An Extended Random-effects Approach to Modeling Repeated, Overdispersed Count Data

Clarice G.B. Demétrio[1], Geert Molenberghs[2] and Geert Verbeke[3]

[1] Departmento de Ciências Exatas, ESALQ, Universidade de São Paulo, Brasil
[2] Center for Statistics, Hasselt University, Diepenbeek, Belgium
[3] Biostatistical Centre, Katholieke Universiteit Leuven, Leuven, Belgium

Non-Gaussian outcomes are often modeled using members of the so-called exponential family. The Poisson model for count data falls within this tradition. The family in general, and the Poisson model in particular, are at the same time convenient since mathematically elegant, but in need of extension since often somewhat restrictive. Two of the main rationales for existing extensions are (1) the occurrence of overdispersion (Hinde and Demétrio 1998, Computational Statistics and Data Analysis 27, 151-170), in the sense that the variability in the data is not adequately captured by the model's prescribed mean-variance link, and (2) the accommodation of data hierarchies owing to, for example, repeatedly measuring the outcome on the same subject (Molenberghs and Verbeke 2005, Models for Discrete Longitudinal Data, Springer), recording information from various members of the same family, etc. There is a variety of overdispersion models for count data, such as, for example, the negative-binomial model. Hierarchies are often accommodated through the inclusion of subject-specific, random effects. Though not always, one conventionally assumes such random effects to be normally distributed. While both of these issues may occur simultaneously, models accommodating them at once are less than common. This paper proposes a generalized linear model, accommodating overdispersion and clustering through two separate sets of random effects, of gamma and normal type, respectively (Molenberghs, Verbeke and Demétrio 2007, LIDA, 13, 513-531). This is in line with the proposal by Booth, Casella, Friedl and Hobert (2003, Statistical Modelling 3, 179-181). The model extends both classical overdispersion models for count data (Breslow 1984, Applied Statistics 33, 38-44), in particular the negative binomial model, as well as the generalized linear mixed model (Breslow and Clayton 1993, JASA 88, 9-25). Apart from model formulation, we briefly discuss several estimation options, and then settle for maximum likelihood estimation with both fully analytic integration as well as hybrid between analytic and numerical integration. The latter is implemented in the SAS procedure NLMIXED. The methodology is applied to data from a study in epileptic seizures. A general framework, encompassing the binary and Poisson types as special cases, is currently under investigation, in particular in the area of the correlation function for which there also are closed-form expressions.