A multinomial generalized linear mixed model for clustered competing risks data

Henrique Aparecido Laureano* Wagner Hugo Bonat*

August 17, 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 data; Within-cluster dependence; Multinomial generalized linear mixed model (GLMM); TMB: Template Model Builder; Laplace approximation; Automatic differentiation (AD).

*Laboratório de Estatística e Geoinformação, Departamento de Estatística, Universidade Federal do Paraná, Curitiba, Brasil. E-mail: laureano@ufpr.br

1 Introduction

Competing risks data, and more generally failure time data, can be modeled in two possible scales: the hazard and the probability scale, with the former being the most popular. The modeling object is the survival experience of the time-to-event data. A competing risks process can be seen as the multivariate extension of a failure time process, having multiple causes competing to be the one responsible for the desired event occurrence, properly, a failure. In Figure 1 a visual aid is provided considering m competing causes.

Failure time data is the branch of Statistics responsible to handle random variables describing the time until the occurrence of an event, a failure (Kalbfleisch and Prentice; 2002; Hougaard; 2000). The time until a failure is called survival experience, and is the modeling object. To accommodate the number of possible causes for a failure there is the competing risks data scheme, described in Figure 1 and the focus of this work. More specifically, its clustered version i.e., with groups of elements sharing some non-observed latent dependence structure.

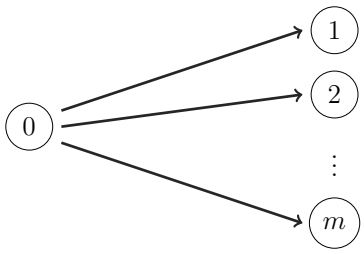

Figure 1: Illustration of competing risks process.

When this framework is applied in real-world situations, we have to be able to handle with the nonoccurrence of the desired event, by any of the competing causes, for, let us say, *logistic reasons* (short-time study and outside scope causes are some examples). This, generally noninformative, nonoccurrence of the event is called censorship.

In its simplest form, it is assumed that all subjects under study are independent. However, if the subjects are structured in a related fashion the independence assumption is unrealistic. When the subjects are organized in clusters (a family, e.g.), the nonindependence is accommodated in terms of a latent/random-effect, shared by all elements of that cluster. This idea opens space to what is called *family studies*. In family studies, the goal is to accommodate and try to understand the relationship between the family elements. In other words, how the occurrence of an event in a subject affects the survival experience for the same or similar event in its familiars.

The survival experiences is usually modeled in the hazard (failure rate) scale, and with the latent within-cluster dependence accommodation we have a frailty model (Clayton; 1978; Valpel et al.; 1979; Liang et al.; 1995; Petersen; 1998). The use of frailty models implies in complicated likelihood functions and inference routines done via elaborated

and slow EM algorithms (Nielsen et al.; 1992; Klein; 1992) or inefficient MCMC schemes (Hougaard; 2000). With multiple survival experiences, the general idea is the same but with even more elaborated likelihoods (Prentice et al.; 1978; Therneau and Grambsch; 2000) or instead with the of mixture model approaches (Larson and Dinse; 1985; Kuk; 1992).

When in the hazard scale, the interpretations are in terms of hazard rates. A less usual scale but with a more appealing interpretation, is to model the survival experiences in the probability scale. For competing risks data, the work on the probability scale is done by means of the cumulative incidence function (CIF) (Andersen et al.; 2012), with the main modeling approach being the subdistribution (Fine and Gray; 1999).

For clustered competing risks data there are some available options but with a lack of predominance. The options vary in terms of likelihood specification, with its majority being designed for bivariate CIFs, where increasing the CIF's dimension is a limitation. Some of the existing options are (i) nonparametric approaches (Cheng et al.; 2007, 2009); (ii) linear transformation models (Fine; 1999; Gerds et al.; 2012); (iii) semiparametric approaches based on composite likelihoods (Shih and Albert; 2009; Cederkvist et al.; 2019), estimating equations (Scheike and Sun; 2012; Cheng and Fine; 2012), copulas (Scheike et al.; 2010), and mixtures (Naskar et al.; 2005; Shi et al.; 2013).

Besides the interpretation, by modeling the CIF it is possible to specify complex within-cluster dependence structures. We follow Cederkvist et al. (2019) and work with a CIF specification based on its decomposition in instantaneous risk and failure time trajectory functions, with both being cluster-specifics and possible correlated. As a modeling framework, we use a generalized linear mixed model (GLMM) specification.

The class of generalized linear models (GLMs) (Nelder and Wedderburn; 1972) is probably the most popular statistical modelling framework. Despite its flexibility, the GLMs are not suitable for dependent data. For the analysis of such data, Laird and Ware (1982) proposed the random effects regression models for longitudinal/repeated-measures data analysis, and Breslow and Clayton (1993) presented the GLMMs for the analysis of non-Gaussian outcomes. In this framework, we can accommodate all competing causes of failure and censorship under a multinomial probability distribution. The within-cluster dependence is accommodated via a multivariate normal distribution, and the cause-specific CIFs via the model's link function. The main goal of this study is to propose a GLMM approach to handle clustered competing risks data with a flexible within-cluster dependence structure. The model specification and the inferential routine are much simpler than the usually used approaches, increasing the practical relevance of our framework. The estimation and inference is made through the efficient computational resources of the R (R Core Team; 2021) package `TMB` (Kristensen et al.; 2016). The latent effects are handled out by means of an efficient Laplace approximation and automatic differentiation.

The main contributions of this article are: (i) introducing the modeling of cause/cluster-specific CIFs of clustered competing risks data into an efficient implementation of the GLMMs framework; (ii) performing a extensive simulation study to check the properties of the maximum likelihood estimator to learn the cause-specific CIF forms and the feasibility of the within-cluster dependence structure.; (iii) providing R code and `C++` implementation for the used GLMMs.

The work is organized as follows. Section 2 presents the CIF specification and the multinomial GLMM. Section 3 presents the estimation and inferential routines. Section 4 presents the performed simulation studies to check the model viability. Finally, the main contributions of the article are discussed in Section 5.

2 Model

2.1 Cluster-specific cumulative incidence function (CIF)

Consider that the observed follow-up time of a subject is given by $T = \min(T^*, C)$, where T^* denote the failure time and C denote the censoring time. Given the possible covariates x, for a cause-specific of failure k, the cumulative incidence function (CIF) is defined as

$$\begin{aligned} F_k(t \mid x) = \mathbb{P}[T \leq t,\ K = k \mid x] &= \int_0^t f_k(z \mid x)\ \mathrm{d}z \\ &= \int_0^t \lambda_k(z \mid x)\ S(z \mid x)\ \mathrm{d}z, \quad t > 0, \quad k = 1,\ \ldots,\ K, \end{aligned}$$

where $f_k(t \mid x)$ is the (sub)density for the time to a type k failure. The subdensity is composed by the cause-specific hazard function or process $\lambda_k(t \mid x)$, representing the instantaneous rate for failures of type k at time t given x and all other failure types (competing causes). If we sum up all cause-specific hazard functions we get the overall hazard function $\lambda(t \mid x)$. From the overall hazard function we arrive in the overall survival function,

$$S(t \mid x) = \mathbb{P}[T > t \mid x] = \exp\left\{-\int_0^t \lambda(z \mid x)\ \mathrm{d}z\right\},$$

the second function that compose the subdensity $f_k(t \mid x)$. A comprehensive reference for all these definitions is the book of Kalbfleisch and Prentice (2002).

To take into consideration our clustered/family structure. We use the same CIF specification of Cederkvist et al. (2019). For two competing causes of failure, the cause-specific CIFs are specified in the following manner

$$F_k(t \mid x,\ u_1,\ u_2,\ \eta_k) = \underbrace{\pi_k(x,\ u_1,\ u_2)}_{\substack{\text{cluster-specific} \\ \text{risk level}}} \times \underbrace{\Phi[w_k g(t) - x\gamma_k - \eta_k]}_{\substack{\text{cluster-specific} \\ \text{failure time trajectory}}}, \quad t > 0, \quad k = 1,\ 2, \quad (1)$$

i.e., as the product of a cluster-specific risk level with a cluster-specific failure time trajectory, resulting in a cluster-specific CIF. What makes these components cluster-specific are $u = \{u_1,\ u_2\}$ and $\eta = \{\eta_1,\ \eta_2\}$, Gaussian distributed latent effects with zero mean and potentially correlated i.e.,

$$\begin{bmatrix} u_1 \\ u_2 \\ \eta_1 \\ \eta_2 \end{bmatrix} \sim \begin{array}{c}\text{Multivariate} \\ \text{Normal}\end{array} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_1} & \text{cov}(u_1,\ u_2) & \text{cov}(u_1,\ \eta_1) & \text{cov}(u_1,\ \eta_2) \\ & \sigma^2_{u_2} & \text{cov}(u_2,\ \eta_1) & \text{cov}(u_2,\ \eta_2) \\ & & \sigma^2_{\eta_1} & \text{cov}(\eta_1,\ \eta_2) \\ & & & \sigma^2_{\eta_2} \end{bmatrix} \right).$$

The cluster-specific survival function is given by $S(t \mid x,\ u,\ \eta) = 1 - F_1(t \mid x,\ u,\ \eta_1) - F_2(t \mid x,\ u,\ \eta_2)$. The second component of Equation 1, the cluster-specific failure time trajectory

$$\Phi[w_k g(t) - x\gamma_k - \eta_k], \quad t > 0, \quad k = 1,\ 2,$$

where $\Phi(\cdot)$ is the cumulative distribution function of a standard Gaussian distribution. With regard to the function $g(t)$, it plays a crucial role since the CIF separation in Equation 1 is only possible with it. It is used a time t transformation given by

$$g(t) = \text{arctanh}\left(\frac{t - \delta/2}{\delta/2}\right), \quad t \in (0,\ \delta), \quad g(t) \in (-\infty,\ \infty),$$

where δ depends on the data and cannot exceed the maximum observed follow-up time τ i.e., $\delta \leq \tau$. With this Fisher-based transformation the value of the cluster-specific failure time trajectory is equal 1, at time δ. Consequently, $F_k(\delta \mid x,\ u,\ \eta_k) = \pi_k(x \mid u)$ and we can interpret $\pi_1(x \mid u)$ and $\pi_2(x \mid u)$ as the cause-specific cluster-specific risk levels, at time δ.

The cluster-specific risk levels are modeled by a multinomial logistic regression model with latent effects i.e.,

$$\pi_k(x, u) = \frac{\exp\{x\beta_k + u_k\}}{1 + \exp\{x\beta_1 + u_1\} + \exp\{x\beta_2 + u_2\}}, \quad k = 1,\ 2, \tag{2}$$

where the β_k's are the coefficients responsible for quantifying the impact of the covariates in the cause-specific risk levels. For individuals from the same cluster/family, at the same time point, the β_ks have the well-known odds ratio interpretation.

A direct understanding of all coefficients/parameters of Equation 1 can be reached via the illustrations in Figure 2. To really understand what is going on, we simplify the model. We still consider just two competing causes but without covariates and we plot just the cluster-specific CIF of one failure cause. In the top-plots of Figure 2 we see that the β's are also related with the curve's maximum value i.e., bigger the β, highest the CIF will be.

The γ_k's are the coefficients responsible for quantifying the impact of the covariates in the cause-specific failure time trajectories i.e., the shape of the cumulative incidence. We see that the γ's are also related with an idea of midpoint and consequently, growth speed. The fact that γ_k enters negatively in the cluster-specific failure time trajectory makes that a negative value causes an advance towards the curve, whereas a positive value causes a delay. Last but not least, the w's. With negative values, we have a decreasing curve and with positive values an increasing curve i.e., we are interested only on the positive side.

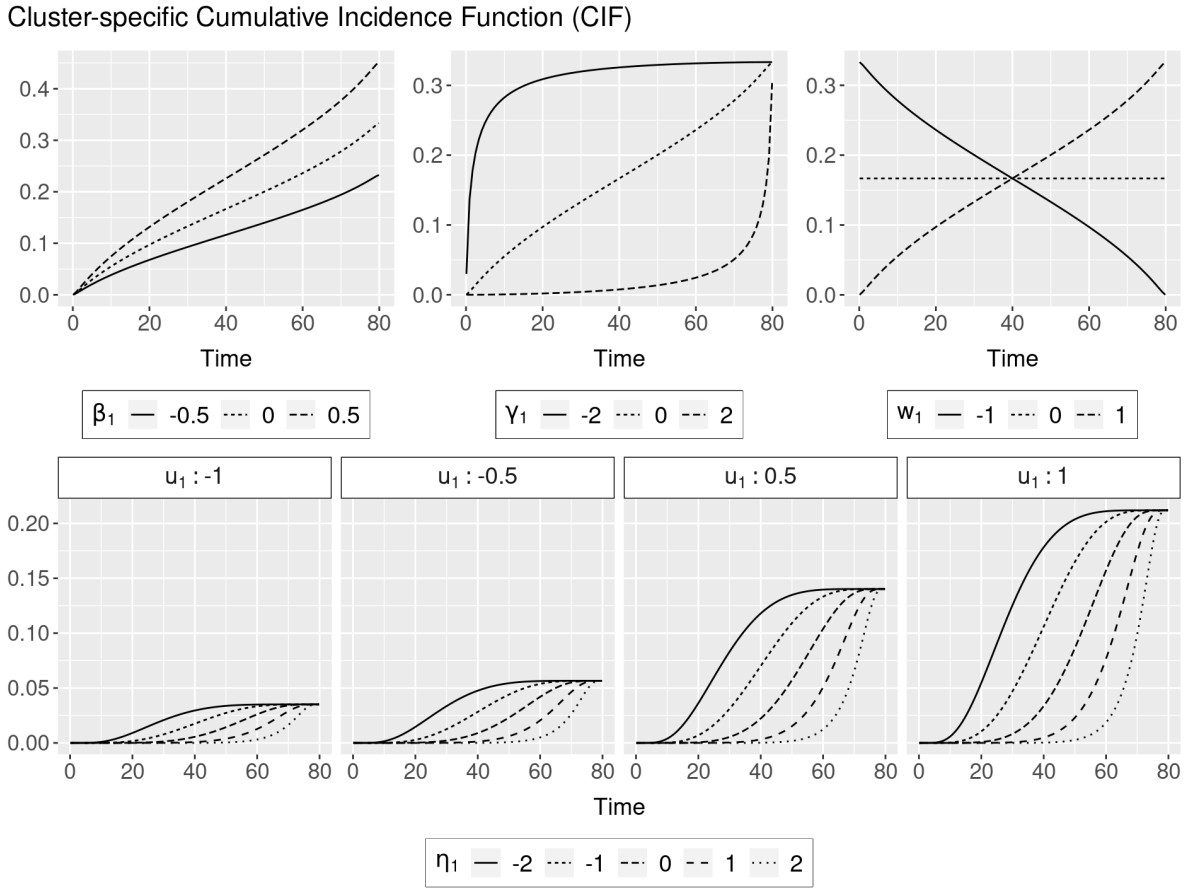

Figure 2: Curve behaviors for different parameter settings, showing then the corresponding parameter effects in a cluster-specific cumulative incidence function (CIF).

Remains to talk about the within-cluster dependence induced by the latent effects in u and η. Unfortunately, they do not have an easy interpretation. To help in the discussion, the bottom-plots on Figure 2 illustrates the cluster-specific CIF for a given failure cause in a model without covariates, let us call it failure cause 1 (in total we have two).

The latent effects u_1 and u_2 always appear together in the cluster-specific risk level, as consequency they have a joint effect on the cumulative incidence of both causes. As we can see in Figure 2, an increase in u_k will increase the risk of failure from cause k. The interpretation of $\text{cov}(\eta_1,\ \eta_2)$ and $\text{cov}(u_1,\ u_2)$ is straightforward, and those values are in most of the cases positive, as said in Cederkvist et al. (2019). With regard to $\text{cov}(u_k,\ \eta_k)$, negative values are the common situation. A negative correlation between

η_k and u_k imply that when η_k decreases, u_k increases and conversely when η_K increases, u_k decreases. In other words, an increased risk level is reached quickly and a decreased risk level is reached later, respectively.

Practical situations with a positive within-cause correlation are hard to find i.e., where an increased risk level is associated with a late onset and vice versa. However, a positive cross-cause correlation between η and u sounds much more realistic i.e., where late onset of one failure cause is associated with a high absolute risk of another failure cause.

The latent effects $\{u_k,\ \eta_k\}$ are assumed independent across clusters and shared by individuals within the same cluster/family.

2.2 Model specification

The multiGLMM for clustered competing risks data is specified in the following hierarchical fashion. By simplicity, we focus on two competing causes of failure but an extension is straightforward.

For two competing causes of failure, a subject i, in the cluster/family j, in time t, we have

$$y_{ijt} \mid \{u_{1j},\ u_{2j},\ \eta_{1j},\ \eta_{2j}\} \sim \text{Multinomial}(p_{1ijt},\ p_{2ijt},\ p_{3ijt})$$

$$\begin{bmatrix} u_1 \\ u_2 \\ \eta_1 \\ \eta_2 \end{bmatrix} \sim \text{MN}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{u_1} & \text{cov}(u_1,\ u_2) & \text{cov}(u_1,\ \eta_1) & \text{cov}(u_1,\ \eta_2) \\ & \sigma^2_{u_2} & \text{cov}(u_2,\ \eta_1) & \text{cov}(u_2,\ \eta_2) \\ & & \sigma^2_{\eta_1} & \text{cov}(\eta_1,\ \eta_2) \\ & & & \sigma^2_{\eta_2} \end{bmatrix} \right)$$

$$\begin{aligned} p_{kijt} &= \frac{\partial}{\partial t} F_k(t \mid x,\ u_1,\ u_2,\ \eta_k) \\ &= \frac{\exp\{x_{kij}\beta_k + u_{kj}\}}{1 + \sum_{m=1}^{K-1} \exp\{x_{mij}\beta_m + u_{mj}\}} \\ &\times w_k \frac{\delta}{2\delta t - 2t^2}\ \phi\left(w_k \text{arctanh}\left(\frac{t - \delta/2}{\delta/2} \right) - x_{kij}\gamma_k - \eta_{kj} \right), \end{aligned} \tag{3}$$

$k = 1,\ 2.$

The probabilities are given by the derivative w.r.t. time t of the cluster-specific CIF. The choice of a multinomial logistic regression model ensures that the sum of the predicted cause-specific CIFs does not exceed 1.

Considering two competing causes of failure, we have a multinomial with three classes. The third class exists to handle the censorship and its probability is given by the complementary to reach 1. This framework in Equation 3 results in what we call multiGLMM, a multinomial GLMM to handle the CIF of clustered competing risks data. For a random

sample, the corresponding marginal likelihood function in given by

$$
\begin{aligned}
L(\theta \; ; \; y) &= \prod_{j=1}^{J} \int_{\Re^4} \pi(y_j \mid r_j) \times \pi(r_j) \, \mathrm{d}r_j \\
&= \prod_{j=1}^{J} \int_{\Re^4} \underbrace{\left\{ \prod_{i=1}^{n_j} \prod_{t=1}^{n_{ij}} \left(\frac{(\sum_{k=1}^{K} y_{kijt})!}{y_{1ijt}! \; y_{2ijt}! \; y_{3ijt}!} \prod_{k=1}^{K} p_{kijt}^{y_{kijt}} \right) \right\}}_{\text{fixed effect component}} \times \\
&\qquad \underbrace{(2\pi)^{-2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} r_j^\top \Sigma^{-1} r_j \right\}}_{\text{latent effect component}} \mathrm{d}r_j \\
&= \prod_{j=1}^{J} \int_{\Re^4} \underbrace{\left\{ \prod_{i=1}^{n_j} \prod_{t=1}^{n_{ij}} \prod_{k=1}^{K} p_{kijt}^{y_{kijt}} \right\}}_{\text{fixed effect}} \underbrace{(2\pi)^{-2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} r_j^\top \Sigma^{-1} r_j \right\}}_{\text{latent effect component}} \mathrm{d}r_j, \qquad (4)
\end{aligned}
$$

where $\theta = [\beta \; \gamma \; w \; \sigma^2 \; \rho]^\top$ is the parameters vector to be maximized. In our framework, a subject can fail from just one competing cause or get censor, at a given time. Thus, the fraction of factorials in the fixed effect component is made only by 0's and 1's. Finally, returning the value 1. The matrix Σ is the variance-covariance matrix, which parameters are given by σ^2 and ρ.

Now, Equation 4 in words. To each cluster/family j we have a product of two components. The fixed effect component, given by a multinomial distribution with its probabilities specified through the cluster-specific CIF (Equation 1) and, the latent effect component, given by a multivariate Gaussian distribution.

To each subject i that composes a cluster j we have its specific fixed effects contribution. The likelihood in Equation 4 is the most general as possible, allowing for repeated measures to each subject. Since all subjects of a given cluster shares the same latent effect, we have just one latent effect contribution multiplying the product of fixed effect contributions. As we do not observe the latent effect variables, r_j, we integrate out in it. With two competing causes of failure, we have four latent effects (a multivariate Gaussian distribution in four dimensions). Consequently, for each cluster, we approximate an integral in four dimensions. The product of these approximated integrals results in the called marginal likelihood, to be maximized in θ.

3 Estimation and inference

Our goal is to estimate the parameter vector $\theta = [\beta \; \Sigma]^\top$ of a mean structure, as in **??**. Besides the role of emphasizing the fact that μ is a function of θ, and that we want to estimate θ, the likelihood function ties the probability densities i.e., the likelihood is the product of the product of probability densities, to each subject i. Since Y_i are mutually

independent, the likelihood for θ can be written as

$$L(\theta \mid y, u) = \prod_{i=1}^{I} \prod_{j=1}^{n_i} f(y_{ij} \mid u_i, \beta, \Sigma)\ f(u_i \mid \Sigma). \tag{5}$$

From standard probability theory is easy to see that in the right-hand side (r.h.s.) we have a joint density, consequently, Equation 5 represents what is called a full or a joint likelihood function. What makes problematic working with this joint likelihood is that we do not have all the necessary information to just maximize it and get the desired parameter estimates. The latent effect u is *latent* i.e., we do not observe it. To handle this we have basically two available paths.

3.1 Marginalization: Laplace approximation

We have a joint density where one of the random variables is not observed, but we are not interested in it, only in the variance parameters inherent in it. Again, from standard probability theory, if we have a joint density we can just integrate out the undesired variable resulting in

$$\begin{aligned} L(\theta \mid y) &= \prod_{i=1}^{I} \int_{\mathcal{R}^{u_i}} \left[\prod_{j=1}^{n_i} f(y_{ij} \mid u_i, \beta, \Sigma)\ f(u_i \mid \Sigma) \right] \mathrm{d}u_i \\ &= \prod_{i=1}^{I} \int_{\mathcal{R}^{u_i}} f(y_i, u_i \mid \theta)\ \mathrm{d}u_i, \end{aligned} \tag{6}$$

a marginal density that keeps the parameters Σ of the integrated variable.

To us, the better option consists in take advantage of the exponential family structure together with the fact that we are dealing with Gaussian latent effects. These ideas converge to an adaptive Gaussian quadrature with one integration point, also called as *Laplace approximation* (Molenberghs and Verbeke; 2005; Shun and McCullagh; 1995; Tierney and Kadane; 1986; Wood; 2015).

With an integral that is analytically intractable, we may approximate it to obtain a tractable closed-form expression allowing the numerical maximization of the resulting marginal likelihood function Bonat and Ribeiro Jr (2016). The Laplace approximation has been designed to approximate integrals in the form

$$\int_{\mathcal{R}^{u_i}} \exp\{Q(u_i)\}\mathrm{d}u_i \approx (2\pi)^{n_u/2}\ |Q''(\hat{u}_i)|^{-1/2}\ \exp\{Q(\hat{u}_i)\}, \tag{7}$$

where $Q(u_i)$ is a known, unimodal bounded function, and $\hat{u}_i$ is the value for which $Q(u_i)$ is maximized.

The second advantage of a Laplace approximation approach in a GLMM is the expo-

nential family structure. In a usual GLMM the response follows a one-parameter exponential family distribution that can be written as

$$f(y_i \mid u_i, \theta) = \exp\left\{y_i^\top (x_i\beta + z_i u_i) - 1_i^\top b(x_i\beta + z_i u_i) + 1_i^\top c(y_i)\right\},$$

where $b(\cdot)$ and $c(\cdot)$ are known functions.

This general and easy to compute expression, together with a (multivariate) Gaussian distribution, highlights the convenience of the Laplace method. The $Q(u_i)$ function to be maximized can be expressed as

$$\begin{aligned} Q(u_i) = {} & y_i^\top (x_i\beta + z_i u_i) - 1_i^\top b(x_i\beta + z_i u_i) + 1_i^\top c(y_i) \\ & - \frac{n_u}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}u_i^\top \Sigma^{-1}\ u_i. \end{aligned} \tag{8}$$

The approximation in Equation 7 requires the maximum $\hat{u}_i$ of the function $Q(u_i)$. As we assume a Gaussian distribution with a known mean for the latent effects, we have the perfect initial guess for a Hessian-based maximization method, as the Newton-Raphson (NR) algorithm.

The NR method consists of an iterative scheme as follows:

$$u_i^{(k+1)} = u_i^{(k)} - Q''(u_i^{(k)})^{-1}\ Q'(u_i^{(k)}), \quad k = 0,\ 1,\ \ldots$$

until convergence, which gives $\hat{u}_i$. At this stage, all parameters θ are considered known. patrao presents the generic expressions for the derivatives required by the NR method, given by the following:

$$\begin{aligned} Q'(u_i^{(k)}) &= \{y_i - b'(x_i\beta + z_i u_i^{(k)})\}^\top - {u_i^{(k)}}^\top \Sigma^{-1}, \\ Q''(u_i^{(k)}) &= -\text{diag}\{b''(x_i\beta + z_i u_i^{(k)})\} - \Sigma^{-1}. \end{aligned}$$

We have the initial guesses at $k = 0$.

Finally, the marginal log-likelihood function returned by the Laplace approximation, to each individual or unit under study i, is as follows:

$$\begin{aligned} l(\theta \mid y_i) = \log L(\theta \mid y_i) = {} & \frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left|\text{diag}\{b''(x_i\beta + z_i\hat{u}_i)\} + \Sigma^{-1}\right| \\ & + y_i^\top (x_i\beta + z_i\hat{u}_i) - 1_i^\top b(x_i\beta + z_i\hat{u}_i) + 1_i^\top c(y_i) \\ & - \frac{n_u}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}\hat{u}_i^\top \Sigma^{-1}\ \hat{u}_i, \end{aligned}$$

that can now be numerically maximized over the model parameters $\theta = [\beta\ \Sigma]^\top$.

4 Simulation studies

5 Discussion

Supplementary material

References

Andersen, P. K., Geskus, R. B., de Witte, T. and Putter, H. (2012). Competing risks in epidemiology: possibilities and pitfalls, *International Journal of Epidemiology* **31**(1): 861–870.

Bonat, W. H. and Ribeiro Jr, P. J. (2016). Practical likelihood analysis for spatial generalized linear mixed models, *Environmetrics* **27**(1): 83–89.

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.

Cederkvist, L., Holst, K. K., Andersen, K. K. and Scheike, T. H. (2019). Modeling the cumulative incidence function of multivariate competing risks data allowing for within-cluster dependence of risk and timing, *Biostatistics* **20**(2): 199–217.

Cheng, Y. and Fine, J. P. (2012). Cumulative incidence association models for bivariate competing risks data, *Journal of the Royal Statistical Society, Series B (Methodological)* **74**(2): 183–202.

Cheng, Y., Fine, J. P. and Kosorok, M. R. J. (2007). Nonparametric Association Analysis of Bivariate Competing-Risks Data, *Journal of the American Statistical Association* **102**(480): 1407–1415.

Cheng, Y., Fine, J. P. and Kosorok, M. R. J. (2009). Nonparametric Association Analysis of Exchangeable Clustered Competing Risks Data, *Biometrics* **65**(1): 385–393.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial rendency in chronic disease incidence, *Biometrika* **65**(1): 141–151.

Fine, J. P. (1999). Analysing competing risks data with transformation models, *Journal of the Royal Statistical Society, Series B (Methodological)* **61**(4): 817–830.

Fine, J. P. and Gray, R. J. (1999). A proportional hazards models for the subdistribution of a competing risk, *Journal of the American Statistical Association* **94**(446): 496–509.

Gerds, T. A., Scheike, T. H. and Andersen, P. K. (2012). Absolute risk regression for competing risks: interpretation, link functions and prediction, *Statistics in Medicine* **31**(29): 3921–3930.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Springer-Verlag, New York.

Kalbfleisch, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, second Edition edn, John Wiley & Sons, Inc., Hoboken, New Jersey.

Klein, J. P. (1992). Semiparametric estimation of random effects using cox model based on the em algorithm, *Biometrics* **48**(1): 795–806.

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.

Kuk, A. Y. C. (1992). A semiparametric mixture model for the analysis of competing risks data, *Australian Journal of Statistics* **34**(2): 169–180.

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

Larson, M. G. and Dinse, G. E. (1985). A Mixture Model for the Regression Analysis of Competing Risks Data, *Journal of the Royal Statistical Society, Series C (Applied Statistics)* **34**(3): 201–211.

Liang, K. Y., Self, S., Bandeen-Roche, K. J. and Zeger, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data, *Lifetime Data Analysis* **1**(1): 403–415.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*, Springer, New York.

Naskar, M., Das, K. and Ibrahim, J. G. (2005). A Semiparametric Mixture Model for Analyzing Clustered Competing Risks Data, *Biometrics* **61**(3): 729–737.

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

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. A. (1992). A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models, *Scandinavian Journal of Statistics* **19**(1): 25–43.

Petersen, J. H. (1998). An Additive Frailty Model for Correlated Life Times, *Biometrics* **54**(1): 646–661.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks, *Biometrics* **1**(1): 541–554.

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

Scheike, T. and Sun, Y. (2012). On cross-odds ratio for multivariate competing risks data, *Biostatistics* **13**(4): 680–694.

Scheike, T., Zhang, Y. S. M. and Jensen, T. K. (2010). A semiparametric random effects model for multivariate competing risks, *Biometrika* **97**(1): 133–145.

Shi, H., Cheng, Y. and Jeong, J. H. (2013). Constrained parametric model for simultaneous inference of two cumulative incidence functions, *Biometrical Journal* **55**(1): 82–96.

Shih, J. H. and Albert, P. S. (2009). Modeling Familial Association of Ages at Onset of Disease in the Presence of Competing Risk, *Biometrics* **66**(4): 1012–1023.

Shun, Z. and McCullagh, P. (1995). Laplace approximation of high dimentional integrals, *Journal of the Royal Statistical Society, Series B (Methodological)* **57**(4): 749–760.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*, Springer-Verlag, New York.

Tierney, L. and Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association* **81**(393): 82–86.

Valpel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography* **16**(1): 439–454.

Wood, S. N. (2015). *Core Statistics*, Institute of Mathematical Statistics, Textbooks, IMS.