

A THEORETICAL APPROACH TO T^2 CONTROL CHART FOR COMPOSITIONAL DATA

M. Vives-Mestres¹, J. Daunis-i-Estadella¹, J.A. Martín-Fernández¹

¹ Dept. Computer Sciences and Applied Mathematics, University of Girona, Campus de Montilivi, Girona, Spain
marina.vives@udg.edu

Compositions are vectors which components represent parts of a whole and have a constant sum. Many compositional data examples can be found in the chemical industry, such as chemical composition, impurity profile or aggregate compositions. The sample space of compositional data is the simplex and specific methods are necessary to deal with their restriction. The log-ratio methodology ([1]) provides a powerful framework to deal with that type of data. Recent advances and applications of this methodology to several fields are presented in [2].

Hotelling T^2 control chart is one of the most familiar multivariate statistical process control tools. Because it is based on the Mahalanobis distance from each sample to the center of the data, it takes into account the covariance structure of the data.

It is not possible to calculate the Mahalanobis distance of a raw composition because of the singularity of the covariance matrix. Due this fact, classical approach recommends eliminating one component -and thus the constant sum- to compute the Mahalanobis distance. As a consequence, the obtained values are not coherent with the compositional nature of the data.

We propose a new approach based on the log-ratio methodology which is equivalent to applying the classical methodology to the compositions expressed in coordinates with respect to an orthonormal basis. Working with coordinates enables to represent the data in real space and reduce dimensionality.

A simulation of a three part compositions is shown on a ternary diagram (Figure 1). Solid lines are the contour lines of the classical Mahalanobis distance calculated using only two parts. Dashed lines are produced by the Mahalanobis distance using the log-ratio coordinates. Note that the log-ratio approach is more consistent with the nature of the data.

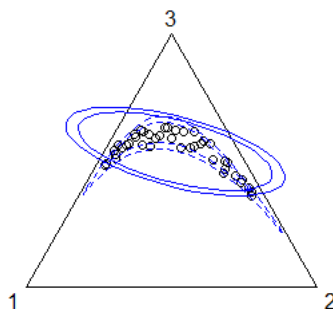


Figure 1: Contour lines of the classical (solid) and log-ratio (dashed) Mahalanobis distance.

References:

- [1] Aitchison, J., 1986. *The Statistical Analysis of Compositional Data*. Monographs on Statistics and Applied Probability. Chapman and Hall Ltd. (Reprinted 2003 with additional material by The Blackburn Press), London (UK). 416 p.
- [2] Pawlowsky, V. and Buccianti, A. (eds), 2011. *Compositional Data Analysis: Theory and Applications*. Chichester (UK), John Wiley & Sons (ISBN: 978-0-470-71135-4; 378p.).