Compositional Data Analysis with R

Matevž Bren

University of Maribor, Faculty of Organizational Sciences,
Kidričeva 55a, 4000 Kranj, Slovenia
Institute of Mathematics, Physics and Mechanics,
Jadranska 19, 1000 Ljubljana, Slovenia
e-mail address: `matevz.bren@fov.uni-mb.si`

Vladimir Batagelj

University of Ljubljana, FMF, dept. of Mathematics
Institute of Mathematics, Physics and Mechanics

R from `http://www.r-project.org/` is 'GNU S' – a language and environment for statistical computing and graphics. The environment in which many classical and modern statistical techniques have been implemented, but many are supplied as packages. There are 8 *standard* packages and many more are available through the `cran` family of Internet sites `http://cran.r-project.org` .

We started to develop a library of functions in R to support the analysis of mixtures and our goal is a `MixeR` package for compositional data analysis that provides support for

operations on compositions: perturbation and power multiplication, subcomposition with or without residuals, centering of the data, computing Aitchison's, Euclidean, Bhattacharyya distances, compositional Kullback-Leibler divergence etc.

graphical presentation of compositions in ternary diagrams and tetrahedrons with additional features: barycenter, geometric mean of the data set, the percentiles lines, marking and coloring of subsets of the data set, theirs geometric means, notation of individual data in the set ...

dealing with zeros and missing values in compositional data sets with R procedures for simple and multiplicative replacement strategy,

the time series analysis of compositional data.

We'll present the current status of `MixeR` development and illustrate its use on selected data sets.