University of Nottingham
UK | CHINA | MALAYSIA

Towards Effective Communication of Uncertain Information

Yu Zhao

20186379

Supervised by Christian Wagner, Brendan Ryan, and Direnc Pekaslan

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

March 2026

Abstract

Uncertain information is pervasive in real-world contexts, due to measurement limitations, the complexity and ambiguity of what is described; this is notably the case for human perceptions, which are often shaped by subjectivity and conceptual vagueness. In the social sciences, Likert and numeric scales are the predominant formats in survey-based research for eliciting responses about human perceptions. These responses are often analysed collectively to measure latent psychological constructs—such as mood, trust, and satisfaction—whose understanding is consequential for decision-making. While efficient to administer and analyse, the single-valued response formats of these scales can constrain the authentic expression of uncertainty and impose false precision.

Over recent decades, fuzzy sets and intervals have been put forward to represent and communicate uncertainty in human perceptions, from elicitation through analysis to decision support. Existing methods across these stages, however, have largely been developed independently, often resting on incompatible assumptions and pursuing divergent goals—for example, some discard uncertainty in favour of simplicity, while others retain it in forms too complex to be interpreted in practice.

Building on the state of the art, this thesis aims to develop and evaluate methods for communicating uncertainty in human-sourced information across an end-to-end workflow, in support of uncertainty-aware decision-making. This thesis further demonstrates the practical value of the proposed methods through their application within service research. This application focus is motivated by the sponsors of this research,[1] together with the practical need to communicate uncertainty in rail passengers' perceptions to stakeholders in support of rail service development.

More specifically, this thesis focuses on interval-valued representations as a unifying basis for communicating uncertainty throughout a conventional service evaluation workflow. The thesis contributions are organised around the key stages of this workflow:

At the elicitation stage, this thesis articulates the necessity of capturing uncertainty in customer responses, as the intangibility and subjectivity of services make customer perceptions difficult to elicit with precision. A UK rail passenger survey is designed to elicit comparable interval- and single-valued responses on the commonly used service evaluation constructs, producing an empirical dataset for systematic comparison across subsequent stages.

Reliability assessment is adopted as the subsequent stage, as it is essential for evaluating whether items intended to measure the same construct exhibit sufficient internal consistency. While Cronbach's α is the most widely used

[1]This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/S023305/1] and by the Rail Safety and Standards Board [project ref. COF-MLD-05] through Horizon Centre for Doctoral Training.

reliability coefficient and has recently been extended to accommodate interval-valued responses, this extension introduces interpretive assumptions that may be misleading in real-world settings. This thesis therefore introduces a novel extension which avoids these assumptions. The behaviour of the proposed coefficient is then evaluated through simulation against the existing extension.

Following reliability assessment, responses are often summarised to provide a succinct overview and facilitate comparison across items. For this purpose, recent research has developed methods to model interval-valued responses as fuzzy sets, which, while containing rich information, can be difficult to interpret given their two-dimensional complexity. This thesis therefore develops and evaluates methods for summarising fuzzy sets as intervals in order to communicate uncertainty while enabling efficient comparison in practice.

Decision-support methods typically serve as the final stage of service evaluation, transforming collected responses into actionable insights. Widely used methods such as Importance–Performance Analysis cannot present uncertainty, and may therefore mislead decision makers with a false sense of precision. The thesis therefore extends these methods with interval-valued representations, and demonstrates these extensions for uncertainty-aware decision-making using the UK rail passenger survey data.

To facilitate wider adoption in practice among researchers and practitioners from a variety of fields, it is important to make novel methods accessible through software support. An open-source R toolkit is thus developed as a final stage to provide an integrated implementation of the methods reviewed and developed in this thesis, with its use demonstrated through a service evaluation workflow using the UK rail passenger survey data.

In summary, this thesis advances the state of the art in the analysis and application of interval-valued representations for communicating human-sourced uncertain information, ultimately strengthening the basis for well-informed decision-making in data-driven contexts.

Acknowledgements

I am deeply grateful to my supervisors, Christian Wagner, Brendan Ryan, and Direnc Pekaslan. Christian has been my supervisor since my master's dissertation, and my understanding of research, my interest in it, and my motivation to pursue a PhD with real-world impact have all been shaped by him. He has always guided me, backed me up, and believed in me over the years, while also being a role model for me as both an excellent academic and a fantastic supervisor. Brendan has supported me consistently since the day we met at my PhD interview and has brought perspectives from outside computer science, especially on industry-facing applications, which have influenced how I think about my work. I am also grateful to Direnc, with whom discussions have always been open and relaxed, and whom I have always felt comfortable turning to whenever I needed advice or help. I have been very lucky with my supervisors, not only academically but also in terms of who they are as people. They are all deeply thoughtful, compassionate, and sincere individuals, and they also have a strong sense of purpose in their academic work. These qualities have had a lasting influence on me.

I am also grateful to EPSRC and Horizon CDT. Their financial support gave me the freedom to focus on my research, attend conferences, and build connections through academic communities. More than that, through the work of Andrea, Laura, Monica, and Adrian, Horizon CDT made values such as equality and inclusion feel real in practice. Through the events they organised, as well as the many thoughtful discussions and courses they offered, they helped us think more systematically about how to grow into independent researchers. They also looked after us, a group of students who often needed reminding and chasing, with enormous patience and care. I am also grateful to the Rail Safety and Standards Board for supporting this work as the industry partner. In particular, I would like to thank Paul for his advice and involvement, and Mel, who was always responsive, never demanding, and helpful in every way she could be. Their support gave me the motivation to

pursue research with a stronger real-world orientation.

I am also grateful to the academics who supported me during my PhD. Within LUCID, I would especially like to thank Chao Chen, Isaac Triguero, and Jon Garibaldi, all of whom have always looked out for me and been generous with their guidance, whether on broader questions of direction or on the details along the way. At the University of Nottingham, I am grateful to Alex Lang, Elena Nichele, John Harvey, and Steve Benford. I would like to thank Hongwei He, Javier Andreu-Perez, John Bateson, Ray Fisk, Vladik Kreinovich, and Zhixing Zhang, whom I met through conferences, for the many inspiring conversations that have influenced my research. I would also like to thank my internal examiner, Steven Furnell, and my external examiner, Scott Ferson, for their thorough reading of my thesis and for making the viva a pleasant and thought-provoking experience.

Looking further back, I would especially like to thank Junpu Fan, who first encouraged me to think critically about science in middle school through his humour and distinctive way of teaching, and Renquan Li, whose engaging teaching at university played an important role in leading me from physics to computer science. I also remain grateful to many other teachers who looked after me throughout my education, including Da Wang, Debin Kong, He Gao, Huixing Zuo, Juanjuan Qin, Lianrong Zhang, Pan Ma, Suying Wei, Xia Feng, Xiaoguang Wang, Xiaoyu Wang, Xingyu Mao, Ying Wang, Zhaoyang Wang, Zheng Wang, and Zhizhong Ding.

Among my friends, I would especially like to thank Te, Jingda, and Ray, who shared both a supervisor and an office with me. I am particularly grateful to Te for all the in-depth discussions and spirited debates we had, to Jingda for all the Dota games and Costco trips, and to Ray for being a steady and reliable presence throughout my PhD. The hotpot meals, barbecues, and card game nights we shared were a lovely part of my PhD life. Thanks are also due to Winnie, the best massage therapist in Nottingham, for helping me stay relaxed and refreshed with every visit. I would also like to thank my cohort

from Horizon CDT: Callum, Daniel, Emma, Gabrielle, Gregor, Matt, Milly, Nasser, Sam, Torran, and Yang, with whom I shared an important part of this journey. I also thank my closest friends outside academia, for being with me through different stages of my life, sharing many experiences along the way, and contributing so much to my growth and happiness.

I am also grateful for the hobbies that have enriched my life outside research. Martial arts have given me the inner strength to be disciplined, tough, and resilient; swimming has kept me healthy and has always been a source of happiness; Dota 2 has given me a way to relax while reinforcing my belief that hope often comes through persistence; and reading has introduced me to different ways of thinking while bringing me a deeper kind of joy, especially through the works of some of my favourite authors, Jin Yong, Terry Pratchett, and Naomi Novik.

I am also grateful to my family, and to the family friends and elders who have cared about me and supported me throughout my life. I thank my husband, my beloved life partner Stefan, whose belief in me and love for me have always reminded me that, beyond work, life itself is also full of joy, found in the meals, films, books, board games, and journeys we have shared. I also thank Tongtong, my sister in spirit, who has always loved me, understood me, and supported me since our teenage years. I thank Xiangxiang, my furry little brother at heart, for the comfort, companionship, and countless moments of joy he has brought me.

Finally, I am especially grateful to my parents. From helping me develop habits of reading and exercise, and encouraging me to think independently when I was young, to giving me the best educational opportunities they could possibly provide as I grew older, they have shaped so much of who I am today. During my PhD, from the challenges of studying abroad to the difficulties of COVID and everything beyond, I have been empowered by their trust, support, and unwavering belief in me, which have given me the confidence to become the person I have always wanted to be.

List of Publications

Published Work

- Y. Zhao, C. Wagner, B. Ryan, D. Pekaslan, and J. Navarro, "Interval Agreement Weighted Average – Sensitivity to Data Set Features," in *Proc. 2024 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2024, pp. 1–8.
- Y. Zhao, C. Wagner, B. Ryan, D. Pekaslan, and V. Kreinovich, "Effectively Communicating Information and Uncertainty from Fuzzy Sets: Why and How," in *Proc. 2025 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2025, pp. 1–7.
- Y. Zhao, C. Wagner, B. Ryan, and D. Pekaslan, "ISurvey: An R Toolkit for Interval-Valued Survey Data Analysis," in *Proc. 2026 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 2026.

Manuscripts Under Review

- Y. Zhao, C. Wagner, V. Kreinovich, B. Ryan, and D. Pekaslan, "Discontinuous Interval-Valued Defuzzification for Non-convex Fuzzy Sets," under review in *IEEE Transactions on Fuzzy Systems*.

Conference Attendance

- **AMA SERVSIG Conference 2024**
  Bordeaux, France, June 2024
  Oral presentation: "Capturing Richer Service Quality Assessments with an Interval-Valued Customer Satisfaction Index"

- **IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2024)**
  Yokohama, Japan, June 2024
  Oral presentation: "Interval Agreement Weighted Average – Sensitivity to Data Set Features"

- **IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2025)**
  Reims, France, July 2025
  Oral presentation: "Effectively Communicating Information and Uncertainty from Fuzzy Sets: Why and How"

- **UK Rail Research and Innovation Network (UKRRIN) Student Conference 2025**
  Sheffield, UK, January 2025
  Oral presentation: "Capturing Richer Service Quality Assessments with an Interval-Valued Customer Satisfaction Index"

- **International Human Factors Rail Conference 2025**
  London, UK, September 2025
  Poster presentation: "Communicating Richer Rail Customer Satisfaction with Interval-Valued Methods"

Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Full Form**
ALC	Averaging Level Cuts
CSI	Customer Satisfaction Index
DU	Disjoint and Union-based defuzzification
DUW	Disjoint, Union-based, and Weighted defuzzification
IA	Interval Average
IAA	Interval Agreement Approach
$IALC$	Interval-valued Averaging Level Cuts
IPA	Importance–Performance Analysis
IV	Interval-Valued
IVD	Interval-Valued Defuzzification (the general class of methods)
IVD	Interval-Valued Defuzzification (the specific method)
IWA	Interval Weighted Average
NU	Nested and Union-based defuzzification
NUW	Nested, Union-based, and Weighted defuzzification
RFH	Relative Frequency Histogram
SV	Single-Valued
$WIVD$	Weighted Interval-Valued Defuzzification

List of Notations

Notation	Description
$\mathbb{R}$	Set of all real numbers
A	Type-1 fuzzy set defined on $X = \mathbb{R}$
$\mu_A(x)$	Membership function of A, where $x \in X$
$h(A)$	Height of A, or h in equations
$\text{supp}(A)$	Support of A
$\text{core}(A)$	Core of A
$p(A)$	Number of peaks of A, or p in equations
A_i	i-th convex fuzzy subset of A
$\boldsymbol{A} = \{A_i\}_{i=1}^{p}$	Set of convex fuzzy subsets of A
$\breve{x}_i$	Position of the i-th trough of A
$\breve{X}_A = \{\breve{x}_i\}_{i=1}^{p-1}$	Ordered set of trough positions of A
$\mathcal{K}_c(\mathbb{R})$	Collection of all nonempty, compact, convex subsets of $\mathbb{R}$
$\overline{A}$	Continuous interval
A^-, A^+	Lower and upper bounds of $\overline{A}$
$m(\overline{A})$	Midpoint of $\overline{A}$
$s(\overline{A})$	Spread of $\overline{A}$
$\bar{\overline{A}}$	Discontinuous (multi-segment) interval
$\overline{A}_i$	i-th sub-interval of $\bar{\overline{A}}$
$q(\bar{\overline{A}})$	Number of sub-intervals in $\bar{\overline{A}}$, or q in equations
$\{\alpha_j\}_{j=1}^{m}$	Set of m α-cut levels, indexed by j
$\bar{\overline{A}}_{\alpha_j}$	α-cut of A at level α_j
$\overline{A}_i^{\alpha_j}$	i-th sub-interval of $\bar{\overline{A}}_{\alpha_j}$
$\overline{\boldsymbol{X}} = \{\overline{X}_i\}_{i=1}^{n}$	Set of n intervals, indexed by i
$d_\theta,\ d_H$	θ-distance and Hausdorff distance
$\overline{X}_H^*$	Fréchet mean under d_H
$\text{Var}_H^*(\overline{\boldsymbol{X}})$	Fréchet variance under d_H
$\alpha_\theta,\ \alpha_H$	Cronbach's α under d_θ and d_H
$\overline{X}_{ij}$	IV response of respondent i for item j
$\overline{T}_i = \sum_{j=1}^{k} \overline{X}_{ij}$	Total score of respondent i across k items, indexed by j
$\hat{I}_j$	Mean importance for attribute j
$\hat{c}_I$	Grand mean of importance across all items

Chapter 1

Introduction

1.1 Information Quantification: A Historical View

Communication is essential for human society to exchange information and support coordinated action [1]. Historically, people typically communicated through qualitative, context-dependent descriptions [2]. For example, individuals used themselves and parts of their body to measure all other objects, which varied from person to person and lacked a universal standard [3]. While these localised methods were practical for immediate, everyday interactions, they were insufficient for consistent information exchange and comparison across contexts [4].

From the late eighteenth century onwards, quantification increasingly became a prominent means of establishing and communicating objectivity, authority, and trust in both science and public life [4]. By introducing standardised metrics, such as the metric system, and applying early statistical tools like probability and mathematical averages to human populations, aspects of the physical world could be measured and communicated with more uniformity, comparability and calculability [2, 5]. Through quantification, stable equivalences were established via shared conventions and classifications, helping information travel across contexts—effectively "making things that hold" [5] (cf. [6]).

As these approaches advanced in the natural sciences, the social sciences faced growing critiques for lacking comparable objectivity [7]. In response, researchers in fields such as psychology increasingly adopted quantitative approaches, extending scientific measurement from physical objects to human perceptions and judgements [8]. Single-valued scales, such as numeric rating scales, Likert scales [9], and Visual Analogue Scales [10], were developed to quantify such subjective assessments by requiring respondents to indicate a precise point on a predefined scale.

Over the twentieth century, these scales became the primary tools used in multi-item surveys to measure psychological constructs[1]—the latent attributes that are not directly observable [11], such as mood, trust, and satisfaction. In this way, quantification was extended from the physical world to human-sourced information, while the growing reliance on precise numerical ratings later raised concerns about a "precision fallacy" [12] (cf. [13]).

1.2 The Precision Fallacy: Uncertainty and Measurement

Uncertainty is pervasive in the natural sciences. Even when measuring well-defined physical phenomena, uncertainty arises from observational and measurement limitations as well as the complexity and variability of real-world settings [14]. During the nineteenth century, scientific objectivity became closely tied to the minimisation of human subjectivity [15] and the pursuit of exact quantification, creating an assumption that valid knowledge must be expressed in precise numerical terms.

[1] In fields such as psychology and the broader social sciences where humans are the default subjects of study, the terms 'psychological construct' and 'construct' are frequently used interchangeably. For brevity, this thesis will use the term 'construct' hereafter.

Since then, efforts to pursue higher precision have continued [16], while, contrastingly, applied work has looked to balance precision against practical utility, seeking to determine a level of precision that is fit for purpose. Accordingly, residual measurement error is acknowledged and accepted when its impact on the intended use of the measurement is negligible [14].

Whereas in the natural sciences uncertainty is often treated as a consequence of limited measurement precision, in the social sciences a reverse misalignment can occur, in which measurements are overly precise relative to the subjective states they aim to capture [12]. This misalignment is driven by inherent uncertainty in self-reported judgements, which commonly arises from subjectivity and the cognitive demands of comprehension, judgement, and response [17]; the fact that human attitudes are often better represented as ranges of positions considered acceptable or objectionable by an individual [18]; and the linguistic and conceptual vagueness of terms used in description [19, 20]. As Russell [19] observed, such inherent imprecision is a hallmark of our "terrestrial life," meaning the strictly precise symbols of traditional logic cannot fully capture human reality.

The single-valued scales used to elicit perceptions largely disregard the inherent uncertainty in subjective states [12, 21]. These scales impose single-valued precision on inherently imprecise perceptions [12]. This practice is an instance of "precision fallacy"—a methodological flaw where the exactness of the measuring method exceeds the precision of the human experience it attempts to measure [12] (cf. [13]). As Cohen [22] pointed out, there is no "royal road" to statistical induction; researchers must rely on informed judgement, which necessitates measurement methods that accommodate human uncertainty rather than forcing a false precision [12].

1.3 Formalising Human Uncertainty: From Fuzzy Sets to Interval-valued Scales

Fuzzy set theory was introduced to formalise uncertainty inherent in human reasoning and language [20]. It has since been widely applied in computational modelling and control systems [23]. While fuzzy-based response formats have long been proposed to elicit uncertainty directly from respondents, their adoption in large-scale survey research has remained limited relative to the widespread use of single-valued scales.

Fuzzy Linguistic Scales, based on Zadeh's notion of linguistic variables [24], were developed in 1975 as an early attempt to incorporate fuzzy representations into survey scales (see Section 2.4.2 for details). Whereas single-valued scales encode each response as a single value and therefore do not explicitly represent the imprecision accompanying many attitudinal judgements, these scales can represent linguistic imprecision through predefined fuzzy sets; however, neither format captures respondent-specific interpretive variation—whereby the same linguistic term may carry different meanings for different respondents [25].

Fuzzy Rating Scales were later proposed in 1988 to elicit respondent-specific uncertainty by allowing respondents to construct their own fuzzy sets [12] (see Section 2.4.3 for details). However, asking lay respondents to specify membership functions increases response burden and can affect response quality in survey settings [21, 26], consistent with broader evidence on cognitive demands in survey responding [17]. In practice, this burden can necessitate pre-training, limiting accessibility and scalability in large-scale deployments [21].

Therefore, interval-valued scales have been developed as a lower-burden alternative for eliciting uncertainty from respondents. Compared with eliciting

full fuzzy sets, indicating an interval (i.e., a range) is a natural and easy-to-understand way for people to express uncertainty [21]. In particular, instead of requiring separate actions to specify lower and upper bounds, the recently developed ellipse-based response format integrates this into a single continuous ellipse-drawing interaction with minimal pre-training [27]. Recent applications of interval-valued scales have highlighted their practical value across multiple domains, including environmental management [28, 29], healthcare [25], consumer and market research [30], and cybersecurity [31]. In addition, the accessibility of interval-valued scales is now enhanced by DECSYS, a free and open-source toolkit for administering ellipse-based surveys digitally, with support for responses on smartphones, tablets, and desktop computers [27]. Therefore, this research adopts interval-valued scales as the primary method for eliciting uncertainty from human respondents in survey-based construct measurement.

1.4 Communicating Uncertainty: From Elicitation to Decision Making

Following information elicitation, the practical value of information is realised when it supports interpretation and purposeful action [32, 33]. For survey data, methods such as Importance–Performance Analysis [34] are often applied to present findings to stakeholders to support data-driven decision-making. In this context, presenting uncertainty is widely recognised as necessary: point estimates alone can create an "illusion of certitude" [35–37], while fulfilling the corresponding "duty to inform" requires making the limits of knowledge explicit [38]. Accordingly, it is important that uncertainty be handled across the analysis stages between elicitation and decision-facing outputs.

In recent decades, interval-valued representations have been used as practical

means of representing uncertainty and imprecision [39, 40] (cf. [41]). There is a growing body of methods that operate on interval-valued data, covering uncertainty summarisation and statistical analysis [39, 42, 43], modelling and visualisation via fuzzy-set representations derived from interval-valued responses [44], reliability assessment for interval-valued questionnaires [45, 46], and decision-support procedures such as association/correlation analysis [47, 48], ranking/evaluation of uncertain quantities [49, 50], and multi-criteria decision analysis under linguistic/uncertain information [51]. A critical limitation, however, is that these methods have largely been developed independently, which can make it challenging to use them within a single, integrated workflow.

For instance, symbolic data analysis often evaluates intervals as geometric sets via the Hausdorff distance [52, 53], whereas interval-valued reliability assessment typically uses a mid-spread decomposition (e.g., via the θ-distance) for computational convenience [46]. Such differences in assumptions can complicate integrated use, as methods from different stages may not be compatible. A further practical tension concerns mismatched aims across stages. Methods designed to represent and model uncertainty commonly aim to preserve uncertainty information for complete understanding, whereas decision-facing tasks often demand a precise ranking; consequently, methods such as defuzzification are used to obtain single-valued scores for ordering [54], thereby discarding the uncertainty elicited and modelled earlier and potentially undermining the original motivation for uncertainty-aware elicitation.

1.5 Aims and Objectives

This research aims to develop and evaluate methods for communicating uncertainty in human-sourced information and to establish an end-to-end integrated workflow, so that uncertainty is effectively captured and preserved

throughout to support well-informed decision-making.

In particular, it focuses on interval-valued representations as a unifying basis for survey-based construct measurement in real-world applications, spanning uncertainty elicitation, reliability assessment, summarisation, and presentation, with the following objectives:

Obj 1: **Motivation and Application Context.** Articulate the necessity of communicating uncertainty in customer responses, with service evaluation adopted as the primary application context for this broader research aim, and establish its links to the existing literature in the field.

Obj 2: **Empirical Data Collection.** Collect first-hand empirical survey data in service evaluation using both single-valued scales and interval-valued scales, to

(a) Enable comparison of responses with and without uncertainty elicitation in subsequent analysis.

(b) Make the dataset publicly available to support further research.

Obj 3: **Reliability Assessment.** Develop and evaluate methods for assessing the internal consistency of interval-valued data, to verify they provide reliable measurement of the intended construct before proceeding to summarisation and presentation.

Obj 4: **Uncertainty Summarisation.** Develop and evaluate methods for uncertainty summarisation, transforming nuanced but complex representations, such as fuzzy sets modelled from interval-valued data, into more accessible interval representations, while retaining key uncertainty information. Specifically:

(a) Define what key information should be retained by establishing desirable properties for such summarisation.

(b) Develop methods that satisfy these properties.

(c) Evaluate whether human interpretation of the summarised representations aligns with the proposed desirable properties, and use the evidence to support or revise the design.

Obj 5: **Decision-Support Presentation.** Extend conventional decision-support methods to enable interval-valued uncertainty presentation, and demonstrate the added insights gained from these extensions in a real-world application, using the data collected in Objective 2.

Obj 6: **Workflow Integration and Software Support.** Integrate the methods developed in Objectives 3–5 into an end-to-end interval-valued survey workflow, and develop an integrated open-source software toolkit to support the latter, to facilitate wider adoption by researchers and practitioners across different fields.

1.6 Thesis Structure and Contributions

While the objectives follow an elicitation-to-decision workflow, the thesis focuses on method development (Chapters 3 and 4), followed by data collection and method extension in a real-world application (Chapter 5), and integrated workflow demonstration with software support (Chapter 6). Specifically, the remainder of the thesis is organised as follows:

Chapter 2 reviews the survey response formats and information representations used throughout the thesis (numbers, intervals, and fuzzy sets), together with their mathematical properties and the state of the art in the corresponding analytical methods. It also reviews survey response scales and introduces

service evaluation as the application domain of this research, including the inherent characteristics of services, the key constructs in this field, and the survey-to-decision workflow together with the methods used at each stage.

Chapter 3 addresses Objective 3 by developing and evaluating methods for assessing the internal consistency of interval-valued data. Existing methods for the reliability assessment of interval-valued data assume that each interval can be appropriately analysed through a midpoint–spread decomposition, which may not hold in practice. A Hausdorff-based extension of Cronbach's α is proposed, which treats each interval as a geometric whole, i.e., without imposing such a decomposition. Its mathematical properties are demonstrated, and its behaviour is examined and compared with existing θ-based approaches through simulation studies.

Chapter 4 addresses Objective 4 by developing and evaluating methods for summarising fuzzy sets as representative intervals. Fuzzy sets capture nuance in uncertainty but are difficult to interpret efficiently in practical communication; conventional defuzzification reduces a fuzzy set to a single value, which is highly efficient but discards the uncertainty that motivated its capture. Desirable properties for the interval summarisation of non-convex fuzzy sets are defined, and novel methods are developed to satisfy these properties. The methods are evaluated against desirable properties using synthetic data, and a mixed-methods user study provides empirical evidence that human interpretation of the summarised representations aligns with the proposed properties.

Chapter 5 addresses Objectives 1 and 2 by collecting comparable interval-valued and single-valued responses through a UK rail passenger survey. This chapter also addresses Objective 5 by extending conventional decision-support methods in service quality research—Gap Analysis, Importance–Performance

Analysis, and Customer Satisfaction Index—to enable interval-valued uncertainty presentation to decision makers. The added insights from these extensions are demonstrated through their application to the collected survey responses.

Chapter 6 addresses Objective 6 by presenting ISurvey, an integrated open-source R toolkit that implements the methods reviewed and developed in this thesis, and demonstrating a complete interval-valued survey workflow using the rail survey data collected in Chapter 5.

Chapter 7 concludes the thesis, summarising contributions, discussing limitations, and outlining directions for future work.

The next chapter reviews the conceptual and methodological background for the remainder of the thesis, providing the basis for the methods and applications developed in later chapters. It also introduces the abbreviations and notations used throughout, which are summarised in the List of Abbreviations and the List of Notations, and re-introduced in full at their first appearance in each subsequent chapter.

Chapter 2

Background

This chapter reviews the information representations, analytical foundations of interval- and fuzzy-set-based methods, survey response formats, and application context that underpin the method development and real-world application in the remainder of the thesis.

Section 2.1 introduces the three information representations used throughout the thesis—numbers, intervals, and fuzzy sets—together with their key definitions and properties. Section 2.2 reviews the main transformations between these representations, with particular attention to interval-valued defuzzification as a bridge from fuzzy sets to intervals. Section 2.3 presents the core analytical framework for interval-valued data, including interval arithmetic, aggregation, and distance-based Fréchet statistics. Section 2.4 reviews the survey response scales relevant to this thesis, including conventional Likert scales and interval-valued scales. Section 2.5 introduces the service research context, including key constructs, conventional evaluation methods, and the UK rail service used as the application setting. Section 2.6 concludes the chapter with a brief summary of the main points.

2.1 Representations of Information

This section introduces three mathematical representations—numbers, intervals, and fuzzy sets—together with their definitions and key mathematical properties.

2.1.1 Numbers

A real number is an element $x \in \mathbb{R}$, where $\mathbb{R}$ denotes the set of all real numbers. In this thesis, real numbers provide the mathematical representation for single-valued data, where each datum is represented by a single element of $\mathbb{R}$. A single-valued dataset therefore consists of n observed or measured real numbers, denoted $\{x_1, \ldots, x_n\}$.

2.1.2 Intervals

Let $\mathcal{K}_c(\mathbb{R})$ denote the collection of all nonempty, compact, convex subsets of $\mathbb{R}$.

An interval $\overline{A} \in \mathcal{K}_c(\mathbb{R})$ can be characterised by its endpoints as

$$\overline{A} = [A^-, A^+], \tag{2.1}$$

where A^- and A^+ denote its lower and upper bounds, with $A^- \leqslant A^+$. The width of $\overline{A}$ is $|\overline{A}| = A^+ - A^-$. When $A^- = A^+$, the interval reduces to a single number (a degenerate interval). Alternatively, an interval can be characterised by its midpoint and spread [55]:

$$m(\overline{A}) = \frac{A^- + A^+}{2}, \quad s(\overline{A}) = \frac{A^+ - A^-}{2}. \tag{2.2}$$

In this thesis, intervals in $\mathcal{K}_c(\mathbb{R})$ are referred to as continuous intervals.[1]

A discontinuous interval (also known as a multi-segment interval) is defined as the union of a finite number of disjoint continuous intervals, which are referred to as its sub-intervals. Let $\bar{\overline{A}}$ be a discontinuous interval with $q(\bar{\overline{A}})$ (simplified to q in mathematical equations for brevity), continuous sub-intervals $\overline{A}_1, \ldots, \overline{A}_q$ such that [50]

$$\bar{\overline{A}} = \bigcup_{i=1}^{q} \overline{A}_i, \tag{2.3}$$

with $\overline{A}_i < \overline{A}_{i+1}$ (i.e., $A_i^+ < A_{i+1}^-$). The width of $\bar{\overline{A}}$ is $|\bar{\overline{A}}| = \sum_{i=1}^{q} |\overline{A}_i|$. When $q(\bar{\overline{A}}) = 1$, $\bar{\overline{A}}$ reduces to a continuous interval, i.e., a continuous interval is a special case of a discontinuous interval.

In this thesis, intervals provide the mathematical representation for interval-valued data, where each datum is either a continuous or a discontinuous interval. An interval-valued dataset therefore consists of n observed or measured intervals, denoted $\{\overline{A}_1, \ldots, \overline{A}_n\}$ when all data are continuous, or more generally $\{\bar{\overline{A}}_1, \ldots, \bar{\overline{A}}_n\}$ when discontinuous intervals may also be present.

2.1.3 Fuzzy Sets

A type-1 fuzzy set A on a universe of discourse X is defined by its membership function $\mu_A : X \to [0, 1]$, which assigns to each element of X a degree of membership in the set [20, 56–60]. There are also Type-2 fuzzy sets, in which the membership values are themselves fuzzy sets, including general type-2 and interval type-2 fuzzy sets [23].

Within the scope of this thesis, the complexity of type-2 fuzzy sets can pose challenges for human-facing communication; therefore, unless otherwise specified, fuzzy set refers to a type-1 fuzzy set defined on $X = \mathbb{R}$ with bounded

[1] Uncertain intervals, in which each endpoint is itself an interval [44], introduce additional complexity and are therefore outside the scope of this thesis.

support.

Properties of Fuzzy Sets

For a fuzzy set A, the following properties are commonly used [20, 56]:

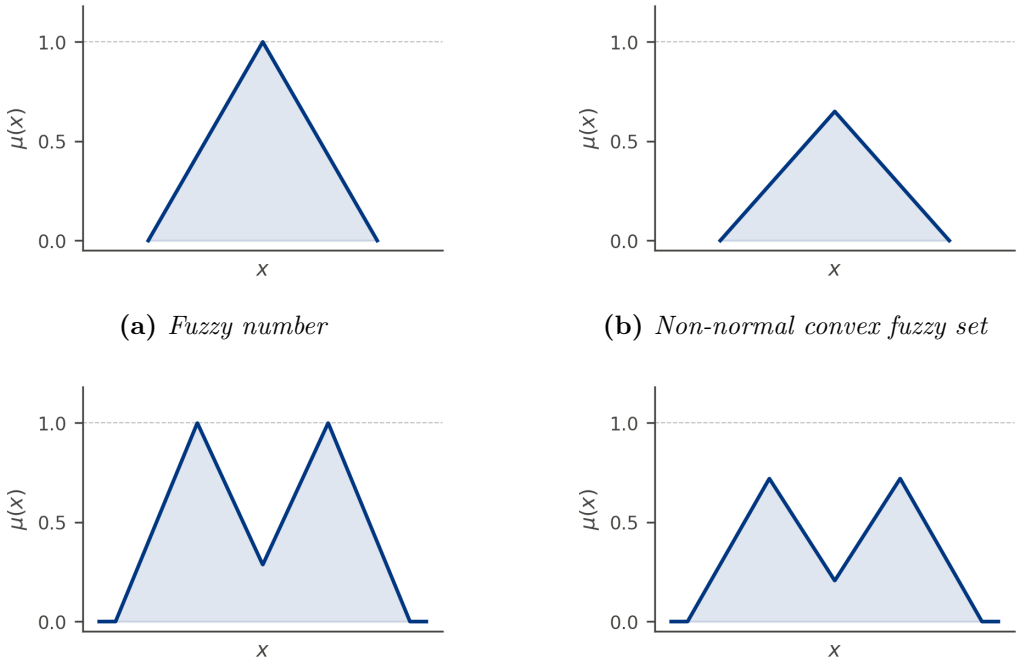
- **Height**: The height of A, denoted $h(A)$ (simplified to h in mathematical equations for brevity), is the supremum of the membership values over all elements in X: $h(A) = \sup_{x \in X} \mu_A(x)$.
- **Core**: The core of A, core(A), contains elements that have full membership in the set: core$(A) = \{x \in X \mid \mu_A(x) = 1\}$.
- **Support**: The support of A, supp(A), includes all elements of X with a non-zero membership: supp$(A) = \{x \in X \mid \mu_A(x) > 0\}$.
- **Normality**: A is normal if its core is non-empty, i.e., there is at least one element x in X such that $\mu_A(x) = 1$.
- **Convexity**: A is convex if and only if $\mu_A(\lambda x_1 + (1 - \lambda)x_2) \geqslant \min(\mu_A(x_1), \mu_A(x_2))$ for any $x_1, x_2 \in X$ and $\lambda \in [0, 1]$ [20].
- **Modality**: A is unimodal if its membership function has a single peak (local maximum), or multimodal if it has multiple peaks. If A has $p(A)$ (simplified to p in mathematical equations for brevity) peaks, it has $p(A) - 1$ troughs (local minima between peaks).

Convex Fuzzy Sets and Fuzzy Numbers

A convex fuzzy set is a fuzzy set satisfying the convexity property. A fuzzy number is a special case of a convex fuzzy set that is also normal and upper semi-continuous with bounded support [56, 61]. Examples of a fuzzy number and a convex fuzzy set are shown in Figs. 2.1a and 2.1b respectively.

Non-convex Fuzzy Sets

A non-convex fuzzy set does not satisfy the convexity property and often exhibits multiple peaks in its membership function. Examples of normal and non-normal non-convex fuzzy sets are shown in Figs. 2.1c and 2.1d respectively.



(a) *Fuzzy number*

(b) *Non-normal convex fuzzy set*

(c) *Normal non-convex fuzzy set*

(d) *Non-normal non-convex fuzzy set*

Figure 2.1: *Examples of fuzzy sets.*

2.2 Transformations Between Representations

This section reviews the pairwise transformations between the three representations—numbers, intervals, and fuzzy sets.

In general, transformations from simpler representations to more complex ones (e.g., from numbers to intervals or fuzzy sets) are often used to express and model uncertainty, whereas those in the other direction are typically used to facilitate efficient computation, ranking, and easier interpretation.

2.2.1 Numbers and Fuzzy Sets

Fuzzification

Fuzzification can be used to transform a number (or a set of numbers) into a fuzzy set. It is commonly used to model uncertainty in linguistic terms and input uncertainty in a fuzzy inference system. For example, a response 'low' on a scale of 'very low' to 'very high' can be fuzzified into a fuzzy set represented by a triangular membership function $(1, 2, 3)$.

Defuzzification

Defuzzification maps a fuzzy set to a number. It is commonly used to simplify the outputs of fuzzy control systems and to obtain comparable scores for fuzzy ranking [23, 56]. Common methods include the Centre of Gravity (also known as Centroid), Mean of Maxima, and Averaging Level Cuts (ALC). In this thesis, these methods are referred to as single-valued defuzzification methods.

Among these, ALC is one of the most widely used α-cut-based single-valued defuzzification methods [49, 62]. For a fuzzy set A with height $h(A)$, ALC is defined as:

$$ALC(A) = \frac{1}{2h} \int_0^h \left(A^-(\alpha) + A^+(\alpha) \right) d\alpha. \tag{2.4}$$

Because α-cuts decompose a fuzzy set into intervals (Section 2.2.3), α-cut-based methods provide a natural basis for extending defuzzification from single-valued to interval-valued outputs, which is the focus of this thesis (Section 2.2.3).

2.2.2 Numbers and Intervals

Number to Interval

A number is often transformed into an interval by representing its measurement error or tolerance bounds, for example by applying a symmetric margin ϵ to obtain $[x - \epsilon, x + \epsilon]$, or by using domain knowledge to assign asymmetric bounds [63].

Interval to Number

An interval $\overline{A} = [A^-, A^+]$ can be reduced to a number by using its midpoint $m(\overline{A})$, lower bound A^-, upper bound A^+, or any value between these bounds, as specified by the chosen method and application context.

2.2.3 Intervals and Fuzzy Sets

Interval to Fuzzy Set

The Interval Approach is one of the main methods for modelling interval-valued data as interval type-2 fuzzy sets [64]. It involves data pre-processing to eliminate outliers and unsuitable data, followed by using the refined data to build unimodal interval type-2 fuzzy set models. The Enhanced Interval Approach extends this by refining data pre-processing for better control of interval widths [65]. The Hao-Mendel Approach improves the process of constructing an interval type-2 fuzzy set by identifying the most common overlap of intervals and building a fuzzy set around it [66].

The Interval Agreement Approach (IAA) is designed to model agreement amongst samples and/or sources using Type-1 or Type-2 fuzzy sets, which

can be non-normal and non-convex [44]. IAA-based methods avoid the need for data preprocessing, model assumptions (e.g., Gaussian distribution), or the elimination of outliers [44]. Given an interval-valued dataset $\{\overline{A}_1, \ldots, \overline{A}_n\}$, with $\overline{A}_i = [A_i^-, A_i^+]$, the Type-1 IAA method constructs a fuzzy set A with membership function [44]

$$\mu_A(x) = \frac{1}{n}\sum_{i=1}^{n} \mu_{\overline{A}_i}(x), \quad \text{where } \mu_{\overline{A}_i}(x) = \begin{cases} 1 & \text{if } A_i^- \leqslant x \leqslant A_i^+, \\ 0 & \text{otherwise.} \end{cases}$$

Here, $\mu_A(x)$ reflects the degree of overlap among input intervals at x. The Efficient IAA is an extension of IAA that simplifies the membership function for enhanced computational efficiency, suitable for larger data sets [67].

Fuzzy Set to Interval

The transformation from a fuzzy set to an interval can be broadly categorised into two approaches. The first approach uses a single α-cut of the fuzzy set as the resulting interval. The second approach decomposes the fuzzy set into a set of α-cuts and then aggregates these cuts to obtain a single interval. In this thesis, methods based on the latter approach are referred to as interval-valued defuzzification methods, to distinguish them from single-valued defuzzification methods. Both interval-valued defuzzification and single-valued defuzzification are types of defuzzification.

Alpha-cuts The α-cut of a fuzzy set A at level $\alpha \in (0, 1]$ is defined as the set of all $x \in X$ such that $\mu_A(x) \geqslant \alpha$:

$$A_\alpha = \{x \in X \mid \mu_A(x) \geqslant \alpha\}. \tag{2.5}$$

As a special case, for $\alpha = 0$, the α-cut is defined as[2]

$$A_0 = \text{supp}(A) = \{x \in X \mid \mu_A(x) > 0\}. \tag{2.6}$$

For a convex fuzzy set and $\alpha \in [0, h(A)]$, the non-empty α-cut A_α can be represented as a continuous interval, denoted

$$\overline{A_\alpha} = [A^-(\alpha), A^+(\alpha)]. \tag{2.7}$$

For a non-convex fuzzy set, a non-empty α-cut A_α can instead be represented as a discontinuous interval, denoted

$$\bar{\overline{A}}_\alpha = \bigcup_{i=1}^{q_\alpha} \overline{A}_i^\alpha, \tag{2.8}$$

In this case, there exists at least one $\alpha \in [0, h(A)]$ such that $\bar{\overline{A}}_\alpha$ consists of multiple disjoint continuous sub-intervals, i.e., $q_\alpha > 1$.

Interval-valued Defuzzification **Interval-valued defuzzification for convex fuzzy sets**

In this thesis, interval-valued defuzzification (in text) refers collectively to interval-valued defuzzification methods as a type of defuzzification; IVD (in math mode) refers to the specific method defined below.

To approximate a fuzzy number A by an interval, the resulting interval $\overline{A} = [A^-, A^+]$ should satisfy the following properties [61]:

1. **Support-subset**: The interval should be a subset of the support of A, i.e., $\overline{A} \subseteq \text{supp}(A)$.

[2] In practice, this is often taken as the smallest closed interval containing all x such that $\mu_A(x) > 0$.

2. **Core-inclusion**: The interval should include the core of A, i.e., $\text{core}(A) \subseteq \overline{A}$.

A nearest interval approximation method satisfying these properties is proposed in [61] by finding the interval with the minimum distance to the α-cuts of the fuzzy number. This method is equivalent to the mean value of a fuzzy number [68] in terms of imprecise probabilities, and to the expected interval [69–71] as the Aumann integral of the α-cut mapping [72], and has been referred to as an interval-valued defuzzification method more recently in [72, 73]. Mathematically, this interval-valued defuzzification method for fuzzy numbers, denoted by IVD, is defined as:

$$IVD(A) = \left[\int_0^1 A^-(\alpha)\, d\alpha, \int_0^1 A^+(\alpha)\, d\alpha\right]. \tag{2.9}$$

The method has been generalised, through normalisation, to handle convex but non-normal fuzzy sets. For a convex, non-normal fuzzy set A, $IVD(A)$ is defined as [50]:

$$IVD(A) = \left[\frac{1}{h}\int_0^h A^-(\alpha)\, d\alpha,\ \frac{1}{h}\int_0^h A^+(\alpha)\, d\alpha\right]. \tag{2.10}$$

When $h(A) = 1$ (i.e., A is normal), Eq. (2.10) reduces to Eq. (2.9) as the normal case is a special instance of the generalised form.

This method can be interpreted as taking the arithmetic average of all points on the left and right boundary functions, respectively, of a normalised convex fuzzy set. Later, this was further extended to a weighted version, in which α is used as the weight, so that points with higher membership degrees make a greater contribution to the resulting interval [74].

This method is referred to as Weighted interval-valued defuzzification, denoted

by $WIVD$, which is defined as:

$$WIVD(A) = \left[\frac{\int_0^h \alpha\, A^-(\alpha)\, d\alpha}{\int_0^h \alpha\, d\alpha},\ \frac{\int_0^h \alpha\, A^+(\alpha)\, d\alpha}{\int_0^h \alpha\, d\alpha} \right]. \tag{2.11}$$

In real-world applications, discrete calculations are commonly used as an alternative to integration, especially for less well-defined membership functions and for computational implementation. The discrete IVD can be interpreted as decomposing a convex fuzzy set, whether normal or non-normal, into a set of continuous intervals via its α-cuts, and then aggregating all non-empty α-cuts using the Interval Average (IA, Section 2.3.2).

For a convex fuzzy set A decomposed into a finite set of representative non-empty α-cuts at levels $\{\alpha_j \mid j = 1, \ldots, m\}$, where $0 < \alpha_j \leqslant h(A)$, the discrete IVD of A is given by:

$$IVD(A) = \left[\frac{1}{m} \sum_{j=1}^{m} A^-(\alpha_j),\ \frac{1}{m} \sum_{j=1}^{m} A^+(\alpha_j) \right]. \tag{2.12}$$

Similarly, the $WIVD$ can be expressed as

$$WIVD(A) = \left[\frac{\sum_{j=1}^{m} \alpha_j A^-(\alpha_j)}{\sum_{j=1}^{m} \alpha_j},\ \frac{\sum_{j=1}^{m} \alpha_j A^+(\alpha_j)}{\sum_{j=1}^{m} \alpha_j} \right]. \tag{2.13}$$

When the number of representative α-levels is sufficiently large and these levels are equally spaced, Eq. (2.12) and Eq. (2.13) provide an approximation to Eq. (2.10) and Eq. (2.11), respectively.

Interval-valued defuzzification for non-convex fuzzy sets

Recent work extends interval-valued defuzzification to non-convex fuzzy sets. Anzilli et al. [50] proposed a general framework that encompasses the IVD method and can generate further variants based on established single-valued

defuzzification methods [62, 75–79], such as centroid and ALC. The framework addresses non-normality through normalisation, and addresses non-convexity by first horizontally aggregating the multiple sub-intervals arising from non-convexity into a single continuous interval at each α-level using a weight parameter $v_i(\alpha_j)$. The resulting intervals across α-levels are then aggregated vertically using a second weight parameter $f(\alpha_j)$, analogously to the aggregation in IVD.

For $\overline{A} = [A^-, A^+]$, the bounds are given by:

$$\begin{aligned} A^- &= \sum_{j=1}^{m} \left[f(\alpha_j) \sum_{i=1}^{q_{\alpha_j}} A_i^-(\alpha_j)\, v_i^-(\alpha_j) \right], \\ A^+ &= \sum_{j=1}^{m} \left[f(\alpha_j) \sum_{i=1}^{q_{\alpha_j}} A_i^+(\alpha_j)\, v_i^+(\alpha_j) \right]. \end{aligned} \tag{2.14}$$

Different parameter choices for $v_i^-(\alpha_j)$, $v_i^+(\alpha_j)$ and $f(\alpha_j)$ yield different variants of the framework [80]. Table 2.1 lists representative variants derived from established single-valued defuzzification methods, where q_{α_j} denotes the number of disjoint sub-intervals at level α_j, $|\overline{A}_i^{\alpha_j}|$ denotes the length of the i-th sub-interval, and $|\overline{A}_{\alpha_j}|$ denotes their total length.

Table 2.1: Representative parameter choices for the interval-valued defuzzification framework.

single-valued defuzzification	$v_i^-(\alpha_j),\ v_i^+(\alpha_j)$	$f(\alpha_j)$	**Source**
Centroid	$\dfrac{1}{q_{\alpha_j}}$	$\dfrac{1}{m}$	[75]
ALC	$\dfrac{\lvert\overline{A}_i^{\alpha_j}\rvert}{\lvert\overline{A}_{\alpha_j}\rvert}$	$\dfrac{1}{m}$	[62]
Area compensation	$\dfrac{\lvert\overline{A}_i^{\alpha_j}\rvert}{\lvert\overline{A}_{\alpha_j}\rvert}$	$\dfrac{\alpha_j}{\sum_{l=1}^{m}\alpha_l}$	[78]

2.3 Interval-valued Data Analysis

2.3.1 Interval Arithmetic

Interval arithmetic is a special case of set arithmetic defined on intervals of the real line.

For two continuous intervals $\overline{A} = [A^-, A^+]$ and $\overline{B} = [B^-, B^+]$, the formulas for the four basic arithmetic operations are as follows, where interval addition is defined by the Minkowski sum:

$$\begin{aligned}\overline{A} + \overline{B} &= [A^-, A^+] + [B^-, B^+] = \{a + b : a \in \overline{A},\, b \in \overline{B}\} \\ &= [A^- + B^-,\, A^+ + B^+]. \end{aligned} \tag{2.15}$$

$$\overline{A} - \overline{B} = [A^-, A^+] - [B^-, B^+] = [A^- - B^+,\, A^+ - B^-], \tag{2.16}$$

$$\begin{aligned}\overline{A} \times \overline{B} &= [A^-, A^+] \times [B^-, B^+] \\ &= [\min(A^-B^-,\, A^-B^+,\, A^+B^-,\, A^+B^+), \\ &\quad \max(A^-B^-,\, A^-B^+,\, A^+B^-,\, A^+B^+)], \end{aligned} \tag{2.17}$$

$$\begin{aligned}\overline{A} \div \overline{B} &= [A^-, A^+] \div [B^-, B^+] = [A^-, A^+] \times [1/B^+,\, 1/B^-], \\ &\quad \text{so long as } 0 \notin \overline{B}. \end{aligned} \tag{2.18}$$

The results of these interval operations are guaranteed to enclose all the values that could possibly be results of the calculations [39, 81, 82].

2.3.2 Interval Central Tendency

Interval Average

The interval average (IA) of n intervals $\overline{A}_1, \ldots, \overline{A}_n$ is defined as:

$$\mathrm{IA} = [y^-, y^+] = \left[\frac{1}{n} \sum_{i=1}^{n} A_i^-,\ \frac{1}{n} \sum_{i=1}^{n} A_i^+ \right]. \tag{2.19}$$

That is, the IA is obtained by taking the arithmetic mean of the lower and upper bounds separately.

Interval Weighted Average

The interval weighted average generalises the IA by incorporating weights. For single-valued weights $w_1, \ldots, w_n$ with $w_i \geqslant 0$, the Interval Weighted Average (IWA) of $\overline{A}_1, \ldots, \overline{A}_n$ is

$$\mathrm{IWA} = [y^-, y^+] = \left[\frac{\sum_{i=1}^{n} w_i\, A_i^-}{\sum_{i=1}^{n} w_i},\ \frac{\sum_{i=1}^{n} w_i\, A_i^+}{\sum_{i=1}^{n} w_i} \right]. \tag{2.20}$$

When the weights are themselves intervals $\overline{w}_i = [w_i^-, w_i^+]$ with $w_i^- \geqslant 0$, the IWA becomes an optimisation problem that can be solved using the Karnik–Mendel algorithms [83]:

$$\mathrm{IWA} = [y^-, y^+] = \left[\min_{w_i \in [w_i^-, w_i^+]} \frac{\sum_{i=1}^{n} A_i^- w_i}{\sum_{i=1}^{n} w_i},\ \max_{w_i \in [w_i^-, w_i^+]} \frac{\sum_{i=1}^{n} A_i^+ w_i}{\sum_{i=1}^{n} w_i} \right]. \tag{2.21}$$

Here, A_i^- and A_i^+ appear in the lower and upper bounds respectively because the weighted average is monotone in the aggregated values; the minimum (maximum) is therefore attained when each A_i is fixed at its lower (upper) bound, reducing the problem to optimisation over the weights alone.

2.3.3 Interval Distances and the Fréchet Framework

While measures of central tendency are relatively straightforward to define for interval-valued data, measures of dispersion such as variance remain challenging.

For a sample of n single-valued data $X = \{x_i\}_{i=1}^n$, the classical variance formula is

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x})^2,$$

where $\hat{x}$ denotes the sample mean.[3] To extend this definition to interval-valued data, one of the issues is it requires a well-defined difference operator with properties analogous to those in the real-valued case. Such an operator is not generally available in $\mathcal{K}_c(\mathbb{R})$. In particular, intervals in $\mathcal{K}_c(\mathbb{R})$ do not generally admit additive inverses under Minkowski addition, since

$$\overline{A} + (-\overline{A}) = \{0\}$$

if and only if $\overline{A}$ is a degenerate interval. For example, based on Eq. (2.15), $[2,2] + [-2,-2] = [0,0] = \{0\}$, whereas $[2,3] + [-3,-2] = [-1,1] \neq \{0\}$.

To address this, a metric space framework has been widely adopted for intervals analysis over the past decades, for example through the Fréchet framework, where, once a distance is defined, variance can be defined as the average squared distance to a Fréchet mean [46, 48, 84].

[3] In this thesis, $\overline{x}$ is reserved to denote an interval-valued datum. To avoid notational conflict, the arithmetic mean of single-valued data is denoted by $\hat{x}$.

Fréchet Framework

In this thesis, Fréchet quantities are defined on $\mathcal{K}_c(\mathbb{R})$ under a given interval distance d [85]. For a sample of n interval-valued data $\overline{\boldsymbol{X}} = \{\overline{X}_i\}_{i=1}^n$, and a given distance d, the Fréchet mean is the interval minimising the average squared distance to all responses:

$$\overline{X}^* = \arg\min_{\overline{X}\in\mathcal{K}_c(\mathbb{R})} \frac{1}{n}\sum_{i=1}^{n} d^2(\overline{X}_i, \overline{X}), \tag{2.22}$$

where $\overline{X}$ denotes a candidate interval in $\mathcal{K}_c(\mathbb{R})$ over which the optimisation is performed.

Accordingly, the corresponding Fréchet variance is

$$\mathrm{Var}^*(\overline{\boldsymbol{X}}) = \frac{1}{n}\sum_{i=1}^{n} d^2(\overline{X}_i, \overline{X}^*). \tag{2.23}$$

Below, the two interval distances used in this thesis are reviewed.

Theta-distance

For intervals $\overline{A}$ and $\overline{B}$, d_θ is defined as [55]

$$d_\theta(\overline{A}, \overline{B}) = \sqrt{(m(\overline{A}) - m(\overline{B}))^2 + \theta(s(\overline{A}) - s(\overline{B}))^2}, \tag{2.24}$$

where $\theta > 0$ controls the relative contribution of location (midpoint) and imprecision (spread). The θ-distance is defined in terms of midpoint and spread differences, which allows closed-form expressions for the corresponding Fréchet mean, variance, and covariance.

Under d_θ, the Fréchet mean is denoted by $\overline{X}_\theta^*$ and coincides with the arithmetic

(Aumann) mean of the interval bounds, as given in Eq. (2.19). The Fréchet variance under d_θ is denoted by $\mathrm{Var}^*_\theta(\overline{\boldsymbol{X}})$ and defined as

$$\mathrm{Var}^*_\theta(\overline{\boldsymbol{X}}) = \frac{1}{n}\sum_{i=1}^{n}(m(\overline{X}_i) - \hat{m})^2 + \theta \cdot \frac{1}{n}\sum_{i=1}^{n}(s(\overline{X}_i) - \hat{s})^2, \tag{2.25}$$

where $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} m(\overline{X}_i)$ and $\hat{s} = \frac{1}{n}\sum_{i=1}^{n} s(\overline{X}_i)$ are the sample means of the interval midpoints and spreads, respectively.

For two interval-valued variables $\overline{\boldsymbol{X}}$ and $\overline{\boldsymbol{Y}}$, the corresponding θ-covariance is defined as

$$\begin{aligned}\mathrm{Cov}^*_\theta(\overline{\boldsymbol{X}}, \overline{\boldsymbol{Y}}) = &\frac{1}{n}\sum_{i=1}^{n}(m(\overline{X}_i) - \hat{m}_X)(m(\overline{Y}_i) - \hat{m}_Y) \\ &+ \theta \cdot \frac{1}{n}\sum_{i=1}^{n}(s(\overline{X}_i) - \hat{s}_X)(s(\overline{Y}_i) - \hat{s}_Y),\end{aligned} \tag{2.26}$$

where $\hat{m}_X$ and $\hat{s}_X$ are the sample means of the midpoints and spreads of $\overline{\boldsymbol{X}}$, and $\hat{m}_Y$ and $\hat{s}_Y$ are those of $\overline{\boldsymbol{Y}}$. Although the corresponding θ-correlation is then naturally obtained by normalising the θ-covariance, its practical use has not yet been evaluated in the current literature, unlike arithmetic correlation approaches.

Hausdorff Distance

For intervals $\overline{A}$ and $\overline{B}$, d_H is defined as

$$d_H(\overline{A}, \overline{B}) = \max\left\{\sup_{a\in\overline{A}}\inf_{b\in\overline{B}}|a-b|,\ \sup_{b\in\overline{B}}\inf_{a\in\overline{A}}|a-b|\right\}. \tag{2.27}$$

When $\overline{A} = \overline{A}$ and $\overline{B} = \overline{B}$, this reduces to

$$d_H(\overline{A}, \overline{B}) = \max\left\{|A^- - B^-|, |A^+ - B^+|\right\}. \tag{2.28}$$

The Hausdorff distance treats each interval as a geometric object, without decomposing it into midpoint and spread components, and extends naturally to discontinuous (multi-segment) intervals [48]. However, the response data considered in this thesis consist only of continuous intervals. Accordingly, references to the Hausdorff distance in this thesis refer to the distance defined in Eq. (2.28).

Under d_H, the Fréchet mean over $\mathcal{K}_c(\mathbb{R})$ is not necessarily unique. Kang et al. [48] prove existence of a Hausdorff Fréchet mean and establish that uniqueness holds when either the midpoint or the width of the solution is fixed. In this thesis, following [48], the interval width is fixed to the sample mean width $\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_i|$ and only the midpoint is optimised. The resulting Hausdorff Fréchet mean and variance are therefore defined as

$$\overline{X}_H^* = \arg \min_{\substack{\overline{X}\in\mathcal{K}_c(\mathbb{R}) \\ |\overline{X}|=\frac{1}{n}\sum_{i=1}^{n}|\overline{X}_i|}} \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{X}_i, \overline{X}), \qquad \mathrm{Var}_H^*(\overline{\boldsymbol{X}}) = \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{X}_i, \overline{X}_H^*). \tag{2.29}$$

Building on this, Kang et al. [48] further define a Hausdorff covariance and correlation for pairs of random intervals, treating each interval as a geometric object in the Hausdorff metric space without imposing algebraic structure or distributional assumptions. This makes the Hausdorff framework applicable to both continuous and discontinuous interval-valued data, where midpoint–spread decomposition methods are not applicable [48]. As this thesis does not further develop the Hausdorff correlation, and its formulation is relatively complex, it is not discussed further here.

2.4 Survey Rating Scales

Surveys are widely used to elicit information from human participants, typically involving multiple items that collectively capture an underlying construct. For example, to measure the construct of trust, a survey might ask participants to rate concise items such as "reliable", "dependable", and "honest". The responses to these specific items are then formally captured through structured rating scales.

A rating scale presents respondents with a continuum or set of ordered response options, from which they select (or specify) the value that best represents their assessment of a given item [11]. The choice of rating scale has direct implications for both the richness of information captured per item and the statistical methods applicable to the resulting data [11, 45]. Likert scales are among the most widely used rating scales; other well-known alternatives include visual analogue scales, fuzzy linguistic scales, fuzzy rating scales, and interval-valued scales [45].

2.4.1 Single-valued Scales

Single-valued scales elicit a single point on the scale per item. Common single-valued scales include Likert scales, which use discrete ordinal categories (e.g., five labelled points from "Very Dissatisfied" to "Very Satisfied"), and visual analogue scales, a continuous scale on which respondents mark a single point along a line anchored at both ends [45]. Examples of a Likert scale and a visual analogue scale are shown in Figs. 2.2 and 2.3 respectively. Single-valued scales are efficient to administer and well supported by established statistical methods. However, they impose a forced precision—respondents must commit to a single category, number, or point even when their true evaluation is

uncertain or spans a range [21]. The choice of the number of scale points can itself affect measurement properties such as reliability [45].

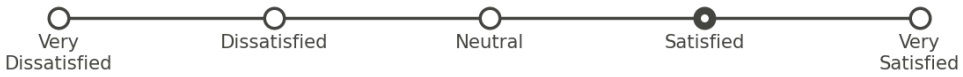


Figure 2.2: *Example of a Likert scale.*

Figure 2.3: *Example of a visual analogue scale.*

2.4.2 Fuzzy Linguistic Scales

Fuzzy linguistic scales use predefined linguistic labels, each associated with a fuzzy membership function (Section 2.1.3), as response options [24, 51]. A typical fuzzy linguistic scale maps each label (e.g., "Low", "Medium", "High") to a triangular or trapezoidal membership function over a numeric domain, and respondents select the label that best describes their perception. For example, the five Likert categories in Fig. 2.2 can each be associated with a triangular fuzzy number, as illustrated in Table 2.2. By associating labels with fuzzy membership functions rather than single points, fuzzy linguistic scales introduce uncertainty representation into the response format. However, fuzzy linguistic scales are constrained by their fixed label sets: the number and shape of the predefined membership functions limit the granularity with which respondents can express their perceptions, and individual differences in how respondents interpret linguistic labels are not captured [45].

Table 2.2: Mapping of linguistic labels to triangular fuzzy numbers on $[1, 5]$.

Label	Very Dissatisfied	Dissatisfied	Neutral	Satisfied	Very Satisfied
Fuzzy number	$(1, 1, 2)$	$(1, 2, 3)$	$(2, 3, 4)$	$(3, 4, 5)$	$(4, 5, 5)$

2.4.3 Fuzzy Rating Scales

Fuzzy rating scales allow respondents to directly specify fuzzy membership functions, for example by drawing or adjusting the shape of a membership function on a graphical interface [12], as illustrated in Fig. 2.4. This addresses the fixed-label limitation of fuzzy linguistic scales, as respondents can indicate both the most representative value and the degree of uncertainty surrounding it. Fuzzy rating scales have been shown to yield higher reliability (as measured by Cronbach's alpha) than Likert scales, visual analogue scales, fuzzy linguistic scales, and interval-valued scales in simulation studies [45]. However, fuzzy rating scales require respondent pre-training to understand and manipulate fuzzy membership functions, and the resulting data are more complex to administer, collect, and analyse at scale. These practical constraints have limited the adoption of fuzzy rating scales outside of controlled experimental settings.

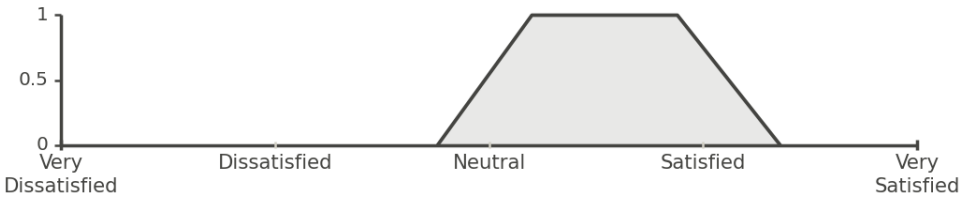


Figure 2.4: *Example of a fuzzy rating scale.*

2.4.4 Interval-valued Scales

Interval-valued scales were proposed as an alternative to single-point response formats [21, 44]. Rather than marking a single point, respondents specify an interval on a continuous scale, for example by circling an area, as illustrated in Fig. 2.5. Crucially, interval-valued scales do not restrict respondents to intervals of a pre-determined size, affording full freedom in expression and enabling the effective capture of both intra- and inter-participant uncertainty

[21].

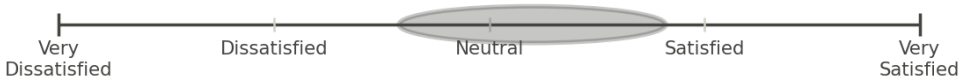


Figure 2.5: *Example of an interval-valued scale.*

Interval-valued scales address the key limitation of fuzzy rating scales—practical complexity—while retaining the ability to capture uncertainty. Specifying an interval requires no pre-training or understanding of fuzzy membership functions; it is a natural and intuitive action that respondents can perform without instruction [21]. At the same time, interval-valued responses capture richer information than single-valued scales: the interval width reflects the respondent's uncertainty, while the interval location reflects their central evaluation.

The validity and usability of interval-valued scales have been established through a series of studies. Interval-valued responses have been shown to capture richer information than single-valued equivalents while maintaining comparable validity [21]. The usability and efficacy of interval-valued scales have been demonstrated across multiple domains, including cybersecurity, consumer research, healthcare, and environmental management [25, 27–31]. Interval-valued scales can be administered using platforms such as DECSYS, a free and open-source system that natively supports interval-valued responses alongside single-valued, single-choice, and free-text response formats [27].

In particular, an ellipse-based response mode was developed [27] in which respondents draw a single continuous ellipse to specify their interval, integrating lower and upper bound specification into one natural interaction with minimal pre-training. Throughout this thesis, interval-valued scales refers specifically to this ellipse-based response mode.

2.5 Service Research

2.5.1 Characteristics

Services differ fundamentally from physical goods in ways that make their quality inherently challenging to evaluate. The services marketing literature commonly identifies three characteristics that differentiate services from goods: intangibility, heterogeneity, and inseparability [86, 87].

Intangibility refers to the fact that services are performances rather than objects. Because of this, precise manufacturing specifications concerning uniform quality can rarely be set, and most services cannot be counted, measured, inventoried, tested, or verified in advance of sale to assure quality. As a result, the firm may find it difficult to understand how consumers perceive their services and evaluate service quality [86, 88, 89].

Heterogeneity refers to the variability inherent in service delivery. Services, especially those with a high labour content, vary in performance from producer to producer, from customer to customer, and from day to day. Consistency of behaviour from service personnel—that is, uniform quality—is difficult to assure, because what the firm intends to deliver may be entirely different from what the consumer receives [86].

Inseparability refers to the fact that production and consumption of many services are inseparable. As a consequence, quality in services is not engineered at the manufacturing plant and then delivered intact to the consumer; rather, quality occurs during service delivery, usually in an interaction between the client and the contact person from the service firm. The service firm may also have less managerial control over quality in services where consumer participation is intense, because the consumer's input becomes critical to the

quality of service performance [86, 87].

2.5.2 Survey and Key Constructs

Given that service evaluation relies heavily on customers' perceptions, judgements, and evaluations, survey-based measures are widely used in both research and in practice. The key constructs commonly measured in this way are as follows.

Expectations and Perceptions Service quality has been conceptualised as the gap between what customers expect from a service and how they perceive the service they receive [86]. Both constructs are typically collected through multi-item survey instruments, most notably SERVQUAL, in which respondents report their expectations and perceptions across multiple service attributes on corresponding scales [90].

Performance Cronin and Taylor [91] argue that a performance-based measure of service quality may be an improved means of measuring the service quality construct, and proposed a performance-only index (SERVPERF). In this approach, service quality is collected as post-experience performance ratings across the same set of attributes, without a separate expectations component [91].

Importance Martilla and James [34] introduced a technique that evaluates the importance and performance of individual attributes to inform the development of effective marketing programmes. In this framework, respondents rate both the importance and the performance of specific service attributes, enabling identification of attributes that underperform or overperform relative to their importance [34].

Satisfaction Service quality and customer satisfaction have been identified as key influences in the formation of consumers' purchase intentions in service environments [92]. Cronin and Taylor [91] found that service quality is an antecedent of consumer satisfaction, and that consumer satisfaction has a significant effect on purchase intentions. In service research, satisfaction is typically collected through survey items asking respondents to rate their overall satisfaction with the service experience or with specific service attributes.

2.5.3 Workflow

Survey Design

Two broad approaches to survey design can be identified in service quality research. The first is research-oriented, adopting or adapting an established academic instrument such as SERVQUAL [90], which provides a multi-item scale with documented construct validity and reliability. However, Parasuraman et al. noted that the instrument provides a "basic skeleton" that "when necessary, can be adapted or supplemented to fit the characteristics or specific research needs of a particular organisation" [90]. The second is practitioner-oriented, with the service provider designing a shorter survey around the specific attributes of direct operational interest, using wording accessible to the target respondents. Such instruments lack the extensive psychometric validation of established scales but are more closely aligned with the provider's decision-making needs.

Psychometric Assessment

Before drawing substantive conclusions, survey results must be assessed for validity and reliability. Validity concerns whether the instrument measures

the intended construct. Convergent validity, in particular, concerns whether conceptually related measures exhibit strong associations, typically assessed via correlation coefficients [11, 93]. Correlation is computed from covariance and variance as

$$\rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}. \tag{2.30}$$

Convergent validity assessment is particularly relevant when a new measurement method is introduced, as it establishes whether the new method produces results consistent with an established measure of the same construct [21]. For practitioner-designed surveys with single-item measures for each attribute, factor analysis and multi-item construct validation are not applicable; convergent validity with respect to an established method therefore becomes the primary means of psychometric validation [11].

Reliability concerns the consistency of measurement. Internal consistency—the degree to which items intended to measure the same construct produce similar responses—is most commonly quantified using Cronbach's α [94]. For a survey with k items, Cronbach's α evaluates internal consistency by comparing the sum of item variances to the variance of the total score:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \mathrm{Var}(X_j)}{\mathrm{Var}\left(\sum_{j=1}^{k} X_j\right)}\right). \tag{2.31}$$

Reliability assessment is a standard requirement for survey-based research regardless of whether the instrument is adapted from an existing scale or designed from scratch, as it provides a baseline assurance that the items are measuring a coherent underlying construct [11]. The extension of reliability assessment to interval-valued data is addressed in Chapter 3.

Descriptive Statistics and Visual Summaries

For single-valued data, descriptive statistics—including measures of central tendency (mean, median, mode) and dispersion (variance, standard deviation, range)—and visual summaries such as relative frequency histograms and box plots provide an initial overview of responses.

Evaluation Methods

The evaluation stage is typically the decision-facing step in the survey workflow, translating measurements into actionable insights for stakeholders through prioritisation, benchmarking, and resource allocation [95]. In service and customer research, three widely used diagnostic methods are reviewed below.

Gap analysis. Gap analysis, rooted in the SERVQUAL framework [86], assesses service shortfalls by computing the difference between perceived performance and expectation for each service attribute:

$$\text{Gap}_j = \hat{P}_j - \hat{E}_j, \tag{2.32}$$

where $\hat{E}_j$ and $\hat{P}_j$ denote the mean expectation and mean perception for attribute j, respectively. A negative gap indicates that perceived performance falls short of expectations, with larger negative values signalling greater dissatisfaction. Gap scores can be computed at the individual attribute level or aggregated across dimensions to provide an overall quality score. The method is widely used to identify specific areas of underperformance and to prioritise service improvements [95].

Importance–performance analysis (IPA). IPA [34] maps service attributes on a two-dimensional grid, with mean importance on the y-axis

and mean performance on the x-axis. The overall means of importance and performance define the crosshairs that partition the grid into four quadrants [95]:

$$\hat{c}_P = \frac{1}{k}\sum_{j=1}^{k} \hat{P}_j, \qquad \hat{c}_I = \frac{1}{k}\sum_{j=1}^{k} \hat{I}_j, \tag{2.33}$$

partitioning the plane into **Concentrate Here** ($\hat{I}_j > \hat{c}_I$, $\hat{P}_j < \hat{c}_P$), **Keep Up the Good Work** ($\hat{I}_j > \hat{c}_I$, $\hat{P}_j \geqslant \hat{c}_P$), **Low Priority** ($\hat{I}_j \leqslant \hat{c}_I$, $\hat{P}_j < \hat{c}_P$), and **Possible Overkill** ($\hat{I}_j \leqslant \hat{c}_I$, $\hat{P}_j \geqslant \hat{c}_P$). IPA provides a visual and intuitive framework for resource allocation, and has been applied in transportation research to identify service aspects most in need of improvement [95].

Customer satisfaction index (CSI). CSI is an importance-weighted satisfaction score, commonly used for benchmarking over time or across providers [96]:

$$\text{CSI} = \frac{\sum_{j=1}^{k} \hat{I}_j \cdot \hat{S}_j}{\sum_{j=1}^{k} \hat{I}_j}, \tag{2.34}$$

where $\hat{I}_j$ and $\hat{S}_j$ are the mean importance and mean satisfaction for attribute j, and k is the number of attributes. By weighting satisfaction scores by relative importance, CSI ensures that attributes considered most important by respondents contribute proportionally more to the overall index. CSI is particularly useful for tracking changes in overall service quality over successive survey waves and for comparing performance across service providers or routes [96]. The extension of these diagnostic methods to interval-valued data is addressed in Chapter 5.

2.5.4 Service Research in the UK Rail Industry

Rail service serves as a highly representative application area for service research, since rail evaluation shares many of the same methodological frameworks.

The UK rail network is one of the most heavily used in Europe, with over 1.6 billion passenger journeys per year. The network is in transition from a mixed public–private system to predominantly public ownership, with most passenger services now operated by publicly owned train operating companies under government-controlled contracts, and the remainder scheduled to transfer to public ownership by 2027. This fragmented structure means that service quality varies across operators, routes, and time periods, making systematic evaluation essential for both regulatory oversight and service improvement.

The primary instrument for monitoring passenger satisfaction has been the National Rail Passenger Survey, conducted by Transport Focus, an independent statutory body [97]. The survey, the largest published rail passenger satisfaction survey in the world [98], surveyed more than 50,000 passengers per year across all train operating companies, and its results served as a key performance indicator for most rail franchises, directly informing operator benchmarking, regulatory decisions, and policy development. In 2025, Transport Focus launched the Rail Customer Experience Survey, a continuous rolling survey aiming to collect approximately 10,000 responses per month. This new survey replaced earlier rail passenger surveys, including the National Rail Passenger Survey and the Rail User Survey, further expanding the scale of single-valued-based passenger feedback [99].

The widespread and large-scale use of single-valued surveys in this domain means that passenger satisfaction data shape investment priorities, franchise specifications, and service improvement programmes across the network. Since survey-based evaluation is deeply embedded in the sector, with results directly consequential for both customer experience and future policy, UK rail services are therefore a highly applicable domain for service research.

2.6 Summary and Remarks

This chapter has reviewed the conceptual and methodological background for the remainder of the thesis. It introduced the three information representations used throughout—numbers, intervals, and fuzzy sets—together with the main transformations between them, with particular emphasis on interval-valued defuzzification as a family of methods used to simplify fuzzy sets to intervals. It then reviewed the core analytical methods for interval-valued data, including interval arithmetic, aggregation, distance measures, and the Fréchet framework. The chapter also outlined the survey response formats relevant to uncertainty-aware elicitation, highlighting interval-valued scales as the main response format adopted in this research. Finally, it introduced service evaluation as the application domain for this research, including the key constructs, workflow considerations, evaluation methods, and the UK rail context that motivates the real-world application in later chapters.

Chapter 3

Reliability Assessment: Hausdorff-based Cronbach's Alpha for Interval-valued Data

Reliability assessment is a standard early step in survey-based research, undertaken to confirm that items intended to measure the same construct exhibit sufficient internal consistency before subsequent analysis proceeds. Recent work has extended the widely used reliability measure, Cronbach's α, to interval-valued data by combining the θ-distance with the Fréchet framework. While valuable, the midpoint–spread (mid–spr) decomposition of the θ-distance may pose limitations in both semantic interpretation and computational procedure. This chapter therefore addresses the reliability assessment stage of the interval-valued workflow by extending Cronbach's α within the Fréchet framework under the Hausdorff distance, to address these limitations through a geometric treatment of interval-valued data.

Section 3.1 motivates the use of the Hausdorff distance by identifying the semantic and computational limitations of the existing θ-based formulation. Section 3.2 recalls the interval distances, the Fréchet framework, and the classical formulation of Cronbach's α that the proposed method builds upon.

Section 3.3 defines the proposed α_H and its setting. Section 3.4 demonstrates its basic mathematical behaviour through targeted simulation tests. Section 3.5 compares α_H with α_θ across six controlled interval-valued response patterns through a simulation study. Section 3.6 discusses the limitations and future work, and Section 3.7 provides a summary and remarks.

3.1 Introduction

The measurement of latent constructs is essential in the social sciences, given that many phenomena of interest are abstract and cannot be observed directly [11]. In practice, such a construct is commonly measured through responses to a set of items. Reliability assessment is often a necessary early step after responses have been collected, most commonly using Cronbach's α [11, 46], to confirm sufficient internal consistency among the items, thereby reducing the risk of misleading interpretations and incorrect conclusions.

Originally developed for traditional single-valued data, such as those collected using Likert scales, Cronbach's α has more recently been extended to interval-valued data under the Fréchet framework using the θ-distance [46]. This distance, while mathematically convenient, is based on a mid–spr decomposition, which raises both semantic and computational concerns.

The mid–spr decomposition is often linked to a semantic interpretation in which the midpoint is treated as a measure of location, often taken as the most representative value, and the spread as a measure of imprecision [46], effectively treating an interval as two separable components. This naturally aligns with a disjunctive view of interval-valued data, where the true value is assumed to be an unknown single point within the interval. However, many real-world interval-valued responses are better interpreted conjunctively, where the interval is treated as a set of admissible values that jointly constitute

the respondent's perception [21, 100]. Under this view, the entire range is meaningful in itself, representing a tolerance zone or a flexibility region. In addition, even when a disjunctive view holds for interval-valued responses and a most representative value can be identified, the midpoint may not coincide with that value. For example, a respondent may answer the same item using both a Likert scale and an interval-valued scale, giving "5–very satisfied" in the former case but "4–satisfied" to "5–very satisfied" in the latter. This is especially plausible because a Likert scale does not allow respondents to choose a value between two adjacent response terms.

The mid–spr decomposition also raises concerns in computation, as it handles midpoints and spreads independently, even though they are meant to jointly represent an interval as a whole. For example, the θ-covariance between two interval-valued variables reduces to exactly the covariance between their midpoints if one of them has a constant spread, i.e., the spread of that variable is the same for every observation. Thus, midpoints alone are used to represent the interval-valued variables, despite the fact that the midpoint is not necessarily the most representative value of the interval. This is further compounded in θ-based correlation, where the spread contributes to the denominator but not the numerator, causing the correlation to be systematically attenuated, as illustrated in Fig. 3.1.

While the precise impact of this attenuation on α_θ remains unknown, it is worth noting that correlation-based methods such as convergent validity assessment are commonly applied alongside Cronbach's α on the same dataset as part of broader psychometric evaluations [11]. Using the θ-distance for α while relying on a different metric for correlation would introduce inconsistency in the underlying distance assumptions across the same analysis, potentially leading to conflicting interpretations of the same data.

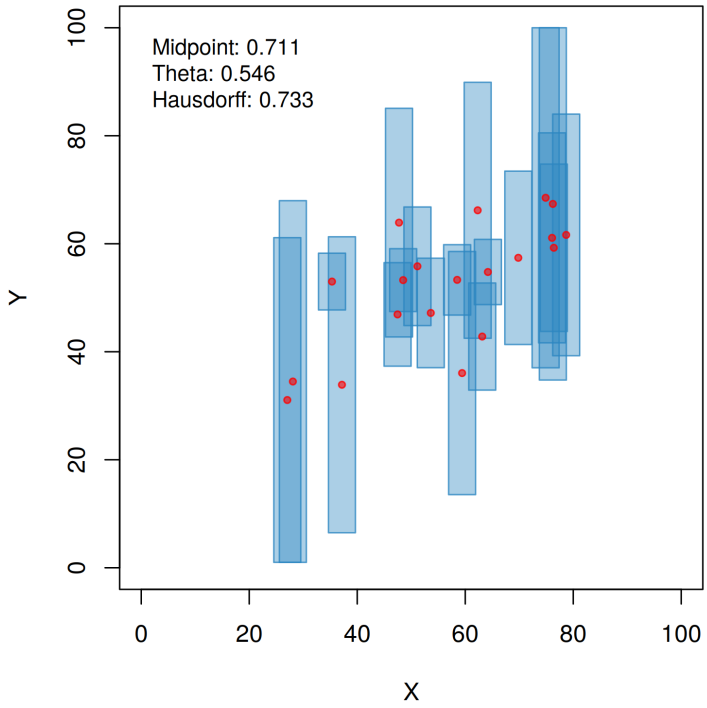


Figure 3.1: *Correlation between two interval-valued variables under midpoint-, θ-, and Hausdorff-based methods.*

The Hausdorff distance offers an alternative metric that addresses these limitations. It treats intervals as bounded sets rather than decomposing them into separable components, aligns naturally with conjunctive interpretations which do not assume a representative value, and retains a geometric interpretation that does not reduce interval-valued responses to single values in the situation described above.

Recently, Kang et al. [48] extended the Fréchet framework to interval-valued data under the Hausdorff distance, defining Hausdorff mean, variance, and covariance. An illustrative example of Hausdorff correlation is shown in Fig. 3.1, in comparison with the θ correlation. However, no Cronbach's α for interval-valued data has yet been developed under the Hausdorff distance.

Accordingly, this chapter makes the following contributions:

1. A Hausdorff-based extension of Cronbach's α for interval-valued data is

proposed, denoted by α_H, using the sample Hausdorff mean and variance defined by Kang et al. [48]. Its mathematical properties, including boundedness, extreme cases, and affine invariance, are demonstrated through simulation.

2. The behaviour of α_H is compared with that of α_θ using a synthetic simulation study that constructs different kinds of interval-valued responses under controlled scenarios reflecting real-world response patterns.

The code for the basic mathematical behaviour tests in Section 3.4 and the simulation study in Section 3.5, including all experimental conditions and parameter settings, is available on GitHub for replication at https://github.com/Carina-YuZhao/.

3.2 Related Work

This section briefly recalls the interval distances d_θ and d_H, the Fréchet mean and variance, and the formulation of Cronbach's α. Their details are reviewed in Sections 2.3.3 and 2.5.3, respectively.

Throughout this chapter, the term interval refers to an element of $\mathcal{K}_c(\mathbb{R})$ (Section 2.1.2).

3.2.1 Interval Distances

For intervals $\overline{A} = [A^-, A^+]$ and $\overline{B} = [B^-, B^+]$, the θ-distance d_θ and the Hausdorff distance d_H are defined as follows.

The θ-distance is defined as [55]

$$d_\theta(\overline{A}, \overline{B}) = \sqrt{\big(m(\overline{A}) - m(\overline{B})\big)^2 + \theta\big(s(\overline{A}) - s(\overline{B})\big)^2}, \tag{3.1}$$

where

$$m(\overline{A}) = \frac{A^- + A^+}{2}, \qquad s(\overline{A}) = \frac{A^+ - A^-}{2},$$

and $\theta > 0$ controls the relative contribution of the midpoint and spread components.

The Hausdorff distance is defined as [48, 101]

$$d_H(\overline{A}, \overline{B}) = \max\left\{|A^- - B^-|, |A^+ - B^+|\right\}. \tag{3.2}$$

3.2.2 Fréchet Framework

For a sample of n interval-valued responses $\overline{\boldsymbol{X}} = \{\overline{X}_i\}_{i=1}^n$, the corresponding sample Fréchet mean and variance under a distance metric d on $\mathcal{K}_c(\mathbb{R})$, for example d_θ or d_H, are defined by

$$\overline{X}^* = \arg\min_{\overline{X} \in \mathcal{K}_c(\mathbb{R})} \frac{1}{n} \sum_{i=1}^n d^2(\overline{X}_i, \overline{X}), \tag{3.3}$$

and

$$\mathrm{Var}^*(\overline{\boldsymbol{X}}) = \frac{1}{n} \sum_{i=1}^n d^2(\overline{X}_i, \overline{X}^*), \tag{3.4}$$

respectively.

Substituting d_θ or d_H into the above definitions gives the corresponding sample Fréchet mean and variance under the θ-distance and the Hausdorff distance, denoted by $(\overline{X}^*_\theta, \mathrm{Var}^*_\theta(\overline{\boldsymbol{X}}))$ and $(\overline{X}^*_H, \mathrm{Var}^*_H(\overline{\boldsymbol{X}}))$, respectively.

It is worth noting that, while the Hausdorff mean may in general be non-unique, Kang et al. [48] have proposed an approach under which uniqueness can be obtained by restricting the optimisation to $\mathcal{K}_c(\mathbb{R})$ and fixing the interval length as the average length of the sample intervals. This chapter follows this approach, and $\overline{X}^*_H$ and $\mathrm{Var}^*_H(\overline{\boldsymbol{X}})$ therefore refer to the resulting sample

Hausdorff Fréchet mean and variance.

3.2.3 Cronbach's α

For a construct consisting of $k \geqslant 2$ items, let X_j denote the single-valued scores on item j in the sample, for $j = 1, \ldots, k$. The sample Cronbach's α [94] is defined by:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \operatorname{Var}(X_j)}{\operatorname{Var}\left(\sum_{j=1}^{k} X_j\right)}\right). \tag{3.5}$$

To extend this coefficient to interval-valued data, Garcia et al. [46] define α_θ by replacing the classical variances in (3.5) with the corresponding Fréchet variances $\operatorname{Var}^*_\theta(\overline{\boldsymbol{X}})$. The mathematical properties of α_θ, including its extremal values and affine invariance, are established in [46].

3.3 Method

3.3.1 Setting and Notation

Let a target construct be measured by $k \geqslant 2$ items administered to $n \geqslant 2$ respondents. For each respondent $i \in \{1, \ldots, n\}$ and item $j \in \{1, \ldots, k\}$, the response is an interval $\overline{X}_{ij}$:

$$\overline{X}_{ij} = [X^-_{ij}, X^+_{ij}]. \tag{3.6}$$

The sample of interval-valued responses for item j is denoted by $\overline{\mathbf{X}}_j = \{\overline{X}_{ij}\}_{i=1}^n$.

The total score for respondent i is defined via the Minkowski sum (Eq. (2.15)):

$$\overline{T}_i \;=\; \sum_{j=1}^{k} \overline{X}_{ij} \;=\; \left[\sum_{j=1}^{k} X_{ij}^{-},\; \sum_{j=1}^{k} X_{ij}^{+}\right]. \tag{3.7}$$

The sample of total scores is denoted by $\overline{\mathbf{T}} = \{\overline{T}_i\}_{i=1}^{n}$.

3.3.2 Definition

For each item sample $\overline{\mathbf{X}}_j = \{\overline{X}_{ij}\}_{i=1}^{n}$, let $\overline{X}_{H,j}^{*}$ denote the sample Hausdorff Fréchet mean obtained by applying the Hausdorff distance in (3.2) to the sample Fréchet mean definition in (3.3). The corresponding sample Hausdorff Fréchet variance, following (3.4), is defined by

$$\mathrm{Var}_H^{*}(\overline{\mathbf{X}}_j) = \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{X}_{ij}, \overline{X}_{H,j}^{*}). \tag{3.8}$$

Likewise, let $\overline{T}_H^{*}$ denote the sample Hausdorff Fréchet mean of the total score sample $\overline{\mathbf{T}} = \{\overline{T}_i\}_{i=1}^{n}$. The corresponding sample Hausdorff Fréchet variance is

$$\mathrm{Var}_H^{*}(\overline{\mathbf{T}}) = \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{T}_i, \overline{T}_H^{*}). \tag{3.9}$$

The sample Hausdorff-based Cronbach's coefficient is then defined by

$$\alpha_H = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \mathrm{Var}_H^{*}(\overline{\mathbf{X}}_j)}{\mathrm{Var}_H^{*}(\overline{\mathbf{T}})}\right), \tag{3.10}$$

provided that $\mathrm{Var}_H^{*}(\overline{\mathbf{T}}) > 0$.

3.4 Basic Mathematical Behaviour

A simulation-based examination was conducted to assess four basic mathematical behaviours of the proposed α_H. Specifically, the tests considered whether α_H equals 1 when all items are identical, whether it approaches 0 when items are generated independently, whether it is invariant under a common affine transformation applied to all items, and whether it reduces to the classical Cronbach's α when all interval-valued responses degenerate to single points. All simulations used $n = 100$ respondents and $k = 5$ items.

3.4.1 $\alpha_H = 1$ for Identical Items

A single interval-valued latent response was generated for each respondent, and this same response was copied across all items. The resulting value was $\alpha_H = 1.000$, matching the expected value exactly and indicating that the coefficient attains 1 when all items are identical.

3.4.2 $\alpha_H \approx 0$ for Independently Generated Items

To simulate uncorrelated interval-valued responses, two candidate bounds were generated independently for each response and then sorted to ensure that the lower bound did not exceed the upper bound. This procedure was repeated 50 times, each time generating a new simulated dataset of n respondents on k items, to examine the distribution of α_H across repeated simulations. Across the 50 repetitions, the mean value of α_H was -0.003, with standard deviation 0.143. The minimum and maximum observed values were -0.335 and 0.264, respectively. These results indicate that, when item responses are simulated without any shared latent structure, the coefficient remains close to 0.

3.4.3 Affine Invariance

A set of correlated interval-valued items was generated as the original data (for later transformation) by first generating one interval for each respondent and then adding independent noise to the lower and upper bounds for each item. A common affine transformation $aX + b$ was then applied to all items, first with $a = 2.5$, $b = -3.7$ and then with $a = -2$, $b = 10$. For $a = -2$, the resulting lower and upper bounds were swapped to maintain the correct ordering. For the original data, the coefficient was $\alpha_H = 0.956$. After both transformations, the value remained $\alpha_H = 0.956$, with a difference of 0.000 in each case. These results indicate that α_H is invariant under a common affine transformation applied to all items.

3.4.4 Reduction to Classical α for Single-valued Data

Single-valued data were generated for all respondents and items. The classical Cronbach's α was computed directly from these responses. The same data were then represented as degenerate intervals and α_H was computed. The classical coefficient was $\alpha = 0.953$, and the Hausdorff-based coefficient was $\alpha_H = 0.953$, with a difference of 0.000. This confirms that α_H reduces exactly to classical α when all responses are single-valued.

Overall, the four simulation tests produced results consistent with the expected basic mathematical behaviour of α_H. Although these findings are in agreement with expectation, further mathematical proofs of these properties remain desirable.

3.5 Simulation Study

In addition to the semantic differences between d_θ and d_H discussed in Section 3.1, this section investigates the practical behavioural differences between α_θ and α_H through a simulation study with controlled interval-valued response patterns. Single-valued responses are first generated as the baseline, as illustrated in Fig. 3.2, and then used to construct the corresponding interval-valued responses.

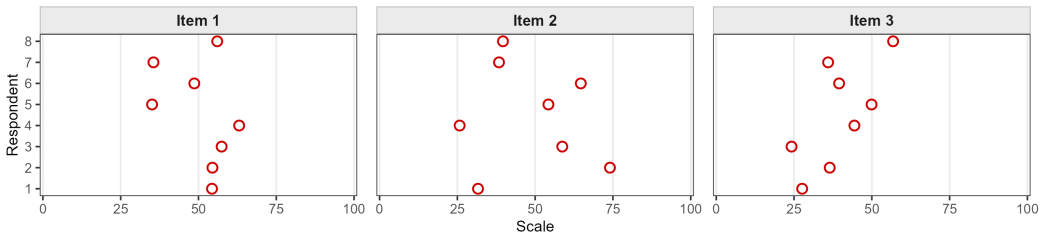


Figure 3.2: *Example single-valued baseline responses ($n = 8$, $k = 3$, high reliability). Each red circle is one respondent's single-valued response for that item.*

3.5.1 Design Principles

To construct interval-valued responses from the baseline single-valued responses, the simulation varies two interval characteristics: width and midpoint position. This is because if interval endpoints are modified directly, changes in width and midpoint may occur simultaneously, making their effects difficult to separate. The simulation therefore controls these two characteristics explicitly and introduces six interval generation modes to examine how α_θ and α_H respond to them.

Specifically, these modes are organised into two groups. The first four modes simulate different width patterns while keeping the interval midpoint aligned with the original single-valued response. The last two modes introduce midpoint displacement relative to the original single-valued response.

1. **Respondent Consistent Width (RCW):** each respondent is assigned a characteristic interval width, which is then applied across all items for that respondent.

2. **Random Width (RW):** interval widths vary independently for each respondent–item pair, with no systematic structure across respondents or items.

3. **Item Consistent Width (ICW):** each item is assigned a characteristic interval width, which is then shared by all respondents for that item.

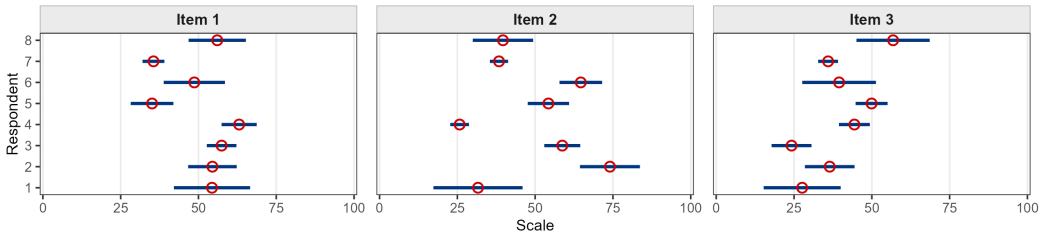
4. **Correlation Width (CW):** interval width is correlated with the corresponding single-valued response value, so that higher single-valued responses produce narrower intervals and lower single-valued responses produce wider intervals.

5. **Asymmetric Midpoint (AM):** each interval is shifted independently relative to the original single-valued response, so that midpoint displacement varies across respondent–item pairs.

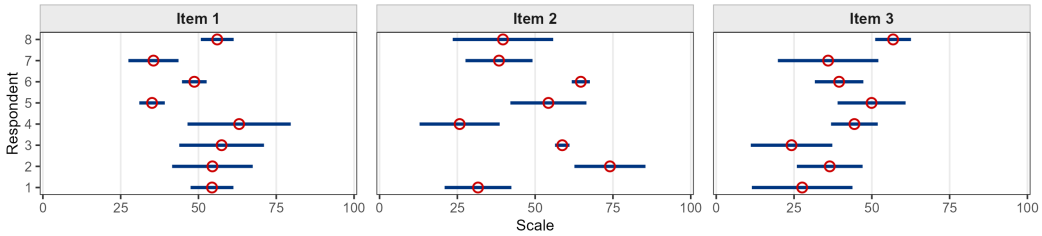
6. **Respondent Asymmetry (RA):** each respondent is assigned a characteristic midpoint shift relative to the original single-valued response, which is then applied consistently across all items for that respondent.

In these two midpoint-shift modes, the interval widths are taken directly from RW, so that comparison with RW isolates the effect of midpoint displacement from that of width structure.

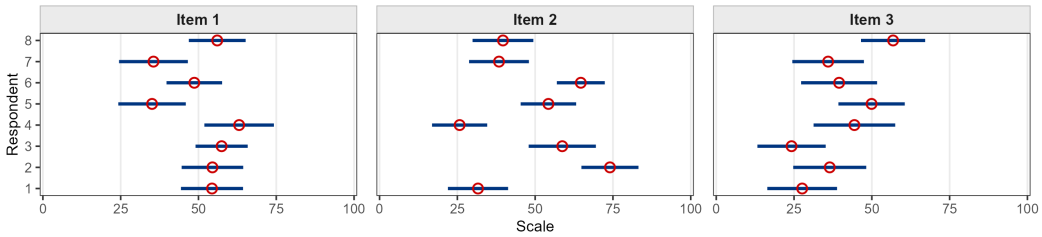
Fig. 3.3 illustrates the six interval generation modes applied to the same single-valued baseline shown in Fig. 3.2. Dark blue segments are the interval-valued intervals; red circles show the original single-valued responses for reference.



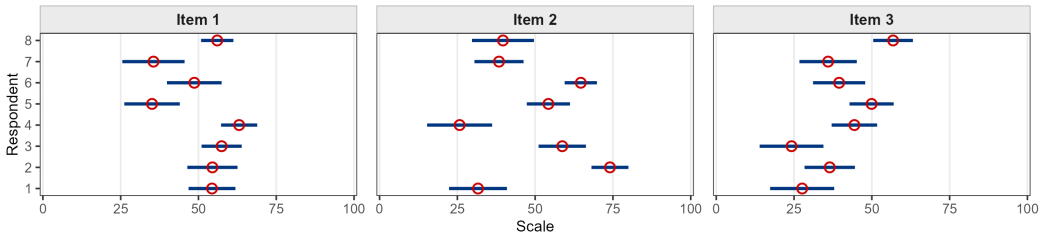
(a) *RCW — Respondent Consistent Width*



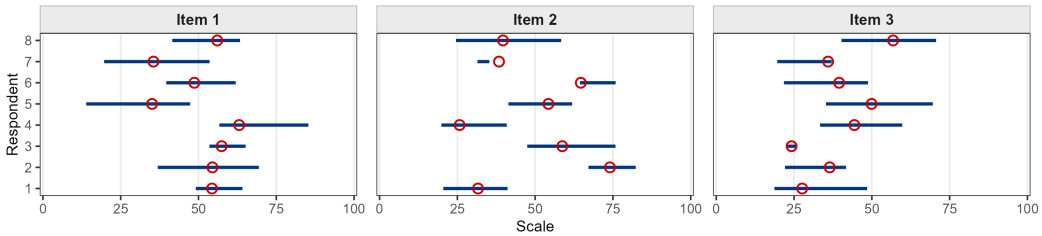
(b) *RW — Random Width*



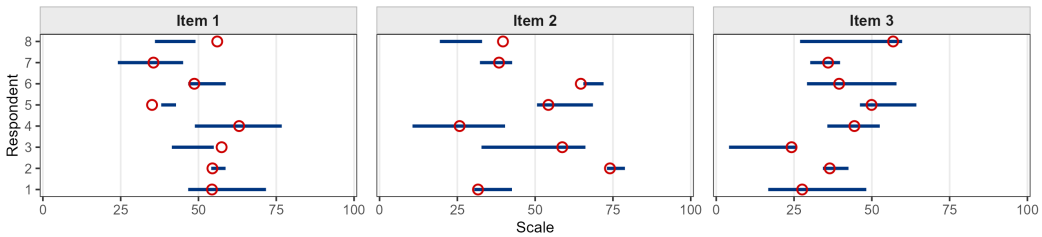
(c) *ICW — Item Consistent Width*



(d) *CW — Correlation Width*



(e) *AM — Asymmetric Midpoint*



(f) *RA — Respondent Asymmetry*

Figure 3.3: *The six interval generation modes applied to the single-valued baseline in Fig. 3.2.*

3.5.2 Experimental Setup

Single-valued responses are generated from a one-factor model, in which all items are driven by a common underlying factor. Let v_{ij} denote the single-valued response of respondent i to item j. For respondent i and item j,

$$v_{ij} = \mu + \sigma(\lambda_j F_i + e_{ij}), \qquad F_i \sim \mathcal{N}(0,1), \quad e_{ij} \sim \mathcal{N}(0,1),$$

where $\mu = 50$ and $\sigma = 10$. Within each condition, the same loading is used for all items, i.e. $\lambda_j = \lambda$.

The numbers of respondents and items are varied as

$$n \in \{100, 200\}, \qquad k \in \{5, 20\}.$$

For each (n, k) configuration, three reliability conditions are considered: high reliability, low reliability, and negative reliability (i.e., $\alpha < 0$). Within each reliability level, the same factor model is used for both values of k, so that any difference between $k = 5$ and $k = 20$ is due to the number of items rather than to a change in the response-generating structure.

The single-valued responses generated under these conditions are then converted into interval-valued responses using the six interval generation modes described above. In total, the simulation comprises $4 \times 3 \times 6 = 72$ experimental conditions. Each condition is replicated 30 times. For each replication, a new full dataset is generated using the random number stream, so that the 30 datasets are different but reproducible under a fixed seed.

3.5.3 Results

Overall relationship between α_θ and α_H. Fig. 3.4 shows the relationship between α_θ and α_H across all 2,160 simulated data sets. The two coefficients show a positive linear relationship, and most points lie close to the line of equality, indicating that α_θ and α_H have similar values under most conditions.

The six modes differ in how the two coefficients compare. RCW is the only mode where α_H is consistently higher than α_θ, forming a cluster above the diagonal; this separation increases as the underlying reliability decreases. Outside RCW, other modes largely remain close to the diagonal, predominantly with $\alpha_\theta \geqslant \alpha_H$. ICW and CW lie close to the equality line. RW, AM, and RA show a pattern of $\alpha_\theta > \alpha_H$, although the differences are modest. In particular, RA has relatively few points below zero for either coefficient compared with the other modes. In the negative-reliability region, all modes except RCW and RA produce similar values.

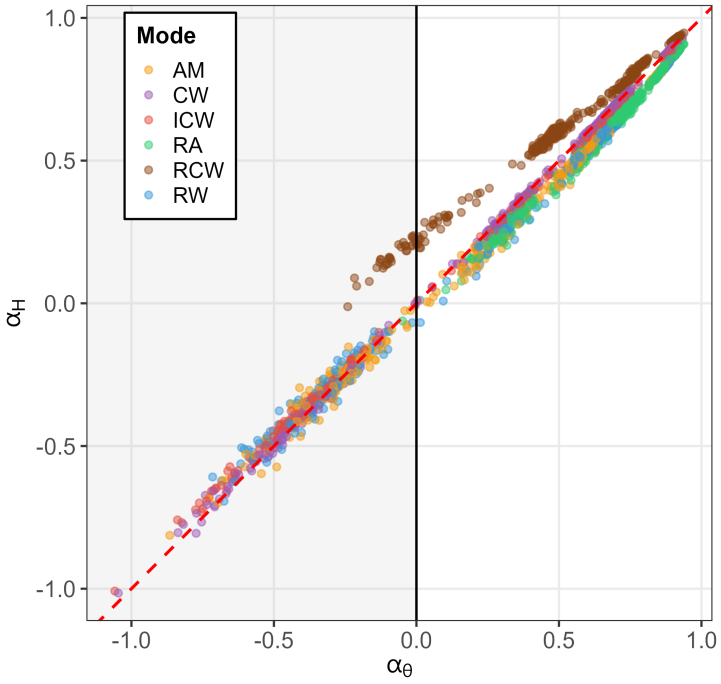


Figure 3.4: *α_θ vs. α_H across all 2,160 simulated data sets. Dashed line: equality.*

Mean behaviour across reliability levels. Fig. 3.5 summarises the average values of α_θ and α_H under the six interval generation modes and three reliability levels.

ICW and CW remain close to the single-valued reference across all reliability levels. RW and AM both fall below the single-valued reference, with AM slightly lower than RW and with a similar α_θ–α_H gap in both modes. RCW shows a relatively large difference between the two coefficients. Across all reliability levels, α_H is higher than α_θ, and both coefficients move above the single-valued baseline for all reliability levels. This pattern is most pronounced under negative reliability. RA also shows a relatively large difference between the two coefficients. Across all reliability levels, α_θ is higher than α_H, and both coefficients move above the single-valued baseline for low and negative reliability levels. This pattern is most pronounced under negative reliability. In addition, under both high and low reliability, RW, AM, and RA appear broadly parallel, indicating that the difference between α_θ and α_H is limited across these three modes.

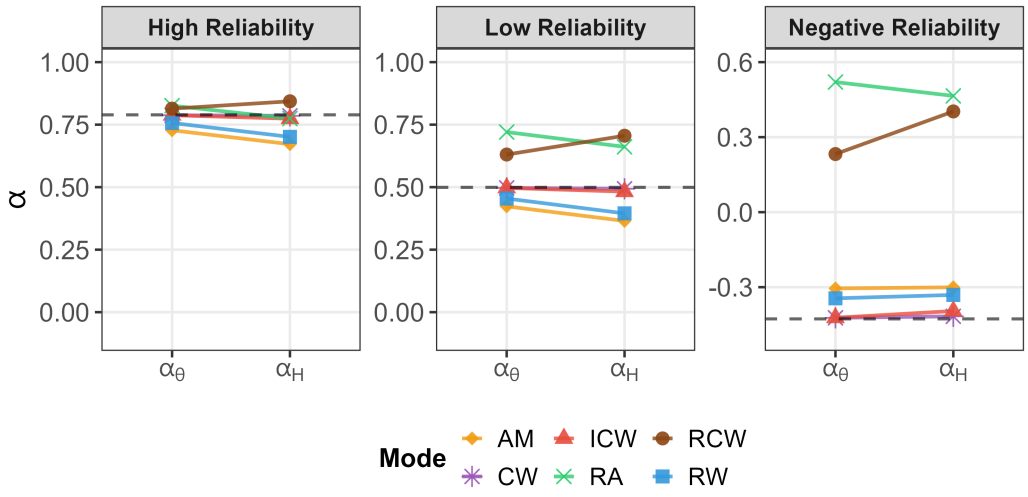


Figure 3.5: *Mean values of α_θ and α_H by interval generation mode and reliability level.*

Effects of sample size and test length. Fig. 3.6 presents the results separated by interval generation mode and (n, k) configuration. Across high and low reliability conditions, increasing the number of items from $k = 5$ to $k = 20$ produces higher reliability estimates in all modes, consistent with the Spearman–Brown formula that reliability tends to increase with test length [102, 103]. Under negative reliability, estimates correctly remain negative and flat for RW, ICW, CW, and AM; in contrast, RCW and RA produce positive estimates that inappropriately increase with k. Differences between $n = 100$ and $n = 200$ are limited. These patterns hold for both α_θ and α_H.

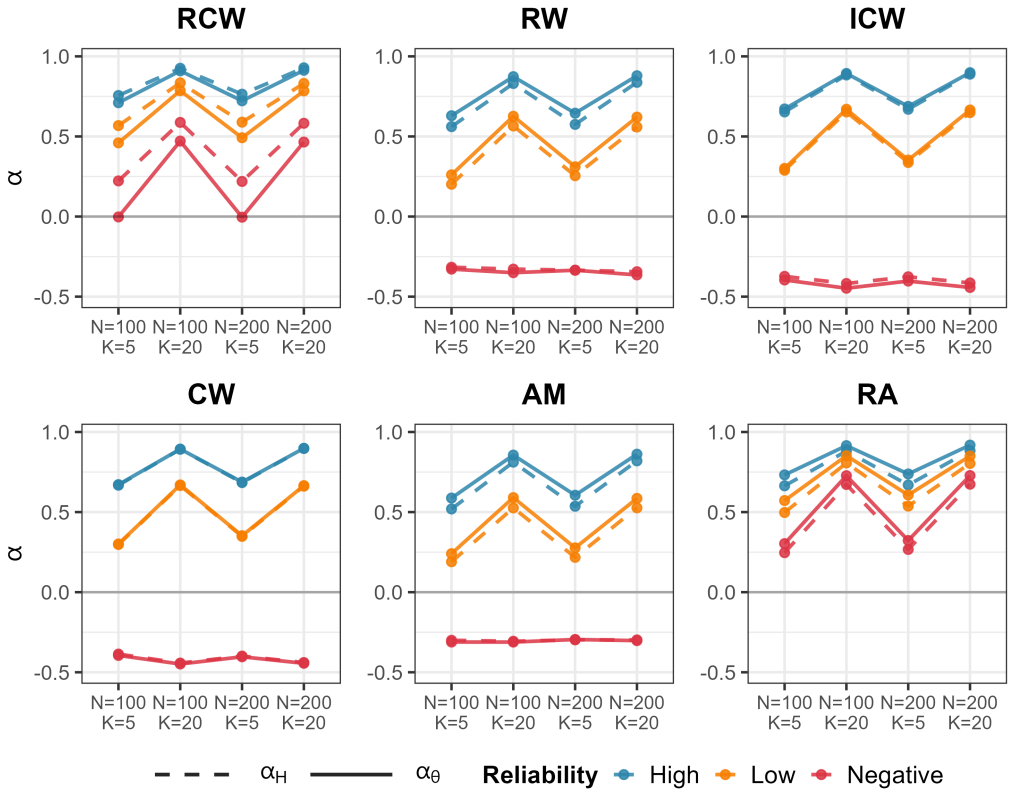


Figure 3.6: *α_θ and α_H across (n, k) configurations, by generation mode.*

3.5.4 Discussion

This discussion considers the simulation results from two perspectives. The first concerns the relationship between α_θ and α_H, focusing on when the two coefficients behave similarly and when they diverge under different interval

generation modes. The second concerns the relationship between the two interval-valued coefficients and the single-valued baseline, focusing on when interval-valued response patterns lead both coefficients to remain close to, or move away from, the corresponding classical Cronbach's α.

α_θ vs. α_H

The simulation shows that α_θ and α_H behave similarly under most interval generation modes, with their main differences arising when interval width carries respondent-level structure.

The main distinction lies in how the two coefficients respond to width variability. When width is tied to item characteristics (ICW) or systematically related to the response value (CW), both coefficients remain close to the single-valued baseline. When width varies randomly at the cell level (RW), both coefficients fall below the single-valued baseline, with α_H decreasing more than α_θ. When width is consistent within respondents (RCW), α_H increases more than α_θ. Taken together, these patterns suggest that α_H is more sensitive to width structure than α_θ: it responds more strongly when width variation is random and also when it is respondent-consistent. By contrast, α_θ appears to attenuate both effects through the mid–spr decomposition.

Midpoint displacement has a smaller differential effect on the two coefficients. The AM and RA modes show that midpoint shifts produce only modest changes in the α_θ–α_H gap relative to the width-matched baseline RW. This is consistent with midpoint variation entering α_θ more directly through the midpoint component, but the resulting difference remains small relative to the width-related patterns observed in RCW.

α_θ and α_H vs. Baseline

Both RCW and RA raise α_θ and α_H above the single-valued baseline because

they introduce a respondent-level source of cross-item consistency that is absent from the original single-valued data. If this consistency reflects a stable and meaningful individual characteristic, then its capture by the interval-valued coefficients may be substantively appropriate. If instead it reflects response style, direct comparison between the interval-valued and single-valued coefficients may overstate reliability gains.

3.6 Limitations

The limitations of this chapter are as follows.

3.6.1 Mathematical Properties

While the mathematical properties of α_H such as the extreme cases and affine invariance have been supported through simulation in Section 3.4, a limitation remains in that formal proofs of these properties have not yet been established. To address this, an initial attempt at proving these properties is provided in Appendix A.

3.6.2 Fixed-width Approach and Sample Level

The present α_H is based on the sample Hausdorff Fréchet framework under the fixed-width approach of Kang et al. [48]. Alternative optimisation choices, such as fixing the midpoint instead, may lead to different behaviour. In addition, the present formulation is restricted to the sample level; the extension to the population case remains for future work.

3.6.3 Comparison with α_θ

Although the simulation studies in this chapter show both similarities and differences between α_H and α_θ, the mathematical reasons behind these patterns remain to be established. The practical meaning of the simulated interval patterns considered in this chapter, such as respondent-consistent width or respondent asymmetry, is also not further discussed in depth in relation to real-world applications. In particular, when an interval-valued coefficient yields a positive value under a negative single-valued baseline, further discussion is needed to determine whether these coefficients capture meaningful internal consistency or an inflated reliability estimate.

3.6.4 Choice of Interval Distance

The present thesis considers only the θ-distance and the Hausdorff distance. The θ-distance was adopted because it is the choice used in existing work, whereas the Hausdorff distance was introduced to address the limitations of mid–spr decomposition. However, no broader analysis was conducted of what other interval distances may be appropriate, what assumptions different choices imply, or under what conditions one distance may be more suitable than another. As a result, although the thesis demonstrates the feasibility of the workflow under these two choices, a more general basis for interval distance selection in interval-valued survey analysis remains to be discussed.

3.7 Summary and Remarks

This chapter introduced α_H, a Hausdorff-distance-based extension of Cronbach's α for interval-valued data, motivated by the semantic and computational limitations of the existing θ-distance-based formulation, which relies on a

mid–spr decomposition that is not always appropriate for real-world interval-valued responses. The basic mathematical behaviour of α_H was demonstrated through simulation, and its behaviour was compared with that of α_θ across six controlled interval generation modes. The two coefficients behave similarly under most conditions, with the main distinction arising under respondent-consistent width, where α_H is more sensitive to width structure than α_θ. The choice between the two coefficients therefore depends primarily on the intended semantics of the interval-valued responses, with respondent-level width consistency being an additional consideration.

Chapter 4

Summarisation: Interval-valued Defuzzification for Non-convex Fuzzy Sets

Fuzzy Sets provide a solid basis for the summarisation of interval-valued survey responses as distributional overviews, but their two-dimensional representation poses challenges when multiple fuzzy sets need to be compared simultaneously. While single-valued defuzzification (Section 2.2.1) methods can simplify fuzzy sets to numbers, such a process discards the uncertainty information that motivated the use of fuzzy sets in the first place. This chapter addresses the summarisation stage of the interval-valued workflow by developing interval-valued defuzzification methods for non-convex fuzzy sets, producing interval summaries that retain key uncertainty information while maintaining communication efficiency. Fig. 4.2 provides illustrative examples of single-valued defuzzification and interval-valued defuzzification applied to convex and non-convex fuzzy sets.

Section 4.1 outlines the motivation, research gaps, and contributions. Section 4.2 reviews the background methods relevant to this chapter. Section 4.3 proposes desirable properties for interval-valued defuzzification of non-convex

fuzzy sets. Section 4.4 introduces two families of interval-valued defuzzification methods proposed to satisfy these properties. Section 4.5 presents an in-depth, two-part evaluation comprising a synthetic data evaluation and a user study. Section 4.6 discusses limitations, and Section 4.7 provides a summary and closing remarks.

4.1 Introduction

Uncertainty is inherently present in real-world settings and is often crucial in communication and decision-making. Fuzzy sets enable the modelling of such uncertainty—particularly vagueness and imprecision in human perception and evaluation—and are widely used across real-world applications.

Fuzzy sets, as outputs, are typically either fed into downstream computational procedures (e.g., clustering and classification) or presented directly to human decision makers for interpretation and decision-making (e.g., fuzzy sets produced by the Interval Agreement Approach (IAA) (Section 2.2.3) to provide a distributional overview of interval-valued survey responses for each aspect in Fig. 4.1).

In the latter case, communication is constrained by decision makers' limited time, cognitive capacity, and varying backgrounds for data interpretation. These constraints become more severe when multiple fuzzy set outputs need to be examined and compared simultaneously.

For example, in customer survey settings, decision makers often need to gain an overall understanding across tens of service aspects and form initial comparisons of how each aspect performs. While each fuzzy set can provide a detailed overview for its respective aspect, comparing tens of them simultaneously can be difficult, as illustrated in Fig. 4.1.

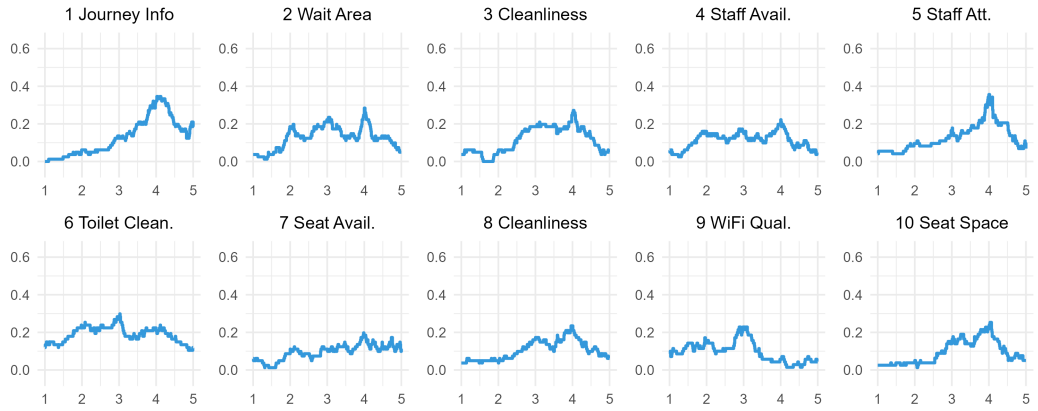


Figure 4.1: *Full fuzzy set distributions for ten survey aspects: potential cognitive overload for decision makers.*

A common approach to address this is to use measures of central tendency as summaries. For single-valued data, these include measures such as mean and median; for fuzzy sets, these include single-valued defuzzification methods such as centroid and Averaging Level Cuts (ALC) (Section 2.2.1).

However, centroid and ALC are both single-valued summaries which cannot communicate the presence or extent of uncertainty. By reducing a fuzzy set to a single point, this effectively discards the uncertainty that is fundamental to fuzzy sets and which motivated the effort to capture and model it in the first place [50, 104, 105]. Such discarding, right before decision-making, can mislead stakeholders suggesting precision where this is not the case.

For example, as shown in Fig. 4.2a, applying centroid to two different trapezoidal fuzzy numbers can yield identical results. A decision maker receiving only the single-valued summary can neither tell that uncertainty is present at all, nor distinguish differing degrees of uncertainty.

Interval-valued summaries are therefore proposed as an alternative to address this limitation. By simplifying a fuzzy set as an interval rather than a single value, uncertainty is retained through the width of the interval while still reducing the complexity of two-dimensional fuzzy sets [50]. This process is referred to as interval-valued defuzzification in contrast to single-valued

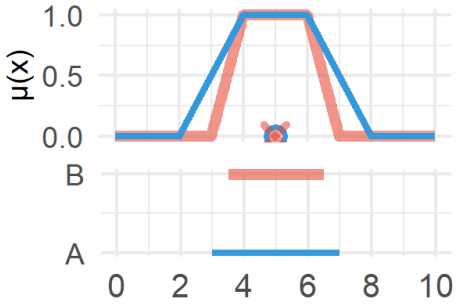
defuzzification.

While interval-valued defuzzification methods were first introduced for fuzzy numbers (see Section 4.2) [61, 68], recent work has emphasised the importance of extending interval-valued defuzzification to multi-modal, i.e., non-convex fuzzy sets [50, 106], given their prevalence in real-world applications, including the output of fuzzy control systems, and methods to model human preferences such as the IAA. Importantly, non-convexity can encode critical information, such as whether consensus can be reached among respondents or whether divergent opinions exist that result in multiple distinct groups.

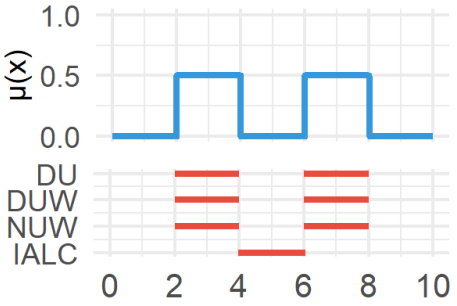
Although recent work [50] has extended interval-valued defuzzification to non-convex fuzzy sets (see Section 4.2), two gaps remain.

First, desirable properties for interval-valued defuzzification of non-convex fuzzy sets remain unspecified. There is no specification of what properties the resulting interval should satisfy to effectively represent the key features of non-convex fuzzy sets. This leaves limited guidance for both method development and the informed interpretation of the resulting summaries. As shown in Figs. 4.2b–4.2d, established methods such as Interval-Valued Averaging Level Cuts ($IALC$) [50, 80] tend to simplify and centralise divergent information into a single, continuous interval, settling on a middle ground between divergent regions without adequately reflecting their disjointness.

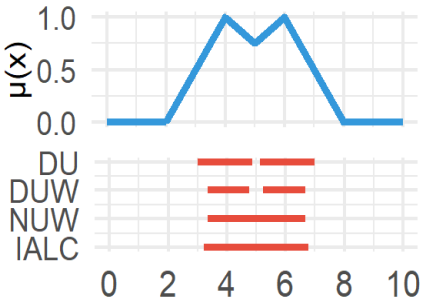
Second, empirical evidence on how human decision makers interpret and use interval-valued summaries is lacking. From the original interval-valued defuzzification methods for fuzzy numbers, to the more recent extensions for non-convex fuzzy sets, the development and discussion of interval-valued defuzzification has been predominantly framed in terms of mathematical properties. If these summaries are intended to effectively communicate uncertainty to human decision makers, then the human view and perspective must also



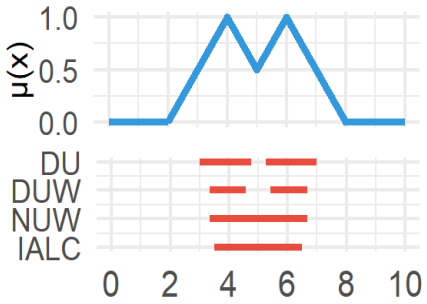
(a) *Trapezoidal fuzzy numbers.*



(b) *Non-convex, non-normal fuzzy set.*



(c) *Non-convex, normal fuzzy set.*



(d) *Non-convex, normal fuzzy set.*

Figure 4.2: *Illustrations of defuzzification methods applied to fuzzy sets with different properties.*

Note: NUW denotes the weighted Nested and Union-based method, and *DU*/*DUW* the Disjoint and Union-based method and its weighted variant, all proposed in this chapter and introduced in Section 4.4; *IALC* denotes the Interval-Valued Averaging Level Cuts [50, 80], reviewed in Section 4.2.

be observed and considered. In particular, it is necessary to understand which fuzzy set features people expect to be retained in simplification, how they interpret interval summaries in uncertainty communication, and whether such expectations align with—or require adjustments to—proposed desirable properties.

Accordingly, this chapter addresses both gaps and makes the following contributions:

1. Desirable properties for interval-valued defuzzification of non-convex fuzzy sets are proposed, and methods developed to satisfy them are presented. Mathematical proofs are provided to establish that these methods satisfy the properties, and synthetic data are then used to evaluate them against established alternatives with respect to these properties.

2. The effectiveness of these methods is evaluated through a mixed-methods user study, providing evidence on the potential of interval-valued summaries for human-facing uncertainty communication.

The code for the synthetic evaluation and the user study survey materials are available on GitHub at https://github.com/Carina-YuZhao/.

4.2 Related Work

This section recalls fuzzy sets and their relevant properties, α-cuts, the interval aggregation methods Interval Average (IA) and Interval Weighted Average (IWA), and the existing interval-valued defuzzification methods for both convex and non-convex fuzzy sets.

4.2.1 Fuzzy Sets and Their Properties

A fuzzy set A on a universe of discourse X is defined by its membership function $\mu_A : X \to [0, 1]$, which assigns to each element of X a degree of membership in the set [20, 56]. Its key properties and notations are reviewed in detail in Section 2.1.3.

Convex Fuzzy Sets and Fuzzy Numbers

A convex fuzzy set is a fuzzy set satisfying the convexity property.

A fuzzy number is a special case of a convex fuzzy set that is also normal and upper semi-continuous with bounded support [56, 61]. Fig. 4.2a shows two trapezoidal fuzzy numbers.

Non-convex Fuzzy Sets

A non-convex fuzzy set does not satisfy the convexity property and often exhibits multiple peaks in its membership function (see Figs. 4.2c and 4.2d). A non-convex fuzzy set may have multiple (disjoint) connected components within its core and/or support, representing non-continuous or contradictory states often encountered in complex information systems.

4.2.2 Alpha-cuts and Interval Aggregation

Alpha-cuts and interval aggregation methods are essential to interval-valued defuzzification: the former decompose a fuzzy set into a set of intervals, and the latter aggregate them into one.

The α-cut of a fuzzy set A at level α is the set of all elements whose membership

is at least α (Section 2.2.3). For a convex fuzzy set, each non-empty α-cut is a continuous interval $\overline{A_\alpha} = [A^-(\alpha), A^+(\alpha)]$. For a non-convex fuzzy set, an α-cut can instead be a discontinuous interval $\overline{\overline{A}}_\alpha$ consisting of multiple disjoint continuous sub-intervals.

The IA of n intervals $\overline{A}_1, \ldots, \overline{A}_n$ is defined as:

$$\text{IA} = \left[\frac{1}{n} \sum_{i=1}^{n} A_i^-,\ \frac{1}{n} \sum_{i=1}^{n} A_i^+ \right]. \tag{4.1}$$

The IWA generalises the IA by incorporating weights. For single-valued weights $w_1, \ldots, w_n$ with $w_i \geqslant 0$:

$$\text{IWA} = \left[\frac{\sum_{i=1}^{n} w_i\, A_i^-}{\sum_{i=1}^{n} w_i},\ \frac{\sum_{i=1}^{n} w_i\, A_i^+}{\sum_{i=1}^{n} w_i} \right]. \tag{4.2}$$

Further details on both alpha-cuts and interval aggregation are provided in Section 2.2.3 and Section 2.3.2, respectively.

4.2.3 Interval-valued Defuzzification for Convex Fuzzy Sets

Grzegorzewski [61] put forward that to approximate a fuzzy number A by an interval $\overline{A} = [A^-, A^+]$, the interval should satisfy the following properties:

1. **Support-subset**: The interval should be a subset of the support of A, i.e., $\overline{A} \subseteq \text{supp}(A)$.
2. **Core-inclusion**: The interval should include the core of A, i.e., $\text{core}(A) \subseteq \overline{A}$.

To this end, Grzegorzewski proposed an interval-valued defuzzification method—denoted IVD in this thesis (Section 2.2.3)—and proved that, for normal fuzzy numbers, it satisfies both properties [61]. This proof can be

extended to IVD and Weighted Interval-valued Defuzzification ($WIVD$) for convex non-normal fuzzy sets. Both methods aggregate α-cut intervals vertically across levels. In continuous form, for a convex fuzzy set A with height $h(A)$ [50]:

$$IVD(A) = \left[\frac{1}{h}\int_0^h A^-(\alpha)\,d\alpha,\ \ \frac{1}{h}\int_0^h A^+(\alpha)\,d\alpha\right], \tag{4.3}$$

$$WIVD(A) = \left[\frac{\int_0^h \alpha\, A^-(\alpha)\,d\alpha}{\int_0^h \alpha\,d\alpha},\ \ \frac{\int_0^h \alpha\, A^+(\alpha)\,d\alpha}{\int_0^h \alpha\,d\alpha}\right]. \tag{4.4}$$

$WIVD$ assigns higher weight to α-cuts at higher levels, giving greater influence to intervals closer to the core [74]. Discrete forms and further details are provided in Section 2.2.3.

4.2.4 Interval-valued Defuzzification for Non-convex Fuzzy Sets

Anzilli et al. [50] proposed a general framework for interval-valued defuzzification of non-convex fuzzy sets (Section 2.2.3). While this framework allows different parameter choices for method variation, from the perspective of this thesis, all variants share the same structural limitation: at each α-level, the disjoint sub-intervals are first aggregated horizontally into a single continuous interval before vertical aggregation, so the output is always a continuous interval. The structural information conveyed by non-convexity—such as the presence of divergent opinion groups—is therefore not represented in the interval-valued output.

Since this limitation is consistent across all variants, $IALC$ [50, 80] is adopted as a representative variant for comparison. $IALC$ is derived from ALC [49, 62], which is among the simplest and most widely used α-level-based single-valued

defuzzification methods. To defuzzify a non-convex fuzzy set A by an interval $\overline{A} = [A^-, A^+]$, $IALC$ is defined as:

$$\begin{aligned} A^- &= \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{q_{\alpha_j}} A_i^-(\alpha_j)\,\frac{|\overline{A}_i^{\alpha_j}|}{|\overline{A}_{\alpha_j}|}, \\ A^+ &= \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{q_{\alpha_j}} A_i^+(\alpha_j)\,\frac{|\overline{A}_i^{\alpha_j}|}{|\overline{A}_{\alpha_j}|}, \end{aligned} \tag{4.5}$$

where q_{α_j} denotes the number of disjoint sub-intervals at level α_j, $|\overline{A}_i^{\alpha_j}|$ is the length of the i-th sub-interval, and $|\overline{A}_{\alpha_j}|$ is their total length.

4.3 Desirable Properties

As introduced in Section 4.2, two fundamental properties 1) and 2) have been defined for the interval-valued defuzzification of fuzzy numbers. In this chapter, these properties are extended to the interval-valued defuzzification of non-convex fuzzy sets to ensure the representation of their support and core, which are two important properties for conveying the information of non-convex fuzzy sets.

Through observation, it is noted that the established interval-valued defuzzification methods for fuzzy sets, including $IALC$ (4.5) introduced in Section 4.2, are not always able to satisfy both properties simultaneously. This is demonstrated through the following counterexample.

Proposition. For a non-convex fuzzy set A which is the union of two triangular fuzzy numbers $A_1 = (a_1, b_1, c_1)$ and $A_2 = (a_2, b_2, c_2)$, where $a_1 < b_1 < c_1 < a_2 < b_2 < c_2$, and which therefore has a disconnected support ($\text{supp}(A) = [a_1, c_1] \cup [a_2, c_2]$) and a core with multiple disjoint components ($\text{core}(A) = \{b_1, b_2\}$), no continuous interval can satisfy both Properties 1) Support-subset and 2) Core-inclusion simultaneously.

Proof. Since $b_1 < b_2$, the smallest continuous interval that includes core(A) and thus satisfies Property 2) Core-inclusion is $[b_1, b_2]$. However, as $b_1 < c_1 < a_2 < b_2$, the interval $[b_1, b_2]$ necessarily includes the sub-interval $[c_1, a_2]$, which is not part of supp(A) and thus violates Property 1) Support-subset. □

Therefore, leveraging discontinuous intervals for the interval-valued defuzzification of non-convex fuzzy sets is proposed so that these properties can be satisfied. It is also important to note that the peaks and non-convex areas of non-convex fuzzy sets carry critical information and that it is intuitive to expect these features to be reflected in simplified representations, such as through areas of discontinuity in the interval-valued defuzzification. Two additional desirable properties are therefore introduced that establish the relationship between a non-convex fuzzy set and its resulting interval-valued defuzzification discontinuous interval. For a non-convex fuzzy set A:

3. **Peak/Sub-interval ratio**: If A has $p(A)$ peaks, then its interval-valued defuzzification interval $\bar{\bar{A}}$ should contain $q(\bar{\bar{A}}) = p(A)$ sub-intervals. That is, the peak-to-sub-interval ratio is 1.

4. **Proportionality**: If the size of the non-convex (trough) area in A—the region around the local minimum between peaks—increases or decreases, the size of the discontinuity/gap in the interval-valued defuzzification interval changes proportionally.

For example, the fuzzy set in Fig. 4.2c has two peaks; thus, it is intuitive that its interval-valued defuzzification should have two sub-intervals. If the non-convex area is enlarged to form the fuzzy set in Fig. 4.2d, the discontinuity between these two sub-intervals should increase correspondingly.

It is worth noting that Property 3) Peak/Sub-interval Ratio can also be described as follows: if a fuzzy set A has $(p(A) - 1)$ troughs, the interval has

$(p(A) - 1)$ discontinuities between its sub-intervals. While peaks are used in this chapter, the choice between peaks and troughs should be based on which concept is more relevant or critical to the audience or specific application.

4.4 Methods

In this section, two families of interval-valued defuzzification methods for non-convex fuzzy sets are introduced. First, the Nested and Union-based methods (NU and its weighted variant NUW, collectively denoted $NU(W)$) are presented. A key limitation of $NU(W)$ is then discussed, which leads to the development of Disjoint and Union-based methods (DU and its weighted variant DUW, collectively denoted $DU(W)$).

4.4.1 Nested and Union-based Interval-valued Defuzzification

Design Rationale

The $NU(W)$ method draws inspiration from the original interval-valued defuzzification methods for fuzzy numbers discussed in Section 4.2. These methods involve aggregating nested intervals, and they have been proven to consistently satisfy Properties 1) Support-subset and 2) Core-inclusion [61]. This process is illustrated in Fig. 4.3b, where the α-cuts result in nested intervals that extend from the core (the smallest nested interval) to the support (the largest nested interval). Consequently, the aggregated interval encompasses the core and remains within the support, thus satisfying these properties.

Since the proposed methods handle non-convex fuzzy sets through α-cut decomposition, the discrete form is used—directly on the sub-intervals produced

at each α-level. For convex fuzzy sets, the α-cuts at each level produce a single interval, and these intervals are inherently nested. For non-convex fuzzy sets, however, α-cuts can produce multiple sub-intervals at each level, and sub-intervals from different levels may belong to different nested subsets of intervals. The key idea of $NU(W)$ is to identify all valid nested subsets of intervals among the α-cut sub-intervals, aggregate each nested subset separately, and combine the results through union—potentially producing a discontinuous interval that reflects the non-convexity of the original fuzzy set.

Method Description and Algorithm

Let A be a non-convex fuzzy set decomposed into a set of α-cut sub-intervals using levels $\{\alpha_j \mid j = 1, \ldots, m\}$, as defined in Section 2.2.3. Specifically, at each level α_j, the α-cut yields a (possibly discontinuous) interval $\overline{\overline{A}}_{\alpha_j}$ comprising one or more sub-intervals. The $NU(W)$ method consists of three steps:

1. **Nested Sets:** Identify all maximal nested subsets of intervals from the α-cut sub-intervals. A nested subset is a sequence of sub-intervals, one from each α-level (where available), such that each sub-interval is contained within the one from the preceding lower α-level.

2. **Aggregate:** Aggregate each nested subset into a single sub-interval using IA (4.1) or IWA (4.2).

3. **Union:** Combine these aggregated sub-intervals through a union operation to form the final discontinuous interval.

The equal-weighted variant is denoted by NU (i.e., using IA in Step 2) and the weighted variant by NUW (i.e., using IWA in Step 2), where NUW places greater emphasis on higher α-levels.

An illustrative example is as follows. At the lowest α-level, due to non-convexity, there are two sub-intervals: $\overline{A}_1^{\alpha_1}$ and $\overline{A}_2^{\alpha_1}$. At a higher α-level, there is one sub-interval $\overline{A}_1^{\alpha_2}$, which is nested within $\overline{A}_2^{\alpha_1}$ but not within $\overline{A}_1^{\alpha_1}$. Therefore, the maximal nested subsets of intervals are $(\overline{A}_1^{\alpha_1})$ and $(\overline{A}_2^{\alpha_1}, \overline{A}_1^{\alpha_2})$. Aggregating each nested subset using IA or IWA produces two sub-intervals, whose union forms the final discontinuous interval.

Algorithm 1 provides a formal description of the $NU(W)$ method. The recursive function FindNested (Algorithm 2) explores all maximal nested subsets of intervals starting from each sub-interval at the lowest α-level.

Algorithm 1 Compute $NU(A)$ or $NUW(A)$

Input: α-levels $\{\alpha_j \mid j = 1, \ldots, m\}$ and α-cut results $\{\bar{\bar{A}}_{\alpha_j}\}_{j=1}^{m}$.
Output: Discontinuous interval $\bar{\bar{A}}$.
Initialise an empty list S for storing maximal nested subsets of intervals (each stored as an ordered sequence of sub-intervals).

Step 1: Nested Sets.
for each sub-interval I in $\bar{\bar{A}}_{\alpha_1}$ **do**
    FindNested(I, 1, [ ], S)
end for

Step 2: Aggregate.
for each nested subset T in S **do**
    Compute the IA or IWA of T to obtain an aggregated sub-interval.
end for

Step 3: Union.
Take the union over all aggregated sub-intervals as the output $\bar{\bar{A}}$.

Geometric Interpretation

It is worth noting that, because a fuzzy set can be uniquely represented by and reconstructed from its α-cuts, $NU(W)$ also admits a geometric interpretation, as demonstrated in Fig. 4.3. Since each nested subset is a collection of nested α-cut intervals, it can be reconstructed—via the correspondence between α-cuts and fuzzy sets—into a convex fuzzy set. The method is therefore

Algorithm 2 Recursive function FINDNESTED

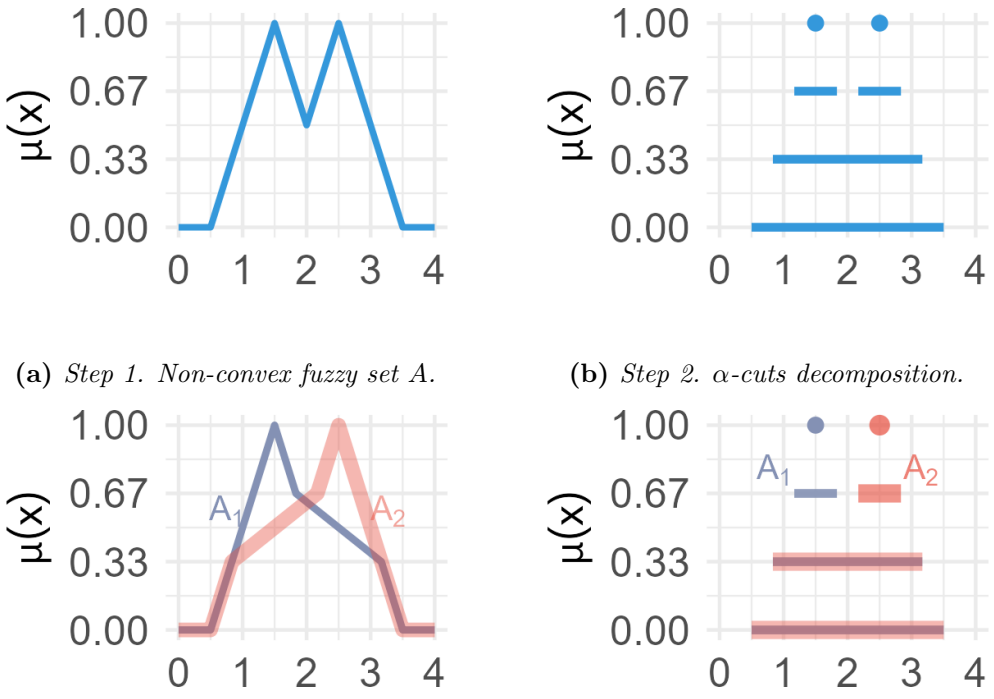
```
1: Function FindNested(CI, CL, subset, S)      ▷ CI: Current Interval, CL: Current Level
2: Append CI to subset.
3: if CL = m or no sub-interval in Ā_{α_{CL+1}} is nested within CI then
4:     Append subset to S.
5:     return
6: else
7:     for each sub-interval J in Ā_{α_{CL+1}} nested within CI do
8:         new_subset ← copy of subset
9:         FindNested(J, CL + 1, new_subset, S)
10:    end for
11: end if
12: End Function
```

equivalent to decomposing a non-convex fuzzy set with $p(A)$ peaks into $p(A)$ convex fuzzy subsets and applying IVD (4.3) or $WIVD$ (4.4) to each subset separately, then taking the union. Crucially, these reconstructed convex subsets may overlap in the domain, because at lower α-levels they share the same continuous α-cut intervals spanning multiple peaks.

With this geometric interpretation, $NU(W)$ can be generalised to a continuous form, in addition to its discrete form as detailed in Algorithm 1. While the discrete form operates on a finite number of α-levels and is used for computational implementation, its continuous form applies to any fuzzy set with a well-defined membership function and is used for mathematical proofs in this chapter for generality.

Limitation

From the geometric interpretation above, $NU(W)$ decomposes a non-convex fuzzy set with $p(A)$ peaks into $p(A)$ convex subsets—one per peak—and applies IVD or $WIVD$ to each, yielding $p(A)$ sub-intervals. If these sub-intervals were pairwise disjoint, the result would satisfy Property 3) Peak/Sub-interval Ratio with $q(\bar{\bar{A}}) = p(A)$. However, since the reconstructed convex subsets



(a) *Step 1. Non-convex fuzzy set A.*

(b) *Step 2. α-cuts decomposition.*

(c) *Step 4. Reconstructed convex fuzzy sets.*

(d) *Step 3. Two nested subsets.*

Figure 4.3: *Geometric interpretation of the NU(W) method.*

can overlap in the domain, their IVD ($WIVD$) results can also overlap and merge upon union, reducing $q(\overline{A})$ below $p(A)$. This is formalised as follows.

Proposition. *There exists a non-convex fuzzy set A for which NU and NUW fail to satisfy Property 3) Peak/Sub-interval Ratio.*

Proof. Let $B_1 = (0, 2, 4)$ and $B_2 = (1, 3, 5)$ denote two triangular fuzzy numbers, with membership functions μ_{B_1} and μ_{B_2}, respectively. Define the non-convex fuzzy set A by

$$\mu_A(x) = \max\{\mu_{B_1}(x), \mu_{B_2}(x)\}.$$

Then A has two peaks, at $x = 2$ and $x = 3$, so $p(A) = 2$.

Its α-cuts are

$$\overline{\overline{A}}_\alpha = \begin{cases} [2\alpha,\ 5-2\alpha], & 0 \leqslant \alpha \leqslant \frac{3}{4}, \\ [2\alpha,\ 4-2\alpha] \cup [1+2\alpha,\ 5-2\alpha], & \frac{3}{4} < \alpha \leqslant 1. \end{cases}$$

Under the geometric interpretation of $NU(W)$, the two maximal nested families reconstruct two convex fuzzy sets A_1 and A_2, whose α-cut intervals are, respectively,

$$\overline{A}_{1,\alpha} = \begin{cases} [2\alpha,\ 5-2\alpha], & 0 \leqslant \alpha \leqslant \frac{3}{4}, \\ [2\alpha,\ 4-2\alpha], & \frac{3}{4} < \alpha \leqslant 1, \end{cases} \qquad \overline{A}_{2,\alpha} = \begin{cases} [2\alpha,\ 5-2\alpha], & 0 \leqslant \alpha \leqslant \frac{3}{4}, \\ [1+2\alpha,\ 5-2\alpha], & \frac{3}{4} < \alpha \leqslant 1. \end{cases}$$

Applying IVD gives

$$\overline{A}_1 = IVD(A_1) = \left[\int_0^1 2\alpha\, d\alpha,\ \int_0^{3/4} (5-2\alpha)\, d\alpha + \int_{3/4}^1 (4-2\alpha)\, d\alpha \right] = \left[1, \frac{15}{4}\right],$$

$$\overline{A}_2 = IVD(A_2) = \left[\int_0^{3/4} 2\alpha\, d\alpha + \int_{3/4}^1 (1+2\alpha)\, d\alpha,\ \int_0^1 (5-2\alpha)\, d\alpha \right] = \left[\frac{5}{4}, 4\right].$$

These intervals overlap, so

$$NU(A) = \overline{\overline{A}} = \overline{A}_1 \cup \overline{A}_2 = [1, 4],$$

which is a single continuous interval. Hence $q(\overline{\overline{A}}) = 1 \neq p(A) = 2$.

Similarly, applying $WIVD$ gives

$$\overline{A}_1 = WIVD(A_1) = \left[\frac{4}{3}, \frac{155}{48}\right], \qquad \overline{A}_2 = WIVD(A_2) = \left[\frac{85}{48}, \frac{11}{3}\right],$$

which also overlap, so

$$NUW(A) = \overline{\overline{A}} = \overline{A}_1 \cup \overline{A}_2 = \left[\frac{4}{3}, \frac{11}{3}\right],$$

again giving $q(\overline{\overline{A}}) = 1 \neq p(A) = 2$.

Therefore, neither NU nor NUW guarantees Property 3) Peak/Sub-interval Ratio. □

This limitation motivates the development of the $DU(W)$ methods, as proposed below.

4.4.2 Disjoint and Union-based Interval-valued Defuzzification

Design Rationale

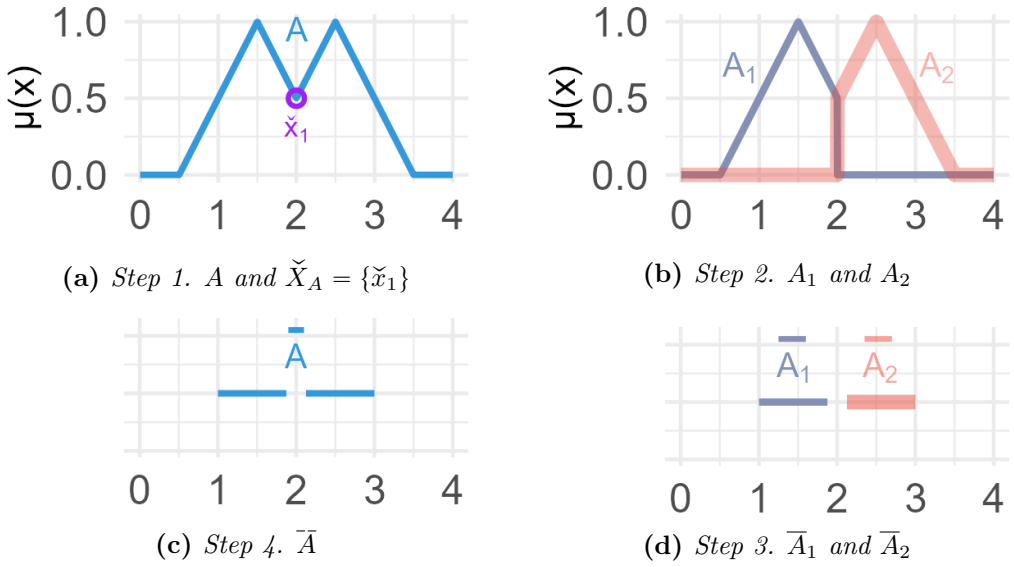
Since $NU(W)$ cannot reliably satisfy Property 3) Peak/Sub-interval Ratio, a different approach is taken that addresses this property directly. According to Property 3) Peak/Sub-interval Ratio, any interval representation of a non-convex fuzzy set with $p(A) > 1$ peaks should reflect its multi-peak feature by comprising exactly $p(A)$ sub-intervals—that is, it should be a discontinuous interval.

To ensure this, A is decomposed into $p(A)$ pairwise quasi-disjoint convex fuzzy subsets, each encapsulating exactly one peak. Here, quasi-disjoint means that any two subsets may intersect at most at a single point, typically a shared trough. These subsets are obtained by placing dividing lines at the midpoints of all troughs.

Established interval-valued defuzzification methods for convex fuzzy sets—specifically IVD (4.3) or $WIVD$ (4.4)—can then be applied to each subset

to obtain disjoint sub-intervals, and their union taken to derive the overall discontinuous interval representation of A, thereby satisfying Property 3) Peak/Sub-interval Ratio.

An illustrative example of the method is provided in Fig. 4.4, and a comparison of this method and *NUW* on the same non-convex fuzzy set is provided in Fig. 4.5.



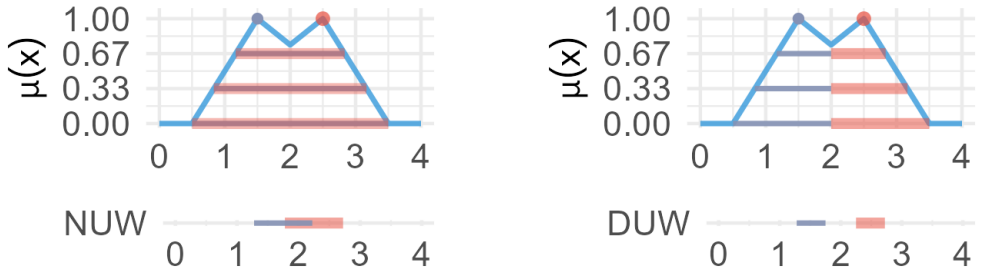
(a) *Step 1.* A *and* $\breve{X}_A = \{\breve{x}_1\}$

(b) *Step 2.* A_1 *and* A_2

(c) *Step 4.* $\overline{A}$

(d) *Step 3.* $\overline{A}_1$ *and* $\overline{A}_2$

Figure 4.4: *Illustrative example of the proposed DU(W) method.*



(a) **(b)**

Figure 4.5: *Comparison of NUW (a) and the proposed method (b) on the same non-convex fuzzy set. In each sub-figure, the top shows the α-cut decomposition and the bottom the resulting sub-intervals.*

Method Description and Algorithm

Let A be a non-convex fuzzy set with $p(A)$ peaks. Between each pair of consecutive peaks there exists a local minimum (a trough, including flat ones), hence there are $(p(A) - 1)$ of them in total. For each trough between the i-th and $(i + 1)$-st peaks, the midpoint of the trough is denoted by $\breve{x}_i$, taken as its position.

1. **Create ordered list of trough positions:** Identify the ordered set of trough positions $\breve{X}_A = \{\breve{x}_i \mid i = 1, \ldots, p - 1\}$ using a first-derivative approach based on finite differences: a trough is detected when the slope changes from negative to positive.

2. **Decompose fuzzy set into convex subsets:** Using the positions in $\breve{X}_A$, together with the domain boundaries (denoted by convention as $\breve{x}_0$ and $\breve{x}_p$), decompose A into $p(A)$ pairwise quasi-disjoint convex fuzzy subsets $\boldsymbol{A} = \{A_i \mid i = 1, \ldots, p\}$. Each subset A_i is defined over the interval $[\breve{x}_{i-1}, \breve{x}_i]$ as

$$\mu_{A_i}(x) = \begin{cases} \mu_A(x), & \text{if } x \in [\breve{x}_{i-1}, \breve{x}_i], \\ 0, & \text{otherwise.} \end{cases}$$

   Thus,

$$A = \bigcup_{i=1}^{p} A_i.$$

3. **Compute sub-intervals for convex subsets:** Apply IVD (4.3) or $WIVD$ (4.4) to each A_i to obtain the corresponding sub-intervals $\overline{A}_i$.

4. **Combine sub-intervals:** The final output $\overline{\overline{A}}$ is obtained by taking the

union of the sub-intervals:

$$\overline{\overline{A}} = \bigcup_{i=1}^{p} \overline{A}_i$$

The equal-weighted variant is denoted by DU (i.e., using IVD in Step 3) and the weighted variant by DUW (i.e., using $WIVD$ in Step 3). It is noted that while the method already handles non-normal fuzzy sets through the normalisation approach used by IVD and $WIVD$ (as discussed in Section 2.2.3), alternative methods for handling non-normality that may be developed in the future can also be incorporated at Step 3.

Algorithm 3 provides a formal description.

Satisfaction of Desirable Properties

Detailed mathematical proofs are provided in Appendix B. The key results are summarised here. Both DU and DUW strictly satisfy Properties 1) Support-subset, 2) Core-inclusion, and 3) Peak/Sub-interval Ratio.

Regarding Property 4) Proportionality, there are two cases. When all peaks are normal, DU strictly satisfies proportionality. The same holds for DUW if the trough area is computed with α-values as weights—that is, following the same weighting approach used in DUW. However, for non-normal fuzzy sets, the proportional relationship holds only between the gap width and the normalised trough area, rather than the raw geometric area. This is further discussed in Section 4.6.

Overall, both DU and DUW satisfy all four desirable properties, with the noted caveat for non-normal fuzzy sets regarding proportionality.

Algorithm 3 Compute $DU(A)$ or $DUW(A)$

Input: Fuzzy set A, represented by domain values $x[1..n]$ and membership values $y[1..n]$.
Output: Discontinuous interval $\overline{A}$.
Initialise empty lists: $\breve{X}_A$ for trough positions, $\boldsymbol{A}$ for decomposed convex subsets, and $\overline{\boldsymbol{A}}$ for sub-intervals.

Step 1: Create ordered list of trough positions.
for $i = 2$ to $n - 1$ **do**
    if $y[i-1] > y[i] < y[i+1]$ **then**
        Add $x[i]$ to $\breve{X}_A$ ▷ Trough
    else if $y[i-1] > y[i] = y[i+1] = \ldots = y[j] < y[j+1]$ **then** ▷ j is the last index of a flat region
        Add midpoint $x[\lfloor (i+j)/2 \rfloor]$ to $\breve{X}_A$ ▷ Flat trough
        Set $i \leftarrow j + 1$
    end if
end for
Add the boundary points $x[1]$ and $x[n]$ to $\breve{X}_A$, and sort all elements in ascending order as $\{\breve{x}_0, \breve{x}_1, \ldots, \breve{x}_p\}$.

Step 2: Decompose fuzzy set into convex subsets.
for $i = 1$ to p **do**
    Define A_i over $[\breve{x}_{i-1}, \breve{x}_i]$:

$$\mu_{A_i}(x) = \begin{cases} \mu_A(x), & \text{if } x \in [\breve{x}_{i-1}, \breve{x}_i], \\ 0, & \text{otherwise.} \end{cases}$$

    Add A_i to $\boldsymbol{A}$.
end for

Step 3: Compute sub-intervals for convex subsets.
for each $A_i \in \boldsymbol{A}$ **do**
    Compute $\overline{A}_i$ using IVD or $WIVD$; add to $\overline{\boldsymbol{A}}$.
end for

Step 4: Combine sub-intervals.
Take the union of all $\overline{A}_i$ in $\overline{\boldsymbol{A}}$ to obtain $\overline{A}$.

4.5 Evaluation

This section presents a two-part evaluation. The evaluation begins with an assessment based on synthetic data sets, evaluating different methods with respect to the four desirable properties outlined in Section 4.3. The methods evaluated include the proposed DU and DUW, as well as established approaches $IALC$ [49, 50, 62, 80] and NUW [106].

Since these methods are intended to provide decision makers with information represented by non-convex fuzzy sets, the synthetic evaluation is complemented with a mixed-methods user study designed to examine how effectively different interval-valued defuzzification methods *communicate* such information.

4.5.1 Synthetic Data Evaluation

Procedure

Four interval-valued defuzzification methods are applied to eight synthetic fuzzy sets to evaluate how these methods respond to variations in the number and size of non-convex areas within the fuzzy sets. These fuzzy sets are illustrated in Figs. 4.6 and 4.7.

The interval-valued defuzzification results produced by each method are evaluated with respect to the four desirable properties, producing the following measures:

1. **Support-subset**: Assesses whether the interval-valued defuzzification result lies within the support of the corresponding fuzzy set. A tick (✓) indicates that the condition is satisfied, and a cross (✗) indicates that it is not.

2. **Core-inclusion**: Assesses whether the interval-valued defuzzification result contains the core of the corresponding fuzzy set. A tick (✓) indicates that the condition is satisfied, and a cross (✗) indicates that it is not.

3. **Peak/Sub-interval Ratio**: Evaluates the ratio between the number of peaks in the fuzzy set and the number of sub-intervals in its interval-valued defuzzification result. The property is satisfied when this ratio equals 1.

4. **Proportionality**: Evaluates whether the ratio between the gap width and the trough area remains consistent across all fuzzy sets. For $IALC$ and DU, the geometric trough area is used, whereas for NUW and DUW, the α-weighted trough area is used to maintain consistency with their weighting approach.

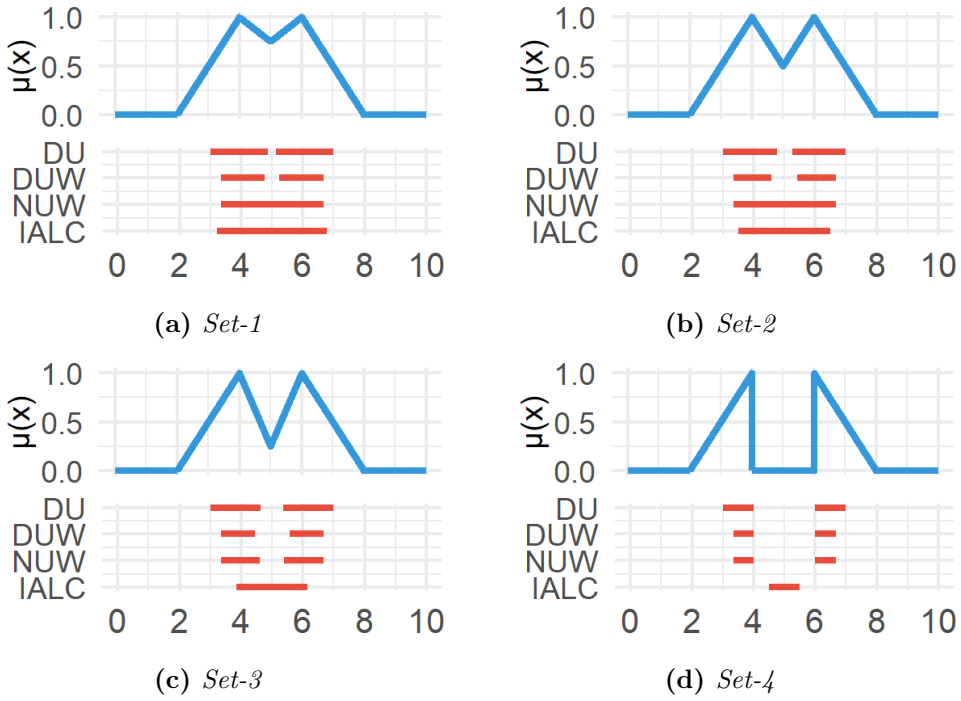
Results and Discussion

The evaluation results are presented in Tables 4.1 to 4.4, with each column showing the results for one of the eight synthetic data sets. Methods that consistently satisfy a given property across all sets are highlighted in green.

As shown in the tables, $IALC$ does not consistently satisfy any of the properties, while NUW consistently satisfies only Properties 1) Support-subset and 2) Core-inclusion. In contrast, both DU and DUW satisfy all four properties.

Table 4.1: Property 1. Support-subset

Set	1	2	3	4	5	6	7	8
$IALC$	✓	✓	✓	✗	✓	✓	✓	✗
NUW	✓	✓	✓	✓	✓	✓	✓	✓
DUW	✓	✓	✓	✓	✓	✓	✓	✓
DU	✓	✓	✓	✓	✓	✓	✓	✓

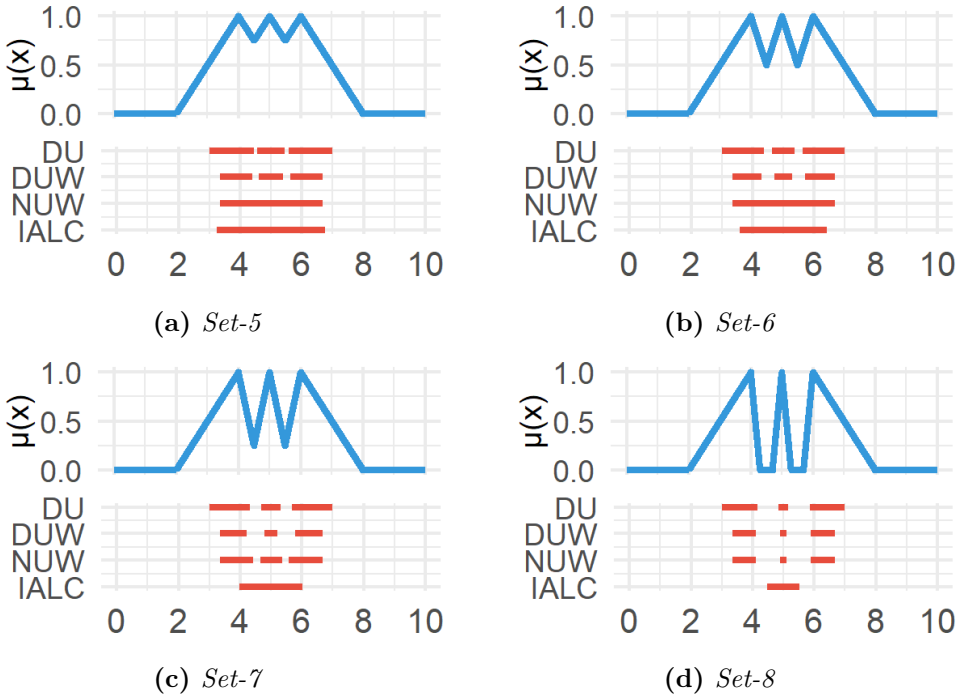


(a) *Set-1* **(b)** *Set-2*

(c) *Set-3* **(d)** *Set-4*

Set	1	2	3	4
IALC	[3.22,6.78]	[3.50,6.50]	[3.84,6.16]	[4.50,5.50]
NUW	[3.33,6.67]	[3.33,6.67]	[3.33,4.60] [5.40,6.67]	[3.33,4.00] [6,6.67]
DUW	[3.33,4.77] [5.23,6.67]	[3.33,4.58] [5.42,6.67]	[3.33,4.44] [5.56,6.67]	[3.33,4.00] [6.00,6.67]
DU	[3.00,4.87] [5.13,7.00]	[3.00,4.75] [5.25,7.00]	[3.00,4.62] [5.38,7.00]	[3.00,4.00] [6.00,7.00]

Figure 4.6: *Set-1 to Set-4: Two-peak synthetic fuzzy sets with increasing trough depth.*



(a) *Set-5* **(b)** *Set-6*

(c) *Set-7* **(d)** *Set-8*

Set	5	6	7	8
IALC	[3.25,6.75]	[3.58,6.42]	[3.97,6.03]	[4.47,5.53]
NUW	[3.33,6.67]	[3.33,6.67]	[3.33,4.42] [4.65,5.35] [5.58,6.67]	[3.33,4.10] [4.90,5.10] [5.90,6.67]
DUW	[3.33,4.39] [4.62,5.39] [5.62,6.67]	[3.33,4.29] [4.71,5.29] [5.71,6.67]	[3.33,4.22] [4.78,5.22] [5.78,6.67]	[3.33,4.10] [4.90,5.10] [5.90,6.67]
DU	[3.00,4.44] [4.56,5.44] [5.56,7.00]	[3.00,4.37] [4.63,5.37] [5.63,7.00]	[3.00,4.31] [4.69,5.31] [5.69,7.00]	[3.00,4.15] [4.85,5.15] [5.85,7.00]

Figure 4.7: *Set-5 to Set-8: Three-peak synthetic fuzzy sets with increasing trough depth.*

In summary, *DU* is the most straightforward method that satisfies all four properties using standard geometric measures, while *DUW* provides a consistent weighted alternative—analogous to the practical use of a weighted average instead of a simple average—particularly for users who prefer to place greater emphasis on the local maxima than on the overall distribution.

It is worth noting that *DUW* produces the same result intervals as *NUW* when the fuzzy set contains only disjoint convex subsets (e.g., Set-4 and Set-8). Therefore, *DUW* is excluded from the user study in the next section to avoid potential confusion for participants, as including both *NUW* and *DUW* would require them to choose between two identical options in some questions.

4.5.2 User Study

In addition to the synthetic data evaluation, a mixed-methods user study is conducted to evaluate how effectively interval-valued defuzzification methods communicate the nuances of non-convex fuzzy sets from the perspective of potential users—such as industry professionals or researchers—who may rely on these representations when making decisions.

This study has two main objectives: (1) to examine which of the three interval-valued defuzzification methods (*IALC*, *NUW*, *DU*) participants tend to select as the most effective for communicating non-convex fuzzy sets, and (2) to gain qualitative insights into when and why users perceive one method as

Table 4.2: Property 2. Core-inclusion

Set	1	2	3	4	5	6	7	8
IALC	✓	✓	✓	✗	✓	✓	✓	✗
NUW	✓	✓	✓	✓	✓	✓	✓	✓
DUW	✓	✓	✓	✓	✓	✓	✓	✓
DU	✓	✓	✓	✓	✓	✓	✓	✓

Table 4.3: Property 3. Peak/Sub-interval Ratio

Set	1	2	3	4	5	6	7	8
$IALC$	2	2	2	2	3	3	3	3
NUW	2	2	1	1	3	3	1	1
DUW	1	1	1	1	1	1	1	1
DU	1	1	1	1	1	1	1	1

Table 4.4: Property 4. Proportionality

Set	1	2	3	4	5	6	7	8
$IALC$	0	0	0	0	0	0	0	0
NUW	0	0	0.71	1	0	0	0.41	1
DUW	1	1	1	1	1	1	1	1
DU	1	1	1	1	1	1	1	1

more effective than others.

Study Participants

A total of 40 participants completed this study in person, recruited through opportunity sampling. The group consisted of 23 academics and 17 industry professionals.

Survey Design

The survey consists of eight questions. Each question is based on one of the eight synthetic fuzzy sets introduced in Section 4.5.1. For each question, a synthetic fuzzy set is shown together with three interval options, generated using the $IALC$, NUW, and DU methods respectively, as illustrated in Fig. 4.8.

To prevent participants from selecting the same method out of habit or for consistency, the three methods were randomly assigned as A, B, or C for each question.

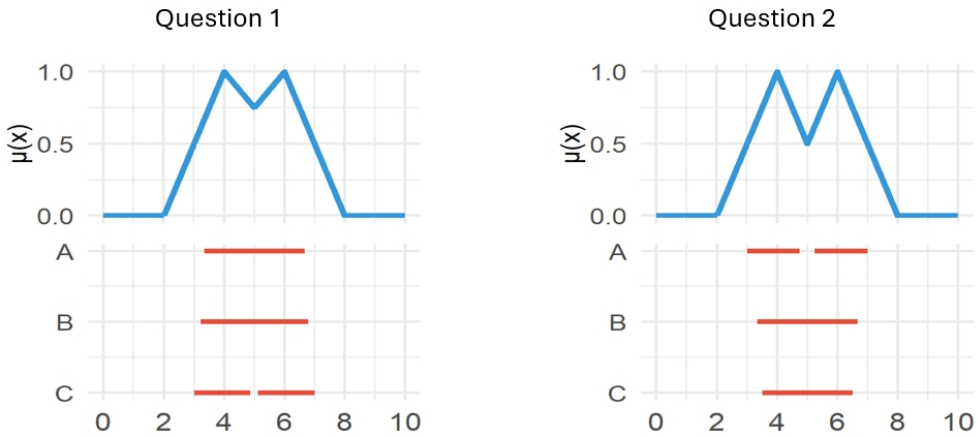


Figure 4.8: *Examples of user study questions.*

Study Procedure

The study procedure was approved by the Ethics Committee at the University of Nottingham. At the beginning of the study, participants were asked to select the interval they believed best *communicates* the displayed distribution (referred to as "distribution" rather than "fuzzy set" for simplicity). They were informed that there is no right or wrong answer. This step addressed Objective 1 by collecting data to identify which interval-valued defuzzification method participants collectively selected as the most effective.

Participants were also instructed to use the Think-Aloud Protocol [107], verbally explaining their reasoning behind each selection. This step provided insight into participants' reasoning processes and addressed Objective 2 by revealing why they perceived one interval as more effective than the others in communicating the distribution.

Therefore, this user study employs a mixed-methods design, collecting both quantitative data (choice responses) and qualitative data (verbal and written explanations).

Results

Table 4.5 presents the quantitative results, showing the proportion of participants who selected each method for each question (proportions are rounded to two decimal places using a largest remainder method to ensure each column sums to 1). The most frequently chosen option is highlighted in green, while the least chosen option is highlighted in red.

Table 4.5: Frequency of method selection.

Question	1	2	3	4	5	6	7	8
$IALC$	0.20	0	0.10	0	0.27	0	0.02	0
NUW	0.10	0.10	0.35	0.47	0.05	0.22	0.23	0.35
DU	0.70	0.90	0.55	0.53	0.68	0.78	0.75	0.65

$IALC$ is the least selected method in 6 out of 8 questions (Set-2 to Set-4, Set-6 to Set-8). NUW is the least selected in 2 questions (Set-1 and Set-5). DU is the most selected method in all 8 questions, indicating a clear participant preference for DU.

Next, the reasons given for why participants favoured or opposed certain methods are summarised.

$IALC$. The main reason participants opposed $IALC$ was its inability to effectively communicate multiple peaks or troughs in the distribution compared to the other two methods. Additionally, a few participants noted that even when only one interval was present, they expected it to widen rather than narrow as the trough area expanded (e.g., from Set-1 to Set-4). Despite this generally negative view, a few participants mentioned that in cases such as Set-1 and Set-5, the trough areas appeared too small and felt more like noise; as a result, they preferred that such minor dips not be reflected and selected $IALC$.

NUW. Although not the most favoured choice in any of the questions, more than one third of participants preferred *NUW* in Set-3, Set-4, and Set-8 because its multiple sub-intervals helped communicate multiple peaks or troughs. Additionally, some participants favoured *NUW* over *DU* in these cases because its narrower sub-interval width better reflected the top peak width of the original distribution. However, participants opposed *NUW* in Set-1 and Set-5, suggesting that it lacked sub-intervals to indicate peaks and was narrower compared to $IALC$, making it less effective in representing the overall peak range.

DU. Participants favoured *DU* the most, with the main reason being that it effectively communicated multiple peaks or troughs through its use of multiple sub-intervals. Compared to *NUW*, when both methods included sub-intervals, some participants preferred *DU* because its wider sub-interval width better reflected the middle area of the original distribution (i.e., membership degree around 0.5). One participant suggested that despite their preference for *DU*, they felt there should be a subtle grey line between sub-intervals to help convey 'continuity' in the distribution, unless the peaks are clearly disjointed, as in Set-4 and Set-8. Such a format would, however, go beyond an interval-valued representation.

Summary and Discussion

Both quantitative and qualitative results indicate that participants were generally in favour of using sub-intervals and gaps to reflect peaks and troughs in the distribution, particularly when the data contained fully disjointed peaks such as Set-4 and Set-8.

Participants preferred certain interval-valued defuzzification methods over

others for two main reasons. First, they preferred cases where the number, position, and width of sub-intervals intuitively correspond to the number, position, and width of peaks in the distribution. Second, they preferred cases where gaps between sub-intervals increase as the trough area becomes larger. Alternatively, if there is only one interval, it should widen accordingly to reflect this change.

Overall, the *DU* method best aligned with participants' preferences due to its ability to intuitively represent peaks and troughs through sub-intervals and the gaps between them. This was further supported by the quantitative results, where *DU* was the most frequently selected method for effectively communicating the given fuzzy sets.

Although *DUW* was not included in the selection for this study, participants' preferences and their reasons for selecting *NUW* in Set-4 and Set-8, as well as their reasons for selecting *DU*, suggest that *DUW* would likely be favoured in a similar manner to *DU*. However, to directly compare participants' preferences between *DU* and *DUW*, further research is needed—ideally in real-world scenarios where participants may exhibit a clearer preference between the average and the weighted average approaches.

4.6 Limitations

Despite the superior performance in both synthetic and user evaluations, the limitations of the proposed methods are discussed below.

4.6.1 Normalisation

Since *DU* and *DUW* rely on IVD and $WIVD$ in Step 3—both of which use normalisation to handle non-normal fuzzy sets—the limitations introduced by

normalisation are inherited by *DU* and *DUW*.

For instance, when a fuzzy set A exhibits multiple peaks of different heights, as depicted in Fig. 4.9a, the two sub-intervals of $\overline{A}$ have the same width as a result of normalisation, which may misleadingly suggest that the original fuzzy set contains two peaks of equal height. This misrepresentation can be concerning, as differences in peak height in fuzzy sets often carry meaningful information—such as varying degrees of preference or agreement—that should be clearly conveyed.

It is emphasised that this issue is inherited from established methods, and that there is a lack of discussion thus far on how the interval's width should be used—whether it should be narrower or wider to reflect non-normality, and why. As a next step, investigating the desirable properties of interval-valued defuzzification for non-normal fuzzy sets would help guide its behaviour in such cases.

In addition, since *DU* and *DUW* do not attempt to reflect peak height, both methods offer a solid foundation for future extensions using complementary methods that go beyond interval representation—such as visualisation approaches that intuitively incorporate peak height information, for example, through colour-coded intervals or annotated height values.

4.6.2 Continuity

Under defuzzification, result continuity generally refers to the property that small changes in the membership values of the output fuzzy set should not cause large changes in the defuzzification results [54]. This feature is particularly important in the case of fuzzy controllers, which require input-output continuity—small changes in input parameters should produce small

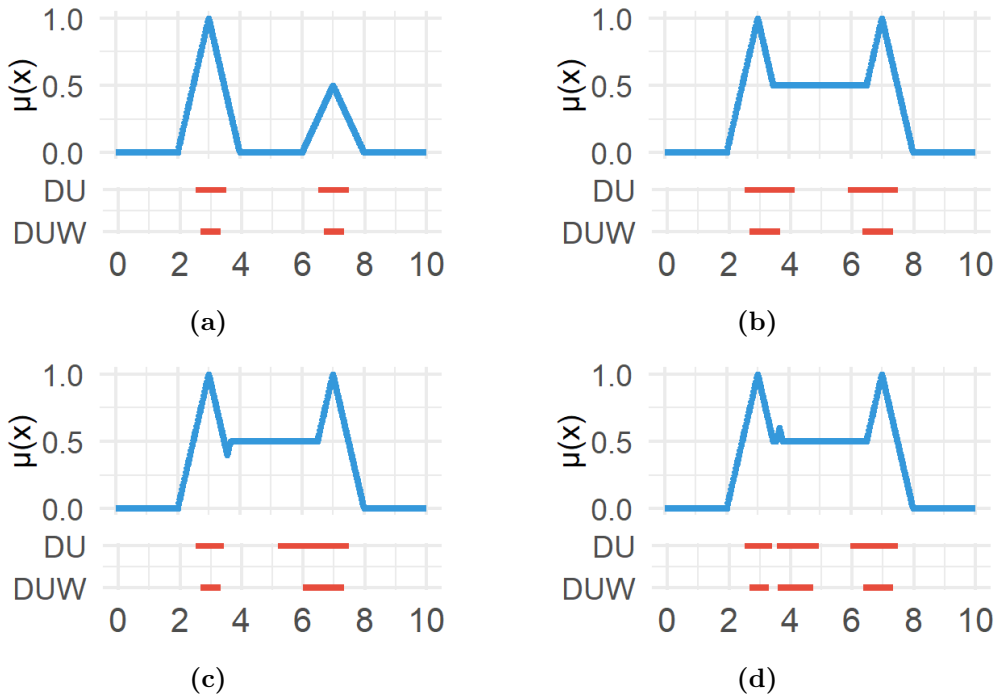


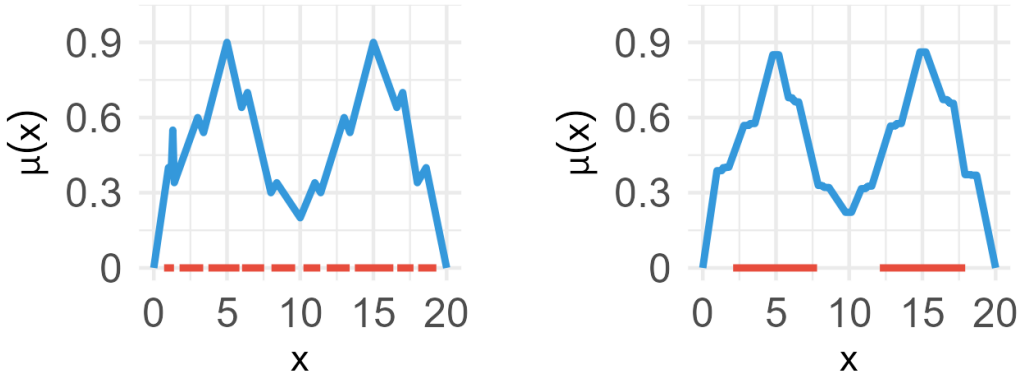
Figure 4.9: *Limitations of the proposed methods: (a) Normalisation issue. (b)–(d) Continuity issue.*

changes in output values [54]. It is therefore desirable for interval-valued defuzzification to also possess this property to ensure consistent and reliable output behaviour.

However, $DU(W)$ lacks continuity in some cases, as depicted in Figs. 4.9b, 4.9c, and 4.9d. This is because $DU(W)$ is highly sensitive to changes in the number and positions of peaks/troughs, as these directly impact how a fuzzy set is decomposed into multiple subsets. While a straightforward solution would be to apply noise-reduction methods during data pre-processing—when such small fluctuations are perceived as unimportant—as discussed in Section 4.6.3 below, methods that address this issue more fundamentally would be desirable in future work.

4.6.3 Noise Reduction

In real-world data-driven applications, a fuzzy set may be 'spiky'—it may exhibit many minor peaks or spikes among its major peaks, e.g., driven by subtle changes in inputs, including unintentional or noise-based fluctuations. In many contexts, it may be more beneficial to focus on the major peaks that signify substantial changes or key features in the data. However, when $DU(W)$ is applied to such 'spiky' fuzzy sets, it reacts to each small variation, resulting in many disjoint intervals across the domain. An example is illustrated in Fig. 4.10a. The resulting sub-intervals do represent features of the fuzzy set in detail but may effectively hide the major trends or characteristics that are of primary interest.



(a) **(b)**

Figure 4.10: *Illustrative examples of (a) a 'spiky' A and $DU(A)$; (b) a 'smoothed' A and $DU(A)$.*

One solution is to treat minor peaks or spikes as noise and apply noise-reduction methods such as the moving average, running mean, or running median. These techniques smooth out small fluctuations, thereby allowing interval-valued defuzzification methods to focus on representing only the major peaks in their interval results. The choice among noise-reduction methods should be guided by the user's specific needs. For example, if the sharp spike in Fig. 4.10a is perceived as noise, methods that are robust to outliers

and extreme values—such as the running median—can be used to obtain a 'smoothed' fuzzy set before applying $DU(W)$. Consequently, the resulting interval focuses on representing the two main peaks, though this may come at the cost of trimming or flattening sharp, meaningful peaks, as shown in Fig. 4.10b.

4.7 Summary and Remarks

Modelling and communicating uncertainty and vagueness in data are critical for informed decision-making. While fuzzy sets provide powerful uncertainty modelling, their complexity often makes them unsuitable for communicating uncertainty to decision makers and wider stakeholders, while also making computation complex. This is particularly the case where the fuzzy sets are derived from real-world data, resulting regularly in fuzzy sets with non-convex and thus complex membership functions.

Traditional single-valued defuzzification approaches, which produce a numeric measure of central tendency, are used widely but fail to track and communicate the uncertainty in the original fuzzy set models. More advanced interval-valued defuzzification methods preserve basic uncertainty information by producing a continuous interval as a measure of central tendency—yet do not have the capability to model or communicate more complex distributions of uncertainty, such as in the case of non-convex membership functions.

In this chapter, desirable properties that interval-valued defuzzification methods should fulfil in order to effectively summarise and communicate complex uncertainty models to decision makers were proposed. Two families of interval-valued defuzzification methods were introduced, with $NU(W)$ reliably satisfying the first two desirable properties, and $DU(W)$ being the first interval-valued defuzzification methods to satisfy the full set of desir-

able properties. This was shown both mathematically and through detailed experiments.

Going beyond technical evaluation, a mixed-methods user study was conducted to explore the human-centric capacity of these methods to faithfully summarise complex fuzzy sets, communicating the underlying uncertainty information to stakeholders and decision makers to a level of nuance not available with other methods. The study showed that DU was the preferred summarisation method among the 40 participants, with qualitative findings suggesting that people generally favour discontinuous intervals to represent non-convexity—particularly when the non-convex features are perceived as prominent.

Future directions include: (1) extensions to indicate peak height information through visual enhancements such as colour-coded intervals or annotated height values, or through interval-valued defuzzification method developments that use interval width to reflect peak-height variations; (2) exploring how the discontinuous intervals produced by $DU(W)$ can be incorporated into existing interval-based ranking and decision-making methods; and (3) further deployment and study of the impact of these tools in real-world decision-making.

Chapter 5

Presentation: Interval-valued Extensions for Service Evaluation

This chapter addresses both the elicitation and presentation stages of the Interval-valued (IV) workflow. At the elicitation stage, an IV survey is conducted to capture uncertainty. At the presentation stage, conventional service evaluation methods are extended to accommodate IV data, so that uncertainty captured in the responses can be retained in decision-support outputs rather than reduced to Single-Valued (SV) summaries. To this end, three widely used service evaluation methods, namely Gap analysis, Importance–Performance Analysis (IPA), and Customer Satisfaction Index (CSI), are extended for the IV survey data. The resulting methods are then demonstrated through a real-world study in the UK rail context, in support of uncertainty-aware decision-making.

Section 5.1 introduces the motivation to capture and present uncertainty in decision-support methods for service quality research. Section 5.2 briefly recalls the related work, including Gap analysis, IPA and CSI, as well as the relevant IV data analysis methods. Section 5.3 presents the proposed

IV extensions of these three decision-support methods. Section 5.4 describes the UK rail survey study, including the study design, participants, and data collection. Sections 5.5 and 5.6 report and discuss the findings, and Section 5.7 outlines the limitations. Abbreviations and notations are summarised in the List of Abbreviations and the List of Notations.

5.1 Introduction

Service evaluation is an important process in the service industry [86, 108], as it enables providers to understand customer needs and expectations, identify service shortfalls and priorities for improvement, and thereby enhance customer satisfaction, loyalty, and long-term profitability [86, 108].

Services are inherently more difficult to evaluate than goods because of characteristics, such as intangibility, heterogeneity, and inseparability [86], as detailed in Section 2.5.1. As a result, service evaluation relies heavily on customers' subjective perceptions and judgements [89, 90]. Response formats such as Likert scales are predominantly used in surveys [95, 109] to collect customer responses on key constructs in this field (e.g., customer expectations, perceptions, and satisfaction; Section 2.5.2). Their widespread use is supported by their convenience and standardised procedure for collecting subjective information from large numbers of customers, together with the relative ease of analysing the resulting survey data using well-established statistical methods [11, 110, 111].

Survey data analysis is often based on evaluation methods designed to convert customer responses into actionable insights, with outputs typically presented in highly summarised and intuitive forms for decision-making. These can be broadly divided into aggregated and disaggregated approaches. The former seeks to combine different attributes into an overall index, as in the CSI, and

is typically used for benchmarking against competitors or for comparisons over time. The latter analyses service attributes individually, as in Gap Analysis and IPA, and is often used for prioritisation.

While it is undoubtedly true that these data collection and evaluation methods have provided valuable insights across a wide range of service contexts over recent decades, current practice has only limited ability to communicate uncertainty. This ability may well be necessary given the difficulty of directly measuring many service constructs and the resulting reliance on subjective customer responses.

For response collection, uncertainty arises from both the question itself and the respondent. On the question side, uncertainty may arise when a question asks multiple things at the same time, when its wording can be interpreted in different ways, or when it explicitly or implicitly asks for a tolerance or flexibility range. On the respondent side, uncertainty may arise when a customer is uncertain between adjacent predefined choices, when memories have faded or the experience was not closely attended to in the first place, or when the customer's preference or judgement is itself a flexible range. Across these cases, the common issue is that the measurement scale is more precise than what is being measured, which leads to distorted data with false precision. One well-documented example in service evaluation is the construct of customer expectation. Given the ambiguity in its definition, it is better represented as a range, i.e., a 'Zone of Tolerance', between the *adequate service* and the *desired service* rather than as a precise point [112].

For response analysis, current evaluation methods operate on and produce SV data as output, without presenting uncertainty, which may lead to an illusion of precision for decision makers. This analysis issue can also be seen in the 'Zone of Tolerance' example: even when both bounds are collected, subsequent

analysis may summarise them into SV indicators [113], despite desired service and adequate service carrying distinct meanings within the 'Zone of Tolerance' framework [112]. Such compression removes the distinct meanings carried by the two bounds. This can be problematic because, across different service contexts and attributes, the managerial significance of adequate service and desired service may differ. For some aspects, meeting adequate service may be sufficient, whereas for other aspects, customers may expect performance much closer to desired service. When this information is reduced to a single score, decision makers may be unable to distinguish between these cases and their different implications.

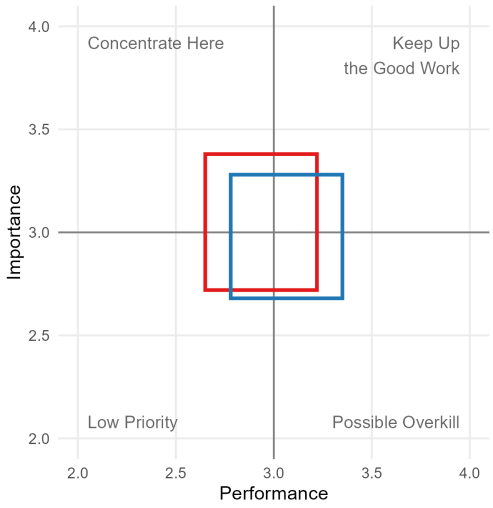
IPA provides another example. As a widely used method in service evaluation, it directly maps attribute-level means to prioritisation quadrants. Fig. 5.1a shows a traditional IPA result, where two service attributes receive completely different prioritisation (i.e., Concentrate Here versus Possible Overkill) despite their means differing only marginally. If uncertainty is presented in the output—for example, in IV representations—the two attributes may cross multiple quadrants and instead be nearly indistinguishable (Fig. 5.1b) or clearly separated (Fig. 5.1c). It is very realistic that a decision-maker, presented with this richer information, would make a different decision on prioritisation in practice.

While the uncertainty in response can now be captured using IV scales as introduced in earlier chapters, conventional evaluation methods are unable to directly operate on IV data, nor can they present the uncertainty in responses to support uncertainty-aware decision-making.

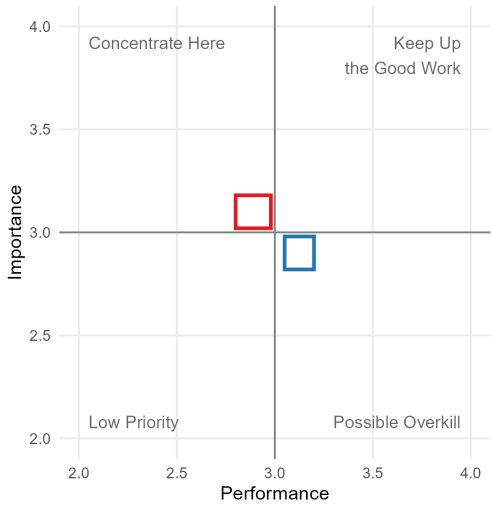
Accordingly, this chapter makes the following contributions:

1. Gap analysis, IPA, and CSI are extended to IV data, enabling uncertainty to be represented directly in decision-support outputs for well-informed

(a) *Traditional IPA*



(b) *IV IPA with overlap attributes.*



(c) *IV IPA with separate attributes.*

Figure 5.1: *Traditional IPA and IV IPA examples.*

decision-making.

2. The added insights from these extensions are demonstrated through a UK rail passenger study, in which comparable SV and IV survey data are collected and analysed.

The collected rail survey data are available on GitHub at https://github.com/Carina-YuZhao/.

5.2 Related Work

This section recalls three methods that are widely used in service evaluation, namely Gap analysis, IPA, and CSI. A few highly relevant IV data analysis methods are also recalled, as they are essential for the methods extension described in Section 5.3.

5.2.1 Service Evaluation Methods

Let the service evaluation involve $k \geqslant 2$ attributes indexed by $j \in \{1, \ldots, k\}$.

Gap analysis assesses service shortfalls by computing the gap for each attribute j [86]:

$$\text{Gap}_j = \hat{P}_j - \hat{E}_j, \tag{5.1}$$

where $\hat{E}_j$ and $\hat{P}_j$ denote the mean expectation and mean perception for attribute j; larger negative gaps indicate lower perceived quality.

IPA maps attributes on a grid of mean importance (y-axis) versus mean performance (x-axis) [34], where $\hat{I}_j$ and $\hat{P}_j$ denote the mean importance and mean performance for attribute j. The quadrant boundaries are set at the grand means across all k attributes, denoted by $\hat{c}_P$ for the grand mean of

performance and $\hat{c}_I$ for the grand mean of importance. These partition the plane into Concentrate Here (Q1-CH, $\hat{I}_j > \hat{c}_I$, $\hat{P}_j < \hat{c}_P$), Keep Up the Good Work (Q2-KU, $\hat{I}_j > \hat{c}_I$, $\hat{P}_j \geqslant \hat{c}_P$), Low Priority (Q3-LP, $\hat{I}_j \leqslant \hat{c}_I$, $\hat{P}_j < \hat{c}_P$), and Possible Overkill (Q4-PO, $\hat{I}_j \leqslant \hat{c}_I$, $\hat{P}_j \geqslant \hat{c}_P$).

CSI provides an importance-weighted satisfaction index, commonly used for benchmarking over time or across providers [96, 114]:

$$\text{CSI} = \frac{\sum_{j=1}^{k} \hat{I}_j \cdot \hat{S}_j}{\sum_{j=1}^{k} \hat{I}_j}, \tag{5.2}$$

where $\hat{I}_j$ and $\hat{S}_j$ are the mean importance and satisfaction for attribute j.

5.2.2 Interval-valued Data Analysis

The IV extensions proposed in Section 5.3 draw on three mathematical constructs reviewed in Chapter 2, recalled here for reference.

Interval subtraction (Section 2.3.1) follows the principle that the result contains all possible outcomes from any pair of elements drawn from the input intervals [81]:

$$\overline{A} - \overline{B} = [A^- - B^+,\ A^+ - B^-]. \tag{5.3}$$

Interval Weighted Average (IWA) with interval weights (Section 2.3.2) extends the standard weighted average to the case where both the values and weights are intervals [83]:

$$\text{IWA}(\{\overline{A}_i\}; \{\overline{w}_i\}) = \left[\min_{w_i \in [w_i^-, w_i^+]} \frac{\sum_i A_i^- w_i}{\sum_i w_i},\ \max_{w_i \in [w_i^-, w_i^+]} \frac{\sum_i A_i^+ w_i}{\sum_i w_i} \right]. \tag{5.4}$$

Fréchet mean (Section 2.3.3) is the interval minimising the average squared

distance to a sample of IV responses $\{\overline{X}_i\}_{i=1}^n$, where each $\overline{X}_i = [X_i^-, X_i^+]$ [48]:

$$\overline{X}^* = \arg\min_{\overline{X}} \frac{1}{n} \sum_{i=1}^{n} d^2(\overline{X}, \overline{X}_i), \tag{5.5}$$

where d is a chosen IV distance metric, such as the Hausdorff distance or the θ-distance. When the Hausdorff distance is used, the mean is denoted $\overline{X}_H^*$; when the θ-distance is used, it is denoted $\overline{X}_\theta^*$.

5.3 Methods

This section extends Gap analysis, IPA, and CSI to their IV counterparts. Let the service evaluation involve $k \geqslant 2$ attributes indexed by $j \in \{1, \ldots, k\}$. The arithmetic mean of SV responses is replaced by the Fréchet mean of IV responses, interval subtraction replaces the arithmetic subtraction in Gap analysis, and the IWA replaces the weighted average in CSI. The choice of distance metric for the Fréchet mean—Hausdorff (d_H) or θ-distance (d_θ)—should be guided by the relative properties of the two metrics discussed in Chapter 3. In the empirical analyses reported in Section 5.4, the Hausdorff distance is used throughout.

5.3.1 Gap Analysis

When both expectation and perception responses are IV, the expectation response can be interpreted as spanning the adequate service level to the desired service level, following the zone of tolerance concept [115]. Correspondingly, the perception response spans the worst to the best perception. Let $\overline{E}_j^* = [E_j^-, E_j^+]$ and $\overline{P}_j^* = [P_j^-, P_j^+]$ denote the Fréchet means (Eq. 5.5) of the expectation and perception responses for attribute j. Since $\overline{P}_j^*$ and $\overline{E}_j^*$ are both intervals, the subtraction in Eq. 5.1 is replaced by interval subtraction

(Eq. 5.3):

$$\overline{\text{Gap}}_j = \overline{P}_j^* - \overline{E}_j^* = [P_j^- - E_j^+,\ P_j^+ - E_j^-]. \tag{5.6}$$

The gap is therefore itself an interval. The lower bound $P_j^- - E_j^+$ is the most pessimistic comparison (worst perception minus desired service level), and the upper bound $P_j^+ - E_j^-$ is the most optimistic (best perception minus adequate service level). If both bounds are negative, even the best perception falls short of the adequate service level. If both bounds are positive, even the worst perception exceeds the desired service level. If the interval contains zero, $\overline{P}_j^*$ and $\overline{E}_j^*$ overlap or touch, indicating that perception may fall short of, meet, or exceed expectation under different endpoint combinations.

5.3.2 Importance–Performance Analysis

When both performance and importance responses are IV, the performance response can be interpreted as spanning the worst to the best perceived performance, and the importance response as spanning a lower to a higher assessed importance. Let $\overline{P}_j^* = [P_j^-, P_j^+]$ and $\overline{I}_j^* = [I_j^-, I_j^+]$ denote the Fréchet means (Eq. 5.5) of the performance and importance responses for attribute j. Each attribute is therefore represented as a rectangle spanning $\overline{P}_j^*$ on the x-axis and $\overline{I}_j^*$ on the y-axis.

The SV grand means $\hat{c}_P$ and $\hat{c}_I$ recalled in Section 5.2.1 are replaced by the Fréchet means of the attribute-level Fréchet means, denoted by $\overline{c}_P^* = [c_P^-, c_P^+]$ for performance and $\overline{c}_I^* = [c_I^-, c_I^+]$ for importance. Accordingly, each boundary line becomes a shaded band spanning the bounds of the corresponding interval. A rectangle fully contained within a single quadrant indicates a stable prioritisation, whereas a rectangle that crosses a boundary band or spans multiple quadrants indicates that the prioritisation is sensitive to uncertainty.

5.3.3 Customer Satisfaction Index

When both satisfaction and importance responses are IV, the satisfaction response can be interpreted as spanning the lowest to the highest satisfaction, and the importance response as spanning a lower to a higher assessed importance. Let $\overline{S}_j^* = [S_j^-, S_j^+]$ and $\overline{I}_j^* = [I_j^-, I_j^+]$ denote the Fréchet means (Eq. 5.5) of the satisfaction and importance responses for attribute j. The weighted average in Eq. 5.2 is then replaced by the IWA with interval weights in Eq. 5.4, using $\{\overline{S}_j^*\}_{j=1}^k$ as the values and $\{\overline{I}_j^*\}_{j=1}^k$ as the weights, yielding $\overline{\text{CSI}} = [\text{CSI}^-, \text{CSI}^+]$ where

$$\text{CSI}^- = \min_{I_j \in [I_j^-, I_j^+]} \frac{\sum_{j=1}^k S_j^- \, I_j}{\sum_{j=1}^k I_j}, \qquad \text{CSI}^+ = \max_{I_j \in [I_j^-, I_j^+]} \frac{\sum_{j=1}^k S_j^+ \, I_j}{\sum_{j=1}^k I_j}. \tag{5.7}$$

Since CSI is primarily used for benchmarking across competitors or over time, comparing two or more IV CSI indices provides richer information through their lower and upper bounds. For example, two IV CSI intervals may have the same midpoint, while one has a higher lower bound and a lower upper bound.

5.4 Application: A UK Rail Survey Study

This section presents a UK rail survey study with the overall aim of evaluating the practical value of the IV extensions introduced above in a real-world service evaluation context. To support this aim, the study has two objectives: first, to collect comparable SV and IV data from UK rail passengers; second, to apply the traditional methods to the SV data and the corresponding extended methods to the IV data, in order to explore and demonstrate the added value of the extensions.

5.4.1 Study Design

Following common practice in this field, the study collects UK rail passengers' ratings on two constructs, importance and satisfaction. Both constructs are measured using the same set of 14 rail service items, with the importance question worded as "how important is:" and the satisfaction question worded as "how satisfied are you with:". To reduce respondent burden, the study uses a between-subjects design for the SV and IV response formats. Thus, each participant is assigned to either the SV or IV version of the survey, while providing both importance and satisfaction ratings for the same set of items.

In the SV version, participants respond using a Likert scale, whereas in the IV version they respond using interval-valued scales, drawing an ellipse on a continuous scale, with the ellipse's endpoints recorded as the interval bounds. Interval-valued scales are introduced in Section 2.4.4 and illustrated in Fig. 5.2. In both cases, the scale ranges from "not important" to "very important" for importance and from "not satisfied" to "very satisfied" for satisfaction. The 14 items are adapted from the East Midlands Railway Survey and are listed in Table 5.1. Importance ratings are mandatory for all 14 items. For satisfaction, six items (marked * in Table 5.1) include a prompt indicating that respondents may skip the item if it is not applicable to their journey.

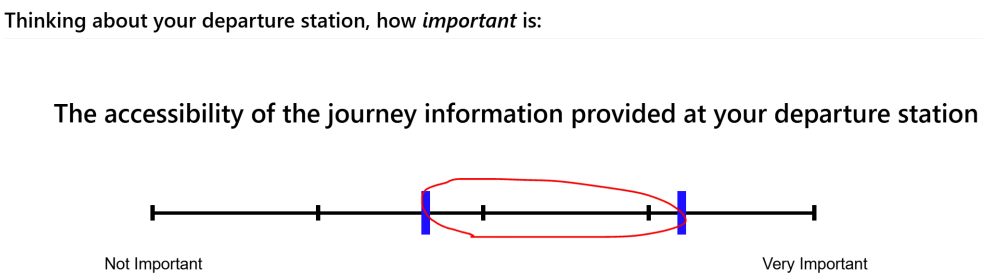


Figure 5.2: *Example of a response to a survey item using the interval-valued scale in DECSYS.*

Table 5.1: Survey items.

ID	Item
Station attributes (1–6)	
1	Accessibility of information about your journey at the station
2	Having somewhere to wait at the station that is comfortable
3	Cleanliness of the station
4	Availability of staff at the station
5	*Attitude of the station staff
6	*Cleanliness of the toilets in the station
On-board attributes (7–14)	
7	Ease of finding a seat on the train
8	Cleanliness of the train
9	*Quality of the WiFi on the train
10	*Amount of space when sitting in a seat on the train
11	Availability of staff on the train
12	*Attitude of staff on the train
13	*Cleanliness of the train toilets
14	Provision of information by the staff on the train

5.4.2 Study Participants

The study involved 166 rail passengers in the UK who had travelled within the seven days prior to completing the survey. Participants were recruited via the online platform Prolific and divided into two groups (85 SV and 81 IV). Details are presented in Table 5.2.

To ensure that all expected frequencies met the minimum requirement of 5 for the chi-square test, Prefer not to say and Other responses were excluded as they cannot be merged into other categories. The following categories were combined before testing: for age, 65 years and older was merged with 55–64 years; for travel purpose, Personal Matters and Essential shopping were merged; and for travel frequency, all categories below Monthly were merged. Chi-square tests then indicated no statistically significant differences between the two groups across gender, age, travel purpose, or travel frequency (all $p > .3$). The two groups were therefore considered comparable for subsequent analysis.

Table 5.2: Demographic summary comparing the SV and IV groups.

Category	**Group / Option**	**SV**	**SV (%)**	**IV**	**IV (%)**
Gender	Female	45	52.9%	44	54.3%
	Male	40	47.1%	35	43.2%
	Other	–	–	1	1.2%
	Prefer not to say	–	–	1	1.2%
Age	18–24 years	7	8.2%	10	12.3%
	25–34 years	31	36.5%	32	39.5%
	35–44 years	18	21.2%	10	12.3%
	45–54 years	17	20.0%	13	16.0%
	55–64 years	10	11.8%	11	13.6%
	65 years and older	2	2.4%	3	3.7%
	Prefer not to say	–	–	2	2.5%
Purpose	Friends / Family	23	27.1%	16	19.8%
	Commuting	19	22.4%	23	28.4%
	Leisure	18	21.2%	24	29.6%
	Business / Work Travel	16	18.8%	14	17.3%
	Personal Matters	8	9.4%	3	3.7%
	Essential shopping	1	1.2%	1	1.2%
Frequency	Daily	9	10.6%	12	14.8%
	Weekly	36	42.4%	41	50.6%
	Monthly	27	31.8%	18	22.2%
	More than six times a year	7	8.2%	8	9.9%
	Less than six times a year	4	4.7%	1	1.2%
	Very rarely or never	1	1.2%	1	1.2%
	Prefer not to say	1	1.2%	–	–

5.4.3 Data Collection Procedure

The study was approved by the Ethics Committee at the University of Nottingham School of Computer Science. The survey was administered online using DECSYS, free software designed for creating and administering IV surveys [27].

Participants were randomly assigned one of the two versions of the survey. All participants received an initial information sheet containing general details about the nature of the study, the estimated completion time (10 minutes), their right to withdraw at any point, and assurances of anonymity. They then received a second information sheet specific to their assigned format, with instructions adapted accordingly for the SV and IV scales. After reviewing the information sheets, participants confirmed that they had read and understood the study information and consented to proceed.

5.4.4 Data Analysis Procedure

The analysis is conducted on the SV and IV data in parallel, so that the two response formats can be compared, with the SV results serving as a baseline.

First, descriptive statistics are computed for both data types, as this is a standard initial step in survey analysis and provides the mean-based inputs required for the three decision-support methods. For the SV data, item-wise means and variances are calculated. For the IV data, Fréchet means and variances under the Hausdorff distance are calculated.

Second, Gap analysis, IPA, and CSI are applied to the SV data, while their IV extensions (Section 5.3) are applied to the IV data.

It should be noted that, in this study, the collected importance ratings are used as expectation ratings in Gap analysis, while the collected satisfaction ratings are used as perception/performance ratings in Gap analysis and in IPA respectively. This follows common practice in service research, as it allows multiple evaluation methods to be applied to the same dataset, with each providing a complementary perspective. The limitations of this practice are discussed further in Section 5.7.

5.5 Results

While the survey collected responses from 85 SV and 81 IV participants, Items S5, S6, S9, S10, S12, and S13 have lower response counts because some respondents chose to skip these items as “not applicable”.

5.5.1 Descriptive Statistics

Tables 5.3 and 5.4 present the descriptive statistics for satisfaction and importance attributes, respectively. Items 1–6 relate to station attributes and Items 7–14 relate to on-board attributes; where item names recur (e.g., Cleanliness, Staff Availability, Staff Attitude, Toilet Cleanliness), they refer to station and on-board contexts respectively.

Table 5.3: Descriptive statistics for satisfaction items.

	N		Mean		Variance	
Item	IV	SV	IV	SV	IV	SV
S1 Journey Info	81	85	[3.64, 4.17]	3.93	1.15	0.71
S2 Wait Area	81	85	[2.99, 3.50]	3.47	1.38	0.99
S3 Cleanliness	81	85	[3.39, 3.85]	3.61	1.43	0.93
S4 Staff Avail.	81	85	[2.98, 3.43]	3.35	1.66	1.23
S5 Staff Att.	73	75	[3.50, 4.01]	3.77	1.24	1.02
S6 Toilet Clean.	67	68	[2.62, 3.38]	3.03	1.76	1.43
S7 Seat Avail.	81	85	[3.42, 3.80]	3.65	1.71	2.18
S8 Cleanliness	81	85	[3.40, 3.82]	3.71	1.34	1.11
S9 WiFi Qual.	70	75	[2.56, 2.95]	3.00	1.57	1.73
S10 Seat Space	79	84	[3.35, 3.73]	3.58	1.07	1.09
S11 Staff Avail.	81	85	[2.85, 3.35]	3.29	1.43	1.21
S12 Staff Att.	65	73	[3.31, 3.82]	3.89	1.46	1.02
S13 Toilet Clean.	60	65	[2.60, 3.20]	3.00	1.74	1.72
S14 Info Prov.	81	85	[3.29, 3.87]	3.80	1.16	0.95

Across both attribute sets, Figs. 5.3 and 5.4 show that most SV means fall within the corresponding IV mean intervals, although some lie above the upper bound. The SV means also tend to sit above the IV midpoints and closer to the upper bounds, indicating a systematic upward offset. This effect is stronger for importance than for satisfaction: for importance, five attributes (I8, I9, I10, I11, I13) have SV means above the upper bound of the IV interval, whereas for satisfaction this occurs for only two attributes (S9, S12).

IV variances are larger than SV variances for both satisfaction and importance attributes (Tables 5.3 and 5.4), with the exception of S7 Seat Availability (IV: 1.71, SV: 2.18) and S9 WiFi Quality (IV: 1.57, SV: 1.73).

Table 5.4: Descriptive statistics for importance items.

	N		Mean		Variance	
Item	IV	SV	IV	SV	IV	SV
I1 Journey Info	81	85	[3.79, 4.28]	4.18	1.32	0.72
I2 Wait Area	81	85	[3.49, 4.10]	4.00	1.35	0.95
I3 Cleanliness	81	85	[3.64, 4.11]	3.95	0.75	0.71
I4 Staff Avail.	81	85	[3.23, 3.71]	3.67	1.42	1.37
I5 Staff Att.	81	85	[3.80, 4.19]	4.07	1.18	1.11
I6 Toilet Clean.	81	85	[3.80, 4.31]	4.29	0.99	1.09
I7 Seat Avail.	81	85	[4.27, 4.67]	4.59	0.65	0.36
I8 Cleanliness	81	85	[3.89, 4.35]	4.47	0.81	0.44
I9 WiFi Qual.	81	85	[3.22, 3.58]	3.73	2.14	1.79
I10 Seat Space	81	85	[3.77, 4.17]	4.22	0.78	0.46
I11 Staff Avail.	81	85	[2.67, 3.00]	3.19	1.63	1.56
I12 Staff Att.	81	85	[3.73, 4.15]	4.12	1.28	1.06
I13 Toilet Clean.	81	85	[3.92, 4.31]	4.44	1.14	0.80
I14 Info Prov.	81	85	[3.53, 3.93]	3.79	1.34	1.19

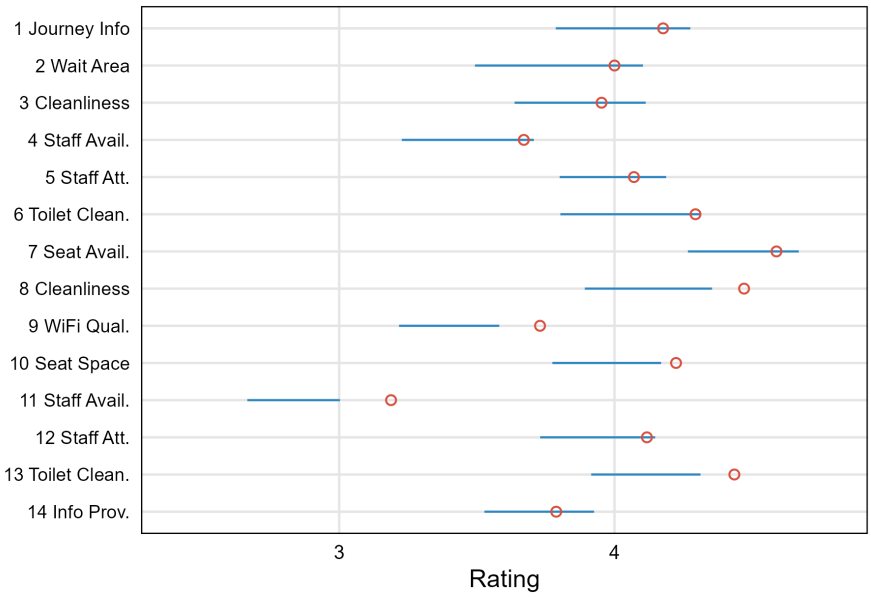


Figure 5.3: *Importance means: IV means as blue segments, SV means as red circles.*

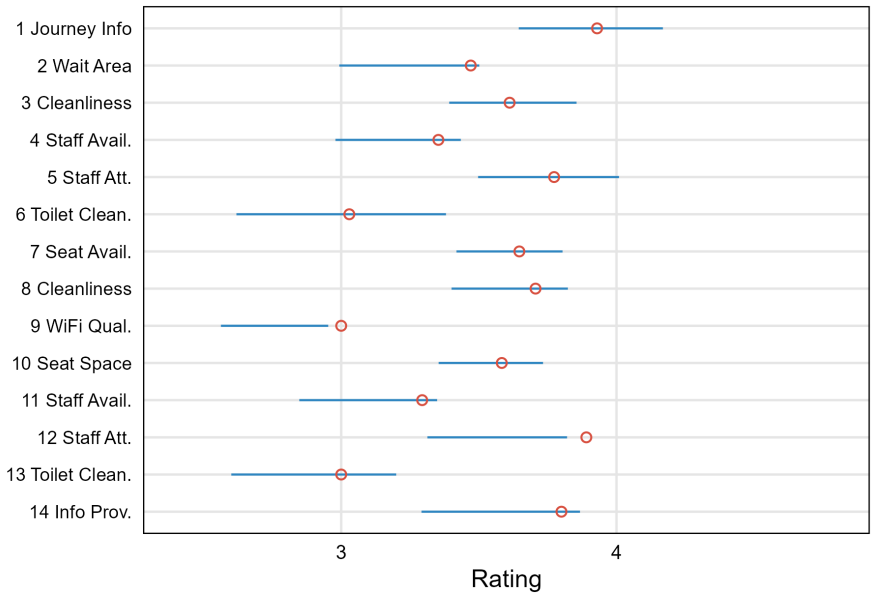


Figure 5.4: *Satisfaction means: IV means as blue segments, SV means as red circles.*

5.5.2 Decision-support Analysis

Gap Analysis. Table 5.5 presents and Fig. 5.5 visualises the SV and IV gap analysis results.

Table 5.5: Gap scores (Satisfaction − Importance) for SV and IV data.

Item	SV Gap	IV Gap
1 Journey Info	−0.25	[−0.63, +0.38]
2 Wait Area	−0.53	[−1.11, +0.01]
3 Cleanliness	−0.34	[−0.72, +0.22]
4 Staff Avail.	−0.32	[−0.73, +0.21]
5 Staff Att.	−0.30	[−0.69, +0.21]
6 Toilet Clean.	−1.26	[−1.69, −0.42]
7 Seat Avail.	−0.94	[−1.25, −0.46]
8 Cleanliness	−0.76	[−0.95, −0.07]
9 WiFi Qual.	−0.73	[−1.02, −0.26]
10 Seat Space	−0.64	[−0.82, −0.04]
11 Staff Avail.	+0.11	[−0.15, +0.68]
12 Staff Att.	−0.23	[−0.83, +0.09]
13 Toilet Clean.	−1.44	[−1.71, −0.72]
14 Info Prov.	+0.01	[−0.63, +0.34]

Under SV, all gaps except those for Items 11 and 14 are negative. Under IV, six attributes (Items 6, 7, 8, 9, 10, and 13) have gap intervals that are entirely

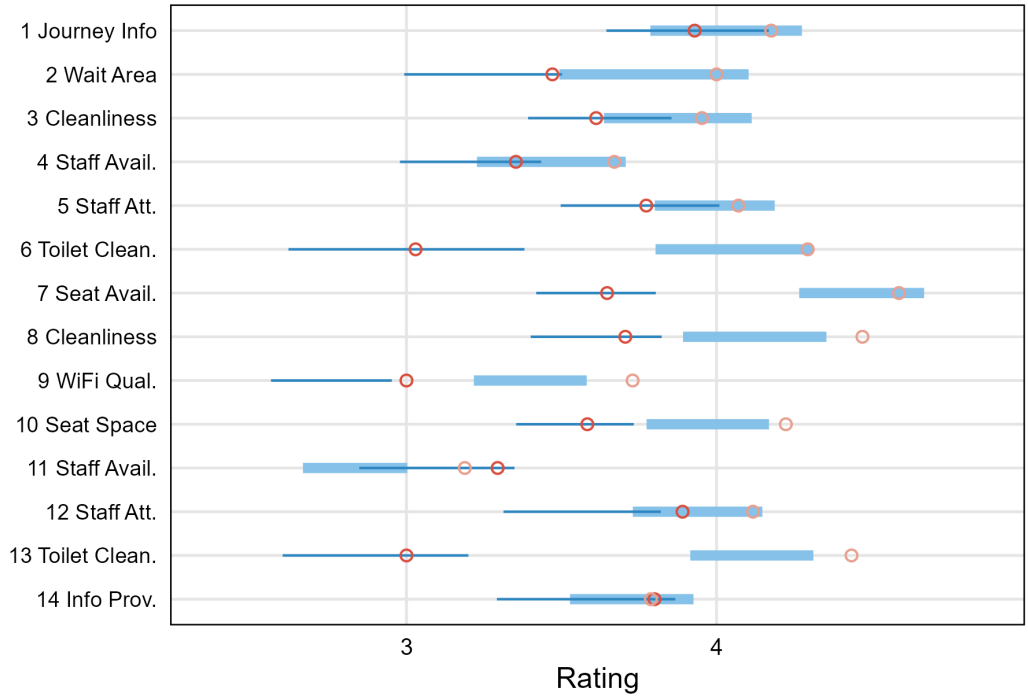


Figure 5.5: *Gap analysis. Dark represents expectation and light represents perception. IV means are shown as interval segments, and SV means are shown as red circles.*

negative. The remaining eight attributes (Items 1, 2, 3, 4, 5, 11, 12, and 14) have gap intervals that cross zero, indicating overlap between the importance and satisfaction intervals, as shown in Fig. 5.5. No item has a gap interval that is entirely positive.

Importance–Performance Analysis. Figs. 5.6 and 5.7 present the SV IPA and IV IPA respectively.

For SV IPA (Fig. 5.6), the Q1-CH quadrant includes attributes 6 and 13. The Q2-KU quadrant includes attributes 1, 7, 8, 10, and 12. The Q3-LP quadrant includes attributes 2, 4, 9, and 11. The Q4-PO quadrant includes attributes 3, 5, and 14.

For the IV IPA (Fig. 5.7), most attribute rectangles overlap with the IV boundaries. Attribute 9 is the only exception, as it is fully contained within Q3-LP. Specifically, attributes 7 and 1 extend from Q2-KU towards Q1-CH

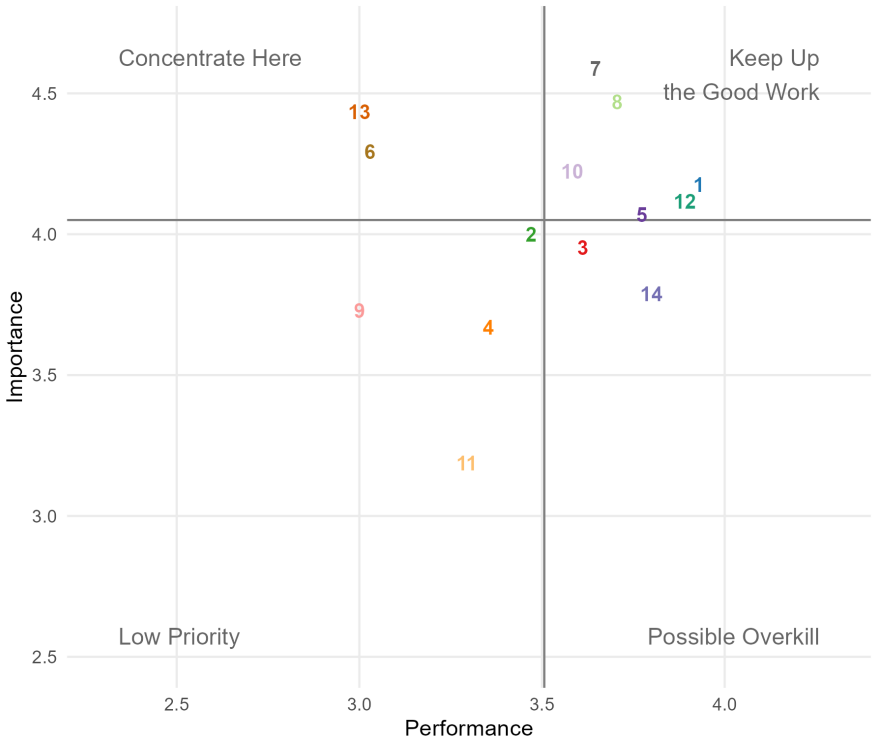


Figure 5.6: *SV IPA. Each attribute is a SV point; quadrant lines are drawn at the grand means.*

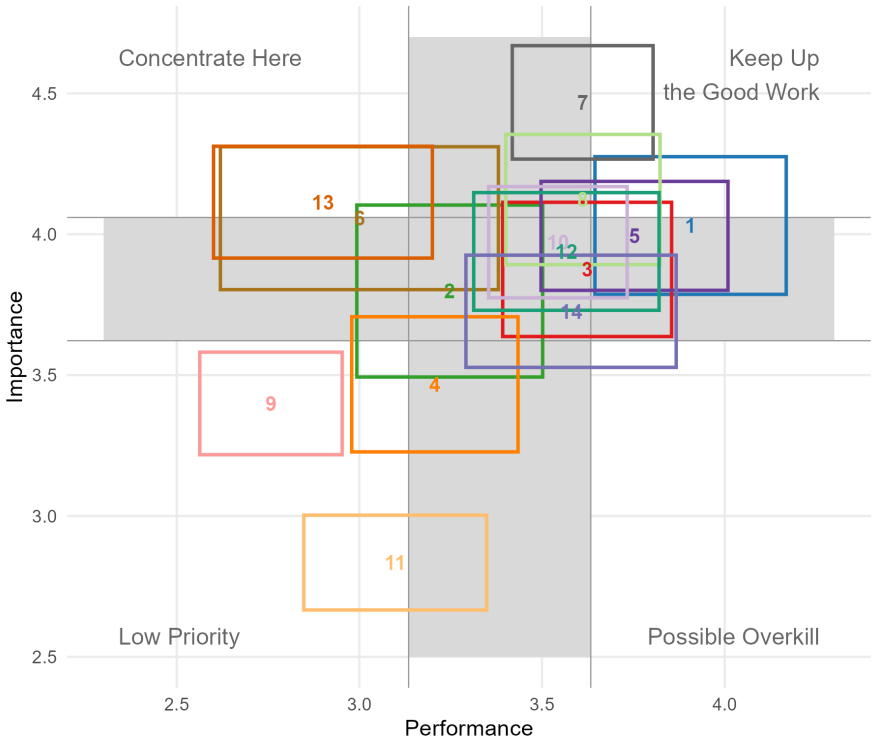


Figure 5.7: *IV IPA. Each attribute is a rectangle; quadrant lines (shaded bands) drawn at the IV grand means.*

and Q4-PO, respectively, while attributes 2 and 11 extend from Q3-LP towards Q1-CH and Q4-PO, respectively. The remaining attributes (3, 4, 5, 6, 8, 10, 12, 13, and 14) all overlap with the central IV boundaries.

Customer Satisfaction Index. Table 5.6 presents the results.

Table 5.6: SV CSI and IV CSI.

Format	CSI
SV	3.51
IV	[3.13, 3.67]

The SV CSI is 3.51, which falls within the IV CSI interval [3.13, 3.67].

5.6 Discussion

The SV and IV results are broadly consistent: most SV means fall within the corresponding IV mean intervals, the SV CSI falls within the IV CSI interval, all SV gaps fall within the corresponding IV gaps, and attributes occupy broadly similar positions under both SV and IV IPA. This consistency suggests that the IV extensions produce results that are compatible with, rather than contradictory to, conventional SV analysis.

At the same time, the IV results provide additional information that is not available from the SV analysis alone. The following subsections discuss the specific insights gained from each IV extension.

5.6.1 Descriptive Statistics

Although most SV means fell within the corresponding IV mean intervals, there are exceptions where the SV means fell beyond the upper bounds of the interval. Moreover, even when they did fall within the intervals, they

were typically located above the IV midpoints and closer to the upper bounds. This upward shift was more pronounced for importance than for satisfaction, suggesting that respondents may express importance at a higher level when using the SV format than when using the IV format. One possible explanation is that, when asked to give a single value for importance, respondents may try to avoid understating their judgement and therefore gravitate towards higher importance ratings. The IV format may help address this by allowing respondents to express lower and upper bounds separately and to select values between scale points, making it possible to indicate high importance without necessarily assigning a maximal single rating.

Variance results show a clearer difference between the two formats. For nearly all attributes, the IV variances are larger than the corresponding SV variances, with S7 Seat Availability and S9 WiFi Quality being the only exceptions. This suggests that, at the descriptive level, the IV format preserves additional response variation that is not captured when responses are represented by a single value.

5.6.2 Gap Analysis

Under the SV format, all gaps except those for Items 11 and 14 are negative, giving the impression that 12 of the 14 attributes fall short of their corresponding importance levels.

Under the IV format, by comparison, only six attributes have gap intervals that are entirely negative. This emphasises that these attributes warrant particular attention, as even the most optimistic interpretation of the gap (upper bound of satisfaction minus lower bound of importance) remains negative.

In addition, the interval gap plot conveys more detailed information than

the IV gap values alone. For example, Item 11 is the only attribute for which both the lower and upper bounds of satisfaction are higher than the corresponding bounds of importance. This indicates that, although the most pessimistic subtraction-based gap remains negative, the satisfaction interval is shifted upward relative to the importance interval at both endpoints. In some application contexts, this may already be sufficient to regard the attribute as adequately matching its assessed importance.

Furthermore, it is intuitive from the interval gap plot that Items 1, 3, 5, and 14 show substantial overlap between the satisfaction and importance intervals. This suggests that, rather than indicating a clear and consistent shortfall, these attributes may reflect cases where satisfaction matches importance in some instances but not in others. For these attributes, further investigation into where satisfaction matches importance and where it falls short may be more informative than interpreting them simply as underperforming.

5.6.3 Importance–Performance Analysis

In contrast to the SV IPA, the IV IPA shows that some attributes cannot be treated as belonging clearly to a single quadrant once uncertainty is taken into account.

A particular example is Attribute 7. Under the SV IPA, it falls within Q2-KU, whereas the IV IPA suggests that its lower bound may extend towards Q1-CH. Given that its importance is notably higher than that of all other attributes, this suggests that effort is still needed to manage its lower bound of performance.

In addition, under the SV IPA, Attributes 3 and 14 are placed in Q4-PO. Under the IV IPA, however, they do not appear to belong clearly to Q4-

PO, and overlap substantially with nearby attributes such as 5, 8, and 12, which themselves also overlap heavily with the dividing boundary lines. This suggests that, for all of these attributes, other indicators may be needed rather than assigning them directly to a single quadrant.

5.6.4 Customer Satisfaction Index

The IV CSI interval [3.13, 3.67] captures the range of possible aggregate satisfaction levels given the uncertainty in both satisfaction and importance responses. The SV CSI is 3.51 and falls above the midpoint of this interval, suggesting that the SV responses may reflect a mildly optimistic tendency relative to the full range of uncertainty expressed in the IV responses.

5.6.5 Insights into Rail Service

Under both SV and IV, the gap and IPA results are consistent in showing that toilet cleanliness at both the station (Item 6) and on the train (Item 13) remains below the desired level, with Item 13 requiring more definite improvement. In particular, the IV results further show that, even under the most optimistic interpretation, both gap intervals remain negative. This is in line with previous rail service research, which has identified toilet cleanliness and station sanitation as areas requiring improvement [116, 117].

Under both SV and IV, seat availability (Item 7) appears as a highly important on-board attribute, and both the SV and IV gap results indicate that it remains below the desired level. The main difference is that, although the SV IPA places Item 7 in Q2-KU, the corresponding IV IPA rectangle extends towards Q1-CH, showing that its performance is less secure once uncertainty is taken into account. This is in line with previous rail service research, which has identified crowdedness and the inability to obtain a seat as important

determinants of passenger dissatisfaction and service quality in rail travel [118].

Under both SV and IV, waiting area (Item 2) lies close to the central axes in the IPA plots. The main difference is that the IV result makes it clearer that this attribute is borderline, as its rectangle also partially belongs to Q1-CH rather than only to Q3-LP. This is in line with previous rail research showing that waiting environments can materially shape station experience and perceived waiting time, and that waiting room facilities may warrant explicit attention rather than being treated as straightforwardly low-priority [117, 119].

For information provision at the station (Item 1), the SV results are mixed, with a negative SV gap but a Q2-KU placement in the SV IPA, whereas the IV results suggest an overall satisfactory performance, as the IV IPA remains in Q2-KU with possible overlap with Q4-PO but not Q1-CH, and the IV gap shows substantial overlap with zero. Taken together, this suggests that the attribute is generally meeting passenger needs, and that any improvement concern relates more to improving its lower bound than to remedying a clear overall shortfall. This is broadly consistent with Great Britain evidence showing that information about train times and platforms was among the better-performing station attributes [120], while rail research also suggests that avoiding poor information is important because it can materially exacerbate passenger dissatisfaction [121].

5.7 Limitations

The limitations of this chapter are as follows.

5.7.1 Between-subjects Design

The real-world study in this chapter adopted a between-subjects design, in which participants were randomly assigned to either the SV or the IV survey. This supports comparison between the two response formats at the group level, but it does not allow direct comparison of how the same individual may respond under both formats. As a result, some observed differences between the SV and IV results may reflect differences between the two groups rather than differences caused only by the response format itself. Future work could therefore use a within-subjects design to allow more direct comparison between SV and IV responses.

5.7.2 Cross-use of Constructs

This study involved cross-use of construct terms in service evaluation. One example in this chapter is the Gap analysis, in which importance ratings were used as a proxy for expectations, following a common practice in the service field. Although the intention in the present study was to support efficient method demonstration, it is worth noting that this practice remains debatable in this field. The use of interval-valued scales may further complicate it, as the bounds of an interval may carry their own meaning, such as adequate service and desired service for expectations, but not necessarily in the same way for importance. This is therefore worth further exploration in future work, for example to examine the relationship between IV responses on expectations and importance, the similarity or dissimilarity between these constructs, and the meaning of their bounds from the respondents' point of view.

5.7.3 CSI

The present study only demonstrated how CSI can be calculated as a single score, while in real-world practice, it is primarily used for benchmarking against competitors and over time; future work should therefore examine IV CSI in these settings.

In addition, in the IV extension, CSI was calculated using the IWA with the Karnik–Mendel algorithm (Section 2.3.2). As a result of cross-item optimisation, it does not directly provide the item-level contribution to the final CSI value as the traditional SV CSI does. One possible solution to address this in future work is to aggregate the lower bounds across items and the upper bounds across items separately, so that item-level contributions are also available in addition to the total score.

5.7.4 Sample Size

The present study was exploratory and based on a relatively small sample of UK rail passengers. While the results provide direct evidence for the potential of IV extensions in assisting uncertainty-aware decision-making, they are not sufficient to support practical conclusions. Further studies with larger samples and ideally in collaboration with rail stakeholders are therefore needed.

5.8 Summary and Remarks

This chapter extended three widely used service evaluation methods—Gap Analysis, IPA, and CSI—to IV data, enabling uncertainty to be represented directly in decision-support outputs. The extensions were demonstrated through a UK rail passenger survey study, in which comparable SV and

IV responses were collected and analysed in parallel. The IV results were broadly consistent with the SV baseline, while providing additional information not available from SV analysis: the IV gap analysis revealed that only six attributes showed gaps that were entirely negative, the IV IPA showed that most attributes could not be assigned clearly to a single quadrant once uncertainty was taken into account, and the IV CSI captured the full range of possible aggregate satisfaction levels given the uncertainty in both satisfaction and importance responses.

Chapter 6

ISurvey: An R Toolkit for Interval Survey Analysis

To facilitate wider adoption in practice among researchers and practitioners from a variety of fields, it is important to make novel methods accessible through software support. In this chapter, an open-source R toolkit is introduced that provides an integrated implementation of the methods reviewed and developed in this thesis.

Section 6.1 briefly introduces the motivation and contributions of the toolkit. Section 6.2 provides an overview of the toolkit design and functions. Section 6.3 demonstrates a complete service evaluation workflow using the UK rail survey data collected in Chapter 5. Section 6.4 summarises the chapter and outlines future directions. Abbreviations and notations are summarised in the List of Abbreviations and the List of Notations.

6.1 Introduction

Surveys are widely used to capture human perceptions, judgements, and evaluations across domains such as marketing, customer experience, the public sector, and healthcare [17, 95, 122]. Recent years have seen growing interest in

Interval-Valued (IV) survey response formats as an alternative to conventional Single-Valued (SV) scales, as they allow respondents to express uncertainty and imprecision as ranges. IV response formats have been deployed across diverse fields, including cybersecurity, consumer research, healthcare, and environmental management [25, 28–31]. Recent developments also include software support such as DECSYS, which supports IV survey elicitation and facilitates its broader adoption in practice [27].

Beyond data elicitation, IV survey responses also require analysis methods tailored to IV data. These needs arise across several stages of conventional survey analysis, including psychometric assessment, descriptive statistics, visual summarisation, and decision-support methods. However, despite methodological progress in these areas, software support for applied IV survey analysis remains limited. Several symbolic data analysis packages [123, 124] support general IV analysis, including parametric modelling, automatic classification, and linear models, but do not support survey-oriented methods such as Cronbach's α.

In contrast, `IntervalQuestionStat` (IQS) is a pioneering publicly available toolkit specifically designed for IV survey analysis [125]; the methods it implements are summarised in Table 6.1. However, it does not yet incorporate recent Hausdorff-distance-based (H-based) developments for descriptive statistics, correlation, and Cronbach's α; visual descriptives such as the Interval Agreement Approach (IAA; Section 2.2.3) [44] and its IV summaries via interval-valued defuzzification (Section 2.2.3 and Chapter 4); or IV extensions of decision-support methods.

`ISurvey` is therefore developed to implement these methods, offering a free and open-source R toolkit designed to enable researchers from diverse backgrounds to conduct IV survey analysis within conventional workflows. Accordingly, this chapter makes the following contributions:

Table 6.1: Methods comparison of IQS and ISurvey

Component	**Methods**	**IQS**	**ISurvey**
Descriptives	Fréchet statistics (θ / H)	✓ / –	✓ / ✓
	Relative Frequency Histogram, IAA, interval-valued defuzzification	–	✓
Psycho-metric	Cronbach's α (θ / H)	✓ / –	✓ / ✓
	Correlation (midpoint, arithmetic, θ, H)	–	✓
Diagnostics	Gap analysis, Importance–Performance Analysis (IPA), Customer Satisfaction Index (CSI)	–	✓

1. Introduce `ISurvey`, which integrates the IV methods reviewed and developed in this thesis to encourage wider adoption and collaboration.

2. Provide a complete workflow demonstration using the UK rail survey data collected in Chapter 5, showing how `ISurvey` supports the full survey analysis process—from data loading through analysis to export—for both IV and SV data. These workflows also serve as a step-by-step tutorial on how uncertainty can be communicated in survey-based studies.

The `ISurvey` toolkit and the complete demonstration code in Section 6.3 are available on GitHub at https://github.com/Carina-YuZhao/.

6.2 Toolkit Overview

While `ISurvey` is able to handle user-defined interval data—for example, imported from Excel or `.csv` files, or specified directly within the software—it is particularly designed to work with survey data exported from DECSYS, which supports IV, SV, and *Choose One* response modes. DECSYS exports data in both `.csv` and `.json` formats, where each record corresponds to a single response, as illustrated by the simplified examples in Table 6.2.

Table 6.2: Examples of selected response modes in DECSYS.

Partic-ipant	Page_ Name	Response Type	Ellipse Scale_min	Ellipse Scale_max	Discrete Scale	Choose One
1	Q1	Ellipse	3.2	4.5	–	–
1	Q2	Discrete	–	–	4	–
1	Gender	Choose One	–	–	–	Male

6.2.1 Design Principles

`ISurvey` is designed to handle both SV and IV responses across all stages of the survey analysis workflow, supporting SV-only, IV-only, IV–SV, and IV–IV analyses, as well as comparison and visualisation. When multiple data sources are provided, items are matched by `Page Name` as identifiers, and response types (IV or SV) are detected automatically, allowing IV and SV responses sharing the same `Page Name` to be processed together. For IV data, all functions default to Hausdorff distance (d_H) within the Fréchet framework, with d_θ also available; SV data are processed automatically using conventional methods.

6.2.2 Function Summary

The main functions of `ISurvey` are grouped according to the survey analysis workflow, from data loading and descriptive analysis to psychometric assessment, diagnostics, and export. Table 6.3 summarises the core functions demonstrated in this chapter.

6.3 Workflow Demonstration

Survey analysis typically follows a structured workflow—descriptive summarisation and visualisation, psychometric assessment, and diagnostic analysis—as reviewed in Section 2.5.3. `ISurvey` extends this workflow for IV data. This

Table 6.3: Main functions in `ISurvey`.

Stage	**Functions**
Data loading and filtering	`load_data()`, `list_choices()`, `filter_by_choice()`
Descriptive statistics	`calc_descriptives()`, `plot_means()`
Visual descriptives	`plot_distributions()`, `plot_ivd_summary()`
Psychometric assessment	`calc_cor()`, `plot_correlation()`, `calc_reliability()`
Diagnostic analysis	`calc_gap()`, `plot_gap()`, `plot_ipa()`, `calc_csi()`
Export	`save_results()`, `save_plot()`

section demonstrates a complete workflow using the UK rail survey data collected and detailed in Chapter 5, where IV and SV responses were collected via a between-subject design using DECSYS. As the purpose here is to demonstrate the toolkit's default settings and multi-method capability, the examples in this section use H-based outputs, consistent with the Hausdorff-based results reported in Chapter 5. The workflow proceeds from Data Loading through Analysis (Visual Descriptives, Psychometric Assessment and Diagnostic Analysis) to Export.

6.3.1 Data Loading

`load_data()` imports IV and SV responses from CSV files, such as those exported from DECSYS, and standardises the raw imports into a consistent format for downstream analysis. By default, a built-in schema matches the standard DECSYS export format. For non-DECSYS datasets, a custom schema can map dataset-specific column names.

```
iv_data <- load_data("Rail_user_survey_interval.csv")
sv_data <- load_data("Rail_user_survey_5-point.csv")
Sat <- paste0("S", 1:14)  # satisfaction items
Imp <- paste0("I", 1:14)  # importance items
LABELS <- c("1 Journey Info", "2 Wait Area", ...)
```

The same Page Names are used across both IV and SV exports, enabling automatic item matching.

`list_choices()` inspects demographic labels stored as `choose_one` responses.

```
list_choices(iv_data)
# $Device: "Desktop" "Mobile phone" "Tablet"
# $Gender: "Female" "Male" "Other"
```

`filter_by_choice()` supports subgroup analysis by filtering respondents based on selected choice values and returning the corresponding sub-dataset.

```
iv_fd <- filter_by_choice(iv_data, c("Female", "Desktop"))
```

6.3.2 Descriptive Statistics

`calc_descriptives()` computes item-wise descriptive summaries for IV and SV data.

```
# method: "both" (default), "theta", "hausdorff"
df_S <- calc_descriptives(iv_data, sv_data, items = Sat,
        method = "both")
df_I <- calc_descriptives(iv_data, sv_data, items = Imp,
        method = "both")
```

```
> df_I$hausdorff["I1",]  # IV H:     [3.79, 4.28]
> df_I$theta["I1",]      # IV theta: [3.76, 4.24]
> df_I$sv["I1"]          # SV:       4.18
```

`plot_means()` visualises item-wise mean summaries. By default, all available means (θ, H, SV) are displayed; specific subsets can be selected via `show`.

Fig. 6.1 provides an intuitive comparison between IV and SV means for the importance items.

```
# sort_by: "none", "min", "max", "mid", "hurwicz"
p2 <- plot_means(df_I, labels = LABELS,
       show = c("theta", "H", "sv"),
       sort_by = "hurwicz", sort_data = "hausdorff",
       sort_order = "desc", gamma = 0.5)
```

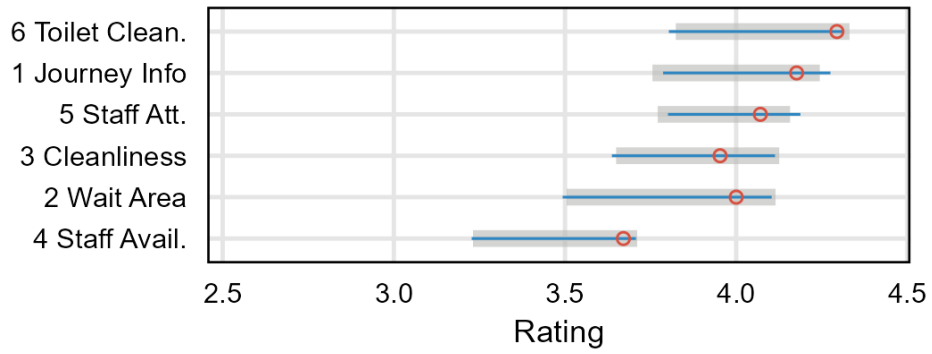


Figure 6.1: *Importance means: SV as red circles, θ-based as wide grey segments, H-based as thin blue segments.*

6.3.3 Visual Descriptives

`plot_distributions()` visualises response distributions across multiple items. For IV data, IAA visualisations are generated together with visualisations produced using Interval-valued Defuzzification (IVD) methods. For SV data, Relative Frequency Histogram (RFH) plots can be generated and overlaid with IAA when comparable. Fig. 6.2 shows IAA fuzzy sets overlaid on SV RFHs for comparing the overall distribution between IV and SV data.

```
# type: "iaa", "rfh", "iaa_rfh", "iaa_ivd"
p3 <- plot_distributions(iv_data, sv_data, items = Imp[1:3],
       type = "iaa_rfh", ncol = 3)
```

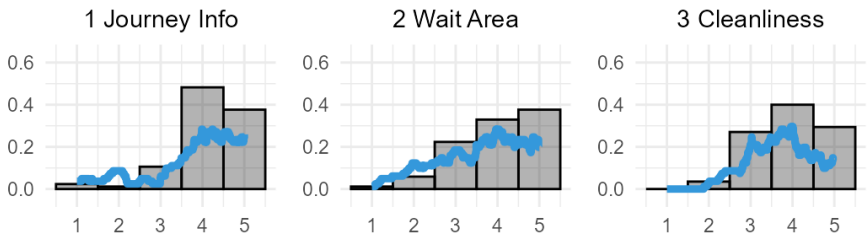


Figure 6.2: *IAA fuzzy sets overlaid on SV RFHs for Items I1–I3.*

```
p4_1 <- plot_distributions(iv_data, items = Imp[1:3],
        type = "iaa_ivd", ncol = 3)
```

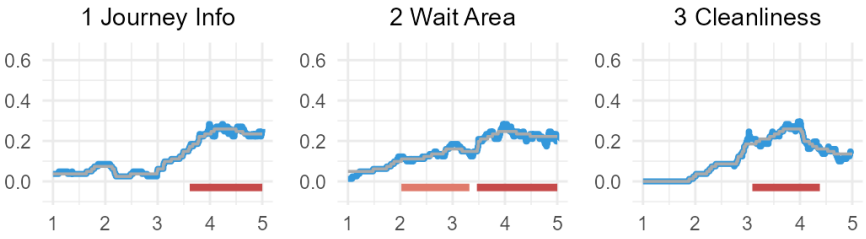
`plot_ivd_summary()` provides an item-wise summary view of IVD outputs. Items can be ranked by defuzzification of their corresponding IAA fuzzy sets. Colour intensity reflects the peak height of each IAA fuzzy set. Fig. 6.3 shows IAA–IVD plots and the IVD summary for cross-item distribution comparison.

```
# ivd_method (see Chapter 4):
# "DUW" (default), "DU", "NUW", "NU", "IALC"
# sort_by: "none", "centroid", "mom", "bisector", "alc"
p4_2 <- plot_ivd_summary(iv_data, items = Imp[1:6],
        labels = LABELS, ivd_method = "DUW",
        sort_by = "centroid", sort_order = "desc")
```

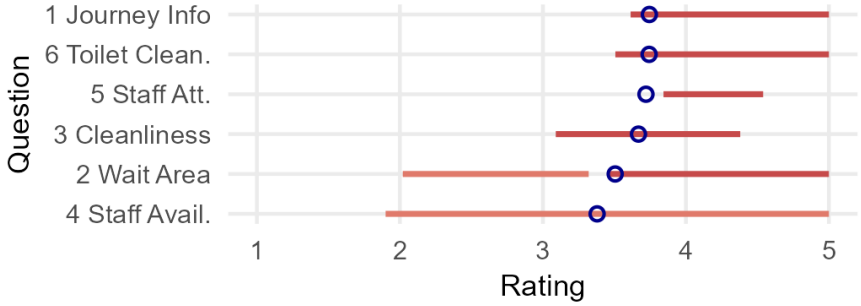
6.3.4 Psychometric Assessment

`calc_cor()` computes correlation coefficients. The `method` parameter specifies the correlation method.

```
# method: "hausdorff" (default), "theta", "midpoint"
> calc_cor(df_I$hausdorff, df_I$sv, method = "hausdorff")
# 0.964
```



(a) *IAA fuzzy sets with IVD summaries for Items I1–I3.*



(b) *IVD summary across Items I1–I6.*

Figure 6.3: *Examples of IAA and IVD visualisations.*

Here, the correlation coefficients between IV and SV means are calculated to assess convergent validity of the IV response format. Since this is a between-subject study, correlation is computed at the group level via SV and IV means.

`plot_correlation()` visualises the correlation, accepting multiple `x` against a single `y` benchmark, as shown in Fig. 6.4.

```
p5 <- plot_correlation(
      x = list(theta = df_I$theta, hausdorff = df_I$hausdorff),
      y = df_I$sv)
```

`calc_reliability()` computes Cronbach's α for IV and SV item sets. The function accepts one or more data sources and a list of item groups, computing α for all combinations.

```
# method: "hausdorff" (default), "theta"
> calc_reliability(iv_data, sv_data,
```

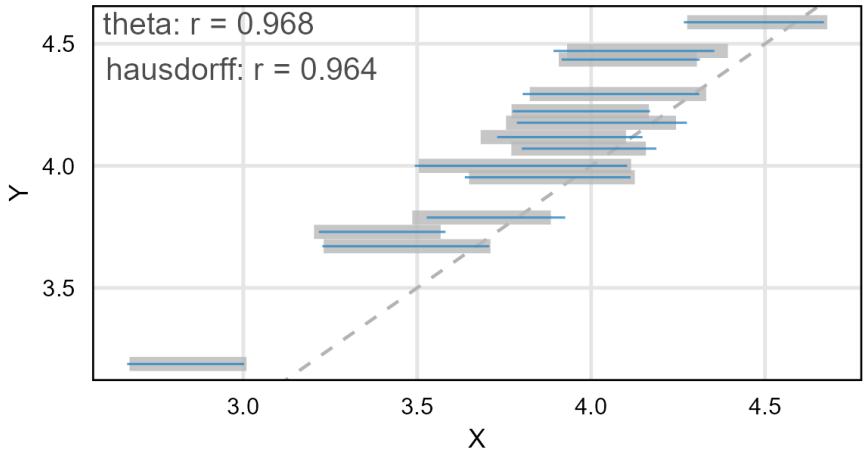


Figure 6.4: *Correlation between IV and SV means for Items I1–I14.*

```
    items = list(S = Sat, I = Imp), method = "hausdorff")
#     data items    method  alpha n_items
# 1 iv_data     S hausdorff 0.899      14
# 2 iv_data     I hausdorff 0.871      14
# 3 sv_data     S        sv 0.899      14
# 4 sv_data     I        sv 0.884      14
```

6.3.5 Diagnostic Analysis

`calc_gap()` computes item-wise gaps as perception minus expectation ($P-E$). For IV data, interval subtraction is applied.

```
> calc_gap(expectation = df_I$hausdorff,
           perception = df_S$hausdorff)
#   Aspect Sat_L ... Imp_L ... Gap_L Gap_U
#        1  3.64 ...  3.79 ... -0.63  0.38
```

`plot_gap()` visualises importance–satisfaction gaps. In Fig. 6.5, importance intervals are thick/light segments and satisfaction intervals are thin/dark segments. SV points (red circles) are overlaid for comparison.

```
p6 <- plot_gap(perception = df_S$hausdorff, expectation = df_I$hausdorff,
      labels = LABELS,
      sv_perception = df_S$sv, sv_expectation = df_I$sv)
```

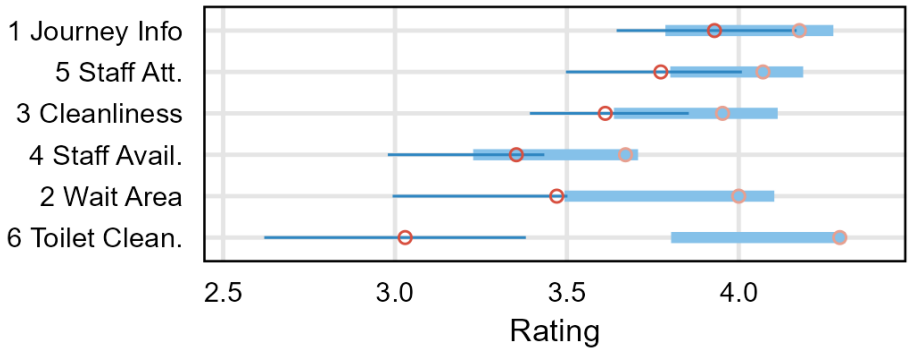


Figure 6.5: *Gap analysis with IV and SV representations.*

`plot_ipa()` generates IPA visualisations. For IV inputs, quadrant boundaries are shaded regions rather than single lines. Fig. 6.6 illustrates SV-based and IV-based IPA.

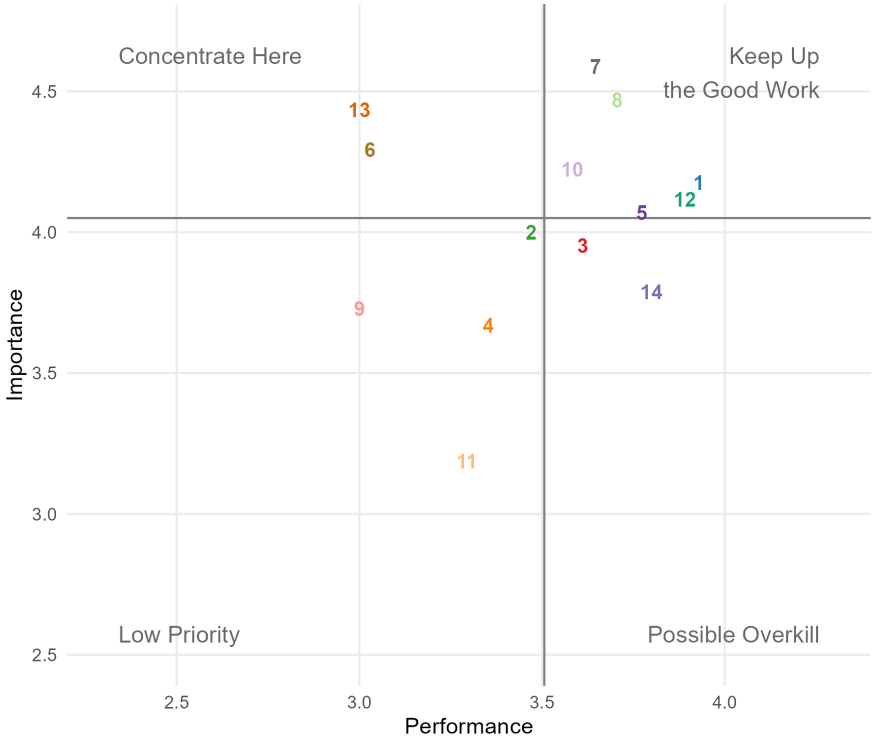
```
p7_1 <- plot_ipa(df_I$sv, df_S$sv)
p7_2 <- plot_ipa(df_I$hausdorff, df_S$hausdorff)
```

`calc_csi()` computes CSI via importance-weighted aggregation. CSI is numeric for SV data and interval for IV data.

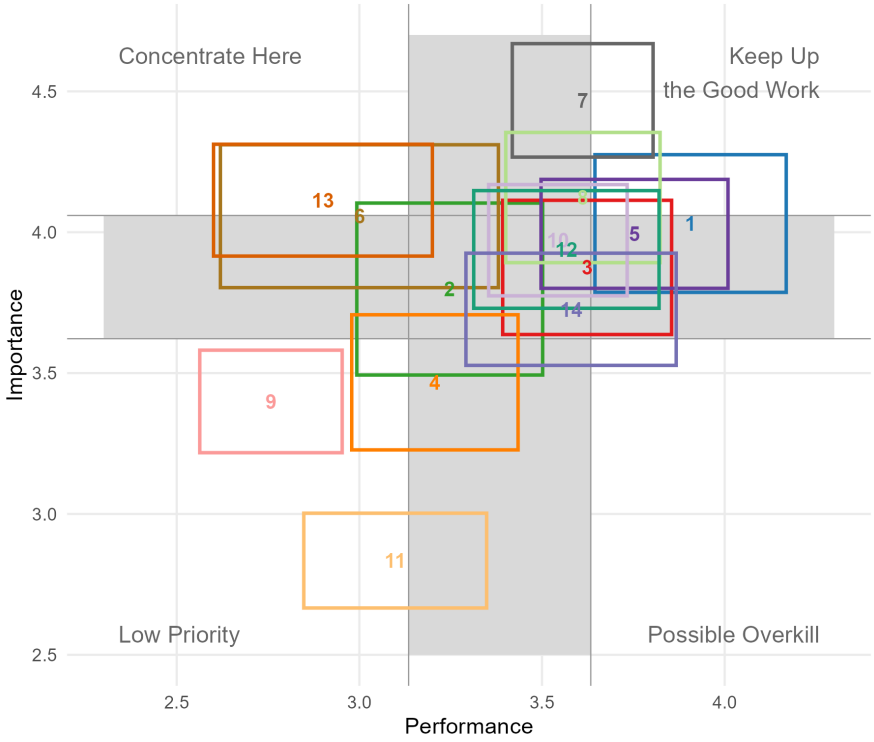
```
> calc_csi(df_I$sv, df_S$sv)
# 3.51
> calc_csi(df_I$hausdorff, df_S$hausdorff)
# [3.13, 3.67]
```

6.3.6 Data Export

`save_results()` exports tabular outputs to Excel for reporting and further analysis. IV data stored as nested lists are unpacked into separate columns.



(a) *SV-based IPA.*



(b) *IV-based IPA with IV quadrant boundaries.*

Figure 6.6: *IPA comparison using SV and IV data.*

```
save_results(df_I, file = "Descriptives", dir = "output/")
```

`save_plot()` saves figures with consistent defaults. Plots are `ggplot2` objects, allowing further customisation.

```
save_plot(p7_2, "IPA_IV", dir = "output/")
```

6.4 Summary and Remarks

This chapter presented `ISurvey`, an R toolkit addressing the limited availability of accessible software for IV survey data analysis. `ISurvey` provides methods including IAA, IVD, Fréchet-based descriptive statistics under d_θ and d_H, correlation measures, and Cronbach's α, as well as IV extensions of Gap analysis, IPA, and CSI. Through the UK rail survey data collected in Chapter 5, this chapter demonstrated a complete workflow—from DECSYS export through analysis to reporting—comparing IV and SV responses to illustrate how the toolkit supports researchers in evaluating whether IV data provide additional insights compared to traditional SV methods.

Future work will focus on three directions. First, the toolkit will be extended to incorporate additional methods as they are developed in the IV analysis literature, including further IV statistical methods, ranking and multi-criteria decision-making methods, and additional visualisation and diagnostic methods from other application domains. Second, a graphical user interface will be developed to further lower the barrier to adoption for researchers and practitioners without programming experience. Third, the toolkit will be applied to additional domains beyond rail services to evaluate its generalisability across different survey contexts.

Chapter 7

Conclusion

7.1 Contributions and Publications

This research focused on developing and evaluating methods for communicating uncertainty in human-sourced information. Motivated by the fact that uncertainty is often inherent in human perceptions, it articulated the necessity of capturing such uncertainty in current survey practice and communicating it throughout an end-to-end workflow to support well-informed decision-making. The practical value of this research was demonstrated through a real-world application in rail service evaluation.

Chapter 1 reviews the historical development and current state of the art in uncertainty communication, identifying research gaps that motivated the adoption of intervals as a unifying basis for uncertainty representation in this thesis, and setting out the research objectives: 1 Motivation and Application Context; 2 Empirical Data Collection; 3 Reliability Assessment; 4 Uncertainty Summarisation; 5 Decision-Support Presentation; and 6 Workflow Integration and Software Support.

Chapter 2 provides the background for the subsequent chapters by reviewing the key terms, mathematical concepts, and recent methodological developments in uncertainty elicitation and analysis, as well as the key concepts and

methods in service evaluation.

Chapter 3 addresses Objective 3 by developing a Hausdorff-distance-based extension of Cronbach's α for assessing the internal consistency of interval-valued data, so that interval-valued survey responses can be verified to have sufficient reliability before proceeding to subsequent analysis.

Chapter 4 addresses Objective 4 by developing and evaluating methods for summarising fuzzy sets as intervals, so that uncertainty information is retained for communication while also remaining relatively simple for interpretation by stakeholders. Related work was published in [74] and [106], with a further paper submitted in [126].

Chapter 5 addresses Objective 1 by drawing on existing literature to articulate the need for capturing and communicating uncertainty in service evaluation, and addresses Objective 2 by collecting comparable single-valued and interval-valued datasets through a UK rail passenger survey and making the data publicly available for further research. It also addresses Objective 5 by extending conventional service evaluation methods to communicate uncertainty and demonstrating the potential benefits of doing so using the collected survey data.

Chapter 6 addresses Objective 6 by integrating the methods developed throughout the thesis into an end-to-end interval survey workflow and implementing these methods in a source-available R toolkit to facilitate their wider adoption by researchers and practitioners. The toolkit and workflow demonstration were published in [127].

Although the development of the methods in this thesis was motivated by, and mainly demonstrated through, an interval-valued survey workflow in service evaluation, the methodological contributions apply more broadly to

interval-valued data and fuzzy sets beyond the survey context. The practical contributions also have wider potential beyond service evaluation, with applicability to other domains where surveys are used to collect subjective human responses.

Taken together, these contributions advance the state of the art in the analysis and application of interval-valued representations for communicating uncertainty in human-sourced information, thereby moving towards better-informed decision-making in data-driven contexts.

7.2 Limitations and Future Work

While detailed limitations are summarised at the end of Chapters 3 to 6, this section draws together a number of broader limitations of the thesis and directions for future work.

7.2.1 Interval Distances and Psychometric Assessment

A broader methodological limitation concerns interval distance choice across the interval-valued survey workflow, as the present thesis focuses primarily on the θ-distance and the Hausdorff distance. In addition, although Chapter 3 shows both similarities and differences in their behaviour, the implications of these differences for reliability and other aspects of psychometric assessment, such as convergent validity, remain unclear.

Future work should therefore proceed in two directions. First, a more comprehensive review of interval distances for interval-valued survey analysis is needed to examine whether the current focus on the θ-distance and the Hausdorff distance is sufficient, or whether further alternatives should be considered. Second, further comparison of the θ-distance and the Hausdorff

distance is needed across psychometric assessment to clarify how and why they differ, how these differences relate to single-valued analysis, and how they should inform method choice and result interpretation in different real-world applications. A next step is therefore to further develop Chapter 3 along these two directions, with the aim of preparing it for publication in a venue such as *Information Sciences*.

7.2.2 Interval-valued Defuzzification

While the interval-valued defuzzification methods developed in Chapter 4 are able to satisfy the desirable properties introduced for interval-valued defuzzification of non-convex fuzzy sets, some issues remain unresolved, in particular those concerning normality and continuity. These issues were noted in Chapter 4, but were not fully addressed within the present thesis, and further work is needed to refine the methods in these respects.

Another aspect that is worth exploring is the statistical meaning of interval-valued defuzzification. In the case of single-valued data, an overall view is often given by a histogram or other distributional representation, whereas a simplified summary is commonly given through a combination of central tendency and dispersion, such as a mean together with a standard deviation, commonly paired as a single point and a symmetrical error bar for data visualisation.

By contrast, while interval-valued defuzzification for convex fuzzy sets is originally viewed as a central tendency measure, with non-convex fuzzy sets and discontinuous intervals it appears that the resulting output reflects not only central tendency, but also aspects of dispersion of the fuzzy set. However, this view has not yet been further discussed or examined in this thesis.

A related direction for future work also concerns whether the same idea can be extended beyond fuzzy sets. In many real-world settings, data are already available in single-valued form together with some corresponding distribution, whether through a histogram or a probability distribution. It remains an open question whether the present approach to summarising a distribution as an interval can be generalised beyond the specific case of fuzzy sets.

7.2.3 Intervals for Service Evaluation

While Chapter 5 articulates in detail the motivation for capturing and presenting uncertainty in service evaluation, and demonstrates the potential of such methods, adopting them not only as an interdisciplinary attempt but also as a fully integrated part of common practice in service research requires further development.

Chapter 5 demonstrated the potential of interval-valued extensions of Gap analysis, Importance–Performance Analysis, and Customer Satisfaction Index to provide additional information beyond single-valued analysis. However, further work is needed to examine this potential from the perspective of decision makers themselves. In particular, experiments are needed to examine whether the presentation of uncertainty through interval-valued outputs leads decision makers to make different decisions, and what the consequences of those differences may be. This would help clarify whether the additional information provided by interval-valued methods leads to meaningful added value in practice.

While part of the work in Chapter 6 has already been presented at conferences in service and rail contexts (see p. vii), further development is still needed before these methods can be more fully established as service methods rather than only as an interdisciplinary application. A next step is therefore to

develop Chapter 5 towards publication in journals such as *Journal of Retailing and Consumer Services*, *Journal of Business Research*, and *Journal of Services Marketing*. Establishing validity and added value more fully in the service domain may then also support wider adoption in related areas such as marketing and beyond.

Bibliography

[1] W. Schramm, "The nature of communication between humans," in *The Process and Effects of Mass Communication*, W. Schramm and D. F. Roberts, Eds. University of Illinois Press, 1971.

[2] A. W. Crosby, *The Measure of Reality: Quantification and Western Society, 1250-1600*. Cambridge University Press, 1997.

[3] W. Kula, *Measures and Men*. Princeton University Press, 1986.

[4] T. M. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press, 1995.

[5] A. Desrosières, *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press, 1998.

[6] L. Thévenot, "Les investissements de forme," in *Conventions économiques*. Paris: Presses Universitaires de France, 1986, pp. 21–71.

[7] K. Danziger, *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge University Press, 1990.

[8] J. Michell, *Measurement in Psychology: A Critical History of a Methodological Concept*. Cambridge University Press, 1999.

[9] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.

[10] R. C. B. Aitken, "Measurement of feelings using visual analogue scales," *Proceedings of the Royal Society of Medicine*, vol. 62, no. 10, pp. 989–993, 1969.

[11] R. F. DeVellis, *Scale Development: Theory and Applications*, 3rd ed. SAGE Publications, 2012.

[12] B. Hesketh, R. Pryor, M. Gleitzman, and T. Hesketh, "Practical applications and psychometric evaluation of a computerized fuzzy graphic rating scale," in *Fuzzy Sets in Psychology*, ser. Advances in Psychology, T. Zetenyi, Ed. Amsterdam: North-Holland, 1988, vol. 56, pp. 425–454.

[13] D. E. Polkinghorne, "Further extensions of methodological diversity for counseling psychology," *Journal of Counseling Psychology*, vol. 31, no. 4, pp. 416–429, 1984.

[14] J. R. Taylor, *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, 2nd ed. University Science Books, 1997.

[15] L. Daston and P. Galison, *Objectivity.* Zone Books, 2007.

[16] R. P. Crease, *World in the Balance: The Historic Quest for an Absolute System of Measurement.* W. W. Norton & Company, 2011.

[17] R. Tourangeau, L. J. Rips, and K. Rasinski, *The Psychology of Survey Response.* Cambridge University Press, 2000.

[18] M. Sherif and C. I. Hovland, *Social Judgment: Assimilation and Contrast Effects in Communication and Attitude Change.* New Haven, CT: Yale University Press, 1961.

[19] B. Russell, "Vagueness," *Australasian Journal of Psychology and Philosophy*, vol. 1, no. 2, pp. 84–92, 1923.

[20] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[21] Z. Ellerby, C. Wagner, and S. B. Broomell, "Capturing richer information: On establishing the validity of an interval-valued survey response mode," *Behavior Research Methods*, vol. 54, no. 3, pp. 1240–1262, 2022.

[22] J. Cohen, "Things i have learned (so far)," *American Psychologist*, vol. 45, no. 12, pp. 1304–1312, 1990.

[23] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions.* Prentice Hall, 2001.

[24] L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning—I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.

[25] J. Navarro, C. Wagner, U. Aickelin, L. Green, and R. Ashford, "Exploring differences in interpretation of words essential in medical expert-patient communication," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2016, pp. 2157–2164.

[26] T. Hesketh, R. Pryor, and B. Hesketh, "An application of a computerized fuzzy graphic rating scale to the psychological measurement of individual differences," *International Journal of Man-Machine Studies*, vol. 29, pp. 21–35, 1988.

[27] Z. Ellerby, J. McCulloch, J. Young, and C. Wagner, "DECSYS – discrete and ellipse-based response capture system," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2019, pp. 1–6.

[28] K. J. Wallace, C. Wagner, and M. J. Smith, "Eliciting human values for conservation planning and decisions: A global issue," *Journal of Environmental Management*, vol. 170, pp. 160–168, 2016.

[29] K. J. Wallace, C. Wagner, D. J. Pannell, M. K. Kim, and A. A. Rogers, "Tackling communication and analytical problems in environmental planning: Expert assessment of key definitions and their relationships," *Journal of Environmental Management*, vol. 317, p. 115431, 2022.

[30] Z. Ellerby, O. Miles, J. McCulloch, and C. Wagner, "Insights from interval-valued ratings of consumer products – a DECSYS appraisal," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2020, pp. 1–8.

[31] S. Kabir, C. Wagner, and Z. Ellerby, "Toward handling uncertainty-at-source in AI – a review and next steps for interval regression," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 1, pp. 3–22, 2024.

[32] C. W. Choo, "The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions," *International Journal of Information Management*, vol. 16, no. 5, pp. 329–340, 1996.

[33] R. L. Ackoff, "From data to wisdom," *Journal of Applied Systems Analysis*, vol. 16, pp. 3–9, 1989.

[34] J. A. Martilla and J. C. James, "Importance-performance analysis," *Journal of Marketing*, vol. 41, no. 1, pp. 77–79, 1977.

[35] J. Dewey, "Logical method and law," *Cornell Law Quarterly*, vol. 10, no. 1, pp. 17–27, 1924.

[36] R. Krzysztofowicz, "The case for probabilistic forecasting in hydrology," *Journal of Hydrology*, vol. 249, no. 1–4, pp. 2–9, 2001.

[37] C. F. Manski, *Public Policy in an Uncertain World: Analysis and Decisions*. Cambridge, MA: Harvard University Press, 2013.

[38] B. Fischhoff, "Communicating uncertainty: Fulfilling the duty to inform," *Issues in Science and Technology*, vol. 28, no. 4, pp. 63–70, 2012.

[39] S. Ferson, V. Kreinovich, J. Hajagos, W. Oberkampf, and L. Ginzburg, "Experimental uncertainty estimation and statistics for data having interval uncertainty," Sandia National Laboratories, Tech. Rep., 2007.

[40] T. Augustin, F. P. A. Coolen, G. de Cooman, and M. C. M. Troffaes, Eds., *Introduction to Imprecise Probabilities*. John Wiley & Sons, 2014.

[41] L. Billard and E. Diday, *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. John Wiley & Sons, 2006.

[42] H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty: Applications to Computer Science and Engineering*, ser. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2012, vol. 393.

[43] G. Schollmeyer and T. Augustin, "Statistical modeling under partial identification: Distinguishing three types of identification regions in regression analysis with interval data," *International Journal of Approximate Reasoning*, vol. 56, pp. 224–248, 2015.

[44] C. Wagner, S. Miller, J. M. Garibaldi, D. T. Anderson, and T. C. Havens, "From interval-valued data to general type-2 fuzzy sets," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 2, pp. 248–269, 2015.

[45] M. A. Lubiano, M. Montenegro, S. Pérez-Fernández, and M. A. Gil, "Analyzing the influence of the rating scale for items in a questionnaire on Cronbach coefficient alpha," in *Trends in Mathematical, Information and Data Sciences*. Cham: Springer, 2023, vol. 445, pp. 377–388.

[46] J. García-García, M. A. Gil, and M. A. Lubiano, "On some properties of Cronbach's α coefficient for IV questionnaires," *Advances in Data Analysis and Classification*, 2024.

[47] M. A. Gil, M. A. Lubiano, M. Montenegro, and M. T. López, "Least squares fitting of an affine function and strength of association for IV data," *Metrika*, vol. 56, pp. 97–111, 2002.

[48] X. Kang, X. Ouyang, H. Liang, and C. Meng, "Hausdorff correlation for IV random objects," *Statistics and Computing*, vol. 35, 2025.

[49] R. R. Yager, "A procedure for ordering fuzzy subsets of the unit interval," *Information Sciences*, vol. 24, no. 2, pp. 143–161, 1981.

[50] L. Anzilli, G. Facchinetti, and G. Mastroleo, "A parametric approach to evaluate fuzzy quantities," *Fuzzy Sets and Systems*, vol. 250, pp. 110–133, 2014.

[51] F. Herrera and E. Herrera-Viedma, "Linguistic decision analysis: steps for solving decision problems under linguistic information," *Fuzzy Sets and Systems*, vol. 115, no. 1, pp. 67–82, 2000.

[52] P. Brito, "Symbolic data analysis: another look at the interaction of data mining and statistics," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 4, pp. 281–295, 2014.

[53] M. Chavent, "A Hausdorff distance between hyper-rectangles for clustering interval data," in *Classification, Clustering, and Data Mining Applications*, D. Banks, F. R. McMorris, P. Arabie, and W. Gaul, Eds. Berlin, Heidelberg: Springer, 2004, pp. 333–339.

[54] D. Z. Saletić, D. M. Velasevic, and N. E. Mastorakis, "Analysis of basic defuzzification techniques," in *Proc. 6th WSEAS International Conference on Circuits, Systems, Communications and Computers (CSCC)*, Rethymnon, Greece, 2002.

[55] C. Bertoluzza, N. Corral, and A. Salas, "On a new class of distances between fuzzy numbers," *Mathware and Soft Computing*, vol. 2, pp. 71–84, 1995.

[56] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*. Upper Saddle River, NJ: Prentice Hall, 1995.

[57] R. Belóhlávek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*. Oxford University Press, 2017.

[58] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*. Florida: Chapman and Hall/CRC, 2019.

[59] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*. Dordrecht: Kluwer, 1999.

[60] D. Dubois and H. Prade, "Gradualness, uncertainty and bipolarity: Making sense of fuzzy sets," *Fuzzy Sets and Systems*, vol. 192, pp. 3–24, 2012.

[61] P. Grzegorzewski, "Nearest interval approximation of a fuzzy number," *Fuzzy Sets and Systems*, vol. 130, pp. 321–330, 2002.

[62] R. R. Yager, "On ranking fuzzy numbers using valuations," *International Journal of Intelligent Systems*, vol. 14, no. 12, pp. 1249–1268, 1999.

[63] Joint Committee for Guides in Metrology, "Evaluation of measurement data — guide to the expression of uncertainty in measurement," JCGM, Tech. Rep. JCGM 100:2008, 2008.

[64] F. Liu and J. M. Mendel, "Encoding words into interval type-2 fuzzy sets using an interval approach," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1503–1521, 2008.

[65] D. Wu, J. M. Mendel, and S. Coupland, "Enhanced interval approach for encoding words into interval type-2 fuzzy sets and its convergence analysis," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 3, pp. 499–513, 2012.

[66] M. Hao and J. M. Mendel, "Encoding words into normal interval type-2 fuzzy sets: HM approach," in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2016, pp. 1–8.

[67] T. C. Havens, D. T. Anderson, and C. Wagner, "Data-informed fuzzy measures for fuzzy integration of intervals and fuzzy numbers," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 5, pp. 1141–1155, 2017.

[68] D. Dubois and H. Prade, "The mean value of a fuzzy number," *Fuzzy Sets and Systems*, vol. 24, pp. 279–300, 1987.

[69] S. Heilpern, "The expected value of a fuzzy number," *Fuzzy Sets and Systems*, vol. 47, no. 1, pp. 81–86, 1992.

[70] M. Jiménez, "Ranking fuzzy numbers through the comparison of its expected intervals," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 4, no. 4, pp. 379–388, 1996.

[71] D. A. Ralescu, "Average level of a fuzzy set," in *Statistical Modeling, Analysis and Management of Fuzzy Data*. Physica-Verlag, 2002, pp. 119–126.

[72] J. Fortin, D. Dubois, and H. Fargier, "Gradual numbers and their application to fuzzy interval analysis," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 2, pp. 388–402, 2008.

[73] A. Pourabdollah, C. Wagner, J. H. Aladi, and J. M. Garibaldi, "Improved uncertainty capture for non-singleton fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 12, pp. 3353–3364, 2020.

[74] Y. Zhao, C. Wagner, B. Ryan, D. Pekaslan, and J. Navarro, "Interval agreement weighted average – sensitivity to data set features," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2024, pp. 1–8.

[75] M. Delgado, M. A. Vila, and W. Voxman, "On a canonical representation of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 93, no. 1, pp. 125–135, 1998.

[76] C. Carlsson and R. Fullér, "On possibilistic mean value and variance of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 122, no. 2, pp. 315–326, 2001.

[77] R. Fullér and P. Majlender, "On weighted possibilistic mean and variance of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 136, no. 3, pp. 363–374, 2003.

[78] P. Fortemps and M. Roubens, "Ranking and defuzzification methods based on area compensation," *Fuzzy Sets and Systems*, vol. 82, no. 3, pp. 319–330, 1996.

[79] L. Anzilli and G. Facchinetti, "The total variation of bounded variation functions to evaluate and rank fuzzy quantities," *International Journal of Intelligent Systems*, vol. 28, no. 10, pp. 927–956, 2013.

[80] ——, "A general fuzzy set representation for decision making," in *Proceedings of the 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSA-EUSFLAT)*. Atlantis Press, 2015.

[81] R. E. Moore, *Interval Analysis*. Prentice-Hall, 1966.

[82] A. Neumaier, *Interval Methods for Systems of Equations*. Cambridge University Press, 1990.

[83] F. Liu and J. M. Mendel, "Aggregation using the fuzzy weighted average as computed by the Karnik–Mendel algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 16, pp. 1–12, 2008.

[84] A. Irpino and R. Verde, "Basic statistics for distributional symbolic variables: A new metric-based approach," *Advances in Data Analysis and Classification*, vol. 9, no. 2, pp. 143–175, 2015.

[85] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Annales de l'Institut Henri Poincaré*, vol. 10, no. 4, pp. 215–310, 1948.

[86] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "A conceptual model of service quality and its implications for future research," *Journal of Marketing*, vol. 49, no. 4, pp. 41–50, 1985.

[87] W. J. Regan, "The service revolution," *Journal of Marketing*, vol. 27, no. 3, pp. 57–62, 1963.

[88] G. L. Shostack, "Breaking free from product marketing," *Journal of Marketing*, vol. 41, no. 2, pp. 73–80, 1977.

[89] V. A. Zeithaml, "How consumer evaluation processes differ between goods and services," in *Marketing of Services*, J. H. Donnelly and W. R. George, Eds. Chicago: American Marketing Association, 1981, pp. 186–190.

[90] A. Parasuraman, V. A. Zeithaml, and L. L. Berry, "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality," *Journal of Retailing*, vol. 64, no. 1, pp. 12–40, 1988.

[91] J. J. Cronin and S. A. Taylor, "Measuring service quality: A reexamination and extension," *Journal of Marketing*, vol. 56, no. 3, pp. 55–68, 1992.

[92] S. A. Taylor and T. L. Baker, "An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions," *Journal of Retailing*, vol. 70, no. 2, pp. 163–178, 1994.

[93] D. T. Campbell and D. W. Fiske, "Convergent and discriminant validation by the multitrait-multimethod matrix," *Psychological Bulletin*, vol. 56, no. 2, pp. 81–105, 1959.

[94] L. J. Cronbach, "Coefficient alpha and the internal structure of tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[95] J. de Oña and R. de Oña, "Quality of service in public transport based on customer satisfaction surveys: A review and assessment of methodological approaches," *Transportation Science*, vol. 49, no. 3, pp. 605–622, 2015.

[96] N. Hill and J. Brierley, *How to Measure Customer Satisfaction*. Routledge, 2017.

[97] Transport Focus, "National rail passenger survey (NRPS) – spring 2020 main report," 2020. [Online]. Available: https://www.transportfocus.org.uk/publication/national-rail-passenger-survey-nrps-spring-2020-main-report/

[98] ——, "National rail passenger survey (NRPS)," 2020, accessed: 2025. [Online]. Available: https://www.transportfocus.org.uk/insight/national-rail-passenger-survey/

[99] ——, "Biggest ever rail passenger satisfaction survey launches nationwide (RCES)," 2025. [Online]. Available: https://www.transportfocus.org.uk/news/biggest-ever-rail-passenger-satisfaction-survey-launches-nationwide/

[100] I. Couso and D. Dubois, "Statistical reasoning with set-valued information: Ontic vs. epistemic views," *International Journal of Approximate Reasoning*, vol. 55, no. 7, pp. 1579–1608, 2014.

[101] F. Hausdorff, *Grundzüge der Mengenlehre*. Leipzig: Veit, 1914.

[102] C. Spearman, "Correlation calculated from faulty data," *British Journal of Psychology*, vol. 3, no. 3, pp. 271–295, 1910.

[103] W. Brown, "Some experimental results in the correlation of mental abilities," *British Journal of Psychology*, vol. 3, no. 3, pp. 296–322, 1910.

[104] O. M. Poleshuk and E. G. Komarov, "New defuzzification method based on weighted intervals," in *Proc. Annual Meeting of the North American Fuzzy Information Processing Society*, 2008.

[105] P. Grzegorzewski and E. Mrówka, "Trapezoidal approximations of fuzzy numbers," *Fuzzy Sets and Systems*, vol. 153, no. 1, pp. 115–135, 2005.

[106] Y. Zhao, C. Wagner, B. Ryan, D. Pekaslan, and V. Kreinovich, "Effectively communicating information and uncertainty from fuzzy sets: Why and how," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2025, pp. 1–7.

[107] K. A. Ericsson and H. A. Simon, *Protocol Analysis: Verbal Reports as Data*, revised ed. MIT Press, 1993.

[108] J. L. Heskett, T. O. Jones, G. W. Loveman, W. E. Sasser, and L. A. Schlesinger, "Putting the service-profit chain to work," *Harvard Business Review*, vol. 72, no. 2, pp. 164–174, 1994.

[109] N. Seth, S. G. Deshmukh, and P. Vrat, "Service quality models: a review," *International Journal of Quality & Reliability Management*, vol. 22, no. 9, pp. 913–949, 2005.

[110] A. T. Jebb, V. Ng, and L. Tay, "A review of key Likert scale development advances: 1995–2019," *Frontiers in Psychology*, vol. 12, p. 637547, 2021.

[111] G. M. Sullivan and A. R. Artino, "Analyzing and interpreting data from Likert-type scales," *Journal of Graduate Medical Education*, vol. 5, no. 4, pp. 541–542, 2013.

[112] V. A. Zeithaml, L. L. Berry, and A. Parasuraman, "The nature and determinants of customer expectations of service," *Journal of the Academy of Marketing Science*, vol. 21, no. 1, pp. 1–12, 1993.

[113] K.-C. Hu, "Evaluating city bus service based on zone of tolerance of expectation and normalized importance," *Transport Reviews*, vol. 30, no. 2, pp. 195–217, 2010.

[114] C. Fornell, M. D. Johnson, E. W. Anderson, J. Cha, and B. E. Bryant, "The american customer satisfaction index: Nature, purpose, and findings," *Journal of Marketing*, vol. 60, no. 4, pp. 7–18, 1996.

[115] A. Parasuraman, L. L. Berry, and V. A. Zeithaml, "Understanding customer expectations of service," *MIT Sloan Management Review*, vol. 32, no. 3, pp. 39–48, 1991.

[116] J. de Oña, R. de Oña, L. Eboli, and G. Mazzulla, "Heterogeneity in perceptions of service quality among groups of railway passengers," *International Journal of Sustainable Transportation*, vol. 9, no. 8, pp. 612–626, 2015.

[117] M. R. Islam, M. T. Ahmed, N. Anwari, M. Hadiuzzaman, and S. Amin, "The aspiration for happy train journey: Commuters' perception of the quality of intercity rail services," *CivilEng*, vol. 3, no. 4, pp. 909–945, 2022.

[118] J. Preston, J. Pritchard, and B. Waterson, "Train overcrowding: Investigating the use of better information provision to mitigate the issues," in *96th Annual Meeting of the Transportation Research Board*, Washington, DC, USA, Jan. 2017.

[119] M. van Hagen, "Waiting experience at train stations," Ph.D. dissertation, University of Twente, Enschede, The Netherlands, 2011, dissertation.

[120] Transport Focus, "Rail passenger satisfaction at a glance: Great Britain – Autumn 2018," Transport Focus, Tech. Rep., Dec. 2018. [Online]. Available: https://www.transportfocus.org.uk/publication/national-rail-passenger-survey-nrps-glance-great-britain-wide-autumn-2018/nrps-autumn-2018-at-a-glance-great-britain/

[121] F. Monsuur, M. Enoch, M. Quddus, and S. Meek, "Modelling the impact of rail delays on passenger satisfaction," *Transportation Research Part A: Policy and Practice*, vol. 152, pp. 19–35, 2021.

[122] R. M. Groves, F. J. Fowler, M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, *Survey Methodology*, 2nd ed. Hoboken, NJ: John Wiley & Sons, 2009.

[123] P. Duarte Silva and P. Brito, *MAINT.Data: Model and Analyse Interval Data*, 2025, R package. [Online]. Available: https://CRAN.R-project.org/package=MAINT.Data

[124] O. Rodriguez, *RSDA: R to Symbolic Data Analysis*, 2025, R package version 3.2.5. [Online]. Available: https://CRAN.R-project.org/package=RSDA

[125] J. García-García, *IntervalQuestionStat: Tools to Deal with Interval-Valued Responses in Questionnaires*, 2022, R package. [Online]. Available: https://CRAN.R-project.org/package=IntervalQuestionStat

[126] Y. Zhao, C. Wagner, V. Kreinovich, B. Ryan, and D. Pekaslan, "Making sense of uncertain and vague information through interval summaries," 2025, manuscript under review.

[127] Y. Zhao, C. Wagner, B. Ryan, and D. Pekaslan, "ISurvey: An R toolkit for interval-valued survey data analysis," in *Proc. IEEE Int. Conf. Fuzzy Systems*, 2026.

Appendix A

Mathematical Properties of α_H

Let $\overline{\mathbf{X}}_j = \{\overline{X}_{ij}\}_{i=1}^n$ denote the sample of interval-valued responses to item j, for $j = 1, \ldots, k$, where n is the number of respondents and $k \geqslant 2$ is the number of items. Using interval addition via the Minkowski sum in Eq. (2.15), the total score for respondent i is defined in Chapter 3 by

$$\overline{T}_i = \sum_{j=1}^{k} \overline{X}_{ij}. \tag{A.1}$$

The sample of total scores is denoted by $\overline{\mathbf{T}} = \{\overline{T}_i\}_{i=1}^n$.

By replacing the classical variance terms in Cronbach's α with the corresponding Hausdorff-based Fréchet variances, computed using the width-fixing approach reviewed in Section 2.3.3, the coefficient α_H is defined in Eq. (3.10) for $\mathrm{Var}_H^*(\overline{\mathbf{T}}) > 0$ as

$$\alpha_H = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \mathrm{Var}_H^*(\overline{\mathbf{X}}_j)}{\mathrm{Var}_H^*(\overline{\mathbf{T}})}\right). \tag{A.2}$$

Throughout this section, we assume that the interval-valued responses lie in $\mathcal{K}_c(\mathbb{R})$, that $k \geqslant 2$, and that $\mathrm{Var}_H^*(\overline{\mathbf{T}}) > 0$. Under these conditions, we prove that the coefficient α_H defined above satisfies the three main mathematical properties considered in Chapter 3, namely: (1) boundedness and its extreme cases, (2) affine invariance, and (3) reduction to the classical Cronbach's α for degenerate intervals.

We recall that for two intervals $\overline{A} = [A^-, A^+]$ and $\overline{B} = [B^-, B^+]$, the Hausdorff distance is reviewed in Section 2.3.3 by

$$d_H(\overline{A}, \overline{B}) = \max\{|A^- - B^-|,\ |A^+ - B^+|\}. \tag{A.3}$$

For a sample of interval-valued data $\overline{\mathbf{X}} = \{\overline{X}_i\}_{i=1}^n$, under the width-fixing

approach reviewed in Section 2.3.3, the Fréchet mean under d_H is given by

$$\overline{X}^*_H = \arg\min_{\overline{X}} \frac{1}{n} \sum_{i=1}^{n} d_H^2(\overline{X}_i, \overline{X}), \tag{A.4}$$

and the corresponding Fréchet variance is

$$\mathrm{Var}^*_H(\overline{\mathbf{X}}) = \frac{1}{n} \sum_{i=1}^{n} d_H^2(\overline{X}_i, \overline{X}^*_H). \tag{A.5}$$

Definition 1. *Under the width-fixing approach, a candidate interval $\overline{V}$ is said to be* feasible *for the minimisation defining the Hausdorff Fréchet mean of a sample $\overline{\mathbf{X}} = \{\overline{X}_i\}_{i=1}^n$ if it satisfies*

$$|\overline{V}| = \frac{1}{n} \sum_{i=1}^{n} |\overline{X}_i|. \tag{A.6}$$

Lemma 1. For any intervals $\overline{A}_1, \ldots, \overline{A}_k, \overline{B}_1, \ldots, \overline{B}_k$, there holds

$$d_H\left(\sum_{j=1}^{k} \overline{A}_j,\ \sum_{j=1}^{k} \overline{B}_j \right) \ \leqslant\ \sum_{j=1}^{k} d_H(\overline{A}_j, \overline{B}_j).$$

Proof. For $j = 1, \ldots, k$, let $\overline{A}_j = [A_j^-, A_j^+]$ and $\overline{B}_j = [B_j^-, B_j^+]$. By interval addition via the Minkowski sum,

$$\sum_{j=1}^{k} \overline{A}_j = \left[\sum_{j=1}^{k} A_j^-,\ \sum_{j=1}^{k} A_j^+ \right], \qquad \sum_{j=1}^{k} \overline{B}_j = \left[\sum_{j=1}^{k} B_j^-,\ \sum_{j=1}^{k} B_j^+ \right].$$

By Eq. (A.3), together with the triangle inequality,

$$\begin{aligned}
d_H\left(\sum_{j=1}^{k} \overline{A}_j,\ \sum_{j=1}^{k} \overline{B}_j \right) &= \max\left\{ \left| \sum_{j=1}^{k} A_j^- - \sum_{j=1}^{k} B_j^- \right|, \left| \sum_{j=1}^{k} A_j^+ - \sum_{j=1}^{k} B_j^+ \right| \right\} \\
&\leqslant \max\left\{ \sum_{j=1}^{k} |A_j^- - B_j^-|, \sum_{j=1}^{k} |A_j^+ - B_j^+| \right\} \\
&\leqslant \sum_{j=1}^{k} \max\left\{ |A_j^- - B_j^-|, |A_j^+ - B_j^+| \right\} \\
&= \sum_{j=1}^{k} d_H(\overline{A}_j, \overline{B}_j).
\end{aligned}$$

□

A.1 Upper Boundedness and Extreme Cases

Proposition 1. Upper Boundedness. There holds $\alpha_H \leqslant 1$.

Proof. By Eq. (A.2), proving $\alpha_H \leqslant 1$ is equivalent to showing that

$$\mathrm{Var}_H^*(\overline{\mathbf{T}}) \leqslant k \sum_{j=1}^{k} \mathrm{Var}_H^*(\overline{\mathbf{X}}_j).$$

Let

$$\overline{V} = \sum_{j=1}^{k} \overline{X}_{H,j}^*.$$

To obtain an upper bound for $\mathrm{Var}_H^*(\overline{\mathbf{T}})$, it remains to verify that $\overline{V}$ is feasible for the minimisation defining the Hausdorff Fréchet mean of $\overline{\mathbf{T}}$.

For each $j = 1, \ldots, k$, since $\overline{X}_{H,j}^*$ is feasible for the minimisation defining the Hausdorff Fréchet mean of $\overline{\mathbf{X}}_j$, by Eq. (A.6), there is

$$\left|\overline{X}_{H,j}^*\right| = \frac{1}{n} \sum_{i=1}^{n} \left|\overline{X}_{ij}\right|.$$

Therefore, by interval addition via the Minkowski sum and Eq. (A.1), there is

$$\begin{aligned} |\overline{V}| &= \left| \sum_{j=1}^{k} \overline{X}_{H,j}^* \right| \\ &= \sum_{j=1}^{k} \left|\overline{X}_{H,j}^*\right| \\ &= \sum_{j=1}^{k} \frac{1}{n} \sum_{i=1}^{n} \left|\overline{X}_{ij}\right| \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{k} \left|\overline{X}_{ij}\right| \\ &= \frac{1}{n} \sum_{i=1}^{n} \left|\overline{T}_i\right|. \end{aligned}$$

Hence, by Definition 1, $\overline{V}$ is feasible for the minimisation defining the Hausdorff Fréchet mean of $\overline{\mathbf{T}}$. Therefore,

$$\mathrm{Var}_H^*(\overline{\mathbf{T}}) = \frac{1}{n} \sum_{i=1}^{n} d_H^2(\overline{T}_i, \overline{T}_H^*) \leqslant \frac{1}{n} \sum_{i=1}^{n} d_H^2(\overline{T}_i, \overline{V}).$$

By Lemma 1 and Eq. (A.1), there is

$$d_H(\overline{T}_i, \overline{V}) = d_H\left(\sum_{j=1}^{k} \overline{X}_{ij}, \sum_{j=1}^{k} \overline{X}^*_{H,j}\right) \leqslant \sum_{j=1}^{k} d_H(\overline{X}_{ij}, \overline{X}^*_{H,j}).$$

Thus, by the Cauchy–Schwarz inequality,

$$d^2_H(\overline{T}_i, \overline{V}) \leqslant k \sum_{j=1}^{k} d^2_H(\overline{X}_{ij}, \overline{X}^*_{H,j}).$$

Averaging over $i = 1, \ldots, n$ gives

$$\frac{1}{n}\sum_{i=1}^{n} d^2_H(\overline{T}_i, \overline{V}) \leqslant k \sum_{j=1}^{k} \frac{1}{n}\sum_{i=1}^{n} d^2_H(\overline{X}_{ij}, \overline{X}^*_{H,j}) = k \sum_{j=1}^{k} \mathrm{Var}^*_H(\overline{\mathbf{X}}_j).$$

Therefore,

$$\mathrm{Var}^*_H(\overline{\mathbf{T}}) \leqslant k \sum_{j=1}^{k} \mathrm{Var}^*_H(\overline{\mathbf{X}}_j),$$

and hence $\alpha_H \leqslant 1$. □

We also recall the following positive homogeneity property. For any intervals $\overline{A}, \overline{B}$ and any $c \in \mathbb{R}$ with $c > 0$, there is

$$|c\,\overline{A}| = c\,|\overline{A}|, \qquad d_H(c\,\overline{A}, c\,\overline{B}) = c\,d_H(\overline{A}, \overline{B}). \tag{A.7}$$

Proposition 2. Upper extreme case. If, for each $i = 1, \ldots, n$,

$$\overline{X}_{i1} = \cdots = \overline{X}_{ik},$$

and these common intervals are not all identical across respondents, then $\alpha_H = 1$.

Proof. Since $\overline{X}_{i1} = \cdots = \overline{X}_{ik}$ for all $i = 1, \ldots, n$, the item samples $\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_k$ are identical. Therefore, there is

$$\overline{X}^*_{H,1} = \cdots = \overline{X}^*_{H,k} \qquad \text{and} \qquad \mathrm{Var}^*_H(\overline{\mathbf{X}}_1) = \cdots = \mathrm{Var}^*_H(\overline{\mathbf{X}}_k).$$

Also, by Eq. (A.1), there is

$$\overline{T}_i = \sum_{j=1}^{k} \overline{X}_{ij} = k\,\overline{X}_{i1} \qquad \text{for all } i = 1, \ldots, n.$$

Since these common intervals are not all identical across respondents, there is

$$\mathrm{Var}^*_H(\overline{\mathbf{X}}_1) > 0.$$

Now let $\overline{V}$ be any feasible interval for the minimisation defining $\overline{T}^*_H$. By Definition 1, there is

$$|\overline{V}| = \frac{1}{n}\sum_{i=1}^{n} |\overline{T}_i| = \frac{1}{n}\sum_{i=1}^{n} |k\,\overline{X}_{i1}| = k \cdot \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{i1}|.$$

Hence, by Eq. (A.7), there exists a unique interval $\overline{W} \in \mathcal{K}_c(\mathbb{R})$ such that

$$\overline{V} = k\,\overline{W}$$

and

$$|\overline{W}| = \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{i1}|.$$

Therefore, by Definition 1, $\overline{W}$ is feasible for the minimisation defining $\overline{X}^*_{H,1}$.

By Eq. (A.7), there is

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{T}_i, \overline{V}) &= \frac{1}{n}\sum_{i=1}^{n} d_H^2(k\,\overline{X}_{i1}, k\,\overline{W}) \\ &= k^2 \cdot \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{X}_{i1}, \overline{W}) \\ &\geqslant k^2\,\mathrm{Var}^*_H(\overline{\mathbf{X}}_1),\end{aligned}$$

with equality when $\overline{W} = \overline{X}^*_{H,1}$. Therefore, there is

$$\overline{T}^*_H = k\,\overline{X}^*_{H,1} \qquad \text{and} \qquad \mathrm{Var}^*_H(\overline{\mathbf{T}}) = k^2\,\mathrm{Var}^*_H(\overline{\mathbf{X}}_1).$$

Hence, by Eq. (A.2), there is

$$\alpha_H = \frac{k}{k-1}\left(1 - \frac{k\,\mathrm{Var}^*_H(\overline{\mathbf{X}}_1)}{k^2\,\mathrm{Var}^*_H(\overline{\mathbf{X}}_1)}\right) = \frac{k}{k-1}\left(1 - \frac{1}{k}\right) = 1.$$

Therefore, $\alpha_H = 1$. □

Proposition 3. Zero case. If

$$\mathrm{Var}^*_H(\overline{\mathbf{T}}) = \sum_{j=1}^{k} \mathrm{Var}^*_H(\overline{\mathbf{X}}_j),$$

then $\alpha_H = 0$.

Proof. By Eq. (A.2), if

$$\mathrm{Var}^*_H(\overline{\mathbf{T}}) = \sum_{j=1}^{k} \mathrm{Var}^*_H(\overline{\mathbf{X}}_j),$$

then

$$\alpha_H = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \mathrm{Var}_H^*(\overline{\mathbf{X}}_j)}{\mathrm{Var}_H^*(\overline{\mathbf{T}})}\right) = \frac{k}{k-1}(1-1) = 0.$$

Therefore, $\alpha_H = 0$. □

Remark. Negative values $\alpha_H < 0$ can occur when

$$\mathrm{Var}_H^*(\overline{\mathbf{T}}) < \sum_{j=1}^{k} \mathrm{Var}_H^*(\overline{\mathbf{X}}_j),$$

for example when items are negatively associated. This parallels the behaviour of the classical Cronbach's α for single-valued data.

A.2 Affine Invariance

We also recall the following affine invariance property. For any intervals $\overline{A}, \overline{B}$, any $a \in \mathbb{R}$, and any $b \in \mathbb{R}\backslash\{0\}$, there is

$$|a + b\,\overline{A}| = |b|\,|\overline{A}|, \qquad d_H(a + b\,\overline{A}, a + b\,\overline{B}) = |b|\,d_H(\overline{A}, \overline{B}). \tag{A.8}$$

Proposition 4. Affine invariance. Let

$$\overline{Y}_{ij} = a + b\,\overline{X}_{ij}, \qquad a \in \mathbb{R}, \quad b \in \mathbb{R}\backslash\{0\},$$

and let $\overline{\mathbf{Y}}_j = \{\overline{Y}_{ij}\}_{i=1}^n$ denote the transformed sample for item j, for $j = 1, \ldots, k$. Let

$$\overline{T}_i^X = \sum_{j=1}^{k} \overline{X}_{ij}, \qquad \overline{T}_i^Y = \sum_{j=1}^{k} \overline{Y}_{ij},$$

and let

$$\overline{\mathbf{T}}^X = \{\overline{T}_i^X\}_{i=1}^n, \qquad \overline{\mathbf{T}}^Y = \{\overline{T}_i^Y\}_{i=1}^n.$$

If $\mathrm{Var}_H^*(\overline{\mathbf{T}}^X) > 0$, then the corresponding Hausdorff-based coefficients computed from $\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_k$ and $\overline{\mathbf{Y}}_1, \ldots, \overline{\mathbf{Y}}_k$ are equal.

Proof. Let α_H^X and α_H^Y denote the values of α_H computed from $\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_k$ and $\overline{\mathbf{Y}}_1, \ldots, \overline{\mathbf{Y}}_k$, respectively.

It is sufficient to show that, under the affine transformation, both the item-level Hausdorff Fréchet variances and the total-score Hausdorff Fréchet variance are multiplied by the same factor b^2.

For each $j = 1, \ldots, k$, since $\overline{X}_{H,j}^*$ is feasible for the minimisation defining the

Hausdorff Fréchet mean of $\overline{\mathbf{X}}_j$, by Eq. (A.6), there is

$$|\overline{X}^*_{H,j}| = \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{ij}|.$$

Hence, by Eq. (A.8), there is

$$\begin{aligned} |a + b\,\overline{X}^*_{H,j}| &= |b|\,|\overline{X}^*_{H,j}| \\ &= |b| \cdot \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{ij}| \\ &= \frac{1}{n}\sum_{i=1}^{n} |a + b\,\overline{X}_{ij}| \\ &= \frac{1}{n}\sum_{i=1}^{n} |\overline{Y}_{ij}|. \end{aligned}$$

Therefore, by Definition 1, $a + b\,\overline{X}^*_{H,j}$ is feasible for the minimisation defining $\overline{Y}^*_{H,j}$.

Now let $\overline{V}^Y$ be any feasible interval for the minimisation defining $\overline{Y}^*_{H,j}$. By Definition 1, there is

$$|\overline{V}^Y| = \frac{1}{n}\sum_{i=1}^{n} |\overline{Y}_{ij}| = |b| \cdot \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{ij}|.$$

Hence, there exists a unique interval $\overline{V}^X \in \mathcal{K}_c(\mathbb{R})$ such that

$$\overline{V}^Y = a + b\,\overline{V}^X$$

and

$$|\overline{V}^X| = \frac{1}{n}\sum_{i=1}^{n} |\overline{X}_{ij}|.$$

Therefore, by Definition 1, $\overline{V}^X$ is feasible for the minimisation defining $\overline{X}^*_{H,j}$.

By Eq. (A.8), there is

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{Y}_{ij}, \overline{V}^Y) &= \frac{1}{n}\sum_{i=1}^{n} d_H^2(a + b\,\overline{X}_{ij},\, a + b\,\overline{V}^X) \\ &= b^2 \cdot \frac{1}{n}\sum_{i=1}^{n} d_H^2(\overline{X}_{ij}, \overline{V}^X) \\ &\geqslant b^2\,\mathrm{Var}^*_H(\overline{\mathbf{X}}_j), \end{aligned}$$

with equality when $\overline{V}^X = \overline{X}^*_{H,j}$. Therefore, there is

$$\overline{Y}^*_{H,j} = a + b\,\overline{X}^*_{H,j} \qquad \text{and} \qquad \operatorname{Var}^*_H(\overline{\mathbf{Y}}_j) = b^2 \operatorname{Var}^*_H(\overline{\mathbf{X}}_j).$$

By interval addition via the Minkowski sum and Eq. (A.1), there is

$$\overline{T}^Y_i = \sum_{j=1}^{k} \overline{Y}_{ij} = \sum_{j=1}^{k} (a + b\,\overline{X}_{ij}) = ka + b \sum_{j=1}^{k} \overline{X}_{ij} = ka + b\,\overline{T}^X_i.$$

Applying the same argument to the total-score samples, there is

$$\operatorname{Var}^*_H(\overline{\mathbf{T}}^Y) = b^2 \operatorname{Var}^*_H(\overline{\mathbf{T}}^X).$$

Since $\operatorname{Var}^*_H(\overline{\mathbf{T}}^X) > 0$ and $b \neq 0$, there is

$$\operatorname{Var}^*_H(\overline{\mathbf{T}}^Y) > 0.$$

Hence, by Eq. (A.2), there is

$$\begin{aligned}
\alpha^Y_H &= \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \operatorname{Var}^*_H(\overline{\mathbf{Y}}_j)}{\operatorname{Var}^*_H(\overline{\mathbf{T}}^Y)}\right) \\
&= \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} b^2 \operatorname{Var}^*_H(\overline{\mathbf{X}}_j)}{b^2 \operatorname{Var}^*_H(\overline{\mathbf{T}}^X)}\right) \\
&= \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k} \operatorname{Var}^*_H(\overline{\mathbf{X}}_j)}{\operatorname{Var}^*_H(\overline{\mathbf{T}}^X)}\right) = \alpha^X_H.
\end{aligned}$$

Therefore, the corresponding Hausdorff-based coefficients are equal. □

A.3 Reduction to Classical Cronbach's α

We also recall that, for any $a, b \in \mathbb{R}$, the Hausdorff distance between the corresponding degenerate intervals reduces to the absolute value:

$$d_H([a,a],[b,b]) = |a - b|. \tag{A.9}$$

Definition 2. *An interval $\overline{X}_{ij} = [X^-_{ij}, X^+_{ij}]$ is said to be* degenerate *if $X^-_{ij} = X^+_{ij}$, equivalently, if $|\overline{X}_{ij}| = 0$. In this case, it is written as*

$$\overline{x}_{ij} = [x_{ij}, x_{ij}],$$

where $x_{ij} \in \mathbb{R}$ denotes the corresponding single-valued observation.

Proposition 5. Reduction to classical Cronbach's α. If every observed interval is degenerate, i.e.

$$\overline{X}_{ij} = \overline{x}_{ij} = [x_{ij}, x_{ij}] \qquad \text{for all } i = 1, \dots, n,\ j = 1, \dots, k,$$

and

$$\mathrm{Var}^*_H(\overline{\mathbf{T}}) > 0,$$

then α_H reduces to the classical sample Cronbach's α computed from the single-valued data $\{x_{ij}\}$.

Proof. Fix $j \in \{1, \dots, k\}$. Since each observed interval $\overline{X}_{ij}$ is degenerate, there is

$$|\overline{X}_{ij}| = |\overline{x}_{ij}| = 0 \qquad \text{for all } i = 1, \dots, n.$$

By Eq. (A.6), any feasible interval for the minimisation defining $\overline{X}^*_{H,j}$ must therefore have width 0. Hence, any such feasible interval can be written as

$$\overline{V} = \overline{v} = [v, v] \qquad \text{for some } v \in \mathbb{R}.$$

By Eq. (A.9), there is

$$d_H(\overline{X}_{ij}, \overline{V}) = d_H(\overline{x}_{ij}, \overline{v}) = d_H([x_{ij}, x_{ij}], [v, v]) = |x_{ij} - v|.$$

Therefore, the minimisation defining $\overline{X}^*_{H,j}$ reduces to

$$\min_{v \in \mathbb{R}} \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - v)^2.$$

Hence, its minimiser is

$$\hat{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij},$$

and therefore

$$\overline{X}^*_{H,j} = [\hat{x}_j, \hat{x}_j] = \overline{\hat{x}}_j.$$

Thus, for each $j = 1, \dots, k$, there is

$$\mathrm{Var}^*_H(\overline{\mathbf{X}}_j) = \frac{1}{n} \sum_{i=1}^{n} d^2_H(\overline{X}_{ij}, \overline{X}^*_{H,j}) = \frac{1}{n} \sum_{i=1}^{n} d^2_H(\overline{x}_{ij}, \overline{\hat{x}}_j) = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \hat{x}_j)^2.$$

Also, by Eq. (A.1), there is

$$\overline{T}_i = \sum_{j=1}^{k} \overline{X}_{ij} = \sum_{j=1}^{k} \overline{x}_{ij} = \left[\sum_{j=1}^{k} x_{ij},\ \sum_{j=1}^{k} x_{ij} \right] = \overline{t}_i,$$

where

$$\bar{t}_i = [t_i, t_i] \qquad \text{and} \qquad t_i = \sum_{j=1}^{k} x_{ij}.$$

Let

$$\hat{t} = \frac{1}{n}\sum_{i=1}^{n} t_i, \qquad \bar{\hat{t}} = [\hat{t}, \hat{t}].$$

Applying the same argument to the total-score sample, there is

$$\mathrm{Var}_H^*(\overline{\mathbf{T}}) = \frac{1}{n}\sum_{i=1}^{n} d_H^2(\bar{t}_i, \bar{\hat{t}}) = \frac{1}{n}\sum_{i=1}^{n}(t_i - \hat{t})^2.$$

Hence, by Eq. (A.2), there is

$$\alpha_H = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k}\mathrm{Var}_H^*(\overline{\mathbf{X}}_j)}{\mathrm{Var}_H^*(\overline{\mathbf{T}})}\right) = \frac{k}{k-1}\left(1 - \frac{\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \hat{x}_j)^2}{\frac{1}{n}\sum_{i=1}^{n}(t_i - \hat{t})^2}\right).$$

This is the classical sample Cronbach's α computed from $\{x_{ij}\}$, since the common variance normalisation cancels in the ratio. Therefore, α_H reduces to the classical sample Cronbach's α. □

Appendix B

Proofs of Satisfying Properties

For a fuzzy set A, let its membership function be a continuous function $\mu_A(x)$ defined on a (bounded or unbounded) interval domain $X = [x^-, x^+]$.

By applying the DU (or DUW) methods to A as described in Chapter 4, the interval-valued defuzzification result of A is a discontinuous interval, which can be represented as the union of multiple continuous sub-intervals:

$$DU(A) \text{ or } DUW(A) = \bar{\bar{A}} = \bigcup_{i=1}^{n} \overline{A_i}. \tag{B.1}$$

In this section, we prove that such $\bar{\bar{A}}$ satisfies the four desirable properties described in Chapter 4, namely: (1) Support Subset, (2) Core Inclusion, (3) Peak/Sub-Interval Ratio, and (4) Proportionality.

In what follows, since we work directly with the membership function of A, for notational simplicity we write $\mu(x) = \mu_A(x)$ and $\overline{A_i} = [x_i^-, x_i^+]$, with $x \in X$.

Let us recall some basic definitions.

Definition 1. *A continuous function $\mu(x)$ on the interval $X = [x^-, x^+]$ is said to be* piece-wise monotonic *if X can be represented as a union of finitely many (bounded or unbounded) sub-intervals on each of which $\mu(x)$ is either (non-strictly) increasing or (non-strictly) decreasing.*

Discussion. All practically used membership functions are piece-wise monotonic. Without loss of generality, in the following text, we will only consider piece-wise monotonic membership functions. Typical examples are convex membership functions.

Definition 2. *We say that a membership function defined on an interval $X = [x^-, x^+]$ is* convex *if for all triples $x' < x < x''$, we have*

$$\mu(x) \geqslant \min(\mu(x'), \mu(x'')).$$

Discussion. It is known that a convex membership function is piece-wise monotonic: it first non-strictly increases and then non-strictly decreases. Many membership functions are of this type, including the popular triangular and trapezoidal ones. Each convex membership function has two monotonicity intervals: it is increasing on the first one and decreasing on the second one.

Definition 3.

- *We say that a function* $\mu(x)$ *has a* local maximum *at a point* x_0 *if for some* $\varepsilon > 0$*, we have* $\mu(x_0) \geqslant \mu(x)$ *for all* $x \in [x_0 - \varepsilon, x_0 + \varepsilon]$.
- *We say that a function* $\mu(x)$ *has a* local minimum *at a point* x_0 *if for some* $\varepsilon > 0$*, we have* $\mu(x_0) \leqslant \mu(x)$ *for all* $x \in [x_0 - \varepsilon, x_0 + \varepsilon]$.

Discussion. For a continuous function, the sets of all local maxima and of all local minima are each unions of disjoint points or finite (possibly infinite) intervals. Since a single point x can be viewed as a degenerate interval $[x, x]$, we can refer to them simply as unions of disjoint intervals; see Fig. B.1.

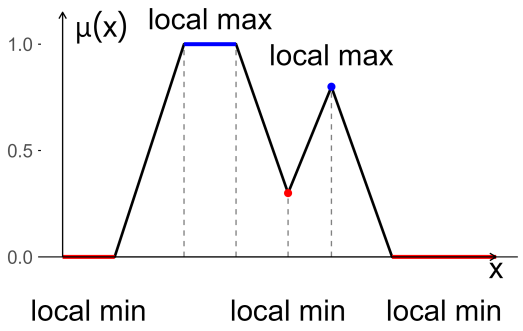


Figure B.1: *Example of local maxima and local minima of a continuous membership function* $\mu(x)$

In particular, for a piece-wise monotonic function, these sets are finite unions of disjoint intervals. Each local maximum interval corresponds to a transition from increasing to decreasing, and each local minimum interval corresponds to a transition from decreasing to increasing. In the following text, by a local minimum or maximum interval, we mean one of these intervals.

For example:

- A triangular membership function increasing from ℓ^- to u and then decreasing from u to ℓ^+ has a local maximum interval $[u, u]$ and two local minimum intervals $(-\infty, \ell^-]$ and $[\ell^+, +\infty)$;
- A trapezoidal membership function increasing from ℓ^- to u^-, then constant $\mu(x) = 1$ from u^- to u^+, and then decreasing from u^+ to ℓ^+ has one local maximum interval $[u^-, u^+]$ and two local minimum intervals $(-\infty, \ell^-]$ and $[\ell^+, +\infty)$.

The *local maximum value* of a local maximum interval is the value of $\mu(x)$ at each point within that interval.

In general, a membership function is convex if and only if it has only one local maximum interval. Every non-convex function has two or more local maximum intervals $[X_i^-, X_i^+]$.

Definition 4. *Let $\mu(x)$ have p local maxima (peaks), denoted by the intervals*

$$[X_1^-, X_1^+], [X_2^-, X_2^+], \ldots, [X_p^-, X_p^+],$$

ordered such that

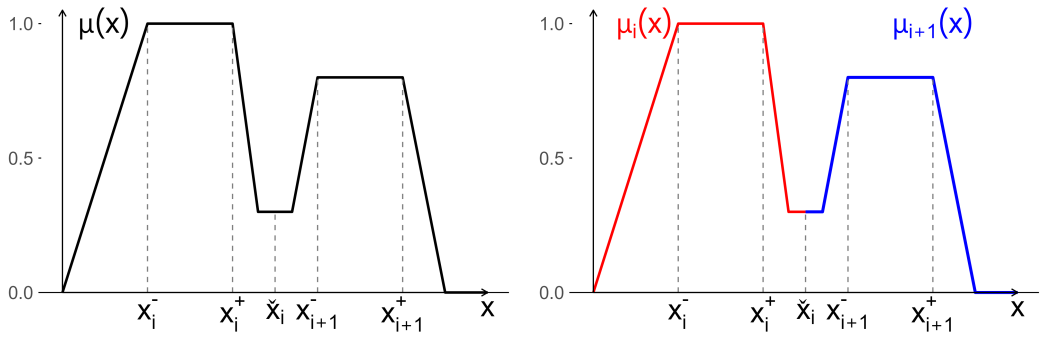
$$X_1^- \leqslant X_1^+ < X_2^- \leqslant X_2^+ < \cdots < X_p^- \leqslant X_p^+.$$

Between each pair of consecutive peaks there exists a local minimum (trough) interval, hence there are $(p-1)$ troughs in total. We denote the trough interval midpoint *between the i-th and $(i+1)$-st peak as $\breve{x}_i$, with the* corresponding membership value *$\breve{\mu}_i = \mu(\breve{x}_i)$, for $i = 1, \ldots, p-1$. Let $\breve{x}_0$ and $\breve{x}_p$ denote the left and right boundaries of the support of $\mu(x)$.*

The $DU(W)$ method decomposes $\mu(x)$ by splitting it at the points $\breve{x}_i$ for $i = 1, \ldots, p-1$. Together with the boundary points $\breve{x}_0$ and $\breve{x}_p$, these split X into p sub-intervals $[\breve{x}_{i-1}, \breve{x}_i]$, $i = 1, \ldots, p$. Let $\mu_i(x)$ denote the restriction of $\mu(x)$ to the interval $[\breve{x}_{i-1}, \breve{x}_i]$. As a result, $\mu(x)$ is represented as the point-wise maximum of p quasi-disjoint convex subsets with membership functions $\mu_1(x), \ldots, \mu_p(x)$:

$$\mu(x) = \max\{\mu_1(x), \ldots, \mu_p(x)\}, \quad x \in X,$$

see Fig. B.2.



(a) **(b)**

Figure B.2: *(a) Illustration of the peak intervals $[X_i^-, X_i^+]$ and the midpoints $\breve{x}_i$ of the trough in a non-convex membership function $\mu(x)$. (b) Decomposition of $\mu(x)$ into quasi-disjoint convex subsets $\mu_i(x)$ and $\mu_{i+1}(x)$ by splitting at the trough midpoints $\breve{x}_i$.*

Proposition 1. Properties 1) *Support-subset* and 2) *Core-inclusion*

are satisfied.

Proof. For each convex subset $\mu_i(x)$ obtained by the $DU(W)$ decomposition, it has a core $\{x \mid \mu_i(x) = 1\}$ and support $\{x \mid \mu_i(x) > 0\}$. From [61], see Section 2.1.2, we know that for each subset:

$$\{x \mid \mu_i(x) = 1\} \subseteq [x_i^-, x_i^+] \subseteq \{x \mid \mu_i(x) > 0\}.$$

Since the support of $\mu(x)$ is the union of the supports of all subsets, we have:

$$\bigcup_{i=1}^{p} [x_i^-, x_i^+] \subseteq \bigcup_{i=1}^{p} \{x \mid \mu_i(x) > 0\} = \{x \mid \mu(x) > 0\}.$$

Similarly, any point where $\mu(x) = 1$ must have $\mu_i(x) = 1$ for at least one subset $\mu_i(x)$. Thus,

$$\{x \mid \mu(x) = 1\} \subseteq \bigcup_{i=1}^{p} \{x \mid \mu_i(x) = 1\} \subseteq \bigcup_{i=1}^{p} [x_i^-, x_i^+].$$

Therefore, both Property 1) and 2) are satisfied. □

Proposition 2. Property 3), *Peak/Sub-Interval Ratio*, is satisfied.

We recall that a membership function $\mu(x)$ with p peaks is decomposed by the $DU(W)$ method into p convex subsets $\mu_1(x), \ldots, \mu_p(x)$. Applying IVD or $WIVD$ to each subset yields p sub-intervals $[x_1^-, x_1^+], \ldots, [x_p^-, x_p^+]$, whose union forms the overall result. To ensure that the number of resulting intervals equals the number of peaks, it remains to prove that these sub-intervals are pairwise disjoint.

Proof. We show that the p intervals $[x_1^-, x_1^+], \ldots, [x_p^-, x_p^+]$ obtained by the $DU(W)$ method are pairwise disjoint, i.e., $x_i^+ < x_{i+1}^-$ for all $i = 1, \ldots, p-1$.

- By construction of the $DU(W)$ decomposition, for the i-th convex subset $\mu_i(x)$, the right endpoint of its α-cut satisfies $x_i^+(\alpha) \leqslant \breve{x}_i$ for all $\alpha > 0$.
- Similarly, for the $(i+1)$-st subset $\mu_{i+1}(x)$, the left endpoint of its α-cut satisfies $x_{i+1}^-(\alpha) \geqslant \breve{x}_i$ for all $\alpha > 0$.
- Since there is a genuine trough between consecutive peaks (i.e., the membership function decreases from the i-th peak and then increases to the $(i+1)$-st peak), there exists at least one $\alpha > \breve{\mu}_i$ such that $x_i^+(\alpha) < \breve{x}_i < x_{i+1}^-(\alpha)$.

When applying IVD, the crisp interval endpoints are obtained as the integrals (means) of the corresponding α-cut boundaries:

$$x_i^+ = \int_0^1 x_i^+(\alpha)\, d\alpha \quad \text{and} \quad x_{i+1}^- = \int_0^1 x_{i+1}^-(\alpha)\, d\alpha.$$

Since $x_i^+(\alpha) \leqslant x_{i+1}^-(\alpha)$ for all $\alpha \in [0,1]$, and strict inequality holds for at least one α, we have

$$x_i^+ = \int_0^1 x_i^+(\alpha)\, d\alpha < \int_0^1 x_{i+1}^-(\alpha)\, d\alpha = x_{i+1}^-.$$

Similarly, when applying $WIVD$, where α-levels serve as weights, we obtain:

$$x_i^+ = \frac{\int_0^1 \alpha\, x_i^+(\alpha)\, d\alpha}{\int_0^1 \alpha\, d\alpha} < \frac{\int_0^1 \alpha\, x_{i+1}^-(\alpha)\, d\alpha}{\int_0^1 \alpha\, d\alpha} = x_{i+1}^-.$$

Therefore, the resulting intervals $[x_i^-, x_i^+]$ are pairwise disjoint, and the $DU(W)$ method thus produces p sub-intervals for a membership function with p peaks.

Hence, Property 3) is satisfied. □

Definition 5. *We say that a membership function $\mu(x)$ is* smaller than or equal to *another membership function $\mu'(x)$ if $\mu(x) \leqslant \mu'(x)$ for all x. Let S be a class of membership functions. We say that $\mu_0 \in S$ is the* smallest *in S if $\mu_0(x) \leqslant \mu(x)$ for all $\mu \in S$.*

Proposition 3. For each piece-wise monotonic membership function $\mu(x)$, in the class of all convex functions $\mu'(x)$ for which $\mu(x) \leqslant \mu'(x)$, there exists the smallest membership function; we denote it by $\overline{\mu}(x)$.

Comment. This smallest membership function has the form

$$\overline{\mu}(x) = \sup\{\min(\mu(x'), \mu(x'')) : x' \leqslant x \leqslant x''\},$$

and is called the *convex hull* of $\mu(x)$; see Fig. B.3.

Definition 6. *Let $\mu(x)$ be a non-convex piece-wise monotonic membership function, and let $[X_i^-, X_i^+]$ and $[X_{i+1}^-, X_{i+1}^+]$ be two neighboring local maxima intervals of this function. Let $m_i(x)$ denote the restriction of $\mu(x)$ to the interval $[X_i^+, X_{i+1}^-]$, and let $\overline{m}_i(x)$ be the convex hull of $m_i(x)$. The i-th* trough area *t_i is defined as the area between the graphs of $m_i(x)$ and $\overline{m}_i(x)$; see Fig. B.2a and Fig. B.4.*

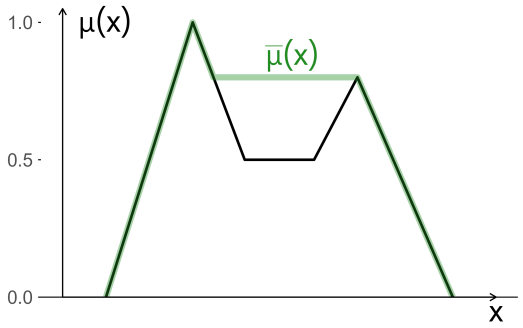
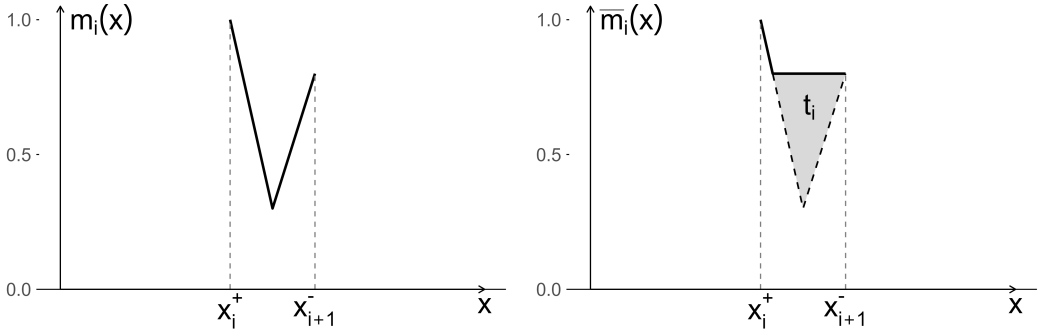


Figure B.3: *Illustration of the convex hull $\overline{\mu}(x)$ of a non-convex membership function $\mu(x)$.*



(a) **(b)**

Figure B.4: *(a) Restriction $m_i(x)$ of a non-convex membership function $\mu(x)$ to the interval $[X_i^+, X_{i+1}^-]$ between two consecutive peaks. (b) The convex hull $\overline{m}_i(x)$ of $m_i(x)$, and the corresponding trough area t_i defined as the area between $m_i(x)$ and $\overline{m}_i(x)$.*

Definition 7. *Let* $[x_1^-, x_1^+], \ldots, [x_n^-, x_n^+]$ *be intervals obtained from* $\mu(x)$ *using the DU(W) method. The i-th* gap width g_i *is defined as* $g_i = x_{i+1}^- - x_i^+$.

Proposition 4. Property 4), *Proportionality*, is satisfied.

- When the local maximum values on both sides of the trough are equal to 1, then $t_i = g_i$.
- In general, $t_i \leqslant g_i$ for all i.

We first discuss the *DU* method, and then extend the result to its weighted alternative *DUW*.

Proof.

Case of *DU*.

1°. Let us first consider the case when the local maximum values on both sides of the trough are 1.

To prove the Proposition for this case, we will:

- first compute the trough area t_i,
- then compute the gap width g_i, and
- then prove that in this case, they are equal to each other.

In this case, the convex hull function $\overline{m}_i(x)$ is equal to 1 for all $x \in [X_i^+, X_{i+1}^-]$; see Fig. B.5. To apply the *DU* algorithm, we consider the midpoint $\breve{x}_i$ of the local minimum interval between the i-th and $(i+1)$-st local maxima. Let $\breve{m}_i = m_i(\breve{x}_i)$ denote the membership value at this midpoint.

1.1°. Let us first compute the area t_i of the trough.

The line $x = \breve{x}_i$ divides the trough into two disjoint parts. Thus, the trough area t_i is equal to the sum $t_i = t_{i-} + t_{i+}$ of the areas t_{i-} and t_{i+} of, correspondingly, the left and the right parts. Let us compute the areas of these two parts.

1.1.1°. Let us start with computing the area t_{i-} of the left part of the trough.

To compute the area t_{i-} of the left part, it makes sense to take into account that the left side of the border of this part is formed by points $(x, m_i(x))$.

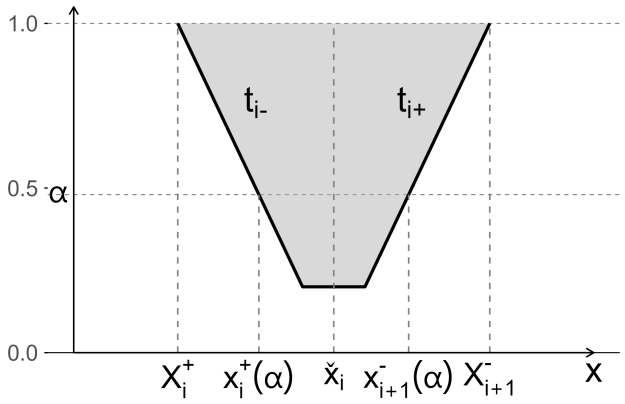


Figure B.5: *Illustration of trough area t_i between two consecutive peaks, divided by the midpoint $\breve{x}_i$ into left and right parts t_{i-} and t_{i+}.*

When writing the border points this way, we view $m_i(x)$ as a function of x, but we can also view these points in the α-cut way, as pairs $(x_i^+(\alpha), \alpha)$ describing this border as a function of α, where $x_i^+(\alpha)$ is the right-end point of the α-cut of the function $m_i(x)$ after it is limited to the decreasing interval $[X_i^+, \breve{x}_i]$:

$$x_i^+(\alpha) \stackrel{\text{def}}{=} \sup\{x \leqslant \breve{x}_i \,\&\, m_i(x) \geqslant \alpha\}. \tag{B.2}$$

This way, the left part of the trough becomes the area under the curve

$$\breve{x}_i - x_i^+(\alpha)$$

corresponding to $\alpha \in [\breve{m}_i, 1]$. By definition of an integral, the area under the curve is exactly the corresponding integral:

$$t_{i-} = \int_{\breve{m}_i}^{1} \left(\breve{x}_i - x_i^+(\alpha)\right) d\alpha. \tag{B.3}$$

1.1.2°. Let us now compute the area t_{i+} of the right part of the trough.

Similarly, the α-cut representation can help us compute the area t_{i+} of the right side of the trough. Namely, we represent each point of the right border of the right side as $(\alpha, x_{i+1}^-(\alpha))$, where $x_{i+1}^-(\alpha)$ is the left-end point of the α-cut of the function $m_i(x)$ after it is limited to the increasing interval $[\breve{x}_i, X_{i+1}^-]$:

$$x_{i+1}^-(\alpha) \stackrel{\text{def}}{=} \inf\{x \geqslant \breve{x}_i \,\&\, m_i(x) \geqslant \alpha\}. \tag{B.4}$$

This way, the right part of the trough becomes the area under the curve

$$x_{i+1}^-(\alpha) - \breve{x}_i$$

corresponding to $\alpha \in [\breve{m}_i, 1]$. By definition of an integral, the area under the

curve is exactly the corresponding integral:

$$t_{i+} = \int_{\breve{m}_i}^{1} \left(x_{i+1}^{-}(\alpha) - \breve{x}_i\right) d\alpha. \tag{B.5}$$

1.2°. Let us now compute the gap width g_i.

The gap width is the width of the gap interval $[x_i^+, x_{i+1}^-]$. The value $\breve{x}_i$ divides this interval into two parts: the left part $[x_i^+, \breve{x}_i]$ and the right part $[\breve{x}_i, x_{i+1}^-]$. Thus, the gap width g_i can be represented as the sum $g_i = g_{i-} + g_{i+}$ of the widths of these two parts: $g_{i-} = \breve{x}_i - x_i^+$ and $g_{i+} = x_{i+1}^- - \breve{x}_i$. To compute the gap width g_i, let us compute the widths of the two parts.

1.2.1°. Let us first compute the width g_{i-} of the left part of the gap.

According to the *DU* method, the value x_i^+ is obtained as the average of the right-end points of the α-cuts of the function $m_i(x)$ limited to the decreasing interval $[X_i^+, \breve{x}_i]$, i.e., the average of the function that is equal:

- to $x_i^+(\alpha)$ for $\alpha \geqslant \breve{m}_i$, and
- to $\breve{x}_i$ for $\alpha \leqslant \breve{m}_i$.

In precise terms, this average is equal to

$$x_i^+ = \int_0^{\breve{m}_i} \breve{x}_i \, d\alpha + \int_{\breve{m}_i}^{1} x_i^+(\alpha) \, d\alpha. \tag{B.6}$$

The value $\breve{x}_i$ can be represented as

$$\breve{x}_i = \int_0^1 \breve{x}_i \, d\alpha = \int_0^{\breve{m}_i} \breve{x}_i \, d\alpha + \int_{\breve{m}_i}^{1} \breve{x}_i \, d\alpha. \tag{B.7}$$

Subtracting (B.6) from (B.7) and taking into account that the difference between the integrals of the two functions is equal to the integral of their difference, we conclude that

$$g_{i-} = \int_{\breve{m}_i}^{1} \left(\breve{x}_i - x_i^+(\alpha)\right) d\alpha. \tag{B.8}$$

By comparing this expression with formula (B.3) for t_{i-}, we can see that $g_{i-} = t_{i-}$.

1.2.2°. Let us now compute the width g_{i+} of the right part of the gap.

Similarly to the previous section, the value x_{i+1}^- is obtained as the average of

the left-end points of the α-cuts of the function $m_i(x)$ limited to the increasing interval $[\breve{x}_i, X_{i+1}^-]$, i.e., the average of the function that is equal:

- to $x_{i+1}^-(\alpha)$ for $\alpha \geqslant \breve{m}_i$, and
- to $\breve{x}_i$ for $\alpha \leqslant \breve{m}_i$.

In precise terms, this average is equal to

$$x_{i+1}^- = \int_0^{\breve{m}_i} \breve{x}_i \, d\alpha + \int_{\breve{m}_i}^1 x_{i+1}^-(\alpha) \, d\alpha. \tag{B.9}$$

Subtracting (B.7) from (B.9) and taking into account that the difference between the integrals of the two functions is equal to the integral of their difference, we conclude that

$$g_{i+} = \int_{\breve{m}_i}^1 \left(x_{i+1}^-(\alpha) - \breve{x}_i\right) d\alpha. \tag{B.10}$$

By comparing this expression with formula (B.5) for t_{i+}, we can see that $g_{i+} = t_{i+}$.

1.3°. Now we can conclude the proof of the first statement of Proposition 4.

Indeed, since $g_{i-} = t_{i-}$ and $g_{i+} = t_{i+}$, we have

$$g_i = g_{i-} + g_{i+} = t_{i-} + t_{i+} = t_i.$$

2°. Let us now consider the general case, when the local maxima values on both sides of the trough may be smaller than 1.

Let us analyse what happens on both sides of the value $\breve{x}_i$.

2.1°. Let us analyse what happens on the left side of the value $\breve{x}_i$.

In this case, according to the *DU* algorithm, to compute x_i^+, we normalise the segment of the membership function $m_i(x)$ limited to the decreasing interval $[X_i^+, \breve{x}_i]$ – i.e., divide all its values by $m_i(X_i^+)$ – and then take the average (as described in Section 1.2 of this proof). When we do this transformation, all the areas are divided by the same value $m_i(X_i^+)$. In particular, the area of the normalised left part of the trough t'_{i-} is now equal to $t'_{i-} = t_{i-}/m_i(X_i^+)$; see Fig. B.6.

Similarly, to Part 1 of this proof, we conclude that the width g_{i-} of the left part of the gap is equal to the area t'_{i-} of the normalised left part of the trough: $g_{i-} = t'_{i-} = t_{i-}/m_i(X_i^+)$. Since $m_i(X_i^+) \leqslant 1$, we conclude that $g_{i-} \geqslant t_{i-}$.

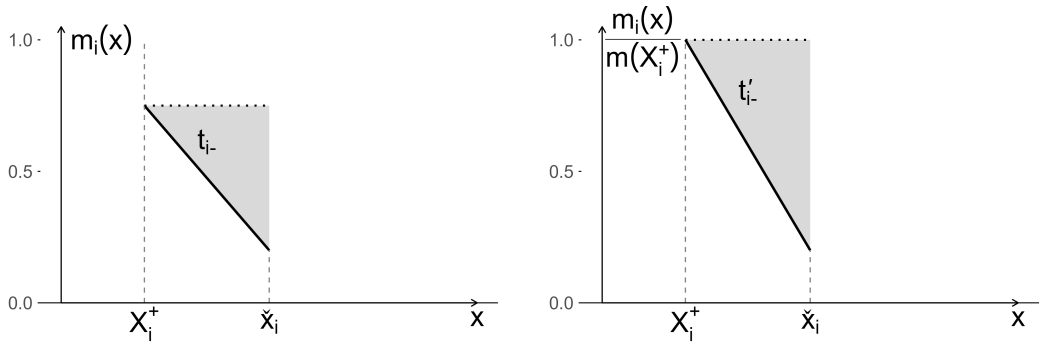


Figure B.6: *(a) Left part of the trough t_{i-} between two consecutive peaks, where the left peak has a maximum value $m_i(X_i^+) < 1$. (b) Normalised version of the same segment, with membership values divided by $m_i(X_i^+)$, giving $t'_{i-} = t_{i-}/m_i(X_i^+)$.*

2.2°. Let us now analyse what happens on the right side of the value $\breve{x}_i$.

To compute x_{i+1}^-, we normalise the segment of the membership function $m_i(x)$ limited to the increasing interval $[\breve{x}_i, X_{i+1}^-]$ – i.e., divide all its values by $m_i(X_{i+1}^-)$ – and then take the average (as described in Section 1.2 of this proof). When we do this transformation, all the areas are divided by the same value $m_i(X_{i+1}^-)$. In particular, the area of the normalised right part of the trough t'_{i+} is now equal to $t'_{i+} = t_{i+}/m_i(X_{i+1}^-)$.

Similarly, to Part 1 of this proof, we conclude that the width g_{i+} of the right part of the gap is equal to the area t'_{i+} of the normalised right part of the trough: $g_{i+} = t'_{i+} = t_{i+}/m_i(X_{i+1}^-)$. Since $m_i(X_{i+1}^-) \leqslant 1$, we conclude that $g_{i+} \geqslant t_{i+}$.

2.3°. Now we can conclude the proof of Proposition 4.

Indeed, since $g_{i-} \geqslant t_{i-}$ and $g_{i+} \geqslant t_{i+}$, we can conclude that

$$g_i = g_{i-} + g_{i+} \geqslant t_{i-} + t_{i+} = t_i.$$

Case of *DUW*.

To distinguish the weighted case from the unweighted case (*DU*), we denote the gap width computed by the *DUW* method as g_i^w and the trough area, computed using the same normalised weights, as t_i^w. We will prove that:

- When the local maximum values on both sides of the trough are equal to 1, then $t_i^w = g_i^w$;
- In general, $t_i^w \leqslant g_i^w$ for all i.

To prove these statements, we use the $WIVD$ method, where the normalised weight function is defined as

$$\hat{w}(\alpha) = \frac{\alpha}{\int_0^1 \alpha' \, d\alpha'} = \frac{\alpha}{\frac{1}{2}} = 2\alpha, \quad \alpha \in [0,1],$$

satisfying

$$\int_0^1 \hat{w}(\alpha)\, d\alpha = \int_0^1 2\alpha \, d\alpha = 1.$$

The weighted variant DUW follows exactly the same geometric procedure as DU, except that all α-cut averages are computed using $\hat{w}(\alpha) = 2\alpha$. Following the same reasoning as in the unweighted case, we compute the weighted trough area and the weighted gap width for the case when both neighboring peaks have maximum value 1, obtaining:

$$t_i^w = g_i^w = \int_{\breve{m}_i}^1 2\alpha \left[x_{i+1}^-(\alpha) - x_i^+(\alpha) \right] d\alpha. \tag{B.11}$$

For the general case, when the peak values are smaller than 1, the same normalisation procedure as in the DU proof applies, leading to proportionally reduced trough areas while the corresponding gap widths remain unchanged. Hence,

$$g_i^w = g_{i-}^w + g_{i+}^w \geqslant t_{i-}^w + t_{i+}^w = t_i^w.$$

Proposition 5. For all i, the gap width g_i^w computed by the DUW method is greater than or equal to the trough area t_i computed without weights: $g_i^w \geqslant t_i$.

We recall from Proposition 4 that the unweighted trough area and the weighted gap width can be expressed as

$$t_i = t_{i-} + t_{i+}, \qquad g_i^w = g_{i-}^w + g_{i+}^w,$$

where

$$t_{i-} = \int_{\breve{m}_i}^1 \left(\breve{x}_i - x_i^+(\alpha) \right) d\alpha, \quad t_{i+} = \int_{\breve{m}_i}^1 \left(x_{i+1}^-(\alpha) - \breve{x}_i \right) d\alpha,$$

and

$$g_{i-}^w = \int_{\breve{m}_i}^1 2\alpha \left(\breve{x}_i - x_i^+(\alpha) \right) d\alpha, \quad g_{i+}^w = \int_{\breve{m}_i}^1 2\alpha \left(x_{i+1}^-(\alpha) - \breve{x}_i \right) d\alpha.$$

To simplify the notation and highlight the geometric meaning, let us denote

$$d_-(\alpha) = \breve{x}_i - x_i^+(\alpha), \qquad d_+(\alpha) = x_{i+1}^-(\alpha) - \breve{x}_i,$$

which represent the horizontal widths of the left and right parts of the trough at each α-level. Both $d_-(\alpha)$ and $d_+(\alpha)$ are non-decreasing functions of α, since as α increases, the α-cut intervals become subsets of those at lower α levels. The weight function $\hat{w}(\alpha) = 2\alpha$ is also non-decreasing on $[0, 1]$.

By Chebyshev's integral inequality, if two functions are both non-decreasing (or both non-increasing) on the same interval $[a, b]$, then

$$\frac{1}{b-a}\int_a^b f(x)\,g(x)\,dx \;\geqslant\; \frac{1}{b-a}\int_a^b f(x)\,dx \;\cdot\; \frac{1}{b-a}\int_a^b g(x)\,dx. \tag{B.12}$$

Applying this inequality to $f(\alpha) = d_-(\alpha)$ and $g(\alpha) = \hat{w}(\alpha) = 2\alpha$ on the interval $[\breve{m}_i, 1]$ yields

$$\int_{\breve{m}_i}^1 2\alpha\, d_-(\alpha)\,d\alpha \;\geqslant\; \frac{\int_{\breve{m}_i}^1 2\alpha\,d\alpha}{1-\breve{m}_i}\int_{\breve{m}_i}^1 d_-(\alpha)\,d\alpha.$$

Since $\int_{\breve{m}_i}^1 2\alpha\,d\alpha = 1 - \breve{m}_i^2$, we obtain

$$g_{i-}^w = \int_{\breve{m}_i}^1 2\alpha\, d_-(\alpha)\,d\alpha \;\geqslant\; (1+\breve{m}_i)\,t_{i-}.$$

An identical argument applied to $d_+(\alpha)$ gives $g_{i+}^w \geqslant (1+\breve{m}_i)\,t_{i+}$. Summing these inequalities and recalling that $0 \leqslant \breve{m}_i < 1$ leads to

$$g_i^w = g_{i-}^w + g_{i+}^w \geqslant t_{i-} + t_{i+} = t_i.$$

Hence, for all i, the weighted gap width is greater than or equal to the unweighted trough area.

In particular, when $\breve{m}_i$ (the trough depth) is fixed, an increase in the trough area t_i implies a corresponding increase in the weighted gap width g_i^w, since both are monotone linear functionals of the width functions $d_-(\alpha)$ and $d_+(\alpha)$.