Interpretation of orthonormal coordinates in case of three-part compositions applied to orthogonal regression for compositional data

S. DONEVSKA[1], E. FIŠEROVÁ[1], and K. HRON[1]

[1] Department of Mathematical Analysis and Applications of Mathematics - Palacký University Olomouc, Czech Republic sandra.donevska01@upol.cz

Orthonormal coordinates are very important tool for compositional data processing using standard statistical methods. Namely, in order to express a D-part composition in the Euclidean real space we use isometric log-ratio (ilr) transformation, which is an isometric mapping from the sample space of compositions, the simplex $\mathcal{S}^D$ with the Aitchison geometry, to the $(D-1)$-dimensional Euclidean real space $\mathbf{R}^{D-1}$. The ilr transformation results in coordinates of an orthonormal basis on the simplex. Advantages coming from this transformation, like the mentioned isometry between $\mathcal{S}^D$ and $\mathbf{R}^{D-1}$, are closely related with the problem of interpreting orthonormal coordinates, constructed by sequential binary partition. Their interpretation can be approached as balances between groups of parts of a composition as well as by expressing their covariance structure by log-ratios of parts of the analyzed composition, i.e. in terms of ratios. Note that if we want to achieve interpretation of results of statistical analysis directly on the simplex (in terms of the original compositional parts), the back-transformation is required.

The aim of the contribution is to analyze the interpretation of two coordinates (balances) obtained by the ilr transformation of three-part compositions. Attention is focused on interpreting coordinates coming from the description of their covariance structure. General conclusions will be used for analysing results from orthogonal regression for compositions. Its main idea is to fit a line explaining the set of n compositional data points in coordinates in such a way that the sum of squared distances from data points to the estimated line is minimal. By using the theory of linear regression models with type II constraints, it is possible to construct confidence bounds or testing hypotheses on regression parameters. However, especially the mentioned parameters cannot be easily interpreted back on the simplex, the interpretation is only possible in sense of the orthonormal coordinates. The theoretical considerations will be illustrated on a real-world example.