Frailty Model for Parametric Analysis of Bivariate Right Censored Data with Competing Risks

Ankit Sinha

Supervisor - Dr.Sudipta Das

July 10, 2024

Abstract

This project delves into the parametric analysis of bivariate right censored data through the application of frailty models. Frailty, which captures the unobserved heterogeneity within a population, is a key aspect of our analysis. Differing from conventional methods, this study does not include any covariate effects. We have generated bivariate right censored data, specifically using gamma frailty models with a Weibull baseline hazard. This approach enables a comprehensive examination of the underlying frailty and its influence on the competing risks present in the data. The results offer valuable insights into the nature and consequences of unobserved heterogeneity in the context of bivariate right censored data.

1 Introduction

Let T_1 and T_2 denote the times of failure for two individuals in a pair, with J_1 and J_2 representing the causes of their respective failures.We assume that J_k (where $k = 1,2$) can take values from the set $\{1, 2, \ldots, L_k\}$, indicating that the individuals may be exposed to different sets of competing risks. However, in certain models, we consider the same set of competing risks for both individuals. Let X_1 and X_2 denote the censoring times for two individuals, assumed to be independent of their respective failure times and not influenced by any common parameters. Thus, our observed data comprises (X_1, X_2, J_1, J_2), where J_1 or J_2 equals 0 if the observation is censored. The sub-distribution function for the k-th individual due to cause j at time t_k is given by $F_j^{(k)}(t_k;\theta) = P(T_k \leq t_k, J_k = j;\theta) = \int_0^{t_k} f_j^{(k)}(u;\theta)du$, where $t_k > 0$, $j = 1, \ldots, L_k$, and

$k = 1, 2$. Here, $f_j^{(k)}(u;\theta)$ represents the sub-density function for the k-th individual due to cause j, and θ denotes the associated parameter vector. The cumulative distribution function for the k-th individual at time t_k is $F^{(k)}(t_k;\theta) = P(T_k \leq t_k;\theta) = \sum_{j=1}^{L_k} F_j^{(k)}(t_k;\theta)$. The survival function is $S^{(k)}(t_k;\theta) = P(T_k > t_k;\theta) = 1 - F^{(k)}(t_k;\theta)$. In the scenario of bivariate failure time data with competing risks, the primary focus is on the joint sub-distribution function $F_{j1j2}(t1, t2;\theta) = P(T1 \leq t1, T2 \leq t2, J1 = j1, J2 = j2;\theta)$. This function encapsulates the probability that both individuals experience failures of type $j1$ and $j2$ at times $t1$ and $t2$, respectively. Modeling such dependencies involves frailty variables, often characterized by Gamma distributions, and the specification of cause-specific baseline hazard functions.

The j-th cause-specific hazard function for the k-th individual at time t_k is defined as $\lambda_j^{(k)}(t_k;\theta) = \lim_{h\to 0+} \frac{P(t_k < T_k < t_k + h, J_k = j;\theta)}{h} = \frac{f_j^{(k)}(t_k;\theta)}{S_{-j}^{(k)}(t_k;\theta)}$, where $t_k > 0$, $j = 1, \ldots, L_k$, and $k = 1, 2$. The correlation in bivariate failure time data with competing risks is addressed using frailty variables. Generally, let $\epsilon^{(k)} = (\epsilon_1^{(k)}, \ldots, \epsilon_{L_k}^{(k)})$ represent the frailty variables associated with different failure types for the k-th individual, where $k = 1, 2$. It's important to note that the parameter vector θ encompasses two sets of parameters: one for the baseline cause-specific hazard functions, denoted by λ, and the other for the frailty models, denoted by ϕ, such that $\theta = (\lambda, \phi)$. We adopt a general multiplicative model for the conditional cause-specific hazard function of the k-th individual, given the frailty vector $\epsilon^{(k)}$, as $\lambda_j^{(k)}(t_k;\lambda \mid \epsilon^{(k)}) = \lambda_{0j}^{(k)}(t_k;\lambda)\epsilon^{(k)}$, for $t_k > 0$, where $\lambda_{0j}^{(k)}(t_k;\lambda)$ represents the cause-specific baseline hazard function for the k-th individual. The corresponding cumulative cause-specific baseline hazard function for the k-th individual is $H_{0j}^{(k)}(t_k;\lambda) = \int_0^{t_k} \lambda_{0j}^{(k)}(u;\lambda)du$.

Furthermore, we consider parametric models for the cause-specific baseline hazard functions $\lambda_{0j}^{(k)}(t_k;\lambda)$ with an associated parameter vector $\lambda \in \Lambda$. These parametric forms of cause-specific baseline hazard functions $\lambda_{0j}^{(k)}(t_k;\lambda)$, where $k = 1, 2$, are defined to be identifiable within $\lambda_{0j}^{(k)}(t_k;\lambda) = \lambda_{0j}^{(k)}(t_k;\tilde{\lambda})$ for all $t_k > 0$, then $\lambda = \tilde{\lambda}$.

1.1 Shared Gamma Frailty Model

We assume that the shared frailty variable ν follows a Gamma distribution with parameters $(\frac{1}{\alpha^2}, \frac{1}{\alpha^2})$, for some $\alpha > 0$. The density function of the Gamma distribution with shape parameter $\frac{1}{\alpha^2}$ and scale parameter α^2 is given by:

$$g(\nu;\alpha) = \frac{1}{\Gamma\left(\frac{1}{\alpha^2}\right)\alpha^{\frac{2}{\alpha^2}}} \nu^{\frac{1}{\alpha^2}-1} e^{-\frac{\nu}{\alpha^2}},$$

where $\nu > 0$ and $\alpha > 0$.

For example, a shared Gamma frailty model can be used to model cancer occurrence times in a father and son, with competing risks referring to different cancer sites, due to shared genetic and environmental factors.

For the shared Gamma frailty model, conditional on the shared frailty ν, we have:

$$S^{(k)}(t_k;\nu \mid \nu) = \exp\left(-\sum_{j=1}^{L_k} H_{0j}^{(k)}(t_k;\nu)\nu\right) = \exp\left(-\nu H_0^{(k)}(t_k;\nu)\right),$$

where $H_0(k)(t_k;\nu) = \sum_{j=1}^{L_k} H_{0j}(k)(t_k;\nu)$, and

$$F_j^{(k)}(t_k;\nu \mid \nu) = \int_0^{t_k} h_{0j(k)}(u_k;\nu)\nu \exp\left(-\nu H_0^{(k)}(u_k;\nu)\right) du_k,$$

for $t_k > 0$, $j = 1, \ldots, L_k$, and $k = 1, 2$.

The unconditional joint sub-distribution function is obtained by integrating with respect to the Gamma density of ν as:

$$F_{j_1 j_2}(t_1, t_2; \nu, \alpha) = \int_0^{t_1} \int_0^{t_2} \left(1+\alpha^2\right) \prod_{k=1}^{2} h_{0,j_k}^{(k)}(u_k;\nu) du_2 du_1 \left(1 + \alpha^2 \sum_{k=1}^{2} H_0^{(k)}(u_k;\nu)\right)^{-2-1/\alpha^2},$$

with the corresponding unconditional joint sub-density function:

$$f_{j_1 j_2}(t_1, t_2; \nu, \alpha) = \frac{\partial^2}{\partial t_1 \partial t_2} F_{j_1 j_2}(t_1, t_2; \nu, \alpha) = \left(1+\alpha^2\right) \prod_{k=1}^{2} h_{0,j_k}^{(k)}(t_k;\nu) \left(1 + \alpha^2 \sum_{k=1}^{2} H_0^{(k)}(t_k;\nu)\right)^{-2-1/\alpha^2},$$

for all $t_k > 0$, $j_k = 1, \ldots, L_k$, and $k = 1, 2$. The unconditional joint survival function is given by:

$$S(t_1, t_2; \nu, \alpha) = \int_0^{\infty} \left(\prod_{k=1}^{2} S_{(k)}(t_k;\nu \mid \nu)\right) g(\nu;\alpha) d\nu = \left(1 + \alpha^2 \sum_{k=1}^{2} H_{0(k)}(t_k;\nu)\right)^{-1/\alpha^2},$$

for all $t_1, t_2 > 0$.

2 Proposed methodology

2.1 Maximum Likelihood Estimation

The observed data is in the form (X_1, X_2, J_1, J_2). Here, $J_k = 0$ indicates that the failure time for the k-th individual is right-censored for $k = 1, 2$. A typical observation from a pair of individuals is (x_1, x_2, j_1, j_2), where $x_1 > 0$ and $x_2 > 0$ are the censoring times, and j_1 and j_2 are the respective causes of failure, with $j_k = 0, 1, \ldots, L_k$ for $k = 1, 2$.

To construct the likelihood function based on observations (x_1, x_2, j_1, j_2) from n independent pairs, we consider the likelihood contributions for different combinations of (j_1, j_2) as follows:

Case 1: $j_1 = 1, \ldots, L_1$ **and** $j_2 = 1, \ldots, L_2$

The likelihood contribution is proportional to

$$\begin{aligned} f(x_1, x_2, j_1, j_2, \nu, \alpha) &= \left[f_{j_1 j_2}(u_1, u_2, \nu, \alpha)\, du_1\, du_2\right] h(x_1, x_2) \\ &\propto f_{j_1 j_2}(x_1, x_2, \nu, \alpha). \end{aligned}$$

Case 2: $j_1 = 0$ **and** $j_2 = 1, \ldots, L_2$

The likelihood contribution is proportional to

$$\begin{aligned} f(x_1, x_2, j_1, j_2, \nu, \alpha) = \bigg[&\int_0^{x_2} f_{j_2}^{(2)}(u_2, \nu, \alpha)\, du_2 \\ &- \sum_{j=1}^{L_1} f_{j j_2}(u_1, u_2, \nu, \alpha)\, du_1\, du_2\bigg] h(x_1, x_2) \\ \propto\, & F_{j_2}^{(2)}(x_2, \nu, \alpha) - \sum_{j=1}^{L_1} f_{j j_2}(x_1, x_2, \nu, \alpha). \end{aligned}$$

Case 3: $j_1 = 1, \ldots, L_1$ **and** $j_2 = 0$

The likelihood contribution is proportional to

$$\begin{aligned} f(x_1, x_2, j_1, j_2, \nu, \alpha) = \bigg[&\int_0^{x_1} f_{j_1}^{(1)}(u_1, \nu, \alpha)\, du_1 \\ &- \sum_{j=1}^{L_2} f_{j_1 j}(u_1, u_2, \nu, \alpha)\, du_2\, du_1\bigg] h(x_1, x_2) \\ \propto\, & F_{j_1}^{(1)}(x_1, \nu, \alpha) - \sum_{j=1}^{L_2} f_{j_1 j}(x_1, x_2, \nu, \alpha). \end{aligned}$$

Case 4: $j_1 = 0$ **and** $j_2 = 0$

The likelihood contribution is proportional to

$$f(x_1, x_2, j_1, j_2, \nu, \alpha) = \left[1 - \int_0^{x_1} \int_0^{x_2} f(u_1, u_2, \nu, \alpha)\, du_1\, du_2\right] h(x_1, x_2) \propto S(x_1, x_2, \nu, \alpha).$$

The likelihood function is optimised with respect to the parameters to find the estimate of the corresponding parameters.

2.2 Simulation

We are attempting to generate 100 samples of bivariate right censored where X_1 and X_2 are the two censoring times following a Weibull baseline hazard function with a shape parameter α_1 and a scale parameter β_1^{-1}. The hazard function for the Weibull distribution is given by:

$$h(t) = \alpha_1 \beta_1 t^{\alpha_1 - 1} e^{-(\beta_1 t)^{\alpha_1}}$$

This hazard function describes the probability density of t, the time to event, under the Weibull model. It captures the shape and scale characteristics of the distribution. The algorithm for generating the data is as follows:

1. Generate a frailty variable ϵ from a Gamma distribution with parameter σ^{-2}.
2. Generate two random variables T_1 and T_2 from Weibull distributions:

   $$T_1 \sim \text{Weibull}(\beta_1, \frac{\alpha_1}{\epsilon} \cdot \frac{1}{\beta_1}), \quad T_2 \sim \text{Weibull}(\beta_2, \frac{\alpha_2}{\epsilon} \cdot \frac{1}{\beta_2})$$

   where the first entities are the shape parameters and the second entities are the scale parameters.
3. Generate two censoring times C_1 and C_2 from exponential distributions with parameter θ.
4. Determine censoring occurrence (1 means censoring has not occured):

   $$j_1 = \begin{cases} 1, & \text{if } T_1 <= C_1 \\ 0, & \text{otherwise} \end{cases}, \quad j_2 = \begin{cases} 1, & \text{if } T_2 <= C_2 \\ 0, & \text{otherwise} \end{cases}$$

   This process is for a single sample point. It will be continued till the desired no. of observations is reached.

3 Results

In our case, we have generated 100 data points with the above algorithm using $\alpha_1 = .45$, $\alpha_2 = 2.57$, $\beta_1 = \beta_2 = 5$ and $\sigma = 0.5$.

Parameters	Estimates	Standard Error
α_1	5.2658	5.2248
α_2	4.2358	3.5487
β_1	4.5256	5.3254
β_2	8.2145	6.2451
σ	0.145	2.2110

Table 1: Parameter Estimates with Standard Errors

4 Conclusion

In this study, we have explored the application of shared frailty models to bivariate failure time data with competing risks. The shared frailty model, incorporating a common frailty variable across individuals irrespective of their failure types, provides a robust framework for understanding dependencies in such data. Through parametric modeling of cause-specific hazard functions and the shared frailty variable, we demonstrated how this approach accommodates different sets of competing risks. Our findings highlight the effectiveness of the shared Gamma frailty model, particularly in contexts where familial or environmental factors contribute to shared risks across individuals. Future research could extend these models to non-Gamma frailty distributions and incorporate more complex dependency structures.Frailty models, particularly shared and correlated gamma frailty models, offer significant advantages in survival analysis by accounting for unobserved heterogeneity. They have been successfully applied in various fields, including cancer research, child survival analysis, and dental research.

References

[1] L. Duchateau and P. Jansenn. *The Frailty Model*. New York: Springer Verlag, 2008.

[2] Biswadeep Ghosh, Anup Dewanji, and Sudipta Das. Parametric analysis of bivariate current status data with competing risks using frailty model. *Technical Report No. ASU/2024/01, Applied Statistics Unit, Indian Statistical Institute, Kolkata*, 2024.

[3] A. Wienke. *Frailty models in survival analysis*. Chapman and Hall/CRC, 1st edition, 2010.