

Fuzzy Statistical Analysis: methods and applications

Pierpaolo D’Urso · María Ángeles Gil

Published online: 29 October 2013
© Sapienza Università di Roma 2013

“Statistical Reasoning is affected by various sources of Uncertainty: randomness, imprecision, vagueness, partial ignorance, etc. Traditional statistical paradigms (such as Statistical Inference, Exploratory Data Analysis, Statistical Learning) are not capable to account for the complex action of Uncertainty in real life applications of Statistical Reasoning.”
(Coppi [1])

This special issue is dedicated to Professor Renato Coppi

The topic of Fuzzy Statistical Analysis, that has occasionally received attention along the 20th century, is nowadays rather appealing. Since the number of direct or potential applications of this analysis is growing steadily, the need for new approaches and methods of analysis is becoming more visible and pressing.

The methodological studies on this topic have been centered on

- the probabilistic tools stating the mathematical foundation for the statistical analysis;
- the statistical data analysis where either data or techniques involve fuzzy modeling.

These methodological developments have an undeniable application-oriented aim and motivation.

Randomness corresponds to the uncertainty associated with a state, an outcome, an event, a datum, a magnitude, an assignment, etc. which can happen with different well-defined values

P. D’Urso (✉)
Sapienza – Università di Roma, Rome, Italy
e-mail: pierpaolo.durso@uniroma1.it

M. A. Gil
Universidad de Oviedo, Oviedo, Spain
e-mail: magil@uniovi.es

or modalities, so that it cannot be exactly predicted in advance. As highlighted by Coppi [1] it is "... the main source of uncertainty possibly investigated in Statistical Reasoning... (quite often limited to the 'data generation process' managed by means of appropriate probabilistic models)."

Fuzziness formalizes the uncertainty "... stemming from the use of vague definitions or vague assessments..." [1], so that it often arises in connection with 'ill-defined' states, outcomes, events, data, magnitudes, assignments, and so on.

"Randomness and fuzziness may act separately or jointly on the various informational ingredients of a statistical reasoning process...". This assertion by Colubi et al. [2] is illustrated in this special issue by collecting seven studies involving uncertainty due to both sources.

The first two papers consider that fuzziness affect the probability assessment. In the first one data are supposed to be identified with exactly known vectorial values, and in the second one data are described by means of fuzzy numbers.

Paper "*Strong Law of Large Numbers with respect to a fuzzy probability measure*", by Dan A. Ralescu, presents the extension of a previous Strong Law of Large Numbers established with respect to a set-valued probability measure to situations where the measures are fuzzy set-valued. An immediate particularization would be that in which data are precisely known, the probabilities are only vaguely known and sample sizes are large (for instance, in Bayesian Inference with a fuzzy-valued prior).

Paper "*Fuzzy Bayesian Inference*", by Reinhard Viertl, displays another approach to statistically deal with situations in a Bayesian framework where both the prior probabilistic and the sample information are assumed to be fuzzy-valued. Bayes' Theorem is generalized to capture this situation, and some Bayesian tools, as the Highest Posterior Density regions for confidence estimation and the predictive distributions are generalized accordingly.

The two following papers deal with a statistical problem and a probabilistic problem with statistical implications where fuzziness is basically incorporated in the procedures.

Paper "*Noise fuzzy clustering of time series by autoregressive metric*", by Pierpaolo D'Urso, Livia De Giovanni, Riccardo Massari and Dario Di Lallo, proposes a robust fuzzy C -medoids clustering model for time series. The model is based on the autoregressive metric by considering the autoregressive representation of the time series. The adopted clustering approach is based on the fuzzy theory.

Paper "*Fuzzy representations of real-valued random variables: Applications to exploratory and inferential studies*", by Ángela Blanco-Fernández, Ana Belén Ramos-Guajardo and Ana Colubi, presents two fuzzy representations of real-valued random variables. When these representations are composed with the variables, the 'mean values' of the resulting fuzzy-valued random elements become either a valuable exploratory tool to visually identify some of the main summary measures of their distributions or a characterizing functional for it. These special features will be valuable for inferential purposes about distributions of random variables.

The last three included papers cover different statistical problems involving fuzzy data.

Paper "*Fuzzy post-retirement financial concepts: an exploratory study*", by Arnold F. Shapiro, prompts the use of fuzzy tools to support final decisions regarding post-retirement, a hot issue today. An approach is suggested to include fuzzy ideas in the financial world, since several aspects related to the post-retirement financial strategies are essentially fuzzy.

Paper "*Weighted Least Squares and Least Median Squares estimation for the fuzzy linear regression analysis*", by Pierpaolo D'Urso and Riccardo Massari, by following an iterative Weighted Least Squares estimation approach, introduces a linear regression model for studying the dependence of a general class of fuzzy response variable on a set of crisp/fuzzy explanatory variables. Some theoretical properties and a generalization of the determination

coefficient are provided to assess the regression goodness of fit. A robust version of the fuzzy regression model based on the Least Median Squares is also given.

Paper “*Bertoluzza et al.’s metric as a basis for analyzing fuzzy data*”, by María Rosa Casals, Norberto Corral, María Ángeles Gil, María Teresa López, María Asunción Lubiano, Manuel Montenegro, Gloria Naval and Antonia Salas, reviews a metric between fuzzy data which have been widely employed for statistical developments from a descriptive and, mainly, an inferential viewpoint. Definitional and topological equivalences of the metric and some statistical methods based on it, are briefly examined, and its impact in the literature is analyzed.

We wish to thank the Editor-in-Chief of METRON, Professor Giovanni M. Giorgi, for his encouragement and support in preparing this special issue. Our gratitude goes also to the contributors of the issue and to all those taking care of the review process.

Last, but not least, we are honoured and pleased to dedicate this special issue as a tribute to Professor Renato Coppi on the occasion of his post-retirement (some of the papers in the issue will certainly be particularly engaging for him). Professor Coppi has recognized many years ago the potential value of combining Fuzzy Logic with Statistical Analysis. Consequently, and because of him being an enthusiastic researcher, he has supplied valuable approaches to statistical problems implicating fuzzy elements; and because of him being a brilliant thinker, he has supplied many valuable reflections. As one of the pioneers on the topic of Fuzzy Statistical Analysis, a nice friend, and teacher/supervisor of several researchers in the field, his influence is now deeply acknowledged.

References

1. Coppi, R.: Management of uncertainty in statistical reasoning: the case of regression analysis. *Int. J. Approx. Reason.* **47**(3), 284–305 (2008)
2. Colubi, A., Coppi, R., D’Urso, P., Gil, M.A.: Statistics with fuzzy random variables. *Metron* **LXV** **3**, 277–303 (2007)