S8

Estructura, Análisis y Estadística Simpliciales

Simplicial Structure, Analysis and Statistics

hrs.: 6

email: `josepantoni.martin@udg.edu`

Organizadores / Organizers:

- Josep Antoni Martín Fernández (Universitat de Girona)

Conferencias / Talks:

Fri 4, 15:30 - 16:30, M1 – Vera Pawlowsky-Glahn (Universitat de Girona):
Geometría del símplex.

Fri 4, 16:30 - 17:30, M1 – Carles Barceló-Vidal (Universitat de Girona):
Calculus on the Simplex.

Fri 4, 18:00 - 19:00, M1 – Juan José Egozcue (Univ. Politècnica de Catalunya):
Espacios de Hilbert de funciones de densidad.

Fri 4, 19:00 - 19:30, M1 – José Antonio Martín-Fernández (Universitat de Girona):
Medidas de diferencia en el símplex.

Fri 4, 19:30 - 20:30, M1 – Karel Hron (Palacky University, Olomouc (Czech Republic)):
Mathematical elements in robust statistics for CODA.

Sat 5, 11:00 - 11:30, M1 – Glòria Mateu-Figueras (Universitat de Girona):
Modelos de probabilidad sobre el símplex.

Sat 5, 11:30 - 12:00, M1 – Raimón Tolosana (Univ. Politècnica de Catalunya):
Momentos en el símplex.

Sat 5, 12:00 - 12:30, M1 – Javier Palarea-Albaladejo (Biomathematics and Statistics Scotland (BioSS), Edinburgh, UK):
Practical issues in statistical modelling of CODA: dealing with zero values.

Practical issues in statistical modelling of CODA: dealing with zero values[**]

J. Palarea-Albaladejo[1], **J.A. Martín-Fernández**[2]

After [1], progress in compositional data analysis has been mainly inspired by the log-ratio approach. It is well-known that both ratio as well as logarithm operation are not applicable to zero values. In fact, the chance of zero components within a composition is left out from the own formal definition of the simplex. However, when we move into the practical side, the presence of zero values in a compositional data set is not unusual. Since the initial approaches were proposed by [1], several contributions have been made to address the zero problem under different assumptions. But the key is that any proposal must satisfy some basic principles in order to respect the vector space structure of the simplex. Three main types of zeros are usually distinguished in the CODA literature: essential or structural zeros, rounded zeros and count zeros.

Essential zeros refer to components which are truly zeros, they mean absolute absence of certain components. They represent the more complicated case, and specific models combining zero and non-zero components in a convenient way are required. Although some proposals have been made (e.g. [2], [3] and [4]), a generally applicable approach is still to be developed. On the other hand, very small values are in practice frequently rounded off to zero so as to fit a pre-fixed decimal precision. Other times they are not even recorded because the measuring devices are not able to detect them, being then included as a zero or as a less-than annotation in the data. Since, in essence, both cases—rounding-off error and existence of detection limit—represent the same kind of situation, we generically name them as rounded zeros. It is the case most often arising in classic CODA application field such as geosciences, environmental sciences, or similar. We focus here on the non-parametric multiplicative replacement method presented by [6], and on the parametric modified EM algorithm introduced in [7] and [8]. Finally, what we call count zeros may arise when working with compositional count data. Count data are vectors accounting for the number of times each category of a categorical vector occurs. Compositional methods come in handy to analyse them where the total sum of occurrences is not relevant. Here, a zero count may not be considered as a true value but a consequence of the sampling, since in another replication of the experiment it may be taking a positive count. We give details on the Bayesian-multiplicative approach proposed in [5] and successfully applied in [9].

In this presentation we review the implications of the presence of zeros in a compositional data set. For each kind of zeros, an overview of the particularities, advantages and drawbacks of the zero treatments cited above is provided. Finally, practical issues and possible ways of improvement are discussed.

Keywords: compositional data, simplex, zeros, detection limits

Mathematics Subject Classification 2000:   62Pxx, 62-07, 65C05

Referencias

[1] J. Aitchison. *The Statistical Analysis of Compositional Data*. Chapman & Hall Ltd., London, 1986 (reprintd in 2003 with additional material by The Blackburn Press).

[2] J. Aitchison and J. Kay. Possible solution of some essential zero problems in compositional data analysis. In *Compositional Data Analysis Workshop - CoDaWork 03, Proceedings*, S. Thió-Henestrosa and J. A. Martín-Fernández (eds.). Universitat de Girona, Girona, 2003.

[3] J. Bacon-Shone. Modelling structural zeros in compositional data. In *Compositional Data Analysis Workshop - CoDaWork 03, Proceedings*, S. Thió-Henestrosa and J. A. Martín-Fernández (eds.). Universitat de Girona, Girona, 2003.

[4] J. Bacon-Shone. Discrete and continuous compositions. In *Compositional Data Analysis Workshop - CoDaWork 08, Proceedings*, J. Daunis-i-Estadella and J. A. Martín-Fernández (eds.). Universitat de Girona, Girona, 2008.

[5] J. Daunis-i-Estadella. Bayesian tools for count zeros in compositional data. In *Compositional Data Analysis Workshop - CoDaWork 08, Proceedings*, J. Daunis-i-Estadella and J. A. Martín-Fernández (eds.). Universitat de Girona, Girona, 2008.

[6] J. A. Martín-Fernández, C. Barceló-Vidal and V. Pawlowsky-Glahn. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology* **35**(3), 253–278, 2003.

[7] J. Palarea-Albaladejo, J. A. Martín-Fernández and J. Gómez-García. Parametric approach for dealing with compositional rounded zeros. *Mathematical Geology* **39**(7), 625–645, 2007.

[8] J. Palarea-Albaladejo and J. A. Martín-Fernández. A modified EM alr-algorithm for replacing rounded zeros in compositional data sets. *Computer & Geosciences* **34**(8), 902–917, 2008.

[9] M. E. R. Pierotti, J. A. Martín-Fernández and O. Seehausen. Mapping individual variation in male mating preference space: multiple choice in a colour polymorphic cichlid fish. *Evolution* **63**(9), 2372–2388, 2009.

[1]Biomathematics and Statistics Scotland
JCMB, The Kingś Buildings, Edinburgh, EH9 3JZ, UK
`javier@bioss.ac.uk`

[2]Dept. Informàtica i Matemàtica Aplicada
Universitat de Girona
Campus Montilivi, Edifici P-IV. E-17071, Girona, Spain
`josepantoni.martin@udg.edu`