the database and has been removed. Nevertheless, we consider that this assumption is acceptable for this initial study.

For illustration, we consider only addition and deletion of records from a database, and that only one addition and one deletion is allowed at a time. We also assume that we have access to the whole population or that we know the size of the whole population. Then, we can define a transition matrix based on assigning a probability of having a deletion (p_d) and a probability of having an addition (p_a). Naturally, these probabilities add less than or equal to one ($p_d + p_a \leq 1$).

Definition 5 Let p_d and p_a be the probability of deleting or adding a record. Then, given an arbitrary database DB_i, where DB_i is a subset of the whole population P (with $|P|$ denoting the size of this population), we define the probability of transition from DB_i to any DB_j as follows (here, p_{ij} stands for $P(Z_{n+1} = DB_j | Z_n = DB_i)$ as above):

$$p_{ij} = \begin{cases} p_d \frac{1}{|DB_i|} & \text{if } c_1 \& c_3 \\ p_a \frac{1}{|P|-|DB_i|} & \text{if } c_2 \& c_3 \\ \frac{1}{|P|} & \text{if } (c_1 \text{ or } c_2) \& c_4 \\ 1-(p_d+p_a) & \text{if } c_5 \& c_3 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where $c_1 - c_5$ are the following conditions:

- c_1: $DB_j \subset DB_i$ and $|DB_i \setminus DB_j| = 1$
- c_2: $DB_i \subset DB_j$ and $|DB_j \setminus DB_i| = 1$
- c_3 : $|DB_i| \notin \{0, |P|\}$
- c_4: $|DB_i| \in \{0, |P|\}$
- c_5: $DB_i = DB_j$

Here, c_4 means that the database DB_i is either empty or no further records can be added to it, and c_3 means that DB_i is not one of such extreme databases.

Lemma 1 *The above definition leads to a valid transition matrix. That is,* $\sum_j p_{ij} = 1$ *for all* j.

Proof Observe that given DB_i, p_{ij} is not zero for databases $DB_j \neq DB_i$ that have either one additional record more (i.e., $DB_i \subset DB_j$ and $|DB_j \setminus DB_i| = 1$) or less (i.e., $DB_j \subset DB_i$ and $|DB_i \setminus DB_j| = 1$) than DB_i. Then, in the general case, DB_i can lead to any of the $|DB_i|$ databases that has just one record less, or DB_i can lead to any of the $|P| - |DB_i|$ databases that have exactly one additional record. So, according to Equation 1, the probability of deleting a record is p_d and the probability of adding a record is p_a. As the probability of DB_i not being modified is $1 - (p_d + p_a)$, it is proved that the definition leads to a row adding to one. Then, we have conditions for the extreme cases in which DB_i is empty or DB_i includes all records. In this case, there are $|P|$ neighboring databases with a probability of transition equal to $\frac{1}{|P|}$. Therefore, this row also adds to one. Therefore, the matrix is a transition matrix. □

In this section, we introduce two definitions of probabilistic metric spaces for databases based on transition matrices and Markov chains.

The first definition considers the distance between two databases in terms of the probability of being transformed into the second one. This approach defines the probabilistic metric space solely based on the transition matrices. We give below both symmetric and asymmetric definitions for the distance distribution functions. See Definition 6. We call this type of space, visited database-based probabilistic metric space (VD-PMS).

The second definition considers the distance between two databases in terms of their evolution. That is, given two databases, will they be similar as time passes? In order to give a formal definition, we need to consider how databases are being modified, and what similarity means for databases. With respect to the later, the model presumes the existence of a standard distance function (a metric space, in fact) on the space of databases. We call this type of space, database distance-based probabilistic metric space (DD-PMS). See Definition 7.

We consider that both types of definitions are relevant for statistical and machine learning and, in particular, for privacy preserving data mining. We are interested in models that are valid today but that will be also valid in the future. So, the first definition states that two models are similar if we can transit from one to the other and the second definition states that two models are similar if they have a *similar future* (a similar machine learning model in the future).

3.1 Visited database-based probabilistic metric spaces

We consider a definition of probabilistic metric spaces for databases based on [5]. The distance between two databases depends on the probability that one database becomes the second one after a sequence of changes (there is a chain between the first to the second) within a given time frame.

The definition is based on transition matrices P on the space of databases. The definition follows.

Definition 6 Let S be a state space representing the space of databases, let P be the transition matrix for S that defines a time-homogeneous Markov chain $(Z_n)_{n\in\mathbb{N}}$. Then, given two states i and j in S we define

$$F_{ij}(t) = P[\text{exists a time } s < t \text{ such that } Z_s = j | Z_0 = i].$$

Formally, let f_{ij}^s denote the probability that with $Z_0 = i$ (i.e., starting the chain from state i), the first time we visit state j is exactly at time s. Then, $F_{ij}(t) = \sum_{s=1}^{t} f_{ij}^s$.

From the point of view of the space of databases, the definition above establishes that the distance between two databases DB_1 and DB_2 for the value t is α (i.e., $P_{12}(t) = \alpha$) if the probability of reaching DB_2 from DB_1 in less than t transitions is α.

We can prove from this definition that $F_{ij}(t_1 + t_2) \geq F_{ik}(t_1)F_{kj}(t_2)$. From this property, we can prove the following theorem. See [5] for a proof. Observe that the formulation of the following theorem in [5] uses stationary to refer to time-homogeneous, using the notation in [2].

Theorem 1 *Let S, P, $(Z_n)_{n\in\mathbb{N}}$ and $F_{ij}(t)$ be defined as in Definition 6. Let $\mathcal{F}$ be the mapping from $S \times S$ into the space of cumulative distribution functions defined by $\mathcal{F}(i, j) = F_{ij}$. Then, $\mathcal{F}$ satisfies properties (i) and (iii) in Definition 1, and properties (i), (ii), and (iv) in Definition 4.*

It is a non-symmetric distance distribution function satisfying (iv) under the t-norm $T = Prod$ (i.e., $T(a, b) = ab$).

The hitting time of a state DB_j starting from state DB_i is the random variable defined by

$$T_{ij} = \min\{n \geq 0 : X_n = DB_j\}$$

with the minimum of the empty set defined as ∞. The probability of hitting state DB_j is defined by

$$h_i^j = P(T_{ij} < \infty).$$

Not all transition matrices lead to Markov chains with probabilities of hitting a state equal to 1. If this is the case, then, the Definition above will lead to a probabilistic metric space with a non-symmetric function. We establish this in the next theorem.

Theorem 2 *Let S, P, $(Z_n)_{n\in\mathbb{N}}$ and $F_{ij}(t)$ be defined as in Definition 6. Then, the pair $(S, \mathcal{F})$ is a probabilistic metric space with a non-symmetric distance function under the t-norm $T = Prod$ when $h_i^j = 1$ for all i, j.*

Definition 6 gives a distance that is not necessarily symmetric. Note that accessing j from i at time t does not mean that it is possible to access i from j in the same time t.

It is possible to define a probabilistic metric space with a symmetric distance function using $F'_{ij} = \sqrt{F_{ij}F_{ji}}$ or $F''_{ij} = 0.5(F_{ij} + F_{ji})$. The first definition is a probabilistic metric space satisfying the Menger inequality with $T = Prod$ and the second one satisfies the Menger inequality with $T = T_m$ (i.e., $T_m(a, b) = \max(a + b - 1, 0)$).

3.2 Computation and example

Definitions above use f_{ij}^s, that as explained above, denotes the probability that with $Z_0 = i$ the first time we visit state j is exactly at time s when we start at $Z_0 = i$. For a given transition matrix P, we can compute f_{ij}^s as follows. If $s = 1$,

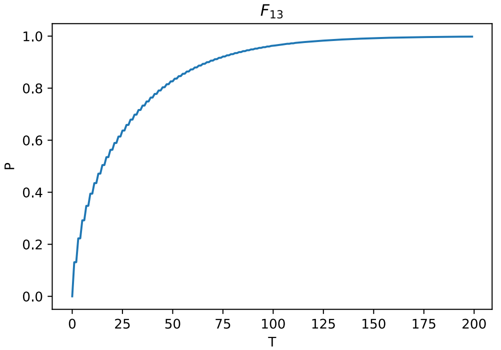


Fig. 1 F_{13} for Example 1 according to Definition 6

$f_{ij}^s = P_{ij}$. If $s > 1$ then we perform the following steps: (i) define $\bar{P}$ as P and assigning $\bar{P}_{rs} = 0$ for all pairs (j, t) and (t, j) with $j \in S$, (ii) compute $\bar{P}^{s-1}$, (iii) compute $f_{ij}^s = \sum_{k \in S} \bar{P}_{ik}^{s-1} P_{kj}$. Finally, (iv) using the expression in Definition 6 we compute $F_{ij}(t)$ as $F_{ij}(t) = \sum_{s \le t} f_{ij}^s$.

The rationale of this definition is that in order to reach j from i in exactly s steps (with $s > 1$) we need to reach any other state $k \neq j$ in exactly $s-1$ steps without reaching j at any moment $s' < s$, and then move from k to j. Probabilities of reaching $k \neq j$ from i in $s-1$ steps without hitting j will be computed using $\bar{P}$. This computation corresponds to compute $\bar{P}^{s-1}$ as noted above.

We give an example of this computation with a very small database. The space is built from a set of 5 records. In this way, we can consider the whole database space that has a size of 2^5 databases.

Example 1 Let DB be the set of all databases that can be generated from 5 records. That is, DB corresponds to the power set of these 5 records. Let P be the transition matrix of DB defined according to Definition 5 using $p_a = p_d = 0.5$. Then, we can compute F_{13} according to Definition 6. F_{13} is the distance between databases $DB_1 = \{a\}$ and $DB_3 = \{c\}$. Figure 1 represents this computation.

3.3 Results on the approximation of distance distribution functions

Computation of the distance introduced in Definition 6 is costly. Because of that we are interested in the approximation of this distance. We can prove the following results in relation to Definition 6.

Lemma 2 *Let DB_1 and DB_2 be two databases. Let us consider the sets $DB_1 \cap DB_2$, $DB_1 \setminus DB_2$, and $DB_2 \setminus DB_1$. Let $t_a = |DB_2 \setminus DB_1|$ be the number of elements we need to add to transit from DB_1 to DB_2, and let $t_d = |DB_1 \setminus DB_2|$ be the elements we need to delete to transit from DB_1 to DB_2. Then, the shortest chain from DB_1 to DB_2 when we only consider addition and deletion of records has length*

$$t_0 = |DB_1 \setminus DB_2| + |DB_2 \setminus DB_1| = t_a + t_d$$

and, therefore,

$$F_{12}(t) = 0$$

for all $t < t_0$.

Let us consider an arbitrary order for the t_a elements we add, and an arbitrary order for the t_d elements we remove. Let i in $\{1, \ldots, t_a\}$ represent the addition of the ith element according to this order and i in $\{t_a + 1, \ldots, t_a + t_d\}$ the removal of the $(i - t_a)$th element according to this order. Using this interpretation, it is clear that any permutation of $\{1, \ldots, t_a + t_d\}$ represents a valid chain with only additions and deletions and with no cycles from DB_1 to DB_2. So, there are $(t_a + t_d)!$ valid chains with no cycles.

We can also prove a lemma similar to Lemma 2 when in addition to addition and deletion we allow transitions that do not change the database. Shortest chains will of course still have length $t_a + t_d$, and from this it also follows: $F_{12}(t) = 0$ for all $t < t_0$.

Let $\mathcal{C}_{12}^t$ denote all valid chains from DB_1 to DB_2 with length t. Then, $\mathcal{C}_{12}^{t_a+t_d}$ will represent all shortest chains. Therefore, $|\mathcal{C}_{12}^{t_a+t_d}| = (t_a + t_d)!$. It is also clear that $\mathcal{C}_{12}^t = \emptyset$ for $t < t_a + t_d$.

Let use denote a chain $c \in \mathcal{C}^t$ by $c_0, c_1, \ldots, c_t$. Here c_i will correspond to a database DB_i. Then, the probability of transiting from DB_0 to DB_t through the chain c is naturally

$$P_c = \prod_{c_r \in c} P_{c_{r-1}, c_r}. \tag{2}$$

Lemma 3 *Using the notation in Lemma 2 and P_c as in Equation 2, we have that when only addition and deletion are allowed, or when addition, deletion and transition without change are allowed, it holds (for t_a and t_b as above)*

$$\begin{aligned} F_{12}(t_a + t_d) = f_{12}^{t_a+t_d} &= \sum_{c \in \mathcal{C}_{12}^{t_a+t_d}} P_c \\ &= \sum_{c \in \mathcal{C}_{12}^{t_a+t_d}} \prod_{c_r \in c} P_{c_{r-1}, c_r}. \end{aligned} \tag{3}$$

We can also prove that when only addition and deletions are allowed, $\mathcal{C}_{12}^t = \emptyset$ for any $t = t_a + t_d + 1 + 2k$ for any k, i.e., given a shortest chain, we can only enlarge this chain both adding and removing k records. So, in this case, for all k it holds $f_{12}^{t_a+t_d+2k+1} = 0$.

When addition, deletion and also non modification are allowed as transitions between databases, we have that for $t = t_a + t_d + 1 + 2k$ (for any k), the chains in $\mathcal{C}^t_{12}$ are the ones in $\mathcal{C}^{t-1}_{12}$ adding a transition corresponding to non-modification (say t_{nm}). Let $p_\emptyset$ denote the probability of a non-modification (this corresponds to the value $1 - (p_d + p_a)$ in Equation 1). Let $c \in \mathcal{C}^t$ be one of such chains with elements $c_0, c_1, \ldots, c_t$. Then, we can insert this t_{nm} transition between any pair of elements of the chain (but not at the end of the chain as we are considering that is exactly at time s that we reach the goal state). This means that there are t options. Given $P(c)$, the probability of the chain $c \in C^t$, the probability of any of these chains is $p_\emptyset \cdot p(c)$. So, as we have t new chains for a given chain c, the probability for this set of chains (say $\tilde{c}$) is $p(\tilde{c}) = t p_\emptyset p(c)$. Then, considering all $\tilde{c}$ generated from all $c \in \mathcal{C}^t$, we have that for $t = t_a + t_d + 1 + 2k$

$$P(\mathcal{C}^t_{12}) = \sum_{c \in \mathcal{C}^{t-1}_{12}} p(\tilde{c}) = \sum_{c \in \mathcal{C}^{t-1}_{12}} t p_\emptyset p(c)$$
$$= t p_\emptyset \sum_{c \in \mathcal{C}^{t-1}_{12}} p(c) = t p_\emptyset P(\mathcal{C}^{t-1}_{12}).$$

Using Expression 3, it is easy to prove the following lemma.

Lemma 4 *Let DB_i and DB_j be two arbitrary databases, and let $\mathcal{R} = \{c\}_c$ be a set of random valid chains $c \in \mathcal{R}$ from DB_i to DB_j with different lengths. Then,*

$$f^s_{ij} \geq \sum_{c:|c|=s+1} P_c$$

and

$$F_{ij}(t) \geq \sum_{c:|c|\leq t+1} P_c.$$

This result implies that the consideration of random valid chains give lower bounds for $F_{ij}(t)$. Therefore, any decision based on a threshold th on a given t (i.e., $F_{ij}(t) \geq th$) valid for a set $\mathcal{R}$ will be also valid if all the set of chains is considered.

The links between triangle functions and t-norms (see e.g., [1], and Def. 7.1.3 and Section 7.1 in [9]) permit us to establish the following lemma. This lemma establishes another lower bound for distance distribution functions for any pair of databases if we can compute exact values for pairs involving a particular database (e.g., a reference one denoted by DB_a below).

Lemma 5 *Let $(S, \mathcal{F}, \tau_T)$ be a probabilistic metric space generated by a t-norm T (i.e., $\tau_T(F, G)(x) = T(F(x), G(x))$ is the triangle function generated by T). Let DB_a be a particular database for which we can calculate exactly the distance distribution function for all $DB_i \in S$. Then,*

$$\mathcal{F}(DB_i, DB_j) \geq T(F_{DB_i,DB_a}(x), F_{DB_a,DB_j}(x))$$

is a lower bound of $\mathcal{F}_{DB_i,DB_j}$.

Proof If T is a t-norm, then the triangle function τ_T generated by T is

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

for distance distribution functions F and G. Then, as $(S, \mathcal{F}, \tau_T)$ is a probabilistic metric space on the space of databases, we know that for all p, q, and r it holds that

$$\mathcal{F}(p, r) \geq \tau_T(\mathcal{F}(p, q), \mathcal{F}(q, r))$$

and, therefore, in particular for $p = DB_i$, $q = DB_a$ and $r = DB_j$ we have that

$$\mathcal{F}(DB_i, DB_j) \geq \tau_T(\mathcal{F}(DB_i, DB_a), \mathcal{F}(DB_a, DB_j))$$
$$= T(\mathcal{F}(DB_i, DB_a), \mathcal{F}(DB_a, DB_j)).$$

□

3.4 Database distance-based probabilistic metric space

We consider an alternative way to define probabilistic metric spaces in which in addition to a transition matrix we use a distance on the state space. The definition is based on [5].

Definition 7 Let S be the database space, let P be the transition matrix for S that defines a time-homogeneous Markov chain $(Z_n)_{n\in\mathbb{N}}$. Let $d : S \times S \to \mathbb{R}^+$ be a distance function on S. Then, for any given time $t \geq 0$, we define the function $F^t_{ij}(x)$ as follows:

$$F^t_{ij}(x) = Pr[d(i, j) < x \text{ at time } t]$$
$$= \sum_{k\in S} P^t_{ik} \left(\sum_{\ell:d(\ell,k)<x} P^t_{j\ell} \right).$$

Informally, for a given time t, this definition implies that the probability level between states i and j at time x is computed in terms of the probability of reaching states k at time t from i and the probability of finding states ℓ from j at most at distance x.

We can prove the following result that is similar to Lemma 4.

Lemma 6 *Let $\mathcal{DB}$ be a collection of databases sampled from the space of databases. Let*

$$\tilde{F}^t_{ij}(x; \mathcal{DB}) = \sum_{DB_k \in \mathcal{DB}} P^t_{ik} \sum_{\ell} : d(\ell, k) < x DB_\ell \in \mathcal{DB} P^t_{j\ell}.$$

Let $\tilde{F}^t_{ij}(x; \mathcal{DB}, \mathcal{R})$ correspond to $\tilde{F}^t_{ij}(x; \mathcal{DB})$ when P^t_{ik} only considers a given set $\mathcal{R}$ of random valid chains as in Lemma 4. Then,

$$\begin{aligned}\tilde{F}^t_{ij}(x; \mathcal{DB}, \mathcal{R}) & \\ &\leq \tilde{F}^t_{ij}(x; \mathcal{DB}) \\ &\leq \sum_{DB_k \in \mathcal{DB}} P^t_{ik} \sum_{\ell : d(\ell,k) < x, DB_\ell \in \mathcal{DB}} P^t_{j\ell}.\end{aligned}$$

The implications of this lemma are similar to the ones of Lemma 4. That is, the consideration of random chains and sets of databases give lower bounds for $F_{ij}(t)$. Therefore, any decision based on considering two databases DB_1 and DB_2 as different based on a threshold th on a given t (i.e., $F_{12}(t) \geq th$) valid for a set $\mathcal{R}$ and for a set of databases $\mathcal{DB}$ will be also valid if all the sets of chains and all databases are also considered. We illustrate this distance with one example that uses the same space of databases we have considered before.

Example 2 Let P be the transition matrix and let DB be the space of databases considered in Example 1. Let DB_1 and DB_3 be the databases considered in Example 1. We can compute F_{13} according to Definition 7. To do so, we use the Jaccard Index to measure the similarities between the databases. Figures 2, 3 and 4 represent these computations with different values of t. We display $F_{1,3}$ as in the previous example, but we also considered the computations for very different databases (i.e., $F_{0,31}$) in Figures 5, 6 and 7. Here, $F_{0,31}$ is the distance between databases $DB_0 = \{\}$ and $DB_{31} = \{a, b, c, d, e\}$.

From the three figures, Figs. 2, 3 and 4, we notice that when t becomes larger, and x is greater than 0.5, the probability become higher.

Figures 5, 6 and 7) show that when we have very different databases, and the Jaccard distance is less than 0.5, we need more transitions in order to increase the probability.

4 Construction of the distance on the space of models

The definitions above permit to extend the probabilistic metric space for databases to models. As discussed in Sect. 2.3, given two models m_1 and m_2 the goal is to define a distance based on the generators G_{m_1} and G_{m_2} of m_1 and m_2. In this

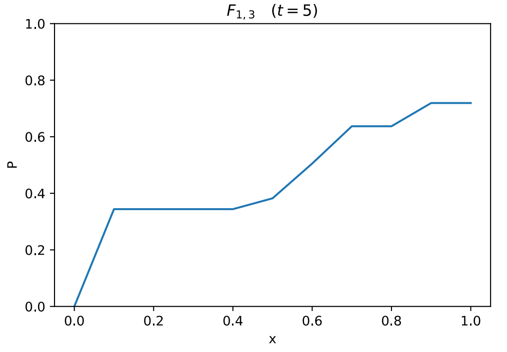


Fig. 2 F_{13} when $t = 5$ for Example 2 according to Definition 7

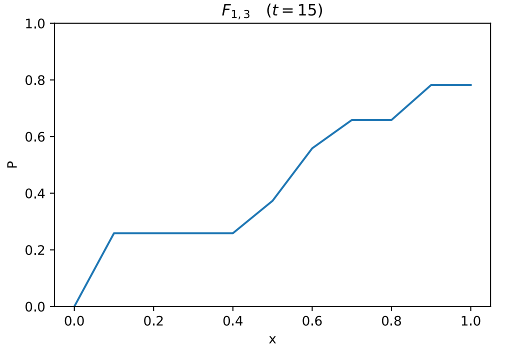


Fig. 3 F_{13} when $t = 15$ for Example 2 according to Definition 7

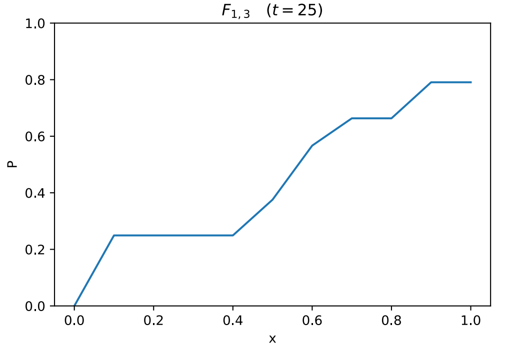


Fig. 4 F_{13} when $t = 25$ for Example 2 according to Definition 7

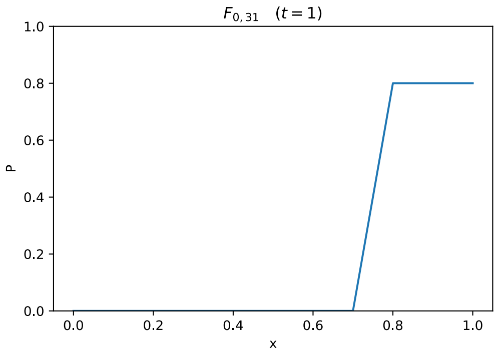


Fig. 5 $F_{0,31}$ when $t = 1$ for Example 2 according to Definition 7

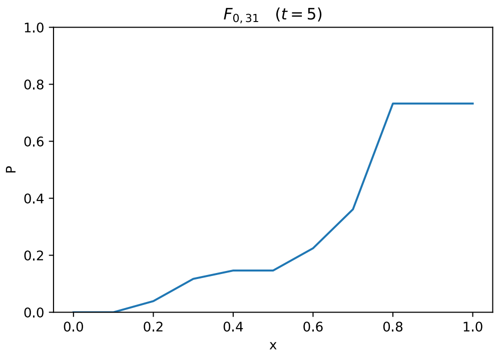


Fig. 6 $F_{0,31}$ when $t = 5$ for Example 2 according to Definition 7

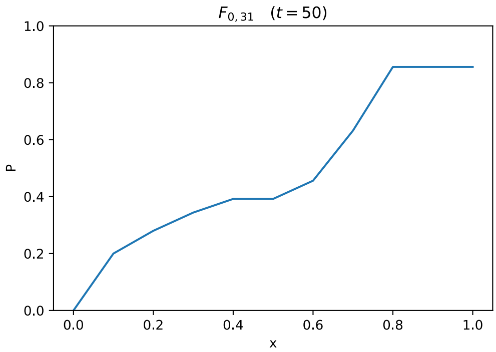


Fig. 7 $F_{0,31}$ when $t = 50$ for Example 2 according to Definition 7

Table 1 Set of models and their corresponding sets of Databases from Example 3

Model m	Gen(m)
1000	(a), (b), (a, b), (a, b, c).
1500	(a, c), (b, c)
2000	(c)

case, instead of a standard distance, we consider a distance distribution function.

Proposition 1 *Let S be the space of databases, let G be an algorithm to generate models from the space of databases S, let $\mathcal{G}$ be the space of models that can be generated by G. Let m_1 and m_2 be two models generated by the application of algorithm G to databases in S. Let Gen_{m_1} and Gen_{m_2} be the set of databases that generate m_1 and m_2. That is, $Gen_{m_1} = \{DB \in S | G(DB) = m_1\}$ and $Gen_{m_2} = \{DB \in S | G(DB) = m_2\}$.*

Let $(S, \mathcal{F})$ be a probabilistic metric space. Then, let $\mathcal{F}$ for pairs of models m_1 and m_2 be defined as follows:

$$\mathcal{F}(m_1, m_2)(x) = \frac{1}{|Gen_{m_1}||Gen_{m_2}|} \sum_{DB_1 \in Gen_{m_1}} \sum_{DB_2 \in Gen_{m_2}} \mathcal{F}_{DB_1, DB_2}(x). \tag{4}$$

Then, $\mathcal{F}$ is a distance distribution function.

Lemma 7 *Using Equation 4 with Definition 6, we obtain a function $\mathcal{F}(m_1, m_2)$ that is a non-symmetric distance distribution function. Using instead definitions F' and F'' above will lead to symmetric distance functions. Using Definition 7 results into a distance distribution function.*

Lemma 8 *When we approximate $\mathcal{F}(m_1, m_2)(x)$ using lower bounds of $\mathcal{F}_{DB_1, DB_2}$ (as considering only some chains and some databases), we will obtain lower bounds of the real $\mathcal{F}(m_1, m_2)(x)$.*

It is relevant to point out that this definition does not necessarily lead to a probabilistic metric space, as condition (iv) in Definition 4 does not always hold. We illustrate this definition above considering databases as in the previous examples based on three records/people and their salaries.

Example 3 Suppose we have three records a, b, and c, with salaries 1000, 1000 and 2000, respectively. The space of databases is the power set of theses records. If we choose G to be the median function to generate the models, then the space of models is $\mathcal{G} = \{1000, 1500, 2000\}$. The models and their generators are listed in Table 1. Figure 8 displays the distance between model m_1=1000 and model m_2=2000 by

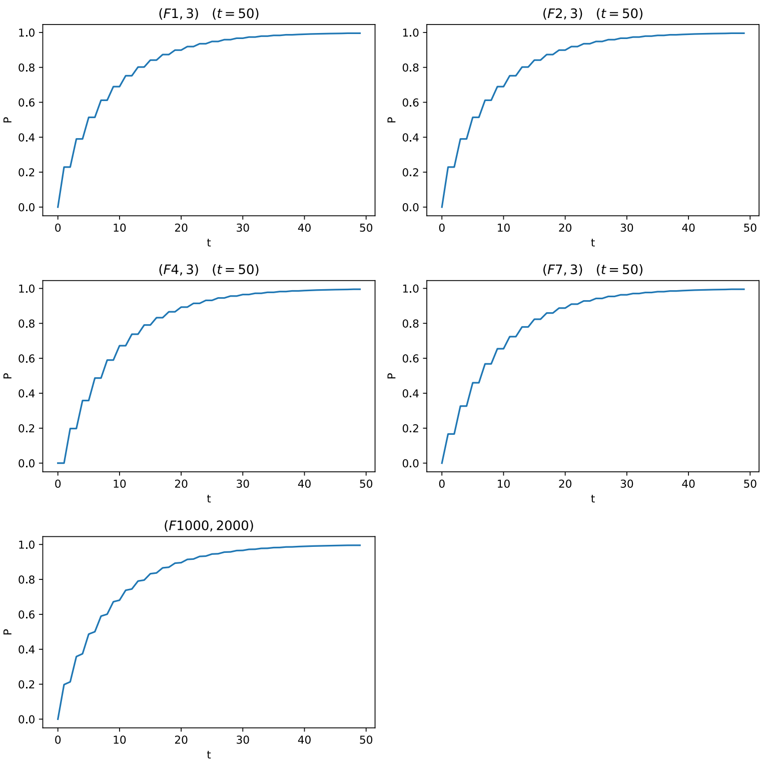


Fig. 8 $F_{1000,2000}$ for Example 3 according to Proposition 1 and Definition 6

using Proposition 1, as well as the distance between pairs of databases according to Definition 6. Similarly, we have also computed the distance between the same pair by using the same proposition, but where the distance between databases follows Definition 7 as illustrated in Fig. 9.

From both Fig. 8, and Fig. 9, we can see that the distance distribution functions for the models and the databases are quite similar.

5 Summary and conclusions

In this paper, we have proposed the use of Markov chains and transition matrices to model transitions between databases, and used them to define a probabilistic metric space for models.

Our goal is to better understand the relationship between data and models. From our perspective, this requires a metric space on the space of models that reflects the relationships between the databases that can generate these models. From a machine learning perspective, a good model is one that has a good accuracy, but also that is not overfitted to data and has some level of generalization. From a data privacy perspective, a good model is one that does not lead to disclosure. This includes not leading to disclosure on the data that has been used to generate the model. In other words, we understand machine and statistical learning as a selection process. We want to select a model with good accuracy that does not have overfitting (and not vulnerable to membership attacks) and that is *near* to models with similar generators. This work is to formalize what *near* means.

As future work, we plan to develop strategies for computing these distances and for defining in practice metric spaces for real-size databases. In this paper, examples have been described for small databases because they are easier to understand but also because when considering a regular size database its power set becomes extremely large. Some initial results on boundary conditions on the distances were given in the paper. We plan to consider how to extend this approach by means of approximating the distances.

We also plan to work on the problem of model selection. Research on graphical visualization of the models and the metric spaces will be appropriate here. Sammon's map as

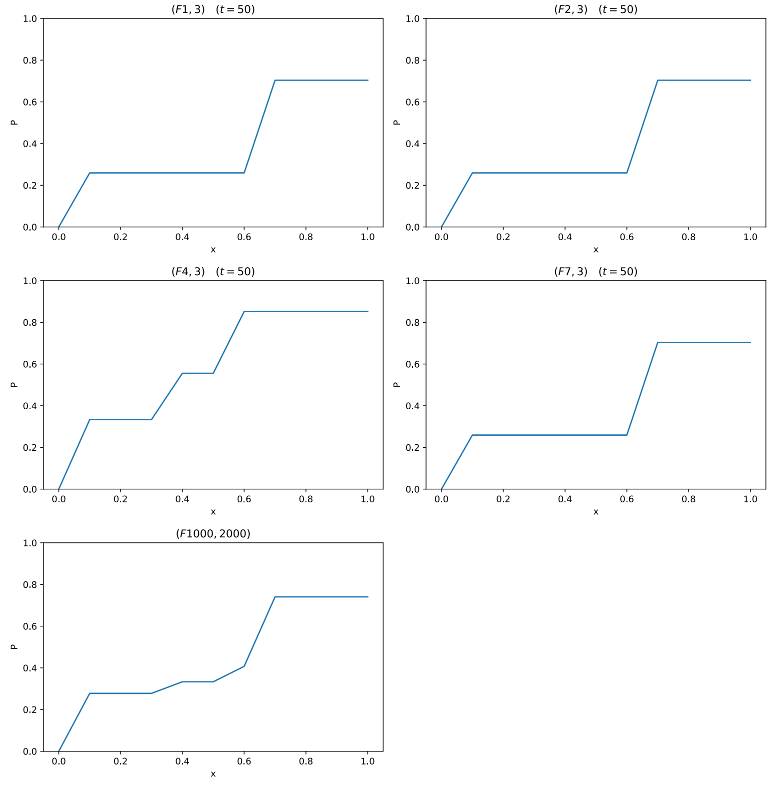


Fig. 9 $F_{1000,2000}$ for Example 3 according to Proposition 1 and Definition 7

well as other multidimensional scaling procedures can help on this purpose.

Acknowledgements This study was partially funded by Vetenskapsrådet project "Disclosure risk and transparency in big data privacy" (VR 2016-03346, 2017-2020), Spanish project TIN2017-87211-R is gratefully acknowledged, and by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

Funding Open access funding provided by Umea University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Alsina, C., Frank, M.J., Schweizer, B.: Associative Functions: Triangular Norms and Copulas. World Scientific, Singapore (2006)
2. Doob, J.L.: Stochastic Processes. Wiley, Hoboken (1953)
3. Eiter, T., Mannila, H.: Distance measures for point sets and their computation. Acta Informatica **34**, 109–133 (1997)
4. Kent, D.C., Richardson, G.D.: Ordered probabilistic metric spaces. J. Austral. Math. Soc. **46**, 88–99 (1989)
5. Marcus, P.S.: Probabilistic metric spaces constructed from stationary Markov chains. Aequationes Mathematica **15**, 169–171 (1977)
6. Moynihan, R.: Probabilistic metric spaces induced by Markov chains. Z. Wahrscheinlichkeitstheorie. Gebiete **35**, 177–187 (1976)
7. Privault, N.: Understanding Markov Chains. Springer, Newyork (2018)
8. Samarati, P.: Protecting respondents identities in microdata release. IEEE Trans. Knowl. Data Eng. **13**(6), 1010–1027 (2001)
9. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. Elsevier, Amsterdam (1983)
10. Senavirathne, N., Torra, V.: Approximating robust linear regression with an integral privacy guarantee. Proc. PST **2018**, 1–10 (2018)
11. Senavirathne, N., Torra, V.: Integral privacy compliant statistics computation. Proc DPM/CBT- ESORICS **2019**, 22–38 (2019)
12. Shokri, R., Stronati, M., Song, C., Shmatikov, V.: Membership inference attacks against machine learning models. Proc. IEEE Symposium on Security and Privacy. (2017). arXiv:1610.05820

13. Torra, V.: Data Privacy: Foundations, New Developments and the Big Data Challenge. Springer, Newyork (2017)
14. Torra, V., Navarro-Arribas, G.: Probabilistic metric spaces for privacy by design machine learning algorithms: modeling database changes, Proc. DPM 2018/CBT 2018. LNCS **11025**, 422–430 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Paper II

On the definition of probabilistic metric spaces by means of fuzzy measures

Yasuo Narukawa, Mariam Taha, and Vicenç Torra.

Fuzzy Sets and Systems, 465, Article 108528 (2023).

On the definition of probabilistic metric spaces by means of fuzzy measures

Yasuo Narukawa[a], Mariam Taha[b], Vicenç Torra[c]

[a]*Department of Management Science, Tamagawa University, 6-1-1 Tamagawagakuen, Machida, Tokyo, 194-8610, Japan*
`nrkwy@eng.tamagawa.ac.jp`
[b]*Dept. Computing Sciences, UmeåUniversity, Sweden*
`mariamt@cs.umu.se`
[c]*Dept. Computing Sciences, UmeåUniversity, Sweden*
`vtorra@ieee.org`

Abstract

Metric spaces are defined in terms of a space and a metric, or distance. Probabilistic metric spaces are a useful extension of metric spaces where the distance is a distribution instead of a number. In this way, we can take into account uncertainty. Then, the triangle inequality is replaced by a condition based on triangle functions on the distributions.

In this paper we introduce F-spaces. This is a new type of probabilistic metric spaces which is based on fuzzy measures (also known as non-additive measures and capacities). We prove some properties that describe which families of fuzzy measures are compatible with which type of triangle functions. Then, we show how we can use Sugeno, Choquet integrals, and, in general, any other fuzzy integral as a tool for building these spaces.

We show how these results can be used to compute distances between functions. We illustrate the example comparing three types of means when applied to a set of databases. The example uses Sugeno λ-measures to illustrate the theoretical results presented in the paper.

Keywords: Fuzzy measures, fuzzy integrals, probabilistic metric spaces.

1. Introduction

Probabilistic metric spaces extend metric spaces in the sense that the metric is no longer a real number but a distribution function. These spaces were studied by several authors, among others Schweizer and Sklar, and much of the results are summarized in their book [18].

It is well known that t-norms and t-conorms extensively used in fuzzy sets and systems, both in theory and applications, have their origins in the studies of probabilistic metric spaces. See e.g. [1, 10, 8] and references therein.

The definition of probabilistic metric spaces is not so easy or straightforward in practice. The literature provides a few results that ensure that a given construction leads to a probabilistic metric space. This is the case of E-spaces, a family of probabilistic metric spaces that are constructed in terms of a set of functions and a probability space. These spaces were studied by Sherwood [20] and Stevens [21]. More precisely, E-spaces define distances between functions between a space Ω and another M. For this we need a probability on Ω and a metric in M.

In this paper we introduce F-spaces, where instead of using probabilities on Ω we use fuzzy measures [27, 7]. We study some of the properties of these new spaces, and show that under some conditions on the measures (i.e., for some types of measures) we obtain a probabilistic metric space under certain t-norms.

We then show how we can exploit Choquet and Sugeno integrals as a basic brick to define these new spaces. It is well known that fuzzy measures, also known as non-additive measures and capacities, permit to represent the interactions between the elements of the reference set. They have been extensively used in applications [6, 2, 3, 4, 5, 9, 16, 15, 27, 30]. When used to define F-spaces, we can represent interactions in the objects of the space. This is not possible in E-spaces.

We also illustrate our results with an example. Its motivation is in the context of machine learning and statistics. More particularly, we are interested in defining a distance between two models or statistics taking into account the possible generators of these models. That is, we consider the space of models (the hypothesis space) with the goal of defining a distance between pairs m_1, m_2 of models in this space. Our objective is that the distance is based on the sets of databases that generate m_1 (i.e., we denote this set by $Gen(m_1)$) and m_2 (i.e., $Gen(m_2)$). Then, we consider probabilistic metric spaces instead of metric spaces. Therefore, the distance $d(m_1, m_2)$ will be a distribution function instead of a real number. As we will see, the distribution function is based on all databases DB in $Gen(m_1)$ and $Gen(m_2)$, and the distances between them. This example is based on previous works [25, 29] where we discuss the convenience of probabilistic metric spaces in the context of model selection, and its special relevance in data privacy. Our approach based on fuzzy measures will permit to model the interaction between the objects (i.e., databases) in the space. As we will see, fuzzy measures will be defined on the space of databases, and, therefore, they can model the interaction (i.e., redundancy and complementarity [23]) in terms of a property (e.g., coverage) of a set of databases.

The structure of this paper is as follows. In Section 2 we review previous definitions and results, mainly related to t-norms, and fuzzy measures and integrals. Then, in Section 3 we review relevant results related to probabilistic metric spaces. Section 4 presents our main results. This includes the definition of F-spaces, and several results showing that a construction using fuzzy measures leads to F-spaces. Section 5 includes some small examples of our results and Section 6 describe the application of our approach. The paper finishes with some conclusions.

2. Preliminaries

This section is divided in two parts. We begin reviewing some results based on t-norms, and then on fuzzy measures.

2.1. t-Norms

We will use in this paper the concept of t-norm [1, 10]. They are functions on $[0,1] \times [0,1]$ that generalize Boolean conjunction. The definition follows:

Definition 1. *A function $\top : [0,1] \times [0,1] \to [0,1]$ is a t-norm if and only if it satisfies the following properties:*

(i) $\top(x,y) = \top(y,x)$ *(symmetry or commutativity)*

(ii) $\top(\top(x,y),z) = \top(x,\top(y,z))$ *(associativity)*

(iii) $\top(x,y) \leq \top(x',y')$ *if $x \leq x'$ and $y \leq y'$ (monotonicity)*

(iv) $\top(x,1) = x$ *for all x (neutral element 1)*

They are usually required to satisfy also continuity and subidempotency ($\top(x,x) < x$ for $x \neq 0$). Such t-norms are called Archimedean t-norms.

Some examples of t-norms follow.

Example 1. *The following functions are t-norms.*

- **Minimum:** $\top(x,y) = \min(x,y)$. *The minimum is often denoted by* $\wedge$. *That is,* $x \wedge y = \min(x,y)$.

- **Algebraic product:** $\top(x,y) = xy$. *This t-norm is denoted by* $\Pi(x,y)$ *following [18].*

- **Bounded difference or Lukasiewicz t-norm:** $\top(x,y) = \max(0, x+y-1)$. *This t-norm is denoted by* $W(x,y)$ *following [18].*

- **Yager family:** $\top_w(x,y) = 1 - \min\left(1, ((1-x)^w + (1-y)^w)^{1/w}\right)$ *for* $w \geq 0$.

- **Drastic:** $\top_d(x,y) = y$ *if* $x = 1$, $\top_d(x,y) = x$ *if* $y = 1$, *and* $\top_d(x,y) = 0$ *otherwise.*

It easy to see that t-norms are proper generalizations for conjunctions, as, for all of them, $\top(0,0) = \top(0,1) = \top(1,0) = 0$ and $\top(1,1) = 1$.

2.2. Fuzzy measures

Let us review in this section fuzzy measures and some of their properties. See e.g. [27, 7, 13, 14] for details. These measures are also known as capacities and fuzzy measures.

Definition 2. *Let* $(\Omega, \mathcal{A})$ *be a measurable space. A set function* μ *defined on* $\mathcal{A}$ *is called a fuzzy measure if and only if*

- $0 \leq \mu(A) \leq \infty$ *for any* $A \in \mathcal{A}$*;*

- $\mu(\emptyset) = 0$*;*

- *If* $A_1, A_2 \subseteq \mathcal{A}$ *and* $A_1 \subset A_2$ *then*

$$\mu(A_1) \leq \mu(A_2)$$

Some definitions and properties of fuzzy measures follow.

Definition 3. *Let* μ *be a fuzzy measure on the measurable space* $(\Omega, \mathcal{A})$. *Then,*

- μ *is additive if* $\mu(A \cup B) = \mu(A) + \mu(B)$ *when* $A \cap B = \emptyset$*;*

- μ *is submodular if* $\mu(A) + \mu(B) \geq \mu(A \cup B) + \mu(A \cap B)$*;*

- *μ is supermodular if $\mu(A)+\mu(B) \leq \mu(A \cup B)+\mu(A \cap B)$. A supermodular measure implies superadditivy. That is, $\mu(A \cup B) \geq \mu(A) + \mu(B)$.*

In this paper we will use, to build our example, a well-known family of measures: Sugeno λ-measures. We use this family of measures because the selection of λ makes easy to build submodular and supermodular measures. The theory is not limited to this type of measures but it is general for all types of fuzzy measures. The definition follows.

Definition 4. *Let Ω be a finite set and let $\lambda > -1$. A Sugeno λ-measure is a fuzzy measure $\mu : 2^{\Omega} \to [0,1]$ such that*

- *$\mu(\Omega) = 1$*
- *if $A, B \subseteq \mathbf{\Omega}$ with $A \cap B = \emptyset$ then*
  $\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B)$

The measure of the singletons $x_i \in \Omega$ is often called a density and it is denoted by $v(x_i)$. For any Sugeno λ-measure μ with values $v(x_i)$, λ should satisfy the following:

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

This means, that given any assignment v, the value λ is completely determined from the equation above. The determination of λ from the values on the singletons is proved and explained in e.g. [11, 22, 27, 26].

Therefore, any fuzzy measure $\mu(A)$ can be defined from the singletons as follows (through first the computation of λ, if unknown).

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

For Sugeno λ-measures it is known that when $\lambda > 0, \mu$ is supermodular. Whereas, when $\lambda > 0, \mu$ is submodular (see e.g. [27, 26]).

Definition 5. *Let φ be a real valued function on a closed interval $[c, d]$. Then,*

- *φ is said to be convex if*

$$\varphi(\beta x + (1-\beta)y) \leq \beta\varphi(x) + (1-\beta)\varphi(y)$$

  for $x, y \in [c, d]$, $0 < \beta < 1$.

- *φ is said to be concave if*

$$\varphi(\beta x + (1-\beta)y) \geq \beta\varphi(x) + (1-\beta)\varphi(y)$$

  for $x, y \in [c, d]$, $0 < \beta < 1$.

Proposition 1. [13] Let μ be a fuzzy measure on $(\Omega, \mathcal{A})$, and $\varphi : [0,1] \to [0,1]$ be a non-decreasing function with $\varphi(0) = 0$ and $\varphi(1) = 1$. Let λ be the Lebesgue measure. Then, the following holds:

1. If φ is convex, then $\varphi \circ \lambda$ is supermodular.
2. If φ is concave, then $\varphi \circ \lambda$ is submodular.

The same applies for distorted probabilities. They are measures of the form $\mu = \varphi \circ P$, where P is a probability distribution.

As we just said, distorted probabilities are fuzzy measures representable by a non-decreasing function φ and a probability distribution P. They were generalized to m-dimensional distorted probabilities [12, 28, 24] which for an appropriate m permit to represent any measure. They are defined in terms of a function φ and m probabilities $P_1, \ldots, P_m$. Distorted probabilities generalize the Sugeno λ-measure defined here, and m-distorted probabilities are such that they permit to extend these restricted type of measures to any one possible.

We introduce now another example of fuzzy measures that will be used later.

Definition 6. *Let $A_0 \neq \emptyset$ be a subset of Ω, then we define the set function μ_{A_0} as $\mu_{A_0}(A) = 1$ if and only if $A_0 \subseteq A$ and $\mu_{A_0}(A) = 0$ otherwise.*

It is easy to prove the following.

Proposition 2. *The set function μ_{A_0} is a fuzzy measure.*

This measure can be represented by a distorted probability but not as a Sugeno λ-measure.

The integration of a function with respect to a fuzzy measure can be done using different types of fuzzy integrals. Choquet and Sugeno integrals are the most important ones, but there are several generalizations as well [15, 16]. Let us represent them by $(C)\int f d\mu$ and $(S)\int f d\mu$, respectively. Given a measure μ and a function f such that the fuzzy integral on the reference set is one, we can build another measure μ' as follows:

$$\mu'(A) = (C)\int_A f d\mu. \tag{2}$$

3. Probabilistic metric spaces

In this section we review the concept of probabilistic metric spaces. To do so we begin with the concept of metric spaces and then the definitions of distance distribution functions and triangle functions. In this way we can introduce Menger spaces. Then, we introduce E-spaces, a type of Menger space.

3.1. Menger spaces

A metric space (see e.g. [19]) is defined in terms of a set S and a function $d : S \times S \to \mathbb{R}^+$ that plays the role of distance on the set S. Here, we understand $\mathbb{R}^+ = [0, \infty)$ and $\overline{\mathbb{R}^+} = [0, \infty]$.

Definition 7. *Let $d : S \times S \to \mathbb{R}^+$, then d is called a metric on S if the following properties hold for $a, b, c \in S$:*

- *$d(a, b) \geq 0$ with equality if and only if $a = b$ (positive property),*
- *$d(a, b) = d(b, a)$ (symmetry property), and*
- *$d(a, b) \leq d(a, c) + d(c, b)$ (triangle inequality property).*

Definition 8. *The pair (S, d) where d is a metric on S is a metric space and $d(a, b)$ is the distance between a and b.*

The pair (S, d) where d is a function $S \times S \to \mathbb{R}^+$ that satisfies the positive property and the triangle inequality (but not the symmetry property) is a quasimetric space. The pair (S, d) where d is a function $S \times S \to \mathbb{R}^+$ that satisfies positive property and the symmetry property (but not the triangle inequality) is a semimetric space.

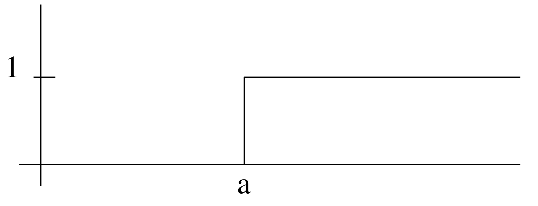


Figure 1: Graphical representation of the distance distribution function ϵ_a.

Probabilistic metric spaces [18, 8] were introduced as a generalization of the concept of a metric. They replace the distance function in a metric space by a distance distribution function. So, the *distance* between a pair of elements in S is not a number but a distribution on these numbers. We introduce this concept below.

Definition 9. *[18] A non-decreasing function F defined on $\overline{\mathbb{R}^+}$ that satisfies (i) $F(0) = 0$; (ii) $F(\infty) = 1$, and (iii) that is left continuous on $(0, \infty)$ is a distance distribution function.*

Δ^+ denotes the set of all distance distribution functions.

In this definition we can understand $F(x)$ as the probability that the distance is less than or equal to x. In this way we can write any classical distance a in terms of a distance distribution function. More particularly, we will use in this case ϵ_a defined as follows. Naturally, ϵ_a is a step function at a. Figure 1 illustrates ϵ_a.

Definition 10. *[18] (Def. 4.1.4) For any a in $\mathbb{R}^+$, we define $\epsilon_a \in \Delta^+$ by*

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

The next step towards the definition of a probabilistic metric space is to consider a counterpart of the triangle inequality. Triangle functions will be used for this purpose. We review them below.

Definition 11. *[18] Let Δ^+ be the set of all distance distribution functions. Then, a binary operation on Δ^+ is a triangle function if it is commutative, associative, and non-decreasing in each place, and has ϵ_0 as the identity.*

There are several families of triangle functions. Two important families establish a connection between triangle functions and t-norms [1]. In particular, for a t-norm $\top$, we have that the function $\tau^{\top}(F,G)(x) = \top(F(x), G(x))$ is a triangle function (see Def. 7.1.3 and Section 7.1 in [18]). The maximal triangle function is $\tau^{\min}$ (Theorem 7.1.4 in [18]). Another family of triangle functions built from a t-norm is the following one. We will use this family in our work. For $x \geq 0$,

$$\tau_{\top}(F,G)(x) = \sup\{\top(F(u), G(v)) | u + v = x\}.$$

See Def. 7.2.1 (Theorem 7.2.4) and Section 7.2 in [18].

We are now in conditions to define probabilistic metric spaces.

Definition 12. *[18] Let $(S, \mathcal{F}, \tau)$ be a triple where S is a nonempty set, $\mathcal{F}$ is a function from $S \times S$ into Δ^+, and τ is a triangle function; then $(S, \mathcal{F}, \tau)$ is a probabilistic metric space (PM space) if the following conditions are satisfied for all p, q, and r in S:*

1. $\mathcal{F}(p,p) = \epsilon_0$
2. $\mathcal{F}(p,q) \neq \epsilon_0$ *if* $p \neq q$
3. $\mathcal{F}(p,q) = \mathcal{F}(q,p)$
4. $\mathcal{F}(p,r) \geq \tau(\mathcal{F}(p,q), \mathcal{F}(q,r))$.

Given a probabilistic metric space $(S, \mathcal{F}, \tau)$, we say that $(S, \mathcal{F})$ is a probabilistic metric space under τ.

A probabilistic pseudometric space (PPM space) $(S, \mathcal{F}, \tau)$ 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 3 is not required we have a probabilistic quasimetric space.

We prefer to use F_{pq} instead of $\mathcal{F}(p,q)$. Then, we express the value of the latter at x simply as $F_{pq}(x)$.

We consider in this paper particular probabilistic metric spaces. The next definition introduces Menger spaces.

Definition 13. *[18] Let $(S, \mathcal{F}, \tau)$ be a probabilistic metric space. Then $(S, \mathcal{F}, \tau)$ is proper if*

$$\tau(\epsilon_a, \epsilon_b) \geq \epsilon_{a+b}$$

for all a, b in $\mathbb{R}^+$.

If $\tau = \tau_{\top}$ for some t-norm $\top$, then $(S, \mathcal{F}, \tau)$ is a Menger space, or equivalently $(S, \mathcal{F})$ is a Menger space under $\top$.

We illustrate this definition with an example that describes several proper Menger spaces. We will use these examples in the next section to prove some results.

Example 2. *Let $(S, \mathcal{F}, \tau)$ be a probabilistic metric space and $a, b \in \mathbb{R}^+$, with $a \geq b$.*

- *Suppose that τ is minimum. That is, $\tau = \wedge$. Then, since $a \wedge b < a + b$, we have $\tau(\epsilon_a, \epsilon_b) \geq \epsilon_{a+b}$. Therefore $(S, \mathcal{F}, \tau)$ is a proper Menger space.*
- *Suppose that τ is the algebraic product. That is, $\tau = \Pi$. Then, since $\epsilon_a \cdot \epsilon_b = \epsilon_a$, we have $\tau(\epsilon_a, \epsilon_b) \geq \epsilon_{a+b}$. Therefore $(S, \mathcal{F}, \tau)$ is also a proper Menger space.*
- *Suppose that τ is the bounded difference. That is, $\tau = W$. Then, since $0 \vee (\epsilon_a + \epsilon_b - 1) = \epsilon_a$, it follows that $(S, \mathcal{F}, \tau)$ is a proper Menger space.*

3.2. E-spaces

This is a family of probabilistic metric spaces [20, 21] that are constructed in terms of a set of functions and a probability space. For any pair of functions, and any x we can compute the measure of the points that are a distance at most x. The definition of the E-spaces uses a probability to measure the set of points. As discussed by Schweizer and Sklar this can be seen as a generalization of just using the Lebesgue measure on the $I = [0, 1]$ interval.

Formally, given functions p and q from I to a metric space (M, d) with $d(a, b) = |a - b|$, and considering the Lebesgue measure λ, we can define for $\Omega = I = [0, 1]$ the following distance distribution function:

$$F_{pq}(x) = \lambda(\{t \in I| \ \ |p(t) - q(t)| < x\}).$$

This expression is a particular case of an E-space. As we see in the next definition, E-spaces are defined considering a probability space on Ω instead of a space with the Lebesgue measure on $I = [0, 1]$. The definition also uses an arbitrary metric space (M, d). The definition follows. In the definition, $L_1^+(\Omega)$ is the set of all positive a.e. finite Lebesgue measurable functions on Ω. Figure 2 gives a representation of E-spaces.

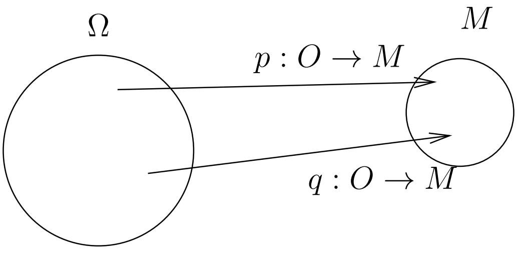


Figure 2: Representation for E-spaces with the probability space $(\Omega, \mathcal{A}, P)$ and the metric space (M, d).

Definition 14. *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 $\mathcal{A}$-measurable and therefore in $L_1^+(\Omega)$.

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

$$F_{pq}(x) = P(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}). \tag{3}$$

Equation 3 implies that $\mathcal{F}$ satisfies Properties 1 and 3 in Definition 12. If $\mathcal{F}$ also satisfies Property 2, then $(S, \mathcal{F})$ is a canonical E-space.

The following can be proven for E-spaces. The proof of this theorem is given in [17] and also in [21].

Theorem 1. *Let $(S, \mathcal{F})$ be an E-space. Then $(S, \mathcal{F})$ is a probabilistic pseudometric space under τ_W. If $(S, \mathcal{F})$ is canonical, then it is a Menger space under W.*

4. Main results

A natural generalization of E-spaces is to consider fuzzy measures for evaluating the set of ω that are at most at a *distance* x. This is proposed in the next definition. We call it F-space (e.g., for Fuzzy Measure, and consecutive letter to E).

Definition 15. *Let $(\Omega, \mathcal{A})$ be a measurable space, and let μ be a fuzzy measure on $(\Omega, \mathcal{A})$. 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 F-space with base $(\Omega, \mathcal{A}, \mu)$ 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 $\mathcal{A}$.
- *(ii) For all p, q in S, $\mathcal{F}(p, q) = F^{\mu}_{pq}$ with*
$$F^{\mu}_{pq}(x) = \mu(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}). \tag{4}$$

We can prove the following.

Theorem 2. *Let $(\Omega, \mathcal{A})$ be a measurable space, let μ be a fuzzy measure on $(\Omega, \mathcal{A})$ and $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$ and target (M, d).*

Then, if μ is a supermodular fuzzy measure on $(\Omega, \mathcal{A})$, it follows that $(S, \mathcal{F})$ is a probabilistic pseudometric space under bounded difference τ_W.

Proof. We prove first that F^{μ}_{pq} as defined in Equation 4 satisfies Property 1 in Definition 12. Observe that if $p = q$ then $p(w) = q(w)$ for all $w \in \Omega$. Therefore, $d(p(w), q(w)) = 0$ for all w in Ω and $\Omega = \{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}$ for all $x > 0$. As $\mu(\Omega) = 1$, then $F^{\mu}_{pq}(0) = 0$ and $F^{\mu}_{pq}(x) = 1$ for all $x > 0$. Therefore, $F^{\mu}_{pq} = \epsilon_0$ and the Property is proven.

The proof that F^{μ}_{pq} satisfies Property 3 is trivial. The symmetry of d naturally implies the one of F^{μ}_{pq}.

We now prove that F^{μ}_{pq} satisfies Property 4. Let us consider any x in $\mathbb{R}^+$. Then, consider u and v such that $u + v = x$ and define the following sets:

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

- $B = \{w \in \Omega | d(q(w), r(w)) < v\}$, and
- $C = \{w \in \Omega | d(p(w), r(w)) < x\}$.

We know that d satisfies the triangle inequality.

Therefore, $C \supseteq A \cap B$ because if we have that for a w it holds $d(p(w), q(w)) = u_0 < u$ (i.e., $w \in A$) and $d(q(w), r(w)) = v_0 < v$ (i.e., $w \in B$) we know that by the triangle inequality

$$\begin{aligned} x &= u + v > u_0 + v_0 \\ &= d(p(q), q(w)) + d(q(w), r(w)) \geq d(p(w), r(w)), \end{aligned}$$

and thus w is in C.

Now, as $C \supseteq A \cap B$, by the monotonicity condition of μ, we have

$$\mu(C) \geq \mu(A \cap B)$$

and by supermodularity

$$\mu(A \cap B) \geq \mu(A) + \mu(B) - \mu(A \cup B).$$

As $1 = \mu(\Omega) \geq \mu(A \cup B)$, we have that

$$\mu(A \cap B) \geq \mu(A) + \mu(B) - \mu(A \cup B) \geq \mu(A) + \mu(B) - 1,$$

and naturally $\mu(A \cap B) \geq 0$. Therefore,

$$\begin{aligned} \mu(C) &\geq \mu(A \cap B) \geq \max(\mu(A) + \mu(B) - 1, 0) \\ &= W(\mu(A), \mu(B)). \end{aligned} \tag{5}$$

Let us consider the expressions for $F^{\mu}_{pq}(u)$, $F^{\mu}_{qr}(v)$, and $F^{\mu}_{pr}(x)$ according to Equation 4 and the sets $\mu(A)$, $\mu(B)$, and $\mu(C)$ as defined above. Then, we have that Equation 5 implies

$$F^{\mu}_{pr}(x) \geq W(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)).$$

Therefore,

$$\begin{aligned} F^{\mu}_{pr}(x) &\geq \sup\{W(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)) | u + v = x\} \\ &= \tau_W(F^{\mu}_{pq}, F^{\mu}_{qr})(x), \end{aligned} \tag{6}$$

and Property 4 holds with τ_W.

As Properties 1, 3, and 4 hold, the theorem is proven. □

The next proposition is immediate from Proposition 1. Recall that a convex function makes the distorted probability supermodular. Therefore, Theorem 2 applies.

Proposition 3. *Let* $(\Omega, \mathcal{A})$ *be a measurable space, and let* P *be a probability on* $(\Omega, \mathcal{A})$*. Let* φ *be an increasing convex function on the closed interval [0,1] with* $\varphi(0) = 0$*,* $\varphi(1) = 1$*. Let* $(S, \mathcal{F})$ *be an F-space with base* $(\Omega, \mathcal{A}, \varphi \circ P)$*.*

Then, $(S, \mathcal{F})$ *is a probabilistic pseudometric space under bounded difference* τ_W*.*

The next theorem is immediate from Example 2.

Theorem 3. *If* $(S, \mathcal{F})$ *is a canonical F-space (i.e.,* $\mathcal{F}$ *satisfies Property 2), then it is a proper Menger space under W.*

Definition 16. *Given the reference set* Ω*, a fuzzy measure such that* $\mu(A) < 1$ *for all* $A \subset \Omega$ *and* $A \neq \Omega$ *is called a* 1^-*-measure*

Theorem 4. *Let* $(S, \mathcal{F})$ *be an F-space. Let* μ *be a* 1^-*-measure on* $(\Omega, \mathcal{A})$*. Then* $(S, \mathcal{F})$ *is a probabilistic pseudometric space under* $\tau_{\top_d}$*.*

Proof. The proof of this theorem follows the proof of Theorem 2. In particular, conditions 1 and 3 are proven in the same way.

Then, we define the sets A, B, and C in the same way as we defined them in Theorem 2. We have, therefore $C \supseteq A \cap B$. By the monotonicity condition, it follows that $\mu(C) \geq \mu(A \cap B)$.

Now, let us consider the following case: $A \neq \Omega$ and $B \neq \Omega$. We can prove that $\top_d(\mu(A), \mu(B)) = 0$ because both $\mu(A) < 1$ and $\mu(B) < 1$. Therefore, it is clear that $\mu(A \cap B) \geq \top_d(\mu(A), \mu(B)) = 0$.

Another case is when $A = \Omega$. Then, $A \cap B = B$. Therefore $\mu(A \cap B) = \mu(B) = \top_d(\mu(A), \mu(B)) = \top_d(1, \mu(B)) = \mu(B)$.

Finally, we have the case in which $B = \Omega$, that is analogous to the previous one and we can also prove $\mu(A \cap B) = \top_d(\mu(A), \mu(B))$.

Therefore, we have that

$$\mu(C) \geq \mu(A \cap B) \geq \top_d(\mu(A), \mu(B)),$$

and we can proceed as we did in Theorem 2 defining $F^\mu_{pq}(u)$, $F^\mu_{qr}(v)$, and $F^\mu_{pr}(x)$, and obtain the equation:

$$F^\mu_{pr}(x) \geq \top_d(F^\mu_{pq}(u), F^\mu_{qr}(v)).$$

From this equation we prove that

$$\begin{aligned} F_{pr}^{\mu}(x) &\geq \sup\{\top_d(F_{pq}^{\mu}(u), F_{qr}^{\mu}(v)) | u+v=x\} \\ &= \tau_{\top_d}(F_{pq}^{\mu}, F_{qr}^{\mu})(x), \end{aligned}$$

holds for $\tau_{\top_d}$. Therefore, the theorem is proven. □

We prove the following theorem that considers fuzzy measures μ_{A_0} introduced in Definition 6.

Theorem 5. *Let $(S, \mathcal{F})$ be an F-space. Let μ_{A_0} be a fuzzy measure defined on $(\Omega, \mathcal{A})$ for a non empty set $A_0 \in \mathcal{A}$. Then $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\min}$.*

Proof. The proof of this theorem also follows the proof of Theorem 2. So, we focus on the proof of condition 4. We also consider A, B, and C as above and that $\mu_{A_0}(C) \geq \mu_{A_0}(A \cap B)$ due to the monotonicity of μ.

As $\mu_{A_0}(A) = 1$ if and only if $A \supseteq A_0$, it is easy to prove that when A_1 and A_2 are such that $\mu_{A_0}(A_1 \cap A_2) = 1$ it means that $A_1 \cap A_2 \supseteq A_0$ and therefore both $A_1 \supseteq A_0$ and $A_2 \supseteq A_0$. Therefore,

$$\mu_{A_0}(A_1 \cap A_2) = 1 = \min(\mu_{A_0}(A_1), \mu_{A_0}(A_2)) = \min(1, 1) = 1.$$

Then, if $\mu_{A_0}(A_1 \cap A_2) = 0$ this means that $A_1 \cap A_2$ does not include A_0. Therefore, it is not possible that both A_1 and A_2 include A_0. At most one of them can include A_0. This means that either $\mu_{A_0}(A_1) = 0$ or $\mu_{A_0}(A_2) = 0$. Because of that

$$\mu_{A_0}(A_1 \cap A_2) = 0 = \min(\mu_{A_0}(A_1), \mu_{A_0}(A_2)) = 0.$$

In a way analogous to previous proofs, we obtain that

$$\begin{aligned} F_{pr}^{\mu}(x) &\geq \sup\{\min(F_{pq}^{\mu}(u), F_{qr}^{\mu}(v)) | u+v=x\} \\ &= \tau_{\min}(F_{pq}^{\mu}, F_{qr}^{\mu})(x), \end{aligned}$$

holds for $\tau_{\min}$. Therefore, the theorem is proven. □

These results show the relationship between the type of probabilistic metric space and the type of t-norm. The stronger the t-norm, the less constraints we have on the measure we can use to build the probabilistic metric space. This is the case of using the drastic t-norm where the measure

	ω_1	ω_2	ω_3
p	0	0	0
q	0	0	1
r	0	1	1
s	1	1	1

Table 1: Functions p, q, r and s from $\Omega := \{\omega_1, \omega_2, \omega_3\}$ into M for Example 3.

is only constrained to be a 1^- measure (see Theorem 4). On the contrary, the less strict the t-norm, the more constraints we have on the measure. This is the case of the minimum where the fuzzy measure is of type μ_{A_0} (see Theorem 5).

The following proposition is a consequence of the fact that given a fuzzy measure μ we can build another one μ' using Expression 2 and the Choquet integral.

Proposition 4. *Let $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$ and target (M, d). Let $f : \Omega \to \mathbb{R}^+$ be a function such that $(C) \int f d\mu = 1$. Then, $(S, \mathcal{F})$ with $\mathcal{F}(p, q) = F_{pq}$ defined by*

$$F_{pq}^{\mu'}(x) = \mu'(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\})$$

is an F-space with base $(\Omega, \mathcal{A}, \mu')$ and target (M, d) for μ' defined using Expression 2.

Proof. To prove this proposition is trivial as $(S, \mathcal{F})$ is an F-space, so (i) in Definition 15 holds, and the construction of $F_{pq}^{\mu'}$ follows the one of Expression 4 in Definition 15 but with μ' instead of μ. □

5. An example

We begin with a toy example corresponding to the application of previous results. We give a larger example in the context of machine learning and statistics in the next section.

Example 3. *Let $\Omega := \{\omega_1, \omega_2, \omega_3\}$, and $\mathcal{A} := 2^\Omega$. Then $(\Omega, \mathcal{A})$ is a measurable space.*

Let $M := [0, \infty)$ and let $d(a, b) := |a - b|$ for $a, b \in M$. Then, (M, d) is a metric space.

Define the functions $p, q, r, s : \Omega \to M$ as given by Table 1. Then, let $S = \{p, q, r, s\}$ and define for $p_1, p_2 \in S$ the following functions: $H_{p_1p_2}(x) := \{\omega| \ \ |p_1(\omega) - p_2(\omega)| < x\}$ for $0 \leq x$. Then we have the sets H_{pq} as given in Table 2.

	$x = 0$	$0 < x \leq 1$	$x > 1$
H_{pq}	$\emptyset$	$\{\omega_1, \omega_2\}$	Ω
H_{pr}	$\emptyset$	$\{\omega_1\}$	Ω
H_{ps}	$\emptyset$	$\emptyset$	Ω
H_{qr}	$\emptyset$	$\{\omega_1, \omega_3\}$	Ω
H_{qs}	$\emptyset$	$\{\omega_3\}$	Ω
H_{rs}	$\emptyset$	$\{\omega_2, \omega_3\}$	Ω

Table 2: Functions H_{p_1,p_2} for $p_1, p_2 \in S$ for Example 3.

Now, let us define a Probability P_1 on $(\Omega, \mathcal{A})$ as follows: $P(\{\omega_1\}) = P(\{\omega_2\}) = P(\{\omega_3\}) = 1/3$. Then, using Equation 3, we construct the functions $\mathcal{F}_1$ as described in Table 3.

	$x = 0$	$0 < x \leq 1$	$x > 1$
F_{1pq}	0	2/3	1
F_{1pr}	0	1/3	1
F_{1ps}	0	0	1
F_{1qr}	0	2/3	1
F_{1qs}	0	1/3	1
F_{1rs}	0	2/3	1

Table 3: Functions F_{1p_1,p_2} for $p_1, p_2 \in S$ for Example 3.

Let $\tau = W$. Then, $(S, \mathcal{F}, \tau)$ is a proper Menger space as we have shown in Theorem 1. In fact, we can see that $\tau(F_{1pq}, F_{1qr}) = 2/3+2/3-1 = 1/3 = F_{pr}$ if $0 < x \leq 1$. We can check the other inequalities in a similar way.

Let us now define a Probability P_2 on $(\Omega, \mathcal{A})$ by $P(\{\omega_1\}) = P(\{\omega_2\}) = 0, P(\{\omega_3\}) = 1$. Then we have functions $\mathcal{F}_2$ as shown in Table 4.

That is, $F_{2pq} = F_{2pr} = F_{2ps} = \epsilon_1$, $F_{2qr} = F_{2qs} = F_{2rs} = \epsilon_0$

Let us now consider a distortion function $\varphi(x) = x^\alpha$ with $\alpha \geq 1$. We can use it to define a fuzzy measure $\varphi \circ P_1$. This permits us to have an F-space

	$x=0$	$0<x\leq 1$	$x>1$
F_{2pq}	0	0	1
F_{2pr}	0	0	1
F_{2ps}	0	0	1
F_{2qr}	0	1	1
F_{2qs}	0	1	1
F_{2rs}	0	1	1

Table 4: Functions F_{2p_1,p_2} for $p_1, p_2 \in S$ for Example 3.

$(S, \mathcal{F}_3)$ with base $(\Omega, \mathcal{A}, \varphi \circ P_1)$. Functions $\mathcal{F}_3$ are described in Table 5.

	$x=0$	$0<x\leq 1$	$x>1$
F_{3pq}	0	$(2/3)^\alpha$	1
F_{3pr}	0	$(1/3)^\alpha$	1
F_{3ps}	0	0	1
F_{3qr}	0	$(2/3)^\alpha$	1
F_{3qs}	0	$(1/3)^\alpha$	1
F_{3rs}	0	$(2/3)^\alpha$	1

Table 5: Functions F_{3p_1,p_2} for $p_1, p_2 \in S$ for Example 3.

Let $\alpha \to \infty$, we have the fuzzy measure $\mu_1 = \varphi \circ P_1$ such that $\mu_1(A) = 0$ if $A \neq \Omega$, and $\mu_1(A) = 1$ if $A = \Omega$. Then we have the F-space $(S, \mathcal{F}_4)$ with base $(\Omega, \mathcal{A}, \mu_1)$. It is easy to see that in this case $F_{4pq} = F_{4pr} = F_{4ps} = F_{4qr} = F_{4qs} = F_{4rs} = \epsilon_1$.

Now, let us regard the set of functions of the set S, that is p, q, r, s as the points in R^3: $P(p), Q(q), R(r), S(s)$. E.g., $P(p) = (0, 0, 0)$ and $Q(s) = (1, 1, 1)$ following the values in Table 1.

Figure 3 represents these points in R^3.

Then, using functions $\mathcal{F}_1$ and $\mathcal{F}_2$, we can classify each pair of points P, Q, R, S.

For example, using $\mathcal{F}_1$ we see that the following pairs $\{P, R\}$ and $\{Q, S\}$ have the same distance. This is represented in Figure 4 (left).

Then, the pairs $\{P, Q\}$, $\{Q, R\}$, and $\{R, S\}$ have also the same distance. This can be observed in Figure 4 (right).

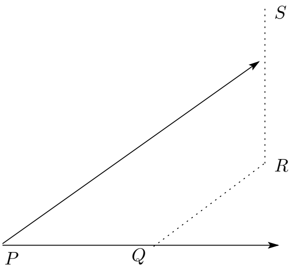


Figure 3: Representation of points $P(p), Q(q), R(r), S(s)$ as in R^3.

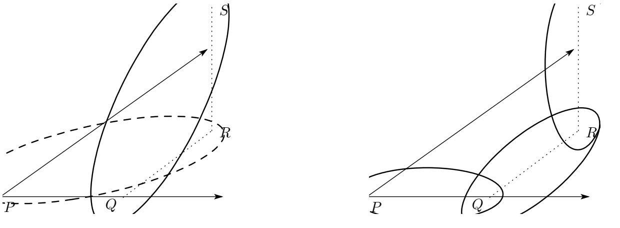


Figure 4: Representation of points $P(p), Q(q), R(r), S(s)$ as in R^3 with pairs grouped according to $\mathcal{F}_1$. On the left the pairs whose distance is 0 for $x = 0$, 1 for $x > 1$ and 1/3 otherwise. On the right the pairs whose distance is 0 for $x = 0$, 1 for $x > 1$ and 2/3 otherwise.

In contrast, the pair $\{P, S\}$ is independent in the view point of P_1. It is the only pair with this particular distance.

When we use $\mathcal{F}_2$ to study the similarity of these objects, we have that on the one hand the pairs $\{P, Q\}$, $\{P, R\}$, and $\{P, S\}$ have same distance. On the other hand, the pairs $\{Q, R\}$, $\{Q, S\}$, $\{R, S\}$ have also the same distance. Therefore, the sets of equivalent functions are different.

6. An application

In this section we describe the application of the results presented in this paper in the context of machine learning and statistics. The motivation is the comparison of algorithms and the models and statistics they build. Our example is small and we will only compare three aggregation functions. This case permits to illustrate some of the theoretical results we have obtained in the previous sections.

The motivation comes from the need to compare models and algorithms taking into account the databases they are applied to. We have shown [25] that probabilistic metric spaces can be useful to define metrics for machine learning. It is known that the application of a deterministic machine learning algorithm to a database produces a machine learning model. Then, different databases would result into different models. Nevertheless, not all models are different. There are models that are more occurrent than others (they appear more frequently). Another aspect to take into account is the following. When there are database changes (e.g., adding or removing people from a database), we may need to retrain the model. This retraining produces the same model or a different one, which can be similar or very dissimilar. In a previous work [29] we considered the construction of probabilistic metric spaces based on Markov chains and transition matrices. That is, we considered that database changes in time can be modeled using a Markov process. The transition matrix was used to build the space. Here we consider a different approach, taking into account a fuzzy measure on the database space. This measure represents our information on a set of databases. For example, the coverage of these databases. We give more details on the interpreation of measures in Subsection 6.3. Then, we need a metric on the space of models. As we have said, for simplicity we only consider three aggregation functions, and, therefore, their output is a positive number. Because of that, we use the absolute value. With respect to the space of databases, we consider a population and then the space of databases are samples of the information of this population. This approach follows our previous work, and is rooted on machine learning where multiple runs are usually considered for a given algorithm. All these elements will permit us to construct a probabilistic metric space, and E-spaces and F-spaces.

Let us consider the formalization of this problem. Suppose the space $(\Omega, \mathcal{A})$ corresponds to the space of possible databases and $\mathcal{A} := 2^{\Omega}$. Then, P and μ are probabilities and fuzzy measures, respectively, on this space.

Our different algorithms are the computation of an aggregated value from the database. I.e., we consider below three types of mean in our examples. Then, the target space (M, d) corresponds to the space of models defined by $M = [0, \infty)$ and the distance between any two models is defined by

index	groups	age	healthy_eating	active_lifestyle	salary
388	O	24	6	5	2878
228	O	54	7	6	3228
112	A	45	6	8	2182
745	A	33	3	6	902
780	O	35	7	3	3923
656	A	40	6	0	4038
932	O	25	6	6	2646
526	AB	40	10	6	4972
626	A	59	3	3	1598
663	B	55	4	4	1948

Table 6: Employee Dataset sample used to build the database space $(\Omega, \mathcal{A})$.

$d(a, b) := |a - b|$ for $a, b \in M$.

Let S be a set of different machine learning algorithms. Each one maps the database space to the model space, thus for any $p_i \in S$, $p_i(DB)$ is the model built from database DB using algorithm p_i. For simplicity, S is composed of only three different functions:

- Arithmetic mean as (p).
- Harmonic mean as (q).
- Geometric mean as (r).

Finally, let $\mathcal{F}$ be a mapping from $S \times S$ into Δ^+.

Following this setup, we next illustrate some toy examples to construct the space $(S, \mathcal{F})$ with respect to different measures and t-norms. The database space for all experiments is constructed from Table 6.

We will consider in the next sections two different cases. First, an additive measure. That is, a probability. This will permit us to obtain results that correspond to an E-space. Then, we consider fuzzy measures, to obtain an F-space. These measures permit to represent a characteristic or property of the space of databases. In our example we will represent interactions between databases which are all either positive or negative (as complementarity or redundancy). We finish the section discussing the meaning of measures in this setting.

6.1. *Case 1: Additive Measure*

In this experiment, the database space is a probability space $(\Omega, \mathcal{A}, P)$ where P is an additive measure defined by $P(DB) = 1/1023$ for any database (DB) in the space.

The model space is built as we described above. Figure 5 demonstrates the histograms of the distances among the three functions.

Now let us define the function $H_{p_1p_2}$ as follows: for any $p_1, p_2 \in S$ and for $x \geq 0$,

$$H_{p_1p_2}(x) = \{DB| \quad |p_1(DB) - p_2(DB)| < x\}.$$

Let us define the function $l(H)$ as the number of elements in H (i.e., cardinality of the set). This function $l(H)$ is given in Table 7. We will define the probability P using $l(H)$ and dividing by the total number of non empty databases (1023 as we have all non empty subsets of 10 records). This is detailed below.

	$x = 0$	$0 < x \leq 200$	$x < 1500$
$l(H_{pq})$	0	152	1023
$l(H_{pr})$	0	441	1023
$l(H_{qr})$	0	436	1023

Table 7: Functions $l(H_{p_1p_2})(x)$ for $p_1, p_2 \in S$.

Now, using Equation 3 and since P is additive, $F_{pq}(x) = \frac{l(H_{pq}(x))}{1023}$. Functions $\mathcal{F}$ are given in Table 8. Since $\mathcal{F}$ satisfies properties 1, 2, and 3 in Definition 12, then $(S, \mathcal{F})$ is a canonical E-space and hence it is a Menger space under W.

	$x = 0$	$0 < x \leq 200$	$x < 1500$
F_{pq}	0	0.148583	1
F_{pr}	0	0.431085	1
F_{qr}	0	0.426197	1

Table 8: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on additive measure P

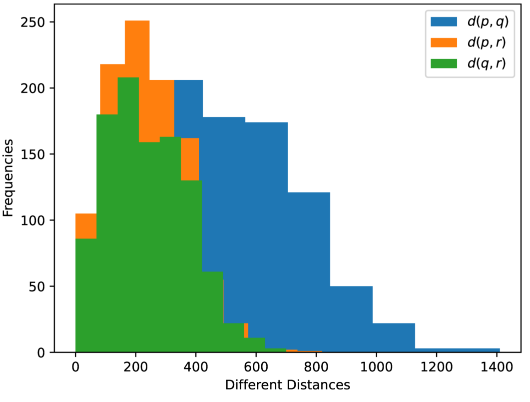


Figure 5: Histogram of the three different distances.

6.2. Case 2: Fuzzy Measures

In the following experiments, we consider the use of fuzzy measures. We will show that the result in our examples are in accordance with what we have proven in previous sections. Because of that, we use two different families of fuzzy measures. We start with a Sugeno λ-measure (μ) followed by the fuzzy measure μ_{A_0}. We use Sugeno λ-measures because they are simple to define on the large space of databases we have, and we can easily define submodular and supermodular measures.

6.2.1. Sugeno λ-measure

In this experiment, we use a Sugeno λ-measure from Definition 4 to build the database space $(\Omega, \mathcal{A}, \mu)$. The construction of the databases space, the Model Space M, and the set S is similar to the previous example. For simplicity, all the singletons measures are assumed to be equal. I.e. $v(x_i) = k$, for all $x_i \in \Omega$. Therefore, solving Equation 1 for k yields:

$$k = \frac{1}{\lambda}(\exp(\frac{1}{n}\ln(1+\lambda) - 1) \tag{7}$$

Suppose $\lambda = 0.5$, we use Equation 7 to compute the values of the measure for the singletons. The functions $\mathcal{F}$ are derived for Table 7, and the results are given in Table 9. Since the measure here is supermodular, the results are aligned with Theorem 2, in which the space $(S, \mathcal{F}, \tau)$ is a probabilistic pseudometric space under bounded difference τ_W. Figure 6 demonstrates the correctness of condition 4 for Definition 12.

The same steps are now repeated but with a submodular measures, i.e., when $\lambda < 0$ and also tested with respect to the t-norm τ_W. While some measures resulted in probabilistic pseudometric spaces, some measures yield probabilistic semimetric spaces (i.e., triangle inequality does not hold). An example of the latter case is when $\lambda = -0.9$. The results for this case are given in Table 10 and Figure 7.

	$x = 0$	$0 < x \leq 200$	$x < 1500$
F_{pq}	0	0.124194	1
F_{pr}	0	0.381992	1
F_{qr}	0	0.377276	1

Table 9: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on Sugeno λ-measure ($\lambda = 0.5$)

	$x = 0$	$0 < x \leq 200$	$x < 1500$
F_{pq}	0	0.321933	1
F_{pr}	0	0.699324	1
F_{qr}	0	0.694664	1

Table 10: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on Sugeno λ-measure ($\lambda = -0.9$)

6.2.2. Fuzzy measure μ_{A_0}

Our last experiment is based on the fuzzy measure μ_{A_0}, which is introduced in Definition 6. Suppose $A_0 = H_{pq}(x)$ when $0 < x \leq 200$ (see Table 7), then we have the following relation:

$$H_{pq} \subseteq H_{qr} \subseteq H_{pr}.$$

The functions F are given in Table 11. It is clear that the space $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\min}$. Therefore, this result is aligned with Theorem 5.

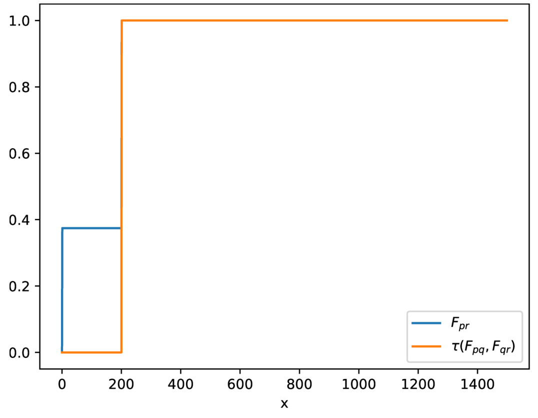


Figure 6: Testing the triangle inequality with t-norm τ_W for $\lambda = 0.5$ in Case 2. We observe that the inequality holds.

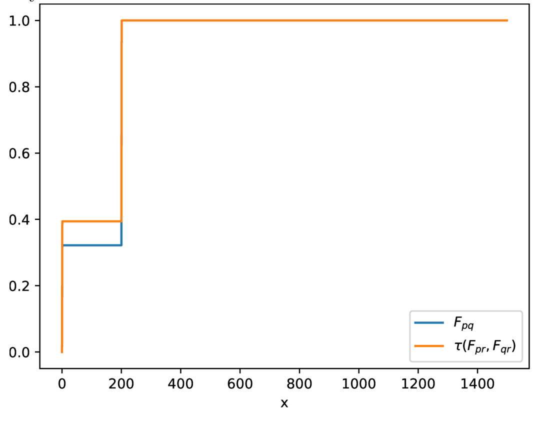


Figure 7: Testing the triangle inequality with t-norm τ_W for $\lambda = -0.9$ in Case 2. We observe that the inequality does not hold.

6.3. The interpreation of the fuzzy measures

Measures (additive and non-additive ones) are defined in this section on the space of databases. They represent a characteristic or property of a set of them. Here, we have considered either the same relevance with the

	$x = 0$	$0 < x \leq 200$	$x < 1500$
F_{pq}	0	1	1
F_{pr}	0	1	1
F_{qr}	0	1	1

Table 11: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based fuzzy measure μ_{A_0}

probability $P(DB) = 1/1023$ or some equal interaction with the Sugeno λ-measure. Fuzzy measures permit to consider more complex interactions. We can consider, for example, the coverage of a database or of a set of databases. Then, naturally, we would have that the larger the set of databases, the larger the coverage. The definition of a new measure using the Choquet integral using Expression 2 (or, similarly, using the Sugeno integral) would permit to define measures that combine a measure representing the coverage with a characteristic of the database itself represented by the function f in the expression.

7. Conclusions

In this paper we have introduced F-spaces, a type of probabilistic metric spaces based on fuzzy measures. They can be seen as an extension of E-spaces, which were based on additive (probability) measures. In contrast, F-spaces use fuzzy measures which permit to take into account that objects of the space are not necessarily independent. We have provided several theoretical results for this type of spaces.

We have illustrated the application of our results with simple examples inspired on machine learning. More precisely, we consider a space of databases, and three different aggregation functions. We build F-spaces considering different types of fuzzy measures on the space of databases. The examples illustrate the theorems and propositions proven in Section 3. In particular, we have shown that supermodular fuzzy measures lead to probabilistic pseudometric spaces that satisfy the appropriate conditions of Definition 12 while for non-supermodular fuzzy measures this is not the case (i.e., triangle inequality does not hold). So the experimental results are consistent with our mathematical results.

As future work we consider studying additional properties of these probabilistic metric spaces, as well as to consider their application in real-size databases with more complex and realistic machine learning algorithms. In relation to the theory, we will consider the requirement of associativity for t-norms and its role in defining probabilistic metric spaces, as suggested by a reviewer. This requires the consideration of metric spaces on the space of models, and defining fuzzy measures for the space of databases. We plan to study fuzzy measures that are appropriate to represent background information on the space of models.

In general, the size of the databases may not be a constraint for the application of this results. Nevertheless, a large size of the space of databases can be. Then, we may need to consider approximations of the distance functions. This is another direction for future work.

Acknowledgements

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

References

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

[2] Bardozzo, F., de la Osa, B., Horanská, L., Fumanal-Idocin, J., delli Priscoli, M., Troiano, L., Tagliaferri, R., Fernández, J., Bustince, H. (2021) Sugeno integral generalization applied to improve adaptive image binarization, Inf. Fusion 68: 37-45.

[3] Beliakov, G., Cao, T., Mak-Hau, V. (2022) Aggregation of interacting criteria in land combat vehicle selection by using fuzzy measures, IEEE Trans. on Fuzzy Systems, in press.

[4] Beliakov, G., Gagolewski, M., James, S. (2022) Reduction of variables and constraints in fitting antibuoyant fuzzy measures to data using linear programming, Fuzzy Sets and Systems, in press.

[5] Bustince, H., Mesiar, R., Fernández, J., Galar, M., Paternain, D., Altalhi, A. H., Pereira Dimuro, G., Bedregal, B. R. C., Takác, Z. (2021) d-Choquet integrals: Choquet integrals based on dissimilarities, Fuzzy Sets Syst. 414 1-27.

[6] Candeloro, D., Mesiar, R., Sambucini, A. R. (2019) A special class of fuzzy measures: Choquet integral and applications, Fuzzy Sets and Systems 355 83-99.

[7] Denneberg, D. (1994) Non Additive Measure and Integral, Kluwer Academic Publishers.

[8] Hadžić, O., Pap, E. (2001) Fixed point theory in probabilistic metric spaces, Springer.

[9] Jin, L., Mesiar, R., Yager, R. R. (2022) Some fuzzy measures constructing paradigm and methods from given measure spaces and evaluation information, IEEE Trans. on Fuzzy Systems, in press.

[10] Klement, E. P., Mesiar, R., Pap, E. (2000) Triangular Norms, Kluwer Academic Publisher.

[11] Leszczyski, K., Penczek, P., Grochulski, W. (1985) Sugeno's fuzzy measure and fuzzy clustering, Fuzzy Sets and Systems 15 147-158.

[12] Narukawa, Y., Torra, V. (2005) Fuzzy measure and probability distributions: distorted probabilities, IEEE Trans. on Fuzzy Systems 13:5 617-629.

[13] Pap, E. (1995) Null-Additive set functions, Kluwer Academic Publishers, Dordorecht.

[14] Pap, E. (ed.) (2002) Handbook of Measure Theory, Vols. I and II, North-Holland.

[15] Pereira Dimuro, G., Fernández, J., Bedregal, B. R. C., Mesiar, R., Sanz, J. A., Lucca, G., Bustince, H. (2020) The state-of-art of the generalizations of the Choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions, Inf. Fusion 57: 27-43

[16] Pereira Dimuro, G., Lucca, G., Bedregal, B. R. C., Mesiar, R., Sanz, J. A., Lin, C.-T., Bustince, H. (2020) Generalized CF1F2-integrals: From Choquet-like aggregation to ordered directionally monotone functions, Fuzzy Sets Syst. 378 44-67.

[17] Schweizer, B., Sklar, A. (1960) Statistical metric spaces, Pacific J. Math. 10 313-334.

[18] Schweizer, B., Sklar, A. (1983) Probabilistic Metric Spaces, Elsevier-North-Holland.

[19] Searcóid, M. O. (2007) Metric spaces, Springer.

[20] Sherwood, H. (1969) On E-spaces and their relation to other classes of probabilistic metric spaces, J. London Math. Soc. 44 441-448.

[21] Stevens, S. S. (1968) Metrically generated probabilistic metric spaces, Fund. Math. 61 259-269.

[22] Tahani, H., Keller, J. M. (1990) Information fusion in computer vision using the fuzzy integral, IEEE Trans. on Systems, Man and Cybernetics 20:3 733-741.

[23] Torra, V. (2000) On a family of fuzzy measures for data fusion with reduced complexity, Proc. 3rd Int. Conf on Information Fusion, Paris, France.

[24] Torra, V., Guillen, M., Santolino, M. (2018) Continuous m-dimensional distorted probabilities, Information Fusion 44 97-102.

[25] Torra, V., Navarro-Arribas, G. (2018) Probabilistic metric spaces for privacy by design machine learning algorithms: modeling database changes, Proc. DPM 2018, LNCS 11025 422-430.

[26] Torra, V., Narukawa, Y. (2007) Modeling decisions: information fusion and aggregation operators, Springer.

[27] Torra, V., Narukawa, Y., Sugeno, M. (eds.) (2013) Non-additive measures: theory and applications, Springer.

[28] Torra, V., Stokes, K., Narukawa, Y. (2012) An Extension of Fuzzy Measures to Multisets and Its Relation to Distorted Probabilities, IEEE Trans. on Fuzzy Systems 20:6 1032-1045.

[29] Torra, V., Taha, M., Navarro-Arribas, G. (2021) The space of models in machine learning: using Markov chains to model transitions. Prog. Artif. Intell. 10(3): 321-332.

[30] Zhang, X., Wang, J., Zhan, J., Dai, J. (2022) Fuzzy Measures and Choquet Integrals based on Fuzzy Covering Rough Sets, in press.

Paper III

Measuring the distance between machine learning models using F-space

Mariam Taha and Vicenç Torra.

Proceedings of the 13th Conference of the European Society for Fuzzy Logic and Technology (EUS-FLAT 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)

Measuring the Distance Between Machine Learning Models Using F-Space

Mariam Taha(✉) and Vicenç Torra

Department of Computing Science, Umeå University, 901 87 Umeå, Sweden
{mariamt,vtorra}@cs.umu.se

Abstract. Probabilistic metric spaces are a natural generalization of metric spaces in which the function that computes the distance outputs a distribution on the real numbers rather than a single number. Such a function is called a distribution function. In this paper, we construct a distance for linear regression models using one type of probabilistic metric space called F-space. F-spaces use fuzzy measures to evaluate a set of elements under certain conditions. By using F-spaces to build a metric on machine learning models, we permit to represent more complex interactions of the databases that generate these models.

Keywords: Fuzzy Measures · Probabilistic Metric Space · Machine Learning

1 Introduction

Probabilistic metric spaces [1] are a natural generalization of metric spaces in which the function that computes the distance outputs a distribution on the real numbers rather than a single number. Such a function is called a distribution function. Constructing a probabilistic metric space (PMS) is not a straightforward process. There are different methods in the literature that aim to construct these spaces. Among them we find the E-spaces [2,3], where the probabilistic metric space is defined in terms of sets of functions and a probability space. These functions map from a probability space into a metric space. Another construction are the F-spaces, which generalize E-spaces by replacing the probability space with a measure space. Hence, the distribution functions are defined in terms of non-additive measures. We have introduced these F-spaces in a previous work [4].

In this paper, our interest lies in measuring the distance between machine learning models taking into account the set of databases that generate these models. We call such sets generators. The distance between models is defined in terms of distribution functions, and the probabilistic metric space is constructed in terms of a F-space. Since we consider models that can be defined in terms

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

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
S. Massanet et al. (Eds.): EUSFLAT 2023/AGOP 2023, LNCS 14069, pp. 307–319, 2023.
https://doi.org/10.1007/978-3-031-39965-7_26

of their generators, the distribution functions are based on these generators and their distances. By using fuzzy measures and F-spaces to construct models distances, it would be possible to model the interactions of the databases in the spaces. Such interactions can be interpreted in terms of coverage of a set of databases or any other properties on the databases. The study of probabilistic metric spaces and their relevance to the problem of model selection were previously studied in [12] where it was linked to data privacy, and in [5] where the authors developed these spaces taking into account transitions that occur among the database and modeled it using Markov chains and transition matrices.

This paper is structured as follows. In Sect. 2 we introduce the definitions that are needed later in the paper. In particular, we review some concepts related to fuzzy measures and their properties. In Sect. 3, we introduce probabilistic metric spaces and F-spaces along with some results and toy examples. In Sect. 4 we illustrate our results. Section 5 concludes the paper with some conclusions and research directions.

2 Fuzzy Measures

Fuzzy measures were first introduced by Sugeno [6]. They are also called capacities, nonadditive measures, and monotone measures. Fuzzy measures are considered as a generalization of classical measures [7–10].

Definition 1. *Let $(\Omega, \mathcal{A})$ be a measurable space. A set function μ defined on $\mathcal{A}$ is called a non-additive measure if an only if*

- $0 \leq \mu(A) \leq \infty$ *for any* $A \in \mathcal{A}$*;*
- $\mu(\emptyset) = 0$*;*
- *If* $A_1 \subseteq A_2 \subseteq \mathcal{A}$ *then*
$$\mu(A_1) \leq \mu(A_2)$$

If in addition $\mu(A) = 1$, then the fuzzy measure is said to be a normalized space. We consider finite sets Ω, and for simplicity we assume $\mathcal{A} = 2^{\Omega}$.

Definition 2. *Let μ be a non-additive measure on the measurable space $(X, \mathcal{A})$. Then,*

- *μ is additive if $\mu(A \cup B) = \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is superadditive if $\mu(A \cup B) \geq \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is subadditive if $\mu(A \cup B) \leq \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is submodular if $\mu(A) + \mu(B) \geq \mu(A \cup B) + \mu(A \cap B)$;*
- *μ is supermodular if $\mu(A) + \mu(B) \leq \mu(A \cup B) + \mu(A \cap B)$;*
- *μ is symmetric if for finite X, when $|A| = |B|$, then $\mu(A) = \mu(B)$.*

A supermodular measure implies superadditivy, while a submodular measure implies subadditivity. When additive fuzzy measures are normalized, they are probability measures. In this paper we will use two families of fuzzy measures in our experiments, Sugeno λ-measures and the non-additive measure μ_{A_0}. Their definitions are as follows.

Definition 3. *Let Ω be a finite set and let $\lambda > -1$. A Sugeno λ-measure is a function $\mu : 2^{\Omega} \to [0, 1]$ such that*

- $\mu(\Omega) = 1$
- *if* $A, B \subseteq \mathbf{X}$ *with* $A \cap B = \emptyset$ *then* $\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B)$

For Sugeno λ-measures, as a convention, the measure of the singletons $\omega_i \in \Omega$ is called a density and it is noted by $v(\omega_i)$. In this case, as the measure is normalized when $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, λ should satisfy the following:

$$\lambda + 1 = \prod_{i=1}^{n} 1 + \lambda v(\omega_i). \tag{1}$$

Once the densities are known, the above polynomial can be used to uniquely determine the value of λ. Then given the densities and λ, the fuzzy measure $\mu(A)$ is defined as:

$$\mu(A) = \begin{cases} v(x_i), & A = x_i \\ \frac{1}{\lambda} \prod_{x_i \in A} (1 + \lambda v(x_i)) - 1), & |A| \neq 1 \quad \& \quad \lambda \neq 0 \\ \sum_{x_i \in A} v(x_i), & \lambda = 0 \quad \& \quad |A| \neq 1 \end{cases}$$

For Sugeno λ-measures, when $\lambda > 0, \mu$ is supermodular. Whereas, when $\lambda > 0, \mu$ is submodular.

Definition 4. *Let A_0 be a subset of Ω, then the set function defined by $\mu_{A_0}(A) = 1$ if and only if $A_0 \subseteq A$, is a non-additive measure.*

3 Probabilistic Metric Space

In this section, we review some concepts related to probabilistic metric spaces and their properties. Following this, we introduce E-space and F-spaces.

Definition 5. *Let $d : S \times S \to \mathbb{R}^+$, then d is called a metric on S if the following properties hold for $a, b, c \in S$:*

- $d(a, b) \geq 0$ *with equality if and only if* $a = b$ *(positive property),*
- $d(a, b) = d(b, a)$ *(symmetry property), and*
- $d(a, b) \leq d(a, c) + d(c, b)$ *(triangle inequality property).*

Definition 6. [11] *The pair (S, d) is called a metric space when d is a metric on the set S. Where $d : S \times S \to \mathbb{R}^+$ plays the role of distance on the set S. Here, we understand $\mathbb{R}^+ = [0, \infty)$ and $\overline{\mathbb{R}^+} = [0, \infty]$.*

When the distance does not satisfy the symmetry condition, we say that (S, d) is a quasimetric space; and when the distance does not satisfy the triangle inequality, we say that (S, d) is a semimetric space. Probabilistic metric spaces are a generalization of metric spaces in which the distance function is replaced by a distribution distance function.

Definition 7. [1] *A distance distribution function F is a nondecreasing function defined on $\overline{\mathbb{R}^+}$ that satisfies (i) $F(0) = 0$; (ii) $F(\infty) = 1$, and (iii) that is left continuous on $(0, \infty)$.*

Therefore, $F(x)$ can be interpreted as the probability that the distance is less than or equal to x. The set of all distance distribution functions is denoted as Δ^+, while the distance distribution function that represents a classical distance is denoted by ϵ_a and is defined as below.

Definition 8. [1] *For any a in $\mathbb{R}^+$, we define $\epsilon_a \in \Delta^+$ by*

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

Next, we introduce the concepts of t-norms and triangle functions in order to construct a probabilistic metric space.

Definition 9. [13] *A function $\top : [0,1] \times [0,1] \to [0,1]$ is a t-norm if and only if it satisfies the following properties:*

- $\top(x, y) = \top(y, x)$ *(symmetry or commutativity)*
- $\top(\top(x, y), z) = \top(x, \top(y, z))$ *(associativity)*
- $\top(x, y) \leq \top(x', y')$ *if* $x \leq x'$ *and* $y \leq y'$ *(monotonicity)*
- $\top(x, 1) = x$ *for all* x *(neutral element 1)*

One example of t-norm is the minimum function $T(x, y) = min(x, y)$ which is denoted by $\wedge$, i.e $T(x, y) = \wedge(x, y)$. Another t-norm is the bounded difference $W(x, y)$ defined as $T(x, y) = max(0, x + y - 1)$.

Definition 10. [1] *A Triangle function T is a binary operation on Δ^+ that for any $F, G, H, K \in \Delta^+$, it satisfies the following:*

- $T(F, \epsilon_0) = F$
- $T(F, G) = T(G, F)$
- $T(F, G) \leq T(H, K)$ *whenever* $F \leq H, G \leq K$
- $T(T(F, G), H) = T(F, T(G, H))$

For a t-norm $\top$, we have that the function $\tau_\top(F, G)(x) = \top(F(x), G(x))$ is a triangle function. Next, we introduce probabilistic metric spaces along with their properties.

Definition 11. [1] *Let $(S, \mathcal{F}, \tau)$ be a triple where S is a nonempty set, $\mathcal{F}$ is a function from $S \times S$ into Δ^+, and τ is a triangle function; then $(S, \mathcal{F}, \tau)$ is a probabilistic metric space (PM space) if the following conditions are satisfied for all p, q, and r in S:*

- $\mathcal{F}(p, p) = \epsilon_0$
- $\mathcal{F}(p, q) \neq \epsilon_0$ *if* $p \neq q$
- $\mathcal{F}(p, q) = \mathcal{F}(q, p)$
- $\mathcal{F}(p, r) \geq \tau(\mathcal{F}(p, q), \mathcal{F}(q, r))$.

For simplicity we will use F_{pq} instead of $\mathcal{F}(p, q)$ and denote the value of the latter at x as $F_{pq}(x)$. Special names are given when some of the above conditions fail. A probabilistic metric space that doesn't satisfy the second condition is called a probabilistic pseudometric space. If the space doesn't satisfy triangle inequality it is called a probabilistic semimetric space, while it is a probabilistic quasi metric space if the symmetry property is invalid.

F-spaces are one family of probabilistic metric spaces. They permit to compute the distance between functions that map from a measurable space to a metric space, where the distance distribution function is defined in terms of measuring those elements that are at most at distance x.

Definition 12. [4] *Let $(\Omega, \mathcal{A})$ be a measurable space, and let μ be a non-additive measure on $(\Omega, \mathcal{A})$. 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 F-space with base $(\Omega, \mathcal{A}, \mu)$ and target (M, d) if*

- *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 $\mathcal{A}$.
- *For all p, q in S, $\mathcal{F}(p, q) = F_{pq}$ with*
$$F_{pq}(x) = \mu(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}). \tag{2}$$

Definition 13. [2,3] *When the measure μ is additive, Definition 12 corresponds to E-spaces.*

Lemma 1. [2,3] *Let $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$ and target (M, d). Then if μ is additive, it is an E-space .*

If F satisfies the first three properties in Definition 11, then $(S, \mathcal{F})$ is a canonical F-space.

The following theorems have been proven in [4], which describe the type of probabilistic metric space when a specific fuzzy measure is used

Theorem 1. [4] *Let $(\Omega, \mathcal{A})$ be a measurable space, let μ be a non-additive measure on $(\Omega, \mathcal{A})$ and $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$.*

Then, if μ is a supermodular non-additive measure on $(\Omega, \mathcal{A})$, it follows that $(S, \mathcal{F})$ is a probabilistic pseudometric space under bounded difference τ_W.

Theorem 2. [4] *Let $(S, \mathcal{F})$ be an F-space. Let μ_{A_0} be a non-additive measure defined on $(\Omega, \mathcal{A})$ for a given set $A_0 \subseteq A$. Then $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\min}$.*

Next, we give an example to show how to construct an F-space.

Example 1. Let $\Omega := \{\omega_1, \omega_2, \omega_3\}$, and $\mathcal{A} := 2^\Omega$. Then $(\Omega, \mathcal{A})$ is a measurable space. Let $M := [0, \infty)$ and let $d(a, b) := |a - b|$ for $a, b \in M$. Then, (M, d) is a metric space. Define the functions $p, q, r : \Omega \to M$ as given in Table 1. Then, let $S = \{p, q, r\}$ and define for $p_1, p_2 \in S$ the following functions: $H_{p_1 p_2}(x) := \{\omega | \ |p_1(\omega) - p_2(\omega)| < x\}$ for $0 \leq x$. Then we have the sets H_{pq} as given in Table 2.

Now let us define a Sugeno λ-measure on $(\Omega, \mathcal{A})$ using Eq. 1 from Definition 3, and solve the equation for $\lambda = 0.4$. Under the assumption that all the densities are equal, we get $v(\omega_i) = 0.296722$. Therefore we construct the functions F as described in Table 3.

Table 1. Functions p, q, and r from $\Omega := \{\omega_1, \omega_2, \omega_3\}$ into M for Example 1.

	ω_1	ω_2	ω_3
p	1	0	0
q	0	0	1
r	0	1	1

Table 2. Functions H_{p_1,p_2} for $p_1, p_2 \in S$ for Example 1.

	$x = 0$	$0 < x \leq 1$	$x > 1$
H_{pq}	$\emptyset$	$\{\omega_2\}$	Ω
H_{pr}	$\emptyset$	$\emptyset$	Ω
H_{qr}	$\emptyset$	$\{\omega_1, \omega_3\}$	Ω

Table 3. Functions F_{p_1,p_2} for $p_1, p_2 \in S$ for Example 1.

	$x = 0$	$0 < x \leq 1$	$x > 1$
F_{pq}	0	0.296722	1
F_{pr}	0	0	1
F_{qr}	0	0.628662	1

If we choose the t-norm $\tau = W$, then $(S, \mathcal{F}, \tau)$ is a canonical space under W as we can see that all inequalities hold in Definition 11.

4 Metrics for Machine Learning Models

In Machine Learning, data are continuously generated, hence models need to be updated to reflect any new insights from the underlying data. However, it has been shown [15] that an adversary can get access to sensitive information by exploiting changes in the models themselves. One of the privacy models that overcome this issue is Integral Privacy [16]. Its goal is that model transformations caused by the training data should not leak any information on the training data. It recommends selecting a machine learning model which can be generated by sufficiently large and diverse datasets. Such models are called recurrent models and can be used to implement Integral privacy. Then, it may also happen that even when two models are different they may be generated from similar data.

The similarity of models in terms of the data that generated them is relevant for model selection. Between two such models we would prefer the one that is more privacy-preserving. From an integral privacy perspective that would be the one with more generators. Similarly, the same applies to the algorithms that produce the models. As we have discussed in [12], probabilistic metric spaces can be useful to define metrics for machine learning models, and thus helps in this process.

In a previous work [5] we considered a simpler approach to construct PMS, where we proposed the use of Markov chains, together with transition matrices to represent, respectively, sequences of changes in databases and the probability of changes of databases to define model similarities. In this paper, we are considering probabilities and fuzzy measures in the space of databases, in order to define metrics on the models.

In this paper, we show how this can be applied to real machine learning models. We consider simple machine learning models such as Linear Regression models. Our goal is to construct distances between such models taking into account the interaction of their generators. We run the experiments on the dataset *Salary_Data* which describes the salaries of employees and their years of experience. Figure 1 illustrates the scatter plot of this dataset [14]. In this experiment, the space $(\Omega, \mathcal{A})$ corresponds to the space of possible databases and $\mathcal{A} := 2^{\Omega}$. Then, P and μ are probabilities and fuzzy measures, respectively, on this space. In order to build the model space, we define the set S as the set of three different linear regression algorithms p, q, and r defined as follows:

– Linear Regression as (p)
– Huber Regression as (q)
– Ridge Regression as (r)

Therefore, given our approximated database space $(\Omega, \mathcal{A})$ together with the set S, we construct the target space (M, d) such that for any $p \in S$, $p(DB)$ is the trained model we obtain after applying one of the linear regression algorithm p on the database DB. Since the problem is a simple linear regression, each model can be characterized by its slope β and y-intercept α. We choose d here to be the Euclidean distance. Hence $(M, d) = (\mathbb{R}^2, d)$ where $d = \sqrt{(\alpha_1 - \alpha_2)^2 + (\beta_1 - \beta_2)^2}, (\alpha_i, \beta_i) \in \mathbb{R}^2$.

Finally, we identify $\mathcal{F}$ as a mapping from $S \times S$ into Δ^+.

Following this setup, we illustrate our experiments to construct the space $(S, \mathcal{F})$ with respect to different measures and t-norms. Since it is impossible to cover the full database space, we used the subsampling method to sample 1000 datasets in order to approximate the full space $(\Omega, \mathcal{A})$ [17]. We ran the

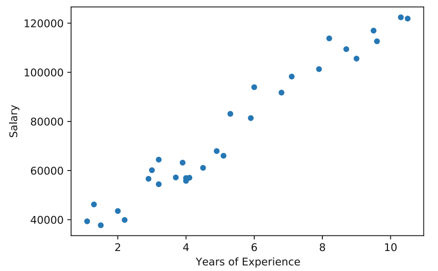


Fig. 1. Scatter Plot of *Salary_Data* dataset

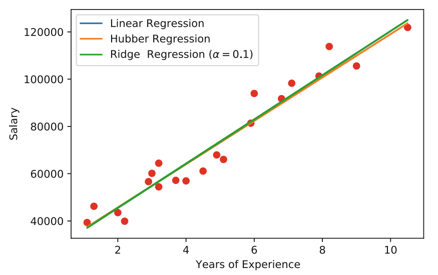


Fig. 2. Three regression models of *Salary_Data* dataset.

experiments under Python and Sklearn library, and the value of the penalty term α in Ridge regression algorithm is chosen to be 0.1 (see Fig. 2).

We will consider in the next sections two different cases. First, an additive measure. That is, a probability. This will permit us to obtain results that correspond to an E-space. Then, we consider fuzzy measures, to obtain a proper F-space. These measures permit us to represent a characteristic or property of the space of databases. In our example, we will represent interactions between databases which are all either positive or negative (as complementarity or redundancy). We finish the section discussing the meaning of measures in this setting.

4.1 Case 1: Additive Measure

In this experiment, the database space is a probability space $(\Omega, \mathcal{A}, P)$ where P is the additive measure defined by $P(DB) = 1/1000$ for any database (DB) in the space. The model space is built as we described above. Figure 3 shows the histograms of the distances among the three functions.

Now let us define the function $H_{p_1 p_2}$ as follows: for any $p_1, p_2 \in S$ and for $x \geq 0$,

$$H_{p_1,p_2}(x) = \{DB| \quad |p_1(DB) - p_2(DB)| < x\}.$$

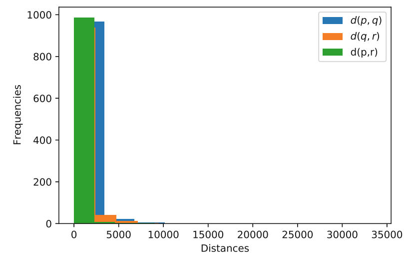


Fig. 3. Histogram of the three different distances.

Let us define the function $l(H)$ as the number of elements in H (i.e., the cardinality of the set). This function $l(H)$ is given in Table 4. We will define the probability P using $l(H)$ and dividing by the total number of the generated databases (i.e. 1000). This is detailed below.

Table 4. Functions $l(H_{p_1 p_2})(x)$ for $p_1, p_2 \in S$.

	$x = 0$	$0 < x \leq 500$	$x < 35000$
$l(H_{pq})$	0	475	1000
$l(H_{qr})$	0	484	1000
$l(H_{pr})$	0	968	1000

Now, using Eq. 2 and since P is additive, $F_{pq}(x) = \frac{l(H_{pq}(x))}{1000}$. Functions $\mathcal{F}$ are given in Table 5. Since $\mathcal{F}$ satisfies the first three properties in Definition 11, then $(S, \mathcal{F})$ is a canonical F-space.

Table 5. Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on additive measure P

	$x = 0$	$0 < x \leq 500$	$x < 35000$
F_{pq}	0	0.475	1
F_{qr}	0	0.484	1
F_{pr}	0	0.968	1

4.2 Case 2: Fuzzy Measures

In the following experiments, we use two different fuzzy measures. We start with a Sugeno λ-measure (μ) followed by the fuzzy measure μ_{A_0}. We use Sugeno λ-measures because they are easy to define and flexible enough to represent both subadditive and superadditive cases. That is, negative and positive interactions.

Table 6. Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on Sugeno λ-measure ($\lambda = 0.5$)

	$x = 0$	$0 < x \leq 500$	$x < 35000$
F_{pq}	0	0.424786	1
F_{qr}	0	0.43365	1
F_{pr}	0	0.961327	1

Table 7. Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on Sugeno λ-measure ($\lambda = -0.96$)

	$x = 0$	$0 < x \leq 500$	$x < 35000$
F_{pq}	0	0.815875	1
F_{qr}	0	0.822323	1
F_{pr}	0	0.995479	1

Sugeno $\boldsymbol{\lambda}$-Measure. In this experiment, we use Sugeno λ-measure from Definition 3 to build the database space $(\Omega, \mathcal{A}, \mu)$. The construction of the databases space, the Model Space M, and the set S is similar to the previous example.

For simplicity, all the singleton measures are assumed to be equal. I.e. $v(x_i) = k$, for all $x_i \in X$. Therefore, solving Eq. 1 for k yields:

$$k = \frac{1}{\lambda}(\exp(\frac{1}{n}\ln(1 + \lambda) - 1) \tag{3}$$

Suppose $\lambda = 0.5$, we use Eq. 3 to compute the values of the measure for the singletons. The functions F are derived for Table 4, and the results are given in Table 6. Since the measure here is supermodular, the results are aligned with Theorem 1, in which the space (S, F, τ) is a probabilistic pseudo metric space under bounded difference τ_W. Figure 4(a) demonstrates the correctness of the triangular inequality of Definition 11. That is, in our case $\mathcal{F}(p, q) \geq \tau(\mathcal{F}(p, r), \mathcal{F}(q, r))$ as the blue line is larger than the orange one. Observe that in most of the domain both distributions are the same.

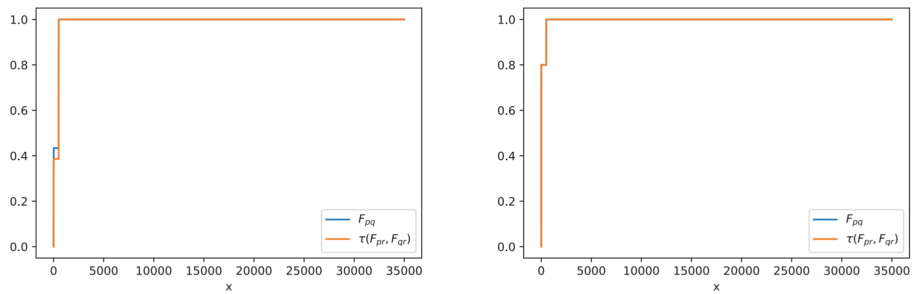
The same steps are now repeated but with submodular measures. I.e., when $\lambda < 0$ and also tested with respect to the t-norm τ_W. While some measures resulted in probabilistic pseudo metric spaces, some measures yield probabilistic semimetric spaces (i.e., triangle inequality does not hold). An example of the latter case is when $\lambda = -0.96$. The results for this case are given by Table 7 and Fig. 4(b). As we can see, in this case:

$$\begin{gathered}\mathcal{F}(p, q) \geq \tau(\mathcal{F}(p, r), \mathcal{F}(q, r)) \\ 0.815 \geq \tau(0.822323, 0.995479) \\ 0.815 \geq max(0, 0.822323 + 0.995479 - 1) = 0.817.\end{gathered}$$

Since 0.815 is not greater than 0.817, thus the inequality does not hold.

Fuzzy Measure $\boldsymbol{\mu_{A_0}}$. Our last experiment is based on the fuzzy measure μ_{A_0}, which is introduced in Definition 4. Let us define A_0 as the set $H_{pq}(x)$ for any

x in the range $0 < x \leq 500$, then in this example the other two sets H_{pr} and H_{qr} are incomparable with respect to inclusion of $H_{pq}(x)$ i.e.: $H_{pr} \nsubseteq H_{pq}$ and $H_{qr} \nsubseteq H_{pq}$. The functions F are given in Table 8. It is clear that the space $(S, \mathcal{F})$ is a probabilistic pseudo metric space under $\tau_{\min}$. Therefore, this result is aligned with Theorem 2.



(a) $\lambda = 0.5$ (b) $\lambda = -0.96$

Fig. 4. Testing the triangle inequality with t-norm τ_W for $\lambda = 0.5$ and $\lambda = -0.96$ in Case 2. We observe that the inequality holds when the $\lambda = 0.5$, and does not hold when $\lambda = -0.96$.

Table 8. Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based fuzzy measure μ_{A_0}

	$x = 0$	$0 < x \leq 500$	$x < 35000$
F_{pq}	0	1	1
F_{pr}	0	0	1
F_{qr}	0	0	1

4.3 The Interpretation of the Fuzzy Measures

Fuzzy measures are set functions that use the monotonicity property instead of additivity. Therefore, naturally, we would have that the larger the set of databases, the larger the coverage. In our experiment, all the fuzzy measures are defined on the space of the databases, where we considered either each database has the same relevance with the probability $P(DB) = 1/1000$ (i.e. the measure is additive), or all have the same interaction under the Sugeno λ-measure. Whereas, in the fuzzy measure μ_{A_0}, the value of the measure is computed with respect to the inclusion relationship of a reference set. Therefore the measure is either zero or one. Fuzzy measures can alternatively be defined using the Choquet integral or Sugeno integral, to define measures that represent the coverage with a characteristic of the database itself [18].

5 Conclusions

In this paper, we have constructed a probabilistic metric space to evaluate the distance between machine learning models built from databases. The probabilistic metric space is based on fuzzy measures and F-space in which the distance distribution function is computed based on functions that map from the database space to the model space. In our case, these functions were represented by different Linear Regression algorithms. Our experiment is based on different measures, both additives and non-additives. In future work, we consider studying additional properties of these probabilistic metric spaces, as well as considering their application in real-size databases. Also, since our experiments are based only on deterministic functions, we would like to expand the study on random functions and hence consider non-deterministic models.

References

1. Schweizer, B., Sklar, A.: Probabilistic Metric Spaces. Elsevier-North-Holland (1983)
2. Sherwood, H.: On E-spaces and their relation to other classes of probabilistic metric spaces. J. London Math. Soc. **44**, 441–448 (1969)
3. Stevens, S.S.: Metrically generated probabilistic metric spaces. Fund. Math. **61**, 259–269 (1968)
4. Narukawa, Y., Taha, M., Torra, V.: On the definition of probabilistic metric spaces by means of fuzzy measures. Fuzzy Sets Syst. **465**, 108528 (2023). https://doi.org/10.1016/j.fss.2023.108528
5. Torra, V., Taha, M., Navarro-Arribas, G.: The space of models in machine learning: using Markov chains to model transitions. Prog. Artif. Intell. **10**(3), 321–332 (2021)
6. Sugeno, M.: Theory of fuzzy integrals and its applications. Ph.D. thesis, Tokyo Institute of Technology (1974)
7. Torra, V., Narukawa, Y., Sugeno, M.: Non-additive Measure-Theory and Applications. Studies in Fuzziness and Soft Computing, vol. 310. Springer, Berlin (2014). https://doi.org/10.1007/978-3-319-03155-2
8. Denneberg, D.: Non-additive Measure and Integral. Kluwer Academic (1994)
9. Wang, Z., Klir, G.J.: Generalized Measure Theory. Springer, Cham (2009). https://doi.org/10.1007/978-0-387-76852-6
10. Torra, V., Narukawa, Y.: Modeling Decisions: Information Fusion and Aggregation Operators. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68791-7
11. Searcóid, M.O.: Metric Spaces. Springer, Heidelberg (2007). https://doi.org/10.1007/1-84628-244-6
12. Torra, V., Navarro-Arribas, G.: Probabilistic metric spaces for privacy by design machine learning algorithms: modeling database changes. In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Livraga, G., Rios, R. (eds.) DPM/CBT 2018. LNCS, vol. 11025, pp. 422–430. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00305-0_30
13. Klement, E.P., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publisher (2000)

14. Salary data. https://www.kaggle.com/datasets/karthickveerakumar/salary-data-simple-linear-regression. Accessed 30 Apr 2022
15. Salem, A., Bhattacharya, A., Backes, M., Fritz, M., Zhang, Y.: Updates leak: data set inference and reconstruction attacks in online learning. In: Proceedings of the 29th USENIX Security Symposium, pp. 1291–1308 (2019)
16. Torra, V., Navarro-Arribas, G.: Integral privacy. In: Foresti, S., Persiano, G. (eds.) CANS 2016. LNCS, vol. 10052, pp. 661–669. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48965-0_44
17. Senavirathne, N., Torra, V.: Integrally private model selection for decision trees. Comput. Secur. **83**, 167–181 (2019)
18. Zhang, X., Wang, J., Zhan, J., Dai, J.: Fuzzy measures and choquet integrals based on fuzzy covering rough sets. IEEE Trans. Fuzzy Syst. **30**(7), 2360–2374 (2022). https://doi.org/10.1109/TFUZZ.2021.3081916

Paper IV

Generalized F-spaces through the lens of fuzzy measures

Mariam Taha and Vicenç Torra.

Fuzzy Sets and Systems, 507, Article 109317 (2025).

Generalized F-Spaces Through the Lens of Fuzzy Measures*

Mariam Taha[1]0000-0003-4233-3519 and Vicenç Torra[1]0000-0002-0368-8037

Dept. of Computing Science, Umeå University, Umeå 901 87, Sweden
{mariamt,vtorra}@cs.umu.se

Abstract. Probabilistic metric spaces are natural extensions of metric spaces, where the function that computes the distance outputs a distribution on the real numbers rather than a single value. Such a function is called a distribution function. F-spaces are constructions for probabilistic metric spaces, where the distribution functions are built for functions that map from a measurable space to a metric space.
In this paper, we propose an extension of F-spaces, called Generalized F-space. This construction replaces the metric space with a probabilistic metric space and uses fuzzy measures to evaluate sets of elements whose distances are probability distributions. We present several results that establish connections between the properties of the constructed space and specific fuzzy measures under particular triangular norms. Furthermore, we demonstrate how the space can be applied in machine learning to compute distances between different classifier models. Experimental results based on Sugeno λ-measures are consistent with our theoretical findings.

Keywords— Fuzzy measures; Probabilistic metric space

1 Introduction

Probabilistic metric spaces (PMS) are spaces where the distance between two points is represented as a probability distribution instead of a single numeric value, as in traditional metric spaces. This allows for a more flexible way to model uncertainty and variability in distances. Probabilistic metric spaces adhere to a set of axioms that ensure the consistent construction of these probabilistic distances. One key axiom is the triangle inequality, which has undergone various developments [1–4], One of the main concepts used to ensure the triangle inequality condition is that of triangular norms [9–11]. In the literature, there are only a few well-established methods for constructing PMSs. Among them, we find E-spaces [12, 13], where the PMS is constructed using sets of functions. These functions map from a probability space (base) into a metric space (target), and then distribution functions are defined in terms of measuring those elements

* This study was partially funded by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

whose distances do not exceed a certain threshold. In a previous work [14], we introduced another construction called F-spaces, which replaces the probability space with a measure space. As a result, the distribution functions are defined using non-additive measures instead of probabilities. This enables us to consider more complex interactions among the objects of the base space. In this paper, we propose a generalization of the F-space construction, where we consider mappings from a measurable space to a probabilistic metric space. Therefore, we require a fuzzy measure for the measurable space and distribution functions for the target space. This construction enables the representation of complex interactions among objects within the base space while accounting for scenarios where the distance between elements is uncertain in the target space. We also demonstrate several results linking the constructed space properties to specific measures under certain triangular norms. We present our findings through an application in machine learning, specifically in modeling similarities between machine learning models by considering the set of databases that these models generate. In this framework, the base space corresponds to the database space, while the model space serves as the target space. As we will discuss later, this approach has significant relevance for data privacy considerations.

The paper is organized as follows: Section 2 reviews previous definitions and results, primarily related to fuzzy measures and triangular norms. Section 3 discusses key findings related to probabilistic metric spaces and highlights some of their constructions. Section 4 presents our main contribution, including the definition of generalized F-spaces followed by several theoretical results. Examples and applications of our results are given in Section 5. Finally, in Section 6, the paper concludes with some final remarks.

2 Preliminaries

In this section, we review some definitions and results related to fuzzy measures and triangular norms.

2.1 Fuzzy Measures

Fuzzy measures, first introduced by Sugeno [16], are also known as capacities, non-additive measures, and monotone measures. They generalize classical measures [17–19] by assigning values to subsets of a universal set to capture the degree to which those subsets contribute to a certain property or concept.

Definition 1. *Let $(\Omega, \mathcal{A})$ be a measurable space. A set function μ defined on $\mathcal{A}$ is called a non-additive measure if and only if*

- $0 \leq \mu(A) \leq \infty$ *for any* $A \in \mathcal{A}$;
- $\mu(\emptyset) = 0$;
- *If* $A_1 \subseteq A_2 \subseteq \mathcal{A}$ *then*
$$\mu(A_1) \leq \mu(A_2)$$

If, in addition, $\mu(A) = 1$, then the fuzzy measure is said to be normalized. We consider finite sets Ω, and for simplicity we assume $\mathcal{A} = 2^{\Omega}$. Fuzzy measures exhibit certain properties that demonstrate how elements and subsets within a set interact and contribute to the overall measure. Below are some of the key properties.

Definition 2. *Let μ be a non-additive measure on the measurable space $(X, \mathcal{A})$. Then,*

- *μ is additive if $\mu(A \cup B) = \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is superadditive if $\mu(A \cup B) \geq \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is subadditive if $\mu(A \cup B) \leq \mu(A) + \mu(B)$ when $A \cap B = \emptyset$;*
- *μ is submodular if $\mu(A) + \mu(B) \geq \mu(A \cup B) + \mu(A \cap B)$;*
- *μ is supermodular if $\mu(A) + \mu(B) \leq \mu(A \cup B) + \mu(A \cap B)$;*
- *μ is symmetric if for finite X, when $|A| = |B|$, then $\mu(A) = \mu(B)$.*

A supermodular measure indicates superadditivity, while a submodular measure suggests subadditivity. When fuzzy measures are additive and normalized, they become equivalent to probability measures.

Next, we introduce different types of fuzzy measures. The Sugeno λ-measure is characterized by a parameter λ that determines how subsets of the set interact.

Definition 3. *Let Ω be a finite set and let $\lambda > -1$. A Sugeno λ-measure is a function $\mu : 2^{\Omega} \to [0, 1]$ such that:*

- *$\mu(\Omega) = 1$*
- *if $A, B \subseteq \mathbf{X}$ with $A \cap B = \emptyset$ then*
  $\mu(A \cup B) = \mu(A) + \mu(B) + \lambda\mu(A)\mu(B)$

For Sugeno λ-measures, as a convention, the measure of the singletons $\omega_i \in \Omega$ is called a density and is denoted by $v(\omega_i)$. In this case, as the measure is normalized when $\Omega = \{\omega_1, \omega_2, ..., \omega_n\}$, λ should satisfy the following:

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda v(\omega_i)). \tag{1}$$

If the densities are known, the above polynomial can be used to uniquely determine the value of λ. Then given the densities and λ, the fuzzy measure $\mu(A)$ is defined as :

$$\mu(A) = \left\{ \begin{array}{cc} v(x_i), & A = \{x_i\} \\ \frac{1}{\lambda} \prod\limits_{x_i \in A} (1 + \lambda v(x_i)) - 1, & |A| \neq 1 \quad \& \quad \lambda \neq 0 \\ \sum\limits_{x_i \in A} v(x_i), & |A| \neq 1 \quad \& \quad \lambda = 0 \end{array} \right\}$$

- If $\lambda > 0$, the measure exhibits superadditivity, meaning the combined effect of the subsets is greater than the sum of their individual effects.

– If $\lambda < 0$, the measure exhibits subadditivity, where the combined effect is less than the sum.
– If $\lambda = 0$, the measure behaves additively, similar to a probability measure.

Definition 4. *Given the reference set Ω, a fuzzy measure such that $\mu(A) < 1$ for all $A \subset \Omega$ and $A \neq \Omega$ is called a 1^--measure.*

Definition 5. *Let A_0 be a subset of Ω, then the set function defined by $\mu_{A_0}(A) = 1$ if and only if $A_0 \subseteq A$, is a non-additive measure.*

In the fuzzy measure μ_{A_0}, the reference set A_0 represents a baseline for the measure, and the value of the measure is computed with respect to the inclusion relationship with this reference set.

2.2 Triangular Norms

Triangular norms (t-norms) are binary operations that generalize the logical conjunction (AND operation) [10, 11]. They are used primarily in the context of fuzzy logic, probabilistic metric spaces, and fuzzy set theory.

Definition 6. *A function $\top : [0,1] \times [0,1] \to [0,1]$ is a triangular norm if and only if it satisfies the following properties:*

– $\top(x,y) = \top(y,x)$ *(symmetry or commutativity)*
– $\top(\top(x,y),z) = \top(x,\top(y,z))$ *(associativity)*
– $\top(x,y) \leq \top(x',y')$ *if $x \leq x'$ and $y \leq y'$ (monotonicity)*
– $\top(x,1) = x$ *for all x (neutral element 1)*

Definition 7. *A t-norm T_1 is stronger than a t-norm T_2, if*

$$T_1(x,y) \geq T_2(x,y) \quad \forall (x,y) \in [0,1] \times [0,1].$$

If T_1 is stronger than T_2, we write $T_1 \geq T_2$.

Definition 8. *A t-norm T is strict if it is continuous on $[0,1] \times [0,1]$ and strictly increasing in each argument on $(0,1]^2$.*

Examples of t-norms are the Minimum, $T(x,y) = min(x,y)$ denoted by Min, the algebraic product $T(x,y) = x.y$ denoted by Π, and the Bounded Difference $T(x,y) = max(0, x+y-1)$ denoted by W, also known as the Lukasiewicz t-norm.

The order of the above t-norms is as follows: $W < \Pi < Min$.

3 Probabilistic Metric Spaces

Probabilistic metric spaces extend the classical concept of a metric space by replacing the standard distance function with a distance distribution function. In these spaces, the distance between two elements is not expressed as a single number, but as a distribution over possible distances. Probabilistic metric spaces were first introduced by K. Menger in 1942 [1] and were further developed by him in the early 1950s [5, 6]. B. Schweizer and A. Sklar [7] expanded the study of these spaces, contributing significantly to the development of the theory. In this section, we review the concepts related to probabilistic metric spaces.

3.1 Metric Spaces

A 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 [21].

Definition 9. *Let $d : S \times S \to \mathbb{R}^+$. Then d is called a metric on S if the following properties hold for $a, b, c \in S$:*

(i) $d(a, b) \geq 0$ *with equality if and only if* $a = b$ *(positive property),*
(ii) $d(a, b) = d(b, a)$ *(symmetry property), and*
(iii) $d(a, b) \leq d(a, c) + d(c, b)$ *(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.

Definition 10. *[7] (Def.4.1.1) A distribution function, is a non-decreasing function F defined on R, with $F(-\infty) = 0$, and $F(+\infty) = 1$. If F is defined on $\mathbb{R}^+$, and it satisfies (i) $F(0) = 0$; (ii) $F(\infty) = 1$, and (iii) it is left-continuous on $(0, \infty)$, then F is a distance distribution function.*

The set of all distance distribution functions is denoted by Δ^+. Distribution functions are usually linked to probabilities, where $F(x)$ is understood as the probability that the distance is less than or equal to x. Any classical distance a can be represented in terms of a distance distribution function using the unit step function ϵ_a as follows.

Definition 11. *[7] (Def. 4.1.4) For any a in $\mathbb{R}^+$, we define the unit step function $\epsilon_a \in \Delta^+$ by:*

$$\epsilon_a(t) = \begin{cases} 0, & 0 \leq t \leq a \\ 1, & a < t \leq \infty \end{cases}$$

This implies $\epsilon_a \leq \epsilon_b$ if and only if $b \leq a$. Triangle functions and Triangular norms play a crucial role in defining PMS; they generalize the triangle inequality in metric spaces.

Definition 12. *[7] (Def.7.1.1) A Triangle function T is binary operation on Δ^+ that, for any $F, G, H, K \in \Delta^+$, it satisfies the following:*

(i) $T(F, \epsilon_0) = F$,
(ii) $T(F, G) = T(G, F)$,
(iii) $T(F, G) \leq T(H, K)$ *whenever* $F \leq H, G \leq K$,
(iv) $T(T(F, G), H) = T(F, T(G, H))$.

Definition 13. *A triangle function T_1 is stronger than a triangle function T_2 (this is denoted $T_2 \leq T_1$), if for all $F, G \in \Delta^+$, and all $x \in \mathbb{R}^+$, $T_2(F, G)(x) \leq T_1(F, G)(x)$.*

We define the set of all binary operations that are non-decreasing as $\mathcal{J}$, and the set of all binary operations on R^+ which are non-decreasing, continuous, and with range R^+ as $\mathcal{L}$. Next, we define one family of triangle functions that are built from a t-norm.

Definition 14. *[7] (Def.7.1.2) Let T be a left-continuous t-norm, then the function T: $\Delta^+ \times \Delta^+ \to \Delta^+$ defined by $T_\tau(F,G)(x) = T(F(x),G(x))$ is a triangle function.*

Definition 15. *[7] (Def.7.2.1) For any T in $\mathcal{J}$, and L in $\mathcal{L}$, $\tau_{T,L}$ is the function on $\Delta^+ \times \Delta^+$ which is defined by: $\tau_{T,L} = sup\{T(F(u),G(v))|L(u,v)=x\}$*

If $L = Sum$, then we drop L in $\tau_{T,L}$ and simply write τ_T. Below, we give an example of a triangle function that is built from a triangular norm.

Example 1. The maximal triangle function is $T_M(F,G)(x) = min(F(x),G(x))$. For any triangle function T we have:

$$T(F,G) \leq T(F,\epsilon_0) = F,$$
$$T(F,G) \leq T(G,\epsilon_0) = G,$$

Hence,

$$T(F,G)(x) \leq \min(F(x),G(x)) = T_M(F,G)(x).$$

We are now in conditions to define probabilistic metric space.

Definition 16. *[7] (Def.8.1.1) Let $(S,\mathcal{F},\tau)$ be a triple where S is a nonempty set, $\mathcal{F}$ is a function from $S \times S$ into Δ^+, and τ is a triangle function; then $(S,\mathcal{F},\tau)$ is a probabilistic metric space (PM space) if the following conditions are satisfied for all p, q, and r in S:*

(i) $\mathcal{F}(p,p) = \epsilon_0$
(ii) $\mathcal{F}(p,q) \neq \epsilon_0$ *if* $p \neq q$
(iii) $\mathcal{F}(p,q) = \mathcal{F}(q,p)$
(iv) $\mathcal{F}(p,r) \geq \tau(\mathcal{F}(p,q),\mathcal{F}(q,r))$.

Given a probabilistic metric space $(S,\mathcal{F},\tau)$, we say that $(S,\mathcal{F})$ is a probabilistic metric space under τ.

A probabilistic pseudometric space (PPM space) $(S,\mathcal{F},\tau)$ is defined as above but does not require condition (ii). When all conditions above apply except condition (iv), we have a probabilistic semimetric space. When all conditions apply except condition (iii), we have a probabilistic quasimetric space.

We shall denote the distribution function $\mathcal{F}(p,q)$ by $\mathcal{F}_{pq}$, therefore $\mathcal{F}_{pq}(x)$ is read as the probability that the distance between p and q is less than x. One of the classifications of probabilistic metric space is based on the properties of the triangle function τ.

Definition 17. *[7] (Definition 8.1.4) Let $(S,\mathcal{F},\tau)$ be a probabilistic metric space. Then $(S,\mathcal{F},\tau)$ is proper if*

$$\tau(\epsilon_a,\epsilon_b) \geq \epsilon_{a+b}$$

for all a,b in $\mathbb{R}^+$.

If $\tau = \tau_\top$ for some t-norm $\top$, then $(S,\mathcal{F},\tau)$ is a Menger space, or equivalently $(S,\mathcal{F})$ is a Menger space under $\top$.

Theorem 1. *[7] (Theorem 8.1.5) If $(S,F,\tau_{T,L})$ is a probabilistic metric space, T is t-norm, L is monotonic, where $L \leq Sum$, then $(S,F,\tau_{T,L})$ is proper.*

3.2 Constructions of Probabilistic Metric Spaces

One of the constructions that can lead to a probabilistic metric space is the E-space, which was first introduced by [12, 13]. E-spaces allow us to compute the distance between functions that map from a probability space to a metric space, where the distance is defined in terms of measuring those elements that are at most at distance x, using a probability measure. In a previous study [14], we introduced another construction, which we call F-space, which permits us to consider more complex interactions using fuzzy measures. We also proved several results that link the type of norms with the properties of the constructed space.

Definition 18. *[14] Let $(\Omega, \mathcal{A})$ be a measurable space, and let μ be a non-additive measure on $(\Omega, \mathcal{A})$. 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 F-space with base $(\Omega, \mathcal{A}, \mu)$ and target (M, d) if*

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

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

belongs to $\mathcal{A}$.

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

$$F^{\mu}_{pq}(x) = \mu(\{\omega \in \Omega | d(p(\omega), q(\omega)) < x\}). \tag{2}$$

Theorem 2. *[14] Let $(\Omega, \mathcal{A})$ be a measurable space, let μ be a non-additive measure on $(\Omega, \mathcal{A})$ and $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$.*

Then, if μ is a supermodular non-additive measure on $(\Omega, \mathcal{A})$, it follows that $(S, \mathcal{F})$ is a probabilistic pseudometric space under the bounded difference τ_W.

4 Generalized F-spaces and Main Results

A natural extension of F-spaces is to generalize the target spaces, allowing for functions that map from a measurable space to a probabilistic metric space. We call this construction a Generalized F-space. An illustration of this space is given in Figure 1. This concept is introduced in the following definition.

Definition 19. *Let $(\Omega, \mathcal{A})$ be a measurable space, and let μ be a non-additive measure on $(\Omega, \mathcal{A})$. Let $(M, \tilde{d}, \tau)$ be a probabilistic metric space where $\tilde{d}$ is a mapping from $M \times M$ into Δ^+, and τ is a triangle function. 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 F-space with base $(\Omega, \mathcal{A}, \mu)$ and target $(M, \tilde{d})$ if*

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

$$\{\omega \in \Omega \mid \tilde{d}(p(\omega), q(\omega)) > \epsilon_x\}$$

belongs to $\mathcal{A}$.

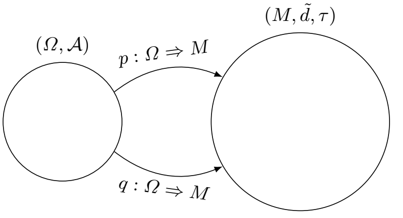


Fig. 1: Generalized F-Space Representation with the measurable space $(\Omega, \mathcal{A})$ and the probabilistic metric space $(M, \tilde{d}, \tau)$.

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

$$F^{\mu}_{pq}(x) = \mu(\{\omega \in \Omega \mid \tilde{d}(p(\omega), q(\omega)) > \epsilon_x\}) \quad (3)$$

We can prove the following.

Theorem 3. *Let $(\Omega, \mathcal{A})$ be a measurable space, let $(M, \tilde{d}, \tau)$ be a proper PMS. Let μ be a non-additive measure on $(\Omega, \mathcal{A})$ and $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A}, \mu)$ and target $(M, \tilde{d}, \tau)$ according to Definition 19.*

Then, if μ is a supermodular non-additive measure on $(\Omega, \mathcal{A})$, it follows that $(S, \mathcal{F})$ is a probabilistic pseudometric space under bounded difference τ_W.

Proof. We first prove that F^{μ}_{pq} as defined in Equation 3 satisfies Property (i) in Definition 16. Observe that if $p = q$ then $p(w) = q(w)$ for all $w \in \Omega$. Therefore, $\tilde{d}(p(w), q(w)) = \epsilon_0$ for all w in Ω and $\Omega = \{\omega \in \Omega | \tilde{d}(p(\omega), q(\omega)) > \epsilon_x\}$ for all $x > 0$ because $\epsilon_0 > \epsilon_x$. Since $\mu(\Omega) = 1$, it follows that $F^{\mu}_{pq}(x) = 1$ for all $x > 0$. If $x = 0$, then $\emptyset = \{\omega \in \Omega \mid \tilde{d}(p(\omega), q(\omega)) > \epsilon_x\}$ and $F^{\mu}_{pq}(0) = 0$. Therefore, $F^{\mu}_{pq} = \epsilon_0$ and the Property is proven.

The proof that F^{μ}_{pq} satisfies Property (iii) is trivial. The symmetry of $\tilde{d}$ naturally implies the symmetry of F^{μ}_{pq}.

We now prove that F^{μ}_{pq} satisfies Property (iv). Let us consider any x in $\mathbb{R}^+$. Then, consider u and v such that $u + v = x$ and define the following sets:

– $A = \{w \in \Omega \mid \tilde{d}(p(w), q(w)) > \epsilon_u\}$,
– $B = \{w \in \Omega \mid \tilde{d}(q(w), r(w)) > \epsilon_v\}$, and
– $C = \{w \in \Omega \mid \tilde{d}(p(w), r(w)) > \epsilon_x\}$.

We know that $\tilde{d}$ satisfies the triangle inequality.

Therefore, $C \supseteq A \cap B$ because if for a w, that it holds $\tilde{d}(p(w), q(w)) = u_0 > \epsilon_u$ (i.e., $w \in A$) and $\tilde{d}(q(w), r(w)) = v_0 > \epsilon_v$ (i.e., $w \in B$). Since $x = u + v$ and τ is proper, then $\epsilon_x = \epsilon_{u+v} \leq \tau(\epsilon_u, \epsilon_v)$. Then, from property (iii) in Definition 12 we know $\tau(u_0, v_0) > \tau(\epsilon_u, \epsilon_v)$. Therefore, $\epsilon_x = \epsilon_{u+v} \leq \tau(\epsilon_u, \epsilon_v) < \tau(u_0, v_0)$. In addition, we also know from triangle inequality in Definition 16 that

$$\tau(u_0, v_0) = \tau(\tilde{d}(p(w), q(w)), \tilde{d}(q(w), r(w))) \leq \tilde{d}(p(w), r(w)).$$

Therefore $\epsilon_x < \tilde{d}(p(w), r(w))$, thus $w \in C$. So, we have proven $C \supseteq A \cap B$.

Now, as $C \supseteq A \cap B$, by the monotonicity condition of μ, we have

$$\mu(C) \geq \mu(A \cap B)$$

and by supermodularity

$$\mu(A \cap B) \geq \mu(A) + \mu(B) - \mu(A \cup B).$$

As $1 = \mu(\Omega) \geq \mu(A \cup B)$, we have that

$$\mu(A \cap B) \geq \mu(A) + \mu(B) - \mu(A \cup B) \geq \mu(A) + \mu(B) - 1,$$

and naturally $\mu(A \cap B) \geq 0$. Therefore,

$$\mu(C) \geq \mu(A \cap B) \geq \max(\mu(A) + \mu(B) - 1, 0) = W(\mu(A), \mu(B)). \tag{4}$$

Let us consider the expressions for $F^{\mu}_{pq}(u)$, $F^{\mu}_{qr}(v)$, and $F^{\mu}_{pr}(x)$ according to Equation 3 and the sets $\mu(A)$, $\mu(B)$, and $\mu(C)$ as defined above. Then, we have that Equation 4 implies for every $x = u + v$

$$F^{\mu}_{pr}(x) \geq W(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)).$$

Therefore,

$$F^{\mu}_{pr}(x) \geq \sup\{W(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)) | u + v = x\} = \tau_W(F^{\mu}_{pq}, F^{\mu}_{qr})(x), \tag{5}$$

and Property (iv) holds with τ_W.

As Properties (i), (iii), and (iv) hold, the theorem is proven. □

Theorem 4. *Let $(S, \mathcal{F})$ be a probabilistic pseudo metric space with base $(\Omega, \mathcal{A})$ and target $(M, \tilde{d}, \tau)$. If F is a canonical F-space (i.e., $\mathcal{F}$ satisfies property (ii) in Definition 16), then it is a proper Menger space under W.*

Proof. If $(S, \mathcal{F})$ is a canonical F-space, this means that all the properties in Definition 16 are satisfied. Then, it is a probabilistic metric space.

From Definition 15 for a t-norm T and L as the sum function we have

$$\tau_{\top}(\epsilon_a, \epsilon_b)(x) = \sup\{\top(\epsilon_a(u), \epsilon_b(v)) \mid u + v = x\}$$

and, therefore,

$$\tau_T(\epsilon_a, \epsilon_b) = \epsilon_{a+b}.$$

So, Equation 5 implies that $(S, \mathcal{F}, \tau)$ is proper, and as we are using t-norm W, $(S, \mathcal{F})$ is a Menger space under $\top$. □

Theorem 5. *Let $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A})$ and target $(M, \tilde{d}, \tau)$. Let μ be a 1^--measure on $(\Omega, \mathcal{A})$. Then $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\top_d}$.*

Proof. The proof of this theorem follows the proof of Theorem 3. In particular, conditions 1 and 3 are proven in the same way.

Then, we define the sets A, B, and C in the same way as we defined them in Theorem 3. We have, therefore $C \supseteq A \cap B$. By the monotonicity condition, it follows that $\mu(C) \geq \mu(A \cap B)$.

Now, let us consider the following case: $A \neq \Omega$ and $B \neq \Omega$. We can prove that $\top_d(\mu(A), \mu(B)) = 0$ because both $\mu(A) < 1$ and $\mu(B) < 1$. Therefore, it is clear that $\mu(A \cap B) \geq \top_d(\mu(A), \mu(B)) = 0$.

Another case is when $A = \Omega$. Then, $A \cap B = B$. Therefore $\mu(A \cap B) = \mu(B) = \top_d(\mu(A), \mu(B)) = \top_d(1, \mu(B)) = \mu(B)$.

Finally, we have the case in which $B = \Omega$, which is analogous to the previous one and we can also prove $\mu(A \cap B) = \top_d(\mu(A), \mu(B))$.

Therefore, we have that

$$\mu(C) \geq \mu(A \cap B) \geq \top_d(\mu(A), \mu(B)),$$

and we can proceed as we did in Theorem 3 defining $F^{\mu}_{pq}(u)$, $F^{\mu}_{qr}(v)$, and $F^{\mu}_{pr}(x)$, and obtain the equation:

$$F^{\mu}_{pr}(x) \geq \top_d(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)).$$

From this equation, we prove that

$$\begin{aligned} F^{\mu}_{pr}(x) &\geq \sup\{\top_d(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)) \mid u + v = x\} \\ &= \tau_{\top_d}(F^{\mu}_{pq}, F^{\mu}_{qr})(x), \end{aligned}$$

holds for $\tau_{\top_d}$. Therefore, the theorem is proven. □

We prove the following theorem that considers non-additive measures μ_{A_0} introduced in Definition 5.

Theorem 6. *Let $(S, \mathcal{F})$ be an F-space with base $(\Omega, \mathcal{A})$ and target $(M, \tilde{d}, \tau)$. Let μ_{A_0} be a non-additive measure defined on $(\Omega, \mathcal{A})$ for a given set $A_0 \subseteq A$. Then $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\min}$.*

Proof. The proof of this theorem also follows the proof of Theorem 3. So, we focus on the proof of property (iv). We also consider A, B, and C as above and that $\mu_{A_0}(C) \geq \mu_{A_0}(A \cap B)$ due to the monotonicity of μ.

As $\mu_{A_0}(A) = 1$ if and only if $A \supseteq A_0$, it is easy to prove that when A_1 and A_2 are such that $\mu_{A_0}(A_1 \cap A_2) = 1$ it means that $A_1 \cap A_2 \supseteq A_0$ and therefore both $A_1 \supseteq A_0$ and $A_2 \supseteq A_0$. Therefore,

$$\mu_{A_0}(A_1 \cap A_2) = 1 = \min(\mu_{A_0}(A_1), \mu_{A_0}(A_2)) = \min(1, 1) = 1.$$

Then, if $\mu_{A_0}(A_1 \cap A_2) = 0$ this means that $A_1 \cap A_2$ does not include A_0. Therefore, it is not possible that both A_1 and A_2 include A_0. At most one of them can include A_0. This means that either $\mu_{A_0}(A_1) = 0$ or $\mu_{A_0}(A_1) = 0$. Thus,

$$\mu_{A_0}(A_1 \cap A_2) = 0 = \min(\mu_{A_0}(A_1), \mu_{A_0}(A_2)) = 0.$$

In a way analogous to previous proofs, we obtain that

$$F^{\mu}_{pr}(x) \geq \sup\{\min(F^{\mu}_{pq}(u), F^{\mu}_{qr}(v)) \mid u + v = x\} = \tau_{\min}(F^{\mu}_{pq}, F^{\mu}_{qr})(x),$$

which holds for $\tau_{\min}$. Therefore, the theorem is proven. □

5 Example and Application in Machine Learning

In this section, we illustrate our results with a toy example first, then it follows an application in Machine Learning.

Example 2. Let the measurable space $(\Omega, \mathcal{A})$ be represented by $\Omega = \{w_1, w_2, w_3\}$, and $\mathcal{A} := 2^{\Omega}$. Let $M = \{1, 2, 4, 6\}$, and for any $a, b \in M$, and assume the distance distribution functions $\tilde{d}(a, b)$ are defined as shown in Figure 2.
Then, choosing the proper t-norm $\tau_{\min}$, the target space can be constructed as $(M, \tilde{d}, \tau_{\min})$. Suppose $S = \{p, q, s\}$ are the functions that map from the measurable space Ω to the target space M as defined in Table 1.

Table 1: Functions p, q and s mapping from $\Omega := \{\omega_1, \omega_2, \omega_3\}$ into M.

	ω_1	ω_2	ω_3
p	4	6	2
q	2	1	4
s	1	4	6

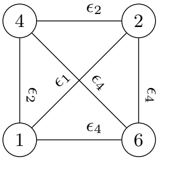


Fig. 2: Distance Distribution functions between elements in M

Therefore for any $p, q \in S$, the distribution function is defined as

$$F^{\mu}_{pq}(x) = \mu(\{\omega_i \in \Omega | \tilde{d}(p(\omega_i), q(\omega_i)) > \epsilon_x\})$$

Let H_{pq} be a set function defined as $H_{pq}(x) = (\{\omega \in \Omega | \tilde{d}(p(\omega), q(\omega)) > \epsilon_x\})$, the values of this function are given in Table 2.

Table 2: Functions $H_{p,q}$ for $p, q \in S$.

	$0 \leq x < 3$	$3 \leq x < 5$	$x \geq 5$
H_{pq}	$\emptyset$	$\{\omega_1, \omega_3\}$	Ω
H_{qs}	ω_1	$\{\omega_1, \omega_2\}$	Ω
H_{ps}	$\emptyset$	ω_1	Ω

Now, let us define Sugeno λ-measures μ on the space $(\Omega, \mathcal{A})$ according to Definition 3, and solve Equation 1 when the measures for the singletons are equal (i.e., $\mu(\{w_i\}) = k$). This yields:

$$k = \frac{1}{\lambda}(\exp(\frac{1}{n}\ln(1+\lambda) - 1) \tag{6}$$

For $\lambda = 0.5$, we get $\mu(\{w_i\}) = 0.289428$, and thus $\mu(\{w_1, w_2\}) = \mu(\{w_1, w_3\}) = 0.620740$. Then, using Equation 3 the distribution functions $F_{p,q}$, for any $p, q \in S$, are constructed as in Table 3.

Table 3: Functions $F_{p,q}$ for $p, q \in S$.

	$0 \leq x < 3$	$3 \leq x < 5$	$x \geq 5$
F_{pq}	0	0.620740	1
F_{qs}	0.289428	0.620740	1
F_{ps}	0	0.289428	1

If we choose the t-norm $\tau = W$, then we can see that all inequalities hold for the space (S, F, τ), therefore the results are aligned with Theorem 3. Now, let us choose the non-additive measure μ_{A_0}, which is introduced in Definition 5. Let us define the sets A_0, and A_1 as follows: $A_0 = \{w_1\}, A_1 = \{w_1, w_3\}$. Then, for the non-additive measures μ_{A_0}, and μ_{A_1}, the functions F are given in Table 4 and Table 5 respectively. In both cases, it is clear that the resulting space $(S, \mathcal{F})$ is a probabilistic pseudometric space under $\tau_{\min}$. Therefore, this result is aligned with Theorem 6.

Table 4: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on fuzzy measure μ_{A_0}

	$0 \leq x < 3$	$3 \leq x < 5$	$x \geq 5$
F_{pq}	0	1	1
F_{qs}	1	1	1
F_{ps}	0	1	1

Table 5: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on fuzzy measure μ_{A_1}

	$0 \le x < 3$	$3 \le x < 5$	$x \ge 5$
F_{pq}	0	1	1
F_{qs}	0	0	1
F_{ps}	0	0	1

5.1 Application in Machine Learning

Over the past decade, Machine Learning (ML) has advanced significantly. Models now require constant updates as data evolves. However, adversaries can exploit these updates to extract sensitive information from models [24]. Data privacy [23] is the field that studies how to prevent the disclosure of sensitive information. Integral Privacy [26, 27] addresses this by ensuring that model transformations during training do not leak information about the model changes or the training data. This approach involves using recurrent models, which are generated from combinations of datasets rather than a single dataset. Therefore, to understand the privacy implications of these models, it is essential to study the relationship between databases and models. Probabilistic metric spaces provide a useful framework for defining metrics in ML [22, 15]. As previously discussed, our focus is on estimating model similarities and defining a distance between models based on their generators. Our approach computes the distance distribution function by representing the database as a measurable space and the model space as a probabilistic metric space. This allows us to account for uncertainties and variations in model performance across different datasets, parameters, and sampling strategies. For example, model performance may fluctuate due to data shifts, noise, or randomness during training. By representing distances probabilistically, we capture these variations, offering a more reliable comparison between models under diverse conditions.

To demonstrate our results in machine learning, we formulate the problem as follows: the measurable space $(\Omega, \mathcal{A})$ corresponds to the space of possible databases, and μ is a non-additive measure on this space. In addition, the target space $(M, \tilde{d}, \tau)$ represents the model space, where M is a set of machine learning models $\{m_1, m_2, ..., m_n\}$. Then, the functions in S represent algorithms that build a machine learning model in M from a database in Ω. In our case $S = \{q_1, q_2, q_3\}$ represents three different classification algorithms that map from the database space into the model space. Therefore, $q_i(w_i) = m_i$, means that m_i is the resulting models after applying the algorithm q_i on database w_i. In this experiment the following algorithms are used to represent S:

- Logistic Regression (p).
- Random Forest (q).
- Support Vector Machines (r).

We conducted the experiment on the well-known Iris dataset, this dataset consists of 150 samples of Iris flowers from three different species. The database space

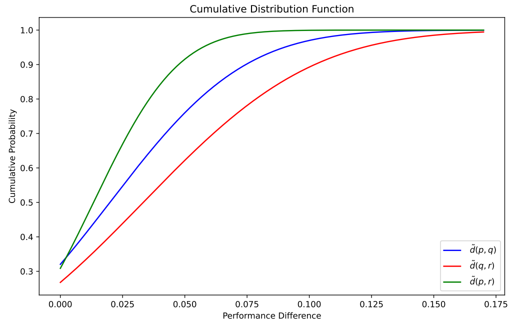


Fig. 3: Distribution functions for each pair of models for Sample 1.

was approximated with 1000 different samples. This means we had 1000 datasets in Ω. Models in the target space are represented by their 10-fold cross-validation accuracy vector of the dataset. Distribution functions $\tilde{d}$ are the empirical cumulative distribution functions of the differences between pairs of models, i.e. if the difference between a pair of models is denoted by $D = (d_1, d_2, ..., d_n)$, then:

$$\tilde{d}(p(w), q(w))(x) = (\tfrac{1}{n}) \textstyle\sum_{i=1}^{n} \mathbb{1}\{d_i \leq x\}$$

Where $\mathbb{1}$ is an indicator function that equals 1 if the condition $d_i \leq x$ is true, and 0 otherwise. The distances are built under the Lukasiewicz t-norm. Figure 3 illustrates the distribution functions for the first sample.

The aim, therefore, is to build the space (S, F) where F are the distance distribution functions between any two algorithms in the set S, these distances are evaluated by measuring those elements in the database space for which their distance is greater than ϵ_x, in the model space. We illustrate the experiment with two different non-additive measures, mainly, the Sugeno λ-measure, and the measure μ_{A_0}.

Sugeno $\boldsymbol{\lambda}$-measure.
We define Sugeno λ-measure from Definition 3 to build the measure on the database space $(\Omega, \mathcal{A}, \mu)$. For simplicity, all the singleton measures are assumed to be equal. i.e. $v(x_i) = k$, for all $x_i \in X$. We can therefore use Equation 6 to compute the measure values for the singletons. Now let us define the function H_{q_1,q_2} as follows: for any $q_1, q_2 \in S$ and for $x \geq 0$, $H_{q_1,q_2}(x) = \{DB \mid \ \tilde{d}(q_1(DB), q_2(DB)| > \epsilon_x\}$. Let us define the function $l(H)$ as the number of elements in H (i.e., the cardinality of the set). This function $l(H)$ is given in Table 6.

Table 6: Functions $l(H_{q_1 q_2})(x)$ for $q_1, q_2 \in S$.

	x=0	$0 < x \leq 0.1$	$0.1 < x \leq 0.117$	$x > 0.117$
$l(H_{pq})$	0	671	825	1000
$l(H_{qr})$	0	557	760	1000
$l(H_{pr})$	0	786	901	1000

Let us consider $\lambda = 0.6$. By applying Equation 6, we find $\mu(\{w_i\}) = 0.000783$. The measure μ and the results in Table 6 allow us to build the functions F, which are presented in Table 7. The distribution functions are visualized in Figure 4. Given that the measure is supermodular, the findings are consistent with Theorem 3, which states that the resulting space (S, F, τ) is a probabilistic pseudometric space under the t-norm τ_W. Figure 5 illustrates the validity of the triangular inequality from Definition 16. That is, in our case, $\mathcal{F}(p,r) \geq \tau(\mathcal{F}(p,q), \mathcal{F}(q,r))$. As we can see in this example, the distribution F_{pr} (blue line) is noticeably larger than the other distribution (orange line), which is obtained from combining F_{pq} and F_{qr} using Lukasiewicz t-norm τ_W.

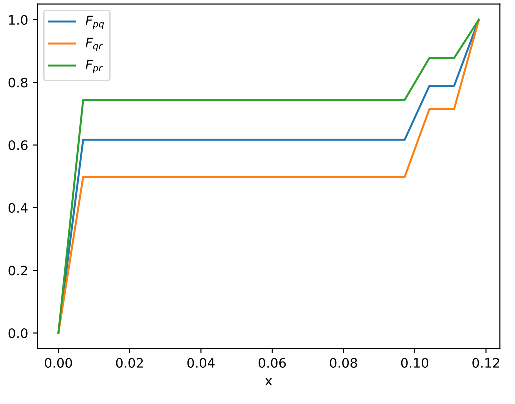


Fig. 4: Distribution functions F for the space (S, F, τ_w).

Fuzzy Measure μ_{A_0}.

We use the fuzzy measure μ_{A_0} introduced in Definition 5. Suppose $A_0 = H_{q,r}(x)$

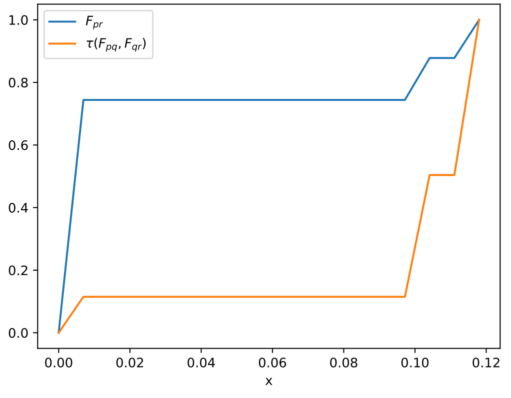


Fig. 5: Validation of Triangle Inequality under t-norm τ_W

Table 7: Functions F_{p_1,p_2} for $p_1, p_2 \in S$ based on Sugeno λ-measure ($\lambda = 0.6$)

	x=0	$0 < x \leq 0.1$	$0.1 < x \leq 0.117$	$x > 0.117$
F_{pq}	0	0.617	0.789	1
F_{qr}	0	0.498	0.715	1
F_{pr}	0	0.744	0.878	1

for $x \in (0, 0.1]$. Then, the only sets which include A_0 are $H_{q,r}$ for $x \in (0.1, 0.117]$, and $H_{q,r}$, $x > 0.117$. The other sets are incomparable under the same domain. Because of this, we obtain the distribution functions $F_{p,q}$, $F_{q,r}$, and $F_{p,r}$ as shown in Table 8. We can see in this table the space (S, F) is a probabilistic pseudo-metric space under τ_{min}, which is consistent with Theorem 6.

Table 8: Functions F_{q_1,q_2} for $q_1, q_2 \in S$ based on Fuzzy measure μ_{A_0}

	x=0	$0 < x \leq 0.1$	$0.1 < x \leq 0.117$	$x > 0.117$
F_{pq}	0	0	0	1
F_{qr}	0	1	1	1
F_{pr}	0	0	0	1

6 Conclusion

In this paper, we proposed Generalized F-space, a construction of probabilistic metric spaces characterized by a set of functions S, and distance distribution functions F based on a non-additive measure μ. This approach extends the traditional F-space by broadening the target space from a conventional metric space to a probabilistic metric space. This generalization not only accounts for potential dependencies among objects in the target space but also incorporates uncertainty when measuring distances within the target space.

We have provided several results demonstrating links between the type of the space and the properties of the fuzzy measure. We illustrated how these results can be utilized in the domain of Machine learning. Specifically, we constructed a database space based on Iris dataset and modeled the set function S with three different classification algorithms. The distances in the model space were defined based on their performance differences, and the resulting Generalized F-space was built under two distinct fuzzy measures: The Sugeno λ measure, which captures similar interactions within the database, and the non-additive measure μ_{A_0}, which depends on the inclusion relationships of a reference set. The experimental results confirmed the consistency of our theoretical findings. In future work, we aim to extend this work to larger datasets and further investigate the properties of the space. A particular focus will be placed on addressing the challenges of model selection and exploring recurrent models in settings where the database space is anonymized.

References

1. Menger, K. Statistical Metrics. *Proc. Nat. Acad. Sci. U.S.A.*, 28 (1942), 535–537.
2. Wald, A. On a Statistical Generalization of Metric Spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 29 (1943), 196–197.
3. Schweizer, B., Sklar, A. Statistical Metric Spaces. *Pacific J. Math.*, 10 (1960), 313–334.
4. Šerstnev, A. N. On the Probabilistic Generalization of Metric Spaces. *Kazan: Gosudarst. Univ., Učenye Zapiski*, 124 (1967), 109–119.
5. Menger, K. Probabilistic Geometry. *Proc. Nat. Acad. Sci. U.S.A.*, 37 (1951), 226–229.
6. Menger, K. Geometric Generale (Chap. VII). *Memorial des Sciences Mathematiques*, No. 124 (1954), Paris.
7. Schweizer, B., Sklar, A. *Probabilistic Metric Spaces*. Elsevier-North-Holland, 1983.
8. Hadžić, O., Pap, E. *Fixed Point Theory in Probabilistic Metric Spaces*. Mathematics and Its Applications, Vol. 536, Kluwer Academic Publishers, Dordrecht/Boston/London, 2001.
9. Šerstnev, A. N. On Triangle Inequalities for Random Metric Space. *Selected Trans. in Math., Statist., and Probability*, 11 (1973).
10. Alsina, C., Frank, M. J., Schweizer, B. *Associative Functions: Triangular Norms and Copulas*. World Scientific, 2006.
11. Klement, E. P., Mesiar, R., Pap, E. *Triangular Norms*. Kluwer Academic Publishers, 2000.
12. Sherwood, H. On E-spaces and Their Relation to Other Classes of Probabilistic Metric Spaces. *J. London Math. Soc.*, 44 (1969), 441–448.
13. Stevens, S. S. Metrically Generated Probabilistic Metric Spaces. *Fund. Math.*, 61 (1968), 259–269.
14. Narukawa, Y., Taha, M., Torra, V. On the Definition of Probabilistic Metric Spaces by Means of Fuzzy Measures. *Fuzzy Sets and Systems*, 465 (2023), 108528. https://doi.org/10.1016/j.fss.2023.108528.
15. Torra, V., Taha, M., Navarro-Arribas, G. The Space of Models in Machine Learning: Using Markov Chains to Model Transitions. *Progress in Artificial Intelligence*, 10 (2021), 321–332. https://doi.org/10.1007/s13748-021-00242-6.
16. Sugeno, M. Theory of Fuzzy Integrals and Its Applications. PhD thesis, Tokyo Institute of Technology, 1974.
17. Torra, V., Narukawa, Y., Sugeno, M. Non-Additive Measure—Theory and Applications. *Studies in Fuzziness and Soft Computing*, Vol. 310, Springer, Berlin, 2014.
18. Denneberg, D. *Non-Additive Measure and Integral*. Kluwer Academic Publishers, 1994.
19. Wang, Z., Klir, G. J. *Generalized Measure Theory*. Springer, 2009.
20. Pap, E. *Null-Additive Set Functions*. Mathematics and Its Applications, Vol. 337, Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.
21. Searcóid, M. O. *Metric Spaces*. Springer, 2007.
22. Torra, V., Navarro-Arribas, G. Probabilistic Metric Spaces for Privacy by Design Machine Learning Algorithms: Modeling Database Changes. *Proc. DPM 2018*, LNCS 11025 (2018), 422–430.
23. Torra, V. *Guide to Data Privacy: Models, Technologies, Solutions*. Springer, 2022. https://doi.org/10.1007/978-3-031-12837-0.
24. Salem, A., Bhattacharya, A., Backes, M., Fritz, M., Zhang, Y. Updates Leak: Data Set Inference and Reconstruction Attacks in Online Learning. *Proc. 29th USENIX Security Symposium*, 2019, 1291–1308.

25. Fisher, R. A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2) (1936), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x.
26. Torra, V., Navarro-Arribas, G. Integral Privacy. *Proc. CANS*, LNCS 10052 (2016), 661–669.
27. Torra, V., Navarro-Arribas, G., Galván, E. Explaining Recurrent Machine Learning Models: Integral Privacy Revisited. *PSD*(2020), 62–73.
28. Zhang, X., Wang, J., Zhan, J., Dai, J. Fuzzy Measures and Choquet Integrals Based on Fuzzy Covering Rough Sets. *IEEE Trans. Fuzzy Syst.*, 30(7) (2022), 2360–2374. https://doi.org/10.1109/TFUZZ.2021.3081916.