

Joint Biplots for CoDa

M. GALLO¹, V. SIMONACCI, M.A. DI PALMA¹

¹Department of Human and Social Sciences – University of Naples - L’Orientale, Italy, mgallo@unior.it

Compositional data (CoDa) consist of vectors of positive values summing to a unit, or in general, to some fixed constant for all vectors. They appear as proportions, percentages, concentrations, absolute and relative frequencies. Sometimes, compositions arise from non-negative data (such as counts, area, weights, volume) that have been scaled by the total of the components because the analyst is not interested in the total sum of the vector.

The multidimensional analysis of this kind of data requires a careful consideration because the sample space for CoDa is the simplex. The first consistent methodological proposal to deal with CoDa was proposed by Aitchison (1986) when he introduced the log-ratio approach. Basically, the idea that this approach conveys is to move from the simplex space to the real space by using log-ratio transformations, applying standard statistical methods, and finally, by means of an inverse log-ratio transformation, to interpret the results in the simplex space. Starting from this paper, pairwise, centered, additive and isometric log-ratio transformations, in short *plr*, *clr*, *alr* (Aitchison, 1986) and *ilr* respectively, are proposed in literature (Egozcue et al., 2003).

In the context of dimension-reducing techniques, Aitchison (1983) proposed applying principal component analysis (PCA) after having applied a centered log-ratio (*clr*) transformation to CoDa. Aitchison and Greenacre (2002) suggested an adaptation of the biplot to CoDa. The biplot is a well established graphical aid in other branches of statistical analysis and can prove to be a useful exploratory and expository tool for compositions.

In literature many papers on dimensional-reduction techniques for CoDa are proposed. Based on log-ratio strategy, Gallo (2012a, 2012b, 2013) recently proposed to use three-mode analysis of compositional data.

Starting from Gallo (2012b), we propose using of *plr* and *clr* joint biplots. Where in some cases the *plr* joint biplot is the only ones that show clearly the correlations.

References

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.

Aitchison, J. (1983). Principal component analysis of compositional data. *Biometrika* 70(1): 57–65.

Aitchison, J. Greenacre, M. (2002). Biplots for compositional data. *Applied Statistics* 51(4): 375–392.

Egozcue, J.J. Pawlowsky-Glahn, V. Mateu-Figueras, G. Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology* 35(3): 279–300.

Gallo, M (2012a). CoDa in Three-Way Arrays and Relative Sample Spaces. *Electronic Journal of Applied Statistical Analysis*, vol. 5, p. 401-406, ISSN: 2070-5948, doi: 10.1285/i20705948v5n3p400

Gallo, M (2012b). Tucker3 model for compositional data. *Submitted*.

Gallo, M (2013). Log-Ratio and Parallel Factor Analysis: An Approach to Analyze Three-Way Compositional Data . In: Araceli N. Proto, Massimo Squillante and Janusz Kacprzyk. *Advanced Dynamic Modeling of Economic and Social Systems*. Studies in Computational Intelligence, vol. 448, p. 209-221, Springer, ISSN: 1860-949X, doi: 10.1007/978-3-642-32903-6.