Contribution ID: **76** Type: **not specified**

Small area models for count data: an area-level EFD model

In this contribution, we focus on small-area compositional data. These data are defined as vectors whose elements are strictly positive and sum to one (e.g., proportions). Compositional data arise in various fields, including medicine, economics, psychology, and environmetrics. They are defined on the D-part simplex (S^D) and require complex techniques for proper analysis.
A traditional approach involves applying log-ratio transformations, which map the D-part simplex onto a (D−1)-dimensional real space. However, this approach has several limitations: parameter estimates are interpretable only in the transformed space, and issues such as skewness, heteroscedasticity, non-normality, and outliers may bias inference.

An alternative solution for regression with compositional responses is the Dirichlet model. It is often implemented using a multinomial logit function to link the response mean vector to covariates, allowing for a straightforward interpretation of regression coefficients in terms of log-odds ratios. Nevertheless, the Dirichlet model has significant drawbacks: it uses only a single parameter to describe the entire variance–covariance matrix (the precision parameter), offers limited flexibility, and implies several forms of simplicial independence. Moreover, it always yields negative covariances, making it unable to model many relevant phenomena (e.g., heavy-tailed or multimodal responses).

We therefore propose using the Extended Flexible Dirichlet (EFD), a structured mixture of Dirichlet components. Unlike general Dirichlet mixture models, the parameters of each EFD component are strictly linked to one another, providing greater flexibility. We employ the EFD for modelling small-area data: we reparameterize the model in terms of mean and precision and incorporate covariates to account for design effects. We estimate the model within a fully Bayesian framework and assess its performance using simulated data.

Special/ Invited session

Classification

Mainly methodology

Keywords

EFD, counts, Bayesian models

Primary author: ARIMA, Serena (University of Salento)

Co-author: POLETTINI, Silvia (University of Rome ”La Sapienza”)

Presenter: ARIMA, Serena (University of Salento)

Track Classification: Other/special session/invited session