HANDLING NEGATIVE CORRELATION AND/OR OVER/UNDERDISPERSON IN GAUSSIAN AND NON-GAUSSIAN HIERARCHICAL DATA

Geert Molenberghs
Interuniversity Institute for Biostatistics and statistical Bioinformatics
(1) I-BioStat, Hasselt University, Hasselt, Belgium
(2) I-BioStat, KU Leuven, Belgium

Abstract

The occurrence and interpretation of negative variance components in the context of linear mixed models is well understood at this point, even though the issue is surrounded by subtle issues for estimation and testing (Verbeke and Molenberghs 2003, Molenberghs and Verbeke 2007). Broadly, negative variance components often point to negative within-cluster correlation. It is even possible to give such linear mixed models a meaningful hierarchical interpretation (Molenberghs and Verbeke 2011). Matters are more complicated when the outcomes are non-Gaussian, either in the context of the generalized linear mixed model, or extensions thereof that allow for flexible modeling of both within-unit correlation as well as overdispersion (Molenberghs *et al.* 2010). An additional complication is that, in practice, not only negative variance components due to negative correlation, but also underdispersion instead of overdispersion can occur, sometimes even jointly. With focus on both continuous and count data, we describe how models can be made sufficiently flexible and, in a number of cases, interpreted hierarchically (Luyts *et al.* 2019).

References

Luyts, M., Molenberghs, G., Verbeke, G., Matthijs, K., Demétrio, C.G.B., and Hinde, J. (2019). A Weibull-count approach for handling under- and/or over-dispersed clustered data structures. *Statistical Modeling,* **19,** 569-589.

Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, **61**, 1-6.

Molenberghs, G. and Verbeke, G. (2011). A note on a hierarchical interpretation for negative variance components. *Statistical Modeling*, **11**, 389-408.

Molenberghs, G., Verbeke, G., Demétrio, C.G.B., and Vieira, A. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325-347.

Verbeke, G. and Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics*, **59**, 254-262.