Journal of Multivariate Analysis 132 (2014) 229–241

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

An improved nonparametric estimator of sub-distribution function for bivariate competing risk models

Takeshi Emura [a,*], Fan-Hsuan Kao [a], Hirofumi Michimae [b]

[a] Graduate Institute of Statistics, National Central University, Taiwan
[b] School of Pharmacy, Department of Clinical Medicine (Biostatistics), Kitasato University, Japan

ARTICLE INFO

Article history:
Received 18 July 2013
Available online 6 September 2014

AMS subject classifications:
62N01
62N02
62G05
62G20
62C05

Keywords:
Bivariate survival function
Right censoring
Survival analysis

ABSTRACT

For competing risks data, it is of interest to estimate the sub-distribution function of a particular failure event, which is the failure probability in the presence of competing risks. However, if multiple failure events per subject are available, estimation procedures become challenging even for the bivariate case. In this paper, we consider nonparametric estimation of a bivariate sub-distribution function, which has been discussed in the related literature. Adopting a decision-theoretic approach, we propose a new nonparametric estimator which improves upon an existing estimator. We show theoretically and numerically that the proposed estimator has smaller mean square error than the existing one. The consistency of the proposed estimator is also established. The usefulness of the estimator is illustrated by the salamander data and mouse data.

1. Introduction

Statistical analysis of competing risks data is common in biology, where individuals experience multiple failure causes. For instance, larvae grown in a cage may experience either metamorphosis or death, whichever comes first [15]. These two failure causes are mutually exclusive in that each larva exhibits only one of the two causes at the time of failure. This type of data is popular, especially in biomedical research involving human and animal subjects [2]. Competing risks models are used to analyze such data. An overview of competing risks data analysis is referred to Crowder [8] and Bakoyannis and Touloumi [5].

In competing risks data analysis, the sub-distribution function plays a fundamental role. Let T be a failure time and $C \in \{1, 2, \dots, \gamma\}$ be the failure cause for γ distinct causes. The *sub-distribution function* (also known as cumulative incidence function) is defined as

$$F_j(t) = \Pr(T \le t, C = j), \quad j = 1, 2, \dots, \gamma.$$

This is the proportion of failure events occurring due to cause j before time t. The sub-distribution function is easy to interpret and is often the target for estimation [8,5,13].

In applications, bivariate competing risks arise naturally. For instance, a pair of larvae in an experimental cage shares unobserved environmental or genetic factors [15]. In analysis of such data, the univariate competing risks models need to be generalized to bivariate models.

* Corresponding author.
E-mail addresses: emura@stat.ncu.edu.tw, takeshiemura@gmail.com (T. Emura).

http://dx.doi.org/10.1016/j.jmva.2014.08.009

The bivariate competing risk models have been recently considered by many authors [4,19,17,18,20]. Under the bivariate competing risks models, the target of estimation is the *bivariate sub-distribution function* (formally defined in Section 2). Several nonparametric estimators have been proposed under various censoring and truncation schemes. Antony and Sankaran [4] and Sankaran et al. [19] developed nonparametric estimators under right-censoring. Sankaran and Antony [18] considered a similar problem where the censoring times are missing. Sankaran and Antony [17] proposed a nonparametric estimator under left-truncation and right-censoring. Shen [20] considered nonparametric estimation under double censoring.

This paper considers nonparametric estimation of the bivariate sub-distribution functions under right censoring as in [4,19]. Note that nonparametric estimation under bivariate competing risks is much more challenging than its univariate counterpart. Especially in small sample sizes, the estimator of Sankaran et al. [19] will generally be a crude step function and will have a large mean squared error (MSE). In light of this problem, the main objective of this paper is to propose a new nonparametric estimator that aims to improve upon the existing estimator. The proposed estimator not only reduces the MSE but also smoothes out the crude step function estimator in some degree.

The paper is organized as follow. Section 2 introduces basic notations and the estimator of Sankaran et al. [19]. Section 3 proposes a new estimator for the bivariate sub-distribution function. Section 4 verifies the consistency of the proposed estimator. Section 5 presents simulations comparing the proposed method with the existing one. Section 6 analyzes the mouse data and the salamander data. Section 7 concludes the paper.

2. Preliminary

This section defines basic notations for bivariate competing risks models and then introduces the nonparametric estimator of Sankaran et al. [19] for estimating a bivariate sub-distribution function.

Let $S(t_1, t_2) = \Pr(T_1 > t_1, T_2 > t_2)$ be the survivor function of bivariate failure times (T_1, T_2). Also, let $(C_1, C_2) \in \{1, 2, \ldots, \gamma_1\} \times \{1, 2, \ldots, \gamma_2\}$ be the corresponding bivariate failure causes. For $(i, j) \in \{1, 2, \ldots, \gamma_1\} \times \{1, 2, \ldots, \gamma_2\}$, the cause-specific hazard is

$$\Lambda_{ij}(dt_1, dt_2) = \frac{\Pr(T_1 \in dt_1, T_2 \in dt_2, C_1 = i, C_2 = j)}{\Pr(T_1 \geq t_1, T_2 \geq t_2)}.$$

Also, the sub-distribution function is

$$F_{ij}(t_1, t_2) = \Pr(T_1 \leq t_1, T_2 \leq t_2, C_1 = i, C_2 = j).$$

The cause-specific hazard and the sub-distribution functions are related through

$$F_{ij}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} S(u^-, v^-) \Lambda_{ij}(du, dv). \tag{1}$$

The above identity is useful for estimating the sub-distribution F_{ij} under right-censoring. If (T_1, T_2) are censored by a pair of independent censoring times (Z_1, Z_2), one observes (Y_1, Y_2) and (δ_1, δ_2), where $Y_k = \min(T_k, Z_k)$ and $\delta_k = \mathbf{I}(T_k = Y_k)$ for $k = 1, 2$, where $\mathbf{I}(\cdot)$ is the indicator function. If $\delta_k = 0$, then we set $C_k = 0$ since the value of C_k is not available. If $H(t_1, t_2) \equiv \Pr(Y_1 > t_1, Y_2 > t_2) > 0$, Eq. (1) becomes

$$F_{ij}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{S(u^-, v^-) F_{ij}^*(du, dv)}{H(u^-, v^-)}, \quad i = 1, 2, \ldots, \gamma_1, j = 1, 2, \ldots, \gamma_2, \tag{2}$$

where

$$F_{ij}^*(t_1, t_2) = \Pr(T_1 \leq t_1, T_2 \leq t_2, \delta_1 = 1, \delta_2 = 1, C_1 = i, C_2 = j).$$

Sankaran et al. [19] used Eq. (2) to estimate F_{ij} based on observations (Y_{1u}, Y_{2u}), (C_{1u}, C_{2u}), and $(\delta_{1u}, \delta_{2u})$, $u = 1, 2, \ldots, n$, which are i.i.d. replications of (Y_1, Y_2), (C_1, C_2), and (δ_1, δ_2). They consider an estimator of $H(t_1, t_2)$ as

$$\hat{H}(t_1, t_2) = \frac{1}{n} \sum_{u=1}^{n} \mathbf{I}(Y_{1u} > t_1, Y_{2u} > t_2),$$

and an estimate of $F_{ij}^*(t_1, t_2)$ as

$$\hat{F}_{ij}^*(t_1, t_2) = \frac{1}{n} \sum_{u=1}^{n} \mathbf{I}(Y_{1u} \leq t_1, Y_{2u} \leq t_2, \delta_{1u} = 1, \delta_{2u} = 1, C_{1u} = i, C_{2u} = j).$$

Under $\hat{H}(t_1, t_2) > 0$, they obtain the nonparametric estimator for $F_{ij}(t_1, t_2)$ as

$$\hat{F}_{ij}(t_1, t_2) = \int_0^{t_1} \int_0^{t_2} \frac{\hat{S}(u^-, v^-) \hat{F}_{ij}^*(du, dv)}{\hat{H}(u^-, v^-)}, \quad i = 1, 2, \ldots, \gamma_1, j = 1, 2, \ldots, \gamma_2. \tag{3}$$

Sankaran et al. [19] proposed to apply the Dabrowska estimator [9] for $\hat{S}$. Other estimators are also available, such as the estimators of Prentice and Cai [16] and Wang and Wells [22]. The strong consistency and weak convergence for $\hat{F}_{ij}$ are studied by [19].