Modeling the cumulative incidence function of clustered competing risks data: a multinomial GLMM approach

Henrique Aparecido Laureano* Wagner Hugo Bonat*

April 16, 2021

Abstract

Clustered competing risks data are a complex failure time data scheme. Its main characteristics are the cluster structure, which implies a latent within-cluster dependence between its elements, and its multiple variables competing to be the one responsible for the occurrence of an event, the failure. To handle this kind of data, we propose a full likelihood approach, based on a generalized linear mixed model instead a usual complex frailty model. We model the competing causes in the probability scale, in terms of the cumulative incidence function (CIF). A multinomial distribution is assumed for the competing causes and censorship, conditioned on the latent effects. The latent effects are accommodated via a multivariate Gaussian distribution. The CIF is specified as the product of an instantaneous risk level function with a failure time trajectory level function. The estimation procedure is performed through the R package TMB (Template Model Builder), an `C++` based framework with efficient Laplace approximation and automatic differentiation routines. A large simulation study is performed, based on different latent structure formulations. The model presents to be of difficult estimation, with our results converging to a latent structure where the risk and failure time trajectory levels are correlated.

Keywords: Clustered competing risks; Within-cluster dependence; Multinomial generalized linear mixed model (GLMM); TMB: Template Model Builder; Laplace approximation; Automatic differentiation (AD).

*Laboratory of Statistics and Goeinformation, Departament of Statistics, Paraná Federal University, Curitiba, Brazil. E-mail: laureano@ufpr.br

1 Introduction

Regression models are the main statistical tool for investigating the relationship between a response variable and a set of explanatory variables. The class of generalized linear models (GLMs) (Nelder and Wedderburn; 1972) is probably the most popular statistical modelling framework to deal with Gaussian and non-Gaussian outcomes. Despite its flexibility, the GLMs are not suitable for . In general, .

The analysis of is generally performed by the . Besides that, other . Some examples are the . Additionally, .

Although these models are useful in many applications, they are usually limited to analyze independent data. In the case of longitudinal data, it is essential that the regression model take into account the longitudinal and/or grouped data structure. According to longitudinal data are repeated measures evaluated on the same subjects over time, that are potentially correlated. Dependent data can also arises in studies with block designs, spatial and multilevel data. For the analysis of such data several methods have been proposed over the last four decades.

Laird and Ware (1982) proposed the random effects regression models for longitudinal data analysis. Breslow and Clayton (1993) presented the generalized linear mixed models (GLMMs) for the analysis of non-Gaussian outcomes. Masarotto and Varin (2012) developed a class of marginal models for modelling dependence structures in the analysis of longitudinal data, time series and spatial based on Gaussian copula models.

The main goal of this study is to propose the . In this paper, we will investigate the as an alternative to . R (R Core Team; 2021) package `TMB` (Kristensen et al.; 2016).

The main contributions of this article are: (i) introducing the unit gamma distribution into the GLMMs framework; (ii) performing a extensive simulation study to check the properties of the the maximum likelihood estimator to deal with longitudinal continuous bounded outcomes; (iii) applying the proposed model in two data sets from different fields of application; (iv) providing R code and `C++` implementation for the unit gamma mixed models.

The work are organized as follows. Finally, the main contributions of the article are discussed in Section .

2 Unit gamma mixed models

3 Estimation and inference

4 Simulation studies

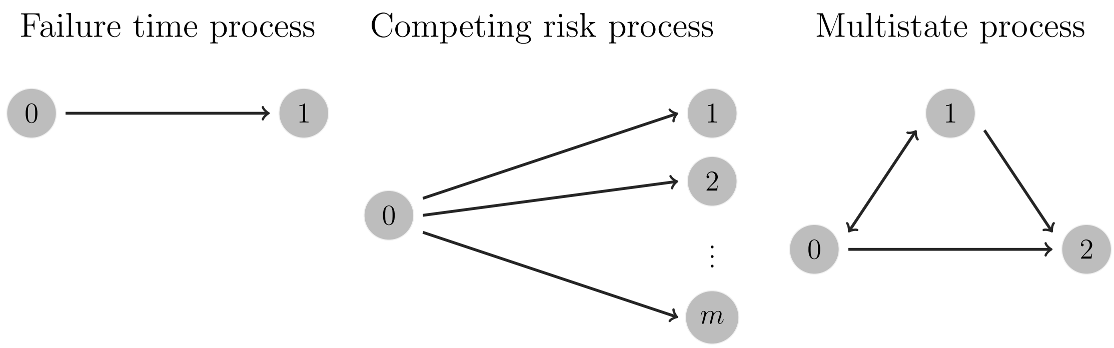

Figure 1: .

Figure 2: .

5 Discussion

Supplementary material

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American statistical Association* **88**(421): 9–25.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. J. and Bell, B. M. (2016). TMB: Automatic Differentiation and Laplace Approximation, *Journal of Statistical Software* **70**(5): 1–21.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**(4): 963–974.

Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression, *Electronic Journal of Statistics* **6**(1): 1517–1549.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models, *Journal of the Royal Statistical Society, Series A* **135**(3): 370–384.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/.