Maximum likelihood estimation for the logistic-normal-multinomial distribution

M. Comas-Cufí[1], J.A. Martín-Fernández[1], G. Mateu-Figueras[1] and J. Palarea-Albaladejo[2]

[1]Universitat de Girona, Girona, Spain; *mcomas@imae.udg.edu*

[2]Biomathematics and Statistics Scotland, Edinburgh, Scotland

Abstract

The logistic-normal-multinomial distribution is the compounding probability distribution resulting from considering the multivariate logistic-normal as the distribution for the probability parameter vector of the multinomial distribution. This distribution can be used to model multivariate count data when only the relative relations between parts are of interest. It may be considered as an alternative to the Dirichlet-multinomial distribution and it can be used to deal with counting zeros (Comas-Cufí and others, 2016).

Let $\mathcal{B}$ be an orthonormal basis of the simplex, the probability mass function in terms of isometric log-ratio coordinates is

$$P\left(\{\boldsymbol{X}=(x_1,\ldots,x_{k+1})\};n,\boldsymbol{\mu},\boldsymbol{\Sigma}\right)=\int_{\mathbf{h}\in\mathbb{R}^k}\mathcal{N}(\mathbf{h};\boldsymbol{\mu},\boldsymbol{\Sigma})\,\frac{n!}{\prod_{\ell=1}^{k+1}x_\ell!}\,\prod_{\ell=1}^{k+1}\mathrm{ilr}_\ell^{-1}(\mathbf{h})^{x_\ell}\,\mathrm{d}\mathbf{h},$$

where n is the multinomial "size" parameter, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are respectively the mean and covariance matrix of the logistic-normal expressed with respect to the basis $\mathcal{B}$ and $\mathcal{N}$ refers to the multivariate normal density function.

Billheimer and others (2001) use a Bayesian approach to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by means of Markov chain Monte Carlo simulation. Xia and others (2013) propose a Monte Carlo EM algorithm to fit the parameters where the expected values are calculated using a Metropolis-Hasting algorithm. Hughes and other (1998) consider the univariate case which can be solved using numerical integration. In all these works, the proposed approach gives satisfactory results when it is applied to a particular case. However, when it is applied in more general settings, problems may appear. In consequence, for general scenarios, a more comprehensive proposal is required.

In this work we consider different alternatives to calculate the expected value in the Monte Carlo EM algorithm and compare their performance. In particular, for the method proposed by Xia and others (2013), we subtitute the Metropolis-Hasting algorithm by a standard Monte Carlo algorithm using different variance reduction techniques. We also analyse how the selection of an specific initial value affects the results obtained by the EM scheme.

References

Billheimer, D., Guttorp, P. & Fagan, W. F. (2001). Statistical Interpretation of Species Composition. *Journal of the American Statistical Association*, **96(456)**, 1205–1214.

Comas-Cufí, M., Palarea-Albaladejo, J., Martín-Fernández, J.A. & MateuFigueras, G. (2016) A parametric approach to count zeros imputation in compositional data sets. *Proceedings of the Royal Statistical Society International Conference.* p.4.

Hughes, G., Munkvold, G.P. & Samita, S. (1998). Application of the logistic-normal-binomial distribution to the analysis of Eutypa dieback disease incidence. *International Journal of Pest Management*, **44(1)**, 35–42.

Xia, M., Chen, J., Fung K. F. & Li H. (2013). A Logistic Normal Multinomial Regression Model for Microbiome Compositional Data Analysis. *Biometrics*, **69**, 1053–1063.