

Likelihood-Based Finite Mixture Models for Ecological Ordinal Data

D. Fernandez^a, R. Arnold^a and S. Pledger^a

^aSchool of Mathematics, Statistics and Operations Research
Victoria University of Wellington
Wellington, New Zealand
daniel.fernandez@msor.vuw.ac.nz

Keywords: Biclustering; Cluster analysis; Community structure and dynamics; Dimension reduction; EM-algorithm; Finite mixture model; Fuzzy clustering; Likert scale; Multispecies models; Ordinal data; Species distribution models; Stereotype model

Abstract: Many of the methods which deal with the reduction of dimensionality in matrices of data are based on mathematical techniques, such as distance-based algorithms or matrix decomposition and eigenvalues. In general, it is not possible to assess objectively the appropriateness of a model via information criteria with these techniques because there is no underlying probability model. Furthermore, ordinal data are very common in ecological data (e.g. Braun-Blanquet scale). It is advantageous in an analysis to properly treat an ordered categorical variable as ordinal, rather than using the common approach of treating such a variable as nominal or continuous. Recent research has developed a set of likelihood-based finite mixture models for a data matrix of binary or count data (Pledger and Arnold, 2014). We present the extension of this work by formulating likelihood-based multivariate methods for ecological ordinal data. This approach applies fuzzy clustering via finite mixtures to the (ordinal) stereotype model. Fuzzy allocation of rows and columns to corresponding clusters is achieved by performing the expectation-maximization (EM) algorithm, and also by Bayesian model fitting, and their performances are compared. Data from two ecological community data sets are used to illustrate the application of this approach and also to demonstrate new and convenient visualisation tools for ordinal data. Finally, we present the results of a simulation study carried out to determine which information criteria are most appropriate for model selection with these particular mixture models when applied to ordinal data.

References

Pledger, S., Arnold, R. (2014). Multivariate methods using mixtures: Correspondence analysis, scaling and pattern-detection. *Computational Statistics and Data Analysis*, 71, 241-261.