Compositional Non-parametric Tests

K. G. van den Boogaart[1,2], R. Tolosana Delgado[1]
[1]Helmholtz Zentrum Dresden Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany; *tolosana@hzdr.de, boogaart@hzdr.de*
[2]TU Bergakademie, Institute for Stochastics, Freiberg, Germany

Abstract

Current compositional methods typically rely on some sort of normality hypothesis for testing. Typical well-known non-parametric tests rely on ranks transforms, which are undefined for multivariate problems. The aim of this contribution is thus to investigate the possibilities for truly multivariate non-parametric tests of location and distribution for compositonal data. The challenge is the to ensure subcompositional coherence, which would bring the possibility to attribute deviations to certain subcompositions.

For the case of tests for a known compositional mean, we propose a bootstrap method, measuring how extreme this mean is with respect to a bootstrap sample of the empirical compositional mean. The extremity is checked in each of the pairwise log-ratios. This ensures a subcompositional coherence in the sense that a rejected hypothesis will always be rejected in at least one subcomposition.

For the case of a two sample test comparing two populations, the same principle can be extended. The mean difference is compared with bootstrap samples of mean differences. The same subcompositional coherence applies.

For multiple samples we can extend the idea of ANOVA of measuring the variability of the group means. The variability is measured in terms of the variation matrix. For each entry of the variatin matrix we quantify its quantile in the bootstrap population. We take then the maximum of the quantiles and bootstrap this maximum. In this way we again get a subcompositionally test for equal mean in all samples. Weighted modifications might improve power of the test in case of unequal sample sizes.

Compositional Non-parametric Regression

R. Tolosana Delgado[1], **K. G. van den Boogaart**[1,2]
[1]Helmholtz Zentrum Dresden Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Freiberg, Germany; *tolosana@hzdr.de, boogaart@hzdr.de*
[2]TU Bergakademie, Institute for Stochastics, Freiberg, Germany

Abstract

Compositional regression is concerned with modelling the dependence of a composition on one or more covariables, or vice versa. State of the art methods typically rely on the assumptions of linearity of the dependence and for tests on the additive logistic normal distribution of the errors. Several different solutions for non-linear regression and tests without normality assumption are available for non-compositional data. Based on them, this contribution derives non-parametric regression models and methods valid for compositional data.

With respect to the non-linear dependence, some sort of regularisation assumption is always required. Different classical approaches can be adapted for compositonal data. LOESS smoothing on pairwise log-ratios or logratio transforms would correspond to some sort of smooth (compositional) derivatives. Regression splines and smoothing splines are already defined in a multivariate way and allow to control the degree of continuity and smoothness by explicit parameters. Piecewise regression needs to be applied to log-ratio transforms and allows to model non-continuous dependence. Geostatistical interpolation or, equivalently, reproducing kernel splines, allow a precise control over the level of continuity and complexity through the variogram.

All methods mentioned admit a multivariate extension which, by virtue of the principle of working in coordinates, automatically give rise to compositional versions of those methods. Moreover, all are either affine equivariant, or else very slight restrictions of them are. Thus, the associated compositional versions deliver results which are: invariant with respect to the choice of basis, scaling invariant, and subcompositionally coherent (in the case of regression with compositional response).

With regard to testing, there are some philosophical difficulties in a classical "zero slope hypothesis". A strict test for dependence could be very misleading when used for model selection in a non-parametric setting. As an alternative we propose to check, whether the prediction by the non-parametric model outperforms the prediction by parametric (constant) one. We propose to compare the jacknifed residuals of the two models. This construction allows to construct all meaningful tests of compositional dependence, namely: global lack of dependence, lack of dependence within a subcomposition, as well as restricted dependence within a subcomposition.