Statistics and Probability Letters 132 (2018) 1–6

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Estimation of the survival function with redistribution algorithm under semi-competing risks data

Jin-Jian Hsieh *, Chia-Hao Hsu

Department of Mathematics, National Chung Cheng University Chia-Yi, Taiwan, ROC

ARTICLE INFO

Article history:
Received 13 February 2017
Received in revised form 25 July 2017
Accepted 8 September 2017
Available online 19 September 2017

Keywords:
Archimedean Copula model
Dependent censoring
Semi-competing risks data
Redistribution algorithm

ABSTRACT

This paper focuses on the estimation of the survival function of the non-terminal event time for semi-competing risks data. Without extra assumptions, we cannot make inference on the non-terminal event time because the non-terminal event time is dependently censored by the terminal event time. Thus, we utilize the Archimedean copula model to specify the dependency between the non-terminal event time and the terminal event time. Under the Archimedean copula assumption, we apply the redistribution method to estimate the survival function of the non-terminal event time and compare it with the copula-graphic estimator introduced by Lakhal et al. (2008). We also apply our suggested approach to analyze the Bone Marrow Transplant data.

© 2017 Published by Elsevier B.V.

1. Introduction

Consider the study of leukemia patients receiving the bone marrow transplants (Klein and Moeschberger, 2003). Let T be the relapse time of leukemia from the bone marrow transplant, and D be the time of death. If T occurs prior to D, the relapse time of leukemia, T, and the death time, D, could be observed. If D occurs prior to T, the relapse time of leukemia, T, could not be observed. Thus, the terminal event time D may censor the nonterminal event time T, but not vice versa. Further, T and D may be correlated. This type of data is referred to as the semi-competing risks data by Fine et al. (2001), which has been investigated by several literatures, such as Lagakos (1976, 1977), Flandre and O'Quigley (1995), Lin et al. (1996), Day et al. (1997), Chang (2000), Rivest and Wells (2001), Wang (2003), Jiang et al. (2005b), Jiang et al. (2005a), Peng and Fine (2005), Peng and Fine (2007), Hsieh et al. (2008), Lakhal et al. (2008), Ding et al. (2009), Hsieh and Huang (2012). In this article, we would like to estimate the survival function of T under the semi-competing risk data. Because T is dependently censored by the terminal event time D, we cannot use the Kaplan–Meier method to estimate the survival function of T. Fine et al. (2001) proposed an estimator of the survival function of T by a straight algebraic calculation under a copula model. Jiang, Fine, Kosorok and Chappell (2005) suggested a self-consistent estimator for the survival function of T. Lakhal et al. (2008) applied the copula-graphic estimator to estimate the survival function of T, which was first introduced by Zheng and Klein (1995). Here, we use the Archimedean copulas to model the dependency of (T, D) on the semi-competing risks data and apply a redistribution method to estimate the survival function of T.

The rest of this article is organized as follows. In Section 2, we introduce the data structure and the copula model. We propose a redistribute-to-the-right algorithm to estimate the survival function of T under semi-competing risks data in Section 3. In Section 4, we display the settings of simulation to assess the empirical performance of our proposed approach and compare the redistribution method with the method by Lakhal et al. (2008). Then, we apply our proposed method to a practical bone marrow transplant data in Section 5. In Section 6, we conclude with some remarks.

* Corresponding author.
E-mail address: jjhsieh@math.ccu.edu.tw (J.-J. Hsieh).

http://dx.doi.org/10.1016/j.spl.2017.09.003
0167-7152/© 2017 Published by Elsevier B.V.

2. Data and models

In this section, we introduce the semi-competing risks data. Let T be the time to non-terminal event and D be the time to terminal event. The non-terminal event time may be censored by the terminal event time. But, the terminal event time is not censored by the non-terminal event time. Let C be a right-censoring variable. Define $X = T \wedge D \wedge C$, $\delta_x = I(T < D \wedge C)$, $Y = D \wedge C$, $\delta_y = I(D < C)$, where $\wedge$ is the minimum operator and $I(\cdot)$ is the indicator function. Denote the observed data by $\{(x_i, y_i, \delta_{x_i}, \delta_{y_i}) : i = 1, \ldots, N\}$. Here, we assume (T, D) are independent of C in the following inference.

To specify the dependence between T and D, we assume (T, D) follow a copula model as

$$\pi(t, d) = P(T > t, D > d) = C_\alpha\{S_T(t), S_D(d)\}, \quad 0 \le t \le d, \tag{1}$$

where $S_T(t)$ and $S_D(t)$ are the marginal survival functions of T and D. $C_\alpha(\cdot)$ is the copula function and α is the association parameter. The association parameter α in (1) does not have a direct interpretation, but it is related to Kendall's tau, defined by

$$\tau = P\{(T_1 - T_2)(D_1 - D_2) > 0\} - P\{(T_1 - T_2)(D_1 - D_2) < 0\},$$

where (T_1, D_1) and (T_2, D_2) are independent replications of (T, D). The advantage of the copula model is that the association and the marginal distributions could be separated from the copula function. The Archimedean copula (AC) model is a popular subclass of the copula model, which can be expressed as

$$C_\alpha(u, v) = \phi_\alpha^{-1}\{\phi_\alpha(u) + \phi_\alpha(v)\}, \quad 0 \leqslant u, v \leqslant 1,$$

where ϕ_α is a non-increasing convex function defined on $(0, 1]$ with $\phi_\alpha(1) = 0$.

3. The redistribution method

In this section, we introduce a redistribution algorithm to estimate $S_T(t)$ for the semi-competing risks data. Let $\{x_{(1)}, \ldots, x_{(k)}\}$ be the ordered distinct time points of $\{x_1, \ldots, x_N\}$, where $0 < x_{(1)} < x_{(2)} < x_{(3)} < \cdots < x_{(k)} < +\infty$. For this data, if $\delta_x = 1$, the non-terminal event time T could be observed and $X = T$. If $\delta_x = 0$ and $\delta_y = 1, X = D$. If $\delta_x = 0$ and $\delta_y = 0, X = C$. The observed data can be summarized as the grouped data $\{(x_{(j)}, n_j, f_j^1, f_j^2) : j = 1, 2, \ldots, k\}$, where $n_j = \sum_{i=1}^N I(X_i = x_{(j)})$ denotes the number of individuals observed at $x_{(j)}$, $f_j^1 = \sum_{i=1}^N I(X_i = x_{(j)}, \delta_{x_i} = 0, \delta_{y_i} = 1)$, and $f_j^2 = \sum_{i=1}^N I(X_i = x_{(j)}, \delta_{x_i} = 0, \delta_{y_i} = 0)$. $N = \sum_{j=1}^k n_j$. For $s = 1, 2, \ldots, k+1$, let I_s denote the interval $(x_{(s-1)}, x_{(s)}]$, where $x_{(0)} = 0, x_{(k+1)} = \infty$. Let $M_s = \sum_{i=1}^N I(T_i \in I_s)$ denote the number of individuals whose non-terminal events occur exactly on I_s. Let $P_s = P(T \in I_s)$ denote the probability of the individuals failed in I_s. To specify the dependence of (T, D), assume (T, D) follow a Archimedean copula (AC) as:

$$P(T > t, D > d) = \phi_\alpha^{-1}[\phi_\alpha(S_T(t)) + \phi_\alpha(S_D(d))],$$

where $S_T(t)$ and $S_D(d)$ are marginal survival functions of T and D. If $\delta_{x_i} = 0$ and $\delta_{y_i} = 1$, define

$$\begin{aligned}
S_T^*(x_{(s)}|x_{(j)}) &= P(T > x_{(s)}|D = x_{(j)}) = \frac{P(T > x_{(s)}, D = x_{(j)})}{P(D = x_{(j)})} \\
&= \phi_\alpha^{-1\prime}[\phi_\alpha(S_T(x_{(s)})) + \phi_\alpha(S_D(x_{(j)}))]\phi_\alpha'(S_D(x_{(j)})) \\
&= \phi_\alpha^{-1\prime}[\phi_\alpha(\sum_{j=s+1}^{k+1} P_j) + \phi_\alpha(S_D(x_{(j)}))]\phi_\alpha'(S_D(x_{(j)})), \\
\pi_{js}^* &= P(x_{(s-1)} < T \le x_{(s)}|D = x_{(j)}) = S_T^*(x_{(s-1)}|x_{(j)}) - S_T^*(x_{(s)}|x_{(j)}),
\end{aligned}$$

where $\phi_\alpha^{-1\prime}(t) = \partial\phi_\alpha^{-1}(t)/\partial t$, $\phi_\alpha'(t) = \partial\phi_\alpha(t)/\partial t$, and $j < s$. Note that $S_T^*(x_{(k+1)}|x_{(j)}) = S_T^*(\infty|x_{(j)}) = 0$. If $\delta_{x_i} = 0$ and $\delta_{y_i} = 0$, define

$$\begin{aligned}
S_T^{**}(x_{(s)}|x_{(j)}) &= P(T > x_{(s)}|D > x_{(j)}) = \frac{\phi_\alpha^{-1}[\phi_\alpha(S_T(x_{(s)})) + \phi_\alpha(S_D(x_{(j)}))]}{S_D(x_{(j)})} \\
&= \frac{\phi_\alpha^{-1}[\phi_\alpha(\sum_{j=s+1}^{k+1} P_j) + \phi_\alpha(S_D(x_{(j)}))]}{S_D(x_{(j)})}, \\
\pi_{js}^{**} &= P(x_{(s-1)} < T \le x_{(s)}|D > x_{(j)}) = S_T^{**}(x_{(s-1)}|x_{(j)}) - S_T^{**}(x_{(s)}|x_{(j)}),
\end{aligned}$$

where $j < s$, and $S_T^{**}(x_{(k+1)}|x_{(j)}) = S_T^{**}(\infty|x_{(j)}) = 0$.

Theorem. *According to the notations in the above, the following equation holds:*

$$P_s = \frac{1}{N}[\sum_{j=1}^{s-1} E(f_j^1)\frac{\pi_{js}^* I(s > 1)}{\sum_{r=j+1}^{k+1}\pi_{jr}^*} + \sum_{j=1}^{s-1} E(f_j^2)\frac{\pi_{js}^{**} I(s > 1)}{\sum_{r=j+1}^{k+1}\pi_{jr}^{**}} + E(n_s - f_s^1 - f_s^2)], \tag{2}$$

where $s = 1, 2, \ldots, k+1$.

Download English Version:

https://daneshyari.com/en/article/5129624

Download Persian Version:

https://daneshyari.com/article/5129624

Daneshyari.com