SORT 48 (2) July-December 2024, 183-208 DOI: 10.57645/20.8080.02.18

Non-parametric estimation of the covariate-dependent bivariate distribution for censored gap times

Ewa Strzalkowska-Kominiak[1], Elisa M. Molanes-López[2] and Emilio Letón[3]

Abstract

In many biomedical studies, recurrent or consecutive events may occur during the follow-up of the individuals. This situation can be found, for example, in transplant studies, where there are two consecutive events which give rise to two times of interest subject to a common random right-censoring time, the first one being the elapsed time from acceptance into the transplantation program to transplant, and the second one the time from transplant to death. In this work, we incorporate the information of a continuous covariate into the bivariate distribution of the two gap times of interest and propose a non-parametric method to cope with it. We prove the asymptotic properties of the proposed method and carry out a simulation study to see the performance of this approach. Additionally, we illustrate its use with Stanford heart transplant data and colon cancer data.

MSC: *62N01, 62N02, 62P10, 62G05.*

Keywords: *Bivariate distribution, Copula function, Covariate, Serial dependence, Random censoring, Kernel estimation.*

1. Introduction

In many biomedical studies, recurrent or consecutive events may occur during the follow-up study of the individuals. In this setting, it is of interest to study the time between consecutive events subject to a common censoring variable. In the literature, these consecutive times are known as gap times. Examples of this situation can be found in the recurrence of breast cancer, bleeding episodes in patients with liver cirrhosis, AIDS studies or transplantation in heart studies. Several authors have already dealt with estimating

[1] Department of Statistics, Universidad Carlos III de Madrid, Spain, email: estrzalk@est-econ.uc3m.es (corresponding author).

[2] Department of Statistics and Operations Research, Facultad de Medicina, Universidad Complutense de Madrid (UCM), Madrid, Spain.

[3] Department of Artificial Intelligence, Universidad Nacional de Educación a Distancia (UNED), Spain.

Received: July 2023

Accepted: December 2023

the distribution function of such gap times (see, e.g., Lin, Sun and Ying (1999), Meira-Machado and Roca-Pardiñas (2011), Serrat and Gómez (2007) or Huang and Louis (1998), among others). Bivariate consecutive data under interval sampling have been considered in Zhu and Wang (2012), where the authors indicated the importance of applying these models to cancer study. Moreover, a bivariate estimation under censoring and with semi-competing risk data, has been studied in Fine, Jiang and Chappell (2001) and Wang (2003). Nevertheless, none of these authors consider the influence of the covariates. In the case of complete data, the estimation of the conditional bivariate distribution has been studied in terms of conditional copula function by Gijbels, Veraverbeke and Omelka (2011). In our paper, we introduce a continuous covariate into the censored gap times setup and propose a new method to estimate the conditional bivariate distribution function. This new methodology is an adaptation and a mixture of the methods proposed by Beran (1981) and van Keilegom (2004), which are further used to estimate as well the bivariate conditional density and the marginal distributions. It provides also a strong basis to study the conditional copulas and measures of conditional dependence.

The paper is organized as follows. In Section 2, we introduce the model. In Section 3, we propose our estimator of the conditional joint distribution function of two gap times. In Section 4, we derive kernel type estimators for the conditional distribution function of the two marginal times. In Section 5, we propose a likelihood based bandwidth selector. We check the behaviour of the proposed methods through a simulation study in Section 6. Finally, we illustrate their use with two real data examples in Section 7.

2. Model description

Let T_1 and T_2 be two consecutive times subject to a common random right-censoring variable, C. Denote by $\tilde{T}_1$ and $\tilde{T}_2$ the observed times, that is, $\tilde{T}_1 = \min(T_1, C)$, $\tilde{T}_2 = \min(T_2, C_2)$, where $C_2 = (C - T_1)1_{\{T_1 \leq C\}}$. Moreover, set $\delta_1 = 1_{\{T_1 \leq C\}}$ and $\delta_2 = 1_{\{T_2 \leq C_2\}}$ as the observed censoring indicators. The following three situations may occur:

a) $T_1 + T_2 \leq C \Rightarrow T_1$ and T_2 are observed, that is, $\tilde{T}_1 = T_1$, $\tilde{T}_2 = T_2$, $\delta_1 = 1$ and $\delta_2 = 1$.

b) $T_1 \leq C < T_1 + T_2 \Rightarrow \tilde{T}_1 = T_1, \delta_1 = 1$, $\tilde{T}_2 = C - T_1$ and $\delta_2 = 0$.

c) $C < T_1 \Rightarrow \tilde{T}_1 = C, \delta_1 = 0$, $\tilde{T}_2 = 0$ and $\delta_2 = 0$.

See Figure 1, where $T = T_1 + T_2$, to get a visual insight of these three possible situations.

The joint cumulative distribution function (cdf) of the pair of consecutive times (T_1, T_2) has been previously studied by several authors: Wang and Wells (1998) under the assumption of independence between the censoring time and the gap times; Visser (1996) considered that censoring may depend upon previous gap times, although requiring discrete censoring time and gap times; de Uña-Álvarez and Meira-Machado (2008), Strzalkowska-Kominiak and Stute (2010) and de Uña-Álvarez and Amorim (2011) studied consecutive survival times with a common censoring variable. In this work, we consider this last setting with the novelty of including extra information given by a one-dimensional continuous covariate, let say X, to gain in estimation efficiency and study its influence on survival times.

To illustrate our setting, we will consider the following examples. The first one is based on the Stanford heart transplant and a colon cancer data set, described in Meira-Machado and Roca-Pardiñas (2011) and Kalbfleisch and Prentice (1980), among others. In this dataset, there are two times of interest, $T_1 =$ time from acceptance into the transplantation program to transplant (in months), and $T_2 =$ time from transplant to death (in months), and some covariates available such as age, year of acceptance and surgery (prior bypass surgery). Additionally, there are other two variables ($delta$ and $status$) that specify if the individual has received a transplant (or not), $delta = 1$ ($delta = 0$), and if he/she has died (or not), $status = 1$ ($status = 0$). Those individuals with both times observed correspond with those that have $delta = 1$ and $status = 1$. In the case that the first time is observed but not the second time, the individuals have $delta = 1$ and $status = 0$. For the rest, with $delta = 0$, none of the two times are observed. The second example is based on a colon cancer dataset, previously studied in Lawless and Yilmaz (2011). For this dataset, the two times of interest are $T_1 =$ time from study registration to recurrence (in years), and $T_2 =$ time from recurrence to death (in years), and the covariates available are treatment and age. Analogously to the first example, there are censoring indicators that state if none, one or both times are censored. In both examples, we aim to assess the effect of age on the vector (T_1, T_2). In the second example, we are also interested in studying the treatment effect. In both cases, the censoring variable C is the time until end of the study which, for a given age, can be assumed independent of the lifetimes of interest.

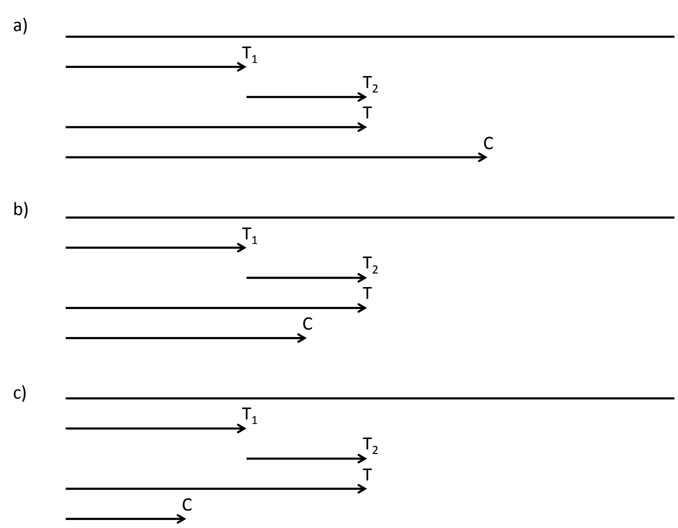

Figure 1. *The three possible situations for T_1, T_2, T and C.*

Based on n i.i.d. replicates, $\{(\tilde{T}_{1i}, \tilde{T}_{2i}, \delta_{1i}, \delta_{2i}), i = 1, ..., n\}$, of the random vector $(\tilde{T}_1, \tilde{T}_2, \delta_1, \delta_2)$, our goal in this paper is to estimate the bivariate cdf of (T_1, T_2) given that $X = x$, that is,

$$\mathbf{F}(y_1, y_2|x) = \mathbb{P}(T_1 \leq y_1, T_2 \leq y_2|X = x), \tag{1}$$

under the following assumption:

A1: (T_1, T_2) is independent of C given X.

Note that under assumption A1, it follows the following condition:

A2: T_2 is independent of $C_2 = (C - T_1)1_{\{T_1 \leq C\}}$ given T_1 and X.

From here on, we denote by $F_T(t)$ the cdf of T, $G(t)$ the cdf of C, and by $G(t|x)$ the conditional cdf of C given X, that is, $G(t|x) = \mathbb{P}(C \leq t|X = x)$. We assume throughout the paper that the densities $\mathbf{f}(y_1, y_2|x)$, $f_X(x)$ and $g(t|x)$, related to $\mathbf{F}(y_1, y_2|x)$, $F_X(x) = \mathbb{P}(X \leq x)$ and $G(t|x)$, exist and are continuous, and that $F_1(y_1|x) = \mathbf{F}(y_1, \infty|x)$ and $F_X(x)$ are differentiable up to order two. Note that $\mathbf{F}$ is identifiable only in the region $\{(t_1, t_2) : t_1 + t_2 \leq \tau_c(x)\}$, where $\tau_c(x)$ is the right hand side of the support of C given $X = x$, that is, $\tau_c(x) = \inf\{t : G(t|x) = 1\}$.

3. Conditional joint estimators

The aim of this section is to propose a nonparametric estimator of (1).

Let's $f_{21}(y_2|t_1, x)$ and $f_1(t_1|x)$ denote the densities related to the distributions $F_{21}(y_2|t_1, x)$ and $F_1(t_1|x)$, respectively, where $F_{21}(y_2|t_1, x) = \mathbb{P}(T_2 \leq y_2|T_1 = t_1, X = x)$. Hence

$$\begin{aligned}\mathbf{F}(y_1, y_2|x) &= \int_0^{y_1}\int_0^{y_2} f(t_1, t_2|x)dt_1dt_2 = \int_0^{y_1}\int_0^{y_2} f_{21}(t_2|t_1, x)f_1(t_1|x)dt_2dt_1 \\ &= \int_0^{y_1} F_{21}(y_2|t_1, x)f_1(t_1|x)dt_1.\end{aligned}$$

Hence, our estimator is based on the following, more general, identity

$$\mathbf{F}(y_1, y_2|x) = \int_0^{y_1} F_{21}(y_2|t_1, x)F_1(dt_1|x),$$

where

$$F_1(y_1|x) = \mathbb{P}(T_1 \leq y_1|X = x). \tag{2}$$

Moreover, under assumption A1, we have that

$$\begin{aligned}\mathbb{P}(T_2 \leq y_2|T_1 = t_1, \delta_1 = 1, X = x) &= \mathbb{P}(T_2 \leq y_2|T_1 = t_1, C \geq t_1, X = x) \\ &= \mathbb{P}(T_2 \leq y_2|T_1 = t_1, X = x)\end{aligned}$$

and therefore

$$F_{21}(y_2|t_1, x) = \mathbb{P}(T_2 \leq y_2|T_1 = t_1, \delta_1 = 1, X = x). \tag{3}$$

In the following subsections, we propose nonparametric estimators of (2), (3) and (1), respectively.

3.1. Estimation of $F_1(t_1|x)$

We propose to estimate $F_1(t_1|x)$ by the Beran (1981) estimator of the conditional cdf of T_1 given $X = x$, denoted by $F_{1n}(t_1|x)$. More precisely, F_{1n} is a standard version of the Beran estimator, defined as

$$F_{1n}(y|x) = 1 - \prod_{i=1}^{n}\left[1 - \frac{B_{in}(x) 1_{\{\tilde{T}_{1i}\le y\}}\delta_{1i}}{\sum_{j=1}^{n} 1_{\{\tilde{T}_{1j}\ge \tilde{T}_{1i}\}} B_{jn}(x)}\right], \tag{4}$$

where

$$B_{in}(x) = \frac{K(\frac{x-X_i}{h_1})}{\sum_{j=1}^{n} K(\frac{x-X_j}{h_1})},$$

and h_1 is a bandwidth parameter. As usual, K is a bounded kernel function with the following properties:

$$\int uK(u)du = 0, \text{ and } \int u^2K(u)du < \infty.$$

3.2. Estimation of $F_{21}(t_2|t_1,x)$

We propose to estimate $F_{21}(t_2|t_1,x)$ by the Beran estimator of the conditional cdf of T_2 given $(T_1,\delta_1,X) = (t_1,1,x)$, denoted by $F_{21n}(t_2|t_1,x)$.

Let define

$$H^*(t|t_1,x) = \mathbb{P}(\tilde{T}_2 \le t|T_1 = t_1, \delta_1 = 1, X = x).$$

and

$$\tilde{H}^*(t|t_1,x) = \mathbb{P}(\tilde{T}_2 \le t, \delta_2 = 1|T_1 = t_1, \delta_1 = 1, X = x).$$

Given that the assumption A1 is valid, the condition A2 holds and therefore, we obtain that

$$1 - H^*(t^-|t_1,x) = (1 - F_{21}(t^-|t_1,x))\mathbb{P}(C_2 \ge t|T_1 = t_1, \delta_1 = 1, X = x)$$

and

$$\tilde{H}^*(t|t_1,x) = \int_0^t \mathbb{P}(C_2 \ge s|T_1 = t_1, \delta_1 = 1, X = x)F_{21}(ds|t_1,x).$$

Hence, the hazard rate of F_{21} equals

$$d\Lambda_{21}(t|t_1,x) = \frac{F_{21}(dt|t_1,x)}{1 - F_{21}(t^-|t_1,x)} = \frac{\tilde{H}^*(dt|t_1,x)}{1 - H^*(t^-|t_1,x)}.$$

Finally, we obtain

$$F_{21n}(t_2|t_1,x) = 1 - \prod_{i=1}^{n}\left[1 - \frac{\tilde{B}_{in}(x,t_1) 1_{\{\tilde{T}_{2i}\le t_2\}}\delta_{2i}}{\sum_{j=1}^{n} 1_{\{\tilde{T}_{2j}\ge \tilde{T}_{2i}\}} \tilde{B}_{jn}(x,t_1)}\right],$$

the Beran estimator based on the restricted sample (that is, only the data with $\delta_1 = 1$), where

$$\tilde{B}_{in}(x,t_1) = \frac{\delta_{1i} K(\frac{x-X_i}{h_1}) K(\frac{t_1-\tilde{T}_{1i}}{h_2})}{\sum_{j=1}^n \delta_{1j} K(\frac{x-X_j}{h_1}) K(\frac{t_1-\tilde{T}_{1j}}{h_2})}$$

and h_1 and h_2 are bandwidth parameters.

3.3. Estimation of $\mathbf{F}(y_1,y_2|x)$

Following the same ideas as in van Keilegom (2004), we propose to estimate $\mathbf{F}$ by plugging in appropriate estimators of F_1 and F_{21}, that is,

$$\mathbf{F}_n(y_1,y_2|x) = \int_0^{y_1} F_{21n}(y_2|t_1,x) F_{1n}(dt_1|x), \tag{5}$$

where $F_{1n}(t_1|x)$ and $F_{21n}(t_2|t_1,x)$ have been previously introduced in Subsections 3.1 and 3.2.

In order to estimate $\mathbf{f}(y_1,y_2|x)$, the density function associated to the bivariate cdf $\mathbf{F}(y_1,y_2|x)$, we consider an alternative way to write $\mathbf{F}_n(y_1,y_2|x)$ as follows

$$\mathbf{F}_n(y_1,y_2|x) = \sum_{i=1}^n W_{1i}(x) 1_{\{\tilde{T}_{1i} \le y_1\}} F_{21n}(y_2|\tilde{T}_{1i},x), \tag{6}$$

where

$$W_{1i}(x) = F_{1n}(\tilde{T}_{1i}|x) - F_{1n}(\tilde{T}_{1i}^-|x)$$

and we set the following weights

$$W_{ik}^B(x) = \mathbf{F}_n(\tilde{T}_{1i},\tilde{T}_{2k}|x) - \mathbf{F}_n(\tilde{T}_{1i}^-,\tilde{T}_{2k}|x) - \mathbf{F}_n(\tilde{T}_{1i},\tilde{T}_{2k}^-|x) + \mathbf{F}_n(\tilde{T}_{1i}^-,\tilde{T}_{2k}^-|x).$$

Taking into account that

$$W_{ik}^B(x) = W_{1i}(x) W_{2ki}(x),$$

where

$$W_{2ki}(x) = F_{2n}(\tilde{T}_{2k}|\tilde{T}_{1i},x) - F_{2n}(\tilde{T}_{2k}^-|\tilde{T}_{1i},x),$$

we propose to estimate $\mathbf{f}(y_1,y_2|x)$, the density function associated to the bivariate cdf $\mathbf{F}(y_1,y_2|x)$, by

$$\mathbf{f}_n(y_1,y_2|x) = \frac{1}{h_2 h_3} \sum_{i=1}^n \sum_{k=1}^n W_{ik}^B(x) K\left(\frac{y_1-\tilde{T}_{1i}}{h_2}\right) K\left(\frac{y_2-\tilde{T}_{2k}}{h_3}\right), \tag{7}$$

where h_3 is an additional bandwidth parameter.

Before we state our first theorem, we set

$$H(t|x) = \mathbb{P}(\tilde{T}_1 \le t|X = x),$$

and define $\tau_1(x)$ and $\tau_2(t_1,x)$ as follows

$$0 \leq \tau_1(x) < \inf\{t : H(t|x) = 1\} < \infty, \quad \tau_2(t_1,x) < \inf\{t : H^*(t|t_1,x) = 1\}.$$

Set

$$A(x) = \{(t_1,t_2) : t_1 \leq \tau_1(x), t_2 \leq \inf_{t \leq t_1} \tau_2(t,x)\}.$$

Theorem 3.1. *Let $(y_1,y_2) \in A(x)$ and $x \in \{u : f_X(u) > 0\}$, where f_X denotes the density of X. Under assumption A1, if $nh_1^5 \to 0$, $nh_1^3 \to \infty$ and $nh_2^5 \to c > 0$, we have*

$$\sqrt{nh_1}(\mathbf{F}_n(y_1,y_2|x) - \mathbf{F}(y_1,y_2|x)) \to \mathcal{N}(0, \sigma_1^2(y_1,y_2|x)),$$

where

$$\sigma_1^2(y_1,y_2|x) = \frac{\rho_1^2(y_1,y_2,x) \int K^2(t)dt}{f_X(x)}$$

and $\rho_1^2(y_1,y_2,x)$ is given in (A.12).

Proof. See Appendix A.

Remark 3.2. *The variance $\sigma_1^2(y_1,y_2|x)$ has a complicated structure since it reflects variability of three leading terms in the expansion of the process from Theorem 3.1 and hence in practical applications it needs to be approximated using bootstrap or jackknife. However, when no censoring is present, it reduces to the well known expression with $\rho_1^2(y_1,y_2,x) = \mathbf{F}(y_1,y_2|x)(1 - \mathbf{F}(y_1,y_2|x))$.*

Remark 3.3. *The estimator $\mathbf{F}_n(y_1,y_2|x)$ given in (6) can be extended to the case where T_1 is, additionally to being censored from the right, left truncated by a random variable Z, independent of T_1 and T_2. In such a case, we only need to replace F_{1n} by the estimator introduced by Iglesias-Pérez and González-Manteiga (1999).*

4. Conditional marginal estimators

Our goal in this section is to estimate $F_1(t|x)$ and $F_2(t|x)$, where F_i denotes the conditional distribution function of T_i given $X = x$, for $i = 1,2$. Remark that $F_1(t|x)$ can be consistently estimated by a standard conditional Beran estimator as given in (4). See Beran (1981) or González-Manteiga and Cadarso-Suárez (1994), for details. Regarding $F_2(t|x)$, due to induced dependent censoring, we cannot use the standard Beran estimator. So we propose a new estimator derived from the proposed bivariate estimator $\mathbf{F}_n(t_1,t_2|x)$. We should recall, however, that we have the asymptotic properties of this estimator only for $(t_1,t_2) \in A(x)$. Hence, similarly to van Keilegom (2004), we define

$$F_2^{\tau_1}(t|x) = \mathbf{F}(\tau_1(x),t|x)$$

that, under the assumption $\inf\{t : F_1(t|x) = 1\} \leq \inf\{t : G(t|x) = 1\}$, can be made arbitrarily close to $F_2(t|x)$. Consequently, we define the following consistent estimator

$$F_{2n}^{\tau_1}(t|x) = \mathbf{F}_n(\tau_1(x), t|x).$$

Since, as mentioned before, the estimation of $F_1(t|x)$ has been widely studied in the literature, we focus in our simulation study on the new estimator that we have proposed for $F_2(t|x)$. In order to measure the performance of this conditional estimator, we consider the Kolmogorov-Smirnov (KS) distance, that is,

$$KS^B(x) = \sup_t |F_{2n}^{\tau_1}(t|x) - F_2^{\tau_1}(t|x)|. \tag{8}$$

5. Likelihood based bandwidth selection

In this section, we derive the likelihood function for two consecutive censored times, T_1 and T_2. To give some insights, we will consider first the case of two independent censored times C_1 and C_2, so that T_1 is censored by C_1 and T_2 is censored by C_2. For this case, the likelihood function is given by

$$\begin{aligned}\ell(\gamma) = \sum_{i=1}^{n} \Big[& \delta_{1i}\delta_{2i} \log(\mathbf{f}(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) + \delta_{1i}(1-\delta_{2i}) \log(g_1(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) \\ & + \delta_{2i}(1-\delta_{1i}) \log(g_2(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) + (1-\delta_{1i})(1-\delta_{2i}) \log(\bar{\mathbf{F}}(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) \Big],\end{aligned}$$

where the density and cdf of (T_1, T_2) given X are assumed to be known up to some parameter γ, and

$$\begin{aligned} g_1(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i) &= \int_{\tilde{T}_{2i}}^{\infty} \mathbf{f}(\tilde{T}_{1i}, t|X_i)dt, \\ g_2(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i) &= \int_{\tilde{T}_{1i}}^{\infty} \mathbf{f}(t, \tilde{T}_{2i}|X_i)dt, \\ \bar{\mathbf{F}}(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i) &= \mathbb{P}(T_1 > \tilde{T}_{1i}, T_2 > \tilde{T}_{2i}|X = X_i). \end{aligned}$$

Taking into account that in our setup there is one common censoring time C so that $(1-\delta_{1i})\delta_{2i} = 0$ and if $1-\delta_{1i} = 1$, then $1-\delta_{2i} = 1$ and $\tilde{T}_{2i} = 0$, we have that $\bar{\mathbf{F}}(\tilde{T}_{1i}, 0|X_i) = 1 - F_1(\tilde{T}_{1i}|X_i)$ and the likelihood function is reduced to

$$\begin{aligned}\ell(\gamma) = \sum_{i=1}^{n} \Big[& \delta_{1i}\delta_{2i} \log(\mathbf{f}(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) + \delta_{1i}(1-\delta_{2i}) \log(g_1(\tilde{T}_{1i}, \tilde{T}_{2i}|X_i)) \\ & + (1-\delta_{1i}) \log(1 - F_1(\tilde{T}_{1i}|X_i)) \Big],\end{aligned} \tag{9}$$

which is a generalization of the discrete case covered in Visser (1996).

Denoting $\mathbb{K}(y) = \int_y^\infty K(z)dz$ and replacing the unknown quantities in (9) by estimators involving three bandwidth parameters, h_1, h_2 and h_3, we derive an estimated log-likelihood function in terms of $\gamma = (h_1, h_2, h_3)$ to work with. Specifically, we denote by $\ell_n(h_1, h_2, h_3)$ the estimated log-likelihood function where $\mathbf{f}$ is estimated by $\mathbf{f}_n$ given in (7) and g_1 is estimated by

$$g_{1n}(y_1, y_2|x) = \frac{1}{h_2} \sum_{i=1}^{n} \sum_{k=1}^{n} W_{ik}^B(x) K\left(\frac{y_1 - \tilde{T}_{1i}}{h_2}\right) \mathbb{K}\left(\frac{y_2 - \tilde{T}_{2k}}{h_3}\right).$$

In order to select the bandwidth parameters in practice, we propose to maximize the estimated log-likelihood function $\ell_n(h_1, h_2, h_3)$. The bandwidth selection is a problematic issue already in the standard Beran estimator (with one variable of interest and one dimensional covariate). An alternative to the proposed method would be to minimize the integrated square error (w.r.t (t_1, t_2, x)). However, our method is computationally much faster, it does not require numerical integration and it provides an estimate of the corresponding bivariate conditional density as a byproduct. Moreover, it can be easily adapted to multidimensional covariates by applying the so called single-index model (see, e.g., Strzalkowska-Kominiak and Cao (2013) for a single-index model under one-dimensional censoring variable).

6. Simulation study

A simulation study is carried out here to check the performance of the new estimator, previously proposed in (5). The scenarios that we consider are based on Huang, Luo and Follmann (2011). More specifically, we start by generating two exponential random variables, (T_1^0, T_2^0), with unit means, linked by a Gaussian copula. This can be done by following the steps:

- Generating n i.i.d. random variables, $A_i \sim \mathcal{N}(0, \sqrt{\rho})$ and $B_{ki} \sim \mathcal{N}(0, \sqrt{1-\rho})$, for $k = 1, 2$ and $i = 1, ..., n$.
- Setting $T_{ki}^0 = -\ln(1 - \Phi(A_i + B_{ki}))$, where Φ is the cumulative distribution function of a standard normal variable.

Since $(A + B_1, A + B_2)$ follows a standardized bivariate normal variable with correlation ρ, it is easy to prove that

$$\mathbf{F}^0(t_1, t_2) = \mathbb{P}(T_1^0 \le t_1, T_2^0 \le t_2) = \Phi_2(\Phi^{-1}(1 - e^{-t_1}), \Phi^{-1}(1 - e^{-t_2})),$$

where Φ_2 denotes the cdf of $(A + B_1, A + B_2)$. Taking into account Sklar's theorem (see Sklar (1959)), it is easy to see that

$$\mathbf{F}^0(t_1, t_2) = \mathbf{C}(F_1^0(t_1), F_2^0(t_2)),$$

where $\mathbf{C}$ refers to the bivariate Gaussian copula with parameter ρ and F_k^0 denotes the cdf of T_k^0, for $k = 1, 2$.

To incorporate the effect of the covariate X in the gap times (T_1, T_2), we define $T_k = \left(T_k^0/a(X)\right)^{1/\kappa}$, for $k = 1,2$, where $a(X) = \exp(\beta X)$. It is easy to check that, conditionally on $X = x$, T_k, for $k = 1,2$, is a Weibull distributed random variable with conditional cdf $F_k(t|x) = 1 - e^{-(t/\lambda)^\kappa}$, for $t > 0$, where the scale parameter equals $\lambda = (1/a(x))^{1/\kappa}$ and the shape parameter equals κ. Observe that the conditional hazard rate of this model is given by $h(t|x) = a(x)\kappa t^{\kappa-1}$ with $a(x) = \exp(\beta x)$. Hence β refers to the parameter of the Cox model with basic hazard defined as $h_0(t) = \kappa t^{\kappa-1}$.

In order to get a copula representation of the conditional cdf of (T_1, T_2) given $X = x$, we use the fact that

$$
\begin{aligned}
\mathbf{F}(t_1,t_2|x) &= \mathbb{P}\left(T_1^0 \le a(X)t_1^\kappa, T_2^0 \le a(X)t_2^\kappa | X = x\right) = \mathbf{F}^0(a(x)t_1^\kappa, a(x)t_2^\kappa) \\
&= \Phi_2(\Phi^{-1}(1-e^{-a(x)t_1^\kappa}), \Phi^{-1}(1-e^{-a(x)t_2^\kappa})) \\
&= \Phi_2(\Phi^{-1}(F_1(t_1|x)), \Phi^{-1}(F_2(t_2|x))),
\end{aligned}
$$

where $F_k(t|x)$ denotes the cdf of T_k given $X = x$, for $k = 1,2$.

Therefore, this procedure leads us to two Weibull variables, T_1 and T_2, linked through the bivariate Gaussian copula. In order to get more flexibility, other copula functions could be used to link T_1 and T_2. In that case, the conditional cdf of (T_1, T_2) given $X = x$ would be as follows $\mathbf{F}(t_1,t_2|x) = \mathbf{C}(F_1(t_1|x), F_2(t_2|x))$, where $\mathbf{C}$ denotes a bivariate copula function.

In Subsection 6.1 we consider the Gaussian copula with fixed parameter $\rho = 0.5$ and variable parameter $\rho = \rho(x) = x/10$. Furthermore, we study in Subsection 6.2 the Clayton copula with parameters $\theta = 0.5$ and $\theta = 5$. Additionally, we consider a cure rate model in Subsection 6.3, where $\mathbf{C}$ denotes a bivariate Gaussian copula with parameter $\rho = 0.5$ and T_2 given $X = x$ comes from an improper distribution $F_2^0(t_2|x) = pF_2(t_2|x)$ with $p = 0.8$. In order to generate values from these copulas see, for example, Cherubini, Luciano and Vecchiato (2004) and Wu, Valdez and Sherris (2007).

Regarding the marginals, we consider $\beta = 0.3$ and three different values for $\kappa = 1.5, 2, 3$. We investigate the behaviour of the estimator under two different scenarios for the distribution of the covariate, namely $X \sim U(0,10)$ and $X \sim \mathcal{N}(5,1)$. Furthermore, $C \sim U(0,\tau_c)$, where we choose two values for τ_c, such that the proportion of subjects with zero events is approximately 10% and 15%, that is, there are $0.1n$ and $0.15n$ subjects with $\delta_{1i} = 0$ and $\delta_{2i} = 0$. Moreover, for those τ_c we have, respectively, 10% and 15% events for which $\delta_{1i} = 1$ but $\delta_{2i} = 0$. Additionally, for the cure rate model, we choose $\tau_c = 5$ so that for $p = 0.8$, we have 9% $\delta_{1i} = \delta_{2i} = 0$ and 25% events with $\delta_{1i} = 1$ but $\delta_{2i} = 0$. In the following subsections, we present the simulation results for three different models.

To simplify the likelihood based selection of the three dimensional bandwidth vector, we consider that $h_j = c_j h$, for $j = 1,2,3$ and maximize the likelihood function over one parameter h. The most natural choice for the constants c_j is $c_1 = \hat{\sigma}(X)$, $c_2 = \hat{\sigma}(\tilde{T}_1)$ and $c_3 = \hat{\sigma}(\tilde{T}_2)$, where $\hat{\sigma}$ denotes the sample standard deviation.

This simulation study is carried out in the open-source software R and shows the performance of our proposed estimator in terms of bias, variance and mean squared error (MSE) that are calculated by resampling using 200 trials. Additionally, we check the estimator of the conditional marginal, $F_2(t|x)$, with the Kolmogorov-Smirnov (KS) distance defined in (8) with $\tau_1 = +\infty$.

6.1. Gaussian copula model

In this subsection, we use the Gaussian copula with a constant parameter $\rho = 0.5$. We access the quality of the estimation for different values of the parameter $\kappa \in \{1.5, 2, 3\}$ and $X \sim U(0,10)$ (Tables 1-3). Furthermore, we check the behaviour of our estimation procedure when $X \sim \mathcal{N}(5,1)$ for $\kappa = 2$ (Tables 4-5). For a given x, the estimators are computed in the middle point of the support for $(t_1,t_2) = (E(T_1), E(T_2))$ and in the right side of the support $(t_1,t_2) = (1,1)$ using 200 trials. In the case of $X \sim U(0,10)$ we choose $x = 5$ and when $X \sim \mathcal{N}(5,1)$ we investigate additionally the behaviour in low density regions for $x = 3$. In continuation, we use a dependent Gaussian copula model with parameter $\rho = \rho(X) = X/10$ and $X \sim U(0,10)$ so that not only the marginals depend on the covariate, but also the correlation between T_1 and T_2 changes with x. The estimators are computed in $(t_1,t_2,x) = (E(T_1), E(T_2), 5)$ so that $\rho(x) = 0.5$ and in $(t_1,t_2,x) = (E(T_1), E(T_2), 8)$ so that $\rho(x) = 0.8$ (see Table 6).

Table 1. *Bias, variance and MSE in (t_1,t_2,x) when $\rho = 0.5$, $\kappa = 1.5$, $X \sim U(0,10)$ and $n = 100, 200$.*

(t_1,t_2,x)		$(0.39,0.39,5)$ $\mathbf{F}(0.39,0.39\|5) \approx 0.51$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	3.9	0.0077	0.006	0.0061	-0.0186	8e-04	0.0012	0.1375
	2.5	0.0082	0.0062	0.0063	-0.0219	0.0011	0.0016	0.1459
200	3.9	-0.0066	0.0043	0.0043	-0.0173	0.0006	0.0009	0.1088
	2.5	-0.0061	0.0042	0.0042	-0.0157	0.0006	0.0009	0.1131

Table 2. *Bias, variance and MSE in (t_1,t_2,x) when $\rho = 0.5$, $\kappa = 2$, $X \sim U(0,10)$ and $n = 100, 200$.*

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.45$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	4.6	-0.0010	0.0064	0.0064	-0.0198	0.0009	0.0013	0.1456
	3.1	-0.0138	0.0060	0.0062	-0.0229	0.0010	0.0015	0.1468
200	4.6	-0.0031	0.0037	0.0038	-0.0144	0.0004	0.0006	0.1076
	3.1	-0.0027	0.0037	0.0037	-0.0145	0.0006	0.0008	0.1123

Table 3. *Bias, variance and MSE in (t_1,t_2,x) when $\rho = 0.5$, $\kappa = 3$, $X \sim U(0,10)$ and $n = 100,200$.*

(t_1,t_2,x)		$(0.56,0.56,5)$ $\mathbf{F}(0.56,0.56\|5) \approx 0.38$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	5.1	-0.0009	0.0046	0.0046	-0.0212	0.001	0.0014	0.1394
	3.9	-0.0031	0.0057	0.0057	-0.022	0.0011	0.0016	0.1529
200	5.1	0.0014	0.0033	0.0033	-0.0133	0.0005	0.0006	0.1126
	3.9	-0.0084	0.0032	0.0033	-0.0146	0.0006	0.0008	0.1126

Table 4. *Bias, variance and MSE in (t_1,t_2,x) with $x = 5$ when $\rho = 0.5$, $\kappa = 2$, $X \sim \mathcal{N}(5,1)$ and $n = 100,200$.*

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.45$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	4.6	-0.0150	0.0037	0.0039	-0.0008	0.0004	0.0004	0.1152
	3.1	-0.0112	0.0039	0.0040	-0.0013	0.0005	0.0005	0.1171
200	4.6	-0.0120	0.0024	0.0026	-0.0029	0.0003	0.0003	0.0875
	3.1	-0.0120	0.0020	0.0021	-0.0027	0.0003	0.0003	0.0929

Table 5. *Bias, variance and MSE in (t_1,t_2,x) with $x = 3$ when $\rho = 0.5$, $\kappa = 2$, $X \sim \mathcal{N}(5,1)$ and $n = 100,200$.*

(t_1,t_2,x)		$(0.46,0.46,3)$ $\mathbf{F}(0.46,0.46\|3) \approx 0.24$			$(1,1,3)$ $\mathbf{F}(1,1\|3) \approx 0.86$			x=3
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	4.6	0.0417	0.0121	0.0138	0.0528	0.0076	0.0104	0.2319
	3.1	0.0225	0.0120	0.0125	0.0414	0.0076	0.0093	0.2395
200	4.6	0.0293	0.0068	0.0076	0.0401	0.0055	0.0071	0.1821
	3.1	0.0450	0.0079	0.0099	0.0454	0.0057	0.0077	0.2009

As can be seen in Tables 1-4, our proposed estimator gives very good results. There is no significant difference in the performance of the estimator for different values of κ nor it changes with change of the distribution of the covariate when $x = 5$ is considered. Introducing a dependent copula model (Table 6) also does not affect negatively the results. As expected the results improve when increasing the sample size and surprisingly they are not very affected by increasing the censoring rate. Additionally, even though the asymptotic properties were proved for compact sets, our new method gives good results even for (t_1,t_2) being near to the right-hand side of the support. The only exception are the low density regions with normal covariate (Table 5), were the quality of estimation declines. This is, however, not surprising since this type of behaviour we observe also with standard Beran estimator.

Table 6. *Bias, variance and MSE in* (t_1,t_2,x) *when* $\rho = x/10$, $\kappa = 2$, $X \sim U(0,10)$ *and* $n = 100,200$.

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.45$			x=5	$(0.46,0.46,8)$ $\mathbf{F}(0.46,0.46\|8) \approx 0.86$			x=8
n	τ_c	Bias	Variance	MSE	KS	Bias	Variance	MSE	KS
100	4.6	-0.0072	0.0056	0.0056	0.1371	-0.0480	0.0038	0.0061	0.1511
	3.1	-0.0040	0.0066	0.0066	0.1468	-0.0437	0.0035	0.0054	0.1525
200	4.6	-0.0051	0.0034	0.0034	0.1098	-0.0270	0.0022	0.0029	0.1093
	3.1	-0.0102	0.0036	0.0037	0.1115	-0.0339	0.0025	0.0037	0.1119

6.2. Clayton copula model

In this subsection, we use the Clayton copula with a constant parameter $\theta = 0.5$ and $\theta = 5$. As in Subsection 6.1, the estimators are computed in the middle point, $(t_1,t_2,x) = (E(T_1),E(T_2),5)$, where $E(T_1) = E(T_2) = 0.46$ and $\mathbf{F}(0.46,0.46|5) \approx 0.41$, and in the right side of the support $(1,1,5)$, where $\mathbf{F}(1,1|5) \approx 0.98$ (see Tables 7 and 8). From these tables, we observe that the behaviour of our estimator is similar to the case of the Gaussian copula model.

Table 7. *Bias, variance and MSE in* (t_1,t_2,x) *when* $\theta = 0.5$, $\kappa = 2$, $X \sim U(0,10)$ *and* $n = 100,200$.

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.41$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	4.6	0.0038	0.0055	0.0055	-0.0218	0.0009	0.0014	0.1355
	3.1	0.0174	0.0055	0.0058	-0.0250	0.0011	0.0017	0.1420
200	4.6	0.0054	0.0032	0.0033	-0.0176	0.0004	0.0007	0.1085
	3.1	0.0072	0.0030	0.0031	-0.0200	0.0006	0.0010	0.1099

Table 8. *Bias, variance and MSE in* (t_1,t_2,x) *when* $\theta = 5$, $\kappa = 2$, $X \sim U(0,10)$ *and* $n = 100,200$.

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.41$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.98$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	4.6	-0.0259	0.0079	0.0086	-0.0131	0.0010	0.0012	0.1568
	3.1	-0.0237	0.0068	0.0074	-0.0144	0.0009	0.0011	0.1541
200	4.6	-0.0140	0.0043	0.0045	-0.0089	0.0005	0.0006	0.1252
	3.1	-0.0080	0.0055	0.0056	-0.0097	0.0006	0.0007	0.1298

6.3. Cure rate model

In this subsection, we use the Gaussian copula with a constant parameter $\rho = 0.5$. The distribution function of T_2 given $X = x$ is improper, that is $F_2^0(t_2|x) = pF_2(t_2|x)$ with $p = 0.8$. The estimators are computed in the middle point, $(t_1,t_2,x) = (0.46, 0.46, 5)$, where $\mathbf{F}(0.46, 0.46|5) \approx 0.38$ and in the right side of the support $(1,1,5)$, where $\mathbf{F}(1,1|5) \approx 0.79$ (see Table 9).

Table 9. *Bias, variance and MSE in (t_1,t_2,x) when $p = 0.8$, $\theta = 0.5$ and $n = 100$*

(t_1,t_2,x)		$(0.46,0.46,5)$ $\mathbf{F}(0.46,0.46\|5) \approx 0.38$			$(1,1,5)$ $\mathbf{F}(1,1\|5) \approx 0.79$			x=5
n	τ_c	Bias	Variance	MSE	Bias	Variance	MSE	KS
100	5	-0.0190	0.0029	0.0033	-0.0312	0.0026	0.0036	0.1123

As can be seen in Table 9, when one of the marginal distributions is improper, our new estimator gives very good results already for $n = 100$. However, from a theoretical point of view, those models are out of scope of the present paper.

7. Examples

7.1. Stanford heart transplant data

In this section, we analyze the example of Stanford heart transplant data, previously introduced in Section 1. There are 103 individuals in this dataset, 45 out of them received transplant ($delta = 1$) and died ($status = 1$), 24 received transplant ($delta = 1$) and were still alive at the end of the study ($status = 0$), for 4 patients the study ended before the transplantation ($delta = 0$ and $status = 0$) and the remaining 30 died before receiving the transplant ($delta = 0$ and $status = 1$). In this real example, the death before receiving the transplant can be considered as a semi-competing risk because it is a termination event that can potentially censor the non terminating event of receiving the transplant (see Zhao and Zhou (2010), among others). In fact, this is the case for those 30 patients that have $delta = 0$ and $status = 1$. Considering that T_3 denotes the time from acceptance into the transplantation program to death (in months), it is easy to adapt our model to this situation by replacing C by $C_1 = \min(C, T_3)$. Under this setting, we set $\delta_1 = delta$ and $\delta_2 = status \times delta$. For the sake of illustration of our methodology, we consider the covariate $X = age$.

As in the previous section, we consider that $h_j = c_j h$, for $j = 1,2,3$ and maximize the likelihood function over one parameter h. As before the constants c_j are defined as follows $c_1 = \hat{\sigma}(X)$, $c_2 = \hat{\sigma}(\tilde{T}_1)$ and $c_3 = \hat{\sigma}(\tilde{T}_2)$, where $\hat{\sigma}$ denotes the sample standard deviation. Using this approach, we obtain the following bandwidth $h = (6.29, 1.57, 10.66)$.

A graphical representation of $\mathbf{F}_n$ is given in Figure 2, with $x = 45$ (the sample mean) and $x = 55$, respectively. In Tables 10 and 11, we collect the estimated $\mathbf{F}(t_1,t_2|x)$ for a fixed x and $(t_1,t_2) \in (1,2,3,4) \times (12,24,36,48,60)$ together with the 95% confidence intervals based on the asymptotic normality proved in Theorem 1, where the bootstrap technique has been used to estimate the standard deviation. In order to show the influence of the covariate X, we also present in Figure 2 the estimator $\mathbf{F}_n^K$ introduced by van Keilegom (2004) which does not take into account the covariate, X. Additionally, for different values of x, Figure 3 shows the estimator of $F_2(t|x)$ derived from $\mathbf{F}_n$ and the standard Beran estimator based on the reduced sample (that is, the individuals for which $\delta_1 = 1$).

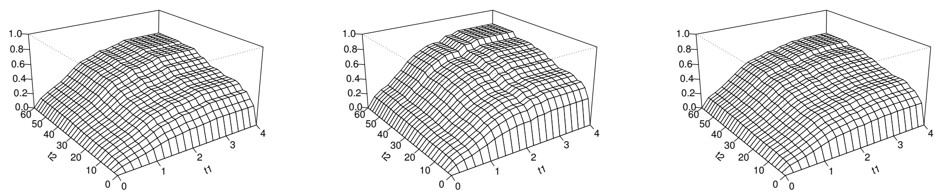

Figure 2. *Estimator of* $\mathbf{F}(t_1,t_2|x)$ *for* $x = 45$ *(left-hand panel) and* $x = 55$ *(middle panel) together with* $\mathbf{F}_n^K(t_1,t_2)$ *(right-hand panel).*

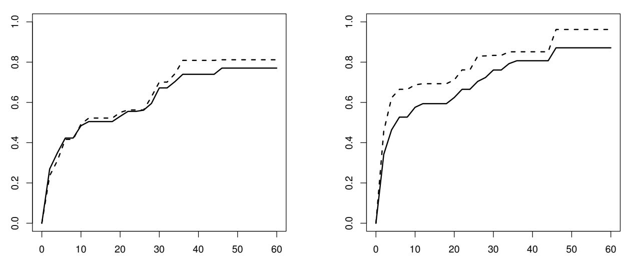

Figure 3. $F_{2n}^{\infty}(t|x)$ *(solid line) and standard Beran estimator on reduced sample (dahed line) where* $x = 45$ *(left-hand panel) and* $x = 55$ *(right-hand panel)*

Table 10. *Estimated* $\mathbf{F}(t_1,t_2|x)$ *for* $x = 45$.

t_1 \ t_2	12	24	36	48	60
1	0.2331	0.2548	0.3314	0.3405	0.3405
	(0.1160,0.3505)	(0.1304,0.3794)	(0.1994,0.4637)	(0.2112,0.4702)	(0.2112,0.4702)
2	0.3847	0.4211	0.5474	0.5628	0.5628
	(0.2516,0.5466)	(0.2847,0.5891)	(0.4302,0.7055)	(0.4592,0.7086)	(0.4592,0.7086)
3	0.4513	0.4944	0.6426	0.6610	0.6610
	(0.2999,0.6029)	(0.3394,0.6497)	(0.5175,0.7679)	(0.5573,0.7648)	(0.5573,0.7648)
4	0.4516	0.4948	0.6431	0.6614	0.6614
	(0.2998,0.6035)	(0.3393,0.6504)	(0.518,0.7683)	(0.5577,0.7653)	(0.5577,0.7653)

Table 11. *Estimated* $\mathbf{F}(t_1, t_2|x)$ *for* $x = 55$.

t_1 \ t_2	12	24	36	48	60
1	0.2837	0.3176	0.3776	0.4111	0.4111
	(0.1553,0.4127)	(0.1796,0.4563)	(0.2284,0.5275)	(0.2683,0.5548)	(0.2683,0.5548)
2	0.4705	0.5277	0.6247	0.6805	0.6805
	(0.3167,0.6414)	(0.3686,0.7057)	(0.4702,0.801)	(0.5541,0.8311)	(0.5541,0.8311)
3	0.5537	0.6215	0.7335	0.7994	0.7994
	(0.3894,0.7192)	(0.4534,0.7909)	(0.5776,0.8908)	(0.6871,0.9133)	(0.6871,0.9133)
4	0.5537	0.6215	0.7335	0.7994	0.7994
	(0.3894,0.7193)	(0.4533,0.791)	(0.5776,0.8908)	(0.6871,0.9133)	(0.6871,0.9133)

We can see, in Figure 2, that the bivariate distribution changes with the covariate, X. Figure 3 shows even more clearly that the age has a big influence on the distribution function. Specifically, the probability to fail increases as the age increases. This is a quite natural effect. Moreover, both from the figures and tables, we observe that the probabilities do not change for $t_2 \in [48, 60]$, which is equivalent to the 4th and 5th year after transplantation. Interestingly, based on Figure 3, the probability of death is growing rapidly in the first months after heart transplantation and stabilizes by the time of 4 years after the surgery. Finally, in Figure 3, we can see that when applying the standard Beran estimator on this reduced sample, we get similar results to our marginal estimator $F_{2n}^{\infty}(t|x)$ when $x = 45$ but this Beran estimator seems to overestimate the distribution when $x = 55$.

7.2. Colon cancer data

In this subsection, we analyze the colon cancer data, previously introduced in Section 1. We investigate the effectiveness of treatment with levamisole plus 5-FU versus placebo as well as influence of the age on the survival times. The data set is a part of the data set "colon" in the R survival library. Similarly as Lawless and Yilmaz (2011) we consider those patients treated with Levamisole plus 5-FU and placebo control. The full data set includes a third treatment group (Levamisole). There were 315 patients assigned to the placebo control group and 304 to the treatment group. By the end of the study, 177 patients (56%) in the placebo group had cancer recurrence, among whom 155 died, whereas in the treatment group 119 (39%) patients had cancer recurrence, among whom 108 died. The maximal observed times until recurrence and from recurrence until death are around 9 and 6 years, respectively. We consider age as a covariate. In Figures 4-6 we plot the bivariate and univariate estimators for both groups and 50, 60, and 70 years old patients, being the respective quartiles in the data set. Figure 7 shows the estimator, $F_{2n}^{\infty}(t|x)$, for the treatment group together with bootstrap-based and normal 95% point wise confidence intervals, respectively.

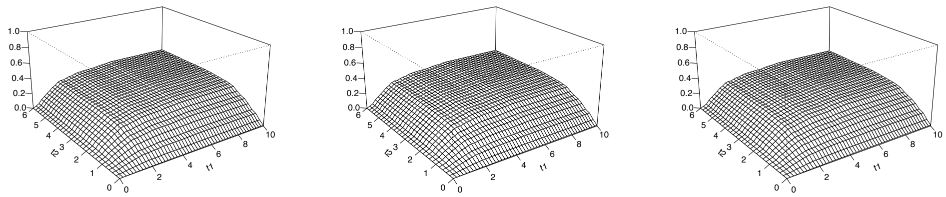

Figure 4. *Estimator of* $\mathbf{F}(t_1,t_2|x)$ *in treatment group for* $x = 50$ *(left-hand panel),* $x = 60$ *(middle panel) and* $x = 70$ *(right-hand panel).*

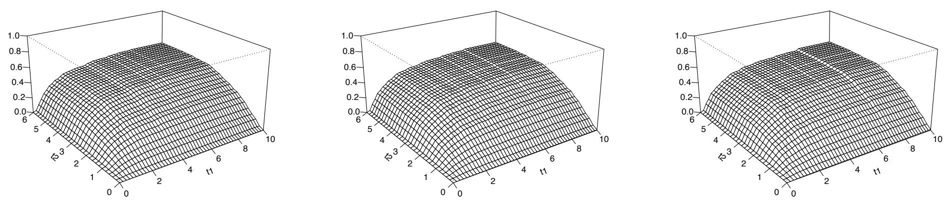

Figure 5. *Estimator of* $\mathbf{F}(t_1,t_2|x)$ *in placebo group for* $x = 50$ *(left-hand panel),* $x = 60$ *(middle panel) and* $x = 70$ *(right-hand panel).*

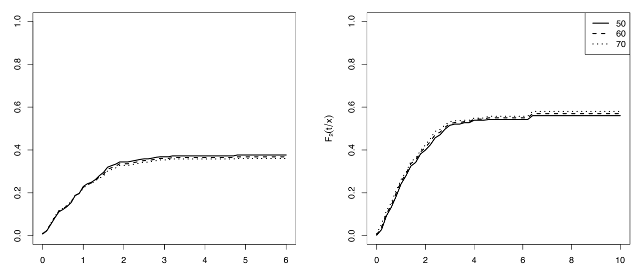

Figure 6. $F_{2n}^{\infty}(t|x)$ *for* $x = 50, 60, 70$ *for treatment group (left-hand panel) and placebo group (right-hand panel).*

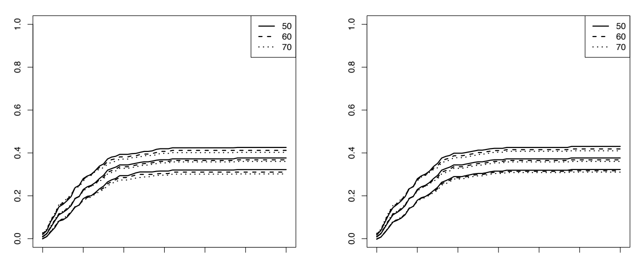

Figure 7. $F_{2n}^{\infty}(t|x)$ *for* $x = 50, 60, 70$ *for treatment group with bootstrap 95% CIs (left-hand panel) and normal 95% CIs (right-hand panel).*

As expected, treatment seems to have big influence on patients survival, increasing the survival time from recurrence. However, based on the confidence intervals for the treatment group, the influence of the age seems to be non significative although a formal test should be developed to confirm this statement.

8. Discussion and future work

In this paper we have introduced a new method to estimate the bivariate conditional distribution of two consecutive censored gap times and the corresponding marginal distributions. This new methodology is an adaptation and a mixture of the methods proposed by Beran (1981) and van Keilegom (2004). It is worth mentioning that a simpler method could be the extension of the Kaplan-Meier based estimator studied by de Uña-Álvarez and Meira-Machado (2008). Although our method is computationally more intensive compared to the Kaplan-Meier based estimator because it requires bootstrapping to estimate the variance, it has an important advantage. While the Kaplan-Meier based estimator requires the assumption of independence between (T_1, T_2) and C, our method only requires the independence given X (allowing some dependence between C and the explanatory variable X). Based on this weaker assumption, we have proved its asymptotic theoretical properties and studied its finite sample behaviour through a simulation study. Additionally, our approach can be extended to a d-dimensional explanatory variable X by using a single-index model (see Strzalkowska-Kominiak and Cao (2013)) avoiding the curse of dimensionality. However, this issue goes beyond the scope of this paper and will be the basis of our future research.

9. Acknowledgments

E.M. Molanes-López acknowledges support to Grant PID2019-106772RB-I00 from the Spanish Government. E. Strzalkowska-Kominiak acknowledges support to Grant PID2022-138114NB-I00 from the Spanish Government. We thank the editor and two referees for their valuable comments that improved our article.

References

Beran, R. (1981). Nonparametric regression with randomly censored survival data. *Technical Report*, University California, Berkeley.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). Copula methods in finance. New York: John Wiley & Sons.

de Uña-Álvarez, J. and Amorim, A.P. (2011). A semiparametric estimator of the bivariate distribution function for censored gap times. *Biometrical Journal* 53, 113–127.

de Uña-Álvarez, J. and Meira-Machado, L.F. (2008). A simple estimator of the bivariate distribution function for censored gap times. *Statistics and Probability Letters* 78, 2440–2445.

Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics* 33, 1380–1403.

Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data. *Biometrika* 88(4), 907–919.

Gijbels, I., Veraverbeke, N. and Omelka, M. (2011). Conditional copulas, association measures and their application. *Computational Statistics and Data Analysis* 55, 1919–1932.

González-Manteiga, W. and Cadarso-Suárez, C. (1994). Asymptotic properties of a generalized Kaplan-Meier estimator with some applications. *Journal of Nonparametric Statistics* 4, 65–78.

Huang, Y. and Louis, T. A. (1998). Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika* 85(4), 785–798

Huang, C.-Y., Luo, X. and Follmann, D.A. (2011). A model checking method for the proportional hazards model with recurrent gap time data. *Biostatistics* 12, 535–547.

Iglesias-Pérez, C. and González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *Journal of Nonparametric Statistics* 10, 213–244.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data, New York: John Wiley & Sons, Inc.

Lawless, J. F. and Yilmaz, Y. E. (2011). Semiparametric estimation in copula models for bivariate sequential survival times. *Biometrical journal* 53(5), 779–796.

Lin, D.Y., Sun, W. and Ying, Z. (1999). Nonparametric estimation of the gap time distribution for serial events with censored data. *Biometrika* 86, 59–70.

Meira-Machado, L. and Roca-Pardiñas, J. (2011). p3state.msm: Analyzing survival data from an illness-death model. *Journal of Statistical Software* 38, Issue 3.

Serrat, C. and Gómez, G. (2007). Nonparametric bivariate estimation for successive survival times. *SORT* 1, 75–96.

Sklar, A. (1959). Fonctions de répartion à n dimensions e leurs marges. *Publications de l'Institut de Statisque de l'Université de Paris* 8, 229–231.

Strzalkowska-Kominiak, E. and Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *Journal of Multivariate Analysis* 114, 74–98.

Strzalkowska-Kominiak, E. and Stute, W. (2010). The statistical analysis of consecutive survival data under serial dependence. *Journal of Nonparametric Statistics* 22, 585–597.

van Keilegom, I. (2004). A note on the nonparametric estimation of the bivariate distribution under dependent censoring. *Journal of Nonparametric Statistics* 16, 659-670.

Visser, M. (1996). Nonparametric estimation of the bivariate survival function with an application to vertically transmitted. *Biometrika* 83, 507–518.

Wang, W. and Wells, M.T. (1998). Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* 85, 561–572.

Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65(1), 257–273.
Wu, F., Valdez, E. and Sherris, M. (2007). Simulating from exchangeable Archimedean copulas. *Communications in Statistics – Simulation & Computation* 36, 1019–1034.
Zhao, X. and Zhou, X. (2010). Applying copula models to individual claim loss reserving methods. *Insurance: Mathematics and Economics* 46, 290–299.
Zhu, H. and Wang, M. C. (2012). Analysing bivariate survival data with interval sampling and application to cancer epidemiology. *Biometrika* 99(2), 345–361.

Appendix A: Asymptotic properties

Proof of Theorem 1. We have

$$\mathbf{F}_n(y_1,y_2|x)-\mathbf{F}(y_1,y_2|x)=A_n(y_1,y_2|x)+B_n(y_1,y_2|x)+C_n(y_1,y_2|x), \tag{A.1}$$

where

$$\begin{aligned}A_n(y_1,y_2|x)&=\int_0^{y_1}(F_{21n}(y_2|t_1,x)-F_{21}(y_2|t_1,x))F_1(dt_1|x)\\B_n(y_1,y_2|x)&=\int_0^{y_1}F_{21}(y_2|t_1,x)(F_{1n}(dt_1|x)-F_1(dt_1|x))\\C_n(y_1,y_2|x)&=\int_0^{y_1}(F_{21n}(y_2|t_1,x)-F_{21}(y_2|t_1,x))(F_{1n}(dt_1|x)-F_1(dt_1|x)).\end{aligned}$$

We now deal with the first term, $A_n(y_1,y_2|x)$, in the right hand side of (A.1). Since, $F_{21n}(y_2|t_1,x)$ is a Beran estimator on the restricted sample, we can use the results from González-Manteiga and Cadarso-Suárez (1994). For $(y_1,y_2)\in A(x)$ and $x\in\{u: f_X(u)>0\}$, we obtain

$$F_{21n}(y_2|t_1,x)-F_{21}(y_2|t_1,x)=\sum_{i=1}^n\tilde{B}_{in}(x,t_1)\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)+R_n(y_2|t_1,x),$$

where

$$\frac{\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)}{1-F_{21}(y_2|t_1,x)}=\left[-\int_0^{\tilde{T}_{2i}\wedge y_2}\frac{d\tilde{H}^*(s|t_1,x)}{(1-H^*(s^-|t_1,x))^2}+\frac{1_{\{\tilde{T}_{2i}\le y_2,\delta_{2i}=1\}}}{1-H^*(\tilde{T}_{2i}^-|t_1,x)}\right].$$

To deal with $R_n(y_2|t_1,x)$, we need the properties of the estimators

$$H_n^*(t|t_1,x)=\sum_{i=1}^n 1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}(t_1,x),$$

$$\tilde{H}_n^*(t|t_1,x)=\sum_{i=1}^n\delta_{2i}1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}(t_1,x),$$

where

$$\tilde{B}_{in}(t_1,x) = \frac{\frac{1}{nh_1h_2}\delta_{1i}K\left(\frac{x-X_i}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1i}}{h_2}\right)}{\frac{1}{nh_1h_2}\sum_{j=1}^{n}\delta_{1j}K\left(\frac{x-X_j}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1j}}{h_2}\right)}.$$

Remark, that

$$\tilde{B}_{in}(t_1,x) = \tilde{B}_{in}^1(t_1,x) + \tilde{B}_{in}^2(t_1,x), \tag{A.2}$$

where

$$\begin{aligned}
\tilde{B}_{in}^1(t_1,x) &= \frac{\frac{1}{nh_1h_2}\delta_{1i}K\left(\frac{x-X_i}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1i}}{h_2}\right)}{\tilde{h}^1(t_1,x)}\\
\tilde{B}_{in}^2(t_1,x) &= \frac{\tilde{B}_{in}(t_1,x)}{\tilde{h}^1(t_1,x)}\left[\tilde{h}^1(t_1,x) - \frac{1}{nh_1h_2}\sum_{j=1}^{n}\delta_{1j}K\left(\frac{x-X_j}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1j}}{h_2}\right)\right]
\end{aligned}$$

and $\tilde{h}^1(t_1,x) = \tilde{h}(t_1|x)f_X(x)$ with $\tilde{h}(t_1|x)$ denoting the density of $\tilde{H}(t_1|x) = \mathbb{P}(\tilde{T}_1 \le t_1, \delta_1 = 1|X = x)$. Hence

$$H_n^*(t|t_1,x) = \sum_{i=1}^{n} 1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}^1(t_1,x) + \sum_{i=1}^{n} 1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}^2(t_1,x).$$

Moreover, since $\frac{nh_1h_2}{\log(n)} \to \infty$, using Theorem 1 and 2 together with Remark 8 from Einmahl and Mason (2005) and the Taylor expansion, we obtain

$$\sup_{t\in\mathbb{R},t_1\le\tau_1(x),x\in I}\left|\sum_{i=1}^{n} 1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}^1(t_1,x) - H^*(t|t_1,x)\right| = O\left(\sqrt{\frac{\log(n)}{nh_1h_2}}\right) + O(h_1^2) + O(h_2^2)$$

almost surely (a.s.) and

$$\sup_{t\in\mathbb{R},t_1\le\tau_1(x),x\in I}\left|\sum_{i=1}^{n} 1_{\{\tilde{T}_{2i}\le t\}}\tilde{B}_{in}^2(t_1,x)\right| = O\left(\left(\frac{\log(n)}{nh_1h_2}\right)^{1/2}\right) + O(h_1^2) + O(h_2^2), \quad \text{a.s.}$$

where $I = \{u : f_X(u) > 0\}$. Similarly, we deal with $\tilde{H}_n^*(t|t_1,x)$. Finally, following the steps of the proof of Theorem 2.3 in González-Manteiga and Cadarso-Suárez (1994), and since $H_n^*(t|t_1,x) \to H^*(t|t_1,x)$ in probability, we can show that

$$\sup_{(t_1,y_2)\in A(x),x\in I}|R_n(y_2|t_1,x)| = O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_1h_2}\right)^{3/4}\right) + O_{\mathbb{P}}(h_1^2) + O_{\mathbb{P}}(h_2^2).$$

Since $nh_1^5 \to 0$, $nh_1^3 \to \infty$ and $nh_2^5 \to c > 0$, uniformly in $(y_1,y_2) \in A(x)$ and $x \in \{u : f_X(u) > 0\}$, we obtain

$$\int_0^{y_1} R_n(y_2|t_1,x)F_1(dt_1|x) = o_{\mathbb{P}}((nh_1)^{-1/2}).$$

Hence, on the set $(y_1, y_2) \in A(x)$ and $x \in \{u : f_X(u) > 0\}$, we have

$$A_n(y_1,y_2|x) = \sum_{i=1}^n \int_0^{y_1} \tilde{B}_{in}(t_1,x)\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)F_1(dt_1|x) + o_{\mathbb{P}}((nh_1)^{-1/2}), \quad \text{(A.3)}$$

where $\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)$ are i.i.d. and $E(\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)|\tilde{T}_{1i} = t_1, \delta_{1i} = 1, X_i = x) = 0$.

We now deal with the second term, $B_n(y_1,y_2|x)$, in the right hand side of (A.1). In order to deal with $B_n(y_1,y_2|x)$, we need to define

$$\tilde{H}(t_1|x) = \mathbb{P}(\tilde{T}_1 \le t_1, \delta_1 = 1|X = x)$$

and its estimator

$$\tilde{H}_n(t_1|x) = \sum_{i=1}^n B_{in}(x) 1_{\{\tilde{T}_{1i} \le t_1\}} \delta_{1i}.$$

Additionally, set

$$G_n(t|x) = 1 - \prod_{i=1}^n \left[1 - \frac{B_{in}(x) 1_{\{\tilde{T}_{1i} \le t\}}(1-\delta_{1i})}{\sum_{j=1}^n 1_{\{\tilde{T}_{1j} \ge \tilde{T}_{1i}\}} B_{jn}(x)} \right].$$

Then, if there are no ties and since $\sum_{i=1}^n B_{in}(x) = 1$, we have that

$$\Delta F_{1n}(\tilde{T}_{1i}|x) = F_{1n}(\tilde{T}_{1i}|x) - F_{1n}(\tilde{T}_{1i}^-|x) = \frac{B_{in}(x)\delta_{1i}}{1 - G_n(\tilde{T}_{1i}^-|x)}.$$

Moreover, for every function $\phi(t_1,x)$ and under A1, we have

$$\int \phi(t_1,x)F_1(dt_1|x) = \int \frac{\phi(t_1,x)}{1 - G(t_1^-|x)} \tilde{H}(dt_1|x). \quad \text{(A.4)}$$

Finally,

$$B_n(y_1,y_2|x) = \int_0^{y_1} F_{21}(y_2|t_1,x) \frac{\tilde{H}_n(dt_1|x)}{1 - G_n(t_1^-|x)} - \int_0^{y_1} F_{21}(y_2|t_1,x) \frac{\tilde{H}(dt_1|x)}{1 - G(t_1^-|x)}.$$

Hence

$$B_n(y_1,y_2|x) = B_n^1(y_1,y_2|x) + B_n^2(y_1,y_2|x) + B_n^3(y_1,y_2|x) + B_n^4(y_1,y_2|x),$$

where

$$B_n^1(y_1,y_2|x) = \int_0^{y_1} \frac{F_{21}(y_2|t_1,x)}{1 - G(t_1^-|x)} (\tilde{H}_n(dt_1|x) - \tilde{H}(dt_1|x))$$

$$B_n^2(y_1,y_2|x) = \int_0^{y_1} \frac{F_{21}(y_2|t_1,x)}{(1 - G(t_1^-|x))^2} [G_n(t_1^-|x) - G(t_1^-|x)] \tilde{H}(dt_1|x)$$

$$B_n^3(y_1,y_2|x) = \int_0^{y_1} \frac{F_{21}(y_2|t_1,x)}{(1 - G(t_1^-|x))^2} [G_n(t_1^-|x) - G(t_1^-|x)] (\tilde{H}_n(dt_1|x) - \tilde{H}(dt_1|x))$$

$$B_n^4(y_1,y_2|x) = \int_0^{y_1} \frac{F_{21}(y_2|t_1,x)[G_n(t_1^-|x) - G(t_1^-|x)]^2}{(1 - G(t_1^-|x))^2(1 - G_n(t_1^-|x))} \tilde{H}_n(dt_1|x).$$

As to $B_n^1(y_1,y_2|x)$, taking into account that $\sum_{i=1}^n B_{in}(x)=1$, we can write it as follows

$$B_n^1(y_1,y_2|x)=\sum_{i=1}^n B_{in}(x)\left(\frac{F_{21}(y_2|\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}^-|x)}\delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}-\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{1-G(t_1^-|x)}\tilde{H}(dt_1|x)\right).$$

Hence

$$B_n^1(y_1,y_2|x)=\sum_{i=1}^n B_{in}(x)\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x), \tag{A.5}$$

where

$$\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x)=\frac{F_{21}(y_2|\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}^-|x)}\delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}-\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{1-G(t_1^-|x)}\tilde{H}(dt_1|x)$$

are i.i.d. and $E(\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x)|X_i=x)=0$.
Furthermore, we use again the results from González-Manteiga and Cadarso-Suárez (1994) for the Beran estimator $G_n(t|x)$. Consequently, for $t_1\le\tau_1(x)$, we obtain

$$G_n(t_1^-|x)-G(t_1^-|x)=\sum_{i=1}^n B_{in}(x)\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)+R_{1n}(t_1,x),$$

where

$$\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)=(1-G(t_1^-|x))\left[-\int_0^{\tilde{T}_{1i}\wedge t_1}\frac{d\tilde{\tilde{H}}(s|x)}{(1-H(s^-|x))^2}+\frac{1_{\{\tilde{T}_{1i}\le t_1,\delta_{1i}=0\}}}{1-H(\tilde{T}_{1i}^-|x)}\right],$$

$$\tilde{\tilde{H}}(s|x)=\mathbb{P}(\tilde{T}_1\le s,\delta_1=0|X=x)$$

and

$$\sup_{t_1\le\tau_1(x),x\in I}|R_{1n}(t_1,x)|=O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_1}\right)^{3/4}\right)+O_{\mathbb{P}}(h_1^2)$$

Hence, on the set $(y_1,y_2)\in A(x)$ and for $x\in I$, we have

$$\begin{aligned}B_n^2(y_1,y_2|x)&=\sum_{i=1}^n\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{(1-G(t_1^-|x))^2}B_{in}(x)\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)\tilde{H}(dt_1|x)\\&\quad+O_{\mathbb{P}}\left(\left(\frac{\log(n)}{nh_1}\right)^{3/4}\right)+O_{\mathbb{P}}(h_1^2).\end{aligned}$$

Consequently, since $nh_1^5\to0$ and $nh_1^3\to\infty$, we obtain

$$\begin{aligned}\sqrt{nh_1}B_n^2(y_1,y_2|x)&=\sqrt{nh_1}\sum_{i=1}^n\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{(1-G(t_1^-|x))^2}B_{in}(x)\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)\tilde{H}(dt_1|x)\\&\quad+o_{\mathbb{P}}(1),\end{aligned} \tag{A.6}$$

where $\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)$ are i.i.d. and $E(\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)|X_i=x)=0$.
Hence, it is easy to show that

$$\sqrt{nh_1}(B_n^3(y_1,y_2|x)+B_n^4(y_1,y_2|x))=o_{\mathbb{P}}(1). \tag{A.7}$$

Regarding the third term, $C_n(y_1,y_2|x)$, in the right hand side of (A.1), we have that

$$\sqrt{nh_1}C_n(y_1,y_2|x)=o_{\mathbb{P}}(1). \tag{A.8}$$

Finally, from (A.3)-(A.8), we obtain

$$\begin{aligned}\sqrt{nh_1}(\mathbf{F}_n(y_1,y_2|x)-\mathbf{F}(y_1,y_2|x)) &= \sqrt{nh_1}\sum_{i=1}^n\int_0^{y_1}\tilde{B}_{in}(x,t_1)\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)F_1(dt_1|x)\\ &+\sqrt{nh_1}\sum_{i=1}^n B_{in}(x)\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x)\\ &+\sqrt{nh_1}\sum_{i=1}^n\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{(1-G(t_1^-|x))^2}B_{in}(x)\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)\tilde{H}(dt_1|x)+o_{\mathbb{P}}(1).\end{aligned} \tag{A.9}$$

The Equation (A.9) is not yet the i.i.d. representation which we aim at. To deal with its first term, using (A.2), we obtain

$$\sqrt{nh_1}\sum_{i=1}^n\int_0^{y_1}\tilde{B}_{in}(x,t_1)\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)F_1(dt_1|x)=T_{1n}+T_{2n},$$

where

$$T_{1n}=\frac{1}{\sqrt{nh_1}h_2}\sum_{i=1}^n\int_0^{y_1}\frac{\delta_{1i}K\left(\frac{x-X_i}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1i}}{h_2}\right)}{\tilde{h}^1(t_1,x)}\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)F_1(dt_1|x)$$

and

$$\begin{aligned}T_{2n}=\sqrt{nh_1}\sum_{i=1}^n\int_0^{y_1}\frac{\tilde{B}_{in}(t_1,x)}{\tilde{h}^1(t_1,x)}&\left[\tilde{h}^1(t_1,x)-\frac{1}{nh_1h_2}\sum_{j=1}^n\delta_{1j}K\left(\frac{x-X_j}{h_1}\right)K\left(\frac{t_1-\tilde{T}_{1j}}{h_2}\right)\right]\\ &\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)F_1(dt_1|x).\end{aligned}$$

As to T_{2n}, recall that $E(\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)|\tilde{T}_{1i}=t_1,\delta_{1i}=1,X_i=x)=0$. Moreover, using (A.2) again, it is easy to show that $T_{2n}=o_{\mathbb{P}}(1)$. As to T_{1n}, by (A.4), change of variables and Taylor expansion, we obtain

$$T_{1n}=\frac{1}{\sqrt{nh_1}}\frac{1}{f_X(x)}\sum_{i=1}^n\delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}K\left(\frac{x-X_i}{h_1}\right)\frac{\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}^-|x)}+O_{\mathbb{P}}(n^{1/2}h_1^{1/2}h_2^2).$$

The properties of the second and third term in (A.9) based on

$$B_{in}(x)=\frac{\frac{1}{nh_1}K\left(\frac{x-X_i}{h_1}\right)}{f_X(x)}+\frac{B_{in}(x)}{f_X(x)}\left[f_X(x)-\frac{1}{nh_1}\sum_{j=1}^nK\left(\frac{x-X_j}{h_1}\right)\right].$$

Finally,

$$\sqrt{nh_1}(\mathbf{F}_n(y_1,y_2|x)-\mathbf{F}(y_1,y_2|x)) = \frac{1}{f_X(x)}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_{ni}(y_1,y_2,x)+o_{\mathbb{P}}(1), \quad \text{(A.10)}$$

where

$$\begin{aligned}\Phi_{ni}(y_1,y_2,x) = & \frac{1}{\sqrt{h_1}}\delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}K\left(\frac{x-X_i}{h_1}\right)\frac{\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}-|x)} \\ & +\frac{1}{\sqrt{h_1}}K\left(\frac{x-X_i}{h_1}\right)\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x) \\ & +\frac{1}{\sqrt{h_1}}\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{(1-G(t_1^-|x))^2}K\left(\frac{x-X_i}{h_1}\right)\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)\tilde{H}(dt_1|x),\end{aligned}$$

$$\frac{\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,t_1,x)}{1-F_{21}(y_2|t_1,x)} = \left[-\int_0^{\tilde{T}_{2i}\wedge y_2}\frac{d\tilde{H}^*(s|t_1,x)}{(1-H^*(s^-|t_1,x))^2}+\frac{1_{\{\tilde{T}_{2i}\le y_2,\delta_{2i}=1\}}}{1-H^*(\tilde{T}_{2i}^-|t_1,x)}\right],$$

$$\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x) = \frac{F_{21}(y_2|\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}^-|x)}\delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}-\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{1-G(t_1^-|x)}\tilde{H}(dt_1|x)$$

and

$$\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x) = (1-G(t_1^-|x))\left[-\int_0^{\tilde{T}_{1i}\wedge t_1}\frac{d\tilde{\tilde{H}}(s|x)}{(1-H(s^-|x))^2}+\frac{1_{\{\tilde{T}_{1i}\le t_1,\delta_{1i}=0\}}}{1-H(\tilde{T}_{1i}^-|x)}\right]$$

Since, for $i=1,..,n$, $\Phi_{ni}(y_1,y_2,x)$ are i.i.d. and $E\Phi_{ni}(y_1,y_2,x)=O(h_1^{5/2})$, the right hand side of (A.10) is a sum of i.i.d. random variables plus remainder of order $o_{\mathbb{P}}(1)$. Hence we obtain

$$\sqrt{nh_1}(\mathbf{F}_n(y_1,y_2|x)-\mathbf{F}(y_1,y_2|x)) \to \mathcal{N}(0,\sigma_1^2(y_1,y_2|x)), \quad \text{(A.11)}$$

where

$$\sigma_1^2(y_1,y_2|x) = \frac{\int K^2(t)dt}{f_X(x)}\rho_1^2(y_1,y_2,x),$$

and

$$\begin{aligned}\rho_1^2(y_1,y_2,x) = Var\Bigg(& \delta_{1i}1_{\{\tilde{T}_{1i}\le y_1\}}\frac{\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,\tilde{T}_{1i},x)}{1-G(\tilde{T}_{1i}^-|x)}+\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x) \\ & +\int_0^{y_1}\frac{F_{21}(y_2|t_1,x)}{(1-G(t_1^-|x))^2}\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x)\tilde{H}(dt_1|x)|X_i=x\Bigg) \quad \text{(A.12)}\end{aligned}$$

and the proof is completed.

Remark that, if there is no censoring, $C\equiv\infty$, it is easy to show that

$$\xi_1(\tilde{T}_{2i},\delta_{2i},y_2,\tilde{T}_{1i},x) = 1_{\{T_{2i}\le y_2\}}-F_{21}(y_2|\tilde{T}_{1i},x),$$

$$\xi_2(\tilde{T}_{1i},\delta_{1i},y_1,y_2,x) = F_{21}(y_2|\tilde{T}_{1i},x)1_{\{T_{1i}\le y_1\}}-\mathbf{F}(y_1,y_2|x)$$

and

$$\xi_3(\tilde{T}_{1i},\delta_{1i},t_1,x) = 0.$$

Hence $\rho_1^2(y_1,y_2,x)=\mathbf{F}(y_1,y_2|x)(1-\mathbf{F}(y_1,y_2|x))$ as desired.