A Weibull-count approach for handling over-dispersed longitudinal data on epileptic patients

Martial Luyts, Geert Molenberghs, Geert Verbeke, Koen Matthijs, Eduardo E. Ribeiro Jr., Clarice G. B. Demétrio, and John Hinde

Interuniversity Institute for Biostatistics and Statistical Bioinformatics, KU Leuven and UHasselt, Leuven and Diepenbeek, Belgium

martial.luyts@kuleuven.be

Interuniversity Institute for Biostatistics and statistical Bioinformatics

Abstract

A random-effects extension of the Weibull-count model of Nakagawa and Osaki (1975) is proposed and applied to the epilepsy dataset. A goodness-of-fit evaluation of the model is provided through a comparison of some-well known hierachical count models, i.e., the Poisson-normal, and double Poisson-normal random effects models. Empirical results show that the proposed extension flexibly fits the data, more specifically, for heavy-tailed, zero-inflated, overdispersed and correlated count data.

Introduction

The epilepsy dataset comes from a randomized, double-blinded, parallel group multicenter study aimed at comparing placebo with a new anti-epileptic drug (AED), in combination with one or two other AEDs. Patients were followed for several weeks during which the number of epileptic seizures experienced in the last week were counted. As result, highly variable longitudinal count data were observed with the presence of extreme values, zero-inflation, and very few observations available at some of the time-points, especially past week 20.

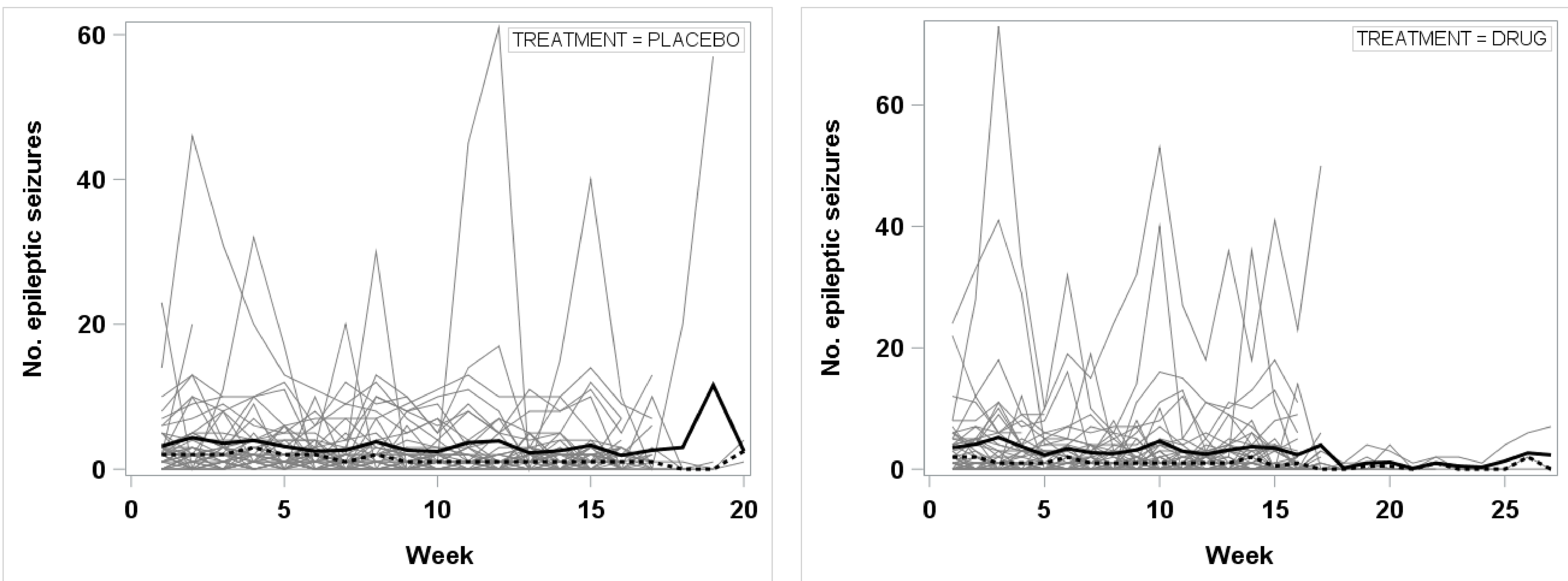

Figure 1: *Epilepsy data (Faught et al., 1996). Subject specific profiles (grey) with corresponding average (solid black) and median (dashed black) profiles of the number of epileptic attacks for every visit, categorized for both treatments.*

Main Objective

- Whether or not the new treatment reduces the number of epileptic seizures.

Materials and Methods

A Weibull-model-based approach for discrete data, here referred as the discrete Weibull (DW) model, first introduced by Nakagawa and Osaki (1975), is considered as underlying model framework. Extensions to hierarchical approaches with dispersion are proposed. In particular, a random-effects extension is considered that takes into account the underlying correlation structure.

Mathematical Section

Let Y_{ij} be the jth discrete outcome measured for cluster (subject) i, $i = 1, \ldots, N$, $j = 1, \ldots, n_i$, and assume to follow a DW distribution (Nakagawa and Osaki, 1975) with parameters $0 < q < 1$ and $\rho > 0$. The probability mass, mean and variance function are respectively given by

$$\mathrm{P}(Y_{ij} = y_{ij}) = q^{y_{ij}^{\rho}} - q^{(y_{ij}+1)^{\rho}}, \quad \mathrm{E}(Y_{ij}) = \mu = \sum_{n=1}^{+\infty} q^{n^{\rho}}, \quad \mathrm{Var}(Y_{ij}) = 2 \cdot \sum_{n=1}^{+\infty} n \cdot q^{n^{\rho}} - \mu - \mu^2,$$

Special cases:

- $\rho = 1$ and $q = 1 - p \rightarrow$ Geometric distribution
  - $q = e^{-\lambda} \rightarrow$ Discrete exponential (DE) distribution (Sato et al., 1999)
- $\rho = 2$ and $q = \theta \rightarrow$ Discrete Rayleigh (DR) distribution (Roy, 2004)
- $\rho \rightarrow +\infty \rightarrow$ DW approaches a Bernoulli distribution with probability q

Characteristics:

To explore the characteristics of the DW model, we compute indexes for dispersion (DI), zero-inflation (ZI) and heavy-tail (HT), which are respectively given by

$$\mathrm{DI} = \frac{\mathrm{Var}(Y_{ij})}{\mathrm{E}(Y_{ij})}, \qquad \mathrm{ZI} = 1 + \frac{\log P(Y_{ij}=0)}{\mathrm{E}(Y_{ij})}, \qquad \mathrm{HT} = \frac{P(Y_i = y_i + 1)}{P(Y_i = y_i)}, \text{ for } y_i \rightarrow \infty.$$

- Over-, under- and equidispersion for, respectively, DI > 1, DI < 1 and DI $= 1$
- Zero-inflation, zero-deflation and no excess of zeros for, respectively, ZI > 0, ZI < 0 and ZI $= 0$
- Heavy-tail distribution for HT $\rightarrow 1$ when $y \rightarrow \infty$

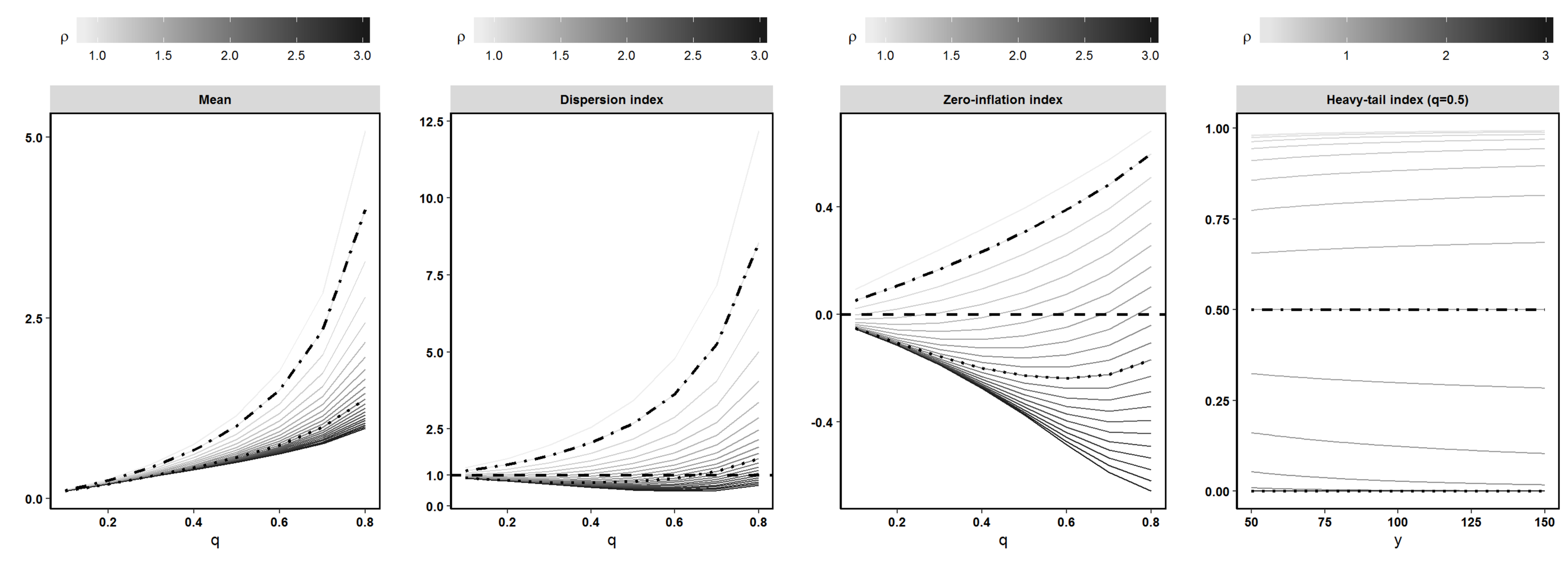

Figure 2: *Characteristic indexes. Dashed, dot dashed and dotted lines represent the Poisson, DE and DR distribution, respectively.*

Regression framework:

$$\ln[-\ln(q_{ij})] = \mathbf{x}_{ij}^{'} \cdot \boldsymbol{\beta} + \mathbf{z}_{ij}^{'} \cdot \mathbf{b}_i \equiv \eta_{ij}, \qquad \mathbf{b}_i \sim \mathrm{N}(\mathbf{0}, D).$$

- Conditional on the random effects, the regression parameter vector $\boldsymbol{\beta}$ can directly be interpreted in terms of the logarithm of the (closed-form) median.
- The random effects vector $\mathbf{b}_i$ here follows a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix D.
- Maximum likelihood principles with numerical integration can be used for estimation.

Analysis

The epilepsy data will be analyzed with the extended DW model, i.e., the discrete Weibull-normal (DWN) model, and compared with the classical log-linear Poisson-normal (PN) and extended double Poisson-normal (DPN) model of Efron (1986).

Let Y_{ij} be the number of epileptic seizures that patient i experiences during week j of the follow-up period, and let t_{ij} be the time-point at which outcome Y_{ij} has been measured, i.e., $t_{ij} = 1, 2, \ldots,$ until at most 27. The following specific choice is made for the linear predictor:

$$\eta_{ij} = \beta_0 + b_i + \beta_0^{'} \cdot T_i + (\beta_1 + \beta_1^{'} \cdot T_i) \cdot t_{ij},$$

where $T_i = 1$ if patient i receives the treatment, and 0 for placebo. Here, $\beta_0^{'}$ and $\beta_1^{'}$ represent differences between treatment and placebo in terms of intercept and slope, respectively. The random intercept b_i is assumed to be normally distributed with mean 0 and variance σ^2, reflecting the between-patient variability within the data.

		PN	DPN*	DWN
Effect	Par.	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
Intercept placebo	β_0	0.8177 (0.1677)	0.8314 (0.1721)	1.4319 (0.2183)
Difference in intercepts	$\beta_0^{'}$	−0.1705 (0.2387)	−0.1582 (0.2451)	−0.2970 (0.3005)
Slope placebo	β_1	−0.0143 (0.0044)	−0.0146 (0.0067)	−0.0297 (0.0098)
Difference in slopes	$\beta_1^{'}$	0.0023 (0.0062)	0.0018 (0.0093)	0.0180 (0.0135)
Ratio of slopes	$1 + \frac{\beta_1^{'}}{\beta_1}$	0.8398 (0.3979)	0.8778 (0.5980)	0.3947 (0.3382)
Std. dev. random effect	σ	1.0755 (0.0857)	1.0458 (0.0875)	1.2658 (0.1063)
	ϕ	−−	0.4355 (0.0169)	−−
	ρ	−−	−−	1.3074 (0.0340)
−2 loglik		6271.9	5652.2	5451.1
AIC		6281.9	5664.2	5463.1

* Over-, under- and equidispersion corresponds to $\phi < 1$, $\phi > 1$ and $\phi = 1$, respectively.

Table 1: Epilepsy dataset. Parameter estimates and standard errors for the (1) Poisson-normal (PN) model, (2) double Poisson-normal (DPN) model, and (3) the discrete Weibull-normal (DWN) model.

Conclusions

- For the epilepsy dataset, DWN is considerably better in terms of likelihood compared to the PN and DPN models.
- DWN model allows inferences directly on the median scale, while a restricted mean scale interpretation is obtained for the PN and DPN model.
- DWN is able to flexibly model highly overdispersed, zero-inflated, heavy-tailed and correlated data, and even underdispersed data.

References

[1] B. Efron. Double exponential families and their use in generalized linear regression. *Journal of the American Statistical Association*, 81(395):709–721, 1986.

[2] E. Faught, B. J. Wilder, R. E. Ramsay, R. A. Reife, L. D. Kramer, G. W. Pledger, and R. M. Karim. Topiramate placebo-controlled dose-ranging trial in refractory partial epilepsy using 200-, 400-, and 600-mg daily dosages. *Neurology*, 46(6):1684–1690, 1996.

[3] T. Nakagawa and S. Osaki. The discrete weibull distribution. *IEEE Transactions on Reliability*, R-24(5):300–301, Dec 1975.

[4] D. Roy. Discrete rayleigh distribution. *IEEE Transactions on Reliability*, 53(2):255–260, June 2004.

[5] H. Sato, M. Ikota, A. Sugimoto, and H. Masuda. A new defect distribution metrology with a consistent discrete exponential formula and its applications. *IEEE Transactions on Semiconductor Manufacturing*, 12(4):409–418, Nov 1999.

Acknowledgements

Financial support from the IAP research network #P7/06 of the Belgian Government (Belgian Science Policy) is gratefully acknowledged. This work was partially supported by CNPq, a Brazilian science funding agency. The research leading to these results has also received funding from KU Leuven GOA project: "New approaches to the social dynamics of long term fertility change".