PARTIALLY MONOTONE TENSOR SPLINE ESTIMATION OF THE JOINT DISTRIBUTION FUNCTION WITH BIVARIATE CURRENT STATUS DATA

BY YUAN WU AND YING ZHANG

University of California San Diego and University of Iowa

The analysis of the joint cumulative distribution function (CDF) with bivariate event time data is a challenging problem both theoretically and numerically. This paper develops a tensor spline-based sieve maximum likelihood estimation method to estimate the joint CDF with bivariate current status data. The I-splines are used to approximate the joint CDF in order to simplify the numerical computation of a constrained maximum likelihood estimation problem. The generalized gradient projection algorithm is used to compute the constrained optimization problem. Based on the properties of B-spline basis functions it is shown that the proposed tensor spline-based nonparametric sieve maximum likelihood estimator is consistent with a rate of convergence potentially better than $n^{1/3}$ under some mild regularity conditions. The simulation studies with moderate sample sizes are carried out to demonstrate that the finite sample performance of the proposed estimator is generally satisfactory.

1. Introduction. In some applications, observation of random event time T is restricted to the knowledge of whether or not T exceeds a random monitoring time C. This type of data is known as current status data and sometimes referred to as interval censoring case 1 [Groeneboom and Wellner (1992)]. Current status data arise naturally in many applications; see, for example, the animal tumorigenicity experiments by Hoel and Walburg (1972) and Finkelstein and Wolfe (1985); the social demographic studies of the distribution of the age at weaning by Diamond, McDonald and Shah (1986), Diamond and McDonald (1991) and Grummer-Strawn (1993); and the studies of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS) by Shiboski and Jewell (1992) and Jewell, Malani and Vittinghoff (1994).

The univariate current status data have been thoroughly studied in the statistical literature. Groeneboom and Wellner (1992) and Huang and Wellner (1995) studied the asymptotic properties of the nonparametrc maximum likelihood estimator (NPMLE) of the CDF with current status data. Huang (1996) considered Cox proportional hazards model with current status data and showed that the maximum

likelihood estimator (MLE) of the regression parameter is asymptotically normal with $\sqrt{n}$ convergence rate, even through the MLE of the baseline cumulative hazard function only converges at $n^{1/3}$ rate.

Bivariate event time data occur in many applications as well. For example, in an Australian twin study [Duffy, Martin and Matthews (1990)], the researchers were interested in times to a certain event such as a disease or a disease-related symptom in both twins. NPMLE of the joint CDF of the correlated event times with bivariate right censored data was studied by Dabrowska (1988), Prentice and Cai (1992), Pruitt (1991), van der Laan (1996) and Quale, van der Laan and Robins (2006). As an alternative, Kooperberg (1998) developed a tensor spline estimation of the logarithm of joint density function with bivariate right censored data. However, asymptotic properties of Kooperberg's estimate are unknown. Shih and Louis (1995) proposed a two-stage semiparametric estimation procedure to study the joint CDF with bivariate right censored data, in which the joint distribution of the two event times is assumed to follow a bivariate Copula model [Nelsen (2006)].

For bivariate interval censored data, the conventional NPMLE was originally studied by Betensky and Finkelstein (1999) and followed by Wong and Yu (1999), Gentleman and Vandal (2001), Song (2001) and Maathuis (2005). A typical numerical algorithm for computing the NPMLE constitutes two steps [Song (2001) and Maathuis (2005)]: in the first stage the algorithm searches for small rectangles with nonzero probability mass; in the second stage those nonzero probability masses are estimated by maximizing the log likelihood with a reduced number of unknown quantities. Sun, Wang and Sun (2006) and Wu and Gao (2011) adopted the same idea used by Shih and Louis (1995) to study the joint distribution of CDF for bivariate interval censored data with Copula models.

This paper studies bivariate current status data, a special type of bivariate interval censored data. Let (T_1, T_2) be the two event times of interest and (C_1, C_2) the two corresponding random monitoring times. In this setting, the observation of bivariate current status data consists of

$$(1.1) \qquad X = \left(C_1, C_2, \Delta_1 = I(T_1 \leq C_1), \Delta_2 = I(T_2 \leq C_2)\right),$$

where $I(\cdot)$ is the indicator function. For bivariate current status data, Wang and Ding (2000) adopted the same approach proposed by Shih and Louis (1995) to study the association between the onset times of hypertension and diabetes for Taiwanese in a demographic screening study. In a study on HIV transmission, Jewell, van der Laan and Lei (2005) investigated the relationship between the time to HIV infection to the other partner and the time to diagnosis of AIDS for the index case by studying some smooth functionals of the marginal CDFs. In both examples, the bivariate event times have the same monitoring time, that is, $C_1 = C_2 = C$. In this paper, we propose a tensor spline-based sieve maximum likelihood estimation of the joint CDF for bivariate current status data in a general scenario in which C_1 and C_2 are allowed to be different. The proposed method is shown to

have a rate of convergence potentially better than $n^{1/3}$ and it can simultaneously estimate the two marginal CDFs along with the joint CDF.

The rest of the paper is organized as follows. Section 2 characterizes the spline-based sieve MLE $\hat{\tau}_n = (\hat{F}_n, \hat{F}_{n,1}, \hat{F}_{n,2})$, where $\hat{F}_n$ is the tensor spline-based estimator of the joint CDF, and $\hat{F}_{n,1}$ and $\hat{F}_{n,2}$ are the spline-based estimators of the two corresponding marginal CDFs. Section 3 presents two asymptotic properties (consistency and convergence rate) of the proposed spline-based sieve MLE. Section 4 discusses the computation of the spline-based estimators. Section 5 carries out a set of simulation studies to examine the finite sample performance of the proposed method and compares it to the conventional NMPLE computed with the algorithm proposed by Maathuis (2005). Section 6 summarizes our findings and discusses some related problems. Section 7 provides proofs of the theorems stated in the early section. Details of some technical lemmas that are used for proving the theorem and their proofs are included in a supplementary file.

2. Tensor spline-based sieve maximum likelihood estimation method.

2.1. *Maximum likelihood estimation.* We consider a sample of n i.i.d. bivariate current status data denoted in (1.1), $\{(c_{1,k}, c_{2,k}, \delta_{1,k}, \delta_{2,k}) : k = 1, 2, \ldots, n\}$. Suppose that (T_1, T_2) and (C_1, C_2) are independent. Then the log-likelihood for the observed data can be expressed by

$$
\begin{aligned}
l_n(\cdot; \text{data}) = \sum_{k=1}^{n} \big\{ & \delta_{1,k}\delta_{2,k} \log P(T_1 \leq c_{1,k}, T_2 \leq c_{2,k}) \\
& + \delta_{1,k}(1 - \delta_{2,k}) \log P(T_1 \leq c_{1,k}, T_2 > c_{2,k}) \\
& + (1 - \delta_{1,k})\delta_{2,k} \log P(T_1 > c_{1,k}, T_2 \leq c_{2,k}) \\
& + (1 - \delta_{1,k})(1 - \delta_{2,k}) \log P(T_1 > c_{1,k}, T_2 > c_{2,k}) \big\}.
\end{aligned}
$$

(2.1)

Denote F the joint CDF of event times (T_1, T_2) and F_1 and F_2 the marginal CDFs of F, respectively. The log-likelihood (2.1) can be rewritten as

$$
\begin{aligned}
l_n(F, F_1, F_2; \text{data}) = \sum_{k=1}^{n} \big\{ & \delta_{1,k}\delta_{2,k} \log F(c_{1,k}, c_{2,k}) \\
& + \delta_{1,k}(1 - \delta_{2,k}) \log\big(F_1(c_{1,k}) - F(c_{1,k}, c_{2,k})\big) \\
& + (1 - \delta_{1,k})\delta_{2,k} \log\big(F_2(c_{2,k}) - F(c_{1,k}, c_{2,k})\big) \\
& + (1 - \delta_{1,k})(1 - \delta_{2,k}) \log\big(1 - F_1(c_{1,k}) - F_2(c_{2,k}) \\
& \hspace{10em} + F(c_{1,k}, c_{2,k})\big) \big\}.
\end{aligned}
$$

(2.2)

A class of real-valued functions defined in a bounded region $[L_1, U_1] \times [L_2, U_2]$ is denoted by

$$
\mathcal{F} = \big\{ (F(s, t), F_1(s), F_2(t)) : \text{for } (s, t) \in [L_1, U_1] \times [L_2, U_2] \big\},
$$

where F, F_1 and F_2 satisfy the following conditions in (2.3):

$$0 \leq F(s, t),$$
$$F(s', t) \leq F(s'', t),$$
$$F(s, t') \leq F(s, t''),$$
$$[F(s'', t'') - F(s', t'')] - [(F(s'', t') - F(s', t')] \geq 0,$$

(2.3)
$$F_1(s) - F(s, t) \geq 0,$$
$$F_2(t) - F(s, t) \geq 0,$$
$$[F_1(s'') - F_1(s')] - [F(s'', t) - F(s', t)] \geq 0,$$
$$[F_2(t'') - F_2(t')] - [F(s, t'') - F(s, t')] \geq 0,$$
$$[1 - F_1(s)] - [F_2(t) - F(s, t)] \geq 0$$

for $s' \leq s''$ with s' and s'' on $[L_1, U_1]$, and $t' \leq t''$ with t' and t'' on $[L_2, U_2]$.

It can be easily argued that if F is a joint CDF and F_1 and F_2 are its two corresponding marginal CDFs, $(F, F_1, F_2) \in \mathcal{F}$. On the other hand, for any $(F, F_1, F_2) \in \mathcal{F}$ there exists a bivariate distribution such that F is the joint CDF and F_1 and F_2 are its two marginal CDFs. Throughout this paper, F_0, $F_{0,1}$ and $F_{0,2}$ are denoted for the true joint and marginal CDFs, respectively. The NPMLE for $(F_0, F_{0,1}, F_{0,2})$ is defined as

(2.4)
$$(\hat{F}_n, \hat{F}_{n,1}, \hat{F}_{n,2}) = \underset{(F, F_1, F_2) \in \mathcal{F}}{\arg \max} \, l_n(F, F_1, F_2; \text{data}).$$

The NPNLE of (2.4) is, in general, a challenging problem both numerically and theoretically. The conventional NPMLE of F is constructed by a larger number of unknown quantities representing the masses in small rectangles. Solving for the NPMLE needs to perform a constrained high-dimensional nonlinear optimization [Betensky and Finkelstein (1999), Wong and Yu (1999), Gentleman and Vandal (2001), Song (2001), Maathuis (2005)]. Though the conventional NPMLE of (2.4) can be efficiently computed using the algorithm developed by Maathuis (2005), it is, however, well known that the conventional NPMLE is not uniquely determined [Song (2001), Maathuis (2005)]. In an unpublished Ph.D. dissertation, Song (2001) showed that the conventional NPMLE of joint CDF with bivariate current status data can achieve a global rate of convergence of $n^{3/10}$ in Hellinger distance, which is slightly slower than that of the NPMLE with univariate current status data.

This paper adopts a popular dimension reduction method through spline-based sieve maximum likelihood estimation. The main idea of the spline-based sieve method is to solve problem (2.4) in a subclass of $\mathcal{F}$ that "approximates" to $\mathcal{F}$ when sample size enlarges. The advantages of the proposed method are that the spline-based sieve MLE is unique, and it is easy to compute and analyze. The univariate spline-based sieve MLEs for various models were studied by Shen (1998),

Lu, Zhang and Huang (2007, 2009), Zhang, Hua and Huang (2010) and Lu (2010). Other problems related to applications of univariate shape-constrained spline estimations have recently been studied as well. For example, Meyer (2008) studied the inference using shape-restricted regression spline functions and Wang and Shen (2010) studied B-spline approximation for a monotone univariate regression function based on grouped data. For analyzing bivariate distributions, the tensor spline approach [de Boor (2001)] has been studied by Stone (1994) in a nonparametric regression setting, by Koo (1996) and Scott (1992) in a multivariate density estimation without censored data and, as noted in Section 1, by Kooperberg (1998) in the bivariate density estimation with bivariate right censored data. Recently, an application of the tensor B-spline estimation of a bivariate monotone function has also been investigated by Wang and Taylor (2004) in a biomedical study.

In this paper, we propose a partially monotone tensor spline estimation of the joint CDF. To solve problem (2.4), the unknown joint CDF is approximated by a linear combination of the tensor spline basis functions, and its two marginal CDFs are approximated by linear combinations of spline basis functions as well. Then the problem converts to maximizing the sieve log likelihood with respect to the unknown spline coefficients subject to a set of inequality constraints.

2.2. *B-spline-based estimation.* In this section, the spline-based sieve maximum likelihood estimation problem is reformulated as a constrained optimization problem with respect to the coefficients of B-spline functions.

Suppose two sets of the normalized B-spline basis functions of order l [Schumaker (1981)], $\{N_i^{(1),l}(s)\}_{i=1}^{p_n}$ and $\{N_j^{(2),l}(t)\}_{j=1}^{q_n}$ are constructed in $[L_1, U_1] \times [L_2, U_2]$ with the knot sequence $\{u_i\}_{i=1}^{p_n+l}$ satisfying $L_1 = u_1 = \cdots = u_l < u_{l+1} < \cdots < u_{p_n} < u_{p_n+1} = \cdots = u_{p_n+l} = U_1$ and the knot sequence $\{v_j\}_{j=1}^{q_n+l}$ satisfying $L_2 = v_1 = \cdots = v_l < v_{l+1} < \cdots < v_{q_n} < v_{q_n+1} = \cdots = v_{q_n+l} = U_2$, where $p_n = O(n^v)$ and $q_n = O(n^v)$ for some $0 < v < 1$.

Define

$$\Omega_n = \left\{ \tau_n = (F_n, F_{n,1}, F_{n,2}) : F_n(s, t) = \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t), \right.$$

$$\left. F_{n,1}(s) = \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s), \quad F_{n,2}(t) = \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(t) \right\},$$

with $\underline{\alpha} = (\alpha_{1,1}, \ldots, \alpha_{p_n,q_n})$, $\underline{\beta} = (\beta_1, \ldots, \beta_{p_n})$ and $\underline{\gamma} = (\gamma_1, \ldots, \gamma_{q_n})$ subject to the following conditions in (2.5):

$$\alpha_{1,1} \geq 0,$$

$$\alpha_{1,j+1} - \alpha_{1,j} \geq 0 \qquad \text{for } j = 1, \ldots, q_n - 1,$$

$$(2.5) \quad \alpha_{i+1,1} - \alpha_{i,1} \geq 0 \qquad \text{for } i = 1, \ldots, p_n - 1,$$

$$(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j}) \geq 0$$
$$\text{for } i = 1, \ldots, p_n - 1, \, j = 1, \ldots, q_n - 1,$$
$$\beta_1 - \alpha_{1,q_n} \geq 0, \qquad \gamma_1 - \alpha_{p_n,1} \geq 0,$$
$$(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \qquad \text{for } i = 1, \ldots, p_n - 1,$$
$$(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \qquad \text{for } j = 1, \ldots, q_n - 1,$$
$$\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.$$

(2.5) is established corresponding to the constraints given in (2.3). Using the properties of B-spline, a straightforward algebra yields $\Omega_n \subset \mathcal{F}$. To obtain the tensor B-spline-based sieve likelihood with bivariate current status data, $\tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega_n$ is substituted into (2.2) to result in

$$\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$$

$$= \sum_{k=1}^{n} \left\{ \delta_{1,k} \delta_{2,k} \log \sum_{i=1}^{p_n} \sum_{j=2}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \right.$$

$$+ \delta_{1,k}(1 - \delta_{2,k}) \log \left\{ \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \right.$$

$$\left. - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \right\}$$

(2.6) $$+ (1 - \delta_{1,k}) \delta_{2,k} \log \left\{ \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k}) \right.$$

$$\left. - \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \right\}$$

$$+ (1 - \delta_{1,k})(1 - \delta_{2,k}) \log \left\{ 1 - \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(c_{1,k}) \right.$$

$$- \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(c_{2,k})$$

$$\left. \left. + \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(c_{1,k}) N_j^{(2),l}(c_{2,k}) \right\} \right\}.$$

Hence, the proposed sieve MLE with the B-spline basis functions is the maximizer of (2.6) over Ω_n.

REMARK 2.1.   The spline-based sieve MLE in Ω_n is the MLE defined in a sub-class of $\mathcal{F}$. Hence, the spline-based sieve MLE is anticipated to have good asymptotic properties if this sub-class "approximates" to $\mathcal{F}$ as $n \to \infty$.

3. Asymptotic properties.

In this section, we describe asymptotic properties of the tensor spline-based sieve MLE of joint CDF with bivariate current status data. Study of the asymptotic properties of the proposed sieve estimator requires some regularity conditions, regarding the event times, observation times and the choice of knot sequences. The following conditions sufficiently guarantee the results in the forthcoming theorems.

Regularity conditions:

(C1) Both $\frac{\partial F_0(s,t)}{\partial s}$ and $\frac{\partial F_0(s,t)}{\partial t}$ have positive lower bounds in $[L_1, U_1] \times [L_2, U_2]$.

(C2) $\frac{\partial^2 F_0(s,t)}{\partial s\, \partial t}$ has a positive lower bound b_0 in $[L_1, U_1] \times [L_2, U_2]$.

(C3) $F_0(s, t)$, $F_{0,1}(s)$ and $F_{0,2}(t)$ are all continuous differentiable up to order p in domain $[L_1, U_1] \times [L_2, U_2]$, $[L_1, U_1]$ and $[L_2, U_2]$, respectively.

(C4) The observation times (C_1, C_2) follow a bivariate distribution defined in $[l_1, u_1] \times [l_2, u_2]$, with $l_1 > L_1, u_1 < U_1, l_2 > L_2$ and $u_2 < U_2$.

(C5) The density of the joint distribution of (C_1, C_2) has a positive lower bound in $[l_1, u_1] \times [l_2, u_2]$.

(C6) The knot sequences $\{u_i\}_{i=1}^{p_n+l}$ and $\{v_j\}_{j=1}^{q_n+l}$ of the B-spline basis functions, $\{N_i^{(1),l}\}_{i=1}^{p_n}$ and $\{N_j^{(2),l}\}_{j=1}^{q_n}$, satisfy that both $\frac{\min_i \Delta_i^{(u)}}{\max_i \Delta_i^{(u)}}$ and $\frac{\min_j \Delta_j^{(v)}}{\max_j \Delta_j^{(v)}}$ have positive lower bounds, where $\Delta_i^{(u)} = u_{i+1} - u_i$ for $i = l, \ldots, p_n$ and $\Delta_j^{(v)} = v_{j+1} - v_j$ for $j = l, \ldots, q_n$.

REMARK 3.1.   (C1) implies that $\frac{dF_{0,1}(s)}{ds}$ and $\frac{dF_{0,2}(t)}{dt}$ have positive lower bounds on $[L_1, U_1]$ and $[L_2, U_2]$, respectively. (C3) implies that both $\frac{\partial F_0(s,t)}{\partial s}$ and $\frac{\partial F_0(s,t)}{\partial t}$ have positive upper bounds in $[L_1, U_1] \times [L_2, U_2]$; $\frac{dF_{0,1}(s)}{ds}$ and $\frac{dF_{0,2}(t)}{dt}$ have positive upper bounds on $[L_1, U_1]$ and $[L_2, U_2]$, respectively.

Let

$$\Omega_{n,1} = \left\{ \tau = (F_n, F_{n,1}, F_{n,2}) : F_n(s, t) = \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \alpha_{i,j} N_i^{(1),l}(s) N_j^{(2),l}(t), \right.$$

$$\left. F_{n,1}(s) = \sum_{i=1}^{p_n} \beta_i N_i^{(1),l}(s), F_{n,2}(t) = \sum_{j=1}^{q_n} \gamma_j N_j^{(2),l}(t) \right\},$$

with $\underline{\alpha} = (\alpha_{1,1}, \ldots, \alpha_{p_n,q_n})$, $\underline{\beta} = (\beta_1, \ldots, \beta_{p_n})$ and $\underline{\gamma} = (\gamma_1, \ldots, \gamma_{q_n})$ subject to the following conditions in (3.1):

$$\alpha_{1,1} \geq 0,$$

$$\alpha_{1,j+1} - \alpha_{1,j} \geq 0 \qquad \text{for } j = 1, \ldots, q_n - 1,$$

$$\alpha_{i+1,1} - \alpha_{i,1} \geq 0 \qquad \text{for } i = 1, \ldots, p_n - 1,$$

$$(\alpha_{i+1,j+1} - \alpha_{i+1,j}) - (\alpha_{i,j+1} - \alpha_{i,j})$$

$$\geq \frac{b_0 \min_{i_1:l \leq i_1 \leq p_n} \Delta_{i_1}^{(u)} \min_{j_1:l \leq j_1 \leq q_n} \Delta_{j_1}^{(v)}}{l^2}$$

(3.1)

$$\text{for } i = 1, \ldots, p_n - 1, j = 1, \ldots, q_n - 1,$$

$$\beta_1 - \alpha_{1,q_n} \geq 0, \qquad \gamma_1 - \alpha_{p_n,1} \geq 0,$$

$$(\beta_{i+1} - \beta_i) - (\alpha_{i+1,q_n} - \alpha_{i,q_n}) \geq 0 \qquad \text{for } i = 1, \ldots, p_n - 1,$$

$$(\gamma_{j+1} - \gamma_j) - (\alpha_{p_n,j+1} - \alpha_{p_n,j}) \geq 0 \qquad \text{for } j = 1, \ldots, q_n - 1,$$

$$\beta_{p_n} + \gamma_{q_n} - \alpha_{p_n,q_n} \leq 1.$$

REMARK 3.2. Note that $\Omega_{n,1}$ is a sub-class of Ω_n due to the change from the forth inequality of (2.5) to that of (3.1). The choice of $\Omega_{n,1}$ is mainly for the technical convenience in justifying the asymptotic properties. In the forth inequality of (3.1), b_0 is the positive lower bound of $\frac{\partial^2 F_0(s,t)}{\partial s\, \partial t}$ stated in (C2). This inequality will guarantee that $\frac{\partial^2 F_n(s,t)}{\partial s\, \partial t}$ also has a positive lower bound which is necessary for the proof of Lemma 0.1 in the supplemental article [Wu and Zhang (2012)]. It is obvious that as sample n increases to infinity, the right-hand side of the forth inequality in (3.1) will approach to 0.

We study the asymptotic properties in the feasible region of the observation times: $[l_1, u_1] \times [l_2, u_2]$. Let $\Omega_n' = \{\tau \in \Omega_{n,1}, \text{ for } (s,t) \in [l_1, u_1] \times [l_2, u_2]\}$ and let $\tau_0 = (F_0(s,t), F_{0,1}(s), F_{0,2}(t))$ with $(s,t) \in [l_1, u_1] \times [l_2, u_2]$. Under (C4), the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over $\Omega_{n,1}$ is actually the maximization of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over Ω_n'. Throughout the study of asymptotic properties, we denote $\hat{\tau}_n$ the maximizer of $\tilde{l}_n(\underline{\alpha}, \underline{\beta}, \underline{\gamma}; \text{data})$ over Ω_n'.

Denote $L_r(Q)$ the norm associated with a probability measure Q which is defined as

$$\|f\|_{L_r(Q)} = (Q|f|^r)^{1/r} = \left(\int |f|^r \, dQ\right)^{1/r}.$$

In the following, $L_r(P_{C_1,C_2})$, $L_r(P_{C_1})$ and $L_r(P_{C_2})$ are denoted as the L_r-norms associated with the joint and marginal probability measures of the observation

times (C_1, C_2), respectively, and $L_r(P)$ is denoted as the L_r-norm associated with the joint probability measure P of observation and event times (T_1, T_2, C_1, C_2).

Based on the L_2-norms, the distance between $\tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega'_n$ and $\tau_0 = (F_0, F_{0,1}, F_{0,2})$ is defined as

$$d(\tau_n, \tau_0)$$
$$= \left(\|F_n - F_0\|^2_{L_2(P_{C_1, C_2})} + \|F_{n,1} - F_{0,1}\|^2_{L_2(P_{C_1})} + \|F_{n,2} - F_{0,2}\|^2_{L_2(P_{C_2})} \right)^{1/2}.$$

THEOREM 3.1. *Suppose* (C2)–(C6) *hold, and* $p_n = O(n^v)$, $q_n = O(n^v)$ *for* $v < 1$; *that is, the numbers of interior knots of knot sequences* $\{u_i\}_1^{p_n+l}$ *and* $\{v_j\}_1^{q_n+l}$ *are both in the order of* n^v *for* $v < 1$. *Then*

$$d(\hat{\tau}_n, \tau_0) \to_p 0 \qquad as \ n \to \infty.$$

THEOREM 3.2. *Suppose* (C1)–(C6) *hold, and* $p_n = O(n^v)$, $q_n = O(n^v)$ *for* $v < 1$; *that is, the numbers of interior knots of knot sequences* $\{u_i\}_1^{p_n+l}$ *and* $\{v_j\}_1^{q_n+l}$ *are both in the order of* n^v *for* $v < 1$. *Then*

$$d(\hat{\tau}_n, \tau_0) = O_p\left(n^{-\min\{pv, (1-2v)/2\}}\right).$$

REMARK 3.3. Theorem 3.2 implies that the optimal rate of convergence of the proposed estimator is $n^{p/(2(p+1))}$, achieved by letting $pv = (1 - 2v)/2$. This rate is equal to $n^{1/3}$ when $p = 2$ and improves as p (the degree of smoothness of the true joint distribution) increases. Nonetheless, the rate will never exceed $n^{1/2}$. The result of Theorem 3.2 also indicates that the proposed method potentially results in an estimate of the targeted joint CDF with a faster convergence rate than the conventional NPMLE method given in Song (2001).

4. Computation of the spline-based sieve MLE. For the B-spline-based sieve MLE, the constraint set (3.1) complicates the numerical implementation. We propose to compute the sieve MLE using I-spline basis functions for the sake of numerical convenience. The I-spline basis functions are defined by Ramsay (1988) as

$$(4.1) \quad I_i^l(s) = \begin{cases} 0, & i > j, \\ \sum_{m=i}^{j} (u_{m+l+1} - u_m) M_m^{l+1}(s)/(l+1), & j - l + 1 \leq i \leq j, \\ 1, & i < j - l + 1 \end{cases}$$

for $u_j \leq s < u_{j+1}$, where M_m^ls are the M-spline basis functions of order l, studied by Curry and Schoenberg (1966), and can be calculated recursively by

$$M_i^1(s) = \frac{1}{u_{i+1} - u_i}, \qquad u_i \leq s < u_{i+1},$$

$$M_i^l(s) = \frac{l[(s - u_i)M_i^{l-1}(s) + (u_{i+l} - s)M_{i+1}^{l-1}(s)]}{(l-1)(u_{i+l} - u_i)}.$$

By the relationship between the B-spline basis functions and the M-spline basis functions [Schumaker (1981)], it can be easily demonstrated that the I-spline basis function defined in (4.1) can be expressed by a sum of the B-spline basis functions

(4.2)
$$I_i^{l-1}(s) = \sum_{m=i}^{p_n} N_m^l(s).$$

Consequently, the spline-based sieve space can be reconstructed using the I-spline basis functions with a different set of constraints:

$$\Theta_n = \left\{ \tau_n = (F_n, F_{n,1}, F_{n,2}) : F_n(s, t) = \sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} I_i^{(1),l-1}(s) I_j^{(2),l-1}(t), \right.$$

$$F_{n,1}(s) = \sum_{i=1}^{p_n} \left\{ \sum_{j=1}^{q_n} \eta_{i,j} + \omega_i \right\} I_i^{(1),l-1}(s),$$

$$\left. F_{n,2}(t) = \sum_{j=1}^{q_n} \left\{ \sum_{i=1}^{p_n} \eta_{i,j} + \pi_j \right\} I_j^{(2),l-1}(t) \right\}$$

with $\underline{\eta} = (\eta_{1,1}, \ldots, \eta_{p_n,q_n})$, $\underline{\omega} = (\omega_1, \ldots, \omega_{p_n})$ and $\underline{\pi} = (\pi_1, \ldots, \pi_{q_n})$ subject to the following conditions in (4.3),

(4.3)
$$\eta_{i,j} \geq 0 \qquad \text{for } i = 1, \ldots, p_n, j = 1, \ldots, q_n,$$
$$\omega_i \geq 0, \qquad i = 1, \ldots, p_n,$$
$$\pi_j \geq 0, \qquad j = 1, \ldots, q_n,$$
$$\sum_{i=1}^{p_n} \sum_{j=1}^{q_n} \eta_{i,j} + \sum_{i=1}^{p_n} \omega_i + \sum_{j=1}^{q_n} \pi_j \leq 1.$$

Then the spline-based sieve log likelihood can be also expressed in I-spline, and the spline-based sieve MLE can be obtained by maximizing the log likelihood in I-spline over Θ_n.

REMARK 4.1. Class Θ_n is actually equivalent to Ω_n, and hence the I-spline-based sieve MLE is the same as the B-spline-based sieve MLE. It is advocated in numerical implementation due to the simplicity of the constraints in class Θ_n.

Given p_n and q_n, the proposed sieve estimation problem described above is actually a restricted parametric maximum likelihood estimation problem with respect to the coefficients associated with the I-spline and the tensor I-spline basis

functions. Jamshidian (2004) generalized the gradient projection algorithm originally proposed by Rosen (1960) using a weighted L_2-norm $\|x\| = x'Wx$ with a positive definite matrix W for the restricted maximum likelihood estimation problems. Because the constraint set (4.3) is made by linear inequalities, the maximization of (2.2) in the I-spline form over Θ_n can be efficiently implemented by the generalized gradient projection algorithm [Jamshidian (2004)] and is described as follows.

First we rewrite (4.3) as $X\theta \leq y$, where $X = (x_1, x_2, \ldots, x_{p_n \cdot q_n + p_n + q_n},$ $x_{p_n \cdot q_n + p_n + q_n + 1})^T$ with $x_1 = (-1, 0, \ldots, 0)^T$, $x_2 = (0, -1, 0, \ldots, 0)^T$, $x_{p_n \cdot q_n + p_n + q_n} = (0, \ldots, 0, -1)^T$, $x_{p_n \cdot q_n + p_n + q_n + 1} = (1, \ldots, 1)^T$; $\theta = (\underline{\eta}, \underline{\omega}, \underline{\pi}) = (\theta_1, \theta_2, \ldots, \theta_{p_n \cdot q_n + p_n + q_n})$; and $y = (0, \ldots, 0, 1)^T$. If some I-spline coefficients equal 0 or all coefficients sum up to 1, then we say their corresponding constraints are active and let $\bar{X}\theta = \bar{y}$ represent all the active constraints and a vector Λ of integers to index the active constraints. For example, if $\Lambda = (2, 1, p_n \cdot q_n + p_n + q_n + 1)$, then the second, first and last constraints become active, and $\bar{X} = (x_2, x_1, x_{p_n \cdot q_n + p_n + q_n + 1})^T$ and $\bar{y} = (0, 0, 1)^T$.

Let $\dot{\tilde{l}}(\theta)$ and $H(\theta)$ be the gradient and Hessian matrix of the log likelihood given by (2.2) in the I-spline form, respectively. Note that $H(\theta)$ may not be negative definite for every θ. We use $W = -H(\theta) + \delta I$, where I is identity matrix, and $\delta > 0$ is chosen as any value that guarantees W being positive definite. With that introduced, the generalized gradient projection algorithm is implemented as follows.

Step 1 (*Computing the feasible search direction*). Compute

$$\underline{d} = (d_1, d_2, \ldots, d_{p_n \cdot q_n + p_n + q_n})$$
$$= \{I - W^{-1}\bar{X}^T(\bar{X}W^{-1}\bar{X}^T)^{-1}\bar{X}\}W^{-1}\dot{\tilde{l}}(\theta).$$

Step 2 (*Forcing the updated θ to fulfill the constraints*). Compute

$$\gamma = \begin{cases} \min\left\{\min\limits_{i:d_i<0}\left\{-\dfrac{\theta_i}{d_i}\right\}, \dfrac{1 - \sum_{i=1}^{p_n \cdot q_n + p_n + q_n}\theta_i}{\sum_{i=1}^{p_n \cdot q_n + p_n + q_n}d_i}\right\}, \\ \qquad\qquad\qquad \text{if} \quad \sum\limits_{i=1}^{p_n \cdot q_n + p_n + q_n}d_i > 0, \\ \min\limits_{i:d_i<0}\left\{-\dfrac{\theta_i}{d_i}\right\}, \qquad \text{else.} \end{cases}$$

Doing so guarantees that $\theta_i + \gamma d_i \geq 0$ for $i = 1, 2, \ldots, p_n \cdot q_n + p_n + q_n$, and $\sum_{i=1}^{p_n \cdot q_n + p_n + q_n}(\theta_i + \gamma d_i) \leq 1$.

Step 3 (*Updating the solution by step-halving line search*). Find the smallest integer k starting from 0 such that

$$\tilde{l}_n(\theta + (1/2)^k\gamma\underline{d}; \cdot) \geq \tilde{l}_n(\theta; \cdot).$$

Replace θ by $\tilde{\theta} = \theta + \min\{(1/2)^k\gamma, 0.5\}\underline{d}$.

Step 4 (*Updating* Λ *and* $\bar{X}$). Modify Λ by adding indexes of new I-spline coefficients when these new coefficients become 0 and adding $p_n \cdot q_n + p_n + q_n + 1$ when the sum of all I-spline coefficients becomes 1. Modify $\bar{X}$ accordingly.

Step 5 (*Checking the stopping criterion*). If $\|\underline{d}\| \geq \varepsilon$, for small ε, go to Step 1; otherwise, compute $\lambda = (\bar{X} W^{-1} \bar{X}^T)^{-1} \bar{X} W^{-1} \dot{\tilde{l}}(\theta)$.

  (i) If the jth component $\lambda_j \geq 0$ for all j, set $\hat{\theta} = \theta$ and stop.

  (ii) If there is at least one j such that $\lambda_j < 0$, let $j^* = \arg\min_{j:\lambda_j < 0}\{\lambda_j\}$, then remove j^*th component from Λ and remove the j^*th row from $\bar{X}$, and go to Step 1.

5. Simulation studies. Copula models are often used in studying bivariate event time data [Shih and Louis (1995), Wang and Ding (2000), Sun, Wang and Sun (2006), Zhang et al. (2010)]

We consider the bivariate Clayton copula function

$$C_\alpha(u, v) = \left(u^{(1-\alpha)} + v^{(1-\alpha)} - 1\right)^{1/(1-\alpha)},$$

with $\alpha > 1$. For the Clayton copula, a larger α corresponds to a stronger positive association between the two random variables. The association parameter α and Kendall's τ for the Clayton copula is related by $\tau = \frac{\alpha-1}{\alpha+1}$.

In the simulation studies, we compare the proposed sieve MLE to the conventional NPMLE, computed using the algorithm developed by Maathuis (2005). As we mentioned previously, this NPMLE is not unique. Only the total mass in each selected rectangle is estimated, therefore the estimated joint CDF is based on where the mass is placed in each rectangle. We denote U-NPMLE and L-NPMLE as the NPMLE for which the probability mass is placed at the upper right and lower left corners of each rectangle, respectively.

The proposed sieve MLE and both U-NPMLE and L-NPMLE are evaluated with various combinations of Kendall's τ ($\tau = 0.25, 0.75$) and sample sizes ($n = 100, 200$). Under each of the four settings, the Monte-Carlo simulation with 500 repetitions is conducted, and the cubic ($l = 4$) I-spline basis functions are used in the proposed sieve estimation method. The event times (T_1, T_2), monitoring times (C_1, C_2) and the knots selection of the cubic I-spline basis functions are illustrated as follows:

  (i) (*Event times*). (T_1, T_2) are generated from the Clayton copula with the two marginal distributions being exponential with the rate parameter 0.5. Under this setting, $\Pr(T_i \geq 5) < 0.1$ for $i = 1, 2$ and $[L_1, U_1] \times [L_2, U_2]$ is chosen to be $[0, 5] \times [0, 5]$.

  (ii) (*Censoring times*). Both C_1 and C_2 are generated independently from the uniform distribution on $[0.0201, 4.7698]$ [$\Pr(0 < T_i < 0.0201) = \Pr(4.7698 < T_i < 5) = 0.01$, for $i = 1, 2$]. The observation region $[l_1, u_1] \times [l_2, u_2] = [0.0201, 4.7698] \times [0.0201, 4.7698]$ is inside $[0, 5] \times [0, 5]$ and the CDFs are bounded away from 0 and 1 inside the observation region.

(iii) (*Knots selection*). As in other spline-based estimations [Lu, Zhang and Huang (2007, 2009), Zhang, Hua and Huang (2010) and Wu and Gao (2011)], the number of interior knots m_n is chosen as $[n^{1/3}] - 1$, where $[n^{1/3}]$ is the integer part of $n^{1/3}$. For moderate sample sizes, say $n = 100, 200$, our experiments show that $m_n = [n^{1/3}] - 1$ is a reasonable choice for the number of interior knots. Therefore, we choose 4 and 5 as the numbers of interior knots for sample sizes 100 and 200, respectively. The number of spline basis functions is determined by $p_n = q_n = m_n + 4$ in our computation. Two end knots of all knot sequences are chosen to be 0 and 5. For each sample of bivariate observation times (C_1, C_2), the interior knots for $\{I_i^{(1),3}\}_{i=1}^{p_n}$ and $\{I_j^{(2),3}\}_{j=1}^{q_n}$ are allocated at the $k/(m_n + 1)$ quantiles $(k = 1, \ldots, m_n)$ of the samples of C_1 and C_2, respectively.

Table 1 displays the estimation biases (*Bias*) and the square roots of mean square errors ($\text{MSE}^{1/2}$) from the Monte-Carlo simulation of 500 repetitions for the pro-

TABLE 1
Comparison of the estimation bias and square root of mean square error among the sieve MLE,
U-NPMLE and L-NPMLE at four selected points

T_1		T_2					
		0.1			**4.6**		
		Sieve	**U-Non**	**L-Non**	**Sieve**	**U-Non**	**L-Non**
Sample size $n = 100$, Kendall's $\tau = 0.25$							
0.1	Bias	−5.00e–3	−1.78e–2	−1.91e–2	2.69e–2	−2.22e–3	−3.93e–2
	$\text{MSE}^{1/2}$	2.75e–2	2.24e–2	1.91e–2	7.32e–2	6.68e–2	5.53e–2
4.6	Bias	2.33e–2	−2.69e–2	−4.19e–2	4.04e–2	1.32e–1	1.09e–1
	$\text{MSE}^{1/2}$	7.18e–2	6.68e–2	5.17e–2	8.24e–2	1.49e–1	1.35e–1
Sample size $n = 200$, Kendall's $\tau = 0.25$							
0.1	Bias	−4.39e–3	−1.85e–2	−1.91e–2	2.26e–2	−2.87e–2	−3.89e–2
	$\text{MSE}^{1/2}$	2.42e–2	1.98e–2	1.91e–2	6.05e–2	5.52e–2	5.04e–2
4.6	Bias	1.65e–2	−3.29e–2	−4.03e–2	2.15e–2	1.10e–1	9.65e–2
	$\text{MSE}^{1/2}$	5.31e–2	5.50e–2	5.30e–2	6.10e–2	1.29e–1	1.21e–1
Sample size $n = 100$, Kendall's $\tau = 0.75$							
0.1	Bias	−1.81e–2	−3.62e–2	−4.33e–2	2.91e–2	−1.95e–2	−4.11e–2
	$\text{MSE}^{1/2}$	4.63e–2	5.36e–2	4.34e–2	7.88e–2	8.19e–2	5.62e–2
4.6	Bias	3.08e–2	−1.90e–2	−4.04e–2	1.98e–2	1.03e–1	8.09e–2
	$\text{MSE}^{1/2}$	8.16e–2	8.45e–2	5.83e–2	6.47e–2	1.22e–1	1.08e–1
Sample size $n = 200$, Kendall's $\tau = 0.75$							
0.1	Bias	−2.03e–2	−4.00e–2	−4.31e–2	2.01e–2	−2.48e–2	−3.81e–2
	$\text{MSE}^{1/2}$	3.86e–2	4.52e–2	4.37e–2	5.87e–2	5.90e–2	5.27e–2
4.6	Bias	2.09e–2	−2.48e–2	−3.93e–2	1.08e–2	8.37e–2	7.20e–2
	$\text{MSE}^{1/2}$	6.00e–2	6.27e–2	5.35e–2	5.24e–2	1.04e–1	9.40e–2

TABLE 2

*Comparison of the overall estimation biases and the overall mean square errors among sieve MLE,
U-NPMLE and L-NPMLE*

		Sample size					
		100			**200**		
τ		**Sieve**	**U-Non**	**L-Non**	**Sieve**	**U-Non**	**L-Non**
0.25	|Bias|	7.56e–3	3.24e–2	4.24e–2	6.70e–3	2.62e–2	3.11e–2
	MSE$^{1/2}$	7.93e–2	1.25e–1	1.26e–1	6.13e–2	1.03e–1	1.03e–1
0.75	|Bias|	1.11e–2	1.50e–2	2.49e–2	7.29e–3	1.33e–2	1.88e–2
	MSE$^{1/2}$	7.40e–2	1.10e–1	1.11e–1	5.77e–2	8.45e–2	8.53e–2

posed sieve MLE (*Sieve*) and both U-NPMLE (*U-Non*) and L-NPMLE (*L-Non*)
of the bivariate CDF at 4 selected pairs of time points (s_1, s_2) near the corners of
the estimation region with different sample sizes and Kendall's τ values. The esti-
mation results at those selected points are comparable among the three estimators.
Table 2 presents the overall estimation bias and mean square error for the three
estimators by calculating the average of absolute values of estimation bias and the
average of square roots of mean square error taking from 2209 pairs of (s_1, s_2)
with both s_1 and s_2 ranging uniformly from 0.1 to 4.7. It appears that the sieve
MLE outperforms its counterparts with a smaller overall bias and a smaller overall
mean square error. The mean square error of the proposed sieve MLE noticeably
decreases as sample size increases from 100 to 200.

For sample size $n = 200$, the estimation biases and the square roots of mean
square error of the sieve MLE and U-NPMLE for the joint CDF from the same
Monte-Carlo simulation are graphed in Figure 1 through Figure 4 for Kendall's
$\tau = 0.25$ and 0.75. These figures clearly indicate that the bias and the MSE of
the sieve MLE are noticeably smaller than that of U-NPMLE inside the closed
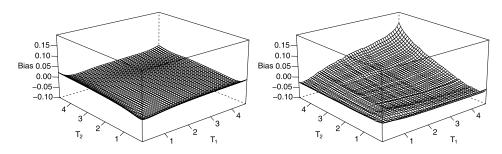region $[0.1, 4.7] \times [0.1, 4.7]$. It is also seen that the bias of the sieve MLE near

FIG. 1.   *Comparison of the estimation bias between the sieve MLE* (*left*) *and the U-NPMLE* (*right*)
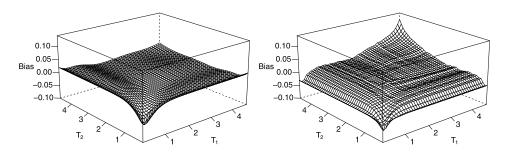*for the joint CDF when sample size $n = 200$, Kendall's $\tau = 0.25$.*

FIG. 2.   *Comparison of the estimation bias between the sieve MLE (*left*) and the U-NPMLE (*right*) for the joint CDF when sample size n = 200, Kendall's τ = 0.75.*

the origin increases as Kendall's τ increases. As a by-product of the estimation methods, the average estimate of the marginal CDF of T_1 from the same Monte-Carlo simulation for both the proposed sieve MLE (*Sieve*) and U-NPMLE (*U-Non*) are also computed and plotted in Figure 5 along with the true marginal CDF (*True*), F_1. Figure 5 clearly indicates that the bias of the proposed sieve MLE for the marginal CDF is markedly smaller than that of the U-NPMLE, particularly near the two end points of interval $[0.1, 4.7]$.
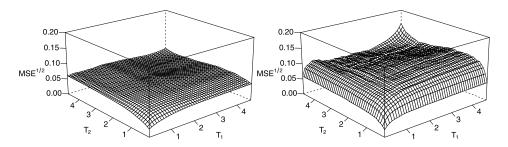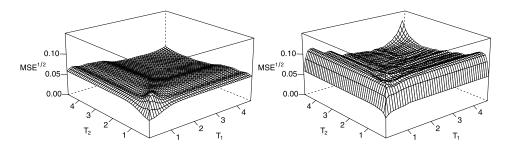

FIG. 3.   *Comparison of the square root of mean square error between the sieve MLE (*left*) and the U-NPMLE (*right*) for the joint CDF when sample size n = 200, Kendall's τ = 0.25.*

FIG. 4.   *Comparison of the square root of mean square error between the sieve MLE (*left*) and the U-NPMLE (*right*) for the joint CDF when sample size n = 200, Kendall's τ = 0.75.*
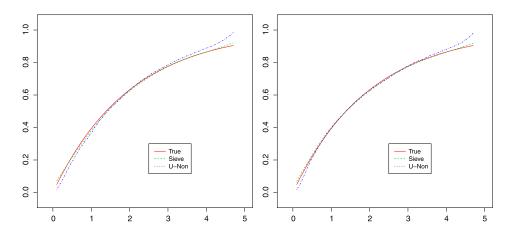
FIG. 5. *Comparison of the estimation bias between the sieve MLE and the U-NPMLE for estimating the marginal CDF of T_1 when sample size $n = 200$ (left: Kendall's $\tau = 0.25$; right: Kendall's $\tau = 0.75$).*

6. Final remarks. The estimation of the joint CDF with bivariate event time data is a challenging problem in survival analysis. Development of sophisticated methods for this type of problems is much needed for applications. In this paper, we develop a tensor spline-based sieve maximum likelihood estimation method for estimating the joint CDF with bivariate current status data. This sieve estimation approach reduces the dimension of unknown parameter space and estimates both the joint and marginal CDFs simultaneously. As a result, the proposed method enjoys two advantages in studying bivariate event time data: (i) it provides a unique estimate for the joint CDF, and the numerical implementation is less demanding due to dimension reduction; (ii) the estimation procedure automatically takes into account the possible correlation between the two event times by satisfying the constraints, which intuitively results in more efficient estimation for the marginal CDFs compared to the existing methods for estimating the marginal CDFs using only the univariate current status data.

Under mild regularity conditions, we also show that the proposed spline-based sieve estimator is consistent and could converge to the true joint CDF at a rate faster than $n^{1/3}$ if the target CDF is smooth enough. Both theoretical and numerical results provide evidence that the proposed sieve MLE outperforms the conventional NPMLE studied in the literature. The superior performance of the proposed method mainly rests on the smoothness of the true bivariate distribution function. In many applications of bivariate survival analysis, this assumption of smoothness is reasonable and shall motivate the use of the proposed method.

Though the development of the proposed method is illustrated with bivariate current status data as it algebraically simplifies the theoretical justification, the proposed method can be readily extended to bivariate interval censored data [Song (2001) and Maathuis (2005)] as well as bivariate right censored data [Dabrowska

(1988) and Kooperberg (1998)] with parallel theoretical and numerical justifications. It is potentially applicable in any nonparametric estimation problem of multivariate distribution function.

While the consistency and rate of convergence are fully studied for the proposed estimator, the study of its asymptotic distribution is not accomplished. With the knowledge of asymptotic distribution of the conventional NPMLE for current status data studied in Groeneboom and Wellner (1992) and Song (2001), it is for sure that the asymptotic distribution of the proposed estimator will not be Gaussian. Discovering the limiting distribution for the proposed estimator remains an interesting yet a very challenging problem for future investigation.

7. Proofs of the theorems. For the rest of this paper, we denote K as a universal positive constant that may be different from place to place and $\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$, the empirical process indexed by $f(X)$.

PROOF OF THEOREM 3.1. We show $\hat{\tau}_n$ is a consistent estimator by verifying the three conditions of Theorem 5.7 in van der Vaart (1998).

For $(s, t) \in [l_1, u_1] \times [l_2, u_2]$, we define Ω by

$$\Omega = \big\{ \tau(s, t) = \big(F(s, t), F_1(s), F_2(t) \big) :$$

$$\tau \text{ satisfies the following conditions (a) and (b)} \big\}:$$

(a) $F(s, t)$ is nondecreasing in both s and t, $F_1(s) - F(s, t)$ is nondecreasing in s but nonincreasing in t, $F_2(t) - F(s, t)$ is nondecreasing in t but nonincreasing in s, and $1 - F_1(s) - F_2(t) + F(s, t)$ is nonincreasing in both s and t,

(b) $F(s, t) \geq b_1$, $F_1(s) - F(s, t) \geq b_2$, $F_2(t) - F(s, t) \geq b_3$ and $1 - F_1(s) - F_2(t) + F(s, t) \geq b_4$, for $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and $b_4 > 0$.

Since (C2) and (C6) hold, Lemma 0.1 in the supplemental article [Wu and Zhang (2012)] implies that there exist $b_1 > 0$, $b_2 > 0$, $b_3 > 0$ and $b_4 > 0$ small enough to guarantee that $\tau_0 \in \Omega$ and $\Omega'_n \in \Omega$. We suppose b_1, b_2, b_3 and b_4, in condition (b) above, are chosen small enough such that Ω contains both τ_0 and Ω'_n.

Denote $\mathcal{L} = \{l(\tau) : \tau \in \Omega\}$ the class of functions induced by the log likelihood with a single observation $x = (s, t, \delta_1, \delta_2)$, where

$$l(\tau) = \delta_1 \delta_2 \log F(s, t) + \delta_1 (1 - \delta_2) \log[F_1(s) - F(s, t)]$$

$$+ (1 - \delta_1) \delta_2 \log[F_2(t) - F(s, t)]$$

$$+ (1 - \delta_1)(1 - \delta_2) \log[1 - F_1(s) - F_2(t) + F(s, t)],$$

with $\delta_1 = 1_{[T_1 \leq s]}$, $\delta_2 = 1_{[T_2 \leq t]}$. We denote $\mathbb{M}(\tau) = Pl(\tau)$ and $\mathbb{M}_n(\tau) = \mathbb{P}_n(l(\tau))$.

First, we verify $\sup_{\tau \in \Omega} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau)| \to_p 0$.

It suffices to show that $\mathcal{L}$ is a P-Glivenko–Cantelli, since

$$\sup_{\tau \in \Omega} |\mathbb{M}_n(\tau) - \mathbb{M}(\tau)| = \sup_{l(\tau) \in \mathcal{L}} |(\mathbb{P}_n - P)l(\tau)| \to_p 0.$$

Let $A_1 = \{\frac{\log F(s,t)}{\log b_1} : \tau = (F, F_1, F_2) \in \Omega\}$, and $\mathcal{G}_1 = \{1_{[l_1,s] \times [l_2,t]}, l_1 \leq s \leq u_1, l_2 \leq t \leq u_2\}$. By conditions (a) and (b), we know $0 \leq \frac{\log F(s,t)}{\log b_1} \leq 1$ and $\frac{\log F(s,t)}{\log b_1}$ is nonincreasing in both s and t. Therefore $A_1 \subseteq \overline{\text{sconv}}(\mathcal{G}_1)$, the closure of the symmetric convex hull of $\mathcal{G}_1$ [van der Vaart and Wellner (1996)]. Hence Theorem 2.6.7 in van der Vaart and Wellner (1996) implies that

$$(7.1) \qquad N(\varepsilon, \mathcal{G}_1, L_2(Q_{C_1,C_2})) \leq K\left(\frac{1}{\varepsilon}\right)^4$$

for any probability measure Q_{C_1,C_2} of (C_1, C_2). By the facts that $V(\mathcal{G}_1) = 3$ and the envelop function of $\mathcal{G}_1$ is 1. (7.1) is followed by

$$\log N(\varepsilon, \overline{\text{sconv}}(\mathcal{G}_1), L_2(Q_{C_1,C_2})) \leq K\left(\frac{1}{\varepsilon}\right)^{4/3},$$

using the result of Theorem 2.6.9 in van der Vaart and Wellner (1996). Hence

$$(7.2) \qquad \log N(\varepsilon, A_1, L_2(Q_{C_1,C_2})) \leq K\left(\frac{1}{\varepsilon}\right)^{4/3}.$$

Let

$$A_1' = \{\delta_1\delta_2 \log F(s,t) : \tau = (F, F_1, F_2) \in \Omega\}.$$

Suppose the centers of ε-balls of A_1 are $f_i, i = 1, 2, \ldots, [K(\frac{1}{\varepsilon})^{4/3}]$, and then for any joint probability measure Q of (T_1, T_2, C_1, C_2),

$$\|\delta_1\delta_2 \log F - \delta_1\delta_2 \log b_1 f_i\|_{L_2(Q)}^2$$

$$= Q\left[\delta_1\delta_2 \log b_1\left(\frac{\log F}{\log b_1} - f_i\right)\right]^2$$

$$= E\left[1_{[T_1 < C_1, T_2 < C_2]} \log b_1\left(\frac{\log F(C_1, C_2)}{\log b_1} - f_i(C_1, C_2)\right)\right]^2$$

$$= E\left\{E\left\{\left[1_{[T_1 < C_1, T_2 < C_2]} \log b_1\left(\frac{\log F(C_1, C_2)}{\log b_1} - f_i(C_1, C_2)\right)\right]^2 \Big| C_1, C_2\right\}\right\}$$

$$= E_{C_1,C_2}\left[F_0(C_1, C_2) \log b_1\left(\frac{\log F(C_1, C_2)}{\log b_1} - f_i(C_1, C_2)\right)\right]^2$$

$$\leq E_{C_1,C_2}\left[\log b_1\left(\frac{\log F(C_1, C_2)}{\log b_1} - f_i(C_1, C_2)\right)\right]^2$$

$$= (\log b_1)^2\left\|\frac{\log F}{\log b_1} - f_i\right\|_{L_2(Q_{C_1,C_2})}^2.$$

Let $\hat{b}_1 = -\log b_1$ then $\delta_1 \delta_2 \log b_1 f_i$, $i = 1, 2, \ldots, [K(\frac{1}{\varepsilon})^{4/3}]$, are the centers of $\varepsilon\hat{b}_1$-balls of A_1'. Hence by (7.2) we have $\log N(\varepsilon\hat{b}_1, A_1', L_2(Q)) \le K(\frac{1}{\varepsilon})^{4/3}$, and it follows that

$$\int_0^1 \sup_Q \sqrt{\log N(\varepsilon\hat{b}_1, A_1', L_2(Q))}\, d\varepsilon \le \int_0^1 \sqrt{K}\left(\frac{1}{\varepsilon}\right)^{2/3} d\varepsilon < \infty.$$

It is obvious that the envelop function of A_1' is $\hat{b}_1$, therefore A_1' is a P-Donsker by Theorem 2.5.2 in van der Vaart and Wellner (1996).

Let

$$A_2' = \{\delta_1(1 - \delta_2)\log(F_1(s) - F(s, t)) : \tau = (F, F_1, F_2) \in \Omega\},$$
$$A_3' = \{(1 - \delta_1)\delta_2 \log(F_2(t) - F(s, t)) : \tau = (F, F_1, F_2) \in \Omega\}$$

and

$$A_4' = \{(1 - \delta_1)(1 - \delta_2)\log(1 - F_1(s) - F_2(t) - F(s, t)) :$$
$$\tau = (F, F_1, F_2) \in \Omega\}.$$

Following the same arguments for showing A_1' being a P-Donsker, it can be shown that A_2', A_3' and A_4' are all P-Donsker classes. So $\mathcal{L}$ is P-Donsker as well. Since P-Donsker is also P-Glivenko–Cantelli, it then follows that $\sup_{l(\tau)\in\mathcal{L}} |(\mathbb{P}_n - P)l(\tau)| \to_p 0$.

Second, we verify $\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \ge K d^2(\tau_0, \tau)$, for any $\tau \in \Omega$.

Note that

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau)$$
$$= P\{l(\tau_0) - l(\tau)\}$$
$$= P\left\{\delta_1\delta_2 \log \frac{F_0}{F} + \delta_1(1 - \delta_2)\log \frac{F_{0,1} - F_0}{F_1 - F}\right.$$
$$+ (1 - \delta_1)\delta_2 \log \frac{F_{0,2} - F_0}{F_2 - F}$$
$$\left. + (1 - \delta_1)(1 - \delta_2)\log \frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right\}$$
$$= P_{C_1, C_2}\left\{F_0 \log \frac{F_0}{F} + (F_{0,1} - F_0)\log \frac{F_{0,1} - F_0}{F_1 - F}\right.$$
$$+ (F_{0,2} - F_0)\log \frac{F_{0,2} - F_0}{F_2 - F}$$
$$\left. + (1 - F_{0,1} - F_{0,2} + F_0)\log \frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right\},$$

and it follows that

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) = P_{C_1,C_2}\Bigg\{ Fm\left(\frac{F_0}{F}\right) + (F_1 - F)m\left(\frac{F_{0,1} - F_0}{F_1 - F}\right)$$

(7.3)
$$+ (F_2 - F)m\left(\frac{F_{0,2} - F_0}{F_2 - F}\right)$$

$$+ (1 - F_1 - F_2 + F)m\left(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right)\Bigg\},$$

where $m(x) = x\log(x) - x + 1 \geq (x-1)^2/4$ for $0 \leq x \leq 5$.

Since F has positive upper bound,

(7.4)
$$P_{C_1,C_2}\left\{ Fm\left(\frac{F_0}{F}\right)\right\} \geq P_{C_1,C_2}\left\{ F\left(\frac{F_0}{F} - 1\right)^2\Big/4\right\} \geq K P_{C_1,C_2}(F_0 - F)^2$$

$$= K\|F_0 - F\|^2_{L_2(P_{C_1,C_2})}.$$

Similarly, we can easily show that

(7.5)
$$P_{C_1,C_2}\left\{ (F_1 - F)m\left(\frac{F_{0,1} - F_0}{F_1 - F}\right)\right\}$$

$$\geq K\|(F_{0,1} - F_1) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})},$$

(7.6)
$$P_{C_1,C_2}\left\{ (F_2 - F)m\left(\frac{F_{0,2} - F_0}{F_2 - F}\right)\right\}$$

$$\geq K\|(F_{0,2} - F_2) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})}$$

and

(7.7)
$$P_{C_1,C_2}\left\{ (1 - F_1 - F_2 + F)m\left(\frac{1 - F_{0,1} - F_{0,2} + F_0}{1 - F_1 - F_2 + F}\right)\right\}$$

$$\geq K\|(1 - F_{0,1} - F_{0,2} + F_0) - (1 - F_1 - F_2 + F)\|^2_{L_2(P_{C_1,C_2})}.$$

So combining (7.4), (7.5), (7.6) and (7.7) results in

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K\big(\|F_0 - F\|^2_{L_2(P_{C_1,C_2})}$$

$$+ \|(F_{0,1} - F_1) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})}$$

$$+ \|(F_{0,2} - F_2) - (F_0 - F)\|^2_{L_2(P_{C_1,C_2})}\big).$$

Let $f_1 = \|F_0 - F\|^2_{L_2(P_{C_1,C_2})}$, $f_2 = \|F_{0,1} - F_1\|^2_{L_2(P_{C_1})}$ and $f_3 = \|F_{0,2} - F_2\|^2_{L_2(P_{C_2})}$. If f_1 is the largest among f_1, f_2, f_3, then

(7.8)
$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq Kf_1 \geq (K/3)(f_1 + f_2 + f_3).$$

If f_2 is the largest, then

$$(7.9) \quad \mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K\big[f_1 + (f_2 - f_1)\big] \geq Kf_2 \geq (K/3)(f_1 + f_2 + f_3).$$

If f_3 is the largest, then

$$(7.10) \quad \mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq K\big[f_1 + (f_3 - f_1)\big] \geq Kf_3 \geq (K/3)(f_1 + f_2 + f_3).$$

Therefore, by (7.8), (7.9) and (7.10), it follows that

$$\mathbb{M}(\tau_0) - \mathbb{M}(\tau) \geq Kd^2(\tau_0, \tau).$$

Finally, we verify $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \geq -o_p(1)$.

Since (C2), (C3) and (C6) hold, Lemma 0.3 in the supplemental article [Wu and Zhang (2012)] implies that there exists $\tau_n = (F_n, F_{n,1}, F_{n,2})$ in Ω'_n such that for $\tau_0 = (F_0, F_{0,1}, F_{0,2})$, $\|F_n - F_0\|_\infty \leq K(n^{-pv})$, $\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv})$ and $\|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv})$. Since $\hat{\tau}_n$ maximizes $\mathbb{M}_n(\tau)$ in Ω'_n, $\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) > 0$. Hence,

$$
\begin{aligned}
\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) &= \mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_n) + \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) \\
(7.11) \qquad &\geq \mathbb{M}_n(\tau_n) - \mathbb{M}_n(\tau_0) = \mathbb{P}_n\big(l(\tau_n)\big) - \mathbb{P}_n\big(l(\tau_0)\big) \\
&= (\mathbb{P}_n - P)\big\{l(\tau_n) - l(\tau_0)\big\} + P\big\{l(\tau_n) - l(\tau_0)\big\}.
\end{aligned}
$$

Define

$$
\begin{aligned}
\mathcal{L}_n = \big\{l(\tau_n) : \tau_n = (F_n, F_{n,1}, F_{n,2}) \in \Omega'_n, \|F_n - F_0\|_\infty \leq K(n^{-pv}), \\
\|F_{n,1} - F_{0,1}\|_\infty \leq K(n^{-pv}), \|F_{n,2} - F_{0,2}\|_\infty \leq K(n^{-pv})\big\}.
\end{aligned}
$$

Since $(a + b + c + d)^2 \leq 4(a^2 + b^2 + c^2 + d^2)$, then for any $l(\tau_n) \in \mathcal{L}_n$, we have

$$
\begin{aligned}
&P\big\{l(\tau_n) - l(\tau_0)\big\}^2 \\
&\leq 4P\left(\delta_1\delta_2 \log \frac{F_n}{F_0}\right)^2 \\
&\quad + 4P\left(\delta_1(1 - \delta_2) \log \frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right)^2 + 4P\left((1 - \delta_1)\delta_2 \log \frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right)^2 \\
(7.12) \qquad &\quad + 4P\left((1 - \delta_1)(1 - \delta_2) \log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right)^2 \\
&\leq 4P_{C_1,C_2}\left(\log \frac{F_n}{F_0}\right)^2 + 4P_{C_1,C_2}\left(\log \frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right)^2 \\
&\quad + 4P_{C_1,C_2}\left(\log \frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right)^2 + 4P_{C_1,C_2}\left(\log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right)^2.
\end{aligned}
$$

The facts that $\|F_n - F_0\|_\infty \leq K(n^{-pv})$ and that F_0 has a positive lower bound result in $1/2 < \frac{F_n}{F_0} < 2$ for large n. It can be easily shown that if $1/2 \leq x \leq 2$, $|\log(x)| \leq K|x - 1|$. Hence $|\log \frac{F_n}{F_0}| \leq K|\frac{F_n}{F_0} - 1|$, and it follows that

$$(7.13) \qquad P_{C_1,C_2}\left|\log \frac{F_n}{F_0}\right|^2 \leq K P_{C_1,C_2}\left|\frac{F_n}{F_0} - 1\right|^2 \leq K P_{C_1,C_2}|F_n - F_0|^2 \to 0.$$

Similar arguments yield

$$(7.14) \quad P_{C_1,C_2}\left|\log \frac{F_{n,1} - F_n}{F_{0,1} - F_0}\right|^2 \leq K P_{C_1,C_2}\left|(F_{n,1} - F_n) - (F_{0,1} - F_0)\right|^2 \to 0,$$

$$(7.15) \quad P_{C_1,C_2}\left|\log \frac{F_{n,2} - F_n}{F_{0,2} - F_0}\right|^2 \leq K P_{C_1,C_2}\left|(F_{n,2} - F_n) - (F_{0,2} - F_0)\right|^2 \to 0$$

and

$$(7.16) \qquad\qquad P_{C_1,C_2}\left|\log \frac{1 - F_{n,1} - F_{n,2} + F_n}{1 - F_{0,1} - F_{0,2} + F_0}\right|^2 \to 0.$$

Combining (7.12)–(7.16) results in $P\{l(\tau_n) - l(\tau_0)\}^2 \to 0$, as $n \to \infty$. Hence

$$(7.17) \quad \begin{aligned} \rho_P\{l(\tau_n) - l(\tau_0)\} &= \{\text{var}_P[l(\tau_n) - l(\tau_0)]\}^{1/2} \\ &\leq \{P[l(\tau_n) - l(\tau_0)]^2\}^{1/2} \to_{n\to\infty} 0. \end{aligned}$$

Since $\mathcal{L}$ is shown a P-Donsker in the first part of the proof, Corollary 2.3.12 of van der Vaart and Wellner (1996) yields that

$$(7.18) \qquad\qquad (\mathbb{P}_n - P)\{l(\tau_n) - l(\tau_0)\} = o_p(n^{-1/2}),$$

by the fact that both $l(\tau_n)$ and $l(\tau_0)$ are in $\mathcal{L}$ and (7.17).

In addition,

$$\left|P\{l(\tau_n) - l(\tau_0)\}\right| \leq P|l(\tau_n) - l(\tau_0)| \leq K\{P[l(\tau_n) - l(\tau_0)]^2\}^{1/2} \to_{n\to\infty} 0.$$

Therefore $P(l(\tau_n) - l(\tau_0)) \geq -o(1)$ as $n \to \infty$. Hence,

$$\mathbb{M}_n(\hat{\tau}_n) - \mathbb{M}_n(\tau_0) \geq o_p(n^{-1/2}) - o(1) \geq -o_p(1).$$

This completes the proof of $d(\hat{\tau}_n, \tau_0) \to 0$ in probability. $\quad\square$

PROOF OF THEOREM 3.2. We derive the rate of convergence by verifying the conditions of Theorem 3.4.1 of van der Vaart and Wellner (1996). To apply the theorem to this problem, we denote $M_n(\tau) = \mathbb{M}(\tau) = Pl(\tau)$ and $d_n(\tau_1, \tau_2) =$

$d(\tau_1, \tau_2)$. The maximizer of $\mathbb{M}(\tau)$ is $\tau_0 = (F_0, F_{0,1}, F_{0,2})$.

(i) Let $\tau_n \in \Omega'_n$ with τ_n satisfying $d(\tau_n, \tau_0) \leq K(n^{-pv})$ and $\delta_n = n^{-pv}$. We verify that for large n and any $\delta > \delta_n$,

$$\sup_{\delta/2 < d(\tau, \tau_n) \leq \delta, \tau \in \Omega'_n} (\mathbb{M}(\tau) - \mathbb{M}(\tau_n)) \leq -K\delta^2.$$

Since $d(\tau, \tau_0) \geq d(\tau, \tau_n) - d(\tau_0, \tau_n) \geq \delta/2 - K(n^{-pv})$, then for large n, $d(\tau, \tau_0) \geq K\delta$. In the proof of consistency, we have already established that $\mathbb{M}(\tau) - \mathbb{M}(\tau_0) \leq -Kd^2(\tau, \tau_0) \leq -K\delta^2$. And as shown in the proof of consistency, $\mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) \leq Kd^2(\tau_0, \tau_n) \leq K(n^{-2pv})$. Therefore, for large n, $\mathbb{M}(\tau) - \mathbb{M}(\tau_n) = \mathbb{M}(\tau) - \mathbb{M}(\tau_0) + \mathbb{M}(\tau_0) - \mathbb{M}(\tau_n) \leq -K\delta^2 + K(n^{-2pv}) = -K\delta^2$.

(ii) We shall find a function $\psi(\cdot)$ such that

$$E\left\{ \sup_{\delta/2 < d(\tau, \tau_n) \leq \delta, \tau \in \Omega'_n} \mathbb{G}_n(\tau - \tau_n) \right\} \leq K \frac{\psi(\delta)}{\sqrt{n}}$$

and $\delta \to \psi(\delta)/\delta^\alpha$ is decreasing on δ, for some $\alpha < 2$, and for $r_n \leq \delta_n^{-1}$, it satisfies

$$r_n^2 \psi(1/r_n) \leq K\sqrt{n} \qquad \text{for every } n.$$

Let

$$\mathcal{L}_{n,\delta} = \left\{ l(\tau) - l(\tau_n) : \tau \in \Omega'_n \text{ and } \delta/2 < d(\tau, \tau_n) \leq \delta \right\}.$$

First, we evaluate the bracketing number of $\mathcal{L}_{n,\delta}$.

Let $\mathcal{F}_n = \{F : \tau = (F, F_1, F_2) \in \Omega'_n, \delta/2 \leq d(\tau, \tau_n) \leq \delta\}$, $\mathcal{F}_{n,1} = \{F_1 : \tau = (F, F_1, F_2) \in \Omega'_n, \delta/2 \leq d(\tau, \tau_n) \leq \delta\}$ and $\mathcal{F}_{n,2} = \{F_2 : \tau = (F, F_1, F_2) \in \Omega'_n, \delta/2 \leq d(\tau, \tau_n) \leq \delta\}$.

Denote $\tau_n = (F_n, F_{n,1}, F_{n,2})$. Lemma 0.5 in the supplemental article [Wu and Zhang (2012)] implies that there exist ε-brackets $[D_i^L, D_i^U], i = 1, 2, \ldots,$ $[(\delta/\varepsilon)^{Kp_n q_n}]$ to cover $\mathcal{F}_n - F_n$. Moreover, Lemma 0.6 in the supplemental article [Wu and Zhang (2012)] implies there exist ε-brackets $[D_j^{(1),L}, D_j^{(1),U}], j = 1, 2, \ldots, [(\delta/\varepsilon)^{Kp_n}]$, to cover $\mathcal{F}_{n,1} - F_{n,1}$, and there exist ε-brackets $[D_k^{(2),L}, D_k^{(2),U}], k = 1, 2, \ldots, [(\delta/\varepsilon)^{Kq_n}]$, to cover $\mathcal{F}_{n,2} - F_{n,2}$.

Denote $F_i^L \equiv D_i^L + F_n$, $F_i^U \equiv D_i^U + F_n$, $F_j^{(1),L} \equiv D_j^{(1),L} + F_{n,1}$, $F_j^{(1),U} \equiv D_j^{(1),U} + F_{n,1}$, $F_k^{(2),L} \equiv D_k^{(2),L} + F_{n,2}$ and $F_k^{(2),U} \equiv D_k^{(2),U} + F_{n,2}$. Let

$$
\begin{aligned}
l_{i,j,k}^U = {} & \delta_1 \delta_2 \log F_i^U + \delta_1(1 - \delta_2) \log\big(F_j^{(1),U} - F_i^L\big) \\
& + (1 - \delta_1)\delta_2 \log\big(F_k^{(2),U} - F_i^L\big) \\
& + (1 - \delta_1)(1 - \delta_2) \log\big(1 - F_j^{(1),L} - F_k^{(2),L} + F_i^U\big)
\end{aligned}
$$

and

$$l_{i,j,k}^L = \delta_1 \delta_2 \log F_i^L + \delta_1 (1 - \delta_2) \log\big(F_j^{(1),L} - F_i^U\big)$$
$$+ (1 - \delta_1)\delta_2 \log\big(F_k^{(2),L} - F_i^U\big)$$
$$+ (1 - \delta_1)(1 - \delta_2)\log\big(1 - F_j^{(1),U} - F_k^{(2),U} + F_i^L\big).$$

Then for any $l(\tau) \in \{\mathcal{L}_{n,\delta} + l(\tau_n)\}$, there exist i, j, k, for $i = 1, 2, \ldots, [(\delta/\varepsilon)^{K p_n q_n}]$, $j = 1, 2, \ldots, [(\delta/\varepsilon)^{K p_n}]$ and $k = 1, 2, \ldots, [(\delta/\varepsilon)^{K q_n}]$, such that $l_{i,j,k}^L \le l(\tau) \le l_{i,j,k}^U$ and the number of brackets $[l_{i,j,k}^L, l_{i,j,k}^U]'s$ is bounded by $(\delta/\varepsilon)^{K p_n q_n} \cdot (\delta/\varepsilon)^{K p_n} \cdot (\delta/\varepsilon)^{K q_n}$.

Note that

$$\|l_{i,j,k}^U - l_{i,j,k}^L\|_\infty$$

$$\le \left\| \log \frac{F_i^U}{F_i^L} \right\|_\infty + \left\| \log \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} \right\|_\infty$$

$$+ \left\| \log \frac{F_k^{(2),U} - F_i^L}{F_k^{(2),L} - F_i^U} \right\|_\infty + \left\| \log \frac{1 - F_j^{(1),L} - F_j^{(2),L} + F_i^U}{1 - F_j^{(1),U} - F_j^{(2),U} + F_i^L} \right\|_\infty.$$

Since for any $\tau \in \Omega_n'$, F has a positive lower bound, then for a small ε, F_i^L can be made to have a positive lower bound as well. Combining with the fact that $F_i^U(s, t)$ is close to $F_i^L(s, t)$ guarantees that $0 \le \frac{F_i^U}{F_i^L} - 1 \le 1$ for $i = 1, 2, \ldots, [(\delta/\varepsilon)^{K p_n q_n}]$. Note that by $\log x \le (x - 1)$ for $0 \le (x - 1) \le 1$, therefore $\log \frac{F_i^U}{F_i^L} \le \frac{F_i^U}{F_i^L} - 1$.

Hence,

$$\left\| \log \frac{F_i^U}{F_i^L} \right\|_\infty \le \left\| \frac{F_i^U}{F_i^L} - 1 \right\|_\infty \le \left\| \frac{1}{F_i^L}(F_i^U - F_i^L) \right\|_\infty \le K \|F_i^U - F_i^L\|_\infty \le K\varepsilon.$$

Similarly, by the definition of Ω_n', we can easily show that

$$\left\| \log \frac{F_j^{(1),U} - F_i^L}{F_j^{(1),L} - F_i^U} \right\|_\infty \le K\varepsilon, \qquad \left\| \log \frac{F_k^{(2),U} - F_i^L}{F_k^{(2),L} - F_i^U} \right\|_\infty \le K\varepsilon$$

and

$$\left\| \log \frac{1 - F_j^{(1),L} - F_j^{(2),L} + F_i^U}{1 - F_j^{(1),U} - F_j^{(2),U} + F_i^L} \right\|_\infty \le K\varepsilon.$$

Hence, the fact that L_2-norm is bounded by L_∞-norm results in

$$(7.19) \qquad N_{[]}\{\varepsilon, \mathcal{L}_{n,\delta}, L_2(P)\} \le N_{[]}\{\varepsilon, \mathcal{L}_{n,\delta}, \|\cdot\|_\infty\} \le (\delta/\varepsilon)^{K p_n q_n}.$$

Next, we show that $P\{l(\tau) - l(\tau_n)\}^2 \le K\delta^2$ for any $l(\tau) - l(\tau_n) \in \mathcal{L}_{n,\delta}$. Since for any $\tau = (F, F_1, F_2)$ with $d(\tau, \tau_n) < \delta$, $\|F - F_n\|_{L_2(P_{C_1, C_2})} \le d(\tau, \tau_n) \le \delta$. Then with (C1), (C3) and (C5), Lemma 0.7 in the supplemental article [Wu and Zhang (2012)] implies that for a small $\delta > 0$ and a sufficiently large n, F and F_n are both very close to F_0 at every point in $[l_1, u_1] \times [l_2, u_2]$. Therefore, F and F_n are very close to each other at every point in $[l_1, u_1] \times [l_2, u_2]$. Then the fact that F_n has a positive lower bound results in $1/2 < \frac{F}{F_n} < 2$. Hence $|\log \frac{F}{F_n}| \le K|\frac{F}{F_n} - 1|$, and it follows that

$$P_{C_1, C_2}\left|\log \frac{F}{F_n}\right|^2 \le K P_{C_1, C_2}\left|\frac{F}{F_n} - 1\right|^2 \le K P_{C_1, C_2}|F - F_n|^2 \le K\delta^2.$$

Again by the definition of Ω'_n, we can similarly show that, given a small $\delta > 0$, when n is large enough, the following inequalities are true:

$$P_{C_1, C_2}\left|\log \frac{F_1 - F}{F_{n,1} - F_n}\right|^2 \le K\delta^2, \qquad P_{C_1, C_2}\left|\log \frac{F_2 - F}{F_{n,2} - F_n}\right|^2 \le K\delta^2$$

and

$$P_{C_1, C_2}\left|\log \frac{1 - F_1 - F_2 + F}{1 - F_{n,1} - F_{n,2} + F_n}\right|^2 \le K\delta^2.$$

Hence for any $l(\tau) - l(\tau_n) \in \mathcal{L}_{n,\delta}$, it is true that $P\{l(\tau) - l(\tau_n)\}^2 \le K\delta^2$. It is obvious that $\mathcal{L}_{n,\delta}$ is uniformly bounded by the structure of the log likelihood. Lemma 3.4.2 of van der Vaart and Wellner (1996) indicates that

$$E_P\|\mathbb{G}_n\|_{\mathcal{L}_{n,\delta}} \le K \tilde{J}_{[\,]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\}\left[1 + \frac{\tilde{J}_{[\,]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\}}{\delta^2\sqrt{n}}\right],$$

where

$$\tilde{J}_{[\,]}\{\delta, \mathcal{L}_{n,\delta}, L_2(P)\} = \int_0^\delta \sqrt{1 + \log N_{[\,]}\{\varepsilon, \mathcal{L}_{n,\delta}, L_2(P)\}}\, d\varepsilon \le K(p_n q_n)^{1/2}\delta,$$

by (7.19). This gives $\psi(\delta) = (p_n q_n)^{1/2}\delta + (p_n q_n)/(n^{1/2})$. It is easy to see that $\psi(\delta)/\delta$ is a decreasing function of δ. Note that for $p_n = q_n = n^v$,

$$n^{2pv}\psi(1/n^{pv}) = n^{2pv}n^v n^{-pv} + n^{2pv}n^{2v}n^{-1/2} = n^{1/2}\{n^{pv+v-1/2} + n^{2pv+2v-1}\}.$$

Therefore, if $pv \le (1 - 2v)/2$, $n^{2pv}\psi(1/n^{pv}) \le 2n^{1/2}$. Moreover, $n^{1-2v} \times \psi(1/n^{(1-2v)/2}) = 2n^{1/2}$. This implies if $r_n = n^{\min\{pv, (1-2v)/2\}}$, then $r_n \le \delta_n^{-1}$ and $r_n^2\psi(1/r_n) \le Kn^{1/2}$.

It is obvious that $\mathbb{M}(\hat{\tau}_n) - \mathbb{M}(\tau_n) \ge 0$ and $d(\hat{\tau}_n, \tau_n) \le d(\hat{\tau}_n, \tau_0) + d(\tau_0, \tau_n) \to 0$ in probability. Therefore, it follows by Theorem 3.4.1 in van der Vaart and Wellner (1996) that $r_n d(\hat{\tau}_n, \tau_n) = O_p(1)$. Hence, by $d(\tau_n, \tau_0) \le K(n^{-pv})$

$$r_n d(\hat{\tau}_n, \tau_0) \le r_n d(\hat{\tau}_n, \tau_n) + r_n d(\tau_n, \tau_0) = O_p(1). \qquad \square$$

Acknowledgments. We owe thanks to the Associate Editor and two anony-mous referees for their helpful and constructive comments and suggestions that helped improve the manuscript from an early version.

SUPPLEMENTARY MATERIAL

Technical lemmas (DOI: 10.1214/12-AOS1016SUPP; .pdf). This supplemental material contains some technical lemmas including their proofs that are imperative for the proofs of Theorems 3.1 and 3.2.

REFERENCES

BETENSKY, R. A. and FINKELSTEIN, D. M. (1999). A non-parametric maximum likelihood esti-mator for bivariate interval censored data. *Stat. Med.* **18** 3089–3100.

CURRY, H. B. and SCHOENBERG, I. J. (1966). On Pólya frequency functions. IV. The fundamental spline functions and their limits. *J. Anal. Math.* **17** 71–107. MR0218800

DABROWSKA, D. M. (1988). Kaplan–Meier estimate on the plane. *Ann. Statist.* **16** 1475–1489. MR0964934

DE BOOR, C. (2001). *A Practical Guide to Splines*, revised ed. *Applied Mathematical Sciences* **27**. Springer, New York. MR1900298

DIAMOND, I. D., MCDONALD, J. W. and SHAH, I. H. (1986). Proportional hazards models for current status data: Application to the study of differentials in age at weaning in Pakistan. *Demography* **23** 607–620.

DIAMOND, I. D. and MCDONALD, J. W. (1991). Analysis of current status data. In *Demographic Applications of Event History Analysis* (J. Trussell, R. Hankinson and J. Tilton, eds.) 231–252. Oxford Univ. Press, Oxford.

DUFFY, D. L., MARTIN, N. G. and MATTHEWS, J. D. (1990). Appendectomy in Australian twins. *Am. J. Hum. Genet.* **47** 590–592.

FINKELSTEIN, D. M. and WOLFE, R. A. (1985). A semiparametric model for regression analysis of interval-censored failure time data. *Biometrics* **41** 933–945. MR0833140

GENTLEMAN, R. and VANDAL, A. C. (2001). Computational algorithms for censored-data prob-lems using intersection graphs. *J. Comput. Graph. Statist.* **10** 403–421. MR1939032

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation*. *DMV Seminar* **19**. Birkhäuser, Basel. MR1180321

GRUMMER-STRAWN, L. M. (1993). Regression analysis of current status data: An application to breat feeding. *J. Amer. Statist. Assoc.* **88** 758–765.

HOEL, D. G. and WALBURG, H. E. JR. (1972). Statistical analysis of survival experiments. *J. Natl. Cancer Inst.* **49** 361–372.

HUANG, J. (1996). Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.* **24** 540–568. MR1394975

HUANG, J. and WELLNER, J. A. (1995). Asymptotic normality of the NPMLE of linear functionals for interval censored data, case 1. *Stat. Neerl.* **49** 153–163. MR1345376

JAMSHIDIAN, M. (2004). On algorithms for restricted maximum likelihood estimation. *Comput. Statist. Data Anal.* **45** 137–157. MR2045465

JEWELL, N. P., MALANI, H. and VITTINGHOFF, E. (1994). Nonparametric estimation for a form of doubly censored data with application to two problems in AIDS. *J. Amer. Statist. Assoc.* **89** 7–18.

JEWELL, N. P., VAN DER LAAN, M. and LEI, X. (2005). Bivariate current status data with univariate monitoring times. *Biometrika* **92** 847–862. MR2234190

KOO, J. Y. (1996). Bivariate B-splines for tensor logspline density estimation. *Comput. Statist. Data Anal.* **21** 31–42.

KOOPERBERG, C. (1998). Bivariate density estimation with an application to survival analysis. *J. Comput. Graph. Statist.* **7** 322–341.

LU, M. (2010). Spline-based sieve maximum likelihood estimation in the partly linear model under monotonicity constraints. *J. Multivariate Anal.* **101** 2528–2542. MR2719879

LU, M., ZHANG, Y. and HUANG, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94** 705–718. MR2410018

LU, M., ZHANG, Y. and HUANG, J. (2009). Semiparametric estimation methods for panel count data using monotone *B*-splines. *J. Amer. Statist. Assoc.* **104** 1060–1070. MR2750237

MAATHUIS, M. H. (2005). Reduction algorithm for the NPMLE for the distribution function of bivariate interval-censored data. *J. Comput. Graph. Statist.* **14** 352–362. MR2160818

MEYER, M. C. (2008). Inference using shape-restricted regression splines. *Ann. Appl. Stat.* **2** 1013–1033. MR2516802

NELSEN, R. B. (2006). *An Introduction to Copulas*, 2nd ed. Springer, New York. MR2197664

PRENTICE, R. L. and CAI, J. (1992). Covariance and survivor function estimation using censored multivariate failure time data. *Biometrika* **79** 495–512. MR1187604

PRUITT, R. C. (1991). Strong consistency of self-consistent estimators: General theory and an application to bivariate survival analysis. Technical report, Dept. Statistics, Univ. Minnesota.

QUALE, C. M., VAN DER LAAN, M. J. and ROBINS, J. R. (2006). Locally efficient estimation with bivariate right-censored data. *J. Amer. Statist. Assoc.* **101** 1076–1084. MR2324147

RAMSAY, J. O. (1988). Monotone regression splines in action. *Statist. Sci.* **3** 425–441.

ROSEN, J. B. (1960). The gradient projection method for nonlinear programming. I. Linear constraints. *J. Soc. Indust. Appl. Math.* **8** 181–217. MR0112750

SCHUMAKER, L. L. (1981). *Spline Functions: Basic Theory*. Wiley, New York. MR0606200

SCOTT, D. W. (1992). *Multivariate Density Estimation*. Wiley, New York. MR1191168

SHEN, X. (1998). Proportional odds regression and sieve maximum likelihood estimation. *Biometrika* **85** 165–177. MR1627289

SHIBOSKI, S. C. and JEWELL, N. P. (1992). Statistical analysis of the time dependence of HIV infectivity based on partner study data. *J. Amer. Statist. Assoc.* **87** 360–372.

SHIH, J. H. and LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51** 1384–1399. MR1381050

SONG, S. (2001). *Estimation with Bivariate Interval-Censored Data*. ProQuest LLC, Ann Arbor, MI. MR2717005

STONE, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.* **22** 118–184. MR1272079

SUN, L., WANG, L. and SUN, J. (2006). Estimation of the association for bivariate interval-censored failure time data. *Scand. J. Stat.* **33** 637–649. MR2300908

VAN DER LAAN, M. J. (1996). Efficient estimation in the bivariate censoring model and repairing NPMLE. *Ann. Statist.* **24** 596–627. MR1394977

VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge Univ. Press, Cambridge. MR1652247

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York. MR1385671

WANG, W. and DING, A. A. (2000). On assessing the association for bivariate current status data. *Biometrika* **87** 879–893. MR1813981

WANG, X. and SHEN, J. (2010). A class of grouped Brunk estimators and penalized spline estimators for monotone regression. *Biometrika* **97** 585–601. MR2672485

WANG, Y. and TAYLOR, J. (2004). Monotone constrained tensor-product B-spline with applicaiton to screen studies. Technical report, Dept. Biostatistics, Univ. Michigan.

WONG, G. Y. C. and YU, Q. (1999). Generalized MLE of a joint distribution function with multivariate interval-censored data. *J. Multivariate Anal.* **69** 155–166. MR1703369

WU, Y. and GAO, X. (2011). Sieve estimation with bivariate interval censored data. *Journal of Statistics*: *Advances in Theory and Applications* **16** 37–61.

WU, Y. and ZHANG, Y. (2012) Supplement to "Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data." DOI:10.1214/12-AOS1016SUPP.

ZHANG, Y., HUA, L. and HUANG, J. (2010). A spline-based semiparametric maximum likelihood estimation method for the Cox model with interval-censored data. *Scand. J. Stat.* **37** 338–354. MR2682304

ZHANG, S., ZHANG, Y., CHALONER, K. and STAPLETON, J. T. (2010). A copula model for bivariate hybrid censored survival data with application to the MACS study. *Lifetime Data Anal.* **16** 231–249. MR2608287

DIVISION OF BIOMEDICAL INFORMATICS
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093
USA
E-MAIL: y6wu@ucsd.edu

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242
USA
E-MAIL: ying-j-zhang@uiowa.edu