On the Connection between the Distribution of Eigenvalues in Multiple Correspondence Analysis and Log-Linear Models

S. Ben Ammou and G. Saporta
Faculté de Droit & des Sciences Economiques & Politiques de Sousse and CEDRIC, Conservatoire National des Arts et Métiers, Paris
saloua.benammou@fdseps.rnu.tn and gilbert.saporta@cnam.fr

Abstract

Multiple Correspondence Analysis (MCA) and log-linear modelling are two techniques of multi-way contingency table analysis having different problematics and fields of applications. Log-linear models are profitable when applied to a small number of variables (Bishop & al.,1975). Multiple Correspondence Analysis is useful in large tables (Lebart & al. 2000)). This efficiency is balanced by the fact that MCA is not able to explicit relations between more than two variables, as can be done by log-linear modelling (Andersen, 1991). The two approaches are complementary.

In this presentation we shall demonstrate that, in MCA, under independence hypothesis each observed eigenvalue is asymptotically normally distributed. These distributions have the same mean, different variances and converge to normal distribution (Ben Ammou ,1996, Ben Ammou & Saporta 1998).
Under some modelling hypothesis, the MCA eigenvalues distribution diagram takes some particular shapes, especially in the case of mutual independence model (theoretically there is only one non trivial, multiple eigenvalue $\lambda = 1/p$, where p is the number of variables), in practice, observed eigenvalues μ_i are different but still close to $1/p$: $\mu_i = 1/p \pm \varepsilon$. Therefore the shape of observed eigenvalues diagram is very peculiar. This shape changes if there is one or more interaction between variables. We can recognize the model fitted by data in some particular cases, especially when the number of interactions is not very large i.e. we can easily identify the observed eigenvalues that are equal (or very close) to 1/p. When the number of interactions increases, we can no more distinguish between eigenvalues theoretically equals to 1/p and those different from 1/p.
Based on these results we propose a simple procedure, fitting progressively log-linear models, where the goodness of fit procedure is based on MCA eigenvalues diagram: the model is inducted by successive utilisations of MCA (non constrained by the number of variables).

The procedure is validated on several data sets from the literature corresponding to various cases: mutual independence, saturated models and graphical models with two way interactions.

References

Andersen, E.B. (1991). *The Statistical Analysis of Categorical Data*. (Second edition), New-York, Springer.

Ben Ammou S. (1996). *Comportement des valeurs propres en Analyse des Correspondances Multiples sous certaines hypothèses de modèles*. Doctoral Thesis, University Paris IX Dauphine.

Ben Ammou, S. Saporta, G. (1998). Sur la normalité asymptotique des valeurs propres en ACM sous l'hypothèse d'indépendance des variables. *Revue de Statistique Appliquée*, vol XLVI, n°3 p 21-35.

Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W (1975). *Discrete Multivariate Analysis : Theory and Practice*. Boston, MIT Press.

Lebart, L., Morineau, A. & Piron, M. (2000) *Statistique exploratoire multidimensionnelle*, 3ème édition, Paris, Dunod.