Hierarchical modelling of patient-reported outcomes data based on the beta-binomial distribution

Dae-Jin Lee

BCAM – Basque Center for Applied Mathematics, Bilbao, Spain

The beta-binomial distribution does not belong to the exponential family and, hence classical regression techniques cannot be used when dealing with outcomes following the mentioned distribution. In this thesis we propose and develop regression models based on the beta-binomial distribution for the analysis of U, J or inverse J-shaped discrete and bounded outcomes. In fact, although this thesis is focused on the analysis of patient-reported outcomes (PROs), which usually follow the mentioned distributional shapes, proposed models can also be extended to several fields. First of all, we make a review and comparison of existing beta-binomial regression approaches in independent data context, concluding that the marginal approach is the most adequate. However, PRO studies are usually carried out in a longitudinal framework, where patients' responses are measured over time. This leads to a multilevel or correlated data structure and consequently, we extend the marginal beta-binomial regression approach to the inclusion of random effects to accommodate the hierarchical structure of the data. We develop the estimation and inference procedure for the model proposal. Furthermore, we compare the performance of our proposal with similar approaches in the literature, showing that it gets better results in terms of reducing the bias of the estimates. We apply the model to a longitudinal Chronic Obstructive Pulmonary Disease study carried out at Galdakao Hospital, reaching clinically and statistically relevant results about the evolution of the patients over time. PROs are usually obtained using rating scale questionnaires consisting of questions or items, grouped into one or more subscales, often called dimensions or domains. Therefore, we also propose a multivariate regression model based on the beta-binomial distribution for the joint analysis of all the longitudinal dimensions provided by different questionnaires. Finally, it is worth mentioning that we have implemented all the proposed regression models in the PROreg R- package which is freely available at CRAN.