

Publicacions més rellevants de la línia de recerca: Anàlisi estadística de les dades composicionals

Referència: Egozcue, J.J. and Pawlowsky-Glahn, V. Groups of parts and their balances in compositional data analysis. *Mathematical Geology*, **37(7)** (2005), pp. 795–828.

Abstract: Amalgamation of parts of a composition has been extensively used as a technique of analysis to achieve reduced dimension, as was discussed during the CoDaWork'03 meeting (Girona, Spain, 2003). It was shown to be a non-linear operation in the simplex that does not preserve distances under perturbation. The discussion motivated the introduction in the present paper of concepts such as a group of parts, balance between groups, and sequential binary partition, which are intended to provide tools of compositional data analysis for dimension reduction. Key concepts underlying this development are the established tools of subcomposition, coordinates in an orthogonal basis of the simplex, balancing element and, in general, the Aitchison geometry in the simplex. Main new results are: a method to analyze grouped parts of a compositional vector through the adequate coordinates in an *ad hoc* orthonormal basis; and the study of balances of groups of parts (inter-group analysis) as an orthogonal projection similar to that used in standard subcompositional analysis (intra-group analysis). A simulated example compares results when testing equal centers of two populations using amalgamated parts and balances; it shows that, in certain circumstances, results from both analysis can disagree..

Referència: Palarea-Albaladejo, J. and Martín-Fernández, J.A. A modified EM algorithm for Replacing Rounded Zeros in Compositional Data Sets. *Computers & Geosciences*, **34(8)** (2008), pp. 902–917.

Abstract: The presence of rounded zeros results in an important drawback for the statistical analysis of compositional data. Data analysis methodology based on log-ratios cannot be applied under these conditions. In this paper rounded zeros are considered as a special kind of missing data. Thus, an EM-type computational algorithm for replacing them is provided. The procedure is based on the additive logistic transformation and assumes an additive logistic normal model for the data. First, the alr transformation moves data from the constrained simplex space to the unconstrained real space. Next, missing transformed data are imputed by using modified EM steps. Last, imputed data are transformed back into the simplex space to obtain a compositional data set free of rounded zeros. Additionally, a sequential strategy is proposed for the case of rounded zeros in all the components of a composition. This work focuses on the properties of the algorithm

and on computational implementation details. Also, its effectiveness on simulated data sets with a range of detection limits is analyzed. Special attention is paid on the effects on the covariance structure of a compositional data set. Results confirm the good behavior of our proposal. Finally, MATLAB routines implementing the algorithm are made available to the reader.

Referència: Mateu-Figueras, G. and Pawlowsky-Glahn, V. A critical approach to probability laws in geochemistry. *Mathematical Geosciences*, **40(5)** (2008), pp. 489–502.

Abstract: Probability laws in geochemistry have been a major issue of concern over the last decades. The lognormal on the positive real line or the additive logistic normal on the simplex are two classical laws of probability to model geochemical data sets due to their association with a relative measure of difference. This fact is not fully exploited in the classical approach when viewing both the positive real line and the simplex as subsets of real space with the induced geometry. But it can be taken into account considering them as real linear vector spaces with their own structure. This approach implies using a particular geometry and a measure different from the usual ones. Therefore, we can work with the coordinates with respect to an orthonormal basis. It could be shown that the two mentioned laws are associated with a normal distribution on the coordinates. In this contribution both approaches are compared, and a real data set is used to illustrate similarities and differences.