

# Statistical reasoning with interval-valued data: problems, results and challenges

Przemyslaw Grzegorzewski

Abstract:

A development in theory and practice of statistical reasoning, machine learning and broadly perceived data analysis is highly associated with surpassing restrictions on available data. Nowadays, unlike in classical statistics, our interests and goals are limited not only to observations expressed by real numbers or vectors but they extend to more general structures like intervals, fuzzy sets and their extensions, functional or symbolic data, etc.

In particular, interval-valued data have drawn an increasing interest in recent years. Intervals appear very useful in modeling uncertainty connected with imprecise or incomplete information. Sometimes even exact values of a variable are hidden deliberately for some confidentiality reasons. But there are also situations when the experimental results appear just as essentially interval-valued data like ranges of fluctuations of some physical measurements, time interval spanned by some activity, etc.

Although intervals perceived as an extension of real data come up so naturally in various applications the way to construct adequate inferential methods cannot reduce to straightforward adaptation of classical statistical tools fitted with the interval arithmetic. Actually, a starting point to construct proper tools for handling interval-valued data is to realize which of the two possible perspectives: ontic or epistemic, should be accepted. Indeed, closed intervals applied for modeling data may deliver two different types of information: the imprecise description of a point-valued quantity (epistemic view) or the precise description of a set-valued entity (ontic view). The adoption of intervals for modeling experimental results does not impose any default interpretation itself. However, this distinction has a strong impact on the concepts and techniques related to statistical reasoning since each view yields its own approach to data analysis and a way of carrying on the statistical inference. Moreover, a generalization of classical statistical procedures into the interval-valued framework may cause considerable computational problems, especially if a sample is large enough.

The aforementioned problems connected with interval-valued data management and processing as well as different methodology leading to analytical solutions which result from the accepted interpretation of the interval-valued data will be discussed during the presentation. Theoretical considerations will be illustrated by examples of statistical tools and procedures that can be used in practice.