

Regularized log-bilinear models.

Julie Josse

Log bilinear models, also known as row-column (RC) association models are well suited to describe contingency tables. The expectation of the counts (on the log scale) is modeled with main row and column effects and an interaction which has a low rank representation. Parameters of the RC models can also be interpreted as describing latent variables in a low-dimensional Euclidean space. However, solving these low-rank log-linear models is non-trivial and standard methods of estimation suffer, especially when the rank is large and the table is sparse. In this talk, I will present a penalized version of the Poisson likelihood to tackle overfitting issues. I will illustrate the method on the specific case of symmetric models to describe a cross-citations matrix where each element corresponds to the number of citations given by a journal to papers from another journal. The analysis of such data often aims at ranking the journals and thus we will offer an alternative to a Bradley terry model.