Journal of Multivariate Analysis 101 (2010) 1903–1909

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Estimating cumulative incidence functions when the life distributions are constrained

Hammou El Barmi [a,*], Matthew Johnson [b], Hari Mukerjee [c]

[a] Department of Statistics and Computer Information Systems, Box 11-220, Baruch College, The City University of New York, One Baruch Way, New York, NY 10010, United States

[b] Teachers College, Columbia University, United States

[c] Wichita State University, United States

ARTICLE INFO

Article history:
Received 5 March 2008
Available online 5 May 2010

AMS 2000 Subject Classifications:
primary 62G05
secondary 60F17
62G30

Keywords:
Competing risks
Cumulative incidence function
Stochastic ordering
Weak convergence

ABSTRACT

In competing risks studies, the Kaplan–Meier estimators of the distribution functions (DFs) of lifetimes and the corresponding estimators of cumulative incidence functions (CIFs) are used widely when no prior information is available for these distributions. In some cases better estimators of the DFs of lifetimes are available when they obey some inequality constraints, e.g., if two lifetimes are stochastically or uniformly stochastically ordered, or some functional of a DF obeys an inequality in an empirical likelihood estimation procedure. If the restricted estimator of a lifetime differs from the unrestricted one, then the usual estimators of the CIFs will not add up to the lifetime estimator. In this paper we show how to estimate the CIFs in this case. These estimators are shown to be strongly uniformly consistent. In all cases we consider, when the inequality constraints are strict the asymptotic properties of the restricted and the unrestricted estimators are the same, thus providing the asymptotic properties of the restricted estimators essentially "free of charge". We give an example to illustrate our procedure.

Published by Elsevier Inc.

1. Introduction

Consider a competing risks study with k populations. For the ith population, let T_i denote the lifetime, assumed continuous, F_i its distribution function (DF), $S_i = 1 - F_i$ its survival function (SF), and let $\delta_i \in \{0, 1, 2, \ldots, r_i\}$ denote the cause of failure or death, where $\{\delta_i = 0\}$ denotes a censored event. The cumulative incidence function (CIF) of failure due to cause j in the ith population is defined by $F_{ij}(t) = P[T_i \le t, \delta_i = j]$ for $1 \le j \le r_i,\ 1 \le i \le k$. Clearly, $\sum_j F_{ij} = F_i$ for all i.

When no prior information is available, the Kaplan–Meier estimators (KMEs) of the DFs and the corresponding estimators of the CIFs are widely used (see [9]). In some cases, utilization of prior information provide better estimators of the DFs. One class of such estimators are provided when the DFs obey some order restriction. Consider the following sequence of four well known order restrictions on T_1 and T_2. We define T_1 to be larger than T_2 in stochastic precedence ordering, $T_1 \ge_{spo} T_2$, if $P(T_1 \ge T_2) \ge 1/2$. We define T_1 to be stochastically larger than T_2, $T_1 \ge_{so} T_2$, if $F_1 \le F_2$. We define T_1 to be larger than T_2 in uniform stochastic ordering, $T_1 \ge_{uso} T_2$, if S_2/S_1 is nonincreasing. If F_1 and F_2 have densities f_1 and f_2, respectively, then T_1 is said to be larger than T_2 in likelihood ratio ordering, $T_1 \ge_{lro} T_2$, if f_1/f_2 is nondecreasing. The following strict implications hold for the four order restrictions:

$$\text{LRO} \implies \text{USO} \implies \text{SO} \implies \text{SPO}.$$

* Corresponding author.
E-mail address: hammou_elbarmi@baruch.cuny.edu (H. El Barmi).

0047-259X/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmva.2010.04.002

Even in the 1-sample case, the empirical likelihood estimator introduced by Owen (see [11]) can be used to improve upon the KM estimator if some functional of the DF is known to obey some inequality, e.g., if the mean is known to be less than or equal to some μ_0.

If the KME and the improved estimator of F_i are denoted by $\hat{F}_i$ and F_i^*, respectively, the usual estimators, $\hat{F}_{ij}$s of the F_{ij}s will not add up to F_i^* unless they are the same, requiring these estimators to be adjusted. The method of adjustment is not obvious. For example, multiplying $\hat{F}_{ij}$ by $F_i^*/\hat{F}_i$ might seem appealing, but it might destroy the monotonicity of the estimators. Since F_i^* is obtained by moving the masses of $\hat{F}_i$ to satisfy the constraints, and since the mass at a point corresponds to failure due to one cause only, our procedure simply makes the corresponding movement of the mass for the CIF corresponding to that particular cause; details are provided in Section 2. It should be noted that the adjustment of $\hat{F}_{ij}$ depends only on $\hat{F}_i$ and F_i^*, i.e., the adjustment of $\hat{F}_{ij}$ does not depend on F_l^* for any $l \neq i$. Hence, the same procedure is applicable to any improvements of $\hat{F}_i$s.

All of the orderings mentioned above are transitive. However, the restricted estimators of the DFs and their properties have been investigated in the literature for the 2-sample case only, except for the stochastic ordering case where the extension to the k-sample case has been thoroughly investigated by El Barmi and Mukerjee [4], hereafter referred to as EBM [4]. In this paper we consider the problem of estimation of the CIFs when k life distributions are stochastically ordered,

$$F_1 \leq F_2 \leq \cdots \leq F_k. \tag{1}$$

It is shown that if the inequality constraints are strict, then the restricted and the unrestricted estimators have the same asymptotic distributions. Thus, the asymptotic inference procedures remain unchanged while we get the improvements in small samples by the use of order restriction "free of charge". In Section 2 we consider the 2-sample case for ease of exposition. In Section 3 we show that the k-sample case is an easy extension of the 2-sample case. In Section 4 we illustrate our methods with a real life example. In Section 5 we make some concluding remarks.

Throughout, we use the left-continuous inverse of F, the DF of a lifetime, given by

$$F^{-1}(p) = \inf\{t \geq 0 : F(t) \geq p\}, \quad 0 \leq p \leq 1, \text{ where } F^{-1}(1) \text{ may be } \infty. \tag{2}$$

We also use the notation

$$\|f\|_a^b \text{ for } \sup_{a \leq t \leq b} |f(t)| \text{ for any function } f.$$

The subscripts i and ij will always stand for the ith population and jth CIF of the ith population, respectively.

2. The 2-sample case

For $i = 1, 2$, let T_i have a continuous DF, F_i, that has a density, f_i, and SF, $S_i = 1 - F_i$, subject to r_i competing risks, along with a possible random right censoring. We identify the risks of failure in the ith population by $\delta_i = 0, 1, \ldots, r_i$, where $\{\delta_i = 0\}$ is the event that an observation has been censored. We assume that $F_1 \leq F_2$.

2.1. The estimators

Let $\{T_{il} : 1 \leq l \leq n_i,\ i = 1, 2\}$ be independent random samples from the two populations, and let C_{il}, with a continuous DF, G_i, denote the censoring time of the lth subject in the ith population. We observe (L_{il}, δ_{il}), where $L_{il} = T_{il} \wedge C_{il}$ and δ_{il} is the cause of failure. We assume that $\{C_{il}, T_{il} : 1 \leq l \leq n_i, i = 1, 2\}$ are independent. Let π_i be the survival function of L_{il}. Then $\pi_i = S_i\bar{G}_i$ from our independence assumption, where $\bar{G}_i = 1 - G_i$. We assume that all observation points are distinct since it occurs with probability 1.

To avoid proliferation of subscripts and possible confusion, we suppress the sample size dependence of the estimators.

The KMEs of the F_is are given by

$$\hat{F}_i(t) = 1 - \hat{S}_i(t) = 1 - \prod_{\{l:L_{i:l} \leq t\}} \left(1 - \frac{1}{n_i - l + 1}\right)^{\delta_{i:l}}, \quad t \geq 0,\ i = 1, 2, \tag{3}$$

where $\{L_{i:l}\}$ are the order statistics from $\{L_{il}\}$ and $\delta_{i:l}$s are the corresponding values of δ. Following the usual practice, we consider the last observation to be uncensored in order to define the estimators for all t.

Estimation of DFs under stochastic ordering has a long history. The nonparametric likelihood estimators (NPMLEs) in the 2-sample uncensored case were found by Brunk et al. [1]. In the censoring case, Dykstra [2] derived the NPMLEs in the 2-sample case. This was extended to the k-sample case by Feltz and Dykstra [5] who provided an iterative algorithm that converges to the NPMLEs. The asymptotic distributions in the 2-sample case were derived by Præstgaard and Huang [12]. These are very complicated and difficult to use for further analyses. Hogg [8] had suggested an alternative procedure that simply uses the least squares estimators of $F_1(t)$ and $F_2(t)$ at each t, subject to the constraint, $F_1(t) \leq F_2(t)$, the so-called isotonic regression of the unrestricted estimators of $F_1(t)$ and $F_2(t)$; see the monograph by Robertson et al. [13] for properties of isotonic regression. Unpublished simulations by El Barmi et al. [3] show that Hogg's [8] estimators appear to have smaller MSE than the NPMLEs at almost all quantiles for almost all distributions they have considered. In the 2-sample