Journal of Multivariate Analysis 100 (2009) 1802–1815

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Asymptotics for non-parametric likelihood estimation with doubly censored multivariate failure times

Dianliang Deng [a,*], Hong-Bin Fang [b]

[a] *Department of Mathematics and Statistics, University of Regina, SK., S4S 0A2, Canada*
[b] *Division of Biostatistics, University of Maryland Greenebaum Cancer Center, MSTF Suite 261, 10 South Pine Street, Baltimore, MD 21201, USA*

ARTICLE INFO

Article history:
Received 16 July 2008
Available online 25 February 2009

AMS 2000 subject classifications:
62G20
62N01
62H12

Keywords:
Multivariate doubly interval-censored
Non-parametric maximum likelihood estimation
Strong consistency
Convergence rate

ABSTRACT

This paper considers non-parametric estimation of a multivariate failure time distribution function when only doubly censored data are available, which occurs in many situations such as epidemiological studies. In these situations, each of multivariate failure times of interest is defined as the elapsed time between an initial event and a subsequent event and the observations on both events can suffer censoring. As a consequence, the estimation of multivariate distribution is much more complicated than that for multivariate right- or interval-censored failure time data both theoretically and practically. For the problem, although several procedures have been proposed, they are only ad-hoc approaches as the asymptotic properties of the resulting estimates are basically unknown. We investigate both the consistency and the convergence rate of a commonly used non-parametric estimate and show that as the dimension of multivariate failure time increases or the number of censoring intervals of multivariate failure time decreases, the convergence rate for non-parametric estimate decreases, and is slower than that with multivariate singly right-censored or interval-censored data.

1. Introduction

Recently the analysis of multivariate, multilevel or clustered failure time data has been of interest because these failure time data arise in various fields such as epidemiology, biomedicine, demography and reliability study. For instance, in the biomedical studies, researchers often undertake lifetime analysis in matched-paired case control studies, studies of time to occurrence of a disease to paired organs, and the examination of duration times of critical stages of multistage disease process.

On the other hand the doubly interval-censored data come up in many disease progression or epidemiological studies. By doubly interval-censored data, it is usually meant that the variable of interest is the time between the infection and the onset of certain disease, and both infection and onset cannot often be directly observed. Only can the information about whether the occurrence of each event lies in the time interval of two consecutive monitoring times be available. The typical example is the time between the infection of HIV virus and the onset of AIDS. Many authors have studied the non-parametric estimation of a failure time distribution for multivariate interval-censored data (see Kim and Xue [1], Goggins and Finkelstein [2], Jones and Rocke [3], Wong and Yu [4] and Yu et al. [5]). Some researchers investigated the non-parametric estimation for the right-censored and interval-censored data in which the occurrence time of the initial event can be exactly observed and observations on the subsequent event are right-censored or interval-censored (see DeGruttola and Lagakos [6], Gómez and Calle, [7], Gómez and Lagakos [8], Sun [9,10]). Further, the statistical inference to the doubly interval-censored data has been

* Corresponding author.
E-mail address: deng@uregina.ca (D. Deng).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.02.007

conducted. Fang and Sun [11] obtained the consistency of the non-parametric maximum likelihood estimation (NPMLE) of a distribution function based on univariate doubly interval-censored failure time data. Moreover, the recent interest is the analysis of multivariate doubly interval-censored data because these data have arisen in the health care field. For example, a longitudinal prospective (1996–2000) oral health screening project was performed in Flanders, Belgium (see Komárek and Lesallre, [12,13]). In this Signal Tandmobiel study, the children born in 1989 were examined on a yearly basis and the primary interest is to investigate the influence of sound versus affected deciduous second molars on the caries susceptibility of the adjacent permanent first molars (teeth 16, 26, 36, 46, respectively). The onset time $U_{i,l}$, $l = 1, \ldots, 4$, is the age of the ith child at which the lth permanent first molar emerged. The failure time, $V_{i,l}$, is the onset of caries of the lth permanent first molar. The variable of interest is the time from tooth emergence to the onset of caries, $T_{i,l}$. Since both the time of tooth emergence $U_{i,l}$ and the onset of caries experience $V_{i,l}$ are only known to lie in an interval of about one year, $T_{i,l}$, $l = 1, \ldots, 4$ is doubly interval-censored and thus the vector $T_i = (T_{i,1}, \ldots, T_{i,4})$ is multivariate doubly interval-censored. Komárek and Lesallre [12] discussed the semi-parametric accelerated failure time model for paired doubly interval-censored data obtained for teeth 16 and 46. Komárek et al. [14] proposed a Bayesian analysis of multivariate doubly interval-censored dental data. Komárek and Lesallre [13] proposed a Bayesian accelerated failure time model with the multivariate doubly interval-censored dental data observed for teeth 16, 26, 36 and 46. In despite of the development of the procedures described above, there does not exist much research on the justification of asymptotic properties of the resulting estimates based on multivariate doubly censored failure time data. Further, as we have known, Groeneboom [15] showed that the NPMLE of a distribution function based on interval-censored failure time data has at least $n^{1/3}$ convergence rate and Deng et al. [16] recently proved that the NPMLE of a distribution function based on univariate doubly interval-censored failure time data has the $n^{3/10}$ convergence rate, which is slower than that of NPMLE based on univariate singly interval-censored data. However, there is no literature for the inference on the joint distribution of multivariate failure times based on doubly interval-censored data, and this motivated the developments presented in this paper.

The aim of the present paper is to study the non-parametric maximum likelihood estimation of a joint distribution and its asymptotic properties for the multivariate doubly interval-censored data. Comparing with the univariate doubly interval-censored data, the inference of asymptotics for multivariate doubly interval-censored data is much more difficult because it involves many variables and different types of interval censoring mechanism. In fact even for bivariate doubly interval-censored data, there are 8 interval censoring variables, 8 indicator variables and two types of interval censoring mechanism. The problem is how to concisely represent the multivariate doubly interval censored data and the censoring mechanism. To avoid the complication of multivariate doubly interval censoring, we first focus on the bivariate doubly interval-censored data and make the efforts for the representation of bivariate doubly interval-censored data, then the similar results for multivariate doubly interval-censored data will be given. By using the designed notation, the procedure proposed by Greoneboom and Wellner [17] and the ϵ-covering number approach (see Van de Geer [18]), we derive the self-consistency equations, the strong consistency and the convergence rate of NPMLE and then give the analogues for multivariate doubly interval-censored data. The results show that the convergence rate of NPMLE depends on not only the dimension of multivariate lifetime variables but also the types of doubly interval censoring mechanism.

The remaining of paper is organized as follows. The notation and assumptions that will be used throughout the paper are introduced in Section 2. In Section 3, the NPMLEs and their strong consistency are discussed. Section 4 deals with the convergence rate for the NPMLEs. The corresponding results for multivariate doubly interval-censored are presented in Section 5 and the proofs are left in Section 6.

2. Notation and assumptions

Consider an epidemiological study that consists of a pair of subjects and in which each subject experiences two related events. Let $\boldsymbol{X} = (X_1, X_2)$ and $\boldsymbol{Y} = (Y_1, Y_2)$ be the vectors of the times of the occurrences of initial and subsequent events, respectively, and the random variable $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$ the vector of survival times of interest. Our goal is to estimate the distribution function $F(\boldsymbol{t})$ of $\boldsymbol{T}$. It is assumed that all of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{T}$ are continuous variables and that $\boldsymbol{X}$ and $\boldsymbol{T}$ are independent (See Gómez and Logakos [7] and Fang and Sun [11]).

Suppose that for both $\boldsymbol{X}$ and $\boldsymbol{Y}$, only interval-censored data are observed. There are different ways to represent interval-censored data. Fang and Sun [11] have given the mechanism which describes the univariate doubly interval-censored data. Following the notations of Fang and Sun [11] in the univariate case, we assume that there exist four vectors of random variables $\boldsymbol{L}$, $\boldsymbol{R}$, $\boldsymbol{U}$ and $\boldsymbol{V}$ that define the observed intervals for $\boldsymbol{X}$ and $\boldsymbol{Y}$. Actually we will observe vectors of random variables $\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\Pi}$ where $\boldsymbol{L} = (L_1, L_2), \boldsymbol{R} = (R_1, R_2), \boldsymbol{U} = (U_1, U_2), \boldsymbol{V} = (V_1, V_2), \boldsymbol{\Delta} = (\Delta_1, \Delta_2) = (I_{\{X_1 \le L_1\}}, I_{\{L_1 < X_1 \le R_1\}})$, $\boldsymbol{\Gamma} = (\Gamma_1, \Gamma_2) = (I_{\{X_2 \le L_2\}}, I_{\{L_2 < X_2 \le R_2\}})$, $\boldsymbol{\Lambda} = (\Lambda_1, \Lambda_2) = (I_{\{Y_1 \le U_1\}}, I_{\{U_1 < Y_1 \le V_1\}})$, $\boldsymbol{\Pi} = (\Pi_1, \Pi_2) = (I_{\{Y_2 \le U_2\}}, I_{\{U_2 < Y_2 \le V_2\}})$. Note that $L_1, R_1, \boldsymbol{\Delta}$ define the observed interval for X_1; $L_2, R_2, \boldsymbol{\Gamma}$ define the observed interval for X_2; $U_1, V_1, \boldsymbol{\Lambda}$ define the observed interval for Y_1; and $U_2, V_2, \boldsymbol{\Pi}$ define the observed interval for Y_2.

From the representation of bivariate doubly interval-censored data, for $p = 1, 2$, X_p and Y_p have their own censoring intervals, which are different from each other. We call this censoring as the different censoring mechanism. However, in some situations, two components of $\boldsymbol{X}$ and $\boldsymbol{Y}$ may share common censoring intervals, respectively. For example, in the paired doubly interval-censored data, X_1, X_2, Y_1 and Y_2 are often measured from the same individual in studies of time to occurrence of a disease to paired organs or in matched-paired case control studies. Therefore, two occurrence times of initial events and two occurrence times of the subsequent events share two common censoring intervals, respectively. That is, X_1

and X_2 share the censoring interval (L, R) and Y_1 and Y_2 share the censoring interval (U, V). In this case, we may assume that $L_1 = L_2, R_1 = R_2, U_1 = U_2$ and $V_1 = V_2$. We call this type of censoring as the common censoring mechanism.

Our goal is to draw the statistical inference about the distribution function $F(\boldsymbol{t})$ of $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$ from the observed random variables $\boldsymbol{L}_i, \boldsymbol{R}_i, \boldsymbol{U}_i, \boldsymbol{V}_i, \boldsymbol{\Delta}_i, \boldsymbol{\Gamma}_i, \boldsymbol{\Lambda}_i$ and $\boldsymbol{\Pi}_i$ $(i = 1, \ldots, n)$ for the bivariate doubly interval-censored data. Actually we will derive the non-parametric likelihood function for F and obtain the NPMLE of F by maximizing this likelihood function. Further we give the self-consistency equation for NPMLE and prove the strong consistency for NPMLE of the distribution function $F(\boldsymbol{t})$. To attain this purpose, we need to introduce some notation and assumptions.

Let $\mathbb{R}_+^m = [0, +\infty)^m$ for any integer m. For any distribution function Q in $\mathbb{R}_+^m$, let μ_Q denote the measure induced by Q. Then for a set A in $\mathbb{R}_+^m$,

$$\mu_Q(A) = \int_A \mathrm{d}\mu_Q(\boldsymbol{z}) = \int \cdots \int_A \mathrm{d}Q(z_1, \ldots z_m) \equiv \int_A \mathrm{d}Q(\boldsymbol{z}).$$

Set $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$, $\boldsymbol{W} = (\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{V})$, $\boldsymbol{\Xi} = (\boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\Pi})$ and consequently, $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, $\boldsymbol{w} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{v})$, $\boldsymbol{\xi} = (\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\pi})$. Now for any $a, b, c \in \mathbb{R}_+$ with $a < b$, define $A_q(a, b)$ and $A_q^c(a, b)(q = 1, 2, 3)$ as

$$A_1(a, b) = [0, a), A_2(a, b) = [a, b), A_3(a, b) = [b, \infty)$$

and

$$A_1^c(a, b) = [0, a \wedge c), A_2^c(a, b) = [a \wedge c, b \wedge c), A_3^c(a, b) = [b \wedge c, c).$$

Then for $\boldsymbol{w} = (\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}_+^8$, $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}_+^4$ with $\boldsymbol{l} < \boldsymbol{r} < \boldsymbol{u} < \boldsymbol{v}$ and $\boldsymbol{x} < \boldsymbol{y}$, and for $q, r, s, t = 1, 2, 3$, define the sets $B_{qr}(\boldsymbol{l}, \boldsymbol{r})$, $B_{qr}^{\boldsymbol{x}}(\boldsymbol{l}, \boldsymbol{r})$, $C_{st}(\boldsymbol{u}, \boldsymbol{v})$ and $C_{st}^{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v})$ in $\mathbb{R}_+^2$, $D_{qrst}(\boldsymbol{w})$ and $C_{qrst}^{\boldsymbol{z}}(\boldsymbol{w})$ in $\mathbb{R}_+^4$ as

$$\begin{aligned}
&B_{qr}(\boldsymbol{l}, \boldsymbol{r}) = A_q(l_1, r_1) \times A_r(l_2, r_2), \qquad C_{st}(\boldsymbol{u}, \boldsymbol{v}) = A_s(u_1, v_1) \times A_t(u_2, v_2),\\
&B_{qr}^{\boldsymbol{x}}(\boldsymbol{l}, \boldsymbol{r}) = A_q^{x_1}(l_1, r_1) \times A_r^{x_2}(l_2, r_2), \qquad C_{st}^{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) = A_s^{y_1}(u_1, v_1) \times A_t^{y_2}(u_2, v_2),\\
&D_{qrst}(\boldsymbol{w}) = B_{qr}(\boldsymbol{l}, \boldsymbol{r}) \times C_{st}(\boldsymbol{u}, \boldsymbol{v}), \qquad D_{qrst}^{\boldsymbol{z}}(\boldsymbol{w}) = B_{qr}^{\boldsymbol{x}}(\boldsymbol{l}, \boldsymbol{r}) \times C_{st}^{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v})
\end{aligned}$$

where $a \wedge b = \min\{a, b\}$ and $\boldsymbol{u} < \boldsymbol{v}$ means that $u_r < v_r$ for $r = 1, 2$. From the definitions of above sets, for any $\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}_+^2$ with $\boldsymbol{l} < \boldsymbol{r} < \boldsymbol{u} < \boldsymbol{v}$,

$$\bigcup_{q,r=1}^{3} B_{qr}(\boldsymbol{l}, \boldsymbol{r}) = \mathbb{R}_+^2, \qquad \bigcup_{q,r=1}^{3} B_{qr}^{\boldsymbol{x}}(\boldsymbol{l}, \boldsymbol{r}) = [0, \boldsymbol{x}],$$

$$\bigcup_{q,r,s,t=1}^{3} D_{qrst}(\boldsymbol{w}) = \mathbb{R}_+^4, \qquad \bigcup_{q,r,s,t=1}^{3} D_{qrst}^{\boldsymbol{z}}(\boldsymbol{w}) = [0, \boldsymbol{z}].$$

Further throughout this paper, we assume:

(A1) $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$ are vectors of non-negative absolutely continuous random variables, and $\boldsymbol{T}$ and $\boldsymbol{X}$ are independent. Let H, F and Φ denote the true cumulative distribution functions of $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$, $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$ and $\boldsymbol{X}$, respectively. Further assume that Φ and F are contained in the class $\mathcal{F}_M := \{\Psi | \text{support}(\boldsymbol{\Psi}) \subset [0, \boldsymbol{M}]; \boldsymbol{\Psi} \ll \boldsymbol{\nu}^2\}$ and H is contained in the class $\mathcal{H}_M := \{\boldsymbol{\Theta} | \text{support}(\boldsymbol{\Theta}) \subset A_M; \boldsymbol{\Theta} \ll \boldsymbol{\nu}^4\}$ where $\boldsymbol{M} = (M, M) \in \mathbb{R}_+^2$ with a given positive constant M, $A_M = \{(\boldsymbol{x}, \boldsymbol{y}) : 0 \le \boldsymbol{x} \le M, \boldsymbol{x} \le \boldsymbol{y} \le 2\boldsymbol{M}\}$ and $\boldsymbol{\nu}^m$ is Lebesgue measure in $\mathbb{R}_+^m$ for any positive integer m. In this case

$$H(\boldsymbol{x}, \boldsymbol{y}) = \int_{[0,\boldsymbol{x}]} F(\boldsymbol{y} - \boldsymbol{z}) \mathrm{d}\mu_\Phi(\boldsymbol{z}), \qquad \Phi(\boldsymbol{x}) = H(\boldsymbol{x}, 2\boldsymbol{M}), \qquad F(\boldsymbol{t}) = \int_{\boldsymbol{y} - \boldsymbol{x} < \boldsymbol{t}} \mathrm{d}\mu_H(\boldsymbol{x}, \boldsymbol{y}).$$

Further, for set $C = A \times B$ in $\mathbb{R}_+^4$,

$$\mu_H(C) = \int_A \mu_F(B - \boldsymbol{z}) \mathrm{d}\mu_\Phi(\boldsymbol{z})$$

where $B - \boldsymbol{z} = \{\boldsymbol{x} - \boldsymbol{z} : \boldsymbol{x} \in B\}$.

(A2) Instead, we observe the random vector $\boldsymbol{W}_i = (\boldsymbol{L}_i, \boldsymbol{R}_i, \boldsymbol{U}_i, \boldsymbol{V}_i)$ with the joint distribution G which satisfies $\mu_G \ll \boldsymbol{\nu}^8$.

(A3) The random vector $\boldsymbol{W}$ is independent of $\boldsymbol{Z}$ and $P(\boldsymbol{L} \le \boldsymbol{R} \le \boldsymbol{U} \le \boldsymbol{V}) = 1$.

(A4) There exist two positive constants η_1 and η_2 such that $P(\boldsymbol{R} - \boldsymbol{L} \ge \eta_1) = 1$ and $P(\boldsymbol{V} - \boldsymbol{U} \ge \eta_2) = 1$.

(A5) The joint density function $g(\boldsymbol{w})$ of $\boldsymbol{W}$, and the marginal density functions $g_{\boldsymbol{LR}}(\boldsymbol{l}, \boldsymbol{r})$ of $(\boldsymbol{L}, \boldsymbol{R})$, $g_{\boldsymbol{UV}}(\boldsymbol{u}, \boldsymbol{v})$ of $(\boldsymbol{U}, \boldsymbol{V})$ have partial derivatives bounded away from both zero and infinity.

(A6) The distribution functions F and Φ have bounded partial derivatives $\frac{\partial^2 F(t_1,t_2)}{\partial t_1 \partial t_2}$ and $\frac{\partial^2 \Phi(x_1,x_2)}{\partial x_1 \partial x_2}$ that are continuous and satisfies $\phi(\boldsymbol{x}) = \frac{\partial^2 \Phi(\boldsymbol{x})}{\partial x_1 \partial x_2} \ge c$, $f(\boldsymbol{t}) = \frac{\partial^2 F(\boldsymbol{t})}{\partial t_1 \partial t_2} \ge c$ for a constant $c > 0$ independent of $\boldsymbol{x}$ and $\boldsymbol{t}$.

(A7) F and H satisfy that for any $1 \le q, r, s, t \le 3$, $\mu_F[C_{st}(\boldsymbol{u}, \boldsymbol{v})] > 0$ and $\mu_H[(D_{qrst}(\boldsymbol{w}))] > 0$ if $0 < \boldsymbol{l} < \boldsymbol{r} < \boldsymbol{u} < \boldsymbol{v} < 2M$.

The above assumptions are necessary for strong consistency of maximum likelihood estimators of a failure time distribution based on doubly interval-censored data and similar to those required for the consistency of the NPMLE of a failure time distribution based on interval-censored data (Groeneboom [15], Yu et al. [5], and Deng and Fang [19]). The condition (A4) means that there are positive time intervals between examination times $\boldsymbol{L}$ and $\boldsymbol{R}$ for the occurrence of the initial events and between examination times $\boldsymbol{U}$ and $\boldsymbol{V}$ for the occurrence of the subsequent events. The assumptions (A4), (A5), (A6) and (A7) are required to ensure the non-singularity of the integral equation appearing in the information calculation. The detailed discussion about these assumptions can be found in Groeneboom and Wellner [17], Groeneboom [15] and Geskus and Groeneboom [20,21].

3. Non-parametric likelihood estimation

Now suppose that $\boldsymbol{T}_1, \boldsymbol{T}_2, \ldots$, are independent identically distributed random variables with distribution function $F(\boldsymbol{t})$, $\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots$, are independent identically distributed random variables with distribution function $H(\boldsymbol{z})$, and $\{(\boldsymbol{W}_i, \boldsymbol{\Xi}_i); i = 1, \ldots, n\}$ are the observed sample from the random variables $(\boldsymbol{W}, \boldsymbol{\Xi})$.

Now, we first discuss the non-parametric maximum likelihood estimation and the strong consistency for the distribution function $H(\boldsymbol{z})$ of random vector $\boldsymbol{Z}$ based on the bivariate doubly interval-censored data.

Note that based on $H(\boldsymbol{z})$ the observed random vector $(\boldsymbol{W}, \boldsymbol{\Xi})$ has the density

$$q_H(\boldsymbol{w}, \boldsymbol{\xi}) = g(\boldsymbol{w}) \prod_{q,r,s,t=1}^{3} \left\{\mu_H[D_{qrst}(\boldsymbol{w})]\right\}^{\delta_q \gamma_r \lambda_s \pi_t} \tag{3.1}$$

where $\delta_3 = 1 - \delta_1 - \delta_2$, $\gamma_3 = 1 - \gamma_1 - \gamma_2$, $\lambda_3 = 1 - \lambda_1 - \lambda_2$ and $\pi_3 = 1 - \pi_1 - \pi_2$. Therefore the conditional likelihood function $C(H)$ for $H(\boldsymbol{z})$ is

$$C(H) = \prod_{i=1}^{n} \prod_{q,r,s,t=1}^{3} \left\{\mu_H[D_{qrst}(\boldsymbol{w}_i)]\right\}^{\delta_{iq} \gamma_{ir} \lambda_{is} \pi_{it}}$$

and the marginal log likelihood function $L(H)$ for $\boldsymbol{Z}$ is

$$\begin{aligned} L(H) &= \sum_{i=1}^{n} \sum_{q,r,s,t=1}^{3} \delta_{iq} \gamma_{ir} \lambda_{is} \pi_{it} \log \mu_H[D_{qrst}(\boldsymbol{w}_i)] \\ &= \sum_{q,r,s,t=1}^{3} \int_{\mathbb{R}^8 \times \{0,1\}^8} \delta_q \gamma_r \lambda_s \pi_t \log \mu_H[D_{qrst}(\boldsymbol{w})] \mathrm{d}\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi}) \end{aligned} \tag{3.2}$$

where $\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi})$ is the empirical probability function obtained from the observations $\{(\boldsymbol{W}_i, \boldsymbol{\Xi}_i); i = 1, \ldots, n.\}$

Now the NPMLE $\hat{H}_n$ of $H(\boldsymbol{z})$ maximizes the function (3.2):

$$\hat{H}_n = \arg\max_{\boldsymbol{\Theta} \in \mathcal{H}_M} L(\boldsymbol{\Theta}).$$

According to the idea of Geskus and Groeneboom [20,21], for $a(\boldsymbol{z}) \in L^2(H)$, we have that

$$E\{a(\boldsymbol{Z}) | (\boldsymbol{W}, \boldsymbol{\Xi}) = (\boldsymbol{w}, \boldsymbol{\xi})\} = \sum_{q,r,s,t=1}^{3} \delta_q \gamma_r \lambda_s \pi_t \frac{\int_{D_{qrst}(\boldsymbol{w})} a(\boldsymbol{z}) \mathrm{d}\mu_H}{\mu_H[D_{qrst}(\boldsymbol{w})]}. \tag{3.3}$$

By letting $a(\boldsymbol{z}') = I_{\{\boldsymbol{z}' \leq \boldsymbol{z}\}}$ for $\boldsymbol{z}' \in \mathbb{R}_+^4$ and taking expectation to both sides of (3.3) with respect to the probability distribution $\mathbb{P}(\boldsymbol{w}, \boldsymbol{\xi})$, we have that

$$H(\boldsymbol{z}) = \sum_{q,r,s,t=1}^{3} \int_{\mathbb{R}^8 \times \{0,1\}^8} \delta_q \gamma_r \lambda_s \pi_t \frac{\mu_H[D^{\boldsymbol{z}}_{qrst}(\boldsymbol{w})]}{\mu_H[C_{qrst}(\boldsymbol{w})]} \mathrm{d}\mathbb{P}(\boldsymbol{w}, \boldsymbol{\xi}).$$

Therefore the self-consistency equation for the non-parametric maximum likelihood estimator $\hat{H}_n(\boldsymbol{z})$ is:

$$\hat{H}_n(\boldsymbol{z}) = \sum_{q,r,s,t=1}^{3} \int_{\mathbb{R}^8 \times \{0,1\}^8} \delta_q \gamma_r \lambda_s \pi_t \frac{\mu_{\hat{H}_n}[D^{\boldsymbol{z}}_{qrst}(\boldsymbol{w})]}{\mu_{\hat{H}_n}[D_{qrst}(\boldsymbol{w})]} \mathrm{d}\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi}) \tag{3.4}$$

where $\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi})$ is the empirical probability function obtained from the observations $\{(\boldsymbol{W}_i, \boldsymbol{\Xi}_i), i = 1, \ldots, n.\}$ The strong consistency of NPMLE $\hat{H}_n(\boldsymbol{z})$ follows from the above self-consistency equation:

Theorem 3.1. *Under the regularity conditions* (A1)–(A7),

$$P\left\{\lim_{n\to\infty} \sup_{\boldsymbol{z}\in\mathbb{R}_+^4} |\hat{H}_n(\boldsymbol{z}) - H(\boldsymbol{z})| = 0\right\} = 1.$$

Further for the determination of $\hat{H}_n(\boldsymbol{z})$, by using the Eq. (3.4) and following the similar procedures given in Betensky and Finkelstein [22], Gentleman and Vandal [23,24] and Bogaerts and Lesaffre [25], one can develop an iterative algorithm to obtain the estimate $\hat{H}_n$.

Next, we consider the non-parametric maximum likelihood estimation and the strong consistency for the distribution function $F(\boldsymbol{t})$ of random variable $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$. Note that the density (3.1) of random vector $(\boldsymbol{W}, \boldsymbol{\Xi})$ can be rewritten into the form

$$q_{F,\Phi}(\boldsymbol{w}, \boldsymbol{\xi}) = g(\boldsymbol{w}) \prod_{q,r,s,t=1}^{3} \left\{ \int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mu_F[C_{st}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{x}]\mathrm{d}\mu_\Phi(\boldsymbol{x}) \right\}^{\delta_q \gamma_r \lambda_s \pi_t}$$

and thus the marginal log likelihood function $L(F, \Phi)$ for F and Φ can be obtained as follows

$$L(F, \Phi) = \sum_{q,r,s,t=1}^{3} \int_{\mathbb{R}^8 \times \{0,1\}^8} \delta_q \gamma_r \lambda_s \pi_t \log \left\{ \int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mu_F[C_{st}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{x}]\mathrm{d}\mu_\Phi(\boldsymbol{x}) \right\} \mathrm{d}\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi}).$$

However, the NPMLE $\hat{\Phi}_n$ can first obtained from the multivariate interval-censored data $\{(\boldsymbol{L}_i, \boldsymbol{R}_i, \boldsymbol{\Delta}_i, \boldsymbol{\Gamma}_i), i = 1, 2, \ldots, n\}$ (see Deng and Fang [19]). Therefore the NPMLE $\hat{F}_n$ of $F(\boldsymbol{t})$ maximizes the function $L(F, \hat{\Phi}_n)$:

$$\hat{F}_n = \underset{\boldsymbol{\Psi} \in \mathcal{F}_M}{\arg\max}\, L(\boldsymbol{\Psi}, \hat{\Phi}_n).$$

Further by defining the conditional expectation $L_F(a(\boldsymbol{T})) = E\{a(\boldsymbol{T})|(\boldsymbol{W}, \boldsymbol{\Xi}) = (\boldsymbol{w}, \boldsymbol{\xi})\}$ for $a(\boldsymbol{T}) \in L^2(F)$, using L_F operator to the function $a(\boldsymbol{z}) = I_{\{\boldsymbol{z} \le \boldsymbol{t}\}}$ and taking expectation with respect to the probability distribution $\mathbb{P}(\boldsymbol{w}, \boldsymbol{\xi})$, the following self-consistency equation for the non-parametric maximum likelihood estimator $\hat{F}_n(\boldsymbol{t})$ of the distribution function $F(\boldsymbol{t})$ for $\boldsymbol{T} = \boldsymbol{Y} - \boldsymbol{X}$ is obtained.

$$\hat{F}_n(\boldsymbol{t}) = \sum_{q,r,s,t=1}^{3} \int_{\mathbb{R}^8 \times \{0,1\}^8} \delta_q \gamma_r \lambda_s \pi_t \frac{\int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mu_{\hat{F}_n}[C_{st}^{\boldsymbol{y}}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{x}]\mathrm{d}\hat{\Phi}_n(\mathbf{x})}{\int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mu_{\hat{F}_n}[C_{st}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{x}]\mathrm{d}\hat{\Phi}_n(\boldsymbol{x})} \mathrm{d}\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi}) \tag{3.5}$$

where $\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi})$ is the empirical probability function obtained from the observations $\{(\boldsymbol{W}_i, \boldsymbol{\Xi}_i), i = 1, \ldots, n.\}$

Now, the strong consistency of the estimator $\hat{F}_n(\boldsymbol{t})$ follows from the self-consistency equation (3.5).

Theorem 3.2. *Under the conditions* (A1)–(A7),

$$P\left\{ \lim_{n\to\infty} \sup_{\boldsymbol{t}\in\mathbb{R}_+^2} |\hat{F}_n(\boldsymbol{t}) - F(\boldsymbol{t})| = 0 \right\} = 1.$$

Also an iterative algorithm can be developed using this self-consistency equation (3.5) to obtain the estimate $\hat{F}_n$.

4. Convergence rates

Now it is of interest to derive the rate of convergence for both $\hat{H}_n$ and $\hat{F}_n$. Before discussing the rate of convergence, we first give the measure for assessing the distance of two distribution functions. The Hellinger distance between two density functions f_1 and f_2 with respect to μ is defined by

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

where μ is a σ-finite dominating measure.

Now the density function ϕ_H of $(\boldsymbol{W}, \boldsymbol{\Xi})$ with respect to the dominating measure $\mu = \mu_1 \times \mu_0$ has the form

$$\phi_H(\boldsymbol{w}, \boldsymbol{\xi}) = \sum_{q,r,s,t=1}^{3} \delta_q \gamma_r \lambda_s \pi_t \mu_H[D_{qrst}(\boldsymbol{w})],$$

where μ_1 is the measure induced by the distribution of $\boldsymbol{W} = (\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{V})$ and μ_0 is the counting measure on $\{0, 1\}^8$.

Now we first give the convergence rate for $\phi_{\hat{H}_n}$.

Theorem 4.1. *Suppose that the conditions* (A1)–(A7) *hold.*
(i) *For the different censoring mechanism, as $n \to \infty$,*

$$h(\phi_{\hat{H}_n}, \phi_H) = O_p\left(n^{-\frac{5}{18}} (\log n)^{\frac{8}{9}} \right).$$

(ii) *For the common censoring mechanism, as $n \to \infty$*

$$h(\phi_{\hat{H}_n}, \phi_H) = O_p\left(n^{-\frac{15}{62}}(\log n)^{\frac{2}{9}}\right).$$

Next we estimate the convergence rate for $\phi_{\hat{F}_n}$. We define ϕ_F as the density of $(\boldsymbol{W}, \boldsymbol{\Xi})$:

$$\phi_F(\boldsymbol{w}, \boldsymbol{\xi}) = \sum_{q,r,s,t=1}^{3} \delta_q \gamma_r \lambda_s \pi_t \int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mu_F[C_{st}(\boldsymbol{u}, \boldsymbol{v}) - \boldsymbol{x}] \mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x}).$$

From Theorem 4.1, the convergence rate for $\phi_{\hat{F}_n}$ can be obtained.

Theorem 4.2. *Under the same conditions as those in Theorem 4.1, we have that as $n \to \infty$,*
(i) *for the different censoring mechanism,*

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}\right);$$

(ii) *for the common censoring mechanism,*

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{15}{62}}(\log n)^{\frac{2}{9}}\right).$$

Now we further specify the convergence rates for NPMLEs $\hat{H}_n$ and $\hat{F}_n$ by using the L_2-distances with respect to the marginals of $G(\boldsymbol{w})$. We first define L_2-distance between two distribution functions $H_1(\boldsymbol{z})$ and $H_2(\boldsymbol{z})$ in $\mathbb{R}_+^4$ with respect to the distribution function $G_{\boldsymbol{LU}}(\boldsymbol{l}, \boldsymbol{u})$ of $(\boldsymbol{L}, \boldsymbol{U})$ as

$$d_{\boldsymbol{LU}}(H_1, H_2) = \left[\int_{\mathbb{R}_+^4} \{H_1(\boldsymbol{l}, \boldsymbol{u}) - H_2(\boldsymbol{l}, \boldsymbol{u})\}^2 \mathrm{d}G_{\boldsymbol{LU}}(\boldsymbol{l}, \boldsymbol{u})\right]^{1/2}.$$

Similarly, $d_{\boldsymbol{LV}}(H_1, H_1)$, $d_{\boldsymbol{RU}}(H_1, H_2)$ and $d_{\boldsymbol{RV}}(H_1, H_2)$ can be defined. Now, the analogue of Corollary 2 in Geskus and Groeneboom [20] is given as what follows.

Theorem 4.3. *Under the same conditions as those in Theorem 4.1, we have that as $n \to \infty$,*
(i) *for the different censoring mechanism,*

$$d(\hat{H}_n, H) = O_p\left(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}\right);$$

(ii) *for the common censoring mechanism,*

$$d(\hat{H}_n, H) = O_p\left(n^{-\frac{15}{62}}(\log n)^{\frac{2}{9}}\right)$$

where $d(\hat{H}_n, H)$ represents anyone of four L_2-distances given above.

Also we have the convergence rate for the NPMLE $\hat{F}_n$.

Theorem 4.4. *Under the same conditions as those in Theorem 4.1, we have that as $n \to \infty$,*
(i) *for the different censoring mechanism,*

$$d_i(\hat{F}_n, F) = O_p\left(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}\right);$$

(ii) *for the common censoring mechanism,*

$$d_i(\hat{F}_n, F) = O_p\left(n^{-\frac{15}{62}}(\log n)^{\frac{2}{9}}\right)$$

where

$$d_1(\hat{F}_n, F) = \left[\int_{\mathbb{R}_+^2}\left(\int_{\mathbb{R}_+^2}[\hat{F}_n(\boldsymbol{u}-\boldsymbol{x}) - F(\boldsymbol{u}-\boldsymbol{x})]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G_{\boldsymbol{U}}(\boldsymbol{u})\right]^{1/2}$$

and

$$d_2(\hat{F}_n, F) = \left[\int_{\mathbb{R}_+^2}\left(\int_{\mathbb{R}_+^2}[\hat{F}_n(\boldsymbol{v}-\boldsymbol{x}) - F(\boldsymbol{v}-\boldsymbol{x})]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G_{\boldsymbol{V}}(\boldsymbol{v})\right].$$

5. Convergence rates for multivariate doubly interval-censored data

Sections 3 and 4 gave the self-consistency equation, strong consistency and convergence rate of NPMLE's $\hat{H}_n$ and $\hat{F}_n$ for two types of bivariate interval censoring mechanism. Now we briefly discuss the analogues for the multivariate doubly interval-censored data. We consider the general censoring mechanism in which, the vectors of interest $\boldsymbol{X} = (X_1, \ldots, X_d)$ and $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ are divided into several groups and each group has a common censoring interval. In fact, vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are partitioned into p groups $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p)$ and $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_p)$ with $\boldsymbol{X}_1 = (X_1, \ldots, X_{d_1}), \ldots, \boldsymbol{X}_p = (X_{d_1+\cdots+d_{p-1}+1}, \ldots, X_{d_1+\cdots+d_{p-1}+d_p})$, $\boldsymbol{Y}_1 = (Y_1, \ldots, Y_{d_1}), \ldots, \boldsymbol{Y}_p = (Y_{d_1+\cdots+d_{p-1}+1}, \ldots, Y_{d_1+\cdots+d_{p-1}+d_p})$ with $d_1 + d_2 + \cdots + d_p = d$ and thus $\boldsymbol{T} = (\boldsymbol{T}_1, \ldots, \boldsymbol{T}_p) = (\boldsymbol{Y}_1 - \boldsymbol{X}_1, \ldots, \boldsymbol{Y}_p - \boldsymbol{X}_p)$. The censoring intervals of each $\boldsymbol{X}_s$ and each $\boldsymbol{Y}_s$ are (L_s, R_s) and (U_s, V_s), respectively ($s = 1, \ldots, p$). In this case, the random variables $\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\Pi}$ are observed, where $\boldsymbol{L} = (L_1, \ldots, L_p), \boldsymbol{R} = (R_1, \ldots, R_p), \boldsymbol{U} = (U_1, \ldots, U_p), \boldsymbol{V} = (V_1, \ldots, V_p), \boldsymbol{\Delta} = (\Delta_1, \ldots, \Delta_d), \boldsymbol{\Gamma} = (\Gamma_1, \ldots, \Gamma_d), \boldsymbol{\Lambda} = (\Lambda_1, \ldots, \lambda_d)$ and $\boldsymbol{\Pi} = (\Pi_1, \ldots, \Pi_d)$ with $\Delta_s = I_{\{X_s < L_t\}}$, $\Gamma_s = I_{\{L_t \le X_s < R_t\}}$, $\Lambda_s = I_{\{Y_s < U_t\}}$, $\Pi_s = I_{\{U_t \le Y_s < V_t\}}$ for $s = d_1 + \cdots + d_{t-1} + 1, \ldots, d_1 + \cdots + d_t$; $t = 1, \ldots, p$. Note that $1 \le p \le d$. If $p = 1$, all X_r's have the common censoring interval (L, R) and all Y_r's have the common censoring interval (U, V). If $p = d$, each X_r has its own censoring interval (L_r, R_r) and each Y_r has its own interval (U_r, V_r) for $r = 1, \ldots, d$.

Further under the similar conditions the self-consistency equation, strong consistency of NPMLE's $\hat{H}_n$ and $\hat{F}_n$ can be obtained for the general doubly interval censoring mechanism. Also the densities ϕ_H and ϕ_F of $(\boldsymbol{L}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\Delta}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda}, \boldsymbol{\Pi})$ can be defined as given above and the convergence rates for $\phi_{\hat{H}_n}$ and $\phi_{\hat{F}_n}$ are

$$h(\phi_{\hat{H}_n}, \phi_H) = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+2d)}{2(1+\alpha+2d+6\alpha d)}} (\log n)^{\frac{d(2\beta-1)}{(1+4d)}}\right) & \text{if } \alpha > 1 \\ O_p\left(n^{-\frac{(1+2d)}{2(1+4d)}} (\log n)^{\frac{2d^2}{(1+4d)}}\right) & \text{if } \alpha = 1; \end{cases}$$

and

$$h(\phi_{\hat{F}_n}, \phi_F) = \begin{cases} O_p\left(n^{-\frac{(1+\alpha)(1+2d)}{2(1+\alpha+2d+6\alpha d)}} (\log n)^{\frac{d(2\beta-1)}{(1+4d)}}\right) & \text{if } \alpha > 1 \\ O_p\left(n^{-\frac{(1+2d)}{2(1+4d)}} (\log n)^{\frac{2d^2}{(1+4d)}}\right) & \text{if } \alpha = 1, \end{cases}$$

where $\alpha = \max\{d_i, 1 \le i \le p\}$ and $\beta = \text{card}\{i : d_i = \alpha\}$ is the cardinal number of the set $\{i : d_i = \alpha\}$.

From Theorem 4.1, the convergence rates are different for two types of bivariate interval censoring mechanism. Now we illustrate the effects of different doubly interval censoring mechanism to the convergence rates of NPMLEs.

We consider the case where $\boldsymbol{X} = (X_1, X_2, X_3, X_4)$ and $\boldsymbol{Y} = (Y_1, Y_2, Y_3, Y_4)$ ($d = 4$). The following table gives the convergence rates for different types of doubly interval censoring mechanism.

From the table we see that $a_{1n} \ll a_{2n} \ll a_{3n} \ll a_{4n} \ll a_{5n}$ where $a_n \ll b_n$ means that $a_n = o(b_n)$. We find that for $d = 4$ the convergence rate a_{1n} with different censoring intervals for each marginal of $(\boldsymbol{X}, \boldsymbol{Y})$ is fastest and a_{5n} is slowest with common censoring intervals for all marginals of $(\boldsymbol{X}, \boldsymbol{Y})$. Also, $n^{-\frac{5}{18}} (\log n)^{\frac{8}{9}} \ll a_{1n}$ and $n^{-\frac{15}{62}} (\log n)^{\frac{2}{9}} \ll a_{5n}$, which mean that for same type of interval censoring mechanism, the convergence rate for lower dimension is faster than that for higher dimension. Further, we give some comments about the convergence rates for the general multivariate doubly interval-censored data.

Partition of $\boldsymbol{X}, \boldsymbol{Y}$	Censoring intervals	α	β	Conv. rate a_n
$\boldsymbol{X}_i = X_i, i = 1, 2, 3, 4$	$(L_i, R_i), i = 1, 2, 3, 4$	1	4	
$\boldsymbol{Y}_i = Y_i, i = 1, 2, 3, 4$	$(U_i, V_i), i = 1, 2, 3, 4$			$a_{1n} = n^{-\frac{9}{34}} (\log n)^{\frac{32}{17}}$
$\boldsymbol{X}_1 = (X_1, X_2)$,	(L_1, R_1),	2	1	
$\boldsymbol{X}_2 = X_3, \boldsymbol{X}_3 = X_4$	$(L_2, R_2), (L_3, R_3)$			$a_{2n} = n^{-\frac{27}{118}} (\log n)^{\frac{4}{17}}$
$\boldsymbol{Y}_1 = (Y_1, Y_2)$,	(L_1, R_1),			
$\boldsymbol{Y}_2 = Y_3, \boldsymbol{Y}_3 = Y_4$	$(U_2, V_2), (U_3, V_3)$			
$\boldsymbol{X}_1 = (X_1, X_2), \boldsymbol{X}_2 = (X_3, X_4)$	$(L_1, R_1), (L_2, R_2)$	2	2	
$\boldsymbol{Y}_1 = (Y_1, Y_2), \boldsymbol{Y}_2 = (Y_3, Y_4)$	$(U_1, V_1), (U_2, V_2)$			$a_{3n} = n^{-\frac{27}{118}} (\log n)^{\frac{12}{17}}$
$\boldsymbol{X}_1 = (X_1, X_2, X_3), \boldsymbol{X}_2 = X_4$	$(L_1, R_1), (L_2, R_2)$	3	1	
$\boldsymbol{Y}_1 = (Y_1, Y_2, Y_3), \boldsymbol{Y}_2 = Y_4$	$(U_1, V_1), (U_2, V_2)$			$a_{4n} = n^{-\frac{3}{14}} (\log n)^{\frac{4}{17}}$
$\boldsymbol{X}_1 = (X_1, X_2, X_3, X_4)$	(L, R)	4	1	
$\boldsymbol{Y}_1 = (Y_1, Y_2, Y_3, Y_4)$	(U, V)			$a_{5n} = n^{-\frac{45}{218}} (\log n)^{\frac{4}{17}}$

Remark. (i) Note that if $\alpha = 1$, then $p = d$ and each X_r and each Y_r have their own censoring intervals (L_r, R_r) and (U_r, V_r), respectively $(r = 1, \dots, d)$. If $\alpha = d$, then $p = 1$ and all X_r's and Y_r's have the common censoring intervals (L, R) and (U, V), respectively $(r = 1, 2, \dots, d)$.

Remark. (ii) In the common doubly censoring interval case, we have that $\alpha = d$, $\beta = 1$ and thus,

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{(1+3d+2d^2)}{2(1+3d+6d^2)}} (\log n)^{\frac{d}{(1+4d)}}\right).$$

If $d = 1$, we have that $\alpha = 1$ and

$$h(\phi_{\hat{F}_n}, \phi_F) = O_p\left(n^{-\frac{3}{10}} (\log n)^{\frac{2}{5}}\right),$$

which coincides with the result in Deng et al. [16].

Remark. (iii) Since $\frac{(1+3d+2d^2)}{2(1+3d+6d^2)}$ and $\frac{1+2d}{2(1+4d)}$ are the monotone decreasing functions of d, the convergence rate decreases as the dimension d of multivariate lifetimes $\boldsymbol{X}$ and $\boldsymbol{Y}$ increases. Moreover, the upper limit of convergence rates is $n^{-3/10}$. The lower limits are $n^{-1/4}$ in the situation where each (X_r, Y_r) has its own censoring interval ($\alpha = 1$) and $n^{-1/6}$ if all (X_r, Y_r)'s have a common censoring interval ($\alpha = d$). Further, note that the convergence rate with singly interval-censored data is $O_p(n^{-\frac{(1+\alpha)(1+d)}{2(1+\alpha+d+3\alpha d)}} (\log n)^{\frac{d(\beta-1)}{2(1+2d)}})$ for $\alpha > 1$ and $O_p(n^{-\frac{(1+d)}{2(1+2d)}} (\log n)^{\frac{d^2}{2(1+2d)}})$ for $\alpha = 1$ (see Deng and Fang [19]). Therefore the convergence rate with multivariate doubly interval-censored data is slower than that with multivariate singly interval-censored data.

Remark. (iv) Also, for the fixed d, $\frac{(1+\alpha)(1+2d)}{2(1+\alpha+2d+6\alpha d)}$ is the monotone decreasing function of α. Hence, the bigger the number of censoring intervals is, the faster the convergence rate is, which is reasonable because the data with more censoring intervals involve much more information than the data with less common censoring intervals about the multivariate lifetime $(\boldsymbol{X}, \boldsymbol{Y})$.

6. Proofs of main results

In this section we give the proofs for the theorems. At first we prove Theorem 3.1. The proof of Theorem 3.2 is same as that of Theorem 3.1 and thus omitted.

Proof of Theorem 3.1. At first note that $L(H)$ is maximized at $\hat{H}_n$. We have that

$$\lim_{\epsilon \downarrow 0} \epsilon^{-1}\{L[(1-\epsilon)\hat{H}_n + \epsilon H] - L(\hat{H}_n)\} \le 0.$$

Thus it follows from the marginal log likelihood function that

$$\int_{\mathbb{R}^8 \times \{0,1\}^8} \sum_{q,r,s,t=1}^{3} \delta_q \gamma_r \lambda_s \pi_t \frac{\mu_H(D_{qrst}(\boldsymbol{w}))}{\mu_{\hat{H}_n}(D_{qrst}(\boldsymbol{w}))} \mathrm{d}\mathbb{P}_n(\boldsymbol{w}, \boldsymbol{\xi}) \le 1.$$

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

In fact, similar to the proof of Lemma 4.3 in Groeneboom and Wellner [17], under (A1)–(A7) it follows that

$$\int_{\mathbb{R}_+^8 \times \{0,1\}^8} \sum_{q,r,s,t=1}^{3} \delta_q \gamma_r \lambda_s \pi_t \frac{\mu_H[D_{qrst}(\boldsymbol{w})]}{\mu_{H'}[D_{qrst}(\boldsymbol{w})]} \mathrm{d}\mathbb{P}(\boldsymbol{w}, \boldsymbol{\xi}) = \int_{\mathbb{R}_+^8} \sum_{q,r,s,t=1}^{3} \frac{[\mu_H[D_{qrst}(\boldsymbol{w})]]^2}{\mu_{H'}[D_{qrst}(\boldsymbol{w})]} \mathrm{d}G(\boldsymbol{w}) \le 1.$$

On the other hand, from the following facts that for real numbers $0 < a_i < 1$, $0 < b_i < 1$; $i = 1, 2, \dots, k$ 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

$$\sum_{q,r,s,t=1}^{3} \mu_H[D_{qrst}(\boldsymbol{w})] = \sum_{q,r,s,t=1}^{3} \mu'_H[D_{qrst}(\boldsymbol{w})] = 1,$$

we have that

$$\int_{\mathbb{R}_+^8\times\{0,1\}^8}\sum_{q,r,s,t=1}^{3}\delta_q\gamma_r\lambda_s\pi_t\frac{\mu_H(D_{qrst}(\boldsymbol{w}))}{\mu_{H'}(D_{qrst}(\boldsymbol{w}))}\mathrm{d}\mathbb{P}(\boldsymbol{w},\boldsymbol{\xi}) > 1$$

unless $H' = H$. From the above result, for all $\boldsymbol{z}$, $H'(\boldsymbol{z}) = H(\boldsymbol{z})$. The proof is complete. □

Next we are in the position to prove Theorem 4.1. To this end, we need the following results. We first give the definition of covering number.

Definition 6.1 (*See Van de Geer [18]*). 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,

$$N(\epsilon,\mathcal{G},Q)=\min\left\{J:\ \text{there exist }\{g_j\}_{j=1}^{J}\text{such that for all } g\in\mathcal{G},\ \min_{j\in\{1,\ldots,J\}}\int(g-g_j)^2\mathrm{d}Q\le\epsilon^2\right\}.$$

Consider a probability space $(\Omega,\mathfrak{B},P)$ and independent identically distributed random vectors $\boldsymbol{Z}_1,\boldsymbol{Z}_2,\ldots,\boldsymbol{Z}_n$, with distribution Φ_0. Suppose that

$$\phi_0=\frac{\mathrm{d}\Phi_0}{\mathrm{d}\mu}\in\mathcal{P}$$

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

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

Let P_n be the measure induced by the empirical distribution of the sample $\{\boldsymbol{Z}_i, i = 1,\ldots n\}$. Define the convex hull $\mathrm{conv}(\mathcal{K})$ of a class $\mathcal{K}$ as the set of all finite convex combinations of elements in $(\mathcal{K})$ and $\overline{\mathrm{conv}}(\mathcal{K})$ as the closure of $\mathrm{conv}(\mathcal{K})$. From Theorem 1.1 and Theorem 2.2 in Van de Geer [18], we have the following proposition.

Proposition 6.1. *Suppose* $\mathcal{K}=\{k(\cdot,\boldsymbol{z}):\boldsymbol{z}\in\mathbb{R}_+^4\}$ *and* $\mathcal{P}=\overline{\mathrm{conv}}(\mathcal{K})$. *Assume that for some sequences* $1\le\rho_n\uparrow\infty$ *and* $0\le\sigma_n\downarrow 0$,

$$\int_{\phi_0>\sigma_n}\frac{K^2}{\phi_0}\mathrm{d}\mu\le\rho_n^2,\quad n=1,2,\ldots,$$

where $K=\sup_{k\in\mathcal{K}}k$ *and for* $\tilde{\mathcal{K}}=\left\{\left(\frac{k(\cdot,\boldsymbol{z})}{\phi_0}\right)I\{\phi_0>\sigma_n\},\boldsymbol{z}\in\mathbb{R}_+^4\right\}$, *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* $\tau_n\ge 0$ *satisfying*

$$\tau_n^2\ge\int_{\phi_0\le\sigma_n}\phi_0\mathrm{d}\mu,\quad n=1,2,\ldots,$$
$$\tau_n\ge n^{-(2+w)/(4+4w)}\rho_n^{w/(2+2w)},\quad n=1,2,\ldots,$$

we have that

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

Proof of Theorem 4.1. We only prove the theorem for the different censoring mechanism. Note that the density ϕ_F of $(\boldsymbol{W},\boldsymbol{\Xi})$, with respect to the dominating measure $\mu=\mu_1\times\mu_0$ is then in the class

$$\mathcal{P}=\left\{\phi_\nu(\boldsymbol{w},\boldsymbol{\xi})=\sum_{q,r,s,t=1}^{3}\delta_q\gamma_r\lambda_s\pi_t\mu_\nu[C_{qrst}(\boldsymbol{w})]:\nu\in\mathcal{F}\right\}$$

where $\mathcal{F}$ is the class of all distributions on $\mathbb{R}_+^4$ and μ_ν is the measure induced by the distribution function ν in $\mathbb{R}_+^4$. Clearly, $\mathcal{P}=\overline{\mathrm{conv}}(\mathcal{K})$, with

$$\mathcal{K}=\left\{\kappa_z(\boldsymbol{w},\boldsymbol{\xi})=\sum_{q,r,s,t=1}^{3}\delta_q\gamma_r\lambda_s\pi_t I_{B_{qr}(\boldsymbol{x})}(\boldsymbol{l},\boldsymbol{r})I_{C_{st}(\boldsymbol{y})}(\boldsymbol{u},\boldsymbol{v}):(\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}_+^4\right\}$$

where $B_{qr}(\boldsymbol{x})=\{(\boldsymbol{l},\boldsymbol{r}):\boldsymbol{x}\in B_{qr}(\boldsymbol{l},\boldsymbol{r})\}$ and $C_{st}(\boldsymbol{y})=\{(\boldsymbol{u},\boldsymbol{v}):\boldsymbol{y}\in C_{st}(\boldsymbol{u},\boldsymbol{v})\}$. Under the condition (A5), one can verify that for any $\boldsymbol{z}_1 = (\boldsymbol{x}_1, \boldsymbol{y}_1)$ and $\boldsymbol{z}_2 = (\boldsymbol{x}_2, \boldsymbol{y}_2)$ in $\mathbb{R}_+^4$,

$$\begin{aligned}
&\int_{\mathbb{R}^8\times\{0,1\}^8} [\kappa_{\boldsymbol{z}_1}(\boldsymbol{w},\boldsymbol{\xi}) - \kappa_{\boldsymbol{z}_2}(\boldsymbol{w},\boldsymbol{\xi})]^2 \mathrm{d}\mu \\
&\quad = \int_{\mathbb{R}^8\times\{0,1\}^8} \left(\sum_{q,r,s,t=1}^{3} \delta_q\gamma_r\lambda_s\pi_t [I_{B_{qr}(\boldsymbol{x}_1)}(\boldsymbol{l},\boldsymbol{r}) I_{C_{st}(\boldsymbol{y}_1)}(\boldsymbol{u},\boldsymbol{v}) - I_{B_{qr}(\boldsymbol{x}_2)}(\boldsymbol{l},\boldsymbol{r}) I_{C_{st}(\boldsymbol{y}_2)}(\boldsymbol{u},\boldsymbol{v})] \right)^2 \mathrm{d}\mu \\
&\quad \le \sum_{q,r=1}^{3} \int_{\mathbb{R}_+^4} |I_{B_{qr}(\boldsymbol{x}_1)}(\boldsymbol{l},\boldsymbol{r}) - I_{B_{qr}(\boldsymbol{x}_2)}(\boldsymbol{l},\boldsymbol{r})| g_{\boldsymbol{LR}}(\boldsymbol{l},\boldsymbol{r}) \mathrm{d}\boldsymbol{l}\mathrm{d}\boldsymbol{r} + \sum_{s,t=1}^{3} \int_{\mathbb{R}_+^4} |I_{C_{st}(\boldsymbol{y}_1)}(\boldsymbol{u},\boldsymbol{v}) - I_{C_{st}(\boldsymbol{y}_2)}(\boldsymbol{u},\boldsymbol{v})| g_{\boldsymbol{UV}}(\boldsymbol{u},\boldsymbol{v}) \mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v} \\
&\quad \le C_0(|x_{11} - x_{21}| + |x_{12} - x_{22}| + |y_{11} - y_{21}| + |y_{12} - y_{22}|),
\end{aligned}$$

where C_0 is the constant which varies from line to line.

Therefore there exists a constant A_0 such that for any probability measure Q on $\mathbb{R}_+^8 \times \{0, 1\}^8$,

$$N(\epsilon, \mathcal{K}, Q) \le A_0\epsilon^{-8}, \text{ for all } \epsilon > 0 \tag{6.1}$$

where $N(\epsilon, \mathcal{K}, Q)$ is the ϵ-covering number of $(\mathcal{K}, Q)$. Application of (6.1), with $\mathrm{d}Q = (1/\phi_F^2)I\{\phi_F > \sigma_n\}\mathrm{d}\mathbb{P}_n/(A^2\rho_n^2)$, gives

$$N(\epsilon, \tilde{\mathcal{K}}, \mathbb{P}_n) \le A^8 A_0 \left(\frac{\rho_n}{\epsilon}\right)^8, \quad \text{for all } \epsilon > 0,$$

on the set $\left\{\int_{\phi_F>\sigma_n} \frac{1}{\phi_F^2}\mathrm{d}\mathbb{P}_n \le A^2\rho_n^2\right\}$. So, for $\int_{\phi_F>\sigma_n} 1/\phi_F \mathrm{d}\mu \le \rho_n^2$, we have that

$$\begin{aligned}
&\limsup_{n\to\infty} P\left(\left(\frac{\epsilon}{\rho_n}\right)^4 N(\epsilon, \tilde{\mathcal{K}}, \mathbb{P}_n) > A^4 A_0\right) \le \limsup_{n\to\infty} P\left(\int_{\phi_F>\sigma_n} \frac{1}{\phi_F^2}\mathrm{d}\mathbb{P}_n > A^2\rho_n^2\right) \\
&\quad \le \limsup_{n\to\infty} P\left(\int_{\phi_F>\sigma_n} \frac{1}{\phi_F}\mathrm{d}\mu_n > A^2\rho_n^2\right) \to 0, \quad \text{as } A \to \infty,
\end{aligned}$$

where $\mathrm{d}\mu_n = \mathrm{d}\mathbb{P}_n/\phi_F \to \mathrm{d}\mu$ as $n \to \infty$.

Next we derive the expressions for $\{\rho_n\}$ and $\{\tau_n\}$. Note that under the condition (A1), we have that

$$\begin{aligned}
\mu_H(D_{qrst}(\boldsymbol{w})) &= \int_{D_{qrst}(\boldsymbol{w})} \mathrm{d}\mu_H(\boldsymbol{z}) \\
&\le C_0 \int_{D_{qrst}(\boldsymbol{w})} \mathrm{d}\boldsymbol{\nu}^4(\boldsymbol{z}) \\
&\le C_0 \int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})} \mathrm{d}\boldsymbol{\nu}^2(\boldsymbol{x}) \int_{C_{st}(\boldsymbol{u},\boldsymbol{v})} \mathrm{d}\boldsymbol{\nu}^2(\boldsymbol{y}) \\
&\le C_0 |A_q(l_1, r_1)||A_r(l_2, r_2)||A_s(u_1, v_1)||A_t(u_2, v_2)|,
\end{aligned}$$

where for $r = 1, 2, 3$, $|A_r(a, b)|$ denotes the length of interval $A_r(a, b)$. Therefore,

$$\begin{aligned}
\phi_H(\boldsymbol{w}, \boldsymbol{\xi}) &= \sum_{q,r,s,t=1}^{3} \delta_q\gamma_r\lambda_s\pi_t \mu_H[D_{qrst}(\boldsymbol{w})] \\
&\le C_0 \sum_{q,r,s,t=1}^{3} I_{\Omega_{qrst}}(\boldsymbol{\xi}) |A_q(l_1, r_1)||A_r(l_2, r_2)||A_s(u_1, v_1)||A_t(u_2, v_2)|
\end{aligned}$$

where $\Omega_{qrst} = \{(\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \boldsymbol{\pi}) : \delta_q = \gamma_r = \lambda_s = \pi_t = 1\}$. Similarly, we have

$$\phi_H(\boldsymbol{w}, \boldsymbol{\xi}) \ge C_0 \sum_{q,r,s,t=1}^{3} I_{\Omega_{qrst}}(\boldsymbol{\xi}) |A_q(l_1, r_1)||A_r(l_2, r_2)||A_s(u_1, v_1)||A_t(u_2, v_2)|.$$

Note that for $0 \le \sigma_n \downarrow 0$, regardless of the order of variables $l_1, l_2, r_1, r_2, u_1, u_2, v_1, v_2$, $\phi_H^{-1}(\boldsymbol{w}, \boldsymbol{\xi})$ is dominated by

$$[l_1(r_2 - l_2)u_1(v_2 - u_2)]^{-1}$$

on the set $\{\phi_H > \sigma_n\}$ and $\phi_H(\boldsymbol{w}, \boldsymbol{\xi})$ is dominated by

$$[l_1(r_2 - l_2)u_1(v_2 - u_2)]$$

on the set $\{\phi_H \le \sigma_n\}$. Now set

$$f(\boldsymbol{w}) = [l_1(r_2 - l_2)u_1(v_2 - u_2)].$$

Then it follows that for $0 \le \sigma_n \downarrow 0$,

$$\begin{aligned}\int_{\phi_H > \sigma_n} \frac{1}{\phi_H} \mathrm{d}\mu &\le C_0 \int_{f(\boldsymbol{w}) > C_0\sigma_n} f^{-1}(\boldsymbol{w}) g(\boldsymbol{w}) \mathrm{d}\boldsymbol{w} \\ &\le C_0 \int_{l_1(r_2 - l_2)u_1(v_2-u_2) > C_0\sigma_n} [l_1(r_2 - l_2)u_1(v_2 - u_2)]^{-1} \mathrm{d}l_1 \mathrm{d}r_1 \mathrm{d}l_2 \mathrm{d}r_2 \mathrm{d}u_1 \mathrm{d}u_2 \mathrm{d}v_1 \mathrm{d}v_2 \\ &\le C_0 \int_{l_1 r_2' u_1 v_2' > C_0\sigma_n} (l_1 r_2' u_1 v_2')^{-1} \mathrm{d}l_1 \mathrm{d}r_2' \mathrm{d}u_1 \mathrm{d}v_2' \le C_0 \left(\log \frac{1}{\sigma_n}\right)^4\end{aligned}$$

and similarly, one can prove that

$$\int_{\phi_H \le \sigma_n} \phi_H \mathrm{d}\mu \le C_0 \sigma_n^2 \left(\log \frac{1}{\sigma_n}\right)^3.$$

Now using Proposition 6.1 with $w = 8$, we have that

$$h(\phi_n, \phi_H) = O_p(n^{-\frac{5}{18}} (\log n)^{\frac{8}{9}})$$

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

$$\begin{aligned}\max_{\phi \in \mathcal{P}} \sum_{i=1}^n \log\{\phi(\boldsymbol{w}_i, \boldsymbol{\xi}_i)\} &= \max_{\nu \in \mathcal{H}_M} \sum_{i=1}^n \log\{\phi_\nu(\boldsymbol{w}_i, \boldsymbol{\xi}_i)\} \\ &= \max_{\nu \in \mathcal{H}_M} \sum_{i=1}^n \sum_{q,r,s,t=1}^3 \delta_{iq}\gamma_{ir}\lambda_{is}\pi_{it} \log\{\mu_\nu[D_{qrst}(\boldsymbol{w}_i)]\} = \max_{\Theta \in \mathcal{H}_M} L(\Theta),\end{aligned}$$

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

Proof of Theorem 4.2. Similar to the proof of Theorem 4.1, we only prove Theorem 4.2 for the different censoring mechanism. Note that by triangle inequality for Hellinger distance,

$$h(\phi_{\hat{F}_n}, \phi_F) \le h(\phi_{\hat{F}_n}, \phi_{\hat{H}_n}) + h(\phi_{\hat{H}_n}, \phi_H).$$

Therefore we need to prove that

$$h(\phi_{\hat{F}_n}, \phi_{\hat{H}_n}) = O_p(n^{-\frac{5}{18}} (\log n)^{\frac{8}{9}}).$$

Towards this end, from the definitions of ϕ_H and ϕ_F, it is sufficient to prove that for any $1 \le q, r, s, t \le 3$,

$$\left[\int_{\mathbb{R}_+^8 \times \{0,1\}^8} \left(\sqrt{\int_{B_{qr}} \mu_{\hat{F}_n}(C_{st} - \boldsymbol{x}) \mathrm{d}\mu_\Phi(\boldsymbol{x})} - \sqrt{\int_{B_{qr}} \mu_{\hat{F}_n}(C_{st} - \boldsymbol{x}) \mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2 \mathrm{d}\mu\right]^{1/2} = O_p(n^{-\frac{5}{18}} (\log n)^{\frac{8}{9}})$$

where for convenience we drop the arguments $(\boldsymbol{l}, \boldsymbol{r})$, $(\boldsymbol{u}, \boldsymbol{v})$ and $(\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{v})$ for $B_{qr}(\boldsymbol{l}, \boldsymbol{r})$, $C_{st}(\boldsymbol{u}, \boldsymbol{v})$ and $D_{qrst}(\boldsymbol{l}, \boldsymbol{r}, \boldsymbol{u}, \boldsymbol{v})$ etc. Now we note that for B_{11} and C_{11},

$$\begin{aligned}&\int_{B_{11}} \mu_F(C_{11} - \boldsymbol{x}) \mathrm{d}\mu_\Phi(\boldsymbol{x}) = \int_{[0,\boldsymbol{l}]} F(\boldsymbol{u} - \boldsymbol{x}) \Phi''_{x_1x_2}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \Phi(l_1, l_2) F(u_1 - l_1, u_2 - l_2) - \int_0^{l_1} \Phi(x_1, l_2) F'_{x_1}(u_1 - x_1, u_2 - l_2) \mathrm{d}x_1 \\ &\quad - \int_0^{l_2} \Phi(l_1, x_2) F'_{x_2}(u_1 - l_1, u_2 - x_2) \mathrm{d}x_2 + \int_0^{l_2}\int_0^{l_1} \Phi(x_1, x_2) F''_{x_1x_2}(u_1 - x_1, u_2 - x_2) \mathrm{d}x_1 \mathrm{d}x_2 \\ &= \mu_\Phi(E_{l_1 l_2}) F(u_1 - l_1, u_2 - l_2) - \int_0^{l_1} \mu_\Phi(E_{x_1 l_2}) F'_{x_1}(u_1 - x_1, u_2 - l_2) \mathrm{d}x_1 \\ &\quad - \int_0^{l_2} \mu_\Phi(E_{l_1 x_2}) F'_{x_2}(u_1 - l_1, u_2 - x_2) \mathrm{d}x_2 + \int_0^{l_2}\int_0^{l_1} \mu_\Phi(E_{x_1 x_2}) F''_{x_1x_2}(u_1 - x_1, u_2 - x_2) \mathrm{d}x_1 \mathrm{d}x_2\end{aligned}$$

where $E_{ab} = [0, a] \times [0, b]$. Now by Cauchy–Schwartz inequality

$$\left[\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\int_{B_{11}}\mu_{\hat{F}_n}(C_{11}-\boldsymbol{x})\mathrm{d}\mu_{\Phi}(\boldsymbol{x})}-\sqrt{\int_{B_{11}}\mu_{\hat{F}_n}(C_{11}-\boldsymbol{x})\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2\mathrm{d}\mu\right]^{1/2}$$
$$=\left[\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\int_{[0,\boldsymbol{l}]}\mu_{\hat{F}_n}([0,\boldsymbol{u}-\boldsymbol{x}))\mathrm{d}\mu_{\Phi}(\boldsymbol{x})}-\sqrt{\int_{[0,\boldsymbol{l})}\mu_{\hat{F}_n}([0,\boldsymbol{u}-\boldsymbol{x}))\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2\mathrm{d}\mu\right]^{1/2}$$
$$\le\left[\int_{\mathbb{R}_+^2}\left(\mu_{\Phi}^{\frac{1}{2}}(E_{l_1l_2})-\mu_{\hat{\Phi}_n}^{\frac{1}{2}}(E_{l_1l_2})\right)^2 g_{\boldsymbol{L}}(\boldsymbol{l})\mathrm{d}\boldsymbol{l}\right]^{1/2}+\left[\int_{\mathbb{R}_+^2}\left(\mu_{\Phi}^{\frac{1}{2}}(E_{l_1l_2})-\mu_{\hat{\Phi}_n}^{\frac{1}{2}}(E_{l_1l_2})\right)^2 g_{L_2}(l_2)\mathrm{d}l_1\mathrm{d}l_2\right]^{1/2}$$
$$+\left[\int_{\mathbb{R}_+^2}\left(\mu_{\Phi}^{\frac{1}{2}}(E_{l_1l_2})-\mu_{\hat{\Phi}_n}^{\frac{1}{2}}(E_{l_1l_2})\right)^2 g_{L_1}(l_1)\mathrm{d}l_1\mathrm{d}l_2\right]^{1/2}+\left[\int_{\mathbb{R}_+^2}\left(\mu_{\Phi}^{\frac{1}{2}}(E_{l_1l_2})-\mu_{\hat{\Phi}_n}^{\frac{1}{2}}(E_{l_1l_2})\right)^2\mathrm{d}l_1\mathrm{d}l_2\right]^{1/2}.$$

From the assumptions (A6) and Theorem 4.3 in Deng and Fang [19],

$$\left[\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\int_{B_{11}}\mu_{\hat{F}_n}(C_{11}-\boldsymbol{x})\mathrm{d}\mu_{\Phi}(\boldsymbol{x})}-\sqrt{\int_{B_{11}}\mu_{\hat{F}_n}(C_{11}-\boldsymbol{x})\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2\mathrm{d}\mu\right]^{1/2}$$
$$=O(n^{-\frac{3}{10}}(\log n)^{\frac{2}{5}})=O_p(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}). \tag{6.2}$$

Further for any $\boldsymbol{a},\boldsymbol{b}\in\mathbb{R}_+^2$ with $\boldsymbol{a}<\boldsymbol{b}$, we have that

$$\mu_F([\boldsymbol{a},\boldsymbol{b}])=\mu_F([0,\boldsymbol{b}])-\mu_F([0,a_1]\times[0,b_2])-\mu_F([0,a_2]\times[0,b_1])+\mu_F([0,\boldsymbol{a}]).$$

Similarly, the same results can be obtained for $\mu_{\Phi}([\boldsymbol{a},\boldsymbol{b}])$. Note that B_{qr} and C_{st} $(q,r,s,t=1,2,3)$ can be expressed into the form of the rectangles $[\boldsymbol{a},\boldsymbol{b}]$ and $[\boldsymbol{c},\boldsymbol{d}]$. Therefore, we have that for any B_{qr} and C_{st},

$$\int_{B_{qr}(\boldsymbol{l},\boldsymbol{r})}\mu_F(C_{st}(\boldsymbol{u},\boldsymbol{v})-\boldsymbol{x})\mathrm{d}\mu_{\Phi}(\boldsymbol{x})=\int_{[\boldsymbol{a},\boldsymbol{b}]}\mu_F([\boldsymbol{c},\boldsymbol{d}]-\boldsymbol{x})\mathrm{d}\mu_{\Phi}(\boldsymbol{x})$$
$$=\int_{[0,\boldsymbol{b}]}-\int_{[0,a_1]\times[0,b_2]}-\int_{[0,a_2]\times[0,b_1]}+\int_{[0,\boldsymbol{a}]}[\mu_F([0,\boldsymbol{d}-\boldsymbol{x}])-\mu_F([0,c_1-x_1]\times[0,d_2-x_2])$$
$$-\ \mu_F([0,c_2-x_1]\times[0,d_1-x_2])+\mu_F([0,\boldsymbol{c}-\boldsymbol{x}])]\,\mathrm{d}\mu_{\Phi}(\boldsymbol{x}). \tag{6.3}$$

Therefore, from (6.2), (6.3) and Cauchy–Schwartz inequality, one can prove that for any B_{qr} and C_{st},

$$\left[\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\int_{B_{qr}}\mu_{\hat{F}_n}(C_{st}-\boldsymbol{x})\mathrm{d}\mu_{\Phi}(\boldsymbol{x})}-\sqrt{\int_{B_{qr}}\mu_{\hat{F}_n}(C_{st}-\boldsymbol{x})\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2\mathrm{d}\mu\right]^{1/2}$$
$$=O_p(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}).$$

The corollary follows. □

Proof of Theorem 4.3. We only prove the theorem for distance $d_{\boldsymbol{RU}}$ under the different censoring mechanism. The proofs for other distances are similar and omitted. From the definition of ϕ_F, for any $q,r,s,t=1,2,3$, we have that

$$\int_{\mathbb{R}_+^8}\left(\mu_{\hat{H}_n}[D_{qrst}(\boldsymbol{w})]-\mu_H[D_{qrst}(\boldsymbol{w})]\right)^2\mathrm{d}G(\boldsymbol{w})$$
$$\le 4\int_{\mathbb{R}_+^8}\left(\sqrt{\mu_{\hat{H}_n}[D_{qrst}(\boldsymbol{w})]}-\sqrt{\mu_H[D_{qrst}(\boldsymbol{w})]}\right)^2\mathrm{d}G(\boldsymbol{w})$$
$$=4\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\delta_q\gamma_r\lambda_s\pi_t\mu_{\hat{H}_n}[D_{qrst}(\boldsymbol{w})]}-\sqrt{\delta_q\gamma_r\lambda_s\pi_t\mu_H[D_{qrst}(\boldsymbol{w})]}\right)^2\mathrm{d}\mu$$
$$\le 4h^2(\phi_{\hat{H}_n},\phi_H). \tag{6.4}$$

Now for distribution function H in $\mathbb{R}_+^4$, we have that

$$\begin{aligned} H(\boldsymbol{r},\boldsymbol{u}) &= \mu_H([0,\boldsymbol{r}]\times[0,\boldsymbol{u}]) = \mu_H([0,r_1]\times[0,r_2]\times[0,u_1]\times[0,u_2]) \\ &= \mu_H[D_{1111}(\boldsymbol{w})\cup D_{1211}(\boldsymbol{w})\cup D_{2111}(\boldsymbol{w})\cup D_{2211}(\boldsymbol{w})] \\ &= \mu_H[D_{1111}(\boldsymbol{w})]+\mu_H[D_{1211}(\boldsymbol{w})]+\mu_H[D_{2111}(\boldsymbol{w})]+\mu_H[D_{2211}(\boldsymbol{w})]. \end{aligned} \tag{6.5}$$

Therefore by (6.4), (6.5) and triangle inequality,

$$\begin{aligned} d_{\boldsymbol{RU}}(\hat{H}_n,H) &= \left[\int_{\mathbb{R}_+^4}\left\{\hat{H}_n(\boldsymbol{r},\boldsymbol{u})-H(\boldsymbol{r},\boldsymbol{u})\right\}^2 \mathrm{d}G_{\boldsymbol{RU}}(\boldsymbol{r},\boldsymbol{u})\right]^{1/2} \\ &= \left[\int_{\mathbb{R}_+^8}\left\{\mu_{\hat{H}_n}([0,\boldsymbol{r}]\times[0,\boldsymbol{u}])-\mu_H([0,\boldsymbol{r}]\times[0,\boldsymbol{u}])\right\}^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\le \left[\int_{\mathbb{R}_+^8}\left\{\mu_{\hat{H}_n}[D_{1111}(\boldsymbol{w})]-\mu_H[D_{1111}(\boldsymbol{w})]\right\}^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\quad + \left[\int_{\mathbb{R}_+^8}\left\{\mu_{\hat{H}_n}[D_{1211}(\boldsymbol{w})]-\mu_H[D_{1211}(\boldsymbol{w})]\right\}^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\quad + \left[\int_{\mathbb{R}_+^8}\left\{\mu_{\hat{H}_n}[D_{2111}(\boldsymbol{w})]-\mu_H[D_{2111}(\boldsymbol{w})]\right\}^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\quad + \left[\int_{\mathbb{R}_+^8}\left\{\mu_{\hat{H}_n}[D_{2211}(\boldsymbol{w})]-\mu_H[D_{2211}(\boldsymbol{w})]\right\}^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\le 16h^2(\phi_{\hat{H}_n},\phi_H). \end{aligned} \tag{6.6}$$

Now, from (6.6) and Theorem 4.1, Theorem 4.3 is obtained. □

Proof of Theorem 4.4. Similar to the proof of Theorem 4.3, from Theorem 4.2 and triangle inequality, we have that

$$\begin{aligned} &\left[\int_{\mathbb{R}_+^2}\left(\int_{\mathbb{R}_+^2}[\hat{F}_n(\boldsymbol{u}-\boldsymbol{x})-F(\boldsymbol{u}-\boldsymbol{x})]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G_{\boldsymbol{U}}(\boldsymbol{u})\right]^{1/2} \\ &= \left[\int_{\mathbb{R}_+^8}\left(\sum_{q,r=1}^{3}\int_{B_{qr}}[\mu_{\hat{F}_n}([0,\boldsymbol{u}-\boldsymbol{x}])-\mu_F([0,\boldsymbol{u}-\boldsymbol{x}])]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &= \left[\int_{\mathbb{R}_+^8}\left(\sum_{q,r=1}^{3}\int_{B_{qr}}\left(\mu_{\hat{F}_n}[C_{11}-\boldsymbol{x}]-\mu_F[C_{11}-\boldsymbol{x}]\right)\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G(\boldsymbol{w})\right]^{1/2} \\ &\le \sum_{q,r=1}^{3}\left(\int_{\mathbb{R}_+^8}\left(\int_{B_{qr}}\left(\mu_{\hat{F}_n}[C_{11}-x]-\mu_F[C_{11}-x]\right)\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G(\boldsymbol{w})\right)^{1/2} \\ &\le 2\sum_{q,r=1}^{3}\left(\int_{\mathbb{R}_+^8\times\{0,1\}^8}\left(\sqrt{\delta_q\gamma_r\lambda_1\pi_1\int_{B_{qr}}\mu_{\hat{F}_n}[C_{11}-x]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}-\sqrt{\delta_q\gamma_r\lambda_1\pi_1\int_{B_{qr}}\mu_F[C_{11}-x]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})}\right)^2 \mathrm{d}\mu\right)^{1/2} \\ &\le 18h(\phi_{\hat{F}_n},\phi_F) = O_p(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}). \end{aligned}$$

Similarly, we can prove that

$$\left[\int_{\mathbb{R}_+^2}\left(\int_{\mathbb{R}_+^2}[\hat{F}_n(\tilde{\boldsymbol{v}}-\boldsymbol{x})-F(\tilde{\boldsymbol{v}}-\boldsymbol{x})]\mathrm{d}\mu_{\hat{\Phi}_n}(\boldsymbol{x})\right)^2 \mathrm{d}G_{\boldsymbol{V}}(\boldsymbol{v})\right]^{1/2} = O_p(n^{-\frac{5}{18}}(\log n)^{\frac{8}{9}}). \quad \square$$

Acknowledgments

Authors are very grateful for referee's comments and suggestions which greatly improve the readability and quality of the current paper. The first author's research is partly supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] M.Y. Kim, X. Xue, The analysis of multivariate interval-censored survival data, Statistics in Medicine 21 (2002) 3715–3726.
[2] W.B. Goggins, D.M. Finkelstein, A proportional hazards model for multivariate interval-censored failure time data, Biometrics 56 (2000) 940–943.
[3] G. Jones, D.M. Rocke, Multivariate survival analysis with doubly-censored data: Application to the assessment of Accutane treatment for fibrodysplasia ossificans progressiva, Statistics in Medicine 21 (2002) 2547–2562.
[4] G.Y.C. Wong, Q.Q. Yu, Generalized MLE of a joint distribution function with multivariate interval-censored data, Journal of Multivariate Analysis 69 (1999) 155–166.
[5] S.H. Yu, Q.Q. Yu, G.Y.C. Wong, Consistency of the generalized MLE of a joint distribution function with multivariate interval-censored data, Journal of Multivariate Analysis 97 (2006) 720–732.
[6] V. DeGruttola, S.W. Lagakos, Analysis of doubly-censored survival data, with application to AIDS, Biometrics 45 (1989) 1–11.
[7] G. Gómez, M.L. Calle, Nonparametric estimation with doubly censored data, Journal of Applied Statistics 26 (1999) 45–58.
[8] G. Gómez, S.W. Lagakos, Estimation of the infection time and latency distribution of AIDS with doubly censored data, Biometrics 50 (1994) 204–212.
[9] J. Sun, Empirical estimation of a distribution function with truncated and doubly interval-censored data and its application to AIDS studies, Biometrics 51 (1995) 1094–1104.
[10] J. Sun, Self-consistency estimation of distributions based on truncated and doubly interval-censored data with applications to AIDS cohort studies, Lifetime Data Analysis 3 (4) (1997) 305–313.
[11] H. Fang, J. Sun, Consistency of nonparametric maximum likelihood estimation of a distribution function based on doubly interval-censored failure time data, Statistics and Probability Letters 55 (2001) 311–318.
[12] A. Komárek, E. Lesaffre, Bayesian semi-parametric accelerated failure tme model for paired doubly interval-censored data, Statistical Modelling 6 (2006) 3–22.
[13] A. Komárek, E. Lesaffre, Bayesian accelerated failure time model with multivariate doubly interval-censored data and flexible distributional assumptions, Journal of the American Statistical Association 103 (2008) 523–533.
[14] A. Komárek, E. Lesaffre, T. Härkänen, D. Declerck, I.J. Virtanen, A Bayesian analysis of multivariate doubly-interval-censored data, Biostatistics 6 (2005) 145–155.
[15] P. Groeneboom, Lectures on inverse problems, in: Lectures on Probability Theory and Statistics (Ecole d'Eté de Probabilités de Saint-Flour XXIV-1994), in: Lectures Notes in Mathematics, vol. 1648, Springer, Berlin, 1996.
[16] D. Deng, H. Fang, J. Sun, Nonparametric estimation for doubly censored failure time data, Journal of Nonparametric Statistics (2008) (submitted for publication).
[17] P. Groeneboom, J.A. Wellner, Information Bounds and Nonparametric Maximum Likelihood Estimation, Birkhäuser Verlag, Basel, 1992.
[18] S. van De Geer, Rates of convergence for the maximum likelihood estimator in mixture models, Journal of Nonparametric Statistics 6 (1996) 293–310.
[19] D. Deng, H. Fang, On nonparametric maximum likelihood estimations of multivariate distribution function based on interval-censored data, Communications in Statistics - Theory and Methods 38 (2008) 54–74.
[20] R.B. Geskus, P. Groeneboom, Asymptotically optional estimation of smooth functionals for interval censoring, Part 2, Statistica Neerlandica 51 (1997) 201–219.
[21] R.B. Geskus, P. Groeneboom, Asymptotically optional estimation of smooth functional for interval censoring, Case 2, The Annals of Statistics 27 (1999) 627–674.
[22] R.A. Betensky, D.M. Finkelstein, A non-parametric maximum likelihood estimator for bivariate interval censored data, Statistics in Medicine 18 (1999) 3089–3100.
[23] R. Gentleman, A.C. Vandal, Computational algorithms for censored data problems using intersection graphs, Journal of Computational & Graphical Statistics 10 (2001) 403–421.
[24] R. Gentleman, A.C. Vandal, Nonparametric estimation of the bivariate CDF for arbitrarily censored data, The Canadian Journal of Statistics 30 (2002) 557–571.
[25] K. Bogaerts, E. Lesaffre, A new, fast algorithm to find the regions of possible support for bivariate interval-censored data, Journal of Computational Graphical Statistics 13 (2004) 330–340.