International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 9(Special Issue), 2009, pp 113-133

Asymptotic Properties of the NPMLE of a Joint Distribution Function Based on Multivariate Current Status Data

Dianliang Deng
Department of Mathematics and Statistics
University of Regina Regina
Canada S4S 0A2
E-mail: deng@uregina.ca

[Received October 1, 2008; Accepted January 15, 2009]

Abstract

In this paper the nonparametric estimation of a joint distribution when the available data is a multivariate current status data, so-called case 1 interval censored data, is considered. The nonparametric maximum likelihood estimator (NPMLE) is proposed, and it is shown that this estimator satisfies a self-consistency property. This self-consistency property is exploited to derive some asymptotic properties of the estimator, such as strong consistency and rates of convergence. The convergence rates are derived under the strong conditions, which include that the joint density of multivariate vector is bounded away from both zero and infinity and that the joint distribution function is a multivariate Archimedean copula.

Keywords and Phrases: Archimedean Copula, Consistency, Convergence Rate, Current Status, Multivariate Lifetime Data, Nonparametric Maximum Likelihood Estimation.

AMS Classification: 62G20; 62N01; 62H12.

1 Introduction

Multivariate interval-censored lifetime data arise in many areas such as epidemiology, bio-medicine, demography, sociology and financial studies. So far there have been a number of literature dealing with the statistical inference on the interval-censored lifetime data. For instance, Wang and Wells (1997, 2000) studied nonparametric estimators of a bivariate survival function under censoring conditions, proposed model selection procedures for censored data generated by the Archimedean copula family and developed estimators of some time-dependent association measures, including estimators of local and global Kendall's tau, local odds ratio, and other measures defined in their papers. Wang and Ding (2000) and Ting and Wang (2004) discussed the associations for bivariate current status data and further developed a nonparametric procedure for assessing marginal independence for current status data using contingency table techniques. Betensky and Finkelstein (1999), Bogaerts and Lesaffre (2004) and Gentleman and Vandal (2001, 2002) discussed the NPMLE based on the bivariate interval-censored data and proposed three algorithms for finding this NPMLE (see also Sun, 2006). Further, Wong and Yu (1999) studied the generalized maximum likelihood estimator (GMLE) of the joint distribution of a multivariate random vector based on the interval-censored data and established the consistency of the GMLE under the assumptions that the multivariate random vector is independent of the multivariate censoring vector and that their components are discrete. Yu, et al (2006) relaxed the assumptions and obtained the strong consistency of the GMLE in the topologies of weak convergence and pointwise convergence.

However, although the strong consistency of the GMLE is derived for the multivariate interval-censored data, to the best of my knowledge, there has not been a study fully investigating the asymptotic properties of these proposed estimators such as self-consistency equation and the rates of convergence. Moreover, it is worth pointing out that in view of complexity of multivariate lifetime data, most of existing results are derived only for bivariate interval-censored lifetime data. It is not feasible to derive the convergence rates of the nonparametric estimators in the multivariate interval-censored data by using the methods proposed in Wong and Yu (1999) and Yu, et al (2006). In an effort to solve this problem, new procedures need to be developed.

The purpose of this paper is to study properties of the nonparametric maximum likelihood estimate (NPMLE) of the joint distribution of a multivariate lifetime vector of interest based on the current status (case 1 interval-censored) data. In particular, results about the self-consistency equation, strong consistency and convergence rate of the NPMLE with multivariate case 1 interval-censoring mechanism are given. The rest of this paper is organized as follows. In Section 2, the assumptions and notation that will be used throughout the paper are introduced for the multivariate lifetime variables with the general case 1 interval censoring mechanism. The self-consistency equation of the NPMLE is obtained in Section 3. In Section 4, the strong consistency for NPMLE is derived for the general multivariate current status data. The convergence rates of NPMLE in two situations are also given in this section. Concluding remarks are addressed in Section 5. The proofs are presented in the Appendix.

2 Notation and Assumptions

Let $\boldsymbol{X} = (X_1, ..., X_d)$ be the vector of lifetime variables of interest with joint distribution function $F(\boldsymbol{x}) = F(x_1, ..., x_d) = P(X_1 \leq x_1, ..., X_d \leq x_d)$ and marginal distribution functions $F_r(x_r)(r = 1, ..., d)$. If $d = 2$, (X_1, X_2) may be the bivariate variable in matched-paired case control studies and studies of time to occurrence of a disease to paired organs. In this situation, only the information about whether the lifetimes of interest lie before or after the corresponding censoring times is known, which means that only censored data are observed for the multivariate lifetime vector $\boldsymbol{X} = (X_1, ..., X_d)$.

Suppose that there exists a vector of random variables $\boldsymbol{C} = (C_1, ..., C_d)$ that decides the current status for $\boldsymbol{X} = (X_1, ..., X_d)$. That is, the random vectors $\boldsymbol{C} = (C_1, ..., C_d)$ and $\boldsymbol{\Delta} = (\Delta_1, ..., \Delta_d)$, where $\boldsymbol{\Delta} = (\Delta_1, ..., \Delta_d) = (I_{\{X_1<C_1\}}, ..., I_{\{X_d<C_d\}})$, are observed. For $r = 1, ..., d$, the variables C_r and Δ_r define the observed current status for each X_r. However, it is not always true that each component of $\boldsymbol{X}$ has its own censoring time. Often several components in $\boldsymbol{X}$ share a censoring time. For example, in the bivariate case ($d = 2$), X_1 and X_2 are often measured from the same individual since the paired data arise in many cases, and thus they share a common censoring time C. Therefore the more general censoring mechanism should be considered in the multivariate situation.

Roughly speaking, the random vector of interest $\boldsymbol{X} = (X_1, ..., X_d)$ is divided into several subvectors and all components of each subvector share a common censoring time. Rigorously, the vector $\boldsymbol{X}$ is partitioned into p subvectors $\boldsymbol{X} = (\boldsymbol{X}_1, ..., \boldsymbol{X}_p)$ with $\boldsymbol{X}_r = (X_{d_1+\cdots+d_{r-1}+1}, ..., X_{d_1+\cdots+d_{r-1}+d_r}), (r = 1, ..., p)$, $d_0 = 0$ and $d_1 + d_2 + \cdots + d_p = d$. The censoring time of components in $\boldsymbol{X}_r$ is a univariate variable C_r $(r = 1, ..., p)$. In this case, the random vectors $\boldsymbol{C}$ and $\boldsymbol{\Lambda}$ are observed, where $\boldsymbol{C} = (C_1, ..., C_p)$ and $\boldsymbol{\Lambda} = (\Lambda_1, ..., \Lambda_d)$ with $\Lambda_s = I_{\{X_s<C_r\}}, (d_1+\cdots+d_{r-1}+1 \le s \le d_1+\cdots+d_r; r = 1, ..., p)$. Note that $1 \le p \le d$. If $p = 1$, all X_s's share one common censoring time C and if $p = d$, each X_s has its own censoring time C_s for $s = 1, ..., d$.

Our goal is to make the statistical inference about the distribution function $F(\boldsymbol{x})$ of $\boldsymbol{X}$ from the observed random vectors $\boldsymbol{C}$ and $\boldsymbol{\Lambda}$. Further we will set up the self-consistency equation for NPMLE of the distribution $F(\boldsymbol{x})$ based on the multivariate current status data and then derive the strong consistency and the convergence rates for NPMLE. Toward these ends, we need to introduce some notation for the multivariate current status data.

Set $\boldsymbol{D} = \{1, 2, ..., d\}$, $D_r = \{i : d_1+\cdots+d_{r-1}+1 \le i \le d_1+\cdots+d_r\}(r = 1, 2, ..., p)$ and $\mathfrak{D} = \{D : D \text{ is a subset of } \boldsymbol{D}\}$. Let μ_F be the measure induced by the distribution function F in $\mathbb{R}_+^d$. For any $D \in \mathfrak{D}$, $\boldsymbol{c} \in \mathbb{R}_+^p$ and $\boldsymbol{x} \in \mathbb{R}_+^d$, define the sets $A_D(\boldsymbol{c})$ and $A_D^{\boldsymbol{x}}(\boldsymbol{c})$ in $\mathbb{R}_+^d$ as

$$A_D(\boldsymbol{c}) = A_D(c_1, ..., c_p) = \bigotimes_{r=1}^{p} \left\{[0, c_r)^{\alpha_r} \times [c_r, \infty)^{\beta_r}\right\},$$

$$A_D^{\boldsymbol{x}}(\boldsymbol{c}) = A_D^{\boldsymbol{x}}(c_1, ..., c_p) = \bigotimes_{r=1}^{p} \left\{ \bigotimes_{i\in D\cap D_r} [0, c_r \wedge x_i) \times \bigotimes_{j\in \bar{D}\cap D_r} [c_r \wedge x_j, x_j] \right\}$$

where $\bar{D} = \boldsymbol{D} - D$, $\alpha_r = \text{card}\{D \cap D_r\}$, $\beta_r = \text{card}\{\bar{D} \cap D_r\}$ (note that $\alpha_r + \beta_r =$

card$\{D_r\} = d_r$), $\bigotimes_{r=1}^{m}[u_r, v_r)^{k_r} = [u_1, v_1)^{k_1} \times \cdots \times [u_m, v_m)^{k_m}$ is a rectangle set of $I\!R_+^{\sum_{r=1}^{m} k_r}$ with $[u_r, v_r)^{k_r} = \underbrace{[u_r, v_r) \times \cdots \times [u_r, v_r)}_{k_r} \in I\!R_+^{k_r}$ and $a \wedge b = \min\{a, b\}$. Then for any $\boldsymbol{c} \in I\!R_+^p$,

$$\bigcup_{D\in\mathfrak{D}} A_D(\boldsymbol{c}) = I\!R_+^d \text{ and } \bigcup_{D\in\mathfrak{D}} A_D^{\boldsymbol{x}}(\boldsymbol{c}) = [0, x_1] \times \cdots \times [0, x_d] \equiv [0, \boldsymbol{x}].$$

Similarly, for any $\boldsymbol{x} \in I\!R_+^d$ and $D \in \mathfrak{D}$, define the set $C_D(\boldsymbol{x})$ in $I\!R_+^p$ as

$$C_D(\boldsymbol{x}) = C_D(x_1, ..., x_d) = \bigotimes_{r=1}^{p}(\max_{s\in D_r\cap D} x_s, \min_{t\in D_r\cap \bar{D}} x_t]$$

Further we have that for any $\boldsymbol{x} \in I\!R_+^d$,

$$\bigcup_{D\in\mathfrak{D}} C_D(\boldsymbol{x}) = I\!R_+^p.$$

Throughout this paper, we assume:

(A1) $\boldsymbol{X} = (X_1, ..., X_d)$ is a vector of non-negative continuous random variables. The joint distribution function $F(\boldsymbol{x})$ of $\boldsymbol{X}$ is contained in the class $\mathcal{F}_M := \{F|\text{support}(F) \subset [0, M]^d;\ F \ll \lambda^d\}$ where M is a given positive constant, λ^d is the Lebesgue measure in $I\!R_+^d$, and $I\!R_+^m = \{(x_1, ..., x_m) : c_r \geq 0, r = 1, ..., m\}$ for any positive integer m.

(A2) $\boldsymbol{C} = (C_1, ..., C_p)$ is a vector of non-negative continuous random variables with joint distribution G. $G(\boldsymbol{c})$ is contained in the class $\mathcal{G}_M := \{G|\text{support}(G) \subset [0, M]^p; G \ll \lambda^p\}$.

(A3) The random vector $\boldsymbol{C}$ are independent of $\boldsymbol{X}$.

(A4) There exists a positive constant η such that $P(\boldsymbol{C} \geq \eta) = 1$.

(A5) The marginal distribution functions $G_r(c_r)$ of $\boldsymbol{C}$ have the bounded continuous densities $g_r(c_r)$ on $[0, M]$.

(A6) F satisfies that for any $D \in \mathfrak{D}$, $\mu_F[A_D(c_1, ..., c_p)] \geq \epsilon > 0$ if $c_r \geq \delta > 0$, $r = 1, ..., p$.

The above conditions are similar to those required for the consistency of the maximum likelihood estimator of a failure time distribution based on interval-censored data (Groeneboom and Wellner 1992 ; Yu et al., 2006). The conditions (A4), (A5) and (A6) are required to avoid the singularity of the integral equation appearing in the information calculation. The condition (A4) means that there exists a positive censoring time. More discussion about these conditions can be found in Geskus and Groeneboom (1996, 1997).

3 NPMLE of $F(\boldsymbol{x})$

In this section, we will study the nonparametric maximum likelihood estimator of the distribution function $F(\boldsymbol{x})$ and its self-consistency equation for the random vector $\boldsymbol{X}$ with the multivariate case 1 censoring mechanism.

At first, in terms of the notation introduced in Section 2, the density function for the random vector $(\boldsymbol{C}, \boldsymbol{\Lambda})$ can be expressed as follows:

$$q_F(\boldsymbol{c}, \boldsymbol{\lambda}) = g(\boldsymbol{c}) \prod_{D \in \mathfrak{D}} \{\mu_F(A_D(\boldsymbol{c})\}^{\{\prod_{r\in D}\lambda_r \prod_{s\in \bar{D}}(1-\lambda_s)\}}.$$

Suppose that $(\boldsymbol{c}_i, \boldsymbol{\lambda}_i)(i = 1, 2, ..., n)$ are the observed sample from the random vector $(\boldsymbol{C}, \boldsymbol{\Lambda})$. Then the conditional likelihood function $L_{\boldsymbol{C}}(F)$ for $F(\boldsymbol{x})$ is

$$L_{\boldsymbol{C}}(F) = \prod_{i=1}^{n} \prod_{D\in\mathfrak{D}} \{\mu_F(A_D(\boldsymbol{c}_i))\}^{\{\prod_{r\in D}\lambda_{ir} \prod_{s\in\bar{D}}(1-\lambda_{is})\}}$$

and thus the log conditional likelihood function $l_{\boldsymbol{C}}(F)$ for the distribution function F is

$$l_{\boldsymbol{C}}(F) = \int_{\mathbb{R}_+^p\times\{0,1\}^d} \sum_{D\in\mathfrak{D}} \left\{ \prod_{r\in D} \lambda_r \prod_{s\in\bar{D}} (1-\lambda_s) \right\} \log\{\mu_F(A_D(\boldsymbol{c}))\} d\mathbb{P}_n(\boldsymbol{c}, \boldsymbol{\lambda}) \qquad (3.1)$$

where $\mathbb{P}_n(\boldsymbol{c}, \boldsymbol{\lambda})$ is the empirical probability function obtained from the observations $\{(\boldsymbol{c}_i, \boldsymbol{\lambda}_i), i = 1, ..., n\}$. Now, the NPMLE $\hat{F}_n$ of the distribution function $F(\boldsymbol{x})$ can be obtained by maximizing the function (3.1):

$$\hat{F}_n = \arg \max_{F\in\bar{\mathcal{F}}_M} l_{\boldsymbol{C}}(F)$$

where

$$\bar{\mathcal{F}}_M := \{F : |\text{support}(F) \subset [0, M]^d; F \text{ is either a continuous ditribution function} \\ \text{a piecewise constant distribution function with a nite number of jumps}\}$$

According to the idea of Geskus and Groeneboom (1996, 1997), one may define the following operator $\boldsymbol{L}_F$ for $a(\boldsymbol{x}) \in L^2(F_1)$ where F_1 is the probability measure in $\mathbb{R}_+^d$

$$[\boldsymbol{L}_F a](\boldsymbol{c}, \boldsymbol{\lambda}) = E\{a(\boldsymbol{X})|(\boldsymbol{C}, \boldsymbol{\Lambda}) = (\boldsymbol{c}, \boldsymbol{\lambda})\} = \sum_{D\in\mathfrak{D}} \left\{ \prod_{r\in D} \lambda_r \prod_{s\in\bar{D}} (1-\lambda_s) \right\} \frac{\int_{A_D(\boldsymbol{c})} a(\boldsymbol{x}) d\mu_F}{\mu_F[A_D(\boldsymbol{c})]}.$$

By applying this operator to the function $a(\boldsymbol{z}) = a(z_1, .., z_d) = I_{\{z_1 \leq x_1, ..., z_d \leq x_d\}}$ and taking expectation with respect to the probability distribution $I\!P(\boldsymbol{c}, \boldsymbol{\lambda})$ of $(\boldsymbol{C}, \boldsymbol{\Lambda})$, we have that

$$F(\boldsymbol{x}) = \int_{I\!R_+^p \times \{0,1\}^d} \sum_{D \in \mathfrak{D}} \left\{ \prod_{r \in D} \lambda_r \prod_{s \in \bar{D}} (1 - \lambda_s) \right\} \frac{\mu_F[A_D^{\boldsymbol{x}}(\boldsymbol{c})]}{\mu_F[A_D(\boldsymbol{c})]} dI\!P(\boldsymbol{c}, \boldsymbol{\lambda}). \tag{3.2}$$

Therefore replacing F and $I\!P(\boldsymbol{c}, \boldsymbol{\lambda})$ with $\hat{F}_n$ and $I\!P_n(\boldsymbol{c}, \boldsymbol{\lambda})$, respectively in (3.2), we have the self-consistency equation for the nonparametric maximum likelihood estimator $\hat{F}_n(\boldsymbol{x})$:

$$\hat{F}_n(\boldsymbol{x}) = \int_{I\!R_+^p \times \{0,1\}^d} \sum_{D \in \mathfrak{D}} \left\{ \prod_{r \in D} \lambda_r \prod_{s \in \bar{D}} (1 - \lambda_s) \right\} \frac{\mu_{\hat{F}_n}[A_D^{\boldsymbol{x}}(\boldsymbol{c})]}{\mu_{\hat{F}_n}[A_D(\boldsymbol{c})]} dI\!P_n(\boldsymbol{c}, \boldsymbol{\lambda}) \tag{3.3}$$

Now we discuss the existence and uniqueness of the NPMLE. Some researchers have developed the algorithms for the NPMLE based on the bivariate interval censored data (see Betensky and Finkelstein, 1999; Gentleman and Vandal 2001, 2002; Bogaerts and Lesaffre, 2004). A similar iterative algorithm can be developed using this self-consistency equation (3.3) to obtain the estimator $\hat{F}_n$. For the uniqueness of NPMLE, let

$$H = \{H_j = (r_{j1}, s_{j1}] \times \cdots (r_{jd}, s_{jd}], j = 1, ..., m\}$$

denote the disjoint hyperrectangles that constitute the regions of possible support of the NPMLE of F, $\alpha_{ij} = I\{H_j \subseteq A_D(\boldsymbol{C}_i)\}$ and $p_j = \mu_F(H_j), i = 1, ..., n, j = 1, ..., m$. Then the likelihood function $L_{\boldsymbol{C}}(F)$ can be rewritten as

$$L(\boldsymbol{p}) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} p_j$$

with $\boldsymbol{p} = (p_1, ..., p_m)^T$, and the NPMLE of F is determined by maximizing $L(\boldsymbol{p})$ over the p_j's subject to $p_j \geq 0$ and $\sum_{j=1}^{m} p_j = 1$. Further, if the rank of $n \times m$ matrix $A = (\alpha_{ij})$ is m, the log-likelihood function $\log L(\boldsymbol{p})$ is strictly concave and thus the vector $\boldsymbol{p}$ is unique. Therefore the NPMLE based on the multivariate case 1 interval-censored data is unique if the log-likelihood function $l_{\boldsymbol{C}}(F)$ is strictly concave.

4 Asymptotic Properties of NPMLE

In this section, we derive the asymptotic properties of NPMLE $\hat{F}_n(\boldsymbol{x})$. From the self-consistency equation (3.3) for the NPMLE $\hat{F}_n$, the following result on the strong consistency of the estimator $\hat{F}_n(\boldsymbol{x})$ can be derived. Its proof is given in Appendix.

Theorem 4.1. *Under the conditions (A1)-(A6),*

$$P\{\lim_{n\to\infty} \sup_{\boldsymbol{x}\in I\!R^d_+} |\hat{F}_n(\boldsymbol{x}) - F(\boldsymbol{x})| = 0\} = 1$$

Next we derive the convergence rate for $\hat{F}_n$ and give some comments on the rates for the different types of case 1 interval censoring mechanism. Before giving the convergence rate for $\hat{F}_n$, we introduce the measure for assessing the distance of two density functions.

The Hellinger distance between two density functions f_1 and f_2 with respect to μ is defined by

$$h(f_1, f_2) = \left(\frac{1}{2}\int(\sqrt{f_1} - \sqrt{f_2})^2 d\mu\right)^{1/2}.$$

where μ is a σ-finite dominating measure.

We now define the density function ϕ_F with respect to the dominating measure $\mu = \mu_1 \times \mu_0$ as follows.

$$\phi_F(\boldsymbol{c}, \boldsymbol{\lambda}) = \sum_{D\in\mathfrak{D}} \left\{\prod_{r\in D} \lambda_r \prod_{s\in \bar{D}} (1-\lambda_s)\right\} \mu_F(A_D(\boldsymbol{c}))$$

where μ_1 is the measure induced by the distribution of $\boldsymbol{C}$ and μ_0 is the counting measure on $\{0,1\}^d$.

We first measure the Hellinger distance between $\phi_{\hat{F}_n}$ and ϕ_F. To this end, the following additional conditions should be assumed.

(S1) The joint density of $\boldsymbol{C}$ is bounded on its support.

(S2) The joint density of $\boldsymbol{X}$ is bounded away from zero and infinity on its support.

Theorem 4.2. *Under the conditions (A1)-(A4), (S1) and (S2), the Hellinger distance $h(\phi_{\hat{F}_n}, \phi_F)$ satisfies*

$$h(\phi_{\hat{F}_n}, \phi_F) = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}}(\log n)^{\frac{\alpha d(\beta-1)}{(1+\alpha)(1+2d)}}\right) & \textit{if } \alpha > 1 \\ O_p\left(n^{-\frac{(1+d)}{(2+4d)}}(\log n)^{\frac{d^2}{2(1+2d)}}\right) & \textit{if } \alpha = 1 \end{cases}$$

where $\alpha = \max\{d_i, 1 \le i \le p\}$ *and* $\beta = \mathrm{card}\{d_i : d_i = \alpha\}$.

It is worth pointing out that (S1) implies (A5) and (S2) implies (A6). On the other hand, (A5) and (A6) imply (S1) and (S2), respectively, provided that components of $\boldsymbol{X}$ and components of $\boldsymbol{C}$ are independent of each other, respectively. Therefore we have

Corollary 4.1. Suppose that components of $\boldsymbol{X}$ and components of $\boldsymbol{C}$ are independent of each other, respectively. Then under the conditions (A1)-(A6), we have that

$$h(\phi_{\hat{F}_n}, \phi_F) = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}}(\log n)^{\frac{\alpha d(\beta-1)}{(1+\alpha)(1+2d)}}\right) & \text{if } \alpha > 1 \\ O_p\left(n^{-\frac{(1+d)}{(2+4d)}}(\log n)^{\frac{d^2}{2(1+2d)}}\right) & \text{if } \alpha = 1. \end{cases}$$

Remarks

(i) Note that if $\alpha = 1$, then $p = d$ and each X_r has its own censoring time $C_r, r = 1, ..., d$. If $\alpha = d$, then $p = 1$ and all X_r's share a common censoring time C.

(ii) In the case that $\boldsymbol{X}$ has a common censoring time, we have that $\alpha = d$, $\beta = 1$ and thus

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{(1+d)^2}{2(1+2d+3d^2)}}\right).$$

Also, if each X_r has its own censoring time, $\alpha = 1$, $\beta = d$ and thus

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{(1+d)}{(2+4d)}}(\log n)^{\frac{d^2}{2(1+2d)}}\right).$$

In particular, if $d = 1$, we have that $\alpha = 1$ and

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{1}{3}}(\log n)^{\frac{1}{6}}\right),$$

which coincides with the results in Groeneboom and Wellner (1992), Van de Geer (1996).

(iii) Since $\frac{(1+d)^2}{2(1+2d+3d^2)}$ and $\frac{1+d}{2(1+2d)}$ are the monotone decreasing functions of d, the convergence rates decrease as the dimensional number of multivariate lifetime $\boldsymbol{X}$ increases in the two extreme settings ($p = 1$ and $p = d$). Generally, the convergence rate decreases as the the dimensional number of $\boldsymbol{X}$ increases since $\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}$ is a monotone decreasing function of d for a fixed value of α. Moreover, the upper limit of convergence rates is $n^{-1/3}$. The lower limits are $n^{-1/4}$ in the situation where each X_i has its own censoring time ($p = d$), and $n^{-1/6}$ if all X_r's share a common censoring time ($p = 1$).

(iv) If the dimension d of multivariate lifetime $\boldsymbol{X}$ is fixed, $\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}$ is a monotone decreasing function of α. Since the small α results in a big number of censoring times, the bigger the number of censoring times is, the faster the convergence rate is, which is reasonable because the data with more censoring times involve much more information than the data with less common censoring times about the multivariate lifetime $\boldsymbol{X}$.

Note that Theorem 4.2 gives the convergence rate for the function $\phi_{\hat{F}_n}$ of NPMLE $\hat{F}_n$. We need to further specify the convergence rate for NPMLE $\hat{F}_n$ using the L_2-distances defined on the spaces induced by the marginal distributions of random vector

$\boldsymbol{C}$. Let

$$\widetilde{\boldsymbol{C}} = (\underbrace{C_1, ..., C_1}_{d_1}, \underbrace{C_2, ..., C_2}_{d_2}, ..., \underbrace{C_p, ..., C_p}_{d_p}).$$

For any set $D = \{i_1, i_2, ..., i_a\} \in \mathfrak{D}$, define the subvectors $\boldsymbol{C}_D$ and $\boldsymbol{X}_D$ of $\widetilde{\boldsymbol{C}}$ and $\boldsymbol{X}$, respectively, as follows:

$$\boldsymbol{C}_D = (C_{i_1}, C_{i_2}, ..., C_{i_a})$$

$$\boldsymbol{X}_D = (X_{i_1}, X_{i_2}, ..., X_{i_a}).$$

Since some components of $\boldsymbol{C}_D$ may be equal, define $\widetilde{\boldsymbol{C}}_D$ as the random vector consisting of the distinct $C_{i_k}, k = 1, 2, ..., a$ in $\boldsymbol{C}_D$. Similar to the Corollary 2 in Geskus and Groeneboom (1997), we have the following theorem.

Theorem 4.3. *Let, for multivariate distribution functions $F_1(\boldsymbol{x})$ and $F_2(\boldsymbol{x})$ in $I\!R^d_+$, and $D \in \mathfrak{D}$, $||F_1 - F_2||_{\boldsymbol{C}_D}$ denote the L_2-distance defined by*

$$||F_1 - F_2||^2_{\boldsymbol{C}_D} = \int \{F_{1\boldsymbol{X}_D}(\boldsymbol{c}_D) - F_{2\boldsymbol{X}_D}(\boldsymbol{c}_D)\}^2 dG_{\widetilde{\boldsymbol{C}}_D}(\tilde{\boldsymbol{c}}_D)$$

where $F_{\boldsymbol{X}_D}$ and $G_{\widetilde{\boldsymbol{C}}_D}$ denote the marginal distributions of subvector $\boldsymbol{X}_D$ and $\tilde{\boldsymbol{C}}_D$, respectively. Then under the same conditions as that in Theorem 4.2, for any $D \in \mathfrak{D}$, we have that as $n \to \infty$,

$$||\hat{F}_n - F||_{\boldsymbol{C}_D} = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}} (\log n)^{\frac{\alpha d(\beta-1)}{(1+\alpha)(1+2d)}}\right) & \text{if } \alpha > 1, \\ O_p\left(n^{-\frac{(1+d)}{(2+4d)}} (\log n)^{\frac{d^2}{2(1+2d)}}\right) & \text{if } \alpha = 1. \end{cases}$$

In particular,

$$||\hat{F}_n - F|| = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}} (\log n)^{\frac{\alpha d(\beta-1)}{(1+\alpha)(1+2d)}}\right) & \text{if } \alpha > 1, \\ O_p\left(n^{-\frac{(1+d)}{(2+4d)}} (\log n)^{\frac{d^2}{2(1+2d)}}\right) & \text{if } \alpha = 1. \end{cases}$$

where

$$||F_1 - F_2||^2 = \int \{F_1(\widetilde{\boldsymbol{c}}) - F_2(\widetilde{\boldsymbol{c}})\}^2 dG(\boldsymbol{c})$$

It should be emphasized that the above results are derived under strong assumptions (S1) and (S2). However, in many situations, the joint densities of $\boldsymbol{X}$ and $\boldsymbol{C}$ do not satisfy these assumptions. We are much more concerned about the convergence rate of $\hat{F}_n$ without (S1) and (S2). By examining the proof of Theorem 4.2 (see Appendix), we found that the key of evaluating the convergence rate for $\hat{F}_n$ is to estimate the values of the tail integrals $\int_{\phi_F > \sigma_n} \frac{1}{\phi_F} d\mu$ and $\int_{\phi_F \le \sigma_n} \phi_F d\mu$. However, it is difficult to assess these values if the joint density function of the multivariate lifetime $\boldsymbol{X}$ is not bounded away from zero and infinity. Therefore, we are making an effort to derive the convergence rate for specified distributions.

We now focus on a class of joint or "couple" multivariate distributions widely used in the statistical inference for dependent multivariate models, which are so-called copulas (see Nelsen (1998), Schwerzer and Sklar (1983)). In fact, a non-normal distribution function $F(\boldsymbol{x})$ of a multivariate variable $\boldsymbol{X}$ can be expressed in terms of its marginals $F_1(x_1), F_2(x_2), ..., F_d(x_d)$, and its associated dependence function A, implicitly defined through the identity $A\{F_1(x), F_2(x_2), ..., F_d(x_d)\} = F(\boldsymbol{x})$, where the mapping A, which is uniquely determined on the unit hypercube whenever $F_1, F_2, ..., F_2$ are continuous, captures the essential features of the dependence among the random variables $X_1, X_2, ..., X_d$. In the bivariate case, one family of dependence functions suitable for statistical analysis was proposed and further illustrated by Plackett (1965), Clayton (1978), Cook and Johnson (1981, 1986) and others. In their studies, analysis was restricted to situations where the dependence function was known to belong to a specific class of bivariate distributions indexed by a one- or two-dimensional parameter. This important class of copulas is Archimedean copulas, which implies that on the unit square, the appropriate dependence function has the form $A(u_1, u_2) = \psi^{[-1]}\{\psi(u_1) + \psi(u_2)\}$ for some convex decreasing function ψ defined on (0,1]. As pointed out by Genest and MacKay (1986a, 1986b), this class of dependence functions is wide and mathematically tractable, and its elements have stochastic properties that make these functions attractive for the statistical treatment of data. Naturally, Archimedean d-copulas can be easily generated in the same way as the bivariate Archimedean copula and be applied in the statistical inference of multivariate dependent variables. Therefore we concentrate the derivation of convergence rate on Archimedean d-copulas. In particular, the convergence rate of NPMLE is obtained for specified Archimedean d-copulas, called the Gumbel-Hougaard family.

Note that Gumbel-Hougaard family of d-copulas has the following dependence function $A_\theta(\boldsymbol{u}) = \exp\{-[(-\ln u_1)^\theta + (-\ln u_2)^\theta + \cdots + (-\ln u_d)^\theta]^{1/\theta}\}$ where $1 \le \theta < \infty$ (see Nelsen, 1998). Now, the following result can be derived provided that there is only one common censoring time C for multivariate lifetime $\boldsymbol{X} = (X_1, X_2, ..., X_d)$. For this purpose, we need the following assumption.

(A6$'$) The marginal densities of $\boldsymbol{X}$ are bounded away from zero and infinity on $[0, M]$.

Theorem 4.4. *Suppose that the conditions (A1)-(A5) and (A6') hold. For the Gumbel-Hougaard family of d-copulas, if* $\alpha = d$*, the Hellinger distance* $h(\phi_{\hat{F}_n}, \phi_F)$ *satisfies*

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p(n^{-\frac{(1+\gamma)(1+d)}{2(1+\gamma+d+3d\gamma)}})$$

where $\gamma = d^{1/\theta}$*. Further, we have*

$$||\hat{F}_n - F|| = O_p(n^{-\frac{(1+\gamma)(1+d)}{2(1+\gamma+d+3d\gamma)}})$$

where

$$||\hat{F}_n - F||^2 = \int \{\hat{F}_n(\underbrace{c, ..., c}_{d}) - F(\underbrace{c, ..., c}_{d})\}^2 dG(c).$$

From Theorem 4.5, we see that for the Gumbel-Hougaard copulas, the convergence rate for NPMLE $\hat{F}_n$ depends on the dependence parameter θ. Further, if $\theta = 1$ in the Gumbel-Hougaard copulas, then $X_1, ..., X_d$ are independent of each other and $\gamma = d$. Therefore from Theorem 4.5,

$$h(q_{\hat{F}_n}, q_F) = O_p(n^{-\frac{(1+d)^2}{2(1+2d+3d^2)}})$$

which coincides with that in Corollary 4.1 (see remark (ii) in Theorem 4.2). The convergence rate of $\hat{F}_n$ increases as θ increases and attains its maximum order for $\theta = \infty$ because the parameter θ measures the association among the variables $X_1, X_2, ..., X_d$, and the big value of θ shows the strong associations and accelerates the convergence rate. Also note that for the situation where $\theta = \infty$, $\gamma = 1$ and the convergence rate is $O(n^{-\frac{(1+d)}{2(1+2d)}})$, which is the same as that in the case where $X_1, X_2, ..., X_d$ have their own censoring times $C_1, C_2, ..., C_d$, respectively.

It is worth pointing out that the convergence rate of NPMLE in the dependent case is derived only for Gumbel-Hougaard copulas. Of course, one can also derive the convergence rate for other specific copulas such as Clayton family and Frank family of d-copulas. But it is of much concern to know the convergence rates for the general Archimedean copulas, even for the general copulas. It is not realistic to obtain the convergence rate of the NPMLE for general multivariate distribution using the current approaches.

5 Concluding Remarks

This paper studied the asymptotic properties of the nonparametric maximum likelihood estimator of the distribution function $F(\boldsymbol{x})$ when the multivariate lifetimes of interest are type 1 interval-censored. Firstly, we described the general current status data for the multivariate lifetimes, in which, the vector of multivariate variables of interest is partitioned into several subvectors, each subvector has its own censoring time. For the multivariate current status data, we introduced some notation, which makes the derivation of asymptotic properties much more clear and concise. Secondly, we established the self-consistency equation and strong consistency for the NPMLE. Moreover, we derived the convergence rates of the NPMLE in two situations and drawn the following conclusions: (i) the convergence rate decreases as the dimensional number of multivariate lifetimes increases or as the number of censoring times decreases provided that $\boldsymbol{X}$ and $\boldsymbol{C}$ have the densities which are bounded away from zero and infinity; (ii) the strong association of the random variables accelerates the convergence rate of NPMLE in the situation where the joint distribution of multivariate variables follows a specified family of Archimedean d-copulas.

Many studies about the multivariate current status data remain to be done. For example, the problems on assessing the association and testing independence among lifetimes $X_1, ..., X_d$ based on the multivariate interval-censored data need to be studied. Further, we are interested in whether the result in Ding and Wang (2004) can be extended with the multivariate current status data. Moreover, it should be pointed out that in one-parameter Archimedean copulas, the association among the random variables can be assessed by estimating the dependent parameter. As we saw in Section 4, in the Gumbel-Hougaard family of copulas the independence is equivalent to the property that the association parameter θ equals zero. We may test the independence in this family of copulas by assessing the dependent parameter.

Another question related to the asymptotic properties of the NPMLE is the asymptotic distribution theory. For instance, does the asymptotic normality hold for the NPMLE $\hat{F}_n$? More generally, we are interested in the asymptotic optimal estimation of smooth functionals for multivariate current status lifetime data. That is, for the functional K of F, does the following hold?

$$\sqrt{n}(K(\hat{F}_n) - K(F)) \to N(0, \sigma_F^2) \qquad \text{in distribution}$$

If so, what is the expression of asymptotic variance σ_F^2?

A Appendix

Now we are in the position to prove the theorems. For the convenience, we assume that A_0 is a positive constant that varies from line to line in the sequel.

The proof of Theorem 4.1

At first note that $l_{\boldsymbol{C}}(F)$ is maximized at $\hat{F}_n$. We have that

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1}\{l_{\boldsymbol{C}}[(1-\epsilon)\hat{F}_n + \epsilon F] - l_{\boldsymbol{C}}(\hat{F}_n)\} \le 0.$$

Thus it follows from the marginal log likelihood function

$$\int_{\mathbb{R}^p \times \{0,1\}^d} \sum_{D \in \mathfrak{D}} \{\prod_{r \in D} \lambda_r \prod_{s \in \bar{D}} (1-\lambda_s)\} \frac{\mu_F(A_D(\boldsymbol{c}))}{\mu_{\hat{F}_n}(A_D(\boldsymbol{c}))} d\mathbb{P}_n(\boldsymbol{c}, \boldsymbol{\lambda}) \le 1$$

Now using the strong law of large number one can show that $\mathbb{P}_n$ converges to $\mathbb{P}$ almost surely. For a fixed $\omega \in \Omega$, by the Helly Compactness Theorem, the sequence of functions $\hat{F}_n(\cdot, \omega)$ has a subsequence $\hat{F}_{n_k}(\cdot, \omega)$ which converges vaguely to a distribution function F'. Then the proof is completed if $F' = F$.

In fact, similar to the proof of Lemma 4.3 in Groeneboom and Wellner (1992), it can be shown that

$$\int_{I\!R^p\times\{0,1\}^d}\sum_{D\in\mathfrak{D}}\{\prod_{r\in D}\lambda_r\prod_{s\in\bar{D}}(1-\lambda_s)\}\frac{\mu_F(A_D(\boldsymbol{c}))}{\mu_{F'}(A_D(\boldsymbol{c}))}dI\!P(\boldsymbol{c},\boldsymbol{\lambda})=\int_{I\!R^p}\sum_{D\in\mathfrak{D}}\frac{\{\mu_F(A_D(\boldsymbol{c}))\}^2}{\mu_{F'}(A_D(\boldsymbol{c}))}dG(\boldsymbol{c})\leq 1.$$

On the other hand, from the following facts that for real numbers $0<a_i<1, 0<b_i<1; i=1,2,...,k$ with conditions $\sum_{i=1}^k a_i=1$ and $\sum_{i=1}^k b_i=1$,

$$\sum_{i=1}^k\frac{a_i^2}{b_i}\begin{cases}=1 & \text{if } a_i=b_i \text{ for all } i\\ >1 & \text{otherwise}\end{cases}$$

and that $\sum_{D\in\mathfrak{D}}\mu_F(A_D(\boldsymbol{c}))=\sum_{D\in\mathfrak{D}}\mu'_F(A_D(\boldsymbol{c}))=1$ (see Section 2), the above expression is larger than one unless $F'=F$. Therefore we have that for all $\boldsymbol{x}$, $F'(\boldsymbol{x})=F(\boldsymbol{x})$. The proof is complete.

In order to prove Theorem 4.2, we need the following results. We first give the definition of ϵ-covering number.

Definition A.1. (see Van de Geer, 1996) Let Q be a measure on $(\mathcal{X},\mathcal{A})$, and $\mathcal{G}\subset\mathcal{L}_2(Q)$. For each $\epsilon>0$, the ϵ-covering number $N(\epsilon,\mathcal{G},Q)$ is defined as the number of balls with radius ϵ, necessary to cover $\mathcal{G}$. Formally,

$$\begin{aligned}N(\epsilon,\mathcal{G},Q)=\min\{J:\ &\text{there exist } \{g_j\}_{j=1}^J \text{ such that for all}\\ &g\in\mathcal{G},\ \min_{j\in\{1,...,J\}}\int(g-g_j)^2dQ\leq\epsilon^2\}.\end{aligned}$$

Consider a probability space $(\Omega,\mathfrak{B},P)$ and independent identically distributed random variables $Y_1,Y_2,...,Y_n$, with distribution Φ. Suppose that

$$\phi_0=\frac{d\Phi}{d\mu}\in\mathcal{P}$$

where μ is a σ-finite dominating measure, and $\mathcal{P}$ is a class of densities with respect to μ. A maximum likelihood estimator $\hat{\phi}_n$ of ϕ_0 satisfies

$$\hat{\phi}_n\in\arg\max_{\phi\in\mathcal{P}}\sum_{i=1}^n\log\phi(Y_i).$$

Let P_n be the measures induced by the empirical distribution of the sample $\{Y_i, i=1,...n\}$. By using Theorem 1.1 and Theorem 2.2 in Van de Geer (1996), we have the following proposition.

Proposition A.1. *Suppose* $\mathcal{K}=\{k(\cdot,\boldsymbol{x}):\boldsymbol{x}\in I\!R_+^d\}$ *and* $\mathcal{P}=\overline{conv}(\mathcal{K})$. *Assume that for some sequences* $1\leq\rho_n\uparrow\infty$ *and* $0\leq\sigma_n\downarrow 0$,

$$\int_{\phi_0>\sigma_n}\frac{K^2}{\phi_0}d\mu\leq\rho_n^2, n=1,2,...,$$

where $K = sup_{k\in\mathcal{K}} k$ *and for* $\tilde{\mathcal{K}} = \left\{ \left(\frac{k(\cdot,\boldsymbol{x})}{\phi_0} \right) I\{\phi_0 > \sigma_n\}, \boldsymbol{x} \in I\!\!R_+^d \right\}$, *we have that*

$$\lim_{A\to\infty} \limsup_{n\to\infty} P\left(\sup_{\delta>0} \left(\frac{\delta}{\rho_n} \right)^w N(\delta, \tilde{\mathcal{K}}, P_n) > A \right) = 0,$$

for some $0 < w < \infty$. *Then for* $\tau_n \geq 0$ *satisfying*

$$\tau_n^2 \geq \int_{\phi_0 \leq \sigma_n} \phi_0 d\mu, n = 1, 2, ...,$$
$$\tau_n \geq n^{-(2+w)/(4+4w)} \rho_n^{w/(2+2w)}, n = 1, 2, ...,$$

we have that

$$\lim_{L\to\infty} \limsup_{n\to\infty} P(h(\hat{\phi}_n, \phi_0) \geq L\tau_n) = 0.$$

Lemma A.1. *Let* $f(x_1, ..., x_m) = x_1^k \cdots x_m^k$ *be the function defined in the unit hypercube* $[0,1]^m$ *of* $I\!\!R_+^m$ *with integer* k *and* $0 \leq \sigma_n \downarrow 0$. *Then we have*

$$\int_{f\geq\sigma_n} f^{-1} dx_1 \cdots dx_m \leq \begin{cases} A_0 \sigma_n^{-1+1/k} \left(\log \frac{1}{\sigma_n} \right)^{m-1} & \text{if } k > 1 \\ A_0 \left(\log \frac{1}{\sigma_n} \right)^m & \text{if } k = 1 \end{cases}$$

and

$$\int_{f\leq\sigma_n} f dx_1 \cdots dx_m \leq A_0 \sigma_n^{1+1/k} \left(\log \frac{1}{\sigma_n} \right)^{m-1}$$

The proof of Theorem 4.2

Note that the density ϕ_F of $(\boldsymbol{C}, \boldsymbol{\Lambda})$, with respect to the dominating measure $\mu = \mu_1 \times \mu_0$ is then in the class

$$\mathcal{P} = \{\phi_\nu(\boldsymbol{c}, \boldsymbol{\lambda}) = \sum_{D\in\mathfrak{D}} \left(\prod_{r\in D} \lambda_r \prod_{s\in\bar{D}} (1-\lambda_s) \right) \mu_\nu(A_D(\boldsymbol{c})) : \nu \in \bar{\mathcal{F}}_M\}$$

where μ_ν is the measure induced by the distribution function ν in $\bar{\mathcal{F}}_M$. Clearly, $\mathcal{P} = \overline{conv}(\mathcal{K})$, with

$$\mathcal{K} = \{k_{\boldsymbol{x}}(\boldsymbol{c}, \boldsymbol{\lambda}) = \sum_{D\in\mathfrak{D}} \left(\prod_{r\in D} \lambda_r \prod_{s\in\bar{D}} (1-\lambda_s) \right) I_{C_D(\boldsymbol{x})}(\boldsymbol{c}) : \boldsymbol{x} \in [0, M]^d\}.$$

Under the condition (S1), one can verify that for any $\boldsymbol{x}$ and $\boldsymbol{y}$ in $[0, M]^d$,

$$\int (k_{\boldsymbol{x}}(\boldsymbol{c}, \boldsymbol{\lambda}) - k_{\boldsymbol{y}}(\boldsymbol{c}, \boldsymbol{\lambda}))^2 d\mu \leq A_0(\sum_{i=1}^{d} |x_i - y_i|).$$

Therefore there exists a constant C_0 such that for any probability measure Q on $I\!R_+^p \times \{0,1\}^d$,

$$N(\epsilon, \mathcal{K}, Q) \leq C_0 \epsilon^{-2d}, \quad \text{for all } \epsilon > 0 \tag{A.1}$$

where $N(\epsilon, \mathcal{K}, Q)$ is the ϵ-covering number of $(\mathcal{K}, Q)$. For some sequences $1 \leq \rho_n \uparrow \infty$ and $0 \leq \sigma_n \downarrow 0$, and for $\tilde{\mathcal{K}} = \left\{ \left(\frac{k_{\boldsymbol{x}}}{\phi_F} \right) I\{\phi_F > \sigma_n\}, \boldsymbol{x} \in I\!R_+^d \right\}$, application of (A.1), with $dQ = ((1/\phi_F^2) I\{\phi_F > \sigma_n\} dI\!P_n/(A^2 \rho_n^2)$, give

$$N(\epsilon, \tilde{\mathcal{K}}, I\!P_n) \leq A^{2d} C_0 \left(\frac{\rho_n}{\epsilon} \right)^{2d}, \quad \text{for all } \epsilon > 0,$$

on the set $\left\{ \int_{\phi_F > \sigma_n} \frac{1}{\phi_F^2} dI\!P_n \leq A^2 \rho_n^2 \right\}$. So, for $\int_{\phi_F > \sigma_n} 1/\phi_F d\mu \leq \rho_n^2$, we have that

$$\begin{aligned}
&\limsup_{n \to \infty} P \left(\left(\frac{\epsilon}{\rho_n} \right)^{2d} N(\epsilon, \tilde{\mathcal{K}}, I\!P_n) > A^{2d} C_0 \right) \\
&\qquad \leq \limsup_{n \to \infty} P \left(\int_{\phi_F > \sigma_n} \frac{1}{\phi_F^2} dI\!P_n > A^2 \rho_n^2 \right) \\
&\qquad \leq \limsup_{n \to \infty} P \left(\int_{\phi_F > \sigma_n} \frac{1}{\phi_F} d\mu_n > A^2 \rho_n^2 \right) \to 0, \quad \text{as } A \to \infty,
\end{aligned}$$

where $d\mu_n = dI\!P_n/\phi_F \to d\mu$ as $n \to \infty$.

Next we derive the expressions for $\{\rho_n\}$ and $\{\tau_n\}$. Note that under the condition (S2), we have that

$$\mu_F[A_D(\boldsymbol{c})] = \int_{A_D(\boldsymbol{c})} d\mu_F(\boldsymbol{x}) \leq A_0 \int_{A_D(\boldsymbol{c}) \bigcap [0,M]^d} d\lambda^d(\boldsymbol{x}) \leq A_0 \prod_{r=1}^{p} c_r^{\alpha_r} (M - c_r)^{\beta_r}.$$

Therefore,

$$\begin{aligned}
\phi_F(\boldsymbol{c}, \boldsymbol{\lambda}) &= \sum_{D \in \mathfrak{D}} \left(\prod_{r \in D} \lambda_r \prod_{s \in \bar{D}} (1 - \lambda_s) \right) \mu_F[A_D(\boldsymbol{c})] \\
&\leq A_0 \sum_{D \in \mathfrak{D}} I_{\Omega(D)}(\boldsymbol{\lambda}) \left(\prod_{r \in D} c_{k_r}^{\alpha_r} \prod_{s \in \bar{D}} (M - c_{k_s})^{\beta_s} \right)
\end{aligned}$$

where $\Omega(D) = \{\lambda : \lambda_r = 1 \text{ for } r \in D; \lambda_s = 0 \text{ for } s \in \bar{D}\}$. Similarly, we have

$$\phi_F(\boldsymbol{c}, \boldsymbol{\lambda}) \geq A_0 \sum_{D \in \mathfrak{D}} I_{\Omega(D)}(\boldsymbol{\lambda}) \left(\prod_{r \in D} c_{k_r}^{\alpha_r} \prod_{s \in \bar{D}} (M - c_{k_s})^{\beta_s} \right)$$

Note that for $\alpha = \max\{d_i : i = 1, ..., p\}$, $\beta = \text{card}\{d_i : d_i = \alpha\}$ and for $0 \leq \sigma_n \downarrow 0$, regardless of the order of variables $c_1, c_2, ..., c_p$, $\phi_F^{-1}(\boldsymbol{c}, \boldsymbol{\lambda})$ is dominated by $(c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha)^{-1}$ on the set $\{\phi_F > \sigma_n\}$ and $\phi_F(\boldsymbol{c}, \boldsymbol{\lambda})$ is dominated by $c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha$ on the set $\{\phi_F \leq \sigma_n\}$, respectively. Therefore from Lemma A.1 and (S1), we have that for $0 \leq \sigma_n \downarrow 0$,

$$\begin{aligned}
&\int_{\phi_F > \sigma_n} \frac{1}{\phi_F} d\mu \leq A_0 \int_{(c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha)^{-1} > C_0 \sigma_n} (c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha)^{-1} g(\boldsymbol{c}) d\boldsymbol{c} \\
&\leq A_0 \int_{(c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha)^{-1} > C_0 \sigma_n} (c_1^\alpha c_2^\alpha \cdots c_\beta^\alpha)^{-1} dc_1 dc_2 \cdots dc_{beta} \\
&\leq \begin{cases} A_0 \sigma_n^{-1+1/\alpha} \left(\log \frac{1}{\sigma_n}\right)^{\beta - 1} & \text{if } \alpha > 1 \\ A_0 \left(\log \frac{1}{\sigma_n}\right)^d & \text{if } \alpha = 1 \end{cases}
\end{aligned}$$

and similarly,

$$\int_{\phi_F \leq \sigma_n} \phi_F d\mu \leq A_0 \sigma_n^{1+1/\alpha} \left(\log \frac{1}{\sigma_n}\right)^{\beta - 1}.$$

Now based on Proposition A.1, we estimate the order of τ_n. Note that for $\alpha = 1$ and $w = 2d$, we have that $\rho_n = (\log \frac{1}{\sigma_n})^{d/2}$ and thus

$$\tau_n = \sigma_n^{1+1/\alpha} (\log \frac{1}{\sigma_n})^{d-1} \geq n^{-\frac{2+2d}{4+8d}} \rho_n^{\frac{2d}{2+4d}} \geq n^{-\frac{2+2d}{4+8d}} (\log \frac{1}{\sigma_n})^{\frac{d}{2}\frac{2d}{2+4d}} \approx n^{-\frac{1+d}{2(1+2d)}} (\log \frac{1}{\sigma_n})^{\frac{d^2}{2(1+2d)}}.$$

Further from the above expression, we have $\log(\frac{1}{\sigma_n}) \approx \log n$. Hence,

$$\tau_n \approx n^{-\frac{1+d}{2(1+2d)}} (\log n)^{\frac{d^2}{2(1+2d)}}.$$

Next for $\alpha > 1$ and $w = 2d$ we have

$$\tau_n = \sigma_n^{1+1/\alpha} (\log \frac{1}{\sigma_n})^{\beta - 1} \geq n^{-\frac{2+2d}{4+8d}} \rho_n^{\frac{2d}{2+4d}} = n^{-\frac{1+d}{2(1+2d)}} [\sigma_n^{-1+1/\alpha} (\log \frac{1}{\sigma_n})^{\beta-1}]^{\frac{1}{2}\frac{d}{1+2d}}.$$

Also it can be derived from the above expression that $\sigma_n \approx n^{-\frac{\alpha(1+d)}{(1+\alpha+d+3\alpha d)}}$, $\log \frac{1}{\sigma_n} \approx \log n$ and thus

$$\tau_n \approx n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}} (\log n)^{-\frac{d(\beta-1)}{2(1+2d)}}.$$

Therefore we have

$$h(\phi_n,\phi_F)=\begin{cases} O_p(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}}(\log n)^{\frac{\alpha d(\beta-1)}{(1+\alpha)(1+2d)}}) & \text{if } \alpha>1\\ O_p(n^{-\frac{(1+d)}{(2+4d)}}(\log n)^{\frac{d^2}{2(1+2d)}}) & \text{if } \alpha=1,\end{cases}$$

where $\phi_n=\arg\max_{\phi\in\mathcal{P}}\sum_{i=1}^n\log\{\phi(\boldsymbol{c}_i,\boldsymbol{\lambda}_i)\}$. Since

$$\begin{aligned}\max_{\phi\in\mathcal{P}}\sum_{i=1}^n\log\{\phi(\boldsymbol{c}_i,\boldsymbol{\lambda}_i)\}&=\max_{\nu\in\bar{\mathcal{F}}_M}\sum_{i=1}^n\log\{\phi_\nu(\boldsymbol{c}_i,\boldsymbol{\lambda}_i)\}\\ &=\max_{\nu\in\bar{\mathcal{F}}_M}\sum_{i=1}^n\sum_{D\in\mathfrak{D}}\left(\prod_{r\in D}\lambda_{ir}\prod_{s\in\bar{D}}(1-\lambda_{is})\right)\log\{\mu_\nu(A_D(\boldsymbol{c}_i))\},\end{aligned}$$

we have $\phi_n=\phi_{\hat{F}_n}$. The proof is complete.

The proof of Theorem 4.4

From the definition of ϕ_F, for any subset D of $\boldsymbol{D}$, we have that

$$\begin{aligned}&\int_{\mathbb{R}_+^p}\left(\mu_{\hat{F}_n}[A_D(\boldsymbol{c})]-\mu_F[A_D(\boldsymbol{c})]\right)^2dG(\boldsymbol{c})\le 4\int_{\mathbb{R}_+^p}\left(\sqrt{\mu_{\hat{F}_n}[A_D(\boldsymbol{c})]}-\sqrt{\mu_F[A_D(\boldsymbol{c})]}\right)^2dG(\boldsymbol{c})\\ &=4\int_{\mathbb{R}_+^p\times\{0,1\}^d}\left(\sqrt{\prod_{r\in D}\lambda_r\prod_{s\in\bar{D}}(1-\lambda_s)\mu_{\hat{F}_n}[A_D(\boldsymbol{c})]}-\sqrt{\prod_{r\in D}\lambda_r\prod_{s\in\bar{D}}(1-\lambda_s)\mu_F[A_D(\boldsymbol{c})]}\right)^2d\mu\\ &\le 4h^2(\phi_{\hat{F}_n},\phi_F)\end{aligned}\tag{A.2}$$

For any subvector $\boldsymbol{C}_D=(C_{i_1},...,C_{i_a})$ of $\tilde{\boldsymbol{C}}$ with $D=\{i_1,...,i_a\}$, we have the marginal distribution $F_{\boldsymbol{X}_D}(\boldsymbol{c}_D)$ of $\boldsymbol{X}_D$=$(X_{i_1},...,X_{i_a})$ as

$$F_{\boldsymbol{X}_D}(\boldsymbol{c}_D)=F_{\boldsymbol{X}_D}(c_{i_1},...,c_{i_a})=\mu_F\left[\bigotimes_{r=1}^a[0,c_{i_r})\times[0,M]^{d-a}\right].\tag{A.3}$$

According to the definition of $A_D(\boldsymbol{c})$, there exists a subset $\mathfrak{D}'$ of $\mathfrak{D}$ such that

$$\begin{aligned}\bigotimes_{r=1}^a[0,c_{i_r})\times[0,M]^{d-a}&=\bigcup_{D'\in\mathfrak{D}'}\bigotimes_{r\in D'}[0,c_r)\bigotimes_{s\in\bar{D}'}[c_s,M]\\ &=\bigcup_{D'\in\mathfrak{D}'}\bigotimes_{r=1}^p\{[0,c_r)^{\alpha'_r}\times[c_r,M]^{\beta'_r}\}=\bigcup_{D'\in\mathfrak{D}'}A_{D'}(\boldsymbol{c})\end{aligned}\tag{A.4}$$

here $\alpha'_r=\text{card}\{D'\cap D_r\}$ and $\beta'_r=\text{card}\{\bar{D}'\cap D_r\}$. Therefore (A.2), (A.3), (A.4) and triangle inequality imply that

$$
\begin{aligned}
&||\hat{F}_n - F||_{C_D} = \left\{ \int_{\mathbb{R}_+^{a'}} \left[\hat{F}_{n\boldsymbol{X}_D}(\boldsymbol{c}_D) - F_{\boldsymbol{X}_D}(\boldsymbol{c}_D) \right]^2 dG_{\widetilde{C}_D}(\tilde{\boldsymbol{c}}_D) \right\}^{1/2} \\
&= \left\{ \int_{\mathbb{R}_+^{a'}} \left[\mu_{\hat{F}_n} \left(\bigotimes_{r=1}^{a} [0, c_{i_r}) \times [0, M]^{d-a} \right) - \mu_F \left(\bigotimes_{r=1}^{a} [0, c_{i_r}) \times [0, M]^{d-a} \right) \right]^2 dG_{\widetilde{C}_D}(\tilde{\boldsymbol{c}}_D) \right\}^{1/2} \\
&= \left\{ \int_{\mathbb{R}_+^{p}} \left[\mu_{\hat{F}_n} \left(\bigcup_{D' \in \mathfrak{D}'} A_{D'}(\boldsymbol{c}) \right) - \mu_F \left(\bigcup_{D' \in \mathfrak{D}'} A_{D'}(\boldsymbol{c}) \right) \right]^2 dG(\boldsymbol{c}) \right\}^{1/2} \\
&= \left\{ \int_{\mathbb{R}_+^{p}} \left[\sum_{D' \in \mathfrak{D}'} \mu_{\hat{F}_n}[A_{D'}(\boldsymbol{c})] - \sum_{D' \in \mathfrak{D}'} \mu_F[A_{D'}(\boldsymbol{c})] \right]^2 dG(\boldsymbol{c}) \right\}^{1/2} \\
&\le \sum_{D' \in \mathfrak{D}'} \left\{ \int_{\mathbb{R}_+^{p}} \left(\mu_{\hat{F}_n}[A_{D'}(\boldsymbol{c})] - \mu_F[A_{D'}(\boldsymbol{c})] \right)^2 dG(\boldsymbol{c}) \right\}^{1/2} \\
&\le 4 \times 3^d h(\phi_{\hat{F}_n}, \phi_F), \qquad\qquad \text{(A.5)}
\end{aligned}
$$

where a' is the number of distinct variables in the subvector $\boldsymbol{C}_D = (C_{i_1}, ..., C_{i_a})$.

Now, Theorem 4.4 follows from (A.5) and Theorem 4.2.

The proof of Theorem 4.5

In the light of proof of Theorem 4.2, it suffices to evaluate the tail integrals $\int_{\phi_F > \sigma_n} \phi_F^{-1} d\mu$ and $\int_{\phi_F \le \sigma_n} \phi_F d\mu$ for the family of the Gumbel-Hougaard copulas. Since $\alpha = d$, $\boldsymbol{X}$ has a common censoring time C. Now define

$$\widehat{\boldsymbol{C}} = (\underbrace{C, C, ..., C}_{d}).$$

Then for any $D = \{i_1, i_2, ..., i_a\}$, define the subvectors $\boldsymbol{C}_D$ and $\boldsymbol{X}_D$ of $\widehat{\boldsymbol{C}}$ and $\boldsymbol{X}$, respectively, as follows:

$$\boldsymbol{C}_D = (C_{i_1}, C_{i_2}, ..., C_{i_a})$$
$$\boldsymbol{X}_D = (X_{i_1}, X_{i_2}, ..., X_{i_a})$$

where all C_{i_k}'s are equal $(= C)(k = 1, 2, ..., a)$. According to the proof of Theorem 4.2, the density ϕ_F has the following form

$$\phi_F(\boldsymbol{c}, \boldsymbol{\lambda}) = \sum_{D \in \mathfrak{D}} \left(\prod_{r \in D} \lambda_r \prod_{s \in \bar{D}} (1 - \lambda_s) \right) \mu_F(A_D(\boldsymbol{c})).$$

Since all X_r's share a common censoring time C, for $D = \{i_1, ..., i_a\}$

$$
\begin{aligned}
&A_D(\boldsymbol{c}) = [0, c)^a \times [c, M]^{d-a} \\
&= [0, c]^a \times \sum_{i=0}^{d-a} (-1)^i \binom{d-a}{i} [0, c)^i \times [0, M]^{d-a-i} = \sum_{i=0}^{d-a} (-1)^i \binom{d-a}{i} [0, c)^{a+i} \times [0, M]^{d-a-i}.
\end{aligned}
$$

For the the Gumbel-Hougaard copula model, the distribution function F of $\boldsymbol{X}$ has the form

$$F(x_1, x_2, ..., x_d) = \exp\{-[(-\ln F_1(x_1))^\theta + (-\ln F_2(x_2))^\theta + \cdots + (-\ln F_d(x_d))^\theta]^{1/\theta}\}$$

and thus under (A6'),

$$\mu_F(A_D(\boldsymbol{c})) = \mu_F([0,c)^a \times [c,M]^{d-a}) = \sum_{i=0}^{d-a}(-1)^i\binom{d-a}{i}\mu_F([0,c)^{a+i}\times[0,M]^{d-a-i})$$
$$= \sum_{i=0}^{d-a}(-1)^i\binom{d-a}{i}\exp\left\{-\left[\sum_{k=1}^{a+i}(-\log F_k(c))^\theta + \sum_{k=a+i+1}^{d}(-\log F_k(M))^\theta\right]^{1/\theta}\right\}$$
$$\le \sum_{i=0}^{d-a}(-1)^i\binom{d-a}{i}\exp\left\{-\left[\sum_{k=1}^{a+i}(-\log(A_0 c))^\theta\right]^{1/\theta}\right\} \le A_0\sum_{i=0}^{d-a}(-1)^i\binom{d-a}{i}c^{(a+i)^{1/\theta}}.$$

Hence,

$$\phi_F(\boldsymbol{c},\lambda) \le A_0\sum_{D\in\mathfrak{D}}\left(\prod_{r\in D}\lambda_r\prod_{s\in\bar{D}}(1-\lambda_s)\right)\left(\sum_{i=0}^{d-a_D}(-1)^i\binom{d-a_D}{i}c^{(a_D+i)^{1/\theta}}\right)$$

where $a_D = \text{card}\{D\}$. Similarly, we have that

$$\phi_F(\boldsymbol{c},\lambda) \ge A_0\sum_{D\in\mathfrak{D}}\left(\prod_{r\in D}\lambda_r\prod_{s\in\bar{D}}(1-\lambda_s)\right)\left(\sum_{i=0}^{d-a_D}(-1)^i\binom{d-a_D}{i}c^{(a_D+i)^{1/\theta}}\right).$$

Now for $0 \le \sigma_n \downarrow 0$, ϕ_F^{-1} is dominated by $c^{-\gamma}$ on the set $\{\phi_F > \sigma_n\}$ and ϕ_F is dominated by c^γ on the set $\{\phi_F \le \sigma_n\}$, respectively and thus

$$\int_{\phi_F>\sigma_n}\phi_F^{-1}d\mu \le A_0\int_{c^\gamma>\sigma_n}c^{-\gamma}dc \le A_0\sigma_n^{-1+1/\gamma} \tag{A.6}$$

and

$$\int_{\phi_F\le\sigma_n}\phi_F d\mu \le A_0\int_{c^\gamma\le\sigma_n}c^{\gamma}dc \le A_0\sigma_n^{1+1/\gamma} \tag{A.7}$$

where $\gamma = d^{1/\theta}$. Therefore by using Proposition A.1, Theorem 4.5 follows from (A.6) and (A.7).

Acknowledgments

This work is partly supported by a grant from the Natural Science and Engineering Research Council of Canada.

References

Betensky, R. A. and D. M. Finkelstain, D. M. (1999). A non-parametric maximum likelihood estimator for bivariate interval censored data. Statistics in Medicine 18, 3089-3100.

Bogaerts, K. and Lesaffre, E. (2004). A new, fast algorithm to find the regions of possible support for bivariate interval-censored data. Journal of Computational Graphical Statistics 13, 330-340.

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

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modeling nonelliptically symmetric multivariate data. Journal of Royal Statistical Society, Ser. B, 43, 210-218.

Cook, R. D. and Johnson, M. E. (1986). Generalized Burr-Parero-logistic distributions with applications to a uranium exploration data set. Technometrics 28, 123-131.

Ding, A and Wang, W. (2004). Testing independence for bivariate current status data. JASA. 99, 145-155.

Genest, C and MacKay, R. J. (1986a). Copules Archimédiennes et familles de lois bidimensionnelles dont les marges sont données. The Canadian Journal of Statistics 14, 145-159.

Genest, C and MacKay, R. J. (1986b). The joy of copulas: bivariate distributions with uniform marginals. The American Statistician 40, 280-283.

Gentleman, R. and Vandal, A. C. (2001). Computational algorithms for censored data problems using intersection graphs. Journal of Computational Graphical Statistics 10, 403-421.

Gentleman, R. and Vandal, A. C. (2002). Nonparametric estimation of the bivariate CDF for arbitrary censored data. The Canadian Journal of Statistics 30, 557-571.

Geskus, R. B. and P. Groeneboom, P. (1996). Asymptotically optional estimation of smooth functionals for interval censoring, Part 1. Statist. Neerlandica 50, 69-88.

Geskus, R. B. and P. Groeneboom, P. (1997). Asymptotically optional estimation of smooth functionals for interval censoring, Part 2. Statist. Neerlandica 51, 201-219.

Groeneboom, P. and Wellner, J. A. (1992). Information Bounds and Nonparametric Maximum Likelihood Estimation. DMV Seminer, Band 19, Birkhäuser, Basel, etc.

Nelsen, R. B. (1998). An Introduction to Copulas; Lecture Notes in Statistics 139, Springer-Verlag.

Plackett,R. L. (1965). A Class of Bivariate Distributions. JASA. 60, 516-522.

Schweizer, B. and Sklar, A. (1983). Probabilistic Metric Spaces, Amsterdam: North-Holland.

Sun, J. (2006). The Statistical Analysis of Interval-censored Failure Time Data, Springer, New York.

Van de Geer, S. (1996). Rates of convergence for the maximum likelihood estimator in mixture models. J. Nonparametric Statist. 6, 293-310.

Wang, W. and Ding, A. (2000). On assessing the association for bivariate current status data. Biometrika 87, 879-893.

Wang, A. and Wells, M. T. (1997). Nonparametric estimators of the bivariate survival function under simplified censoring conditions. Biometrika, 84, 863-880.

Wang, A. and Wells, M. T. (2000). Model selection and semiparametric inference for bivariate failure-time data. JASA. 94, 62-72.

Wong, G. Y. C. and Q. Q. Yu, Q. Q. (1999). Generalized MLE of a joint function with multivariate interval-censored data. Journal of Multivariate Analysis 69, 155-166.

Yu, S. H., Yu, Q. Q. and Wong, G. Y. C. (2006). Consistency of the generalized MLE of a joint distribution function with multivariate interval-censored data, Journal of Multivariate Analysis, 97, 720-732.