

Fuzzy Management Methods

Series Editors: Andreas Meier · Witold Pedrycz · Edy Portmann

Rédina Berkachy

The Signed Distance Measure in Fuzzy Statistical Analysis

Theoretical, Empirical and Programming
Advances

 Springer

Fuzzy Management Methods

Series Editors

Andreas Meier, Fribourg, Switzerland

Witold Pedrycz, Edmonton, Canada

Edy Portmann, Bern, Switzerland

With today's information overload, it has become increasingly difficult to analyze the huge amounts of data and to generate appropriate management decisions. Furthermore, the data are often imprecise and will include both quantitative and qualitative elements. For these reasons it is important to extend traditional decision making processes by adding intuitive reasoning, human subjectivity and imprecision. To deal with uncertainty, vagueness, and imprecision, Lotfi A. Zadeh introduced fuzzy sets and fuzzy logic. In this book series "Fuzzy Management Methods" fuzzy logic is applied to extend portfolio analysis, scoring methods, customer relationship management, performance measurement, web reputation, web analytics and controlling, community marketing and other business domains to improve managerial decisions. Thus, fuzzy logic can be seen as a management method where appropriate concepts, software tools and languages build a powerful instrument for analyzing and controlling the business.


More information about this series at <http://www.springer.com/series/11223>

Rédina Berkachy

The Signed Distance Measure in Fuzzy Statistical Analysis

Theoretical, Empirical and Programming
Advances

 Springer

Rédina Berkachy 
Applied Statistics And Modelling (ASAM)
Department of Informatics,
Faculty of Management, Economics
and Social Sciences
University of Fribourg
Fribourg, Switzerland

ISSN 2196-4130

ISSN 2196-4149 (electronic)

Fuzzy Management Methods

ISBN 978-3-030-76915-4

ISBN 978-3-030-76916-1 (eBook)

<https://doi.org/10.1007/978-3-030-76916-1>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

*To my beloved mom Marcelle and dad
Georges, my precious sister Rana, my brother
Roy and my sister-in-law Mathilda, to
Georges and Maria, and to my wonderful
husband Salim,*

Foreword

At the beginning of her thesis work, R dina Berkachy thoroughly examined the empirical distributions of defuzzified fuzzy quantities. In particular, she focused her attention on the signed distance measure as a defuzzification tool. She compared this measure with other methods. Although the signed distance measure was already well known, surprisingly, little was done to clearly understand the impact on the defuzzified distribution. For an applied statistician, it is essential to have in mind the potential repercussions of the tool used on data. The mean, variance, skewness and kurtosis were particularly observed. It was not very unexpected to find that this distance measure tends to normalise the data, in the sense that the crisp data obtained by defuzzification tend to have some characteristics of a normal distribution.

The signed distance measure was applied in global and individual evaluations of linguistic questionnaires. Based on the normality assumption, confirmed by the precedent results, testing hypotheses on individual evaluations could be made rightfully as in classical statistics. It was the start of promising research.

Several distance measures are commonly used in fuzzy statistics. The signed distance is one among them. It is not a new concept and has been widely applied, for example, in linguistic questionnaires analyses. Indeed, R dina Berkachy proposed two new declinations of this measure, proved their properties and used them in different contexts or procedures.

Fuzzy inference methods are central in the work of R dina Berkachy. It is worth mentioning the use of fuzzy logic principles in the formulation of fuzzy statistical hypotheses. Doing so, not only the data could be vague or imprecise but also the hypotheses. The development of fuzzy confidence intervals using the likelihood function is a remarkable achievement. The approach appears to be general because the technique is generalisable to many situations. By means of a bootstrap algorithm, one can simulate the distribution of the fuzzy likelihood ratio and construct a proper confidence interval. Then, this interval can be used, for example, to test hypotheses. Furthermore, R dina Berkachy showed how to properly use the signed distance measure to defuzzify fuzzy decisions and fuzzy p-values and wrote the decision rules.

Another welcomed result is the proposed fuzzy analysis of variance, in which the method is generalised to the multi-ways analysis of variance. It is shown that the decomposition of fuzzy treatment effects is possible, and proper testing can be done after simulation of the distributions by a bootstrap algorithm. Particularly appealing, the application of the method is straightforward.

Rédina Berkachy's book proposes a lot of procedures and methods, which could discourage a reader. Fortunately, all of them are illustrated by numerous examples. An R package, named FuzzySTs, is now also available on CRAN. The reader can install it and run the functions to perform the analyses. On the other hand, the book is well structured, and since a pedagogical point of view was adopted in writing, its content is very accessible to everyone. My dearest hope is a warm welcome from the community, fast dissemination of the results and the acquisition of new fans of fuzzy statistics.

Prof. Dr. Laurent Donzé

Preface

This book intends to present some advances in fuzzy statistical analyses. A particular distance between fuzzy numbers, called the signed distance, seems to be appealing because of its directional property. It has the ability of describing the direction of travel between two fuzzy numbers. In addition, it has been often used as a fuzzy ranking tool or a defuzzification operator. Despite the fact that this distance appears to have interesting properties, it presents serious drawbacks. To overcome these problems, this book develops improved versions of it given by two L_2 metrics for which the first one is directional and preserves the properties of the signed distance, and the other one is non-directional. Both sophisticated distances have the advantage of taking into consideration the shapes of the fuzzy numbers and their possible irregularities. The core aim of this book is therefore to apply these novel distances in a series of statistical approaches defined on the set of fuzzy numbers.

Based on the proposed distances, this book provides a methodology of testing hypotheses in the fuzzy environment. This method of testing relies on the estimation of fuzzy confidence intervals, where we consider not only the data as fuzzy, but also the hypotheses. As such, the defended distances are used at different stages of the process of the inference test: in the conception of the models themselves or as defuzzification operators of the obtained fuzzy decisions. Moreover, since the traditional way of estimating fuzzy confidence intervals is in some sense limited in terms of the chosen parameters and the involved distributions, this book presents a practical procedure of estimation of such intervals based on the likelihood method. This new procedure is seen as general, since it can be used with any type of parameter and without the obligation of defining a particular distribution a priori. Analogously to this testing model, a definition of the fuzzy p-value described in the same setups with its corresponding decision rule is given. In this context, a defuzzification of this fuzzy p-value can be of good use to get a crisp interpretable decision.

Furthermore, this book presents two conceptually different applications in which the defended distances are involved. On a first stage, a novel methodology of assessment of linguistic questionnaires is developed on two distinct levels: the

global and the individual ones. The proposed procedure has the great possibility of being able to treat the sampling weights and the eventual missingness occurring in the concerned data sets. The second application consists an extension of the multi-ways analysis of variance to the space of fuzzy sets. The decision related to the corresponding test statistic can be made according to two decision rules: a heuristic one preserving the fuzzy nature of the sums of squares, and another crisp one based on the defended distances.

To illustrate these approaches, multiple empirical and simulation studies are displayed using synthetic and real data sets. Note that a prominent objective of these studies is to investigate empirically some different statistical contexts, and explore theoretically the use of the newly introduced distances compared to known ones from one side and to the results obtained from the conventional statistical theory from another side. A recurrent general finding is that the classical approach seems to be a particular case of the fuzzy one. As such, the mentioned fuzzy approaches seem to be potentially promising.

A coherent R package covering all the previously mentioned concepts with complete documentation and some use cases are finally described. This package is developed from scratch in compliance with all the theoretical tools presented in a user-friendly comprehensive programming environment with a big panoply of calculation methods.

Fribourg, Switzerland

Rédina Berkachy

Acknowledgements

This book is a revised version of my Ph.D. thesis presented to the Faculty of Management, Economics and Social Sciences at the University of Fribourg. Enrolling in these Ph.D. studies has truly been a life-changing experience, and this would not have been possible without the support and guidance I received from many people.

Originating from a very beautiful village called Deir El Ahmar in the valley of Lebanon, I decided in 2013 to go beyond my frontiers and be adventurous as always. When I first embarked in Switzerland to start my Ph.D., I did not imagine my life will change drastically, leaving not only my books and my souvenirs in my beloved country, but also my heart and my thoughts.

First of all, let me express my gratitude to my mentor Prof. Dr. Laurent Donzé who gave me the opportunity to be part of the Applied Statistics and Modelling Group at the University of Fribourg. I would like to thank him for his constant support and patience, for all his motivation and encouragement. Thank you for cheering me up whenever I felt down, for all our friendly discussions around coffee breaks, and for believing in me.

I am also grateful to Prof. Dr. Marino Widmer who stood by me in every difficult step I faced. Thank you from the bottom of my heart for always motivating me, and for always telling me not to surrender. You gave me strength. Thank you.

I would like also to acknowledge the reviewers of this manuscript for accepting to do this task. Thank you for providing me with constructive and interesting enhancements of my book. In addition, I am particularly thankful to the dean's office in the Faculty of Management, Economics and Social Sciences at the University of Fribourg who encouraged me throughout the book process.

The publication of this manuscript is supported by the FMSquare association represented by Prof. Dr. Andreas Meier and Prof. Dr. Edy Portmann. I am very thankful for their support during the whole process.

My sincere thanks also go to Prof. Mathias Rossi and Prof. Pascale Voirin who encouraged me in the beginning of this thesis and helped to make it happen. I must also thank Dr. Jean-Vincent Le Bé from Nestlé beverages who trusted in me and collaborated with me on a very interesting project.

I would like to express my very profound gratitude to my friends Nayla Sokhn, Zeina Youssef-Aebli, Rana Bassil, Marielle Hayeck, Elie-Jacques Fares, Karl Daher, Rita Abboud, Clauda Mhanna and Joanne Mrad, and to my colleagues and friends, Christine Schaeren, Martina Raggi and Layal Lettry who all kept their sense of humour when I lost mine. Thank you for all the encouragement and support during this difficult period.

I would like finally to say a heartfelt thank you to my beloved family: my mom Marcelle and my dad Georges for their constant prayers and thoughts, for always believing in me and cheering me up. I am also indebted to my twin-sister Rana who did not miss any chance to come to Switzerland (and unfortunately travelled a lot because of me) whenever I felt down. I also thank my brother Roy and my sister-in-law Mathilda for always standing by me with their support in each step and their constant humour. Thank you to little Georges and Maria for giving me the inspiration and the patience. Seeing them growing up so fast is a blessing.

Last but not least, I would like to thank my wonderful husband Salim for his endless loyalty, for his unfailing love, support, and care. Thank you for cheering me up and standing by me especially when I completely lost hope. Because of you, I have gained self-confidence and found my life's passion. Thank you also for your invaluable advices throughout the writing process, they were crucial.

This accomplishment would not have been achieved without the encouragement and motivation extended by all these people. Thank you all from the bottom of my heart.

Contents

1	Introduction	1
	References	6
 Part I Theoretical Part		
2	Fundamental Concepts on Fuzzy Sets	13
	2.1 Fuzzy Sets and Fuzzy Numbers	13
	2.2 Common Fuzzy Numbers	18
	2.3 Vector of Fuzzy Numbers and Fuzzy Vector	20
	2.4 Logical Operations on Fuzzy Sets	22
	2.5 Extension Principle	24
	2.6 Arithmetic Operations on Fuzzy Sets	26
	References	32
3	Fuzzy Rule-Based Systems	35
	3.1 Fuzzification	37
	3.2 Inference Rules	38
	3.3 Defuzzification	42
	References	45
4	Distances Between Fuzzy Sets	47
	4.1 The ρ_1 and ρ_2 Distances	49
	4.2 The Bertoluzza Distance	50
	4.3 The $\delta_{p,q}^*$ and ρ_p^* Distances	51
	4.4 The “mid/spr” Distance	52
	4.5 The “ ϕ -wab/ldev/rdev” Distance	53
	4.6 The Signed Distance	54
	4.7 The Generalized Signed Distance	58
	Appendix	71
	A Proofs of Chap. 4	71
	References	84

5	Fuzzy Random Variables and Fuzzy Distributions	87
5.1	Fuzzy Events and Their Probabilities	88
5.2	Fuzzy Random Variables	90
5.2.1	The Kwakernaak–Kruse and Meyer Approach	90
5.2.2	The Féron–Puri and Ralescu Approach	92
5.2.3	Synthesis	93
5.3	Fuzzy Probability Distributions of Fuzzy Random Variables	93
5.3.1	The Kwakernaak–Kruse and Meyer Approach	94
5.3.2	The Féron–Puri and Ralescu Approach	95
5.3.3	Synthesis	96
5.4	Expectation and Variance	97
5.4.1	The Kwakernaak–Kruse and Meyer Approach	98
5.4.2	The Féron–Puri and Ralescu Approach	99
5.4.3	Synthesis	101
5.5	Estimators of the Fuzzy Distributions Parameters	101
5.5.1	The Mean	102
5.5.2	The Sample Moments	104
5.6	Simulation	108
	References	112
6	Fuzzy Statistical Inference	115
6.1	Testing Hypotheses by the Classical Approach	117
6.2	Fuzzy Hypotheses	120
6.3	Fuzzy Confidence Intervals	122
6.3.1	Fuzzy Confidence Intervals for Pre-defined Parameter	123
6.3.2	Fuzzy Confidence Intervals with the Likelihood Function	127
6.3.3	Bootstrap Technique for the Approximation of the Likelihood Ratio	135
6.3.4	Simulation Study	137
6.3.5	Numerical Example	140
6.4	Fuzzy Hypotheses Testing	143
6.4.1	Fuzzy Hypotheses Testing Using Fuzzy Confidence Intervals	143
6.4.2	Defuzzification of the Fuzzy Decisions	150
6.4.3	Fuzzy p-Values	155
6.4.4	Defuzzification of the Fuzzy p-Value	160
6.5	Applications: Finanzplatz	162
6.5.1	Application 1	163
6.5.2	Application 2	166
6.5.3	Classical Versus Fuzzy Hypotheses Testing	171
	References	174

Conclusion Part I

Part II Applications

7 Evaluation of Linguistic Questionnaire 181

 7.1 Questionnaire in Fuzzy Terms 183

 7.2 Readjustement of Weights 187

 7.3 Global and Individual Evaluations 188

 7.4 Indicators of Information Rate of a Data Base 190

 7.5 Application: Finanzplatz 192

 7.6 Analyses by Simulations on the Individual Evaluations 197

 7.7 Conclusion 202

 Appendix 203

 B Description of the Finanzplatz Data Base 203

 C Results of the Applications of Chap. 7 205

 References 210

8 Fuzzy Analysis of Variance 213

 8.1 Classical Multi-Ways Analysis of Variance 216

 8.2 Introduction to Fuzzy Regression 220

 8.3 Fuzzy Multi-Ways Analysis of Variance 223

 8.3.1 Particular Cases 236

 8.3.2 Ordering Effect and Sequential Sums 239

 8.4 Applications 240

 8.4.1 Application 1: Nestlé Data Set 241

 8.4.2 Application 2: Finanzplatz Data Set 246

 8.5 Conclusion 254

 Appendix 255

 D Description of the Nestlé data base 255

 References 256

Part III An R Package for Fuzzy Statistical Analysis: A Detailed Description

9 fuzzySTs: Fuzzy Statistical Tools: A Detailed Description 261

 9.1 Description of the Package 264

 9.1.1 Dependencies 264

 9.1.2 List of Functions 265

 9.1.3 Description of the Functions 266

 Appendix 328

 E Intersection Situations Between the Fuzzy Confidence Intervals and the Fuzzy Hypothesis 328

 F Possible Positions of the Areas A_l and A_r Under the of \tilde{t} on the Left and Right Sides of the Median \tilde{M} 338

 References 340

Conclusion 341

Acronyms 345

List of Figures

Fig. 2.1	Membership function of \tilde{X} describing the term very young—Example 2.1.1	15
Fig. 2.2	Fuzzy numbers of different symmetrical shapes—Example 2.2.1	21
Fig. 2.3	Example with logical operations—Example 2.4.1	23
Fig. 2.4	Example of the extension principle—Example 2.5.1	26
Fig. 2.5	Sum of the fuzzy numbers \tilde{X} and \tilde{Y} —Example 2.6.1	29
Fig. 2.6	Product of the fuzzy numbers \tilde{X} and \tilde{Y} —Example 2.6.1	30
Fig. 2.7	Quotient of the fuzzy numbers \tilde{X} and \tilde{Y} —Example 2.6.1	30
Fig. 3.1	The fuzzy process	36
Fig. 3.2	Membership functions of the variable waiting time at a traffic light	38
Fig. 3.3	Membership functions of the variable estimated number of surrounding vehicles at the traffic light stop	40
Fig. 3.4	Membership functions of the output variable expected delay for a meeting caused by traffic conditions	41
Fig. 3.5	Membership functions of the resulting combined fuzzy set	41
Fig. 3.6	Decomposition of the membership surface in the process of defuzzification by the centroid	44
Fig. 4.1	Chronological review of distances of fuzzy quantities	48
Fig. 4.2	Membership functions of \tilde{Y} and its nearest symmetrical trapezoidal fuzzy number \tilde{S} —Example 4.7.2	69
Fig. 4.3	Membership functions of \tilde{X} , \tilde{Y} and the nearest mid-way symmetrical trapezoidal fuzzy number \tilde{S}^* between \tilde{X} and \tilde{Y} —Example 4.7.2	70
Fig. 5.1	The skewness— Case 1 and Case 2 with a 10000 sized-sample—Sect. 5.6	112
Fig. 6.1	The fuzzy hypotheses—Sect. 6.2	120
Fig. 6.2	Fuzzy hypothesis \tilde{H} and its boundary \tilde{H}^* —Example 6.2.1	121

Fig. 6.3	The fuzzy null and alternative hypotheses—Example 6.2.2	122
Fig. 6.4	Fuzzy confidence interval for the mean $\tilde{\Pi}$ —Example 6.3.1	126
Fig. 6.5	Fuzzy log-likelihood function for the mean and the intersection with the upper and lower bounds of the fuzzy parameter \tilde{X} —Example 6.3.2	141
Fig. 6.6	Fuzzy confidence interval by likelihood ratio method $\tilde{\Pi}_{LR}$ —Example 6.3.2	142
Fig. 6.7	The membership functions of the fuzzy null hypothesis and the fuzzy confidence intervals—Example 6.4.1	149
Fig. 6.8	The membership functions of the fuzzy decisions—Example 6.4.2	154
Fig. 6.9	Fuzzy p -value \tilde{p} for the mean—Example 6.4.3	160
Fig. 6.10	Membership functions of the fuzzy null hypothesis and the fuzzy confidence intervals—Sect. 6.5.1	164
Fig. 6.11	Membership functions of the fuzzy null hypothesis and the fuzzy confidence intervals by LR—Sect. 6.5.1	165
Fig. 6.12	Fuzzy p -value of the test—Sect. 6.5.2	168
Fig. 6.13	Fuzzy p -values of the different tests given—Table 6.6	170
Fig. 7.1	The histogram of the distribution of the weighted individual evaluations of the Finanzplatz linguistic questionnaire—Sect. 7.5	195
Fig. 7.2	The QQ-plot of the distribution of the weighted individual evaluations of the Finanzplatz linguistic questionnaire—Sect. 7.5	195
Fig. 7.3	Boxplot of the weighted individual evaluations classified by company size—Sect. 7.5	197
Fig. 7.4	Boxplot of the weighted individual evaluations classified by company branch—Sect. 7.5	197
Fig. 7.5	QQ-plot of the distributions of the weighted individual evaluations with and without NA—Sect. 7.5	198
Fig. 7.6	Order of the skewness of the distributions of the individual evaluations with respect to the different distances (the red ones are the L_3 group, the blue ones are the L_2 group, and the gray one is the L_1 group)	200
Fig. 8.1	QQ-plot of the bootstrap distribution of the model test statistic and the theoretical Fisher distribution—Sect. 8.4.2	250
Fig. 8.2	QQ-plots of the bootstrap distributions of the test statistics related to the factors size and branch, respectively, with the theoretical Fisher distribution—Sect. 8.4.2	251
Fig. 8.3	Fuzzy means of the treatment and the residuals sums of squares related to the variable “branch” of the Mult-FANOVA model—Sect. 8.4.2	252

Fig. 8.4	QQ-plots of the bootstrap distributions of the test statistics related to the factors size and branch, respectively, with the theoretical Fisher distribution—Sect. 8.4.2	253
Fig. 9.1	Positions (a) 1 and (b) 2 where $A_l > A_r$	338
Fig. 9.2	Positions (a) 3, (b) 4, and (c) 5 where $A_l \leq A_r$	339