Double Generalized Linear Model From a Bayesian Perspective

Afranio M. C. Vieira[1], Roseli A. Leandro[2] and Clarice G. B. Demetrio[2]

[1] Departamento de Estatistica, Universidade de Brasilia, BRAZIL
[2] Departamento de Ciencias Exatas, ESALQ - Universidade de Sao Paulo, BRAZIL

The variability modelling under experimental designs became popular due to the Genichi Taguchi works, that highlights the importance to deal with control and noise factors in industrial experiments, as a way to reduce loss and to optimize the quality of related processes and products. If factors are acting over the variation, the usual normal regression models (that are based on constant variance) will be based on inflated variation measures, guiding to misleading conclusions. For count and proportion data, these noise factors can generate an overdispersion effect, underestimating the variability and, consequently, indicating erroneous significatively effects [3].
Some statistical models were proposed, based on the classical generalized linear models for a joint modelling strategy [4], where the extended quasi-likelihood function was used in the estimation algorithm. However, this approach is highly dependent of asymptotic results and so, large samples are required to produce reliable inference. A Bayesian joint modelling for data with normal distribution that independs of large samples was proposed by [1]. It allows the use of prior knowledge about the control and noise effects and is adequated for many small sample agricultural experiments.
In this work we propose a double generalized linear model from a Bayesian perspective, focusing in the case of proportion data where the overdispersion can be modeled through a random effect that depends of some noise factors. It was implemented in `R` code using the `BRugs` library, to obtain a sample of the joint posterior distribution using MCMC algorithms like Gibbs Sampling and Metropolis-Hastings, which use the posterior full conditional densities. The Gelman-Rubin procedure was used as convergence diagnosis of the Markov chains [2], through the`coda` library for the `R` computing environment [5]. An application to an apple tissue culture data is shown, where the candidate models where selected through the *deviance information criteria*. The Bayesian approach is quite feasible to apply a double GLM, even if few prior information is available, generating valuable information for the researcher about its experimental results.

[1]The first and third authors were partially supported by The Brazilian National Council for Scientific and Technological Development (CNPq).

References

[1] Cepeda E, Gamerman D (2000) Bayesian modeling of variance heterogeneity in normal regression models. Brazilian Journal of Probability and Statistics 14:207-221

[2] Gelman A, Rubin DR (1992) Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 7:457-511

[3] Hinde J, Demetrio CGB (1998) Overdispersion: models and estimation. Computational Statistics & Data Analysis 27:151-170

[4] Nelder JA, Lee Y (1991) Generalized linear models for the analysis of Taguchi-type experiments. Applied Stochastic Models and Data Analysis, 7:107-120

[5] R Development Core Team (2007) R: A language and environment for statistical computing, version 2.6.0. Vienna: R Foundation for Statistical Computing, ISBN 3-900051-07-0, URL `http://www.r-project.org`