Taking into account spatial dependence in multivariate analysis

Sébastien Ollier, Anne-Béatrice Dufour, Jean Thioulouse, Daniel Chessel

Laboratoire de Biométrie et de Biologie Evolutive. UMR CNRS 5558, Villeurbanne, France.

Abstract

R is a valuable tool to develop statistical methods using its package structure and the open sources. Indeed, it allows combining classes and methods from different packages in order to generalize their own capabilities. The present topic deals with the relationship between classes and methods from both packages.**spdep** and **ade4**, which implement methods for spatial data analysis (mainly lattice/area style) and multivariate analysis respectively. Combining objects from both packages leads to the development of a new statistical method to take into account spatial dependence in multivariate analysis.

spdep is a package defined by Bivand as “a collection of functions to create spatial weights matrix **W** from polygon continuities, from point patterns by distance and tessellations for permitting their use in spatial data analysis”. Two classes of objects have been established: “nb” a list of vectors of neighbour indices, and “listw” a list containing a “nb” member and a corresponding list of weights for a chosen weighting scheme. The package contains in addition many functions to test spatial autocorrelation and estimate spatial autoregressive model but it only supports univariate data analysis.

ade4 is precisely a package dedicated to multivariate analysis of Ecological and Environmental Data in the framework of Euclidean Exploratory methods. It is a collection of functions to analyse the statistical triplet (**X**,**Q**,**D**) where **X** is a data set of variables or a distance matrix; **Q** and **D** are diagonal matrices containing column and row weights, respectively. The singular value decomposition of this triplet gives principal axes, principal components and, column and row coordinates. All these elements are gathered in a list defining the **du**ality **di**agram class called “dudi”. Each basic statistical method corresponds to the analysis of a particular triplet. For instance, **ade4** implement the principal component analysis on correlation matrix *via* a function called “dudi.pca”. In that case, **X** is a table containing normalized quantitative variables, **Q** is the identity matrix $\mathbf{I}_p$ and **D** is equal to $\frac{1}{n}\mathbf{I}_n$.

An illustration of the relationship between **spdep** and **ade4** is shown through the “multispati” function. This function combines objects of class ”listw “ with objects of class “dudi” to analyse the statistical quadruplets (**X**,**Q**,**D**,**W**). It allows first, the generalisation of the univariate Moran test and second, to take into account spatial constraints in Euclidean Exploratory methods. During this talk, the implementation of multivariate spatial correlations analysis proposed by Wartenebrg (1985) will be described as a particular use of the “multspati” function.

Wartenberg, D. E. 1985. Multivariate spatial correlations: a method for exploratory geographical analysis. Geographical Analysis **17**:263-283.