ESTIMATION AND MODEL SELECTION IN DIRICHLET REGRESSION

A. Camargo[1], J. M. Stern[1], M. S. Lauretto[2]
(1) Institute of Mathematics and Statistics, University of Sao Paulo
(2) School of Arts, Sciences and Humanities, University of Sao Paulo
andreuler@yahoo.com.br, marcelolauretto@usp.br

Abstract

Compositional data consist of vectors whose components are the proportions or percentages of some whole. The peculiarity of this family is that their sum is constrained to be some constant (usually 1). Hence, the corresponding sample space is the simplex, that is quite different from the real Euclidean space associated with unconstrained data. Therefore, attempts to apply statistical methods for unconstrained data often lead to inappropriate inference. Some statistical models for compositional data have been developed since the 70s, particularly for regression analysis.

Here we focus on the Dirichlet Covariate Model, suggested by Campbel and Mosimann (1987). In this model, one considers $\mathbf{y} = (y_1, \ldots, y_D)$ to be a $1 \times k$ positive vector having Dirichlet distribution $\mathcal{D}(\alpha_1, \ldots, \alpha_D)$. In this approach, a Dirichlet regression model is readily obtained by allowing the parameters of a Dirichlet distribution to change with a covariate. For a given covariate vector $\mathbf{x} = (x_1, \ldots, x_C)$, each parameter α_j may be written as a positive-valued function $\lambda_j(\mathbf{x})$ of the covariates $\mathbf{x}$. Thus, a different Dirichlet distribution is modeled for every value of the covariates, resulting in a conditional Dirichlet distribution with $\mathbf{y}|\mathbf{x} \sim \mathcal{D}(\lambda_1(\mathbf{x}) \ldots \lambda_D(\mathbf{x}))$.

In this work, we introduce a new method for estimating the parameters of the Dirichlet Covariate Model, considering $\lambda_j(\mathbf{x})$ as polynomials. We also propose a model selection approach based on the Full Bayesian Significance Test (Pereira and Stern, 1999). Problems of interest are, for example, variable selection and choice of polynomial order.

References:

[1] G.Campbell, J.Mosimann. Multivariate methods for proportional shape. ASA Proceedings of the Section on Statistical Graphics, 10-17 (1987).

[2] R.H.Hijazi, R.W.Jernigan. Modelling Compositional Data Using Dirichlet Regression Models. Journal of Applied Probability & Statistics **4**, 77-91 (2009).

[3] C.A.B.Pereira, J.M.Stern. Evidence and Credibility: Full Bayesian Significance Test for Precise Hypotheses. Entropy **1**, 99-110 (1999).

Key Words: Dirichlet regression, model selection