II Joint Conference of the Belgian, Royal Spanish and Luxembourg Mathematical Societies
Logroño, June 6–8, 2016

An applied approach to robust statistical analysis of the location of interval-valued data

Beatriz Sinova[1,3] and **Stefan Van Aelst**[2,3]

The need for statistical analysis of interval-valued data arises with the increasing number of real-life experiments whose outputs are imprecise and require intervals for modelling such imprecision. This imprecision can result from different situations: data can be essentially interval-valued, like interval-censored data, but they can also correspond to aggregate information because the magnitude of interest is the fluctuation of a real-valued attribute over a given time period or collection of individuals or due to confidentiality reasons, or they can even originate as a consequence of uncertain or incomplete information. In order to illustrate the importance of interval-valued data in knowledge fields as different as Finance, Chemistry or Social Sciences, some real-life examples will serve as motivation.

The generalization of statistical techniques and procedures to cover the interval-valued setting is usually not straightforward because of the peculiarities of the considered space. The space of non-empty compact intervals, $\mathcal{K}_c(\mathbb{R})$, is not linear with the usual interval arithmetic, but forms a closed convex cone. One of the consequences is that there is no 'difference operation' that is always well-defined and preserves the main properties of the difference between real values in connection with the sum. In practice, such a drawback is overcome to some extent by replacing differences in the statistical developments by suitable distances between values in $\mathcal{K}_c(\mathbb{R})$.

Most of the statistical techniques already adapted for interval-valued data are based on the Aumann mean as location measure. Despite its numerous handy properties, the Aumann mean is an extension of the concept of mean of a random variable, from which it inherits the high sensitivity to outliers and data changes, as will be illustrated through some real-life examples. Thus, a more robust alternative measure should be used instead of the Aumann mean in situations involving data contamination, so frequently encountered in applications. The aim of this work will be to provide robust tools to summarize the location of interval-valued data and to introduce them from an applied point of view, highlighting the advantages of their use and showing their suitable performance by means of the motivating real-life studies. Finally, some concluding remarks about future research lines related to this work will be given.

[1]Department of Statistics and Operational Research and D. M.,
University of Oviedo, 33007 Oviedo, Spain
sinovabeatriz@uniovi.es

[2]Department of Mathematics, KU Leuven, 3001 Leuven, Belgium
stefan.vanaelst@kuleuven.be

[3]Department of Applied Mathematics, Computer Science and Statistics,
Ghent University, 9000 Gent, Belgium