Spatial regression model for compositional data in discrete space with spatial auto-correlation and spatial cross-correlation

T. Yoshida[1] **and M. Tsutsumi**[1]
[1]University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan;
yoshida.takahiro@sk.tsukuba.ac.jp

Abstract

Compositional data of geographical space, such as land-use compositional data, can have potential spatial correlation. In this regard, there are two aspects to be considered: 1) Spatial auto-correlation, which occurs when a category in a position is related to the same category in a different position, and 2) Spatial cross-correlation, which occurs when a category in a position is related to different categories in different positions, such as rice paddies tend to be located near water area.

This study proposes a spatial econometric model to examine the spatial association between compositional categories in discrete space where spatial auto-correlation and spatial cross-correlation arises. Traditionally, compositional data analysis uses geostatistical approaches to examine spatial association (*e.g.*, Pawlosky-Glahn and Olea, 2004; Martin and others, 2016). However, quite a few studies have used a spatial econometric approach (*e.g.*, Leininger and others, 2013). Geostatistics assumes that spatial processes occur in continuous and infinite space, whereas, spatial econometrics assumes that spatial processes occur in discrete and finite space. Spatial econometrics has accumulated the know-how on treating spatial data as data in discrete space, such as lattice data. Further, it provides us with useful tools with spatial association to extend the applications of spatial compositional data analysis to lattice data. In this study, we first empirically analyze the above mentioned two aspects of spatial association in compositional data using Japanese land-use data (1km $\times$ 1km grid data). We detect spatial cross-correlation using a bivariate measure of local spatial association, known as bivariate Moran's I (Lee, 2001), which captures the topological relationship between a category x_i in grid i and a neighborhood and different category y_j in grid j (not equal to i). Second, we develop a spatial regression model considering spatial cross-correlation in compositional data. In particular, following Leininger and others (2013), we combine the log-ratio transformation with the multivariate conditionally auto-regressive model. Finally, we compare the proposed model with the existing models (Leininger and others, 2013) in terms of their predictive accuracy using a simulation analysis whereby we assume a particular data generation process for each level of bivariate spatial correlation.

References

Lee, S.-I. (2001). Developing a bivariate spatial association measure: An integration of Pearson's r and Moran's I. *Journal of Geographical Systems 3*(4), pp. 369–385.

Leininger, T.J., A.E. Gelfand, J.M. Allen, and J.A. Silander Jr. (2013). Spatial regression modeling for compositional data with many zeros. *Journal of Agricultural, Biological, and Environmental Statistics 18*(3), pp. 314–334.

Martins, A.B.T., W.H. Bonat, and P.J. Ribeiro Jr. (2016). Likelihood analysis for a class of spatial geostatistical compositional models. *Spatial Statistics 17*, pp. 121–130.

Pawlowsky-Glahn, V. and R.A. Olea (2004). *Geostatistical Analysis of Compositional Data*. London: Chapman and Hall.