Session abstract

Session: Compositional data analysis

September 20, 10:30-12:00

Organizer: Michael Greenacre, Universitat Pompeu Fabra, Barcelona, Spain

Compositional data can be defined as vectors of positive components, where the relative importance of these components is what ultimately matters to the research questions. This relative importance is distilled in ratios, or, more precisely, the logarithms of these ratios, i.e. logratios, for their appropriate statistical treatment. The compositional data tradition started treating data with a fixed sum (compositions expressed as proportions, percentages, ppm, etc.) but has recently spanned all fields in which relative importance is of interest, be the data constrained to a fixed sum or not. Applications include chemical and geological compositions, time budgets, genomics, microbiomics, geology, financial ratios, pollution studies, dietary studies, text mining and content analysis, and so on. Standard statistical learning methods have to be adapted to this new scenario, as well as the interpretation of their results. The session presents three advancements in these domains.

The first contribution in this sessıon shows that for high-dimensional data found typically in genomic and microbiome studies the simplest logratio transformation, the additive logratio, can almost perfectly reproduce the exact logratio geometry and thus provide a practical solution that is easy to interpret. The second contribution involves mining the prediction of a criterion variable using logratios between pairs of components as predictors, where the issue of interpretation is crucial. The third contribution presents a new approach to compositional regression where the sampling units are not treated in a single simplex space but in a simplicial complex after combining the units into groups.

- Michael Greenacre - Universitat Pompeu Fabra, Barcelona, Spain - *Compositional data analysis of high-dimensional biological datasets: a revalidation of the additive logratio transformation*

- Germà Coenders - Universitat de Girona, Girona, Spain - *Three approaches to supervised learning for compositional data with pairwise logratios*

- Andrej Srakar – Institute for Economic Research (IER) and University of Ljubljana, Ljubljana, Slovenia - *Combinatorial regression in abstract simplicial complexes*