The Extended Flexible Dirichlet model: a simulation study

Roberto Ascari[1], Sonia Migliorati[2], and Andrea Ongaro[2]

[1] Department of Statistics and Quantitative Methods (DISMEQ) Via Bicocca degli Arcimboldi, 8, Università di Milano-Bicocca, Milano, Italy
(E-mail: r.ascari@campus.unimib.it)

[2] Department of Economics, Management and Statistics (DEMS) P.zza dell'Ateneo Nuovo, 1, Università di Milano-Bicocca, Milano, Italy
(E-mail: sonia.migliorati@unimib.it; andrea.ongaro@unimib.it)

Abstract. Compositional data are prevalent in many fields (e.g. environmetrics, economics, biology, etc.). They are composed by positive vectors subject to a unit-sum constraint (i.e. they are defined on the *simplex*), proportions being an example of this kind of data. A very common distribution on the simplex is the Dirichlet, but its poor parametrization and its inability to model many dependence concepts make it unsatisfactory for modeling compositional data. A feasible alternative to the Dirichlet distribution is the Flexible Dirichlet (FD), introduced by Ongaro and Migliorati [1]. The FD is a generalization of the Dirichlet that enables considerable flexibility in modeling dependence as well as various independence concepts, though retaining many good mathematical properties of the Dirichlet. More recently, the Extended Flexible Dirichlet (EFD, [2]) distribution has been proposed in order to generalize the FD. The EFD preserves a finite mixture structure as the FD, but it exhibits some relevant advantages over the FD, such as a more flexible cluster structure and a (even strong) positive dependence for some pairs of variables. The aim of this contribution is twofold. First we propose and investigate sophisticated EM algorithms for parameters estimation, with particular emphasis on the initialization problem, which is a crucial issue. Furthermore, we devise a simulation study to evaluate the performances of the MLE of the parameters as well as of a procedure proposed to compute their standard errors.
Keywords: Compositional Data, Dirichlet Mixture, EM algorithm.

References

1. A. Ongaro and S. Migliorati. A generalization of the Dirichlet distribution. *Journal of Multivariate Analysis*, 114, 412–426, 2013.
2. A. Ongaro and S. Migliorati. A Dirichlet mixture model for compositions allowing for dependence on the size. In M. Carpita, E. Brentari and E.M. Qannari (Eds), *Advances in Latent Variables Methods, Models and Applications*, Springer, 2014.
3. S. Migliorati and A. Ongaro. Estimation issues in the Extended Flexible Dirichlet model. *16th ASMDA Conference Proceedings*, Greece, 2015.