

SIMPLICIAL REGRESSION. THE NORMAL MODEL

JUAN JOSÉ EGOZCUE

Dept. Matemàtica Aplicada III, U. Politècnica de Catalunya (UPC), Barcelona, Spain.
Email: juan.jose.egozcue@upc.edu

JOSEP DAUNIS-I-ESTADELLA

Dept. Informàtica i Matemàtica Aplicada, U. de Girona (UdG), Girona, Spain.
Email: pepus@ima.udg.edu

VERA PAWLOWSKY-GLAHN

Dept. Informàtica i Matemàtica Aplicada, U. de Girona (UdG), Girona, Spain.
Email: vera.pawlowsky@udg.edu

KAREL HRON

Dept. of Mathematical Analysis and Application of Mathematics, Palacký U. (UPOL), Olomouc,
Czech Republic.
Email: hronk@seznam.cz

PETER FILZMOSE

Dept. of Statistics and Probability Theory, Vienna U. of Technology (TU-Wien, Vienna, Austria.
Email: P.Filzmoser@tuwien.ac.at

SUMMARY

Regression models with compositional response have been studied from the beginning of the log-ratio approach for analysing compositional data. These early approaches suggested the statistical hypothesis of logistic-normality of the compositional residuals to test the model and its coefficients. Also, the Dirichlet distribution has been proposed as an alternative model for compositional residuals, but it leads to restrictive and not easy-to-use regressions. Recent advances on the Euclidean geometry of the simplex and on the logistic-normal distribution allow re-formulating simplicial regression with logistic-normal residuals. Estimation of the model is presented as a least-squares problem in the simplex and is formulated in terms of orthonormal coordinates. This estimation decomposes into simple linear regression models which can be assessed independently. Marginal normality of the coordinate-residuals suffices to check influence of covariables using standard regression tests. Examples illustrate the proposed procedures.

Keywords and phrases: Aitchison geometry, normal distribution on the simplex, isometric log-ratio transformation (ilr), orthonormal coordinates, log-ratio analysis.

Subject Classification: 62J05, 62J02, 86A32, 91B42.