A Marginalized Model for Zero-Inflated, Overdispersed, and Correlated Count Data

SAMUEL IDDI1 and GEERT MOLENBERGHS1,2
1I-BioStat, Katholieke Universiteit Leuven, Belgium
2I-BioStat, Universiteit Hasselt, Belgium
Contact: samuel.iddi@med.kuleuven.be

Abstract

Iddi and Molenberghs (2012) merged the attractive features of the so-called combined model (CM) of Molenberghs et al (2010) and the marginalized multilevel model (MMM) of Heagerty (1999) for hierarchical non-Gaussian data with overdispersion. In this model, the fixed-effect parameters retain their marginal interpretation. Lee et al (2011) also developed an extension of Heagerty (1999) to handle zero-inflation from count data, using the hurdle model. To bring together all of these features, a marginalized, zero-inflated, overdispersed model for correlated count data is proposed. Using an empirical dataset, it is shown that the proposed model leads to important improvements in model fit.

Methodology

- Denote $Y_{ij}$ the $j$th ($j = 1,2,\ldots, n_i$) outcome measured for cluster (subject) $i = 1,2,\ldots, N$.
- $Y_{ij} = (Y_{ij1}, Y_{ij2}, \ldots, Y_{ijm})'$ follow Poisson distributions with mean number of events $\lambda_{ij}$.

1. Combined Overdispersed and Marginalized Multilevel Model (COMMM)

The COMMM model (Iddi and Molenberghs, 2012) is of the form:

$$\log(\lambda_{ij}) = \alpha_1 + \beta_j + \theta_{ij},$$

where $\alpha_1$, $\beta_j$, and $\theta_{ij}$ are the intercept, slope, and random effect, respectively.

- The full design is $X_i = (x_{i1}, x_{i2}, \ldots, x_{im})$ and $Z_i = (x_{i1}, x_{i2}, \ldots, x_{im}, \theta_i) \sim \text{Gamma}(\alpha_0, \sigma_0)$ and $b_j \sim N(0, \sigma^2)$.

- The connector function $\Delta_{ij}$ is solved from the equation:

$$\Lambda_{ij} = \exp(\alpha_1 + \beta_j + \theta_{ij}).$$

- $\Phi_{ij}(\lambda)$ and $\Phi_{ij}(\theta)$ are the cumulative distribution functions of $\theta_{ij}$ and $\beta_j$, respectively.

2. Zero-Inflated, Overdispersed, and Marginalized Multilevel Model (ZICOMMM)

We propose a ZICOMMM model as follows:

- Assume $Y_{ij} \sim \left\{ \begin{array}{ll} 0 & \text{with probability } \pi_{0ij}, \\ \text{Poisson}(\lambda_{ij}) & \text{with probability } (1 - \pi_{0ij}). \end{array} \right.$

- Marginal specification:

$$P(Y_{ij} = y_{ij}) = \pi_{0ij} + (1 - \pi_{0ij}) \exp(-\pi_{0ij})\sum_{z_{ij}} \exp(z_{ij} - \lambda_{ij})$$

- Next, a conditional specification:

$$P(Y_{ij} = y_{ij}|\theta_{ij}, b_j) = \pi_{0ij} + (1 - \pi_{0ij}) \exp(-\pi_{0ij})\sum_{z_{ij}} \frac{\exp(z_{ij} - \lambda_{ij})}{\lambda_{ij}^{y_{ij}} y_{ij}!}$$

- Estimation

- Based on maximum likelihood through partial marginalization.
- The observed data likelihood of the $i$th subject conditioned on the two random effects is:

$$f(\beta, \alpha, D, \phi) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(y_{ij}|\theta_{ij}, b_j) f(b_j|D) db_j,$$

where

$$f(y_{ij}|\theta_{ij}, b_j) = \int f(y_{ij}|\theta_{ij}, b_j, f(\theta_{ij}|D) d\theta_{ij},$$

from which we can derive the likelihood for $\beta, \alpha, D, \phi$ as

$$L(\beta, \alpha, D, \phi) = \prod_{i=1}^{N} \prod_{j=1}^{n_i} f(y_{ij}|\beta, \alpha, D, \phi).$$

- If $Y_{ij}$ follows Poisson with mean composing of a normal and a conjagate gamma distributed random effect terms, then:

$$f(y_{ij}|\theta_{ij}, b_j) = \frac{1}{\theta_{ij}^{y_{ij}} y_{ij}!} \exp\left(-\frac{y_{ij}}{\theta_{ij}}\right) \frac{1}{\theta_{ij}^{\sigma^2/2}} \exp\left(-\frac{\theta_{ij}}{2\sigma^2}\right)\frac{1}{\theta_{ij}^{\sigma^2/2}} \exp\left(-\frac{\theta_{ij}}{2\sigma^2}\right)$$

- Easy to fit within the SAS NLMIXED procedure.

Application in Epilepsy Clinical Trial

- Randomized, double-blind, parallel group multi-center study, Faught et al (1996)
- Comparing placebo with a new Anti-Epileptic Drug (AED) + one or two other AED’s
- Randomization after 12-weeks stabilization period (45 patients to placebo, 44 to the new treatment)
- Number of epileptic seizures measured for up to 27 weeks of follow-up
- Research interest is whether or not the additional new treatment reduces the number of seizures

Model and Results

- Let $Y_{ij}$ be the number of epileptic seizures experienced by the ith patient at the jth oc-
- Let $Y_{ij}$ be the number of epileptic seizures experienced by the jth patient at the ith oc-
- Marginal mean model for the Poisson process:

$$\log(\lambda_{ij}) = \alpha_1 + \beta_j + \theta_{ij},$$

- The combined model version, $X_j = \theta_j + \beta_j$.

- The marginal model for the zero-inflated probabilities:

$$\pi_{0ij} = \beta_0 + \beta_1 j.$$